Timing Analysis and Optimization Techniques for High Performance Integrated Circuits

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Electrical Engineering and Computer Science

By

Ahmed Shebaita

EVANSTON, ILLINOIS

June 2008
ABSTRACT

Timing Analysis and Optimization Techniques for High Performance Integrated Circuits

Ahmed Shebaita

As manufacturing technology moves toward fundamental limits of silicon CMOS processing, it is increasingly important to utilize the full potential of available transistors and interconnects. While manufacturing technology faces fundamental limits inherent in physical laws or material properties, design and verification technology faces fundamental limitations inherent in the computational intractability of design optimizations and in the broad and unknown range of potential applications within various design processes. In this research, we explore how design and verification technology can enable the implementation of single-chip microelectronic systems that take full advantage of manufacturing technology with respect to such criteria as complexity, performance, and power dissipation. One limitation is that the integrated circuit (IC) analysis and verification process involves practical tradeoffs among multiple objectives. For example, there is a need to apply a correct full model that is computationally very efficient in terms of both run time and resource usage. A second limitation is in the integrated circuit design process which again involves practical tradeoffs among multiple objectives such as performance, power dissipation, reliability, etc.
In today’s highly competitive environment, even small differences in the quality of one process versus another can be the difference between success and failure, and a major improvement in quality can lead to an entirely new generation of commercial tools and services.

This research handles the verification, modeling and optimization problems of current technologies.

For analysis and verification purposes, new techniques to handle both the inductance effects and the nonlinear gate behavior effects are presented. A new technique of time shifted moment matching (TSMM) performs moment matching (for expansion around \( s=0 \)) on a time-shifted version of the original signal. Existing model order reduction techniques were highly improved by applying the time shifted moment’s concept. Including inductance in static timing analysis for better delay and rise time estimations become possible by both introducing a generalized driving point admittance which takes the inductive shielding into consideration and introducing a more general waveform shape that accounts for the nonmonotonic behavior due to inductance effects. To account for the nonlinear behavior of the gate in static timing analysis, while preserving the minimum complexity and utmost flexibility, a novel representation of the gate is introduced which linearizes the gate around a given load. Hence, moment propagation and uniform treatment of the gates and interconnect becomes possible for the first time.

For modeling purposes, this research handles one of the most critical phenomena that appear in the multi GHZ chips, namely skin effect. Figures of merit to determine when skin effect becomes important and require volume discretization modeling in the GHZ range is presented. This derived figure of merit is shown to depend solely on the interconnect dimensions and spacing
and is independent of the type of materials used or technology scaling. The model has put a much tighter bound on which interconnect requires to have the volume discretized model than the existing models.

Finally, a novel optimization technique that minimizes a cost function of both the propagation delay and power consumption of CMOS tapered buffers is presented. A slight increase in the threshold voltage is shown to have an exponential effect in reducing the total power dissipation. The corresponding increase in the propagation delay is compensated for by increasing the number of buffer stages such that there is still an overall significant improvement in both the propagation delay and the total power dissipation.
ACKNOWLEDGEMENTS

The completion of my Ph.D. was only possible with the support and help from many precious people around me. First, I am extremely grateful and indebted to my advisor, Prof. Yehea Ismail. I deeply thank him for having his trust in me, and accepting me as his Ph.D. student. Throughout the course of my Ph.D. studies, he always has been very encouraging, patient, and has given me valuable guidance and directions not only as my Ph.D. advisor, but also as a mentor of life. I hope to continue the relationship with Prof. Ismail even after my graduation, and I also wish him all the best for his career and life.

I would like to express my gratitude to Dr. Dusan Petranovic for his valuable guidance and support during my studies. He always has been very encouraging and supportive to me all the way starting from being my internship mentor to being on my Ph.D. committee. I would like also to thank several EECS faculty members, especially, Prof. Hai Zhou, Prof. Seda Ogrenci Memik, Prof. Robert Dick, and Prof. Gokhan Memik for their advice, help, and support during my studies, and also for being on my Ph.D. committee. I would like to express my gratitude also I would like to thank DiaaeIdin Khalil and Frank Huebbers for being wonderful lab mates, and for making the working environment very enjoyable. I also want to thank some of those who have already graduated from this lab, Dr. Chirayu Amin, Dr. Maged Ghoneima, and Dr. Ja-Chun Ku for their helpful advice and support whenever I needed them. I sincerely wish to keep the relationship with these people as well. Moreover, I would like to express my gratitude to Semiconductor Research Corporation for the funding that supported my Ph.D. studies.
I would like to thank my wife and kids for the unlimited support and help they offered me. I
would also like to deeply thank my parents and my brothers for their unmeasurable amount of
love, trust, and support. Without them, I would have not been able to carry myself up to this
point. Above all, I thank God for his amazing plans and unconditional love for me, and providing
all these precious relationships around me for his glory and purpose.
CONTENTS

ABSTRACT .................................................................................................................................3

ACKNOWLEDGEMENTS .........................................................................................................6

LIST OF FIGURES ..................................................................................................................11

LIST OF TABLES ....................................................................................................................15

CHAPTER 1 INTRODUCTION ................................................................................................17

Chapter 2 EXPANDING THE FREQUENCY RANGE OF AWE VIA TIME SHIFTING .................................................................................................................................23

2.1 Introduction ........................................................................................................................23

2.2 Time Shifted Moment Matching .......................................................................................26

2.2.1 Calculating the Inertial Delay In Terms of the Circuit Moments ................................27

2.2.2 Calculating the Time Shifted Moments and Time Shifted Response .......................28

2.2.3 Calculating the Original Response ................................................................................31

2.3 Simulation Results ............................................................................................................33

2.4 CONCLUSIONS .................................................................................................................38

CHAPTER 3 INCLUDING INDUCTANCE IN STATIC TIMING ANALYSIS .......................39

3.1 Introduction and Background ...........................................................................................40

3.2 Driving Point Admittance and Characteristic Impedance Calculations .......................43

3.3 Estimating the Gate Delays and Effective Capacitances For an Interconnect Experiencing Late Oscillations .......................................................................................................49
3.4 Estimating the Gate Delay and Effective Capacitances for An Interconnect Experiencing Early Oscillation.................................................................56
3.5 Conclusions.................................................................................................64

CHAPTER 4 A NOVEL MOMENT-BASED METHODOLOGY FOR ACCURATE AND EFFICIENT STATIC TIMING ANALYSIS..............................................66
4.1 Introduction .................................................................................................66
4.2 Load-Based Linear Gate Model.................................................................70
4.3 Gate Characterization....................................................................................74
4.4 Path Delay Propagation...............................................................................78
4.5 Conclusions..................................................................................................84

CHAPTER 5 IMPORTANCE OF VOLUME DESCRETIZATION OF SINGLE AND COUPLED INTERCONNECTS..........................................................86
5.1 Background.................................................................................................86
5.2 Qualitative Behavior of Interconnect Impedance under Skin Effect..........89
5.3 Importance of Volume Discretization for Single Wire...............................92
5.3.1 Volume Discretization Figure of Merit for Single Wire..........................93
5.3.2 Delay Error in Using the Interconnect DC Model Versus the 3D model...96
5.3.3 Error in Impedance Due to Using DC Model...........................................98
5.3.4 Effect of Scaling the Interconnect Dimensions.......................................101
5.3.5 Effect of Interconnect Aspect Ratio.......................................................104
5.3.6 Impact of Physical Constant.................................................................106
5.4 Importance of Volume Discretization for Coupled Wire.........................107
5.4.1 Impact of Inductive Coupling ................................................................. 107
5.4.2 Volume Discretization Figure of Merit in Case of Loop Inductance (case 3) ................................................................................................................................. 110
5.4.3 Volume Discretization Figure of Merit in Case of Loop Inductance (case 3) ................................................................................................................................. 113
5.5 Error Formulation .............................................................................................. 115
5.6 Conclusions ........................................................................................................ 117

Chapter 6 VARIABLE THRESHOLD VOLTAGE DESIGN SCHEME FOR CMOS TAPPERED BUFFERS ................................................................................................................................. 119
6.1 Introduction ......................................................................................................... 119
6.2 Background .......................................................................................................... 121
6.3 Constant Vth Tapered Buffer Design ................................................................. 124
6.4 Variable Vth Tapered Buffer Design ..................................................................... 128
6.5 Variable Vth Tapered Buffer Design Scheme Effects on Other Performance Criteria ................................................................................................................................. 134
6.6 Conclusions ......................................................................................................... 135

CHAPTER 7 CONCLUSION ........................................................................................... 136

CHAPTER 8 FUTURE WORK ....................................................................................... 138

BIBLIOGRAPHY .......................................................................................................... 144

PUBLICATIONS ........................................................................................................... 150
LIST OF FIGURES

Figure 1-1. Signal behavior in using the RC model versus the RLC model..........................18
Figure 1-2 Skin effect resistance .........................................................................................19
Figure 1-3 Skin effect inductance .........................................................................................19
Figure 1-4 Normalized power as a function of delay penalty for various technology nodes.................................................................................................................................20
Figure 2-1. Highly inductive signals have a large inertial delay. Their time shifted versions have less high frequency components than the original one.................................................................25
Figure 2-2 A simple example to illustrate skin effect............................................................28
Figure 2-3 The effect of zero order approximation of x(t) from t=0 to t=t0 results in a discontinuity at t=t0........................................................................................................................................32
Figure 2-4 TSMM performs better than AWE for circuits with moderate inductive effects. (circuit #2, q=5)................................................................................................................................................35
Figure 2-5 For circuits with high inductance, AWE misses the signal whereas TSMM captures it very well (circuit #4, q=9)................................................................................................................................................36
Figure 2-6. AWE misses the signal here from a rise time perspective. TSMM captures the signal accurately (circuit #4, q=3)................................................................................................................................................36
Figure 2-7. AWE misses the signal here from a rise time perspective. TSMM captures the signal accurately (circuit #4, q=3)................................................................................................................................................37
Figure 2-8. This signal shows that TSMM performs much better than AWE for the same order. (q=6)..................................................................................................................................................37
Figure 3-1. The gate output voltage waveform of a CMOS gate driving an interconnect with
(a) \( R_d > R_0 \) and (b) \( R_d < R_0 \) for different interconnect models, RLC, RC, and Ct
..........................................................................................................................................................41

Figure 3-2. The general rules in calculating the admittance coefficients at the parent node
knowing those of the leaves ..................................................................................................................43

Figure 3-3. Upstream traversal for (a) lumped and (b) series circuits ..............................................44

Figure 3-4. Distributed RLC interconnect ..........................................................................................45

Figure 3-5. The equivalent circuit of a CMOS gate driving an RLC interconnect for the period of
time \( 0 < t < t_s + 2T_0 \) ...................................................................................................................46

Figure 3-6. Late oscillation versus early oscillation of an RLC interconnect .................................48

Figure 3-7. An effective capacitance that captures the effects of a \( \pi \)-model ..............................50

Figure 3-8. The voltage waveform at the output of a CMOS gate driving an RLC interconnect
for an initially injected voltage (a) higher than 0.8VDD and (b) less than 0.8VDD
..........................................................................................................................................................51

Figure 3-9. The voltage waveform at the output of a CMOS gate driving an RLC interconnect
experiencing early oscillation ............................................................................................................57

Figure 3-10. An effective capacitance that captures the effects of an RLC interconnect
..........................................................................................................................................................60

Figure 4-1. Gate current source model driving a load interconnect .................................................70

Figure 4-2. Gate output waveform using a gate model with a single set of precomputed linear
model coefficients ...............................................................................................................................71

Figure 4-3. Gate output of piecewise linear gate approximation .....................................................73
Figure 4-4. Gate output waveform for arbitrary input signal ........................................74
Figure 4-5. The setup used in the moment characterization...........................................74
Figure 4-6. PWL approximation of the SPICE produced waveform of $v_o(t)$ ..................76
Figure 4-7. The setup used in characterizing for gate input capacitance.........................77
Figure 4-8. Gate output waveform for various number of moments used in their approximation.................................................................79
Figure 4-9. The voltage waveform at the end of a single stage for various tree interconnects........................................................................................................80
Figure 4-10. Waveform at the end of the 5-stages paths with various loads................................................................................................................................83
Figure 5-1. Rectangular interconnect 3D model............................................................87
Figure 5-2 Effective cross sectional area of a rectangular interconnect under skin effect........................................................................................................................................88
Figure 5-3 Resistance and inductance behavior with frequency........................................90
Figure 5-4. The behavior of $R(f)$ and $\omega L(f)$ up to $f=100$ GHZ..............................................91
Figure 5-5. The two frequency points $\omega_1$ and $\omega_2$ that characterize skin effect..............92
Figure 5-6. The minimum $\frac{l}{w+t}$ that results in no skin effect........................................95
Figure 5-7 Delay experimental setup................................................................................96
Figure 5-8: Delay error between DC model and 3D model for different $\frac{l}{w+t}$ ..................98
Figure 5-9. Complete 3D Model Versus the DC Model.....................................................99
Figure 5-10. Percentage errors in impedance magnitude due to using the DC model.............100
Figure 5-11: Percentage errors in phase due to using the DC model..................................101
Figure 5-12 The change in $\omega_1$ and $\omega_2$ with.............................................................102
Figure 5-13. The behavior of $\omega_1$ and $\omega_2$ for Different Interconnect Widths..................103
Figure 5-14. Effect of increasing the aspect ratio on skin effect.............................................105
Figure 5-15 Interconnects with Proximity Effects.................................................................108
Figure 5-16 Plot of $g(\alpha, \beta)$..................................................................................111
Figure 5-17 Delay Setup Experiment.....................................................................................113
Figure 5-18 Delay error between DC model and 3D model for different wire dimensions.....114
Figure 6-1. Split-capacitor model of tapered buffer.............................................................122
Figure 6-2. $h(N)$ for various $C_L/C_g$.............................................................................125
Figure 6-3. Normalized total power as function of $C_L/C_g$................................................126
Figure 6-4. Normalized power delay square product.........................................................127
Figure 6-5. The behavior of $g(v_{th})$ with varying $v_{th}$......................................................129
Figure 6-6. The behavior of $s(v_{th})$ with varying $v_{th}$......................................................130
Figure 6-7. The behavior of $l(v_{th})$ with varying $v_{th}$......................................................130
Figure 6-8. Normalized power delay squared product ......................................................131
Figure 8-1. Stage Delay vs. Interconnect Width.................................................................139
Figure 8-2. Power grid circuit elements representation.....................................................141
## LIST OF TABLES

Table 2-1: Circuits used for Testing………………………………………………………………..33
Table 2-2: Comparison of AWE using time shifted moment matching (TSMM) with traditional AWE……………………………………………………………………………………………33
Table 3-1. Delay errors at the gate output for an RLC line model having an initial injected voltage less than 0.8VDD versus the RC model as compared to SPICE ........................................................................................................................................52
Table 3-2. Delay errors at the gate output for RLC tree interconnect having an initial injected voltage less than 0.8VDD versus the RC model as compared to SPICE ........................................................................................................................................53
Table 3-3. Delay errors at the gate output for RLC line interconnect model having an initial injected voltage higher than 0.8VDD versus the RC model as compared to SPICE ........................................................................................................................................54
Table 3-4. Delay errors at the gate output for RLC tree interconnect model having an initial injected voltage higher than 0.8VDD versus the RC model as compared to SPICE ........................................................................................................................................56
Table 3-5. Delay errors at the gate output for RLC line interconnect model having an initial injected voltage less than 0.5VDD versus the RC model as compared to SPICE ........................................................................................................................................63
Table 3-6. Delay errors at the gate output for RLC tree interconnect model having an initial injected voltage less than 0.5VDD versus the RC model as compared to SPICE ........................................................................................................................................64
Table 4-1. Numerical results……………………………………………………………………………………………………84
Table 5-1. Maximum error in impedance magnitude and phase (t=w)………………………………101
Table 5-2. Maximum errors due to using the DC model for different aspect ratios, w=1
\( \mu \) and \( \mu n \)…………………………………………………………………………………………………………………………104
Table 5-3 Effect of changing \( \sigma \) and \( \mu \) on the impedance error……………………………………….106
Table 5-4. The values of \( \frac{d}{l} \) above which the 3D model is not needed…………………………………111
Table 6-1. Effects of using \( N=N_{opt} \) & \( N=N_{opt} + 1 \) on power and delay…………………………127
Table 6-2. Effects of varying \( v_{th} \) on delay and power…………………………………………………………132
CHAPTER 1

INTRODUCTION

A lot of challenges are posed by deep-submicron technologies. Although deep submicron microelectronic technologies enable greater degrees of semiconductor integration, such integration makes the design, verification, and test challenges more difficult. Designers try to handle the design and verification complexity through models and methodologies to raise the abstraction level, but process and device challenges still take their toll on the overall design and test. Consequently, microelectronic circuit design and test are often the first areas affected by aggressive scaling in deep-submicron technologies. Such effects are mainly the increased frequency effects, which led to the introduction of new effects such as inductance and skin effects, the increased nonlinear behavior of the CMOS gates, and the significant increase in the power density which puts an upper limit on any further increase in the speed.

On-chip inductance has currently become more important with faster on-chip rise times and wider wires. Wide wires are frequently encountered in clock distribution networks and in upper metal layers. These wires are low resistance lines that can exhibit significant inductive effects. Furthermore, performance requirements are pushing the introduction of new materials such as copper interconnect for low resistance interconnect and new dielectrics to reduce the interconnect capacitance. These technological advances increase the importance of inductance.

In the limiting case, the advent of high critical-temperature Superconductors has created the possibility of realizing high density, extremely high-speed interconnects for integrated circuits. These superconductive interconnects have zero dc resistance.
and are highly sensitive to inductance. Ignoring on-chip inductance can cause significant errors in current deep submicron technologies. Figure 1-1 shows the error in signal behavior when modeling an interconnect with an RC versus the complete RLC model.

![Graph showing signal behavior comparison between RC and RLC models](image)

**Figure 1-1. Signal behavior in using the RC model versus the RLC model**

Thus, it is becoming more and more important to include inductance effects in circuit simulations at the least possible computational overhead.

The increasing inductance effects together with other nonlinear effects led to the fact that modeling the gate as a black box that outputs a saturated ramp waveform as a response to a saturated ramp input, increasingly deviates from reality as VLSI process technology evolves. Another sophisticated models such as the current source models suffer from the increased complexity and running time. New models that better represent the gate while at the same time preserve the minimum complexity feature becomes crucial in timing analysis.
Another consequence of moving to the multi-gigahertz range is the introduction of skin and proximity effects. With skin effect, currents tend to crowd toward the surface of conductors at high frequencies. Because the effective area where current flows decreases with frequency, the wire resistance increases with frequency as shown in Figure 1-2. Similarly, skin effect inductance decreases with frequency since the effective geometrical mean distance associated with the current distribution increases with frequency as shown in Figure 1-3. Due to mutual inductance between wires, the current flow pattern in wide buses changes with frequency too. This phenomenon is called proximity effect.

In general, it is very expensive to include the interconnect 3D models in circuit simulations even after using several reduction techniques. Therefore, it is very important to know when 3D models result in significantly different circuit behavior as compared to using the simple DC model.
On the other hand, with the increase in frequency, number of transistors, number and length of interconnects; the power density has been exponentially increasing during the last 20 years. This increase in power density is translated to a rise in the chip junction temperature which creates reliability threats such as gate-oxide breakdown, negative bias temperature instability (NBTI), accelerated electro migration and ultimately, failure in the long run. Also heating impacts the performance by reducing transistor switching speeds, increasing interconnect resistance, and generally reducing the overall signal quality. Figure 1-4 shows the interconnect power per unit length as a function of delay penalties for various technology nodes [38].

Figure 1-4: Normalized power as a function of delay penalty for various technology nodes
Thus, the development of new design techniques becomes crucial to maximize the use of deep submicron technologies while maintaining an acceptable power consumption level.

This research handles the previously mentioned challenges as follows. Chapter 2 introduces new technique of time shifted moment matching (TSMM) which allows accurate estimation of an RLC circuit response at no extra cost. The TSMM technique performs moment matching (for expansion around $s=0$) on a time-shifted version of the original signal. As compared to other well-known techniques (such as AWE, [3]), TSMM offers distinct advantages. The 50% delay and rise time are determined with much more accuracy for a given approximation order. Moreover, the solutions have significantly improved accuracy as compared to AWE, especially for moderate to highly inductive signals.

Chapter 3 deals with the problem of how to include the inductance effects in static timing analysis while preserving the conventional gate representation as being a lookup table that gives the output propagation delay and slew as response to the input slew and output effective capacitance. Based on a generalized driving point admittance and a generalized waveform shape assumption, the effective load Capacitances expressions of RLC interconnects are derived accurately to estimate both the propagation delay and transition time at the output of a CMOS gate. The new effective capacitance calculation technique poses no extra complexity as compared to the RC based approaches but can accommodate inductance. Simulation results show that the error in propagation delays and rise times when neglecting inductance can be over 4-5 times as compared to the proposed approach.

