

Design and Modeling of High Speed Global On-Chip Interconnects

by

Guoqing Chen

Submitted in Partial Fulfillment
of the
Requirements for the Degree
Doctor of Philosophy

Supervised by
Professor Eby G. Friedman

Department of Electrical and Computer Engineering
The College
School of Engineering and Applied Sciences

University of Rochester
Rochester, New York

2007

Dedication

This work is dedicated to my parents, Mr. Xiusheng Chen and Mrs. Yalan Zhang, my wife Ning, and my daughter Ariana.

Curriculum Vitae

Guoqing Chen was born in Beijing, China in 1975. He received the B.S. (with honors) and M.S degrees in electronic engineering from Tsinghua University, Beijing, China, in 1998 and 2001, respectively. In 2002, he received his second M.S. degree in electrical engineering from University of Rochester, Rochester, NY. He is currently a Ph.D. candidate in the area of high performance VLSI/IC design at the University of Rochester.

In the summers of 2004 and 2005, he was with Manhattan Routing, Inc., New York City, NY, where his work was focused on the development of an EDA tool for the timing closure procedure in the IC design process. His research interests include high speed interconnect design and modeling, signal integrity, and low power circuit design.

Acknowledgments

The time I spent at the University of Rochester is definitely one of the most important and cherishable periods in my life. During these years of studies, I received invaluable support and encouragement from my colleagues, friends, and family, for whom I would like to make grateful acknowledgments.

First of all, I would like to express my deep appreciation to my academic advisor, Professor Eby G. Friedman, for his kind and patient mentorship on my academic and personal growth. His vast knowledge and nice personality made my research work successful and highly enjoyable. It is my great pleasure to perform the Ph.D. studies under his supervision.

I thank Professors Mitsunori Ogihara, Philippe M. Fauchet, David H. Albonesi, and Daniel Stefankovic for their service on my committee and valuable suggestions for this dissertation. Special thanks to Professors Philippe M. Fauchet, David H. Albonesi, and other members in the Nanoscale Interdisciplinary Research Team, Mikhail Haurylau, Hui Chen, Nicholas A. Nelson, and Jidong Zhang for their collaborations

and support in the on-chip optical interconnect project. I'm grateful to Tor Ekenberg and Erhan Ergin for providing me the opportunity of summer internship in Manhattan Routing, Inc., where I obtained real industrial experience and enjoyed the colorful life in New York City.

I would like to thank those previous and current members in the High Performance VLSI/IC Design and Analysis Lab, Andrey Mezhiba, Dimitris Velenis, Volkan Kursun, Boris Andreev, Weize Xu, Magdy El-Moursy, Junmou Zhang, Mikhail Popovich, Vasilis Pavlidis, Jonathan Rosenfeld, Emre Salman, Renatas Jakushokas, and Ioannis Savidis, for their help and accompanyship. I would also like to thank RuthAnn Williams for her support in preparing all the paperwork in these years.

Finally, I would like to express my great gratitude to my wife Ning and my daughter Ariana for their support to my work and the happiness they bring to my life. This gratitude is also extended to my family, friends, and relatives in China for their understanding and encouragement which accompanies me through my life.

This work is supported in part by the Semiconductor Research Corporation under Contract No. 2003-TJ-1068 and 2004-TJ-1207, the National Science Foundation under Contract Nos. CCR-0304574 and CCF-0541206, grants from the New York State Office of Science, Technology & Academic Research to the Center for Advanced Technology in Electronic Imaging Systems, and by grants from Intel Corporation, Eastman Kodak Company, Manhattan Routing, and Intrinsix Corporation.

Abstract

Interconnect has become a dominant factor in deep submicrometer (DSM) integrated circuits (ICs). With increasing levels of on-chip integration, more functional units are integrated onto a single die, such as a multi-core microprocessor and a system-on-chip. Global interconnect, which acts as a communication media among these functional units, plays an increasingly important role and can significantly limit the performance of advanced systems.

With decreasing on-chip clock periods, the timing characteristics of on-chip signals need to be determined and controlled more precisely. Accurate interconnect models are therefore critical to the IC design process. In this dissertation, two global interconnect models are presented. Closed-form expressions of the signal waveform are developed, which achieve good agreement with Spectre simulations.

During the interconnect design process, multiple design criteria are considered, such as delay, power, bandwidth, and noise. Repeaters are widely used in digital ICs to reduce interconnect delay and signal transition time with the penalty of additional

power and area. A repeater insertion methodology is presented for achieving a tradeoff among different design criteria. Closed-form expressions for the number and size of the power optimal repeaters are developed.

With the scaling of CMOS technology, the requirements of different design criteria have become more stringent. It is increasingly difficult for conventional copper interconnect to satisfy these requirements. On-chip optical interconnect is shown to be a promising substitute for electrical interconnect in future advanced architectures. Critical lengths at which optical interconnect becomes advantageous are shown to be approximately one tenth of the chip edge length at the 22 nm technology node.

The focus of the IC design process in the DSM regime has shifted from logic optimization to interconnect optimization. The research presented in this dissertation provides several interconnect design and modeling methods to support this interconnect-centric design strategy.

Contents

Dedication	ii
Curriculum Vitae	iii
Acknowledgments	iv
Abstract	vi
List of Tables	xiii
List of Figures	xv
1 Introduction	1
2 On-Chip Electrical Interconnect	9
2.1 Design Flows for DSM ASICs	10
2.2 Interconnect Design Criteria	12
2.2.1 Delay	13
2.2.2 Power Dissipation	14
2.2.3 Noise	15
2.2.4 Bandwidth	16
2.2.5 Physical Area	18
2.3 Interconnect Characteristics	18
2.3.1 Resistance	19
2.3.2 Capacitance	24

2.3.3	Inductance	26
2.4	Interconnect Models	30
2.4.1	Single Interconnect	30
2.4.2	Parallel Coupled Interconnects	39
2.4.3	Model Order Reduction	42
2.5	Design Methodologies for Interconnect	46
2.5.1	Constructing an Interconnect Tree	47
2.5.2	Wire Sizing, Shaping, and Spacing	48
2.5.3	Repeater Insertion	50
2.5.4	Shielding Techniques	53
2.5.5	Net-Ordering and Wire Swizzling	53
2.6	Conclusions	55
3	An <i>RLC</i> Interconnect Model Based on Fourier Analysis	56
3.1	Introduction	56
3.2	Single Interconnect Model	58
3.2.1	Interconnect Transfer Function	59
3.2.2	Fourier Series Representation of Input Signal	62
3.2.3	Far End Time Domain Response	64
3.2.4	The 50% Delay and Overshoots/Undershoots	66
3.2.5	Model Verification and Discussion	71
3.3	Distributed <i>RLC</i> Trees	79
3.3.1	Transfer Function of Distributed <i>RLC</i> Trees	79
3.3.2	Examples	81
3.4	Multiple Coupled Interconnect Lines	87
3.4.1	Decoupling Multiconductor Systems	88
3.4.2	Far End Response	91
3.4.3	Model Verification and Discussion	95
3.5	Conclusions	98

4	Transient Response of a Distributed RLC Interconnect Based on Direct Pole Extraction	99
4.1	Introduction	99
4.2	Special Cases of a Single Interconnect System	101
4.2.1	RC interconnect	102
4.2.2	RLC interconnect with a zero R_d	106
4.2.3	Step and Ramp Response	109
4.3	Distributed RLC Interconnect with Driver Resistance	113
4.3.1	System Transform	114
4.3.2	Improve the Accuracy of the Poles	117
4.3.3	Model Accuracy and Efficiency	121
4.4	Frequency Dependent Effects	123
4.5	Conclusions	130
5	Effective Capacitance of RLC Loads for Estimating Short-Circuit Power	132
5.1	Introduction	132
5.2	Effective Capacitance of an RLC Load	133
5.2.1	π -Model Representation of RLC Interconnects	134
5.2.2	Effective Capacitance for Short-Circuit Power	136
5.3	Model Verification	142
5.4	Conclusions	147
6	Low Power Repeaters Driving RC and RLC Interconnects with Delay and Bandwidth Constraints	148
6.1	Introduction	148
6.2	Power Dissipation in an RC Interconnect with Delay and Bandwidth Constraints	150
6.2.1	Delay and Transition Time Model of RC Interconnects	150
6.2.2	Power Dissipation Components in Interconnects with Repeaters	156
6.2.3	Power Dissipation with Delay Constraints	160
6.2.4	Power Dissipation with Bandwidth Constraints	168

6.2.5	Power Dissipation with both Delay and Bandwidth Constraints	171
6.3	Effects of Inductance on the Repeater Insertion Methodology	173
6.3.1	Timing Model of <i>RLC</i> Interconnects	173
6.3.2	Effects of Inductance on the Repeater Design Space	176
6.3.3	Power Dissipation with Delay and Bandwidth Constraints . .	180
6.4	Conclusions	185
7	Predictions of CMOS Compatible On-Chip Optical Interconnect	187
7.1	Introduction	187
7.2	Electrical Interconnect	189
7.2.1	Delay Optimal Design	189
7.2.2	Delay Uncertainty Model	192
7.3	On-Chip Optical Data Path	193
7.3.1	Transmitters	194
7.3.2	Waveguides	197
7.3.3	Receivers	198
7.4	Comparison between Electrical and Optical Interconnects	201
7.4.1	Delay Uncertainty	202
7.4.2	Delay	210
7.4.3	Power	212
7.4.4	Bandwidth Density	213
7.4.5	Discussion	215
7.5	Potential Challenges in Optical Interconnects	216
7.6	Conclusions	218
8	Conclusions	219
9	Future Research	223
9.1	Effect of Repeaters on Delay Uncertainty	224
9.2	Figure of Merit to Characterize the Importance of Frequency Depen- dent Effects	225

9.3	Design Methodology for Optical Clock Distribution Networks	227
9.4	3-D Integration with Optical Interconnects	228
9.5	Summary	229
Bibliography		230
Appendices		
A	Minimizing P_{total} with a Delay Constraint for RC Interconnects	249
B	Modeling of MOSFET Transistors	253
B.1	Threshold voltage	254
B.1.1	Effect of L variation	257
B.1.2	Effect of T_{ox} variation	258
B.1.3	Effect of N_{sub} variation	259
B.1.4	Effect of T variation	259
B.1.5	Effect of V_{dd} variation	260
B.2	Mobility	260
B.3	I-V characteristics	264
B.4	Transconductance and output resistance	265
C	Publications	267

List of Tables

1.1	Scaling trends in semiconductor device dimensions.	4
3.1	Comparison of the 50% delay of Fb3 and Fb5 with SPICE and a single pole model. The input signal parameters are $T = 500$ ps, $\tau = 50$ ps, and $V_{dd} = 1.5$ volts. The interconnect parameters are $l = 2$ mm and $h = 1\mu\text{m}$	73
3.2	Comparison of overshoots/undershoots of Fb3 and Fb5 with SPICE simulations. The input signal parameters are $T = 500$ ps, $\tau = 50$ ps, and $V_{dd} = 1.5$ volts. The interconnect parameters are $l = 2$ mm and $h = 1\mu\text{m}$	73
3.3	Interconnect lengths shown in Fig. 3.12 normalized to l_x	82
3.4	Load capacitances shown in Fig. 3.12 normalized to C_x	83
3.5	The 50% delays at nodes $N5$ and $N7$ as shown in Fig. 3.12 with different circuit parameters.	83
3.6	The transition times at nodes as $N5$ and $N7$ shown in Fig. 3.12 with different circuit parameters.	86
3.7	Comparison of the maximum crosstalk noise of Fb3 and Fb5 with SPICE simulations. The input signal parameters are $T = 500$ ps, $\tau = 50$ ps, and $V_{dd} = 1.5$ volts.	97
5.1	Short-circuit energy dissipation during a full signal switch.	145
6.1	Device parameters of BPTM 45 nm model. $V_{dd} = 1.1$ volts.	151

6.2	Minimum power with delay constraints obtained analytically as compared with SPICE simulations. $f = 1$ GHz.	166
6.3	Different power components dissipated in the repeaters. $f = 1$ GHz. .	167
7.1	Predictive model of future silicon based electro-optical modulators. .	196
7.2	Parameters and 3σ variations.	206
7.3	Delay (ps) and 3σ value of a 10 mm optical data path.	208
7.4	Delay comparison between electrical and optical interconnects. . . .	212
7.5	Power consumption (mW) in optical and electrical interconnects. . . .	213
B.1	Parameters used in the model of the threshold voltage.	255

List of Figures

1.1	Total number of transistors in lead Intel microprocessors.	2
1.2	Die area and minimum feature size of transistors in lead Intel microprocessors.	3
1.3	Scaling of transistor and interconnect.	4
2.1	A conventional ASIC design flow.	11
2.2	A data path in a synchronous digital system.	13
2.3	Components of dynamic power dissipation due to different capacitance sources: gate capacitance, diffusion capacitance, and interconnect capacitance.	15
2.4	Interconnect coupling noise.	16
2.5	Timing diagram of a data waveform.	18
2.6	Cross section of an on-chip copper interconnect.	20
2.7	Scattering of electrons at the interconnect surface and grain boundary.	21
2.8	Current distribution in the cross section of an interconnect at high frequencies. Darker color indicates higher current density.	23
2.9	Skin depth of Cu as a function of frequency.	24
2.10	Current distribution in the cross section of two parallel wires at high frequencies due to the proximity effect.	29
2.11	Return current distribution at different frequencies.	30
2.12	Lumped interconnect models.	31
2.13	Circuit models of transmission lines.	33
2.14	Model of orthogonal layer.	34

2.15	Normalized frequency spectrum of a saturated ramp signal.	36
2.16	Normalized integral of the frequency spectrum of a saturated ramp signal.	37
2.17	Modeling a frequency dependent impedance with lumped elements. . .	38
2.18	Decoupling multiple parallel coupled interconnects.	40
2.19	Impulse and step responses of RC trees.	43
2.20	An example of an A -tree.	48
2.21	Shaping interconnect to minimize delay.	49
2.22	Staggering repeaters to reduce the worst case delay and crosstalk noise.	51
2.23	Buffered interconnect tree.	52
2.24	Examples of net-ordering and wire swizzling.	54
3.1	Equivalent circuit model of a distributed RLC interconnect.	59
3.2	The amplitude transfer function of different models of an RLC inter- connect.	61
3.3	The amplitude transfer functions of an RLC interconnect with different inductive effects.	63
3.4	Normalized amplitude of odd order harmonics.	65
3.5	Comparison of the time domain response of Fb3 and Fb5 with SPICE.	66
3.6	Examples of pseudo-undershoots. The input signal parameters are $T = 500$ ps, $\tau/T = 0.1$, and $V_{dd} = 1.5$ volts. The driver and load parameters are (a) $R_d = 100\ \Omega$ and $C_l = 500$ fF, (b) $R_d = 100\ \Omega$ and $C_l = 50$ fF, (c) $R_d = 60\ \Omega$ and $C_l = 500$ fF, and (d) $R_d = 60\ \Omega$ and $C_l = 500$ fF.	71
3.7	The effect of initial conditions on the periodic signals. $l = 5$ mm. . . .	74
3.8	The 50% delay versus interconnect length. $w = 2\ \mu\text{m}$, $T = 500$ ps, and $\tau = 50$ ps.	75
3.9	The 50% delay for different τ/T . $w = 2\ \mu\text{m}$, $l = 2$ mm, $T = 500$ ps, $V_{dd} = 1.5$ volts, $R_d = 30\ \Omega$, and $C_l = 50$ fF.	76
3.10	Normalized amplitude of harmonics with different τ/T	76