Chapter 4 introduces a novel model for CMOS gates which is more accurate and at the same time more efficient than the existing models. This model is based on linearizing the gate around
a given load, which allows for moment propagation and uniform treatment of the gates and
interconnects. It is shown that despite the nonlinear overall gate behavior, a linearized gate
model with the model parameters as functions of the load is very accurate. This method is an
order to two orders of magnitude faster then current source based one, while it maintains
accuracy within 5\% of SPICE.

Chapter 5 presents figures of merit and error formulae to determine which interconnects require
volume discretization in the GHZ range. Most of the previous work simply assumes that when
skin depth becomes smaller than the wire cross section dimensions, volume discretization
becomes essential. However, careful analysis in this chapter shows that this assumption is invalid
and a much tighter figure of merit is derived to characterize when volume discretization of single
and coupled wires is required. This derived figure of merit is shown to depend solely on the
interconnect dimensions and spacing and is independent of the type of the materials used or
technology scaling.

Chapter 6 proposes a novel low power, low delay design scheme for CMOS tapered buffers. A
slight increase in the threshold voltage is shown to have an exponential effect in reducing the
total power dissipation. The corresponding increase in the propagation delay is compensated for
by increasing the number of buffer stages such that there is still an overall significant reduction
in the total power dissipation. As compared to the constant threshold voltage design based on a
cost function of \( PT^2 \), the proposed scheme can lead to either a power dissipation reduction of
about 70\% while maintaining the same delay, or up to 30\% in power dissipation with 10\%
propagation delay reduction, respectively. Closed form expressions that give the optimum
threshold voltage and number of stages are presented.
CHAPTER 2

EXPANDING THE FREQUENCY RANGE OF AWE VIA TIME SHIFTING

The new technique of time shifted moment matching (TSMM) is introduced in this chapter. The TSMM technique performs moment matching (for expansion around $s=0$) on a time-shifted version of the original signal. As compared to other well-known techniques (such as AWE,[3]) SMM offers distinct advantages. The 50% delay and rise time are determined with much more accuracy for a given approximation order. Moreover, the solutions have significantly improved accuracy as compared to AWE, especially for moderate to highly inductive signals. TSMM is able to achieve the approximation capability of PVL [4] and PRIMA [5] with the runtime of AWE.

2.1 Introduction

Moment matching techniques, e.g., [1]-[5] are currently one of the most popular linear circuit simulation techniques. The moments of a transfer function of order $n$ results from expanding the transfer function into a Taylor series around $s=0$ as given by

$$H(s) = \frac{1+a_1s+...+a_ms^m}{1+b_1s+...+b_ns^n} = m_0 + m_1s + m_2s^2 + \cdots \quad (2-1)$$

The $i^{th}$ moment $m_i$ of the transfer function is the coefficient of $s^i$ in the Taylor series expansion. To illustrate the relation between the moments, poles, and residues of the transfer function, (2-1) can be expressed as a sum of partial fractions as given by
\[ H(s) = \frac{k_1}{s - p_1} + \frac{k_2}{s - p_2} + \ldots + \frac{k_n}{s - p_n} \quad (2-2) \]

where \( p_i \) is the \( i^{th} \) pole of the transfer function and \( k_i \) is the corresponding residue. By expanding each term in (2-2) into powers of \( s \), the moments of \( H(s) \) can be expressed as:

\[
m_i = -\left( \frac{k_1}{p_1^{i+1}} + \frac{k_2}{p_2^{i+1}} + \ldots + \frac{k_n}{p_n^{i+1}} \right), \quad i = 0 : 2n - 1 \quad (2-3)
\]

This favorable reciprocal relation between the moments and the poles stresses the dominant poles with smaller magnitudes. These dominant poles are of most interest when evaluating the transient response. This characteristic makes the moments very popular in circuit simulation. Moreover, the moments around \( s=0 \) can be calculated very easily for tree structured and tree-like interconnect in linear time with the number of elements in the circuit ([3],[11],[12]).

Asymptotic Waveform Evaluation (AWE,[3]) employs moment matching by calculating the first \( 2n \) moments of the transfer function around \( s=0 \) to determine the first \( n \) dominant poles and the corresponding residues of the transfer function. The terms representing poles with magnitude larger than \( p_n \) \( (p_1<p_2<\ldots<p_n) \) are neglected. Hence, the first \( n \) most dominant poles and the corresponding residues can be calculated by solving the set of \( 2n \) non-linear equations with \( 2n \) variables in (2-3). A straightforward and widely used algorithm to solve this problem is presented in [3]. Once the stable poles and the corresponding residues are available, the waveform which mimics the unit step response can be drawn as a sum of exponentials.

Tens of moments are required for accurate estimation of RLC circuit responses and existing techniques suffer numerical instability, inaccuracies, and/or long runtimes. AWE despite being computationally efficient, suffers from inaccuracy problems with moderate to highly inductive
circuits. On the other hand PVL and PRIMA use implicit moment matching to enhance the accuracy but result in longer run times. Therefore, interconnect timing analyzers must be modified to efficiently handle complex RLC networks present in high speed digital circuits.

![Image](image.png)

**Figure 2-1.** Highly inductive signals have a large inertial delay. Their time shifted versions have less high frequency components than the original one.

The response waveforms for such general linear RLC circuits can have non-monotonic shape and a relatively large inertial delay as shown in Figure 2-1. The sharp transition of the signal that corresponds to this latency contains high frequency components that are relatively difficult to represent. This difficulty is despite the fact that the beneficial information that results from representing the signal in this region of time is just the numerical value of the inertial delay $t_0$. The reason for the difficulty in modeling the response from $t=0$ to $t=t_0$ is that all existing model order reduction techniques use a sum of exponential (SOE) approximation as given by
\[
x(t) \approx \tilde{x}_n(t) = \sum_{i=1}^{n} \left( \frac{k_i}{p_i} e^{p_i t} - \frac{k_i}{p_i} \right) u(t).
\]

(2-4)

where \( k_i \) and \( p_i \) are generally complex numbers and \( u(t) \) is the unit step function. The values of \( p_i \) in the above equation are closely related to the dominant eigenvalues or poles of the original circuit and the values \( k_i \) are the residues. The time domain exponential functions

\[
e^{p_i t} u(t).
\]

(2-5)

with negative real part of \( p_i \) (required for stability) have their highest slope at \( t=0 \). Therefore, trying to represent the inertial latency of an RLC circuit using the sum of exponentials requires a large number of poles. As the signal becomes more inductive, the problem becomes much more severe in a way that the approximation completely misrepresents the original signal.

The method presented in this chapter eliminates the need for extra poles to represent the inertial delay. Hence, all the poles will be available to capture both the 50% delay and rise time accurately. Figure 2-1 shows the original signal and its time shifted version. Time shifting removes the inertial latency as well as the high frequency components associated with it.

The rest of the chapter is organized as follows. The new time shifted moment matching technique is presented in section 2.2. Results are provided in section 2.3 for mild to highly inductive circuits. Finally, the chapter is concluded in section 2.4.

### 2.2 Time Shifted Moment Matching

In this section, the theory for approximating RLC circuit responses accurately using few moments will be presented. Section 2.2.1 presents the details for calculating the inertial delay
using the circuit moments. Section 2.2.2 describes the details about calculating the time shifted circuit moments and representing a shifted version of the original signal in time domain using the corresponding new poles and residues. The method is general enough to be extended easily for other moment matching techniques. The details about obtaining the original signal back from the time shifted time domain version are presented in section 2.2.3.

2.2.1 Calculating the Inertial Delay In Terms of the Circuit Moments

It is important to calculate the value of the inertial delay $t_0$ for a general $RLC$ circuit. In general, $t_0$ will be different at different nodes of the circuit. Hence, the objective of this section is to calculate $t_0$ as a function of the moments at the node of interest. Analogy to the special case of a low loss transmission line will be used to derive $t_0$ for a general $RLC$ circuit.

In the case of a low loss transmission line, the inertial delay $t_0$ is the time required by the signal to fly across the line and is given by [7]-[12]

$$t_0 = \frac{l}{c_0} = l \sqrt{\frac{\mu \varepsilon}{L_0}} = \sqrt{LC},$$  

(2-6)

where $l$ is the length of the transmission line, $c_0$ is the local speed of light, $\varepsilon$ is the permittivity, and $\mu$ is the permeability, $L$ and $C$ are the total interconnect inductance and capacitance, respectively.

To illustrate how to calculate the inertial delay $t_0$ for a general $RLC$ circuit consider the simple single section $RLC$ circuit depicted in Figure 2-2. This circuit has a second order transfer function that is given by
\[ \hat{H}(s) = \frac{1}{s^2 LC + sRC + 1} = \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2}. \] \tag{2-7}

where

\[
\zeta = \frac{1}{2} \frac{RC}{\sqrt{LC}} \quad \text{and} \quad \omega_n = \frac{1}{\sqrt{LC}} \tag{2-8}
\]

Figure 2-2 A simple example to illustrate skin effect

By comparing (2-6) and (2-8), the inertial delay \( t_0 \) is given by \( 1/\omega_n \). The expansion of the transfer function in equation (2-7) is given by

\[
\hat{H}(s) = 1 - s \left( \frac{2\zeta}{\omega_n} \right) + s^2 \left( \frac{-1 + (2\zeta)^2}{\omega_n^2} \right) - \cdots = 1 + m_1 s + m_2 s^2 + \cdots. \tag{2-9}
\]

where \( m_1 \) and \( m_2 \) are the first and second moments of the transfer function at the node of interest. Hence, the damping factor and natural frequency as functions of the circuit moments are given by

\[
\zeta = \frac{-m_1}{2 \sqrt{m_1^2 - m_2}} \quad \text{and} \quad \omega_n = \frac{1}{\sqrt{m_1^2 - m_2}} \tag{2-10}
\]
Hence, the inertial delay can be calculated in terms of the moments of the original system as

\[ t_0 = \sqrt{m_1^2 - m_2}. \]  

(2-11)

Hence, for a system with a non-monotonic response, a second order approximation to the inertial delay can be found if the first and second moments of the system are known.

### 2.2.2 Calculating the Time Shifted Moments and Time Shifted Response

Equation (2-9) shows the transfer function expansion into a Taylor series in terms of the circuit moments. Shifting the time-domain signal by the inertial delay \( t_0 \) as given by (2-11) is equivalent to multiplying the transfer function by \( e^{st_0} \):

\[
\tilde{H}(s) = H(s)e^{st_0} = (m_0 + m_1s + \cdots + m_2s^2 + \cdots) \left(1 + t_0s + \frac{t_0^2s^2}{2!} + \cdots\right).
\]  

(2-12)

where

\[
\tilde{m}_i = \sum_{j=0}^{i} m_j \frac{t_0^{i-j}}{(i-j)!}.
\]  

(2-13)

Equation (2-13) can be written in matrix format as:
\[
\begin{bmatrix}
\tilde{m}_0 \\
\tilde{m}_1 \\
\tilde{m}_2 \\
\vdots \\
\tilde{m}_{2n}
\end{bmatrix} =
\begin{bmatrix}
1 & 0 & 0 & \cdots & 0 \\
t_0 & 1 & 0 & \cdots & 0 \\
t_0^2 & t_0 & 1 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
t_0^{2n} & \frac{t_0^{2n-1}}{2!} & \frac{t_0^{2n-2}}{(2n-1)!} & \cdots & 1
\end{bmatrix}
\begin{bmatrix}
m_0 \\
m_1 \\
m_2 \\
\vdots \\
m_{2n}
\end{bmatrix}.
\tag{2-14}
\]

Equation (2-14) represents a computationally efficient method to obtain the frequency domain moments of the time shifted signal. Now the task is to find the time domain response whose moments are \( \tilde{m}_i \). The AWE algorithm cannot be applied directly because shifting the original signal by a pre-determined value \( t_0 \) as given by (2-11) may result in a non-zero initial condition. A slight modification to the algorithm is applied to solve the problem. The transfer function is expressed as

\[
\tilde{H}(s) = \frac{\tilde{k}_1}{s - \tilde{p}_1} + \frac{\tilde{k}_2}{s - \tilde{p}_2} + \ldots + \frac{\tilde{k}_n}{s - \tilde{p}_n}.
\tag{2-15}
\]

Thus, the corresponding step response for a non-zero initial condition \( \tilde{x}_0 \) is given by:

\[
\tilde{x}(s) = \frac{\tilde{k}_1}{s(s - \tilde{p}_1)} + \frac{\tilde{k}_2}{s(s - \tilde{p}_2)} + \ldots + \frac{\tilde{k}_n}{s(s - \tilde{p}_n)} + \frac{\tilde{x}_0}{s}.
\tag{2-16}
\]

A comparison of the moments results in:
\[
\tilde{m}_0 = \left( \frac{\tilde{k}_1}{\tilde{p}_1} + \frac{\tilde{k}_2}{\tilde{p}_2} + \ldots + \frac{\tilde{k}_n}{\tilde{p}_n} \right) + \tilde{x}_0
\]
\[
\tilde{m}_i = \left( \frac{\tilde{k}_1}{\tilde{p}_{i+1}} + \frac{\tilde{k}_2}{\tilde{p}_{i+1}} + \ldots + \frac{\tilde{k}_n}{\tilde{p}_{i+1}} \right), i = 1:2n.
\]

Instead of using the equations for \(\tilde{m}_0\) to \(\tilde{m}_{b_k-1}\) to solve for the \(n\) poles and \(n\) residues, the equations for \(\tilde{m}_1\) to \(\tilde{m}_{2n}\) are used. The poles and residues for this system can be calculated using standard methods such as AWE. Once the poles and residues are known, the initial condition \(\tilde{x}_0\) is extracted using the equation for \(\tilde{m}_0\):

\[
\tilde{x}_0 = \tilde{m}_0 + \left( \frac{\tilde{k}_1}{\tilde{p}_1} + \frac{\tilde{k}_2}{\tilde{p}_2} + \ldots + \frac{\tilde{k}_n}{\tilde{p}_n} \right).
\]

Hence the time shifted, time domain version of the unit step response has the form

\[
\tilde{x}(t) = \left[ \tilde{m}_0 + \sum_{i=1}^{n} \frac{\tilde{k}_i}{\tilde{p}_i} e^{\tilde{p}_i t} \right] u(t).
\]

### 2.2.3 Calculating the Original Response

Once the time shifted, time domain version of the unit step response is available, obtaining an approximation to the original response \(x(t)\) in time domain is very easy and is given by

\[
x(t) = \tilde{x}(t - t_0) \cdot u(t - t_0).
\]
Figure 2-3 The effect of zero order approximation of $x(t)$ from $t=0$ to $t=t_0$ results in a discontinuity at $t=t_0$.

Note that this is an approximation of the exact response. The derived $x(t)$ has a discontinuity at $t=t_0$ as shown in Figure 2-3. This discontinuity is due to the zero order approximation of the signal from $t=0$ to $t=t_0$. As inductive effects increase, the slope at $t_0$ becomes extremely large and the effect of this approximation becomes infinitesimal.

To improve the shape of the signal before $t_0$, a correction term can be added. This correction term extends the signal backwards from $t_0$ with a straight line whose slope equals to the slope of the signal $x(t)$ at $t=t_0$ multiplied by a factor $\alpha$ whose value depends on $\tilde{x}_0$. The correction to (2-20) is given by

$$x(t) = \tilde{x}(t-t_0) \cdot u(t-t_0)$$

$$+ \left( \beta \cdot (t-t_0) + \tilde{x}_0 \right) \cdot \left( u(t-t_x) - u(t-t_0) \right)$$

(2-21)

where
\[ \beta = \alpha \cdot \frac{d\bar{x}}{dt} \bigg|_{t=0} \quad \text{and} \quad t_x = t_0 - \frac{\bar{x}_0}{\beta} \]

with \( \alpha = 2.5\bar{x}_0 + 0.5 \)

### 2.3 Simulation Results

In this section, simulation results will be presented to show that the time shifted moment matching performs significantly better than non-time shifted moment matching for many different circuits. The damping factor, \( \zeta \), for these circuits ranges from 0.5 (mildly inductive) down to 0.1 (highly inductive). The method was tested on circuits shown in Table 2.1.

**Table 2-1: Circuits used for Testing**

<table>
<thead>
<tr>
<th>Circuit #</th>
<th>Circuit description</th>
<th>Damping factor (( \zeta ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Small netlist (extracted)</td>
<td>0.52</td>
</tr>
<tr>
<td>2</td>
<td>Large netlist (extracted)</td>
<td>0.41</td>
</tr>
<tr>
<td>3</td>
<td>Transmission line</td>
<td>0.13</td>
</tr>
<tr>
<td>4</td>
<td>10-line coupled bus</td>
<td>0.10</td>
</tr>
</tbody>
</table>

**Table 2-2. Comparison of AWE using time shifted moment matching (TSMM) with traditional AWE**

<table>
<thead>
<tr>
<th>Circuit #</th>
<th>order ( q )</th>
<th>rise time error</th>
<th>delay error</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AWE</td>
<td>TSMM</td>
<td>AWE</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>23%</td>
<td>13%</td>
</tr>
</tbody>
</table>
Table 2-2 shows a comparison of AWE using time shifted moment matching (TSMM) with traditional AWE. The table shows 10%-90% rise time and 50% (VDD/2) delay errors at receiver nodes for the circuits listed in Table 2-1. The comparison was done for different approximation orders $q$. For circuits with moderate to high inductive effects, and hence high inertial delay, traditional AWE fails to capture the delay and rise time accurately. The dominant poles are able either to capture delay or to capture rise time but not both. This occurs because AWE tries to

<table>
<thead>
<tr>
<th></th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>16%</td>
<td>9%</td>
<td>4.4%</td>
<td>2.3%</td>
</tr>
<tr>
<td>4</td>
<td>11%</td>
<td>4.7%</td>
<td>3.3%</td>
<td>2.1%</td>
</tr>
<tr>
<td>5</td>
<td>4%</td>
<td>&lt;0.1%</td>
<td>1%</td>
<td>&lt;0.1%</td>
</tr>
<tr>
<td>7</td>
<td>6%</td>
<td>1%</td>
<td>5%</td>
<td>&lt;0.1%</td>
</tr>
<tr>
<td>3</td>
<td>9.6%</td>
<td>6.1%</td>
<td>19.8%</td>
<td>12.2%</td>
</tr>
<tr>
<td>4</td>
<td>7.5%</td>
<td>1.9%</td>
<td>2.5%</td>
<td>1%</td>
</tr>
<tr>
<td>5</td>
<td>11%</td>
<td>1.1%</td>
<td>7%</td>
<td>1%</td>
</tr>
<tr>
<td>7</td>
<td>6%</td>
<td>1%</td>
<td>5%</td>
<td>&lt;0.1%</td>
</tr>
<tr>
<td>3</td>
<td>80%</td>
<td>23%</td>
<td>20%</td>
<td>11%</td>
</tr>
<tr>
<td>5</td>
<td>49%</td>
<td>11%</td>
<td>17%</td>
<td>7%</td>
</tr>
<tr>
<td>7</td>
<td>30%</td>
<td>9%</td>
<td>43%</td>
<td>1.2%</td>
</tr>
<tr>
<td>9</td>
<td>17%</td>
<td>5%</td>
<td>89%</td>
<td>1%</td>
</tr>
<tr>
<td>3</td>
<td>&gt;100%</td>
<td>27%</td>
<td>34%</td>
<td>13%</td>
</tr>
<tr>
<td>5</td>
<td>61%</td>
<td>15%</td>
<td>20%</td>
<td>11%</td>
</tr>
<tr>
<td>7</td>
<td>41%</td>
<td>10%</td>
<td>&gt;100%</td>
<td>&lt;0.1%</td>
</tr>
<tr>
<td>9</td>
<td>19%</td>
<td>7%</td>
<td>&gt;100%</td>
<td>&lt;0.1%</td>
</tr>
</tbody>
</table>
match the high frequency components associated with large inertial delay, which uses up a relatively large number of poles.

A sample of the signals is plotted in Figure 2-4 to Figure 2-8. TSMM performs better than AWE even for mildly inductive circuits. For circuits with moderate to high inductance, it is clear from the results that AWE can completely lose the track of the signal. The TSMM technique is able to enhance the frequency range of AWE and hence improve the results significantly for such circuits. For low inductive circuits, $t_0$ tends to zero and TSMM falls back to AWE.

Figure 2-4 TSMM performs better than AWE for circuits with moderate inductive effects. (circuit #2, $q=5$)
Figure 2-5 For circuits with high inductance, AWE misses the signal whereas TSMM captures it very well (circuit #4, q=9).

Figure 2-6. AWE misses the signal here from a rise time perspective. TSMM captures the signal accurately (circuit #4, q=3).
Figure 2-7. Another example where TSMM captures the signal better than AWE. \((q=5)\)

Figure 2-8. This signal shows that TSMM performs much better than AWE for the same order. \((q=6)\)
2.4 CONCLUSIONS

The new concept of time shifted moment matching was introduced in this chapter to expand the frequency range of methods such as AWE. The introduction of the time shift, which removes the inertial delay of the signal, removes the computational complexity required to obtain the high frequency components to represent the sharp transition of the signal around the inertial delay point $t_0$. The computational efficiency in addition to the improvement of accuracy in rise time and delay prediction makes this technique more appealing for use in circuit simulators. The solution has significantly improved accuracy as compared to AWE for the same approximation order, especially for moderate to highly inductive signals.
CHAPTER 3

INCLUDING INDUCTANCE IN STATIC TIMING ANALYSIS

In this chapter analytical expressions are derived for effective load capacitances of \( RLC \) interconnects to accurately estimate both the propagation delay and transition time at the output of a CMOS gate. The new effective capacitance calculation technique poses no extra complexity as compared to the \( RC \) based approaches but can accommodate inductance. These new expressions are derived based on generalized driving point admittance. The generalized driving point admittance takes inductance into consideration and hence accounts for the inductive shielding that in some cases can even exceed the resistive shielding in current technologies. Another improvement in the new effective capacitance calculation method is the utilization of a more general waveform shape that accounts for the non-monotonic behavior due to inductance effects. It is shown throughout the chapter that two effective capacitances are required for accurate estimation of the propagation delay and rise time with an \( RLC \) interconnect load. Simulation results show that the error in propagation delays and rise times when neglecting inductance can be over 60% as compared to an \( RLC \) model in realistic interconnects. On the other hand, simulations show that the propagation delay and rise time maximum errors associated with the proposed approach are less than 10% as compared to SPICE.
3.1. Introduction and Background

It has been well established that interconnect effects must be accounted for to ensure accurate static timing analysis. Traditionally, gate level static timing analyzers have subdivided the path delay as the sum of the gate delay and the wire or interconnect delay. Regardless of the method used to compute gate delays, an accurate characterization of the loading due to the interconnect at the output of the gate is required.

Historically, interconnect has been modeled as a single lumped capacitance that is the summation of all the interconnect capacitances, $C_t$. With the scaling of technology and increased chip sizes, the cross-sectional area of wires has been scaled down while global interconnect length has increased. The resistance of the interconnect has therefore increased in significance, requiring the use of more accurate models [18]. Qian et al in [13] presented a more accurate effective capacitance modeling. This model took into consideration that a part of the load capacitance is shielded from the gate due to increased interconnect resistance. This shielding resulted in decreasing the value of the effective capacitance that the gate is driving and therefore the gate delay and rise time has decreased. Qian’s approach yields an effective capacitance value that is less than that of the total net capacitance. This model and other extensions, e.g., [13]-[15], have been successfully used in the design of high speed microprocessors [17].

However, these $RC$ based approximations are not always accurate enough since inductive effects are becoming significant in many cases with faster on-chip rise times and longer wire lengths. In addition, many core chips and 3D integration are becoming more common. Thru-Silicon-Vias (TSV’s) are expected to be used extensively to cross between different device layers [18]. These TSV’s have large cross-sectional area and pass through the substrate without having active
devices in the middle. Thus, these interconnects, which are an integral part of the timing of the chip, are expected to be highly inductive. Also, System-in-Package (SiP) is becoming more common with significant amount of interconnects at the package level [19]. So, for a system partitioned across several dies that are integrated on a package, the wires on the package between the dies are an essential part in the timing of the system. These wires are well known to be highly inductive.

Increasing inductance effects can lead to an oscillatory behavior in the voltage waveform [21]-[25] which can lead to a gate delay that is higher than the delay of a gate driving the total net capacitance, $C_t$, as shown in Figure 3-1-a. This behavior occurs when the initial voltage wave injected has a value less than $0.5V_{DD}$, i.e., the time at which the voltage non-monotonic behavior starts, $t_s$, is less than the 50% delay. This happens when the driver gate impedance, $R_d$, is higher than the interconnect characteristic impedance, $R_0$.

![Figure 3-1. The gate output voltage waveform of a CMOS gate driving an interconnect](image)

with (a) $R_d > R_0$ and (b) $R_d < R_0$ for different interconnect models, $RLC$, $RC$, and $C_t$
It is clear from Figure 3-1-a that despite the fact that the voltage waveform at the output of a gate driving an $RLC$ interconnect has a steep slope, the voltage waveform reflection before the 50% point creates an additional delay which increases the total delay even as compared to the delay calculated using the old lumped capacitance assumption. Thus, the error in using the effective capacitance value while assuming an $RC$ interconnect can be even higher than the error in simply assuming a pure capacitive interconnect. It can also be seen in Figure 3-1-a that the $RLC$ model of the interconnect can lead to faster transition time than that of the $RC$ model. This behavior implies that one effective capacitance cannot account for both the faster transition time and longer propagation delay. Two different effective capacitances are required to accurately estimate both the propagation delay and transition time if the initial voltage wave injected has a value less than $0.5V_{DD}$. On the other hand, if the initial voltage wave injected has a value higher than $0.5V_{DD}$, the time at which voltage non-monotonic behavior starts, $t_s$, is higher than the 50% delay. This behavior occurs when $R_d$ is less than $R_0$. In this case, the gate delay is less than what an $RC$ model estimates as shown in Figure 3-1-b. Note that in the case shown in Figure 3-1-b, inductance effects create additional shielding on top of the resistive shielding, resulting in an overall faster signal as compared to the $RC$ model. Thus, the values of the effective capacitance in this case would decrease as compared to the total interconnect capacitance even more than when using an $RC$ model. A simple fudge factor cannot be used on top of $RC$ modeling to account for inductance since as is shown in the previous examples, $RC$ can overestimate or underestimate the delay depending on the ratio $R_0/R_d$.

The rest of the chapter is organized as follows. The computation of the interconnect driving point admittance and characteristic impedance is presented in section 3.2. Section 3.3 presents the
effective capacitance and gate delay calculation at the output of a CMOS gate driving an RLC interconnect that has an initially injected voltage higher than $0.5V_{DD}$ ($R_0 > R_d$). The effective capacitances and the gate delays at the output of a CMOS gate driving an RLC interconnect when $R_0 < R_d$ are computed in section 3.4. Finally, the chapter is concluded in section 3.5.

3.2. Driving Point Admittance and Characteristic Impedance Calculations

In this section, the driving point admittance computations in [20] are generalized to account for RLC loads. Starting with the leaf nodes of an RLC interconnect structure and working back to the source, the Taylor series expansion of the parent driving point admittance is computed in a finite sequence of steps.

![Figure 3-2](image-url)

Figure 3-2. The general rules in calculating the admittance coefficients at the parent node knowing those of the leaves
Figure 3-2 presents the three main rules that should be followed in order to compute the Taylor series expansion coefficients of the driving point admittance of a parent node knowing those of the leaves. These rules are based on two facts:

- For an upstream traversal of a lumped circuit to ground as that shown in Figure 3-3-a, the parent node admittance can be computed in terms of the leaf admittance, $Y_L$, and the circuit admittance, $Y_H$, by $Y_P = Y_L + Y_H$. Note that each impedance is represented as a polynomial in $s$ and hence this step requires a single polynomial addition.

- For an upstream traversal of a series circuit as shown in Figure 3-3-b, the parent node admittance can be computed in terms of the leaf admittance, $Y_L$, and the circuit admittance, $Y_H$, by $Y_P = \frac{Y_L \times Y_H}{Y_L + Y_H}$. The coefficients of $Y_P$ are produced by equating the coefficients of the polynomial equation $Y_P \times (Y_L + Y_H) = Y_L \times Y_H$.

Figure 3-3. Upstream traversal for (a) lumped and (b) series circuits

In order to apply a similar approach to that in [13], an equivalent $RC$-$\pi$ model for a general $RLC$ interconnect needs to be derived. This goal is accomplished by first studying the special case of a
transmission line and generalizing the results to any RLC interconnect. This generalization is achieved by expressing transmission line parameters such as the characteristic impedance $R_0$ in terms of the input impedance $y$ coefficients. It is shown later that the expressions in terms of the $y$ coefficients work very well with a general interconnect structure.

Figure 3-4. Distributed RLC interconnect

In the special case of an RLC transmission line as shown in Figure 3-4 with total resistance, $R_t$, total capacitance $C_t$, and total inductance $L_t$, it was shown in [19]-[20] that the characteristic impedance is complex with a negative imaginary part. Therefore, for the period of time $0 < t < t_s+2T_0$, the characteristic impedance looks like a resistance in series with a capacitance, where $T_0$ is the time required by the signal to travel to the end of the interconnect. The characteristic impedance can be expressed as:

$$Z_0 = R_0 - j \frac{1}{\omega C_0},$$  \hspace{1cm} (3-1)$$

where the values of $R_0$ and $C_0$ in the high frequency GHZ range, are given by [21]:

$$R_0 = \sqrt{\frac{L_t}{C_t}},$$  \hspace{1cm} (3-2)$$
Figure 3-5. The equivalent circuit of a CMOS gate driving an RLC interconnect for the period of time $0 < t < t_s + 2T_0$

The time of flight, $T_0$ was also derived in [21] in terms of the total inductance and the total capacitance and can be given by

$$T_0 = \sqrt{L_t C_t}$$

(3-4)

In order to represent (3-2), (3-3) and (3-4) in terms of the input admittance coefficients, the input admittance of a transmission line can be computed by utilizing the rules in Figure 3-2 which yields an input admittance given by

$$Y_p = (y_{l_t} + C_t) s + \left( y_{l_t} - R \left[ y_{l_t}^2 + C_t y_{l_t} + \frac{1}{3} C_t^2 \right] s^2 \right) \left( y_{l_t} - R \left[ 2y_{l_t} + C_t y_{l_t} + \frac{1}{3} C_t^2 \right] s \right)$$

$$\left( y_{l_t}^3 + \frac{4}{3} C_t y_{l_t}^2 + \frac{2}{3} C_t^2 y_{l_t} + \frac{2}{15} C_t^3 \right) - L \left[ y_{l_t}^2 + C_t y_{l_t} + \frac{1}{3} C_t^2 \right] s^3,$$

(3-5)

where $y_{l_t}$ is the $i^{th}$ term of the load admittance at the end of the line. In the time between $t=0$ and $t=t_s + 2T_0$, the load has no effect on the input waveform since the signal has not come back yet from the load. Thus, the line can be modeled as an unloaded line for this period of time. The input admittance of an unloaded distributed RLC transmission line is given by

$$C_0 = \frac{2\sqrt{L_t C_t}}{R}$$

(3-3)
\[ Y_p = y_{p1}s + y_{p2}s^2 + y_{p3}s^3 \]  

\[(3-6)\]

Where

\[ y_{p1} = C \]
\[ y_{p2} = -\frac{1}{3}RC^2 \]
\[ y_{p3} = \frac{2}{15}R^2C^3 - \frac{1}{3}LC^2 \]  

\[(3-7)\]

Thus, \( T_0, R_0 \) and \( C_0 \) of any general RLC interconnect structure can be expressed as:

\[ T_0 = \sqrt{\frac{\frac{6}{5}y_{p2}^2 - y_{p3}}{y_{p1}}} \]  

\[(3-8)\]

\[ R_0 = \sqrt{\frac{3\frac{6}{5}y_{p2}^2 - y_{p3}}{y_{p1}^2}} \]  

\[(3-9)\]

\[ C_0 = -\frac{2}{3} \frac{y_{p1}^2}{y_{p2}} \sqrt{\frac{3\frac{6}{5}y_{p2}^2 - y_{p3}}{y_{p1}}} \]  

\[(3-10)\]

The relation between \( R_0 \) and \( R_d \) determines the value of the initial voltage wave injected to the interconnect. In other words, their ratio determines whether the interconnect experiences a late oscillation \((R_0>R_d)\), after the 50% point, or early oscillation \((R_0<R_d)\), before the 50% point, as shown in Figure 3-6. Note that \( R_d \) does not need to be exact. It only serves as a figure to determine whether early oscillation or late oscillation takes place.
Figure 3-6. Late oscillation versus early oscillation of an RLC interconnect

The driver resistance, $R_d$, can be computed from characterized gate timing tables as shown in (3-11) knowing $t_d(C_t, t_{ii})$ which is the propagation delay that corresponds to the total capacitance $C_t$ and the input transition time, $t_{ii}$

$$R_d = \frac{t_d(C_t, t_{ii})}{0.7(C_t + C_{out})} \quad (3-11)$$

where $C_{out}$ is the driver gate output capacitance and can be computed as shown in (3-12)

$$C_{out} = \frac{C_t}{\frac{t_d(C_t, t_{ii})}{t_d(0, t_{ii})} - 1} \quad (3-12)$$

The simplification process explained in the previous section ignores proximity effect completely. However, proximity effect can be prominent for very close wires. When this effect is significant, measures must be taken to handle it while keeping the simplicity of the equivalent circuits.
3.3. Estimating the Gate Delays and Effective Capacitances For an Interconnect Experiencing Late Oscillations

A CMOS gate driving an \(RLC\) interconnect that experience a late oscillation has the equivalent circuit shown in Figure 3-5 for \(0 < t < t_s + 2T_0\). Note that \(C_{out}\) is not included explicitly since it is part of the gate characterization. Thus, the capacitance that drives the same average current as the \(R_0C_0\) segment load would accurately represent the effective capacitance that estimates the propagation delay.

Referring to Figure 3-7, the effective capacitance that drive the same average current as an \(RC\) \(\pi\)-model is given by [13]

\[
C_{\text{eff}} = C_2 + C_1 \left[1 - \frac{R_2C_1}{2} \left( \frac{t_d - t_s}{2} \right) \left( 1 - e^{-\frac{t_s}{R_2C_1}} \right) \right] \quad (3-13)
\]

Where

\[
t_d = t_d + \frac{t_{\text{d}}}{2} \quad (3-14)
\]

\[
t_x = t_d + \frac{t_{\text{d}}}{2} - 0.5t_{\text{r}0} \quad (3-15)
\]
The expression in (3-13) was derived based on approximating the voltage waveform shape as a quadratic shape starting from the initial voltage to the 20% point. Then form the 20% point to the 50% point the voltage waveform shape was assumed to be linear. This approximation holds very well with the late oscillation case shown in Figure 3-6. Thus, by direct application of (3-13), the $R_0 C_0$ segment in Figure 3-5 would have an effective capacitance given by

$$C_{eff} = C_0 \left[ 1 - \frac{R_0 C_0}{t_D - t_x/2} + \frac{(R_0 C_0)^2}{t_D - t_x/2} e^{\frac{R_0 C_0}{t_D - t_x/2} \left( t_D - t_x/2 \right)} \right]$$

(3-16)

From (3-14), (3-15) and (3-16) it can be seen that the effective capacitance is a function of the delay and the transition time of the output waveform which are not known a priori. Similar to the iterative approach that is adopted in calculating the effective capacitance of an $RC \pi$-model [13], the effective capacitance for the $R_0 C_0$ segment can be calculated from the following iterative procedure:

1. Set the load capacitance value equal to the total interconnect capacitance $C_i$. 
2. Use the load capacitance value with the gate characterization to obtain the propagation delay and the output transition time.

3. Using $t_d$ and $t_{o0}$ obtained in step 2, calculate a $C_{eff}^{R,C_0}$ using (3-13), (3-15) and (3-16)

4. If the value of $C_{eff}^{R,C_0}$ is still changing, set the load capacitance value equal to $C_{eff}^{R,C_0}$ and go to step 2

The initial value of the load capacitance is set to $C_t$ for which convergence is then achieved in 2-3 iterations. The same value of $C_{eff}^{R,C_0}$ can be used to estimate both the propagation delay and the transition time if the value of initial voltage wave injected to the interconnect is higher than $0.8V_{DD}$ as shown in Figure 3-8-a

![Figure 3-8.](image)

**Figure 3-8.** The voltage waveform at the output of a CMOS gate driving an RLC interconnect for an initially injected voltage (a) higher than $0.8V_{DD}$ and (b) less than $0.8V_{DD}$

On the other hand, if the initial voltage wave injected to the interconnect is less than $0.8V_{DD}$ as shown in Figure 3-8-b, then using the same effective capacitance value, $C_{eff}^{R,C_0}$, underestimates the
transition time by $2T_0$. Hence, if the initially injected voltage wave has a value that is higher than $0.5V_{DD}$ and less than $0.8V_{DD}$ i.e., the relation between $R_d$ and $R_0$ is $0.5 < \frac{R_0}{R_d + R_0} < 0.8$, a correction factor of the value $2T_0$ should be added to the transition time estimated using $C_{off}^{R_dC_0}$.

The proposed methodology was tested on a number of interconnects on upper layers in 90 $nm$ and 65 $nm$ technologies, including both transmission lines and tree structures.

Table 3-1 shows the propagation delays and transition times that result in modeling line interconnects as RLC interconnect versus RC interconnects when $0.5 < \frac{R_0}{R_d + R_0} < 0.8$. The error in both models is relative to SPICE.

**Table 3-1. Delay errors at the gate output for an RLC line model having an initial injected voltage less than 0.8$V_{DD}$ versus the RC model as compared to SPICE**

<table>
<thead>
<tr>
<th></th>
<th>SPICE</th>
<th>RLC</th>
<th>Error</th>
<th>RC</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$l=3mm$</td>
<td>$w=0.5um$</td>
<td>$t_{50}$=36.3ps</td>
<td>38 ps</td>
<td>4.6%</td>
<td>39ps</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$t_r$=24.5 ps</td>
<td>26.7ps</td>
<td>9.3%</td>
<td>15 ps</td>
</tr>
<tr>
<td>$l=2mm$</td>
<td>$w=0.5um$</td>
<td>$t_{50}$=31ps</td>
<td>33 ps</td>
<td>6.4%</td>
<td>36ps</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$t_r$=22 ps</td>
<td>23.7ps</td>
<td>7.7%</td>
<td>13 ps</td>
</tr>
<tr>
<td>$l=2.5mm$</td>
<td>$w=1.25um$</td>
<td>$t_{50}$=35ps</td>
<td>38ps</td>
<td>7.6%</td>
<td>41ps</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$t_r$=36ps</td>
<td>37ps</td>
<td>3%</td>
<td>18 ps</td>
</tr>
<tr>
<td>$l=2.25mm$</td>
<td>$w=3.27um$</td>
<td>$t_{50}$=32ps</td>
<td>34 ps</td>
<td>5.3%</td>
<td>40ps</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$t_r$=37 ps</td>
<td>40ps</td>
<td>8.7%</td>
<td>19 ps</td>
</tr>
<tr>
<td>$l=4mm$</td>
<td>$t_{50}$=63ps</td>
<td>65 ps</td>
<td>5%</td>
<td>54ps</td>
<td>15%</td>
</tr>
</tbody>
</table>
Table 3-2 shows the propagation delays and transition times that result in modeling tree interconnects using $RLC$ models versus $RC$ models when $0.5 < \frac{R_0}{R_d + R_0} < 0.8$. The error in both models is relative to SPICE.

**Table 3-2. Delay errors at the gate output for $RLC$ tree interconnect having an initial injected voltage less than $0.8V_{dd}$ versus the $RC$ model as compared to SPICE**

<table>
<thead>
<tr>
<th>l=2.5mm</th>
<th>tr=50 ps</th>
<th>45ps</th>
<th>10%</th>
<th>39.4 ps</th>
<th>-21%</th>
</tr>
</thead>
<tbody>
<tr>
<td>w=0.6um</td>
<td>t50=42ps</td>
<td>44.2 ps</td>
<td>5%</td>
<td>47ps</td>
<td>13%</td>
</tr>
<tr>
<td></td>
<td>tr=30 ps</td>
<td>32ps</td>
<td>6.1%</td>
<td>16 ps</td>
<td>-47%</td>
</tr>
</tbody>
</table>

It is clear from Table 3-1 that using the $RC$ model overestimates the propagation delay due to neglecting the inductance shielding effect by assuming only $RC$ behaviour. On the other hand, the $RC$ model underestimates the transition time. This is because the $RC$ model does not account for the non-monotonic behaviour shown Figure 3-8-b.

Table 3-2 shows the propagation delays and transition times that result in modeling tree interconnects using $RLC$ models versus $RC$ models when $0.5 < \frac{R_0}{R_d + R_0} < 0.8$. The error in both models is relative to SPICE.

**Table 3-2. Delay errors at the gate output for $RLC$ tree interconnect having an initial injected voltage less than $0.8V_{dd}$ versus the $RC$ model as compared to SPICE**

| l=3mm | t50=32.4ps | 35ps | 8% | 54ps | 68% |
| w=3um | tr=72ps | 69ps | -4.7% | 35.7 ps | -51% |
| t=1.5um |
| l=2mm | t50=27ps | 29 ps | 7.4% | 43ps | 59% |
| w=3um | tr=65 ps | 62ps | 4.5% | 30.1 ps | -53% |
| t=1.2um |
| l=4mm | t50=34ps | 36.2 ps | 6% | 47.3ps | 31% |
It is clear from Table 3-2 that the $RC$ model for tree interconnects leads to more delay errors than the $RC$ models for line interconnects. The reason is that the signal has longer time of flight with the increased capacitance. Hence, the monotonic approximation of the $RC$ model becomes more erroneous.

Table 3-3 shows the propagation delays and transition time that result in modeling line interconnects using $RLC$ models versus $RC$ models when $\frac{R_0}{R_d + R_0} > 0.8$.