3.11	The effects of signal frequency on the accuracy of the proposed model. $w = 2\mu\text{m}$, $l = 2\text{mm}$, $\tau/T = 0.1$, $V_{dd} = 1.5\text{ volts}$, $R_d = 30\Omega$, and $C_l = 50\text{ fF}$. (a) the 50% delay, (b) Overshoot.	78
3.12	A distributed <i>RLC</i> tree.	80
3.13	An example of a shielded clock wire structure.	82
3.14	Time domain response at the leaves of the tree shown in Fig. 3.12. $l_x = 1\text{ mm}$, $\tau = 50\text{ ps}$, $T = 500\text{ ps}$, $V_{dd} = 1.5\text{ volts}$, $R_d = 10\Omega$, and $C_x = 20\text{ fF}$. (a) Node <i>N5</i> , (b) Node <i>N7</i>	85
3.15	Time domain response at node <i>N7</i> in Fig. 3.12 evaluated by the Fourier series based model with different n_f as compared with SPICE simula- tions. $l_x = 1\text{ mm}$, $\tau = 5\text{ ps}$, $T = 500\text{ ps}$, $V_{dd} = 1.5\text{ volts}$, $R_d = 10\Omega$, and $C_x = 20\text{ fF}$	88
3.16	Geometric characteristics of five parallel interconnect lines	94
3.17	Amplitude transfer functions of a five line system.	95
3.18	Comparison of the far end response from Fb3 and Fb5 with SPICE in a five line coupled system. The input signal parameters are $T = 500\text{ ps}$, $\tau = 50\text{ ps}$, and $V_{dd} = 1.5\text{ volts}$	96
4.1	Distributed interconnect with a lumped capacitive load and driver re- sistance.	102
4.2	Graphic view of the roots of (4.4), $R_T = C_T = 1$	103
4.3	Analytic solution of (4.3) as compared with the exact solution for dif- ferent values of R_T and C_T	105
4.4	Graphic view of the roots of (4.15), $C_T = 1$	107
4.5	Analytic solution of (4.14) as compared with the exact solution for different values of C_T	108
4.6	Wire geometry of an example circuit, where the signal wire is shielded by two ground lines.	109

4.7	Comparison between the analytic expression (4.21) and the exact transfer function. The wire length is 5 mm and the load capacitance is $C_L = 50$ fF. a) RC interconnect case, $R_d = 30 \Omega$. b) RLC interconnect with a zero R_d	111
4.8	Step and ramp response obtained analytically as compared with Spectre simulations. (a) Step response, RC , (b) Ramp response, RC , (c) Step response, RLC , and (d) Ramp response, RLC	112
4.9	Transient response of a transmission line obtained with the proposed model, four-pole model, and Spectre simulations. $t_r = 50$ ps and $C_L = 50$ fF. (a) $R_d = 20 \Omega$. (b) $R_d = 300 \Omega$	116
4.10	Mapping between the approximated poles and the exact poles, $R_d = 100 \Omega$	118
4.11	Pseudo-code for computing the exact poles. Function <i>Newton_Raphson</i> () is the Newton-Raphson converging process starting with the input argument.	119
4.12	Transient response of transmission line obtained with the improved analytic method as compared with Spectre simulations. $m = 2$. (a) $R_d = 20 \Omega$. (b) $R_d = 300 \Omega$	122
4.13	Comparison of the 50% delay, 10%-to-90% output rise time, and the normalized overshoot obtained from the proposed model and Spectre simulations. $R_d = 20 \Omega$, $C_L = 50$ fF, and $l = 5$ mm. (a) Delay and rise time. (b) Overshoot.	124
4.14	Comparison of the 50% delay and 10%-to-90% output rise time obtained from the proposed model and Spectre simulations. $t_r = 50$ ps. (a) 50% delay. (b) 10%-to-90% output rise time.	125
4.15	A segment of interconnect with length Δl	126
4.16	Frequency dependent impedance of an interconnect with a length of 5 mm. (a) Resistance. (b) Inductance.	128
4.17	Comparison of the output signal waveforms with and without the frequency dependent effect, $R_d = 10 \Omega$, $C_L = 50$ pF, and $t_r = 50$ ps. . . .	130

4.18	Comparison of transfer functions with and without the frequency dependent effect, $R_d = 10\ \Omega$ and $C_L = 50\ \text{pF}$	131
5.1	Reduction of an RLC interconnect network.	135
5.2	An example of a distributed RLC tree.	137
5.3	Effect of inductance on short-circuit current. $t_r = 0.5\ \text{ns}$	137
5.4	Current components in a CMOS inverter.	138
5.5	Effective capacitance as a function of t_{ev} . $C_n = 120.1\ \text{fF}$, $C_f = 965.9\ \text{fF}$, $R_\pi = 15.9\ \Omega$, and $L_\pi = 0.96\ \text{nH}$	141
5.6	Short-circuit current with different output loads.	142
5.7	Short-circuit energy with different loads. $L_{int} = 0.74\ \text{pH}/\mu\text{m}$	143
5.8	Effect of inductance on the effective capacitance.	144
5.9	Short-circuit energy with multiple switching inputs. $C_n = 100\ \text{fF}$, $C_f = 800\ \text{fF}$, $R_\pi = 200\ \Omega$, and $L_\pi = 3\ \text{nH}$	146
6.1	Repeater insertion in a long RC interconnect line.	150
6.2	Total delay for an RC interconnect driven by repeaters. $R = 0.31\ \Omega/\mu\text{m}$, $C = 0.223\ \text{fF}/\mu\text{m}$, $l = 5\ \text{mm}$, and $h = 50$	154
6.3	Repeater design space with delay constraint. $R = 0.31\ \Omega/\mu\text{m}$, $C = 0.223\ \text{fF}/\mu\text{m}$, and $l = 10\ \text{mm}$	160
6.4	Total power dissipation in an interconnect with repeaters as a function of h and k . $f = 1\ \text{GHz}$, $R = 0.31\ \Omega/\mu\text{m}$, $C = 0.223\ \text{fF}/\mu\text{m}$, and $l = 10\ \text{mm}$	161
6.5	The ratio of C_{eff} to C_{stage} . $R = 0.31\ \Omega/\mu\text{m}$, $C = 0.223\ \text{fF}/\mu\text{m}$, and $l = 10\ \text{mm}$	162
6.6	Power dissipation with constant delay. $f = 1\ \text{GHz}$, $T_{req} = 1\ \text{ns}$, $R = 0.31\ \Omega/\mu\text{m}$, $C = 0.223\ \text{fF}/\mu\text{m}$, and $l = 10\ \text{mm}$	165
6.7	The effect of α_s on the optimal repeater size h_p . $R = 0.31\ \Omega/\mu\text{m}$, $C = 0.223\ \text{fF}/\mu\text{m}$, $l = 10\ \text{mm}$, $f = 1\ \text{GHz}$, and $T_{req} = 1\ \text{ns}$	168
6.8	Repeater design space with bandwidth constraints. $R = 0.31\ \Omega/\mu\text{m}$, $C = 0.223\ \text{fF}/\mu\text{m}$, and $l = 10\ \text{mm}$	169

6.9	Power dissipation and 50% delay at the edge of the design space with bandwidth constraint. $B_{req} = 1$ Gb/s, $R = 0.31 \Omega/\mu\text{m}$, $C = 0.223$ fF/ μm , and $l = 10$ mm.	172
6.10	The design space and power dissipation at the edge of the design space with both delay and bandwidth constraints.	174
6.11	Inductance effect for different driver size and interconnect length. $W = 20W_{min}$ and $L = 1$ pH/ μm	177
6.12	Inductance values with difference current return paths.	177
6.13	Effects of inductance on the repeater design space satisfying bandwidth constraints. $B_{req} = 2$ Gb/s, $l = 10$ mm, and $W = 10W_{min}$	178
6.14	Effects of inductance on the interconnect delay with repeaters. $l = 10$ mm, $k = 10$, $h = 100$, and $W = 10W_{min}$	180
6.15	Effects of inductance on repeater design space satisfying delay constraints. $T_{req} = 700$ ps, $l = 10$ mm, and $W = 10W_{min}$	181
6.16	Effects of inductance on short-circuit current in repeaters. $l = 10$ mm, $k = 10$, $h = 150$, and $W = 10W_{min}$	182
6.17	Effects of inductance on the short-circuit power in repeaters. $l = 10$ mm, $k = 10$, and $W = 10W_{min}$	182
6.18	Effects of inductance on the minimum interconnect power while satisfying a delay constraint. $l = 15$ mm, $W = 10W_{min}$, and $T_{req} = 1$ ns.	184
6.19	Effects of inductance on the minimum interconnect power while satisfying a bandwidth constraint. $l = 15$ mm, $W = 10W_{min}$, and $B_{req} = 2$ Gb/s.	185
7.1	Repeater insertion in an RLC interconnect.	190
7.2	Minimum delay per unit length as a function of interconnect width.	191
7.3	An on-chip optical interconnect data path.	194
7.4	Circuit model of an optical receiver.	198
7.5	Detector response time versus electrode width. A ($100 \mu\text{m} \times 100 \mu\text{m}$), B ($50 \mu\text{m} \times 50 \mu\text{m}$), C ($20 \mu\text{m} \times 20 \mu\text{m}$), and D ($10 \mu\text{m} \times 10 \mu\text{m}$).	200

7.6	Delay distribution of a 10 mm electrical interconnect at the 45 nm technology node.	209
7.7	Comparison of standard deviation of delays of electrical and optical interconnects.	209
7.8	A timing diagram of data and clock waveforms.	211
7.9	Comparison of bandwidth density of electrical and optical interconnects.	214
7.10	Normalized critical length beyond which optical interconnect is advantageous over electrical interconnect.	215
9.1	Design space for repeaters in global interconnect.	225
9.2	Variations in impedance over the frequency range of interest.	226
9.3	Optical-electrical clock distribution network.	228

Chapter 1

Introduction

The invention of the integrated circuit (IC) in the early 1960's enabled the development of a vast number of microelectronic applications over the past half century, such as personal computers and cell phones. After decades of IC technology evolution, CMOS technology has become the dominant technology in the digital IC market due to the low static power and excellent scalability of CMOS.