**Table 3-3. Delay errors at the gate output for $RLC$ line interconnect model having an initial injected voltage higher than 0.8$V_{DD}$ versus the $RC$ model as compared to SPICE**

<table>
<thead>
<tr>
<th>l=2mm</th>
<th>SPICE</th>
<th>RLC</th>
<th>Error</th>
<th>RC</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>w=3um</td>
<td>t50=30ps</td>
<td>31.3</td>
<td>4.3%</td>
<td>34.5</td>
<td>14%</td>
</tr>
<tr>
<td>t=1.5um</td>
<td>tr=10.3ps</td>
<td>10</td>
<td>-3%</td>
<td>11.6</td>
<td>13%</td>
</tr>
</tbody>
</table>

<p>| l=3mm  | t50=32.4ps | 32.3  | 0.3%  | 35    | 9%    |
| w=0.5um| tr=11.5ps  | 11    | -4%   | 15    | 20%   |</p>
<table>
<thead>
<tr>
<th>$t=1.25,\mu m$</th>
<th>$l=2,\text{mm}$</th>
<th>$t50=24,\text{ps}$</th>
<th>23 ps</th>
<th>4%</th>
<th>27ps</th>
<th>17%</th>
</tr>
</thead>
<tbody>
<tr>
<td>$w=0.5,\mu m$</td>
<td>$t=0.5,\mu m$</td>
<td>$tr=9,\text{ps}$</td>
<td>9.7ps</td>
<td>7.7%</td>
<td>10.5 ps</td>
<td>16%</td>
</tr>
<tr>
<td>$l=2.25,\text{mm}$</td>
<td>$w=3.27,\mu m$</td>
<td>$t=1.55,\mu m$</td>
<td>$t50=30,\text{ps}$</td>
<td>31 ps</td>
<td>3%</td>
<td>34ps</td>
</tr>
<tr>
<td>$t=1.55,\mu m$</td>
<td>$tr=11,\text{ps}$</td>
<td>11.5ps</td>
<td>4%</td>
<td>12 ps</td>
<td>9%</td>
<td></td>
</tr>
<tr>
<td>$l=4,\text{mm}$</td>
<td>$w=0.5,\mu m$</td>
<td>$t=1.,\mu m$</td>
<td>$t50=63,\text{ps}$</td>
<td>65 ps</td>
<td>3%</td>
<td>70ps</td>
</tr>
<tr>
<td>$t=1.,\mu m$</td>
<td>$tr=20,\text{ps}$</td>
<td>19ps</td>
<td>-5%</td>
<td>25 ps</td>
<td>25%</td>
<td></td>
</tr>
<tr>
<td>$l=2.5,\text{mm}$</td>
<td>$w=0.6,\mu m$</td>
<td>$t=1.56,\mu m$</td>
<td>$t50=42,\text{ps}$</td>
<td>44.2 ps</td>
<td>5%</td>
<td>47ps</td>
</tr>
<tr>
<td>$t=1.56,\mu m$</td>
<td>$tr=10,\text{ps}$</td>
<td>9.2ps</td>
<td>-8%</td>
<td>11.5 ps</td>
<td>15%</td>
<td></td>
</tr>
</tbody>
</table>

It is clear from Table 3-3 that using the \textit{RC} model overestimates both the propagation delay and transition time due to neglecting the inductive shielding effect by assuming only \textit{RC} behaviour. In this case, the voltage waveform reaches its 80\% point before the oscillatory behaviour starts as shown in Figure 3-8-a.

Table 3-4 shows the propagation delays and transition time that result in modeling tree interconnects as \textit{RLC} interconnect versus \textit{RC} interconnects when \( \frac{R_0}{R_d+R_0} > 0.8 \).
Table 3-4. Delay errors at the gate output for RLC tree interconnect model having an initial injected voltage higher than $0.8V_{DD}$ versus the RC model as compared to SPICE

<table>
<thead>
<tr>
<th>L=3mm</th>
<th>W=3um</th>
<th>T=1.5um</th>
<th>SPICE</th>
<th>RLC</th>
<th>Error</th>
<th>RC</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>t50=31.4ps</td>
<td>33ps</td>
<td>4.18%</td>
<td>40.6ps</td>
<td>20%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tr=10.8ps</td>
<td>11.5ps</td>
<td>6.4%</td>
<td>15.9ps</td>
<td>38%</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>L=2mm</th>
<th>W=3um</th>
<th>T=1.5um</th>
<th>SPICE</th>
<th>RLC</th>
<th>Error</th>
<th>RC</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>t50=24.8ps</td>
<td>26ps</td>
<td>4.8%</td>
<td>30.5ps</td>
<td>23%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tr=7ps</td>
<td>7.2ps</td>
<td>2.8%</td>
<td>10ps</td>
<td>43%</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>L=4mm</th>
<th>W=0.5um</th>
<th>T=1.25um</th>
<th>SPICE</th>
<th>RLC</th>
<th>Error</th>
<th>RC</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>t50=32.2ps</td>
<td>34.2ps</td>
<td>6.2%</td>
<td>38.7ps</td>
<td>14%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tr=11.2ps</td>
<td>11.4ps</td>
<td>1.7%</td>
<td>15.4ps</td>
<td>38%</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>L=2mm</th>
<th>W=0.5um</th>
<th>T=1.25um</th>
<th>SPICE</th>
<th>RLC</th>
<th>Error</th>
<th>RC</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>t50=27ps</td>
<td>28ps</td>
<td>3.7%</td>
<td>31ps</td>
<td>11%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tr=9ps</td>
<td>9.1ps</td>
<td>1.1%</td>
<td>11ps</td>
<td>22%</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3.4. Estimating the Gate Delay and Effective Capacitances for an Interconnect Experiencing Early Oscillation

The general rising waveform shape of the output voltage of a CMOS gate driving an RLC interconnect that experience early oscillations is shown in Figure 3-9. For this case, the formula in (3-13) cannot be used to estimate the effective capacitance since the waveform cannot be
assumed linear between the 20% point and the 50% point. Starting at the initial voltage $V_i$, the waveform is quadratic up to the 20% point ($V_i$ is equal to zero for rising waveform, and equal to $V_{DD}$ for a falling waveform). Then from the 20% point to the 80% the waveform follows a third order behavior due to the reflection. Thereafter, the remaining part of the curve experiences a number of overshoots and undershoots that fade gradually and finally the voltage curve settles to $V_{DD}$ for a rising waveform and zero for a falling waveform. These overshoots and undershoots are best approximated by an exponential tail, especially for the purpose of current computation. The intuition behind this exponential approximation is that the net current flowing in a capacitor, that results from an overshoot followed by an equivalent undershoot, is approximately zero. Hence, the current that results from an exponentially fading overshoots and undershoots is equivalent to the current that results from an exponential tail that has the same starting and ending point.

![Figure 3-9. The voltage waveform at the output of a CMOS gate driving an RLC interconnect experiencing early oscillation](image)

Figure 3-9. The voltage waveform at the output of a CMOS gate driving an RLC interconnect experiencing early oscillation
Thus, the output voltage waveform equation is assumed as

\[
V_{out}(t) = \begin{cases} 
-ct(t - t_0) & 0 < t < t_i \\
a(t - t_y)^3 + b(t - t_y) + d & t_i < t < t_f \\
V_{DD} (1 - ke^{-\alpha t}) & t_f < t 
\end{cases}
\]  

(3-17)

where \( t_0 \) is the output 0\% point shown in Figure 3-9, \( t_i \) is the output 20\% point, and \( t_f \) is the output 80\% point. The other constants denoted by \( a, b, c, d, t_y, k \), and \( \alpha \) are solved for, assuming that the values of \( t_0, t_i \), and \( t_f \) are estimated. The output voltage value at \( t_i \) equals 0.2\( V_{DD} \). Also, the voltage waveform and its first derivative are continuous at \( t_i \), therefore

\[-ct_i(t - t_0) = 0.2 \times V_{DD} \]  

(3-18)

\[a(t_i - t_y)^3 + b(t_i - t_y) + d = 0.2 \times V_{DD} \]  

(3-19)

\[3a(t_i - t_y)^2 + b = c(t_{si} - 2t_i) \]  

(3-20)

The output voltage values at the 50\% point denoted by \( t_D \) and \( t_f \) are 0.5\( V_{DD} \) and 0.8\( V_{DD} \), respectively. Thus,

\[a(t_D - t_y)^3 + b(t_D - t_y) + d = 0.5 \times V_{DD} \]  

(3-21)

\[a(t_f - t_y)^3 + b(t_f - t_y) + d = 0.8 \times V_{DD} \]  

(3-22)

The values of the five constants \( a, b, c, d, \) and \( t_y \), can be obtained by solving (3-18) through (3-22) as

\[c = \frac{0.2V_{DD}}{t_i(t_i - t_{si})} \]  

(3-23)

\[t_y = \frac{A}{B} \]  

(3-24)

Where
The values of the constants $k$ and $\alpha$ can be obtained assuming that the output voltage waveform and its first derivative are continuous at $t_f$ which yields

$$V_{DD} \left( 1 - ke^{-\alpha t_f} \right) = 0.8 \times V_{DD}$$  \hspace{1cm} (3-30)

$$V_{DD} \times \alpha ke^{-\alpha t_f} = 3a \left( t_f - t_y \right)^2 + b$$  \hspace{1cm} (3-31)

Thus, $k$ and $\alpha$ are given by

$$\alpha = 3a \left( t_f - t_y \right)^2 + b$$  \hspace{1cm} (3-32)

$$k = 0.2V_{DD} \times e^{-\alpha t_f}$$  \hspace{1cm} (3-33)
Hence, the complete output voltage waveform can be estimated knowing the 20%, 50%, and 80% delay points.

Similar to the approach presented in [13] to determine the effective capacitance of a given load. The effective capacitance that was to determine the effective capacitance that results in the same propagation delay as the $RLC$ interconnect is the capacitance that drives the same average current as that of the interconnect. Referring to Figure 3-10, the same approach can be adopted knowing the gate output driving point admittance.

![Diagram](image)

**Figure 3-10. An effective capacitance that captures the effects of an $RLC$ interconnect**

The current flowing in the effective capacitance can be expressed as:

$$I_C(s) = V_{out}(s) \times sC_{eff}$$

$$= i_{C0} + i_{C1}s + i_{C2}s^2 + ...$$

(3-34)

where $V_{out}(s)$ is the Laplace transform of (3-17) The current components $i_{C0}$ and $i_{C1}$ are given by

$$i_{C0} = \left(2a\left(t_D^3 - t_i^3\right) + V_{DD}\right)C_{eff}$$

(3-35)

$$i_{C1} = h \times C_{eff}$$

(3-36)

Where
The current flowing in the original interconnect load can be approximated as

\[ I_Y(s) = V_{out}(s) \times Y(s) \]  

(3-38)

where \( Y(s) \) is the gate output driving point admittance whose Taylor series expansion takes the form

\[ Y(s) = y_1s + y_2s^2 + y_3s^3 + ... \]  

(3-39)

Thus, the current flowing in the original interconnect can be approximated by

\[ I_Y(s) = V_{out}(s)\left(y_1s + y_2s^2 + y_3s^3 + ...ight) \]
\[ = i_{y0} + i_{y1}s + i_{y2}s^2 + ... \]

(3-40)

The current components \( i_{y0} \) and \( i_{y1} \) are given by

\[ i_{y0} = \left(2a(l_D^3 - l_i^3) + V_{DD}\right)y_1 + 3a(l_i^2 - t_f^2)y_2 \]  

(3-41)

\[ i_{y1} = h_y y_1 + \left(2a(l_D^3 - l_i^3) + V_{DD}\right)y_2 + 3a(l_i^2 - t_f^2)y_3 \]  

(3-42)

Thus, the effective capacitance that has the same average current as that of the interconnect which gives the same propagation delay can derived by equating (3-35) and (3-41)and is given by

\[ C_{eff}^d = \frac{\left(2a(l_D^3 - l_i^3) + V_{DD}\right)y_1 + 3a(l_i^2 - t_f^2)y_2}{2a(l_D^3 - l_i^3) + V_{DD}} \]  

(3-43)
Accurate estimation of the transition time requires higher frequency component matching.

Thus, the effective capacitance that gives accurate transition time estimation is obtained by matching $i_{C1}$ and $i_{y1}$ and is given by

$$C_{eff}^{tr} = \frac{h.y_1 + \left(2a(t_3^3 - t_i^3) + V_{DDD}\right)y_2 + 3a(t_i^2 - t_f^2)y_3}{h}$$

(3-44)

Thus, as in [13] the effective capacitances are calculated iteratively as follows:

1. Calculate the first three coefficients of the interconnect admittance at the gate output point, $y_1$, $y_2$, and $y_3$
2. Set $C_{eff}^{d}$ and $C_{eff}^{tr}$ to an initial value equal to the total capacitance, $C_t$.
3. Use $C_{eff}^{d}$ and $C_{eff}^{tr}$ to obtain the propagation delay and transition time, respectively, using the gate characterization tables.
4. Calculate the 0% point, $t_0$, the 20% point, $t_i$, 50% point, $t_D$ and the 80% delay point, $t_f$, as $t_d+0.5t_{tt}, t_d+0.83t_{tt}, t_d+0.5t_{tt}, t_d+0.5t_{tt}$ and $t_d+0.5t_{tt}+0.5t_{tt}$, respectively.
5. Substitute in (3-43) and (3-44) to update the value of $C_{eff}^{d}$ and $C_{eff}^{tr}$, respectively.
6. If the value either of $C_{eff}^{d}$ or $C_{eff}^{tr}$ is changing update the capacitance value and go to step 3.

The iteration procedures as described above makes the benefit from each of the effective capacitances in the calculation of the other which speed up the convergence. Empirically, we have found that the effective capacitances values converge in no more than 2-3 iterations.
Table 3-5 shows the propagation delays and transition time that result in modeling line interconnects as $RLC$ interconnect versus $RC$ interconnects when $R_0 < R_d$. The error in both models is relative to SPICE.

Table 3-5. Delay errors at the gate output for $RLC$ line interconnect model having an initial injected voltage less than $0.5V_{DD}$ versus the $RC$ model as compared to SPICE

<table>
<thead>
<tr>
<th>l=2mm</th>
<th>SPICE</th>
<th>RLC</th>
<th>Error</th>
<th>RC</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>w=3um</td>
<td>t50=67ps</td>
<td>70ps</td>
<td>4.48%</td>
<td>58ps</td>
<td>13.4%</td>
</tr>
<tr>
<td></td>
<td>tr=54ps</td>
<td>49ps</td>
<td>9%</td>
<td>44 ps</td>
<td>18.5%</td>
</tr>
<tr>
<td>t50=63ps</td>
<td>59.8 ps</td>
<td>4.75%</td>
<td>55ps</td>
<td>12.7%</td>
<td></td>
</tr>
<tr>
<td>w=3um</td>
<td>tr=45 ps</td>
<td>44.482ps</td>
<td>1%</td>
<td>42 ps</td>
<td>6%</td>
</tr>
<tr>
<td>t50=58ps</td>
<td>59 ps</td>
<td>1.6%</td>
<td>53.3ps</td>
<td>9%</td>
<td></td>
</tr>
<tr>
<td>w=0.5um</td>
<td>tr=43 ps</td>
<td>39ps</td>
<td>9%</td>
<td>37 ps</td>
<td>12%</td>
</tr>
<tr>
<td>t50=55ps</td>
<td>58ps</td>
<td>7.6%</td>
<td>47ps</td>
<td>13%</td>
<td></td>
</tr>
<tr>
<td>w=1um</td>
<td>tr=40ps</td>
<td>37ps</td>
<td>9.3%</td>
<td>32 ps</td>
<td>15%</td>
</tr>
<tr>
<td>t50=62ps</td>
<td>59 ps</td>
<td>5%</td>
<td>54ps</td>
<td>12.7%</td>
<td></td>
</tr>
<tr>
<td>w=3um</td>
<td>tr=44 ps</td>
<td>44ps</td>
<td>1%</td>
<td>41 ps</td>
<td>6%</td>
</tr>
<tr>
<td>t50=63ps</td>
<td>65 ps</td>
<td>5%</td>
<td>54ps</td>
<td>15%</td>
<td></td>
</tr>
<tr>
<td>w=0.5um</td>
<td>tr=50 ps</td>
<td>45ps</td>
<td>10%</td>
<td>39.4 ps</td>
<td>21%</td>
</tr>
</tbody>
</table>

It is clear from Table 3-5 that using the $RC$ model underestimates the propagation delay due to neglecting the inductance shielding effect by assuming only $RC$ behaviour.
Table 3-6 shows the propagation delays and transition time that result in modeling tree interconnects as $RLC$ interconnect versus $RC$ interconnects when $R_d > R_l$. The error in both models is relative to SPICE.

**Table 3-6. Delay errors at the gate output for $RLC$ tree interconnect model having an initial injected voltage less than $0.5V_{DD}$ versus the $RC$ model as compared to SPICE**

<table>
<thead>
<tr>
<th></th>
<th>SPICE</th>
<th>RLC</th>
<th>error</th>
<th>RC</th>
<th>error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$l=1$mm $w=3\mu m$</td>
<td>$t_{50}=71.7\text{ps}$</td>
<td>68.8ps</td>
<td>4.18%</td>
<td>66ps</td>
<td>8%</td>
</tr>
<tr>
<td></td>
<td>$t_{r}=55\text{ps}$</td>
<td>57ps</td>
<td>1.8%</td>
<td>59ps</td>
<td>9%</td>
</tr>
<tr>
<td>$l=2$mm $w=0.5\mu m$</td>
<td>$t_{50}=54.8\text{ps}$</td>
<td>58ps</td>
<td>8%</td>
<td>50.5ps</td>
<td>9%</td>
</tr>
<tr>
<td></td>
<td>$t_{r}=50\text{ps}$</td>
<td>49ps</td>
<td>1%</td>
<td>61ps</td>
<td>21%</td>
</tr>
<tr>
<td>$l=4$mm $w=0.5\mu m$</td>
<td>$t_{50}=56.3\text{ps}$</td>
<td>54.2ps</td>
<td>3.57%</td>
<td>60ps</td>
<td>9%</td>
</tr>
<tr>
<td></td>
<td>$t_{r}=53.4\text{ps}$</td>
<td>57ps</td>
<td>4.1%</td>
<td>66ps</td>
<td>24.3%</td>
</tr>
<tr>
<td>$l=2$mm $w=0.5\mu m$</td>
<td>$t_{50}=52\text{ps}$</td>
<td>49ps</td>
<td>5.57%</td>
<td>46ps</td>
<td>11.54%</td>
</tr>
<tr>
<td></td>
<td>$t_{r}=48\text{ps}$</td>
<td>52ps</td>
<td>8.3%</td>
<td>55ps</td>
<td>12.5%</td>
</tr>
</tbody>
</table>

### 3.5. Conclusions

Accurate analytical expressions for effective load capacitances of $RLC$ interconnects that estimate both the propagation delay and transition time at the output of a CMOS gate were derived. The new effective capacitances calculation was done at no extra complexity compared to the $RC$ based approaches. It was shown throughout the chapter that simply using single effective capacitances does not capture both propagation delay and rise time accurately when
inductance becomes important. Simulation results showed that the error in propagation delays and rise times when neglecting inductance can be over 60% as compared to an $RLC$ model. On the other hand, simulations showed that the propagation delay and rise time maximum errors associated with the proposed approach are less than 10% as compared to SPICE. The obtained results will enable accurate STA for general $RLC$ interconnects.
CHAPTER 4
A NOVEL MOMENT-BASED METHODOLOGY FOR
ACCURATE AND EFFICIENT STATIC TIMING ANALYSIS

A novel methodology for accurate and efficient static timing analysis is presented in this chapter. The methodology is based on gate linearization around a given load, which allows for moment propagation and uniform treatment of the gates and interconnects. It is shown that despite the highly nonlinear overall gate model, a linearized gate model with the model parameters as functions of the load is very accurate. A gate and input capacitance characterization is proposed which provides for accuracy, efficiency and flexibility in the path performance calculation. To illustrate the concept and prove its merits, multiple examples are presented. The method is an order to two orders of magnitude faster then current source based one, while it maintains accuracy within 5% of SPICE.

4.1. Introduction

Static timing analysis (STA) is one of the crucial steps in gate level design flows and improving the performance and accuracy of STA has been the focus of recent research work in both, academia and industry, e.g., [33]-[39]. Traditionally, STA has been based on a circuit decomposition (decoupling) into nonlinear gates and linear interconnects, which were treated separately. In this methodology, the gates are pre-characterized by dynamic simulation using various input ramp times and output load capacitances and the 2-D tables for the output delays and the output rise times are generated and stored for the delay reading purposes during the path delay calculation [27], [28]. The interconnects are treated as linear networks and, typically, the
impulse response moment based methods [26], [29], and [30] are used to reduce the interconnect networks and to propagate the delays and ramp times to the inputs of the next stages in the IC paths. The interface between nonlinear gates and linear interconnects is facilitated using effective load calculation at the driving points of the gates [27], [28], which account for increasingly resistive and inductive [27], [28], [31], and [32] interconnects.

In order to improve accuracy, various EDA vendors were increasing the resolution of characterization information [33], and [34] but the models essentially remained input rise time and output load dependant and require iterative $C_{eff}$ calculation.

The current source models [35]-[39] were introduced to improve accuracy in predicting increasingly nonlinear waveform shapes resulting from narrow and dense interconnect in new shrinking technologies. The initial current source model [35] consisted of a DC current source which depended on the gate arc’s input and output voltages, along with the arc’s input and output capacitances for transient effects modeling.

The model was later extended to include nonlinear output and input capacitance models as well as multi input switching effects [37]-[39]. All of these extensions improve the model accuracy but, at the same time, significantly increase the run time making current source gate based STA much less efficient.

This chapter proposes a novel model for STA, which linearizes the gate models around a given output load, allowing for uniform and efficient model-based approach to be applied for accurate stage and path delay calculation. The gate characterization produces the sets of parameters of the linearized gate model for various output loads (as functions of the first two admittance moments of the load). The input capacitance of a gate is also characterized for various gate output loads.
providing a linearized interface between the interconnects and the gates. Essentially, linearizing the gate for a given load is reasonable, because it is reasonable to assume that there exists some linear system for a given load that will reproduce the same gate output signal as the nonlinear gate. A single linearized gate model will not be able to replicate the non-linear gate response for all loads, but as shown in this chapter, linearization works very well for a single load.

Once the linearized gate model is determined for a given load, the stage output signal moments can be calculated using the convolution of the interconnect and gate moments, i.e. transfer function multiplication. This way switching between the frequency and time domains during the waveform propagation is avoided, preserving high efficiency as well as good accuracy. Only the moments are propagated. There is no need for ramp time or waveform approximation at the path stage inputs. The time domain representations (the waveforms) would be extracted at the specified nodes only. The proposed model has a built in flexibility which allows to trade-off accuracy and performance by including the desirable number of moments in the moment propagation, depending on the needed waveform accuracy. Note that the gate characterization complexity is totally independent of the number of moments chosen. The propagation of only the two to three moments is sufficient to capture accuracy associated with conventional delay and ramp time calculation.

This methodology provides a solution for the otherwise difficult problem of integrating nonlinear gates and linear interconnect. The method accurately and efficiently accounting for the impact of the gate load on the gate input impedance (through the gate’s miller capacitance) which can significantly affect accuracy and/or performance of STA.
Effective load based methods are inherently inaccurate and hardly applicable in state of the art technologies, while current source based approach has to pay significant performance penalty in order to accurately account for the interconnect effects. The application of the proposed methodology for the path delay calculation is straightforward. One has to first, going from the circuit leafs to the roots, calculate the driving point admittances of all the gates along the path, and then in a single forward traversal accurately calculate the delays by simply propagating the path moments. To prove of the concept and demonstrate the accuracy and efficiency of the proposed method, the method has been applied to a number of examples.

By uniform impulse response moment based treatment of the gates and interconnects, the proposed methodology provides a different direction (opportunity for efficient and accurate work) in a range of applications, such as noise analysis, inductive effects analysis, floor-planning with a time and noise budgeting, cell characterization, variation aware analysis, performance corner based analysis, and full Statistical STA.

The rest of the chapter is organized as follows. Section 4.2 introduces the load specific gate linearization idea and show that it produces accurate results for arbitrary input waveforms and various loads. Section 4.3 explains the proposed gate characterization approach which enables efficient moment propagation. Section 4.4 then explains how the proposed methodology works for the stage and path delay calculation and propagation and show experimental results that confirm efficiency and accuracy of the proposed approach. Finally, the chapter is concluded in section 4.5.
4.2. Load-based linear gate model

This section illustrates the gate linearization idea. First, the gate current source model is linearized and it shows that using a single set of the coefficients of a linearized gate model across different loads would not produce accurate results. Instead, gate model linearization around a given load is shown to be necessary to accurately capture the gate output waveform. An efficient gate characterization methodology is used to produce 2D tables for the gate linearized parameters and input capacitances as functions of gate driving point admittance coefficients.

A current source based gate model [25], and [26] which includes a voltage-controlled current source as well as an intrinsic gate output capacitance is shown in Figure 4-1, together with an interconnect tree that the gate drives.

![Gate current source model driving a load interconnect.](image)

The current source model also includes a 2-D lookup table \( I(V_i, V_x) \) which gives the gate output current for a pair of gate input and output voltages. Similar to the work in [42] A linear interpolation to the gate current lookup table is applied and approximated \( I(V_i, V_x) \) as a linear function of \( V_i \) and \( V_x \).
\[ I(V_i, V_x) = a_0 + a_1 v_i(t) + a_2 v_x(t) \] (4-1)

The gate output signal is then derived based on the linearized model coefficients \( a_0, a_1, \) and \( a_2, \) and the input moments to reconstruct the gate output waveform. The waveforms, obtained for four different loads are compared to SPICE simulations of the nonlinear gates. As can be observed in Figure 4-2, using only a single pre-computed set of the coefficients \( a_0, a_1 \) and \( a_2 \) to approximate the behavior of the gate under four different loading conditions leads to very inaccurate results.

Figure 4-2. Gate output waveforms using a gate model with a single set of pre-computed linear model coefficients
It is clear from Figure 4-2 that a CMOS gate is not unconditionally linear. More appropriate, however, would be to assume that the gate is conditionally linear in the neighborhood of a specific load. In other words, each load should have a separate set of coefficients $a_0$, $a_1$ and $a_2$ in order for the linear assumption to be valid. These sets of coefficients $a_0$, $a_1$ and $a_2$ are the ones that would have the same gate output signal moments ($m_1^x$, $m_2^x$, and $m_3^x$, for example) as the original gate when it drives the same interconnect.

It should be pointed out that characterizing the gate with the linear model coefficients is equivalent to characterizing it with the output waveform moments, and from now on throughout this chapter the gate is characterized based on the output moments.

To show that this piece-wise linearization approach has a merit we characterized the gate for various loads. The first three output gate moments $m_1^x$, $m_2^x$, and $m_3^x$ (or any number of moments, whatsoever), can be computed using a single SPICE simulation for each load. The loading effect of the interconnect is approximated using the coefficients of the driving point admittance [31]. Similarly, the gate input capacitance was characterized for various loads, again as a function of the driving point admittance coefficients. Thus, a gate model consists of the lookup tables which give the first three output gate moments and input gate capacitance for each set of load parameters. Using the proposed gate model we obtained very accurate output gate waveforms, shown on Figure 4-3.
Figure 4-3. Gate outputs for piece-wise linear gate approximation

To show that the proposed linearization works for arbitrary inputs, we compared the gate output waveforms obtained for the various input waveforms with SPICE simulations of the non-linear gates. Our waveforms (in black) were obtained by convolution of the inputs and the pre-characterized gate. The gate input (red) and output (blue) waveforms are shown in Figure 4-4.
4.3. Gate Characterization

As briefly described in the previous section, the gate characterization methodology we propose produces accurate results with 2-D tables, where the two variables are only the first two coefficients of the gate driving point admittance, $y_1$ and $y_2$. These tables contain the gate output impulse response moments as well as the gate input capacitance.

The setup used in characterizing for the moments is shown in Figure 4-5.

![Figure 4-5. The setup used in the moment characterization](image)
It consists of the gate under test denoted by $G$ and a typical interconnect. A typical load with capacitance $C_L$ is also attached to the interconnect. The effect of the interconnect together with the load, $C_L$, on the gate behavior is accurately represented by the first two admittance coefficients $y_1$ and $y_2$ i.e., the gate output response to different interconnect structures will be similar provided that these different interconnect structures have the same $y_1$ and $y_2$.

The gate under test is driven by a step input, $u(t)$, so that the response at the gate output, $v_u(t)$, is the gate unit step response that corresponds to the set of admittance coefficients $y_1$ and $y_2$. The gate output impulse response, $h(t)$, can be obtained by directly differentiating $v_u(t)$ with respect to time i.e.,

$$h(t) = \frac{\partial v_u(t)}{\partial t} \quad (4-2)$$

Any desirable number of the gate output impulse response moments can be derived by numerical integration of $h(t)$. The Laplace transform $H(s)$ of $h(t)$ is

$$H(s) = L[h(t)] = \int_{0}^{\infty} h(t)e^{-st}dt \quad (4-3)$$

Expanding the exponent in (4-3) around $s=0$ results in

$$H(s) = m_0^H + m_1^H s + m_2^H s^2 + \ldots = \int_{0}^{\infty} h(t) \left(1 - st + \frac{1}{2!} s^2 t^2 + \ldots \right) dt. \quad (4-4)$$

Since $h(t) = 0$ for $t > t_n$, (4-4) can be rewritten as

$$H(s) = \sum_{i=0}^{n} m_i^H s^i = \sum_{i=0}^{n} \frac{(-1)^i s^i}{i!} \int_{0}^{t_n} t^i v_u^{(i)}(t) dt. \quad (4-5)$$
Thus, the formula for the impulse response moments can be given by

\[ m_i^H = \frac{(-1)^i}{i!} \int_0^{t_n} t^i h(t) dt. \]  

(4-6)

The SPICE produced waveform for \( v_u(t) \) can approximated by piecewise linear, PWL, segments as shown in Figure 4-6.

Each piece \( v_{uk}(t) \) is defined by

\[ v_{uk}(t) = a_k \cdot t + b_k. \]  

(4-7)

Substituting in (4-2) then (4-6) yields

\[ m_i^H = \frac{(-1)^i}{i!} \sum_{k=1}^{n} a_k (t_{i+1}^k - t_{i}^k) \]  

(4-8)

As will be shown later, the gate output response to any non-step input such as ramps or exponentials can be obtained by doing simple moment multiplication.
The Taylor series expansion around $s=0$ of the frequency domain representation of any time domain input $x(t)$ that approaches a constant value asymptotically is given by

$$X(s) = \frac{m_0^X}{s} + m_1^X s + m_2^X s^2 + \ldots + m_q^X s^{q-1}$$  (4-9)

When the input $X(s)$ is applied to a circuit with impulse response given by $H(s)$, the response $R(s)$ is given by

$$R(s) = X(s)H(s) = \sum_{i=0}^{q} m_i s^{i-1}$$

where $m_i = \sum_{i=0}^{q} m_k^i m_{i-k}^H$  (4-10)

The gate input capacitance exhibits non trivial variations with varying the gate output load. The intuition behind this behavior is that varying the gate output load leads to significant variations in the gate output rising and falling times which in turn varies the Miller Coupling Factor of the coupling capacitance between the gate input and output. Thus, the gate input capacitance needs also to be characterized using the gate output admittance coefficients.

Gate input capacitance characterization can be done using the setup shown in Figure 4-7.

![Figure 4-7. The setup used in characterizing for the gate input capacitance](image-url)
For each set of $y_1$-$y_2$ the first moment of the signal $s(t)$, $m_1^s$, is computed using a single SPICE simulation and applying the same technique described earlier in this section. Knowing $m_1^s$, the input capacitance $C_{in}$ can be computed by [26]

$$C_{in} = \frac{m_1^s}{R} \quad (4-11)$$

Hence, the result of the gate characterization procedure is a 2-D lookup table for each gate. The variables of the lookup table are the admittance coefficients $y_1$-$y_2$. With each selection set of $y_1$-$y_2$, the lookup table gives the gate impulse response moments, $m_i^H$'s, together with the gate input capacitance, $C_{in}$, that corresponds to the selection set $y_1$-$y_2$.

The equation (4-10) can be used to compute the gate output response moments when any non-step input is applied.

4.4. Path Delay Propagation

In Section 4-2 we showed that the idea of linearization in the neighborhood of a given gate load is valid and produces accurate results (Figure 4-3 and Figure 4-4). In this section, we will present how the methodology works for stage and path calculations and demonstrate, through several examples, its accuracy, efficiency, and flexibility.

The gate characterization methodology we propose produces accurate results with 2-D tables, where the two variables are only the first two coefficients of the gate driving point admittance. Any desirable number of the gate output impulse response moments can be derived by numerical integration of SPICE produced output waveform and get even more accurate waveform at the gate output.
More accurate gate input capacitance characterization can also be done by using a simple network instead of a single capacitance, but again with a penalty of more storage resources. The tables composed in terms of the two gate driving point admittance coefficients are sufficient for accurate waveform reproduction for arbitrary $RC$ dominated interconnects.

In order to show the flexibility of the approach, we show how we can easily make the tradeoff between performance/storage and accuracy. Namely, the input waveform with a pronounced plateau can also cause nonlinearity at the gate output, which can be accurately reproduced by just storing and reading more moments. In Figure 4-8, we show how, going from the top left to the bottom right, the accuracy of the output waveform can be improved by increasing number of moments used. The spice simulation is shown using dashed blue lines, and our waveforms are in black.

![Figure 4-8. Gate output waveforms for various number of moments used in their approximation](image-url)
For a stage waveform propagation, where the stage is defined as a path segment from the input of the gate to the input of the next one down the path, we have to first calculate the gate driving point admittance, using the algorithm originally presented in [31] and read the gate output moments from the 2D table using the first two admittance coefficients as inputs. We then convolute the moments obtained from the table with the moments of the gate input waveform and with the interconnect moments calculated. Both the driving point admittance and the interconnect moments will be calculated using the input capacitances of the gates at the end of the interconnect, obtained also using corresponding 2D tables, with the inputs to these tables being the first two admittance coefficients of the load of the driven gate.

The results, for typical tree type interconnect networks are shown on Figure 4-9.

Figure 4-9. The waveforms at the end of a single stage for various tree interconnects
In order to propagate the worst arc delay, the delay approximation based on the calculated output moments as in [40,] and [41] can be effectively used in this methodology.

The algorithm for the path waveform propagation, or delay and ramp time calculation, is shown bellow.

**Algorithm**

1. Starting with the leafs of the last stage, calculate gate driving point admittance using algorithm in [29] for RC dominated interconnect or [33] for RLC interconnect

2. Using the first two coefficient of the corresponding driving point admittance read the values for the gate input capacitance from the pre-characterized table.

3. Proceed towards the root of the path by again calculating the driving point admittances for the stages using the gate input capacitances obtained in step 2

4. Continue till the FF at the starting point of the path is reached. At this point, all the driving point admittances are calculated accurately including the miller capacitance effect.

5. Calculate all the interconnect transfer function moments along the path.

6. Beginning at the input of the FF, start the forward traversal. Multiply the interconnect moments with the gate output moments obtained from a 2-D table and the first two coefficients of the driving point admittance of the gate calculated above.

7. Continue the procedure till the end of the path is reached
8. **Reconstruct the waveform at the end of the path using AWE, or some other similar procedure, to convert the obtained moments to a time domain signal**

9. **Read delay or perform the other needed measurement of the waveform**

As can be seen from the algorithm, one first needs to calculate all the gates driving point admittances starting from the end of the path and going towards the root. This is needed since the input gate capacitances depend on the gate’s loads. Note that traditional static timing analysis methodologies do not take this important effect into account and, consequently, might introduce very significant errors. Once the driving point admittances and the input capacitances are accurately determined, the forward traversal of the paths is needed to calculate the interconnect moments, to read the output gate moments and to propagate (convolute) them in order to get the moments at the output.

Figure 4-10 shows the waveforms obtained, for four different interconnects, at the end of a 5-stage path by propagating moments to the circuit output using our methodology.
The novel approach was examined using different CMOS gates. A lot of simulations were made for each gate when driving various interconnect structures and different gates and being driven by different input waveforms. Table 4-1 shows the average and maximum error in the computed moments at the gate output versus the actual signal moments calculated using SPICE.
### Table 4-1. Numerical results

<table>
<thead>
<tr>
<th></th>
<th>Avg % error</th>
<th>Max. % error</th>
</tr>
</thead>
<tbody>
<tr>
<td>INV_small</td>
<td>2%</td>
<td>3%</td>
</tr>
<tr>
<td>INV_normal</td>
<td>1.6%</td>
<td>2.1%</td>
</tr>
<tr>
<td>INV_large</td>
<td>1%</td>
<td>1.3%</td>
</tr>
<tr>
<td>NAND2_small</td>
<td>4.3%</td>
<td>5.2%</td>
</tr>
<tr>
<td>NAND2_normal</td>
<td>3.1%</td>
<td>5.1%</td>
</tr>
<tr>
<td>NAND2_large</td>
<td>2.7%</td>
<td>4.0%</td>
</tr>
<tr>
<td>NOR2_small</td>
<td>2.2%</td>
<td>3.3%</td>
</tr>
<tr>
<td>NOR2_normal</td>
<td>2.15%</td>
<td>3.13%</td>
</tr>
<tr>
<td>NOR2_large</td>
<td>1.86%</td>
<td>1.95%</td>
</tr>
<tr>
<td>XOR2_small</td>
<td>2.5%</td>
<td>3.4%</td>
</tr>
<tr>
<td>XOR2_normal</td>
<td>2.3%</td>
<td>3.3%</td>
</tr>
<tr>
<td>XOR2_large</td>
<td>2.0%</td>
<td>3.2%</td>
</tr>
</tbody>
</table>

### 4.5. Conclusions

An accurate and efficient methodology for static timing analysis was presented. The methodology is based on the assumption that the gate can be linearized around a given load and characterized for the output moments and input gate capacitance for various loads expressed in terms of a gate driving point admittances. The linearization ad the proper characterization would then allow the application of convolution (moment multiplication) and effective waveform propagation through the network. Accuracy can be controlled by the appropriate number of
moments. We have demonstrated the accuracy, flexibility and efficacy of the algorithm using it on various typical examples for a stage and path calculation.

The methodology has advantages compared to the classical, delay and ramp time based approaches since it is more accurate and more efficient. It is also much more efficient than the recently proposed current source based approaches, because those approaches have to pay high performance price in order to achieve the desired accuracy.

We also believe that this novel approach, which allows for uniform representation of the gate and interconnect waveforms in terms of moments, would open the opportunities for various improvements in static timing analysis, static noise analysis, performance corner calculation, cell characterization and similar applications.
CHAPTER 5

IMPORTANCE OF VOLUME DISCRETIZATION OF SINGLE AND COUPLED INTERCONNECTS

This chapter presents figures of merit to determine which interconnects require volume discretization in the GHZ range. Most of the previous work focused mainly on efficient modeling of volume discretized interconnects using several integration and reduction techniques. However, little work has been done to characterize when using the simple DC model has an impact on critical circuit metrics such as delay, impedance, etc. Most of the previous work simply assumes that when skin depth becomes smaller than the wire cross section dimensions, volume discretization becomes essential. However, careful analysis in this chapter shows that this assumption is invalid and a figure of merit is derived to characterize when volume discretization of single and coupled wires is required. This derived figure of merit is shown to depend solely on the interconnect dimensions and spacing and is independent of the type of the materials used or technology scaling.

5.1. Background

A lot of work has been done to 3D model interconnects efficiently. Several models aim at finding inductance and resistance of interconnects as functions of frequency, e.g., [43]-[44]. Although these models are accurate, they are difficult to use with most available simulators [44]. Several other models try to find frequency independent lumped-element circuits to replace the
original frequency dependent elements. Among these models are the volume filament model [47], compact circuit models [48], and the reduced decoupled R-L model [49]. For example, Figure 5-1 shows the volume filament model and its equivalent circuit. As can be seen, volume discretization makes circuit simulation very expensive. In general, it is very expensive to include the interconnects 3D models in circuit simulations even after using several reduction techniques. Therefore, it is very important to know when 3D models result in significantly different circuit behavior as compared to using the DC model.

![Rectangular interconnect 3D model.](image)

**Figure 5-1. Rectangular interconnect 3D model.**

Most of the previous work assumes that using the 3D model leads to significant changes in the circuit performance when the operating frequency is high enough such that the skin depth starts to be comparable to the interconnect cross section dimensions. In such cases, the interconnect effective cross section area becomes less than that of the geometrical cross section as shown in Figure 5-2. This condition can be expressed mathematically as:

\[
\Delta < 0.5 \min(w, t), \tag{5-1}
\]
Where $\Delta$ is the skin depth and is given by

$$\Delta = \frac{1}{\sqrt{\pi f \mu \sigma}},$$

(5-2)

The critical frequency above which volume discretization becomes mandatory is given by:

$$f = \frac{4}{\min(w,l)^2 \mu \pi \sigma},$$

(5-3)

Based on (5-3), the frequency boundary after which the 3D model should be used is 4 GHz for an interconnect width of 1 $\mu$m. Signal harmonicas in current technologies far exceed 4 GHz and wires wider than 1 $\mu$m are frequent in power and clock distribution networks and in global interconnect. Therefore, this metric has led to the conclusion that in current technologies, including volume discretization for on chip interconnects is essential.

![Figure 5-2 Effective cross sectional area of a rectangular interconnect under skin effect](image)

This conclusion, despite being widely used, is not correct. This boundary was chosen based on the fact that when skin depth starts to be less than the interconnect cross section, the resistance starts to increase significantly and the DC model is expected to be erroneous. However, at such
high frequencies the inductive impedance dominates the total impedance and inductance is practically constant independent of frequency.

The analysis presented in this chapter starts by deriving a figure of merit that characterizes when volume discretization of the wires is required. This figure of merit is derived based on the behavior of both the resistive and inductive impedances over the entire frequency range. The study shows that simply using the DC model of the wire leads to minimal error at certain wires dimensions and spacing.