With increasing functionality and performance requirements, on-chip integration levels and system clock frequencies have increased exponentially. This trend is commonly referred to as *Moore's Law* [1]. The original form of Moore's law is that the number of transistors on a monolithic die with the lowest manufacturing costs per transistor doubles roughly every year [2]. This prediction was revised in 1975 as "the number of transistors on the most complex ICs would double every two years" [3]. The number of transistors in the lead Intel microprocessors is shown in Fig. 1.1 [4, 5], which agrees quite well with Moore's law.

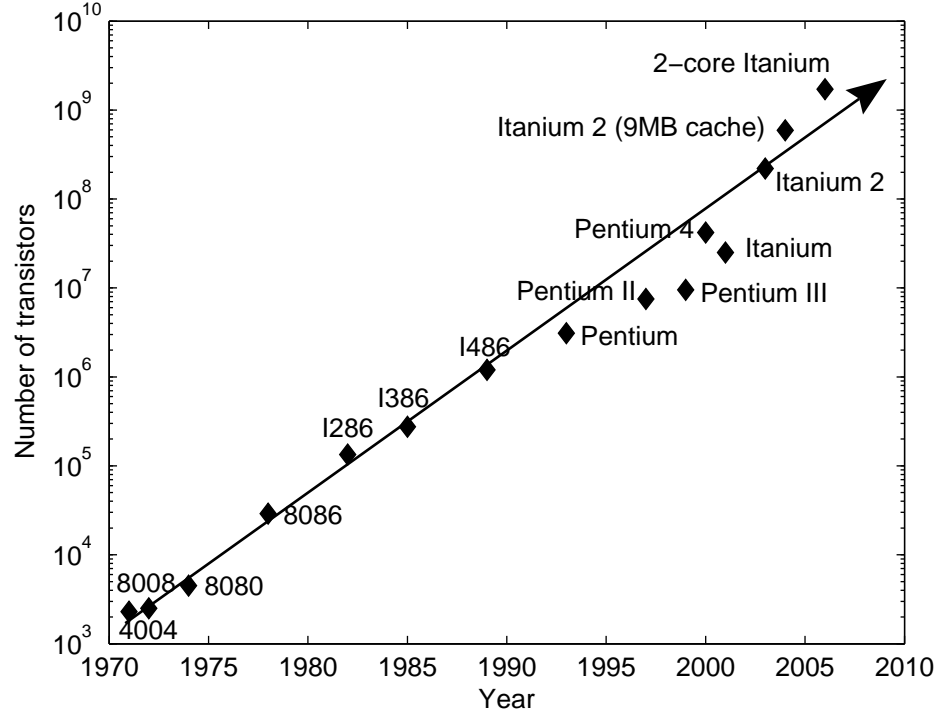


Figure 1.1: Total number of transistors in lead Intel microprocessors.

The increase in on-chip integration is due to larger die areas and smaller transistor dimensions. From 1971, the die area of the lead Intel microprocessors has increased by 14% per year, as shown in Fig. 1.2 [5, 6]. This trend slowed down in the mid 1990's due to concerns about power dissipation and yield. The die area is predicted to be fixed for future technologies, as described in the International Technology Roadmap for Semiconductors (ITRS) [7]. The minimum feature size of transistors in the lead Intel microprocessors has decreased from $10\mu\text{m}$ in 1971 to 90 nm in 2006. This trend is expected to continue into the next decade [7]. The scaling of transistors and interconnects is illustrated in Fig. 1.3.

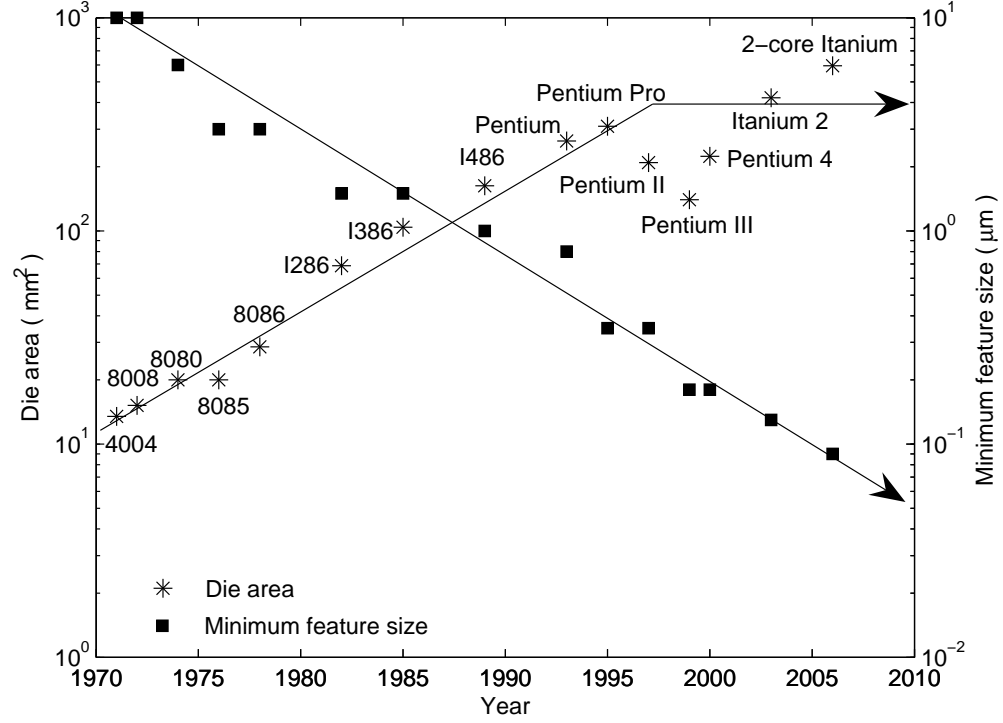


Figure 1.2: Die area and minimum feature size of transistors in lead Intel microprocessors.

In Table 1.1, the predicted geometric dimensions for different technology nodes are listed for both transistors and interconnects [7]. Each technology node lasts approximately two to three years. Note that the node names are determined by half of the metal pitch of the bit lines in typical dynamic random access memory (DRAM) circuits rather than the printed gate lengths of the transistors. This dimension decreases by a half every two technology nodes.

With scaling, the performance of the transistor is improved due to smaller parasitic capacitive loads and higher drain currents. The interconnect delay, however,

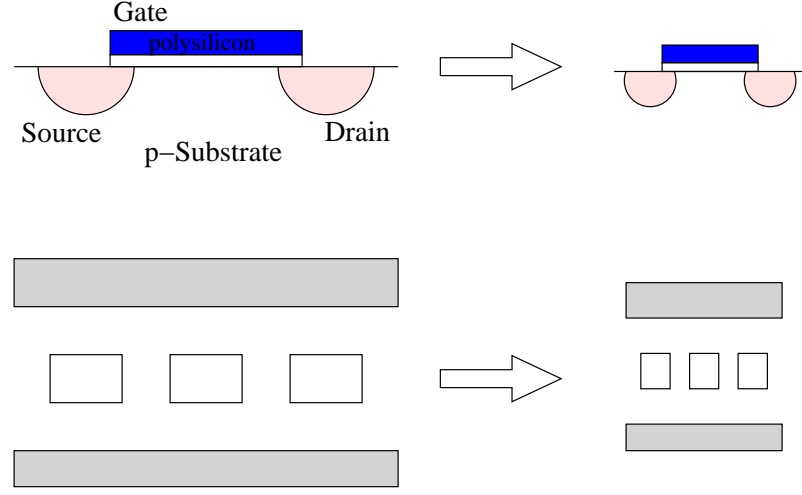


Figure 1.3: Scaling of transistor and interconnect.

Table 1.1: Scaling trends in semiconductor device dimensions.

Year	2004	2007	2010	2013	2016
Technology node (nm)	90	65	45	32	22
Printed gate length (nm)	53	35	25	18	13
Equivalent gate oxide thickness (nm)	1.2	0.9	0.7	0.6	0.5
Local metal wire pitch (nm)	214	152	108	76	54
Local metal wire aspect ratio	1.7	1.7	1.8	1.9	2
Intermediate metal wire pitch (nm)	275	195	135	95	65
Intermediate metal wire aspect ratio	1.7	1.8	1.8	1.9	2
Global metal wire pitch (nm)	410	290	205	140	100
Global metal wire aspect ratio	2.1	2.2	2.3	2.4	2.5

increases significantly due to the large increase in the resistance per unit length, which approximately doubles with each new technology node. Interconnect delay now dominates gate delay in deep submicrometer (DSM) technologies [7]. With increasing on-chip integration, the interconnections among the cells also increase significantly,

requiring additional metal resources. The number of metal layers in current state-of-the-art technologies is nine [8] and is expected to increase to 14 by the 22 nm technology node [7]. The additional metal layers will further increase the dynamic power dissipated by the interconnects due to the greater amount of interconnect capacitance. As the power supply voltage decreases, the noise margin of digital CMOS circuits also decreases, making the circuit more sensitive to injected noise. One of the primary noise sources in ICs is due to the interconnect, including interconnect coupling noise, IR and Ldi/dt drops in the power/ground grid, and jitter/skew in the clock distribution network. If these sources of noise exceed the noise margin, a malfunction can occur in the circuit. The design of on-chip interconnects, therefore, has become an essential issue in high speed ICs.

Due to the increasing complexity of ICs, modern digital circuits are generally realized by computer-aided-design (CAD) tools. The design procedure is highly automated and a variety of design flows have been developed over the last two decades. The dominant behavior of interconnects, however, greatly affects the circuit design process and the development of related CAD algorithms. The design focus has therefore shifted from logic optimization to interconnect optimization, requiring redesigned CAD tools to support an interconnect-centric design flow [9]. In these tools, interconnects need to be accurately modeled. The modeling of on-chip interconnects has

become more challenging due to the smaller physical dimensions, higher signal frequencies, and more complicated network structures. A number of non-ideal effects need to be included, such as the inductive behavior and frequency dependent effects.

As interconnect becomes more important, it is essential to understand the manner in which on-chip interconnect affects the circuit design process and how to manage related interconnect effects. In Chapter 2, a review of the background of on-chip interconnect is provided, including a description of IC design flows, the modeling and analysis of interconnects, design criteria, and interconnect design methodologies.

Accurate and efficient *RLC* interconnect models are critical to the design of high performance DSM circuits. Based on a Fourier series analysis, an analytic interconnect model is presented in Chapter 3, which is suitable for periodic signals, such as a clock signal. No approximation is made to the transfer function of the interconnect. The far end response is approximated by the summation of several sinusoids. Since the solution is the steady state response to a periodic signal, initial conditions are considered. The model is verified by SPICE simulations and successfully extended to *RLC* trees and multiple transmission lines. The computational complexity of the model is linear with the model order.

In Chapter 4, an alternative solution for the transient response at the far end of a transmission line is proposed, which is based on a direct pole extraction of the system. Closed-form expressions of the poles are developed for two special interconnect

systems: an RC interconnect and an RLC interconnect with a zero driver resistance. By performing a system conversion, the poles of an interconnect system with general circuit parameters are determined. The Newton-Raphson method is used to further improve the accuracy of the poles. Based on these poles, closed-form expressions for the step and ramp response are determined. Higher accuracy can be obtained with additional pairs of poles. The computational complexity of the model is proportional to the number of pole pairs. Frequency dependent effects are also successfully included in the proposed method and excellent match is observed between the proposed model and Spectre simulations.

Since power dissipation has become a fundamental design criterion in the IC design process, accurate and efficient power estimation is critical in designing low power circuits. In Chapter 5, an effective capacitance of a distributed RLC load is presented for accurately estimating short-circuit power. Both resistive and inductive shielding of interconnects are considered and no iterations are required to determine the effective capacitance. The proposed method has been verified with Cadence Spectre, and can be used in look-up tables or k -factor based models to estimate short-circuit power dissipation in CMOS gates with complex interconnect loads.

Repeater insertion is an efficient method for reducing interconnect delay and signal transition times in integrated circuits. In Chapter 6, a repeater insertion methodology is proposed for achieving the minimum power in an RC interconnect while satisfying

delay and bandwidth constraints. These constraints determine a design space for the number and size of the repeaters. The minimum power is shown to occur at the edge of the design space. Closed-form solutions for the minimum power satisfying a delay constraint are developed. The effects of inductance on the delay, bandwidth, and power of an *RLC* interconnect with repeaters are also analyzed.

As CMOS technology is scaled, the design requirements of delay, power, bandwidth, and noise due to the on-chip interconnects have become more stringent. New design challenges are emerging, such as delay uncertainty induced by process and environmental variations. It has become increasingly difficult for conventional copper interconnect to satisfy these design requirements. On-chip optical interconnect has therefore been considered as a potential substitute for long distance global electrical interconnect. In Chapter 7, predictions of the performance of CMOS compatible optical devices are made based on current state-of-the-art optical technologies. Electrical and optical interconnects are compared for various design criteria based on these predictions.

In Chapter 8, the research described in the dissertation is summarized. Proposed future research is presented in Chapter 9, including a delay uncertainty constrained repeater insertion methodology, a figure of merit to characterize the importance of frequency dependent effects, methodologies for optical-electrical clock distribution networks, and 3-D integration via optical interconnects.

Chapter 2

On-Chip Electrical Interconnect

Due to the importance of interconnects in current and future ICs, significant research has been published over the past several decades, covering different areas such as parasitic extraction, interconnect models, and interconnect design methodologies. In this chapter, a brief review of the background of on-chip electrical interconnect is provided. In Section 2.1, a typical design flow for application-specific integrated circuits (ASIC) is described. Challenges in DSM technologies due to interconnect dominant behavior are discussed. In Section 2.2, different design criteria that need to be considered during the interconnect design procedure are described. The impedance characteristics of interconnect are presented in Section 2.3; specifically, the resistance, capacitance, and inductance. Interconnect models and design methodologies are reviewed in Sections 2.4 and 2.5, respectively. Finally, some conclusions are offered in Section 2.6.

2.1 Design Flows for DSM ASICs

A conventional design flow for ASICs is shown in Fig. 2.1 [10]. A typical design process can be divided into two steps: functional design (front-end) and physical design (back-end). The functional design phase includes functional specification, VHDL/Verilog coding in the register transfer level (RTL), and logic synthesis. A gate level netlist is generated as the result of logic synthesis. Functional design is implemented during the front-end design process. The back-end physical design process converts a gate level netlist into a layout, including floorplaning, module placement, and interconnect routing. From the physical layout, parasitic impedances are extracted. A post-layout timing analysis tool is used to detect any timing violations. Necessary corrections are made in the physical layout or gate level netlist to fix these violations. This design flow is successful for those technologies where gate delays dominate. The timing of the circuits is determined by the gate types and loads. The effect of the interconnect parasitic impedances typically produces only a few timing violations in a medium speed application, making the design flow efficient.