The rest of the chapter is organized as follows. The qualitative behavior of interconnect impedance at high frequency is studied in section 5.2. The derivation of the figure of merit that characterizes the importance of volume discretization for a single wire and the simulation results that verifies the proposed figure of merit are presented in section 5.3. The single wire case is studied in details to gain intuitive understanding of the behavior of filamented interconnects. Section 5.4 presents the derivation of a modified figure of merit that includes the effect of coupling and loop inductance. The simulation results that verify the figure of merit are also presented in section 5.4. The Formulations that quantify the error introduced by using DC model instead of an accurate model with volume filamentations are presented in section 5.5. Finally, the chapter is concluded in section 5.6

5.2. Qualitative Behavior of Interconnect Impedance under Skin Effect

The behavior of the resistance and inductance with frequency for a rectangular interconnect is studied based on the volume 3D model shown in Figure 5-1. The interconnect is divided into
such that the cross section dimension of each filament is significantly smaller than the skin depth. This discretization ensures that the current is constant within each filament.

The simulations results for the resistance and inductance of a typical interconnect are shown in Figure 5-3. These results match well the resistance and inductance formulae derived in [52]-[56]. Note that the resistance changes significantly while the inductance is practically constant over the entire frequency range.

In the low frequency region where the skin depth is larger than the wire cross section, both the inductance and the resistance are practically constant, and equal to their DC values. At this frequency region, the resistance dominates the impedance but has a very low rate of change, which means that skin effect is negligible at this frequency region. Thereafter, at higher frequencies the skin depth becomes comparable to the wire cross section, the current starts to concentrate along the perimeter of the wire cross section and the resistance thereby increases. This current concentration also modifies the magnetic field in the space between the conductors.
and within each conductor resulting in a slow decay of the total inductance. At higher frequencies, the inductive impedance starts to dominate the total impedance as shown in Figure 5-4, and the inductance becomes practically constant. At even higher frequencies the resistance increases as the square root of frequency while the inductive impedance increases as frequency. Thus, the inductive impedance further dominates the total impedance.

![Figure 5-4. The behavior of $R(f)$ and $\omega L(f)$ up to f= 100 GHZ](image)

The skin impedance can be described in terms of two frequency points as shown in Figure 5-5. The first frequency point $\omega_1$ is the frequency at which the inductance starts to dominate the total impedance. The second frequency point $\omega_2$ is the frequency at which the resistance starts to increase with a high rate (when skin depth becomes comparable to the wire cross section dimensions).
Both $\omega_1$ and $\omega_2$ are functions of the interconnect dimensions. In cases when $\omega_2$ is greater than $\omega_1$, the impedance is always dominated by a slowly varying element with frequency, i.e., the impedance dominant term is always the one whose DC model has minimal error. This behavior occurs because the frequency region at which the resistance starts to increase at a high rate has the inductive impedance as the dominant factor.

5.3. Importance of Volume discretization for Single Wire

In this section, the figure of merit that characterizes when volume discretization becomes important for a single is derived and the experimental results are shown to verify its correctness. Section 3.3.1 shows that for a single wire case, the DC model of an interconnect can be used with minimal error if the ratio between the interconnect length, $l$, to the summation of its width, $w$, and thickness, $t$, is greater than 7. Section 3.3.2 presents delay error introduced in using the DC model versus the 3D model for different values of $\frac{l}{w+t}$. Section 3.3.3 explores the error in
the total impedance $Z_{DC}$ versus $Z_{3D}(f)$ of an interconnect having a cross section of constant width and thickness while having a different $\frac{l}{w+t}$. The effect of scaling the interconnect dimensions, while having a constant $\frac{l}{w+t}$ is studied in section 3.3.4. Section 3.3.5 shows impact of changing the aspect ratio of the interconnect cross section on the total error. The effect of varying physical constants such as the conductivity and magnetic permeability on the introduced error is studied in section 3.3.6.

5.3.1. Volume discretization Figure of Merit for Single Wire

Volume discretization is considered to be of minimal importance under the interconnect dimension conditions that make $\omega_2 >> \omega_1$. Based on simulation results and theoretical analysis, as will be shown in subsequent sections, less than 5% error is guaranteed under the interconnect dimensions conditions that makes $\omega_R > 5 \omega_L$. The frequency point $\omega_1$ occurs when the resistance value equals that of the inductive impedance as shown in Figure 5-5. Thus, $\omega_1$ is given by:

$$\omega_1 = \frac{R(f)}{L(f)}, \quad (5-4)$$

If $\omega_2$ is greater than $\omega_1$, this intersection point occurs when the resistance is almost at its DC value as shown in Figure 5-5. Note also that inductance is almost constant. Hence, the DC values of the resistance and inductance are used in (5-4). The DC value of the resistance is given by

$$R_{DC} = \frac{l}{\sigma wt}, \quad (5-5)$$

where $\sigma$ is the wire conductivity. The DC inductance is [49] -[53]
\[ L_{DC} \approx \frac{\mu}{2\pi} \left( \ln \left( \frac{2l}{w+t} \right) + 0.5 + 0.2235 \left( \frac{w+t}{l} \right) \right), \quad (5-6) \]

where \( \mu \) is the magnetic permeability of SiO\(_2\). Hence, \( \omega_1 \) is given by

\[ \omega_1 \approx \frac{l}{2\pi \sigma wt} \left( \ln \left( \frac{2l}{w+t} \right) + 0.5 + 0.2235 \left( \frac{w+t}{l} \right) \right), \quad (5-7) \]

Simplifying (5-7) yields,

\[ \omega_1 = \frac{2\pi}{\mu \sigma wt \left( \ln \left( \frac{l}{w+t} \right) + 1.2 + 0.2235 \left( \frac{w+t}{l} \right) \right)}. \quad (5-8) \]

The frequency point \( \omega_2 \) occurs when the skin depth is equal to the minimum of the interconnect width and thickness. Hence, \( \omega_2 \) can be calculated from,

\[ \frac{1}{\sqrt{\pi f \mu \sigma}} = 0.5 \min(w,t). \quad (5-9) \]

Hence

\[ \omega_2 = \frac{8}{\min(w,t)^2 \mu \sigma}, \quad (5-10) \]

Substituting in the condition \( \omega_2 > 5 \omega_1 \) results in

\[ \ln \left( \frac{l}{w+t} \right) + 0.2235 \left( \frac{w+t}{l} \right) \geq \frac{1.9 \times \min^2(w,t)}{w \times t} \quad (5-11) \]

Since \( \frac{\min^2(w,t)}{w \times t} \) is less than 1. Skin effect is negligible if

\[ \ln \left( \frac{l}{w+t} \right) + 0.2235 \left( \frac{w+t}{l} \right) \geq 1.9. \quad (5-12) \]
The minimum value of \( \frac{l}{w+t} \) that satisfies (5-12) can be obtained graphically by finding the minimum value \( u \) at which the function \( f(u) = \ln u + 0.2235 \left( \frac{1}{u} \right) \) exceeds 1.9 as shown in Figure 5-6.

![Figure 5-6. The minimum \( \frac{l}{w+t} \) that results in no skin effect](image)

The minimum value of \( \frac{l}{w+t} \) that satisfies (5-12) is approximately 7. Hence, using the DC model of a single wire gives minimal error at any frequency if:

\[
\frac{l}{w+t} \geq 7
\]  

(5-13)

However, to prove the uniqueness of this value, the function \( f(u) \) should be monotonically increasing for any \( u \geq 7 \). Differentiating \( f(u) \) yields,

\[
\frac{\partial f(u)}{\partial u} = \frac{1}{u} - \frac{0.2235}{u^2}
\]  

(5-14)
The condition at which the function \( f(u) \) is monotonically increasing is thus, \( \frac{\partial f(u)}{\partial u} > 0 \).

Substituting from (5-14) yields,

\[
\frac{1}{u} - \frac{0.2235}{u^2} > 0
\]

This condition is always satisfied for any \( u > 0.2235 \), which proves the uniqueness of the figure of merit defined in (5-13).

Note that the condition that \( \omega_2 \) is greater than \( \omega_1 \) can also be interpreted as the condition that makes the inductive impedance exceeds the resistance at the frequency point \( \omega_2 \), which is the point at which the resistance starts to increase.

\[
\omega_2 > \omega_1 \Leftrightarrow \omega_2 > \frac{R}{L} \Leftrightarrow \omega_2 L > R
\]  

(5-16)

### 5.3.2. Delay Error in Using the Interconnect DC Model Versus the 3D model

The experimental setup used in examining the impact of using interconnect DC model on delay is shown in Figure 5-7.

**Figure 5-7 Delay experimental setup**
The error in delay when an interconnect is modeled using the DC model instead of the complete 3D volume model is studied for different values of $\frac{l}{w+l}$. The interconnect shown in Figure 5-7 has $l=w=2 \, \mu m$, and the input source is a step input with slope of less than 0.01 $ns$, which includes very high frequency components. The output voltages at two interconnects having $\frac{l}{w+l}$ equals to 1 and 3 are shown in Figure 5-8(a) and (b), respectively, and show errors of at least 50% in the output signal delay for low $\frac{l}{w+l}$. Also, the 3D model behaves more as an $RC$ interconnect while the DC model tends to behave as an $RLC$ circuit. This behavior is due to the increase in the resistance value of the 3D model due to skin effect. However, this difference between the two models is negligible for $\frac{l}{w+l} > 7$ as shown in Figure 5-8(c) and (d). The delay error is less than 1% for $\frac{l}{w+l} > 7$. In addition, the DC model of the interconnect captures the signal waveform shape of the 3D model with negligible error.
Figure 5-8: Delay error between DC model and 3D model for different \( \frac{l}{w+l} \).

5.3.3. Error in Impedance Due to Using DC Model

Figure 5-9 (a) shows the complete 3D model of an interconnect, while Figure 5-9(b) shows the DC model of the same interconnect.
According to our figure of merit, skin effect is not important at any frequency when \( \frac{l}{w+t} \geq 7 \). When this condition is satisfied, the 3D impedance of the wire should be the same as the DC impedance at any frequency. The equivalence of these two impedances can be expressed as:

\[
\frac{Z_{3D}(f)}{Z_{DC}(f)} = \frac{\sqrt{R^2(f) + (\omega L(f))^2}}{\sqrt{R_{DC}^2 + (\omega L_{DC})^2}} \cong 1
\]

(5-17)

\[
\tan^{-1}\left(\frac{\omega L(f)}{R(f)}\right) - \tan^{-1}\left(\frac{\omega L_{DC}}{R_{DC}}\right) \cong 0
\]

(5-18)

Five cases are studied assuming \( w=t=2.5 \text{ um} \) in each case. The interconnect length is varied such that \( \frac{l}{w+t} \) varies between 10-40. Figure 5-10 shows the error in the impedance magnitude versus frequency. The phase error versus frequency is shown in Figure 5-11. It is clear from Figure 5-10 and Figure 5-11 that the entire frequency range can be divided into four main regions. In region I where \( \omega < \omega_1 \), the error increases monotonically at a slow rate. This increase is mainly due to the increase in the resistance value. In region II where \( \omega_1 < \omega < \omega_2 \), the inductance starts to dominate.

Figure 5-9. Complete 3D Model Versus the DC Model
the impedance, and the resistance is still changing slowly with frequency. Thus, the error rate, and eventually the error value decreases. In region III $\omega_2 < \omega < 2\omega_2$. This is the region with the highest error rate. In this region the rate of change of resistance with frequency attains its maximum. However, the error is still small because the ratio between the inductive impedance and the resistance increases at a higher rate. In region IV where $2\omega_2 < \omega$, the inductive impedance is by far dominating the total impedance, which leads to a significant decrease in the error rate since the inductance is very slowly varying with frequency.

![Figure 5-10. Percentage errors in impedance magnitude due to using the DC model](image-url)
5.3.4. Effect of Scaling the Interconnect Dimensions

The figure of merit introduced in section 3.3.1 predicts that scaling the interconnect dimensions does not affect the impact of skin effect as long as \( \frac{l}{w+t} \) remains constant. Table 5-1 shows the total impedance magnitude and phase percentage error in using the 3D model versus the DC model for different interconnects with square shape cross sections.

Table 5-1. Maximum error in impedance magnitude and phase (t=w)

<table>
<thead>
<tr>
<th></th>
<th>% error in magnitude</th>
<th>% error in phase</th>
</tr>
</thead>
<tbody>
<tr>
<td>w=</td>
<td>w=2µ</td>
<td>w=3µ</td>
</tr>
<tr>
<td>l=2.5(w+t)</td>
<td>9.1</td>
<td>9.0</td>
</tr>
<tr>
<td>l=5(w+t)</td>
<td>7.5</td>
<td>7.25</td>
</tr>
<tr>
<td>l=7.5(w+t)</td>
<td>6.8</td>
<td>6.4</td>
</tr>
<tr>
<td>l=10(w+t)</td>
<td>5.9</td>
<td>6.27</td>
</tr>
<tr>
<td>$l$</td>
<td>12.5(w+t)</td>
<td>15(w+t)</td>
</tr>
<tr>
<td>-----</td>
<td>------------</td>
<td>----------</td>
</tr>
<tr>
<td>$w$</td>
<td>5.6</td>
<td>5.59</td>
</tr>
<tr>
<td>$w$</td>
<td>5.4</td>
<td>5.43</td>
</tr>
<tr>
<td>$w$</td>
<td>5.2</td>
<td>5.22</td>
</tr>
<tr>
<td>$w$</td>
<td>4.9</td>
<td>4.88</td>
</tr>
<tr>
<td>$w$</td>
<td>4.8</td>
<td>4.8</td>
</tr>
<tr>
<td>$w$</td>
<td>4.7</td>
<td>4.7</td>
</tr>
<tr>
<td>$w$</td>
<td>3.4</td>
<td>3.4</td>
</tr>
</tbody>
</table>

The experimental results presented in Table 5-1 shows that the percentage error in both impedance magnitude and phase is not varying with scaling the interconnect dimensions while having constant $\frac{l}{w+t}$ which verifies the independence of the figure of merit of absolute interconnect dimensions.

Figure 5-12 The change in $\omega_1$ and $\omega_2$ with $\frac{l}{w+t}$
The physical explanation of this trend for a constant aspect ratio is that increasing the length, width, and thickness of the interconnect such that \( \frac{l}{w + t} \) is kept constant decreases the frequency point \( \omega_2 \) since the skin depth becomes comparable to the wire width at lower frequencies. However, this scaling in the wire dimensions decreases the resistance and at the same time increases the inductive impedance. Thus, the frequency point \( \omega_1 \) decreases as well. Figure 5-12 shows the behavior of the two frequency point with increasing the interconnect dimensions while having a constant \( \frac{l}{w + t} \).

Figure 5-13 shows the resulting \( \omega_1 \) and \( \omega_2 \) when applying the complete 3D model simulation to different interconnects having square shape cross section. Note that the intersection point is at \( \frac{l}{w + t} = 7 \) which agrees very well with our metric.

![Figure 5-13. The behavior of \( \omega_1 \) and \( \omega_2 \) for Different Interconnect Widths](image-url)
5.3.5. Effect of Interconnect Aspect Ratio

The interpretation of (5-11) is that changing the aspect ratio of an interconnect cross section while having a constant \( \frac{t}{w+t} \) decrease the error in both magnitude and phase.

### Table 5-2. Maximum errors due to using the DC model for different aspect ratios, \( w/l \)

<table>
<thead>
<tr>
<th>( l = )</th>
<th>( w=t )</th>
<th>( w=2t )</th>
<th>( w=3t )</th>
<th>( w=4t )</th>
<th>( w=t )</th>
<th>( w=2t )</th>
<th>( w=3t )</th>
<th>( w=4t )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( l=2.5(w+t) )</td>
<td>9.0</td>
<td>7.5</td>
<td>7</td>
<td>6.7</td>
<td>3.1</td>
<td>2.5</td>
<td>2</td>
<td>1.8</td>
</tr>
<tr>
<td>( l=5(w+t) )</td>
<td>7.5</td>
<td>6.2</td>
<td>5.8</td>
<td>5.6</td>
<td>2.5</td>
<td>2.0</td>
<td>1.9</td>
<td>1.7</td>
</tr>
<tr>
<td>( l=7.5(w+t) )</td>
<td>6.8</td>
<td>5.5</td>
<td>5.2</td>
<td>5.1</td>
<td>2.1</td>
<td>1.7</td>
<td>1.5</td>
<td>1.3</td>
</tr>
<tr>
<td>( l=10(w+t) )</td>
<td>5.9</td>
<td>5.1</td>
<td>4.9</td>
<td>4.7</td>
<td>1.8</td>
<td>1.5</td>
<td>1.4</td>
<td>1.2</td>
</tr>
<tr>
<td>( l=12.5(w+t) )</td>
<td>5.6</td>
<td>4.9</td>
<td>4.6</td>
<td>4.5</td>
<td>1.6</td>
<td>1.3</td>
<td>1.1</td>
<td>1.0</td>
</tr>
<tr>
<td>( l=15(w+t) )</td>
<td>5.4</td>
<td>4.7</td>
<td>4.4</td>
<td>4.3</td>
<td>1.55</td>
<td>1.1</td>
<td>1.0</td>
<td>0.95</td>
</tr>
<tr>
<td>( l=17.5(w+t) )</td>
<td>5.2</td>
<td>4.5</td>
<td>4.3</td>
<td>4.1</td>
<td>1.47</td>
<td>1</td>
<td>0.91</td>
<td>0.8</td>
</tr>
<tr>
<td>( l=20(w+t) )</td>
<td>4.9</td>
<td>4.2</td>
<td>4.0</td>
<td>3.9</td>
<td>1.44</td>
<td>1</td>
<td>0.9</td>
<td>0.8</td>
</tr>
<tr>
<td>( l=22.5(w+t) )</td>
<td>4.8</td>
<td>4.2</td>
<td>3.9</td>
<td>3.8</td>
<td>1.37</td>
<td>0.96</td>
<td>0.87</td>
<td>0.73</td>
</tr>
<tr>
<td>( l=25(w+t) )</td>
<td>4.7</td>
<td>4.1</td>
<td>3.9</td>
<td>3.8</td>
<td>1.3</td>
<td>0.96</td>
<td>0.88</td>
<td>0.73</td>
</tr>
<tr>
<td>( l=50(w+t) )</td>
<td>3.4</td>
<td>3.2</td>
<td>3.1</td>
<td>3.0</td>
<td>0.8</td>
<td>0.72</td>
<td>0.7</td>
<td>0.64</td>
</tr>
</tbody>
</table>
Table 5-2 shows that the error decreases with increasing the aspect ratio of the interconnect cross section. The explanation of this trend is that the frequency point $\omega_2$ is inversely proportional to $w \times t$ while the frequency point $\omega_1$ is inversely proportional to $\min^2(w,t)$. This means that increasing the aspect ratio leads to increasing the ratio between $\omega_1$ and $\omega_2$ which means that the ratio between $\omega_1$ and $\omega_2$ increases. Substituting for $\frac{\min^2(w,t)}{w \times t}$ with values other than 1 in (5-11) lead to different minimum value for $\frac{l}{w+t}$ to satisfy (5-11) as shown in Figure 5-14.

![Figure 5-14. Effect of increasing the aspect ratio on skin effect](image-url)

The behavior shown in Figure 5.14 shows that choosing $\frac{l}{w+t} > 7$ as the criteria based on which skin effect is considered negligible, is thus a safe upper bound.
5.3.6. Impact of Physical Constants

The interpretation of (5-11) is that changing the aspect ratio of an interconnect cross section while having a constant $\frac{l}{w+l}$ decrease the error in both magnitude and phase.

The effect of having different wires with different conductivity and/or magnetic permeability is also studied for constant $\frac{l}{w+l}$ and constant aspect ratio interconnects. Table 5-3 shows the magnitude and phase error for $w=t=1 \ \mu m, \ \frac{l}{w+l} = 10$ interconnect and having a conductivity that varies from $0.1\sigma_0$ to $10\sigma_0$, and magnetic permeability that varies from $0.1\mu_0$ to $10\mu_0$, where the values of $\sigma_0$ and $\mu_0$ are $3.5 \times 10^7$ and $12.5664 \times 10^{-7}$, respectively.

Table 5-3 Effect of changing $\sigma$ and $\mu$ on the impedance error

<table>
<thead>
<tr>
<th></th>
<th>% error in magnitude</th>
<th>% error in phase</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$0.1\sigma_0$</td>
<td>$\sigma_0$</td>
</tr>
<tr>
<td>$0.1\mu_0$</td>
<td>5.95</td>
<td>5.9</td>
</tr>
<tr>
<td>$\mu_0$</td>
<td>5.92</td>
<td>5.9</td>
</tr>
<tr>
<td>$5\mu_0$</td>
<td>6</td>
<td>5.97</td>
</tr>
<tr>
<td>$10\mu_0$</td>
<td>5.9</td>
<td>5.96</td>
</tr>
</tbody>
</table>

It is clear that the neither the magnitude error nor the phase changes with varying the conductivity or the magnetic permeability. The reason is that changing the conductivity and/or the magnetic permeability changes both $\omega_1$ and $\omega_2$ such that the ratio between the two frequencies remains the same for constant $\frac{l}{w+l}$ and constant aspect ratio.
5.4. **Importance of Volume discretization for Coupled Wire**

In this section, the figure of merit that characterizes when volume discretization becomes important for coupled wires is derived and the experimental results are shown to verify its correctness. Section 3.4.1 shows the impact of coupling on the interconnect circuit behavior. The modified figure of merit that includes the effect of inductive coupling in characterizing the importance of volume discretization is derived in section 3.4.2. Section 3.4.3 shows the experimental results that verifies the figure of merit.

**5.4.1. Impact of Inductive Coupling**

Inductive coupling results in transfer of energy from one circuit component to another through the shared magnetic field. Inductive coupling effect diminishes as distance between wires increases. In the extreme situation where wires are far apart, the inductive coupling between wires become negligible and the figure of merit that characterize the relevance of volume discretization derived in the previous section can be applied to each wire. However, measures must be taken to handle inductive coupling when it becomes significant.
Figure 5-15 Interconnects with Proximity Effects

Figure 5-15 illustrates two coupled interconnects by a mutual inductance $M$. The equivalent circuit of the two interconnects shown in Figure 5-15(a) is illustrated in Figure 5-15(b). Because of coupling, the equation of each circuit contains a term depending on the current change in the other circuit as shown in (5-19).

$$v_1 = i_1 R + L \frac{\partial i_1}{\partial t} + M \frac{\partial i_2}{\partial t}$$
$$v_2 = i_2 R + L \frac{\partial i_2}{\partial t} + M \frac{\partial i_1}{\partial t}$$

Thus, (5-19) can be analyzed based on three different possible cases of wire switching.

Case 1
In this case it is assumed that there is no change in \( i_2 \), while \( i_1 \) is following a unit step change in the voltage source \( v_1 \). In this case (5-19) can be rewritten as:

\[
\begin{align*}
\frac{\partial i_1}{\partial t} + M \frac{\partial i_1}{\partial t} &= v_1 - i_1 R + L \frac{\partial i_1}{\partial t} + M \times 0 \\
\frac{\partial i_1}{\partial t} + M \frac{\partial i_1}{\partial t} &= v_2 - i_2 R + L \times 0 + M \frac{\partial i_1}{\partial t}
\end{align*}
\]  

(5-20)

This case simply falls back to a single wire case and the previous figure of merit is valid

**Case 2**

In this case it is assumed that the two wires switch similarly, resulting in a change in both \( i_2 \) and \( i_1 \) in the same direction. In this case (5-19) can be rewritten as:

\[
\begin{align*}
\frac{\partial i_1}{\partial t} + (L + M) \frac{\partial i_1}{\partial t} &= v_1 - i_1 R + (L + M) \frac{\partial i_1}{\partial t} \\
\frac{\partial i_2}{\partial t} + (L + M) \frac{\partial i_2}{\partial t} &= v_2 - i_2 R + (L + M) \frac{\partial i_2}{\partial t}
\end{align*}
\]  

(5-21)

For this case, using the DC model for the wires results in even less error than the single wire case. This behavior is due to the increase in the inductance value which leads to a corresponding decrease in the frequency point \( \omega_p \). Thus the figure of merit in (5-13) is also valid as an upper bound for this case

**Case 3**

In this case, that the two wires switch oppositely, resulting in a change in both \( i_1 \) and \( i_2 \) but in opposite directions. In this case (5-19) can be rewritten as:

\[
\begin{align*}
\frac{\partial i_1}{\partial t} + (L - M) \frac{\partial i_1}{\partial t} &= v_1 - i_1 R + (L - M) \frac{\partial i_1}{\partial t} \\
\frac{\partial i_2}{\partial t} + (L - M) \frac{\partial i_2}{\partial t} &= v_2 - i_2 R + (L - M) \frac{\partial i_2}{\partial t}
\end{align*}
\]  

(5-22)
This case results in a decrease in the total inductance value which leads to an increase in the frequency point \( \omega_1 \). Hence, the figure of merit for a single wire should be modified to include the effect the mutual inductance as will be shown in the next subsection. Note that this case is equivalent to using the loop inductance.

5.4.2. Volume Discretization Figure of Merit in Case of Loop Inductance (case 3)

The figure of merit that includes the inductive coupling between adjacent wires is derived in this section. This criteria depends solely on the interconnect dimensions and spacing.

Inductive coupling affects the volume discretization figure of merit derived in the previous section in the case of oppositely switching interconnects. The total inductance is reduced to \( L-M \).

Thus, the frequency point \( \omega_f \) should be redefined as:

\[
\omega_1 = \frac{R(f)}{L(f) - M(f)}
\]  \hspace{1cm} (5-23)

Denoting the distance between the center axes of the interconnects by \( d \), \( M \) is given by [49]-[53]

\[
M \approx \frac{\mu}{2\pi l} \left[ \ln \left( \frac{l}{d} + \sqrt{1 + \frac{l^2}{d^2}} \right) - \sqrt{\left( 1 + \frac{d^2}{l^2} \right)^2 + \frac{d}{l}} \right]
\]  \hspace{1cm} (5-24)

Substituting in the condition \( \omega_2 > \omega_1 \) with \( \omega_2 \) as given by (5-10), results in

\[
g(\alpha, \beta) = \ln(\alpha) + \frac{0.235}{\alpha} \ln \left( \beta + \sqrt{1 + \beta^2} \right) + \sqrt{1 + \frac{1}{\beta^2} - \frac{1}{\beta} - 1.9 \frac{\min^2(w,t)}{wt}} \geq 0
\]  \hspace{1cm} (5-25)

where \( \alpha = \frac{l}{w + l} \) and \( \beta = \frac{l}{d} \). Figure 5-16 shows the plot of (5-25) at \( \frac{\min^2(w,t)}{wt} = 1 \), for different values of \( \alpha \) and \( \beta \).
It is shown in Figure 5-16 that there exist certain interconnect dimensions and spacing at which the interconnect DC model can be used with high accuracy. The values of \( \frac{d}{t} \) above which the complete 3D model is not needed for various \( \frac{l}{w+t} \) and aspect ratios are shown in Table 5-4.

**Table 5-4 The values of \( \frac{d}{t} \) above which the 3D model is not needed**

<table>
<thead>
<tr>
<th>( l/w ) Ratio</th>
<th>1:1</th>
<th>1:2</th>
<th>1:3</th>
<th>1:4</th>
<th>1:5</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.5</td>
<td>0.1662</td>
<td>0.142</td>
<td>0.125</td>
<td>0.1</td>
</tr>
<tr>
<td>50</td>
<td>0.041</td>
<td>0.0333</td>
<td>0.026</td>
<td>0.0227</td>
<td>0.02</td>
</tr>
<tr>
<td>100</td>
<td>0.05</td>
<td>0.0182</td>
<td>0.013</td>
<td>0.011</td>
<td>0.01</td>
</tr>
<tr>
<td>200</td>
<td>0.014</td>
<td>0.01</td>
<td>0.0067</td>
<td>0.0059</td>
<td>0.005</td>
</tr>
<tr>
<td>300</td>
<td>0.008</td>
<td>0.0067</td>
<td>0.0045</td>
<td>0.0038</td>
<td>0.003</td>
</tr>
</tbody>
</table>
It can be depicted from the results shown in Table 5-4 that as $\frac{l}{w+t}$ increases, $d$ can get smaller without the need for filamentation. This behavior is because increasing $\frac{l}{w+t}$ leads to a corresponding increase in the value of the self inductance $L$ which in turn gives more space for the mutual inductance $M$ to increase while still satisfying (5-25). Another observation from the results in Table 5-4 is that as the aspect ratio deviates from 1, $d$ is allowed to get smaller. This is intuitive since as the aspect ratio decreases the term $\frac{\min(2(w,t))}{wt}$ decreases and hence the ratio $\frac{d}{l}$ can get smaller without the need for filamentation while still satisfying (5-25).

The general figure of merit that determines the importance of volume discretization in case of prominent inductive coupling can be obtained by applying curve fitting to the experimental results shown in Table 5-4 and is given by

$$\frac{l}{w+t} - 0.83 \frac{l}{d} > 0.0027 \frac{7.9 \min(2(w,t))}{wt}$$  \hspace{1cm} (5-26)$$

In the worst case when the aspect ratio is 1, the figure of merit is given by

$$\frac{l}{w+t} - 0.83 \frac{l}{d} > 7$$  \hspace{1cm} (5-27)$$
5.4.3. Volume Discretization Figure of Merit in Case of Loop Inductance (case 3)

The experimental setup used in examining the delay error that results when using the DC model versus the 3D model is shown in Figure 5-17.

The error in delay is studied for different values of $l$, $w$, $t$ and $d$. Figure 5-18(a) shows the error in delay at aspect ratio $\frac{l}{w+t}=10$ and $\frac{d}{l}=0.0625$. This error is significantly reduced when the aspect ratio changes from 1 to 0.5 as shown in Figure 5-18 (b). The reduction in delay error when increasing the wire spacing, $d$, and increasing $\frac{l}{w+t}$ are shown in Figure 5-18 (c) and Figure 5-18(d), respectively. Figure 5-18 shows very small error even at the edge of the figure of merit. This is because of the capacitive effect which filters out the high frequency components. Hence, our figure of merit is conservative.
Figure 5-18 Delay error between DC model and 3D model for different wire dimensions
5.5. Error Formulation

The error that might arise in using the DC model of an interconnect is mainly due to the change of the resistance and mutual inductance from their DC values. Thus, there are two types of errors, the resistance error, $E_R$, and the mutual inductance error, $E_M$. The resistance error arises due to the change in the effective value of the interconnect cross sectional area at frequencies higher than $\omega_R$. Thus, the resistance error is only valid for operating frequencies $\omega_0 > \omega_R$ and can be given by

\[
E_R = \left( \frac{R(f) - R_{\text{DC}}}{R(f)} \right) = 1 - \frac{A_{\text{eff}}}{A_{\text{DC}}} 
= 1 - \frac{2\Delta(w + t - 2\Delta)}{wt}
\]  

(5-28)

Substituting from (5-2) into (5-28) yields

\[
E_R \approx 1 - 2 \left( 1 - \frac{1}{2} \sqrt{\frac{\omega_R}{\omega_0}} \right) \sqrt{\frac{\omega_R}{\omega_0}}
\]

(5-29)

The error as presented in (5-29) is a tight upper bound assuming the resistance is completely dominating the total impedance. A modification of (5-29) that takes into consideration the contribution of the inductive impedance is

\[
\hat{E}_R \approx \left[ 1 - 2 \left( 1 - \frac{1}{2} \sqrt{\frac{\omega_R}{\omega_0}} \right) \sqrt{\frac{\omega_R}{\omega_0}} \right] \frac{R}{\sqrt{R^2 + \left( \omega_0 L_{\text{eff}} \right)^2}}
\]

(5-30)

The mutual inductance error arises from the change in the effective distance between the centre axes of the interconnects, $d$, from being $s+w$ at low frequencies to approximately $s+\Delta$ at high
frequencies, where $s$ is the interconnect-to-interconnect spacing. Hence, the mutual inductance error, $E_M$, can be given by

$$E_M = \frac{M(f) - M_{pc}}{M(f)} \approx 1 - \frac{\ln\left(\frac{2l}{s + w}\right)}{\ln\left(\frac{2l}{s + \Delta}\right)}$$

(5-31)

Substituting from (5-2) into (5-31) yields

$$E_M = \frac{1}{\ln\left(\frac{2l}{d}\right)} \ln\left(\frac{s + w}{s + \Delta}\right)$$

(5-32)

A modification of (5-32) that takes into consideration the contribution of the other circuit elements in the total impedance is

$$\hat{E}_M = \frac{1}{\ln\left(\frac{2l}{d}\right)} \ln\left(\frac{s + w}{s + \frac{1}{2}w \sqrt{\frac{\omega_R}{\omega_0}}}\right)$$

(5-33)

It can be deduced that an upper bound for the total error due to skin effect can be given by

$$\hat{E}_{skin} = \sqrt{\hat{E}_R^2 + \hat{E}_M^2}$$

(5-34)

It has to be mentioned here that (5-34) can be used to estimate the error in using the DC model of an interconnect whether or not the figure of merit is satisfied.

The figures of merit insure that at frequencies above $\omega_R$ when (5-30) becomes applicable; the value of the inductive impedance is at least 5 times that of the resistance which in turn results in
negligible values for the resistance error at any frequency. In addition, for (5-30) to get satisfied the ratio between $d$ and $w+t$ should be

$$d > 0.83(w+t)$$

(5-35)

This condition sets a minimum value for $s$ to be comparable to the interconnect cross section dimension. Also the figure of merit in (5-30) sets a minimum value for the ration between the interconnect length and its cross section dimensions. With these limitations, the error in (5-33) will also be negligible at any frequency. An example of $s=w=t$, the figure of merit in (5-30) deduce that $l$ should be at least 70 times higher than the interconnect cross sectional dimensions. Substituting by these values in (5-33) would give an error that is less than 10% at frequencies higher than 100 GHZ. Hence, the satisfaction of the figures of merit implies that the total error in (5-34) is negligible.

5.6. Conclusions

The chapter characterized the relevance of volume discretization in the GHZ range. It is shown that comparing the skin depth to the interconnect cross section dimensions cannot solely identify when to use volume discretization. A figure of merit that characterizes the importance of volume discretization for a single wire was then derived based on both the frequency point at which the skin depth starts to be comparable to the interconnect cross section dimensions and the frequency at which the resistance starts to have negligible impact on the total impedance. The experimental results verified the figure of merit in terms of both the signal delay error and total interconnect impedance error. Moreover, it was shown that the variation of other parameters such as scaling of interconnect dimensions and having different physical constants do not have any impact on the introduced error which agrees with the prediction of the figure of merit. The impact of
coupling on determining the importance of volume discretization is studied. A modified figure of merit was also derived that includes the inductive coupling effect. The experimental results also verified the modified figure of merit. Finally, error formulae that quantifies the error were also presented.
CHAPTER 6

Variable Threshold Voltage Design Scheme for CMOS Tapered Buffers

This chapter proposes a low power-low delay design for CMOS tapered buffers. A slight increase in the threshold voltage is shown to have an exponential effect in reducing the total power dissipation. The corresponding increase in the propagation delay is compensated for by increasing the number of buffer stages such that there is still an overall significant reduction in the total power dissipation. As compared to the constant threshold voltage design based on a cost function of \( PT^2 \), the proposed scheme can lead to either a power dissipation reduction of about 70% while maintaining the same delay, or up to 30% in power dissipation with 10% propagation delay reduction, respectively. Closed form expressions that give the optimum threshold voltage and number of stages are presented.

6.1. Introduction

For gigascale integration (GSI) networks, both the number and length of global interconnects rapidly increase. This has led to the fact that the power consumption of on chip busses is about 40% of the total power dissipated by the whole chip. In deep submicron (DSM) technologies, 60% of the interconnect power is short circuit and leakage power with only 40% dynamic power [57]- [59]
Historically, the supply voltage ($V_{DD}$) has been scaled down in order to keep the dynamic power consumption under control. Hence, the transistor threshold voltage ($v_{th}$) has also been scaled down to maintain a high drive current and achieve performance improvement. However, the threshold voltage scaling results in an exponential increase in the leakage current and hence the leakage power [66]. In addition, decreasing the threshold voltage causes a substantial increase in the short circuit power. Thus, the minimization of the on-chip interconnect power has become a dominant part of the effort to contain the increase in the overall circuit power consumption.

Unfortunately, trading speed for low power is not possible anymore. This fact is basically due to the increase in the demand for high-speed, computationally efficient applications, especially in portable systems like cellular phones and personal digital assistances (PDAs). The development of new design techniques becomes crucial to maximize the use of deep submicron technologies while maintaining an acceptable power consumption level.

In this chapter, we introduce a new design scheme for CMOS tapered buffers. Allowing the threshold voltage to increase over $0.2V_{DD}$ reveals a global optimum design point which can have less total power dissipation and less propagation delay. The total power dissipation is much more sensitive to the threshold voltage variations than the propagation delay, especially in the neighborhood of the $v_{th}=0.2V_{DD}$ point. This behavior can lead to a significant reduction in the power dissipation with minimal increase in the propagation delay. In addition, the propagation delay penalty can be compensated for by a slight increase in the number of buffers. Thus, an overall saving in both power dissipation and propagation delay can be achieved as compared to a cost function based on $PT^2$ metric. Attaining a decrease in both the power dissipation and
The propagation delay is made possible due to the fact that the power dissipation is much less sensitive to the variation in the number of stages than the propagation delay [63].

The rest of the chapter is organized as follows. Section 6.2 gives a background on the tapered buffer design. A qualitative analysis on constant threshold voltage design scheme is presented in section 6.3. The new variable threshold voltage design scheme is introduced in section 6.4. Section 6.5 discusses the effects of the proposed scheme on other performance criteria such as total area and reliability. Finally, the chapter is concluded in section 6.6.

6.2. Background

In CMOS integrated circuits, large capacitive loads are often encountered, both on chip and off chip. In order to drive these large capacitive loads at high speed, buffer circuits are required which must quickly source and sink relatively large current while not degrading the performance of previous stages.

The tapered buffer structure was first introduced by Lin and Linholm, in 1975 [60]. This structure consists of a series of inverters where each transistor channel width is $\beta$ times larger than that of the previous inverter. The output current drive to output capacitance ratio remains fixed for each stage in the buffer, giving each inverter stage equal rise, fall, and delay times.

Thus, using the split-capacitor model of a tapered buffer system shown in Figure 6-1, as denoted by Li, Haviland, and Tuszyński [61] the optimum tapering factor, $\beta_{opt}$, and the optimum number of stages, $N_D$, that minimize the propagation delay can be given by [61].

$$N_D = \frac{\ln\left(\frac{C_L}{C_g}\right)}{\ln(\beta_{opt})}$$

(6-1)
\[ \beta_{\text{opt}} \left[ \ln(\beta_{\text{opt}}) - 1 \right] = \frac{C_0}{C_g} \]  

(6-2)

**Figure 6-1** Split-capacitor model of tapered buffer

The Li split-capacitor model was used in [62] with the \( \alpha \)-power short-channel transistor model resulting in the following propagation delay through a tapered buffer system

\[ T_D = h(N) \times g(v_i) \]  

(6-3)

where

\[ h(N) = N \left( C_0 + C_g \left( \frac{C_I}{C_g} \right)^{1/N} \right) V_{DD} \]  

and

(6-4)

\[ g(v_{ih}) = \frac{1}{I_{D_0}} \left[ \left( \frac{9}{8} + \frac{V_{D_h}}{0.8V_{DD}} \right) \ln \left( \frac{10V_{D_h}}{eV_{DD}} \right) \left( 0.5 - \frac{1 - v_i}{1 + \alpha} \right) + 0.5 \right] \]  

(6-5)
In the alpha-power model, $I_{D0}$ is the saturation current at $V_{GS} = V_{DD}$, $V_{D0}$ is the drain saturation voltage at $V_{GS} = V_{DD}$, and $\alpha$ is the velocity saturation index. Approximated formulae describing both $V_{D0}$ and $I_{D0}$ were derived in [63] and are given by:

$$V_{D0} = \frac{2v_{sat}L_0}{\mu_0 \left(1 + \frac{C_{D0}}{C_{ox}}\right)} \left(V_{GS} - V_{th}\right) \left(1 + \left(V_{GS} - V_{th}\right) \left(\frac{\mu_0}{4.4 \times 10^7 t_{ox}}\right)\right)$$ (6-6)

$$I_{D0} = W_0 C_{ox} v_{sat} (V_{DD} - V_{th})$$ (6-7)

where $\mu_0$ is the electron mobility, $L_0$ is the minimum channel length, $W_0$ is the minimum channel width, $v_{sat}$ is the saturation velocity, $C_{D0}$ is the depletion layer capacitance, and $t_{ox}$ is the oxide thickness.

The power dissipation in a CMOS tapered buffer has got three main components. These components are dynamic power, $P_{dyn}$, short circuit power, $P_{S.C.}$, and leakage power, $P_l$. The total power dissipation is thus

$$P_t = P_{dyn} + P_{S.C.} + P_l$$ (6-8)

The total dynamic power dissipation for a tapered buffer is given by [62], [64]

$$P_{dyn} = V_{DD} f \frac{f}{N} d(N) h(N)$$ (6-9)

where $f$ is the operating frequency, $N$ is the number of stages and

$$d(N) = \left(\frac{C_L}{C_s} - 1\right) \left(\left(\frac{C_L}{C_s}\right)^{\frac{1}{IN}} - 1\right)$$ (6-10)

The short circuit power is given by [62], [65]
\[ P_{sc} = P_{dyt}(v_{th}) \]  

where

\[
s(v_{th}) = \left[ \frac{9}{8} + \frac{V_{th}}{V_{DD}} \ln \left( \frac{10V_{th}}{eV_{DD}} \right) \right] \times \frac{1}{\alpha + 1} \frac{1}{2^{\alpha-1} (1-v_i)^{\alpha+1}} \]  

The leakage power of the buffers is given by

\[ P_{leak} = l(v_{th})d(N)V_{DD} \]

where

\[ l(v_{th}) = \mu_0 C_{Dh} V_{th}^2 \times e^{C_{ox}(V_{th} - V_{DD})/C_{Dh}} \times \left[ 1 - e^{-V_{th}/V_T} \right] \]

and \( V_T \) is the thermal voltage.