With interconnect becoming increasingly important, the interconnect delay needs to be considered during the functional design process. Due to the lack of placement and routing information, the interconnect delay is approximated with statistical fanout-based wire load models. The circuit design based on these inaccurate delay models can produce a large number of timing violations. Design iterations are usually

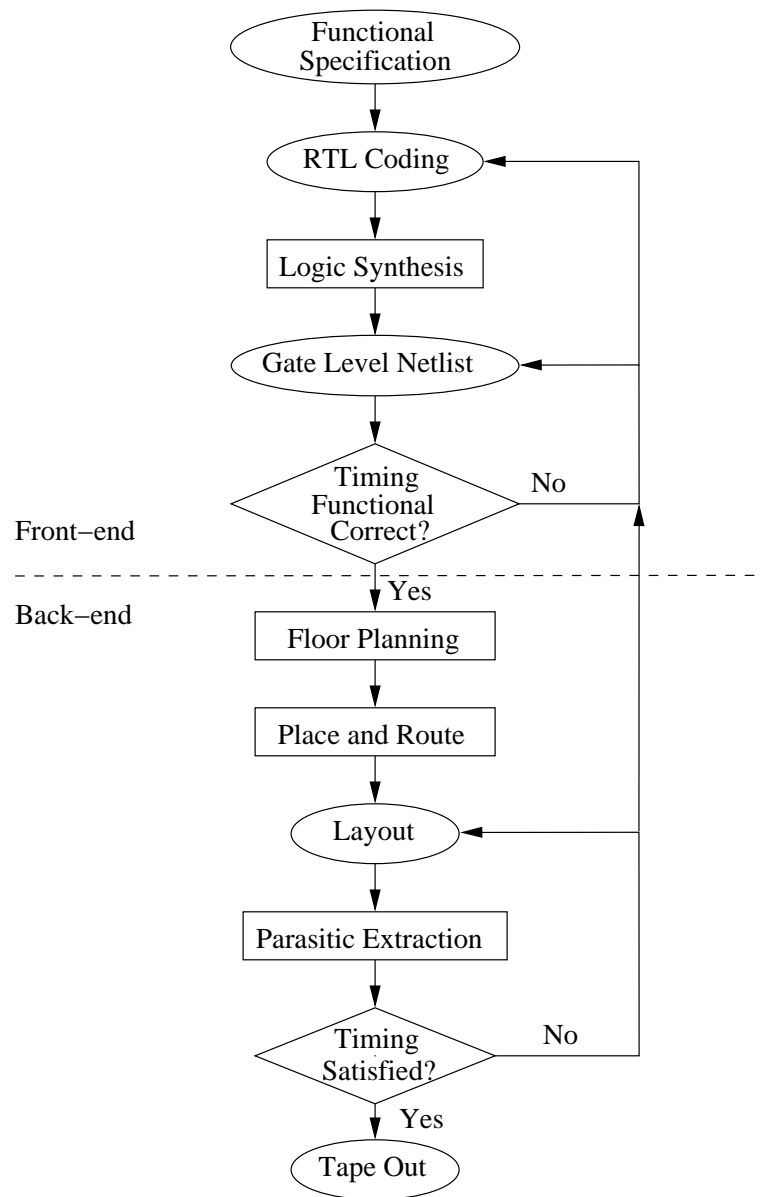


Figure 2.1: A conventional ASIC design flow.

required to achieve timing closure. A method to alleviate this problem is to introduce physical information earlier into the logic synthesis stage. An initial floorplan is created before the synthesis procedure to provide an estimate of the location of

the cells as well as the interconnect lengths. A timing model based on this estimation is significantly more accurate, making the synthesis process more efficient and resulting in a placed gate level netlist. This synthesis procedure is called physical synthesis [11]. In the DSM regime, the functional and physical design processes are no longer separated, requiring tight integration of the front-end and back-end design processes.

Interconnect plays an important role in both the physical synthesis and timing verification stages in the design flow. Requirements placed on the interconnect analysis are different in these two stages. During the synthesis process, since the detailed routing information is not available, higher efficiency with reasonable accuracy is preferred, such as closed-form models. In the post-layout verification stage, realistic timing information describing the entire IC is determined, requiring both high efficiency and high accuracy.

2.2 Interconnect Design Criteria

Since interconnect has become a dominant issue in high performance ICs, the focus of the circuit design process has shifted from logic optimization to interconnect optimization. Multiple criteria should be considered during the interconnect design process, such as delay, power dissipation, noise, bandwidth, and physical area. These criteria are individually discussed in the following subsections.

2.2.1 Delay

Interconnect delay is a primary design criterion due to the close relationship to the speed of a circuit. Early interconnect design methodologies [12, 13] focused primarily on delay optimization. A typical data path in a synchronous digital circuit is shown in Fig. 2.2. In the case of zero clock skew, the minimum allowable clock period is [14]

$$T_{p_min} = T_{C-Q} + T_{int} + T_{logic_max} + T_{setup}, \quad (2.1)$$

where T_{C-Q} is the time required for the data to leave the initial register after the clock signal arrives, T_{int} is the interconnect delay, T_{logic_max} is the maximum logic gate delay, and T_{setup} is the required setup time of the receiving register. From (2.1), by reducing T_{int} , the clock period can be decreased, increasing the overall clock frequency of the circuit (assuming the data path is a critical path).

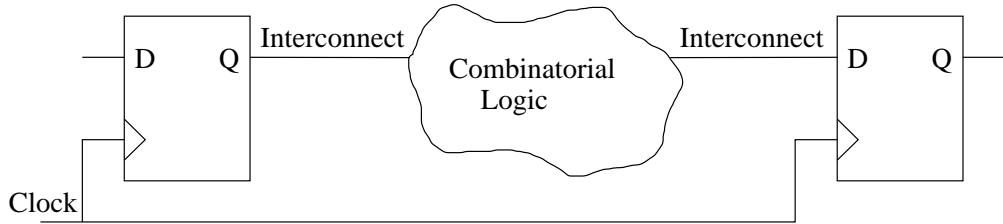


Figure 2.2: A data path in a synchronous digital system.

In advanced microprocessors, multiple computational cores can be fabricated on the same die [5]. Communication among these cores and on-chip memories generally requires multiple clock cycles. Sometimes the computational core enters an idle state

waiting for the required data or control signals from other regions of the IC. The computational resource of these cores, therefore, cannot be efficiently utilized due to the large amount of multi-cycle communication. By reducing the interconnect delay, the speed of the system, *i.e.*, the computational efficiency of the cores, can be improved at the architecture level.

2.2.2 Power Dissipation

Due to higher clock frequencies and on-chip integration levels, power dissipation has significantly increased. The on-chip power dissipation of current state-of-the-art microprocessors is on the order of hundreds of watts and the power density has exceeded the power density of a kitchen hot plate [15]. In Fig. 2.3, the components of dynamic power due to different capacitance sources are shown for a state-of-the-art microprocessor [16]. The dynamic power due to the interconnect capacitance can be greater than 50% of the total dynamic power. Furthermore, the repeaters and pipeline registers inserted in the interconnect introduce additional dynamic, leakage, and short-circuit power [17]. High power dissipation increases the packaging cost due to heating problems and shortens the battery life in portable applications. Power dissipation, therefore, is another important criterion in interconnect design.

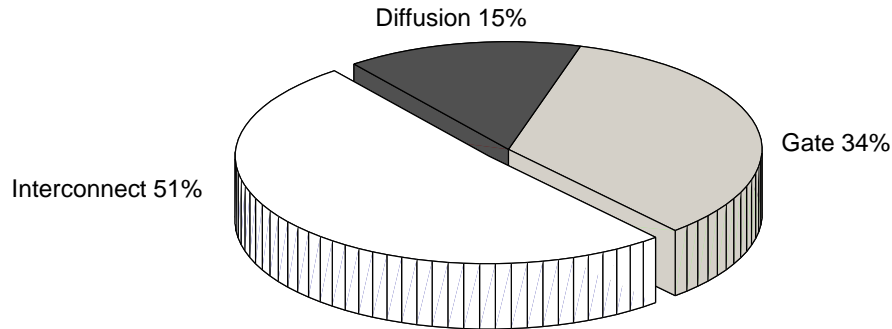


Figure 2.3: Components of dynamic power dissipation due to different capacitance sources: gate capacitance, diffusion capacitance, and interconnect capacitance.

2.2.3 Noise

With interconnect scaling, coupling capacitance between (and among) interconnects dominates the ground capacitance. Furthermore, inductive coupling has to be considered due to increasing signal frequencies, making coupling noise more significant (and complicated). Interconnect coupling induced noise can be classified into two categories: voltage level noise and delay uncertainty, as shown in Fig. 2.4. Noise may cause a malfunction in the circuit if the noise level is greater than a certain threshold, thereby reducing yield.