### 6.3. Constant \( V_{th} \) Tapered Buffer Design

One strategy to include the effect of power dissipation in buffer design is to have a cost function, \( F \), defined as

\[ F = P^k_p \times T^k_t \]

where \( k_p \) is the power dissipation weighting exponent and \( k_t \) is the propagation delay weighting exponent. The values of these parameters determine how much weight is placed on propagation delay over total power dissipation and vice versa. The most common choice that compromise the relative importance of power dissipation over propagation delay, is \( k_p=1 \) and \( k_t=2 \) [67]. The constant threshold voltage tapered buffer design can therefore be translated to the problem of finding the number of stages that minimizes \( (6-15) \) at \( v_{th}=0.2V_{DD} \).
The effect of varying the number of stages on the propagation delay, for various $C_L/C_g$ while having a constant threshold voltage, is shown in Figure 6-2.

From the shape of Figure 6-2, it is clear that the propagation delay dramatically increases when using a value of $N$ smaller than $N_D$. Thereafter, the propagation delay slightly increases for $N$ greater than $N_D$.

The effect of varying the number of stages on the power dissipation for various $C_L/C_g$ while having a constant threshold voltage is shown in Figure 6-3.
Figure 6-3. Normalized total power as function of $C_L/C_g$

Unlike Figure 6-2, the total power dissipation graph shown in Figure 6-3 depicts no local minima. The shape of the graph demonstrates a steady increase in power dissipation for increasing values of $N$. Most importantly, the absolute rate of change of power dissipation with the number of stages is much less than that of the propagation delay.

In Figure 6-4, the power delay square product for various $C_L/C_g$ ratios is plotted. From this graph it is shown that $N_{opt}$, which represent the number of stages that minimizes the power delay squared product, is typically significantly less than $N_D$. 
Figure 6-4: Normalized power delay square product

The percentage of increase in propagation delay and decrease in total power dissipation when using $N_{opt}$ or $N_{opt} + 1$ versus $N_D$ are shown in Table 6-1.

Table 6-1: Effects of using $N = N_{opt}$ & $N = N_{opt} + 1$ on power and delay

<table>
<thead>
<tr>
<th>$C_L/C_g$</th>
<th>$N_D$</th>
<th>$N_{opt}$</th>
<th>$N = N_{opt}$</th>
<th>$N = N_{opt} + 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>% increase in delay</td>
<td>% decrease in power</td>
</tr>
<tr>
<td>$&lt;10^2$</td>
<td>4</td>
<td>&lt;3</td>
<td>&gt;40%</td>
<td>&gt;60%</td>
</tr>
<tr>
<td>$10^3$</td>
<td>5</td>
<td>3</td>
<td>37.44%</td>
<td>62.54%</td>
</tr>
<tr>
<td>$10^4$</td>
<td>7</td>
<td>4</td>
<td>38.4%</td>
<td>64%</td>
</tr>
<tr>
<td>$10^6$</td>
<td>9</td>
<td>5</td>
<td>39.44%</td>
<td>65.54%</td>
</tr>
<tr>
<td>$10^7$</td>
<td>11</td>
<td>6</td>
<td>39.4%</td>
<td>67.5%</td>
</tr>
</tbody>
</table>
Table 6-1 shows that decreasing the number of stages from $N_D$ to $N_{opt}$ cuts off about 70% of the total power but at the expense of the propagation delay which increases by about 38%. However, increasing the number of stages slightly above $N_{opt}$ leads to a substantial decrease in the propagation delay with only a slight increase in the power dissipation as compared to $N_{opt}$. This behavior is due to the fact that the total power dissipation is less sensitive to the total number of stages than the propagation delay for $N<N_D$ as shown in Figure 6-2 and Figure 6-3.

Applying curve fitting, the optimum number of stages for $V_{th}=0.2V_{DD}$ can be approximated by

$$N_{opt} = \left\lfloor N_D \left(1 - e^{-0.3 \left(\frac{I}{k_f}\right)}\right) - 1 \right\rfloor$$

(6-16)

6.4. **Variable $V_{th}$ Tapered Buffer Design**

The global optimum design point can be achieved by allowing the threshold voltage to vary. As will be shown later in this section, choosing the threshold voltage as another variable is mainly because the propagation delay is much less sensitive to the $V_{th}$ variation than the power dissipation.

Note in (6-1)(6-1) that $h(N)$ and $g(V_{th})$ are two independent functions such that $h(N)$ only expresses the effect of the total number of tapered buffers $N$ on the propagation delay while the effect of varying the threshold voltage is completely characterized by $g(V_{th})$. The effect of varying $V_{th}$ on $g(V_{th})$ for various $C_L/C_g$ ratios is shown in Figure 6-5.
From the shape of the graph shown in Figure 6-5, the effect of increasing the threshold voltage between $0.2V_{DD}$ and $0.4V_{DD}$ causes the propagation delay to increase with a very small rate. Thereafter, the propagation delay increases quickly as $v_{th}$ goes beyond $0.4V_{DD}$. The explanation of this behavior can be made clear by approximating $\frac{g(v_{th})}{g(0)}$ as

$$
g(v_{th})/g(0) = \begin{cases} 
\frac{k_1}{V_{DD}-V_{th}} \cdot \frac{-\ln(V_{DD}-V_{th})}{V_{DD}-V_{th}} & v_{th} < 0.4V_{DD} \\
\frac{1}{(V_{DD}-V_{th})} \cdot \frac{\ln(V_{DD}-V_{th})}{V_{DD}-V_{th}} & v_{th} > 0.4V_{DD}
\end{cases} \quad (6-17)
$$

where $k_1$ is a technology dependent constant and equals to 1.7 for 65 nm technology. It is clear from (6-17) that the slope of $g(v_{th})$ varies significantly as $v_{th}$ exceeds $0.4V_{DD}$. As with the propagation delay expression, $s(v_{th})$ characterizes the effect of varying the threshold voltage on the short circuit power and $l(v_{th})$ characterizes the effect of varying the threshold voltage on
leakage power. The effects of varying $v_{th}$ on $s(v_{th})$ and $l(v_{th})$ are shown in Figure 6-6 and Figure 6-7, respectively.

**Figure 6-6: The behavior of $s(v_{th})$ with varying $v_{th}$**

**Figure 6-7: The behavior of $l(v_{th})$ with varying $v_{th}$**
Unlike Figure 6-5, in which $g(v_{th})$ has a small slope in the range between $0.2V_{DD}$ and $0.4V_{DD}$. Figure 6-6 and Figure 6-7 show that the power dissipation decreases with a high rate in the threshold voltage range between $0.2V_{DD}$ and $0.4V_{DD}$. Thereafter, increasing the threshold voltage beyond $0.4V_{DD}$ seems to have negligible effects on minimizing the total power dissipation. The behavior of $l(v_{th})$ is intuitive due to the exponential dependence of $l(v_{th})$ on $v_{th}$ as shown in (6-14).

On the other hand $s(v_{th})$ in (6-12) varies as $(V_{DD}-v_{th})ln(V_{DD}-v_{th})$ for $0<v_{th}<0.5V_{DD}$.

The previous behavior of the propagation delay and power with varying the threshold voltage in the range between $0.2V_{DD}$ and $0.4V_{DD}$ gives the highest reduction in power with a minimal penalty in delay.

Plotting the power delay square product in (6-15) for $C_l/C_g=10^5$ is shown in Figure 6-8.

![Figure 6-8. Normalized power delay squared product](image)

From the shape of the graph shown in Figure 6-8, it is clear that the cost function exhibits an upward concavity. Therefore, an optimum number of stages and optimum threshold voltage
exists for any $C_L/C_g$. Another interesting property is that the number of stages that gives better solution than that of $V_{th}=0.2V_{DD}$ lies in the range $N_{opt}<N<N_{D}$. That’s there is no threshold voltage that gives a lower value for $F$ at either $N<N_{opt}$ or $N>N_{D}$. The intuition behind this fact is that increasing the number of stages beyond $N_{D}$ leads to an increase in both the propagation delay and power dissipation. On the other hand, the propagation delay penalty incurred by reducing the number of stages below $N_{opt}$ is not mitigated by a substantial reduction in power dissipation, and therefore, minimal incentive exists to decrease $N$ below $N_{opt}$.

The percentage of increase in the propagation delay and decrease in the total power dissipation are depicted in Table 6-2 for $N = N_{opt}+1$ and $N=N_{D}-1$. For each choice of the number of the stages there exists a single $V_{th}$ value that minimizes the cost function at this value of $N$.

Table 6-2 shows that the closer the number of stages to $N_{opt}$ the higher the reduction in power dissipation. On the other hand, the reduction in propagation delay reaches its maximum as the number of stages gets closer to $N_{D}$.

**Table 6-2 Effects of varying $V_{th}$ on delay and power**

<table>
<thead>
<tr>
<th>$C_L/C_g$</th>
<th>$N=N_{opt}+1$</th>
<th>$N=N_{D}-1$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$N$</td>
<td>$V_{th_{opt}}$</td>
</tr>
<tr>
<td>$10^3$</td>
<td>$&lt;3$</td>
<td>0.35</td>
</tr>
<tr>
<td>$10^4$</td>
<td>3</td>
<td>0.35</td>
</tr>
<tr>
<td>$10^4$</td>
<td>4</td>
<td>0.36</td>
</tr>
</tbody>
</table>
It is clear that the results of Table 6-2 outperform those of Table 6-1. Thus, allowing the threshold voltage to vary, results in a decrease in both power dissipation and propagation delay compared to the $v_{th}=0.2V_{DD}$ case. Moreover, as the weight of the power dissipation, $k_p$, increases, it becomes more crucial to allow $v_{th}$ to vary.

The expression that gives the threshold voltage that minimizes the cost function in (6-15) as a function of the number of stages, $C_L/C_g$ ratio, and the weighting factors, $k_p$ and $k_t$, can be approximated as

$$v_{th_{opt}} = \left(0.35 - 0.07 \frac{N - N_{opt} - 1}{N_D - N_{opt}} \right) \left[1 + \ln \left( \frac{k_t}{2k_p} \right) \right]$$

(6-18)

where $N_D$ and $N_{opt}$ are given by (6-1) and(6-16), respectively. Finally, it should mentioned that the percentage increase in propagation delay and the percentage decrease in the total power dissipation almost follow a linear behavior as the number of stages changes between $N_{opt}+1$ and $N_D-1$. So, based on Table 6-2 the expected increase in delay and decrease in power dissipation over the $N=N_D$ case can be calculated for any number of stages.
6.5. Variable \( V_{th} \) Tapered Buffer Design Scheme Effects on Other Performance Criteria

The global optimum design point can be achieved by allowing the threshold voltage to vary. As will be shown later in this section, choosing the threshold voltage as another variable is mainly because the propagation delay is much less sensitive to the \( V_{th} \) variation than the power dissipation. The variable \( V_{th} \) design scheme was shown to have the tendency of having higher threshold voltages than \( 0.2V_{DD} \) and higher number of stages than \( N_{opt} \). Increasing the number of stages has a positive effect on the average bond breaking current density. The average bond breaking current density in a transistor is a measure of hot carrier degradation experienced by transistor. For a CMOS tapered buffer, the average bond-breaking current density can be given by [62]:

\[
\langle J_{BB} \rangle = f \left( C_0 + C_g \left( \frac{C_L}{C_g} \right)^{1/N} \right) \langle J_{BB_0} \rangle \tag{6-19}
\]

where \( \langle J_{BB_0} \rangle \) describes the average bond breaking current density of a saturated minimum size transistor. It is thus clear that the average bond breaking exponentially decreases with increasing the number, which in turn means that the expected device lifetime increases exponentially. Nevertheless, the increase in the threshold voltage decreases \( \langle J_{BB_0} \rangle \) tremendously.

The only drawback in utilizing the variable \( V_{th} \) scheme over a \( PT^2 \) cost function with constant \( V_{th} \) is the increase in area. However, in current DSM technologies, the number of transistors per chip
is not a limiting factor which makes this issue of minimal importance with respect to other benefits.

6.6. Conclusions

This chapter characterized the effect of varying the threshold voltage on the design for CMOS tapered buffers. A qualitative behavior study showed that a slight increase in the threshold voltage has a substantial effect in reducing the total power dissipation with only a minimal propagation delay penalty. The corresponding increase in the propagation delay is compensated for by increasing the number of buffers such that there is still an overall significant reduction in the power dissipation. Further increase in the number of buffers results in a less delay and less power. As compared to the constant threshold voltage design, the proposed scheme led a power dissipation reduction of about 70% while having the same propagation delay and up to 30% and 10% reductions in power dissipation and propagation delay, respectively. Approximate formulae that give the threshold voltage and optimum number of stages were obtained by curve fitting and match the simulation results. Finally, the effects on other performance criteria were also studied.
CHAPTER 7

CONCLUSIONS

In this dissertation, we discussed some of the key issues in timing analysis and optimizations for high performance integrated circuits.

A new technique of time shifted moment matching (TSMM) which allows accurate estimation of $RLC$ circuit response at no extra cost was introduced. The TSMM technique performs moment matching (for expansion around $s=0$) on a time-shifted version of the original signal. As compared to other well-known techniques (such as AWE, [59]), TSMM offers distinct advantages. The 50% delay and rise time are determined with much more accuracy for a given approximation order. Moreover, the solutions have significantly improved accuracy as compared to AWE, especially for moderate to highly inductive signals.

Including the inductance effects in static timing analysis, while preserving the conventional gate representation, was also introduced. Based on a generalized driving point admittance and a generalized waveform shape assumption, the effective load Capacitances expressions of $RLC$ interconnects were derived accurately to estimate both the propagation delay and transition time at the output of a CMOS gate. The new effective capacitance calculation technique poses no extra complexity as compared to the $RC$ based approaches but can accommodate inductance. Simulation results showed that the error in propagation delays and rise times when neglecting inductance can be over 4-5 times as compared to the proposed approach.
Chapter 4 introduces a novel model for CMOS gates which is more accurate and at the same
time more efficient than the existing models. Unifying the timing analysis environment for both
the interconnects and the gate became possible for the first time by linearizing the gate around a
given load. This approach allows for moment propagation and uniform treatment of the gates and
interconnects. It is shown that despite the highly nonlinear overall gate model, a linearized gate
model with the model parameters as functions of the load is very accurate. A gate and input
capacitance characterization is also proposed which provides for accuracy, efficiency and
flexibility in the path performance calculation. This method is an order to two orders of
magnitude faster then current source based one, while it maintains accuracy within 5% of SPICE.
The importance of including the skin and proximity effects in current interconnect models is
judged by introducing much tighter figures of merit that characterized the importance of skin and
proximity effects. The introduced figures of merit characterized when 3D models result in
significantly different circuit behavior as compared to using the simple DC model.
Finally, the issue of increased power dissipation in current CMOS circuits such as the CMOS
tapered buffers was resolved by introducing a variable threshold voltage design scheme which
relaxed the power-delay tradeoff and allowed for an enhancement in both criteria. As compared
to the constant threshold voltage design, the proposed scheme led a power dissipation reduction
of about 70% while having the same propagation delay and up to 30% and 10% reductions in
power dissipation and propagation delay, respectively.
CHAPTER 8

FUTURE RESEARCH

It has been shown throughout this dissertation that when VLSI technology advances from one generation to the next, design methodologies and simulation tools need to be developed to cope with the posed challenges due to the increase in frequency, increase in number and length of interconnects, increase in number of transistors, …etc. This in turn has led to the increase in the importance of both inductance effects and skin effects for interconnect wires, on one hand, and the increasing size of circuits to be dynamically simulated on the other hand. Also, power dissipation became one of the biggest obstacles in the way of the advance in VLSI technology. Thus, future research in these directions includes:

1. **Applying the Novel Gate Model for Accurate Performance Corners Estimation and Statistical Waveform Analysis.**

The problem of finding the best and worst case delays, i.e., the performance corners has gained a lot of attention during the past few years. It was shown in [70] that the performance corners do not necessarily coincide with the process corners. For example Figure 8-1 shows an example of a stage delay simulated for various interconnect widths where $p_{min}$ and $p_{max}$ in case 1 through case 3 represent the process corners for three different scenarios.
It is clear from Figure 8-1 that the performance corners will coincide with the process corners in case 1 only. However, case 2 and case 3 show that assuming that the performance corners coincide with the process corners lead to erroneous results.

The novel CMOS gates representation shown in Chapter 4 can be utilized to accurately estimate the performance corners at any node throughout any interconnect path. It was shown in Chapter 4 that for each output load, the gate can be accurately represented by the first few moments. Thus, the signal moments at any node in the interconnect path can easily extracted by doing simple moment multiplication. Knowing that the propagation delay can be directly related to the first three signal moments, for example, by [71]

\[
    t_{50\%} = -1.76 m_1 + 0.292 \frac{m_2}{m_1} + 0.0911 \frac{m_3}{m_1^2} \tag{8-1}
\]
Finding the maxima and minima of (8-1) with respect to the varying parameters can lead to accurate estimation of the performance corners.

Nevertheless, the CMOS gate novel representation can lead to the migration from statistical timing analysis to statistical waveform analysis. Propagating the probability density function of the signal moments propagates all the possible variations that can take place within the signal. This can lead into far more accurate results comparing to simply propagating the probability density function of the 50% delay only.

2. Decouple Power Ground Mesh

Efficient power distribution network analysis and design is a key factor in the performance of modern VLSI circuits. Power grids are usually modeled as a network of resistors, capacitors, and mutually coupled inductors. Load current sources are modeled as current sources connected at nodes where vias connect the power grid metal wires running cross each other [72] as shown in Figure 8-2
Using only resistors or even resistors and capacitors does not accurately capture the supply voltage variations and hence is not accurate in the analysis of power grids in advanced technologies where the supply voltage is very low and the supply voltage variations are more pronounced. However, the power grid complete model that involves $R, L, M, C$ is very difficult to analyze mainly due to the exponential number of mutual inductances, $M$. This difficulty led to complicated algorithms and methodologies to design and verify the power grid network with excessive runtime and computations. Thus, decoupling the power lines is of utmost importance to achieve both performance and accuracy. Assuming the return current in the nearest few neighbors and applying loop inductance approximation can serve in inductance decoupling. This approximation is getting very close to reality with the current increase in frequencies.
Similarly, a valid approximation is to assume that the power lines are capacitively coupled to the nearest neighbors only. These approximations are based on the locality nature of the power grid that has been proved in the literature.

3. System in Package Interconnect Analysis

System in a Package is a semiconductor device that incorporates multiple chips that make up a complete electronic system into a single package. Electronic devices like mobile phones conventionally consist of several individually packaged IC’s handling different functions, e.g., logic circuits for information processing, memory for storing information, and I/O circuits for information exchange with the outside world. In a System-in-a-Package, all of these individual chips are assembled into a single package, allowing tremendous space savings and significant down-sizing of electronic gadgets. The ability to take existing chips to come up with a totally new system in a single package has one clear advantage: it drastically reduces development time and risk to bring new products to the market more quickly. With SiP technology, vendors are able to cram multiple flash devices, SRAMs, DRAMs, microcontrollers, ASICs, DSPs, and passive components into very thin packages that can fit into sleeker, more stylish, and yet more complex electronic gadgets. SiP’s also simplify the process of assembling the final application module by requiring simpler PCB lay-outs, since the complex interconnections required by the system are being taken care of inside the SiP [73].

The challenge in SIP manufacturing lies in the assembly process itself. It requires the ability to assemble and interconnect several dies both vertically and horizontally. The average length of the interconnect lines from chip to chip is about several centimeters, depending on the die size
and the number of dies integrated. To achieve high speed propagation, thicker dielectric and metal lines needed instead of thin film interconnections, so that we do have

- \( LC \) transmission rather than \( RC \) charging lines which require special analysis, model order reduction techniques, and design criterias.
- Significant skin effect behaviors which require the application of tight figures of merit and using compact models for representation.
- Significant thermal issues which requires accurate analysis and design.
- High crosstalk level at the required interconnect-density which requires accurate analysis and design.

### 4. Application of the Variable Threshold Voltage Design Scheme to other Circuits

The effect of varying the threshold voltage on the design for CMOS tapered buffers was characterized in this dissertation. A qualitative behavior study showed that a slight increase in the threshold voltage has a substantial effect in reducing the total power dissipation with only a minimal propagation delay penalty. The corresponding increase in the propagation delay was compensated for by increasing the number of buffers such that there is still an overall significant reduction in the power dissipation. Further increase in the number of buffers resulted in a less delay and less power. Thus, appealing as it seems, this design scheme is required to be applied to other circuits such as parallel tapered buffers, interconnect repeaters, buffered latches, …etc.
BIBLIOGRAPHY


[14] S. Nassif and Z. Li, “A more effective $C_{eff}$” in Proc of the Sixth International Symposium on Quality Electronic Design (ISQED’05).


[71] Yehea I. Ismail, Chirayu S. Amin, “Computation of signal threshold crossing times directly from higher order moments,” *ICCAD* pp. 246-253 Nov. 2004


PUBLICATIONS

Refereed Journal Publications


Submitted Journal Paper


Refereed Conference Publications


Submitted Conference Papers