In addition to coupling effects, delay uncertainty can also be caused by other factors, such as process variations (on both interconnects and the inserted repeaters

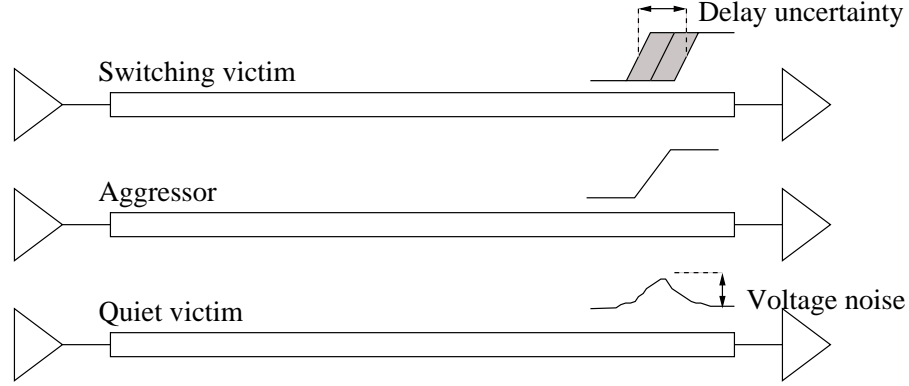


Figure 2.4: Interconnect coupling noise.

or pipeline registers), temperature variations, and power/ground noise. Delay uncertainty is both spatially dependent (due to process variations) and temporally dependent (due to coupling, temperature variations, and power/ground noise). Timing margins are assigned to manage this delay uncertainty, thereby increasing the clock period and reducing the overall performance of the circuits. When delay uncertainty exceeds these margins, setup or hold violations may occur, reducing the yield.

2.2.4 Bandwidth

The concept of bandwidth originates from the telecommunications field [18]. For on-chip applications, bandwidth is used to measure the data transmitting capacity for global interconnects, *i.e.*, the number of bits transmitted through an interconnect per second. A higher bandwidth reduces the total time required to transmit a certain amount of data, thereby increasing the performance of the system. A bit period can be divided into two parts. One part is dedicated to the transition time, while

the other part is the steady state part during which the data can be latched at the receiving register. Assuming the steady state part occupies at least half of the bit period, the maximum bandwidth is related to the rise/fall time as

$$B = \frac{1}{2t_r} \text{ (bit/s)}. \quad (2.2)$$

In [19, 20, 21, 22], the bandwidth is assumed to be proportional to the reciprocal of the delay. This assumption, however, is only valid for RC lines, where there is approximately a linear relationship between the 50% delay and the 10% to 90% transition time.

The bandwidth of an interconnect is also affected by the delay uncertainty. A timing diagram of a data waveform with delay uncertainty is shown in Fig. 2.5, where T_{un} is the delay uncertainty and T_s is the required steady state period. Note that the delay uncertainty increases the bit period T_{bit} . By including the delay uncertainty, (2.2) can be rewritten as

$$B = \frac{1}{T_{bit}} = \frac{1}{2t_r + T_{un}}, \quad (2.3)$$

where T_s is assumed to be the same as t_r .

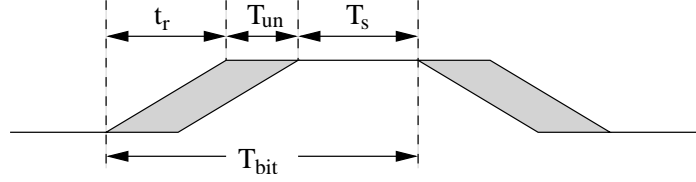


Figure 2.5: Timing diagram of a data waveform.

2.2.5 Physical Area

With technology scaling, billions of transistors can now be integrated onto a single monolithic die. The number of interconnects has therefore also significantly increased. The die size, however, is expected to remain approximately fixed for future technologies as predicted in [7]. The number of metal layers, therefore, needs to be increased to provide sufficient metal resources for interconnect routing. Increasing the number of metal layers, however, increases the fabrication cost. Furthermore, buffers and pipeline registers inserted along the interconnects make the constraint on silicon area more stringent. The area criterion, therefore, should be considered during the interconnect design processes, such as wire sizing and repeater insertion.

2.3 Interconnect Characteristics

The impedance characteristics of on-chip interconnect include the resistance, capacitance, and inductance. These parameters can be extracted from the geometry of the interconnect structures, as illustrated in the following subsections.

2.3.1 Resistance

For a conductor with a rectangle cross-section, the resistance is described by the following expression,

$$R = \rho \frac{l}{WH}, \quad (2.4)$$

where ρ is the material resistivity. l , W , and H are the length, width, and thickness of the interconnect, respectively. In present DSM CMOS technologies, copper has been adopted to replace aluminum as the primary interconnect material due to the lower resistivity of copper as compared to aluminum. At 20°C, the bulk resistivities of Cu and Al are $1.7 \mu\Omega\text{-cm}$ and $2.7 \mu\Omega\text{-cm}$, respectively. Due to specialized processing and operating conditions of the on-chip copper interconnect, certain non-ideal effects need to be considered, making the effective resistivity deviate from the idea bulk resistivity.

a) Diffusion barrier

For on-chip Cu interconnect, a thin and highly resistive barrier layer is built on three sides of the interconnect to prevent Cu from diffusing into the surrounding dielectric, as shown in Fig. 2.6. This barrier layer consumes part of the cross sectional area allocated to the interconnect. The effective resistivity ρ_b due to this barrier

induced reduction in the cross sectional area is [23]

$$\rho_b = \frac{\rho_0}{1 - \frac{A_b}{WH}}, \quad (2.5)$$

where ρ_0 is the bulk resistivity at a given temperature, and A_b is the cross sectional area occupied by the barrier layer.

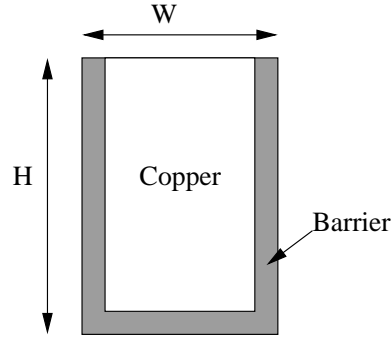


Figure 2.6: Cross section of an on-chip copper interconnect.

b) Surface and grain boundary scattering

When the dimensions of the interconnect are scaled deep into the DSM regime, the resistivity of the interconnect increases as the wire dimensions shrink. This behavior is due to surface and grain boundary scattering [24], as illustrated in Fig. 2.7.

The electron mean-free path λ of copper is 42.1 nm at 0°C [24]. When any dimension of the wire shrinks to the order of λ , the electrons will experience more collisions at the surface, increasing the effective resistivity. The resistivity of a thin

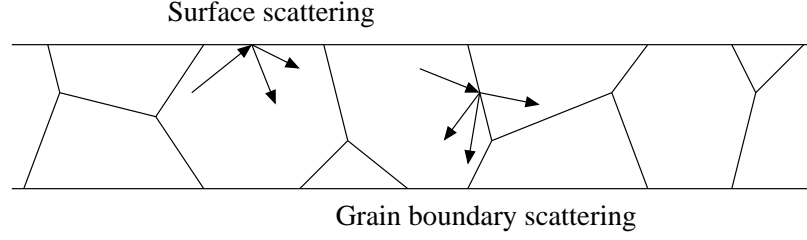


Figure 2.7: Scattering of electrons at the interconnect surface and grain boundary.

film structure can be characterized by the following expressions [23, 25],

$$\rho_s = \frac{\rho_0}{1 - \frac{3(1-p)}{2k} \int_1^\infty \left(\frac{1}{x^3} - \frac{1}{x^5} \right) \frac{1-e^{-kx}}{1-pe^{-kx}} dx}, \quad (2.6)$$

which can be further simplified to [24]

$$\rho_s = \frac{\rho_0}{1 - \frac{3(1-p)}{8k}}, \quad k \gg 1, \quad (2.7)$$

where $k = d/\lambda$ is the ratio of the thickness of the thin film to the electron mean-free path, and p is the fraction of the electrons that are elastically scattered at the surface. A typical value of p for copper is 0.47 [24]. Note that in (2.6) and (2.7), only one dimension (thin film structure) surface scattering is considered. For thin wires with two-dimensional surface scattering effect, the effective resistivity is larger [26]. A reduced k is used in [27] to consider this two-dimensional surface scattering effect.

Grain boundaries in a polycrystalline interconnect act like partially reflecting planes, as illustrated in Fig. 2.7. Grain sizes are usually scaled linearly with the

wire dimensions [28]. When the grain size is comparable to the electron mean-free path, the electrons suffer greater grain boundary scattering, further increasing the resistivity. The effect of grain-boundary scattering can be characterized as [27]

$$\rho_g = \frac{\rho_0}{3[\frac{1}{3} - \frac{1}{2}\alpha_g + \alpha_g^2 - \alpha_g^3 \ln(1 + \frac{1}{\alpha_g})]}, \quad (2.8)$$

where

$$\alpha_g = \frac{\lambda p_g}{d_g(1 - p_g)}. \quad (2.9)$$

d_g is the grain diameter, and p_g is the grain boundary reflection coefficient with a value ranging between 0 and 1.

c) Temperature effect

The resistivity of copper increases approximately linearly with temperature [29] and can be characterized as

$$\rho_t = \rho_0(1 + \beta\Delta T), \quad (2.10)$$

where β is the temperature coefficient of resistivity (TCR) and ΔT is the difference in temperature from a reference temperature. For bulk Cu, β is 0.43%/°C at 20°C [29]. Since the electron mean-free path λ will decrease with increasing temperature, the k in (2.6) will be larger, resulting in a smaller ratio of ρ_s/ρ_0 . The TCR for thin-film interconnect, therefore, is smaller than that of bulk Cu [25].

d) High frequency effects

At sufficiently high frequencies, the current density in an interconnect is no longer uniform, as shown in Fig. 2.8. The current tends to flow near the interconnect surface. This phenomenon is called the skin effect [30]. The effective cross sectional area of the interconnect is reduced, thereby increasing the interconnect resistance.



Figure 2.8: Current distribution in the cross section of an interconnect at high frequencies. Darker color indicates higher current density.

The skin depth is the distance below the conductor surface where the current density drops to $1/e$ of that at the surface, and is determined as

$$\delta(f) = \sqrt{\frac{\rho}{\pi\mu f}}, \quad (2.11)$$

where μ is the permeability in the conductor. Expression (2.4) actually characterizes the DC resistance, and is no longer accurate when δ is smaller than the wire cross sectional dimension. The skin depth of bulk Cu as a function of frequency at 20°C is shown in Fig. 2.9. As the frequency increases to tens of GHz, the skin depth enters the DSM region and decreases slowly.

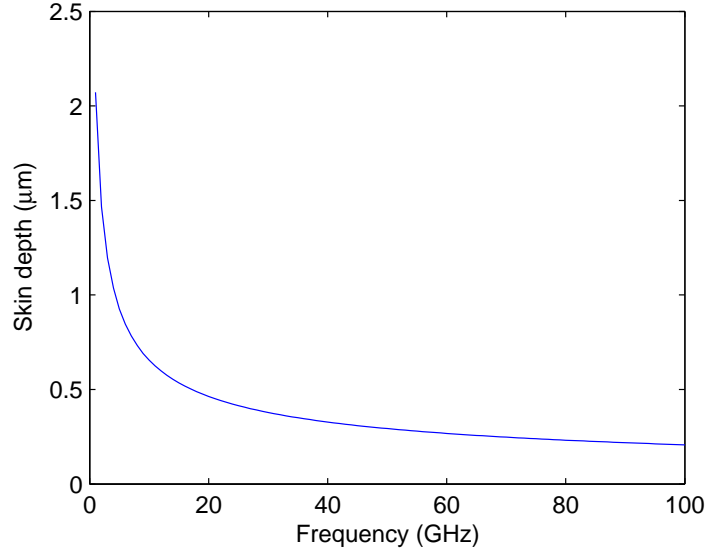


Figure 2.9: Skin depth of Cu as a function of frequency.

Whether to consider these non-ideal effects depends upon the accuracy requirements of the models and the operating regime of the circuits. Often more than one effect needs to be simultaneously considered. For example, the skin effect and surface scattering effect when simultaneously considered is known as the anomalous skin effect (ASE) [26, 31]. In the ITRS [7], the requirement for the effective resistivity of copper interconnect is $2.2 \mu\Omega\text{-cm}$ for all of the technology nodes.

2.3.2 Capacitance

Since interconnect delay dominates gate delay in the DSM regime, the requirement on the accuracy of parasitic extraction of the interconnect impedances increases. 2-D or 3-D extraction is generally required. A 3-D field solver, such as FastCap [32],

can provide accurate capacitance results, however, with large timing and memory requirements. With increasing integration, the number and geometric complexity of the on-chip interconnects drastically increases. It is, therefore, not practical to apply a field solver to an entire IC.

Modern 3-D on-chip capacitance extraction can be divided into three steps [33]. Initially, test patterns are measured or simulated with a 2-D or 3-D field solver. The generated data are used to derive closed-form formulae [34] or to build look-up tables. The geometric parameters of the interconnects are extracted next. Finally, the geometric parameters are matched to the test patterns, and the capacitance values are obtained through formulae or look-up tables. Due to the short-range nature of electrostatic interaction, only the nearest neighbors are considered during the process of capacitance extraction. The capacitance matrices, therefore, are fairly sparse.

Interconnect capacitance is composed of two components, the capacitance between the interconnect and adjacent metal layers or substrate C_g , and the coupling capacitance between neighboring interconnects in the same layer C_c . C_c is expected to dominate C_g in the DSM regime due to the increasing aspect ratio and decreasing wire spacing. In early stage interconnect design and analysis, adjacent layers are generally treated as a ground plane for capacitance extraction. By numerical fitting, closed-form capacitance expressions have been derived for parallel lines above one ground plane or between two ground planes in [35, 36, 37].

2.3.3 Inductance

As compared with resistance and capacitance, the interconnect inductance is significantly more difficult to extract. One reason for this difficulty is due to the loop-based inductance definition,

$$L_{ij} \equiv \frac{\psi_{ij}}{I_j}, \quad (2.12)$$

where ψ_{ij} is the magnetic flux in loop i induced by the current I_j in loop j . To form a loop, the current return paths need to be identified. The current distribution in a circuit, however, *a priori* depends on the interconnect characteristics. The effect of inductance in wide global interconnects in top metal layers is more significant than that of local interconnects in lower metal layers. Since the wires in adjacent layers are generally orthogonal, adjacent layers can no longer be treated as a ground plane as in capacitance extraction.

Another reason for the difficulty in inductance extraction is due to long range inductive coupling effects. Artificially restricting the inductance extraction to nearby geometries not only introduces inaccuracy but may also results in unstable models [33]. The pattern matching method used for capacitance extraction, therefore, can not be used for inductance extraction due to the complex geometries surrounding the wire.

a) Partial inductance

One way to avoid determining *a priori* the current return path is to use the concept of partial inductance [38]. In determining the partial inductance, the flux area extends from the conductor to infinity. The loop inductance of a closed loop can be uniquely determined by the partial self-inductance of each segment of the loop and the partial mutual inductance between any pair of those segments. The partial inductance is used in partial element equivalent circuit (PEEC) models [39], which can be used to accurately simulate a circuit. Partial inductance nonlinearly depends upon the interconnect length. This behavior is the result of inductive coupling among different segments of the same line [30]. For a loop formed by two closely placed parallel interconnects (where the length of the loop is more than ten times longer than the loop width), the loop inductance depends linearly on the length of the loop.

Note that the inductance of a wire not forming a closed loop has no physical meaning [38]. When applying the concept of partial inductance in circuit models, all of the wires that form the current loops should be included, *e.g.*, the reference ground lines. The current return paths are determined from circuit simulation. The PEEC model generally results in huge and dense inductance matrices, increasing the computational complexity of the simulation. Various methods have been presented to sparsify the inductance matrices [40], such as the shell technique [41], the halo technique [42], and the K matrix technique [43].

b) Loop-based inductance

As an alternative to the PEEC model, a loop-based inductance model is preferred in well-designed interconnect structures, such as shielded buses and clock distribution networks. In early design stages, a good assumption regarding the current return path is the nearby power/ground networks, since these tracks are generally wide with low resistive impedance. FastHenry [44] is a commonly used numerical tool for extracting the partial or loop inductance of simple interconnect structures. By estimating the distribution of the return current, more accurate loop-based inductance models have been developed [45, 46, 47, 48].

c) High frequency effects

Inductance is also a function of frequency due to the variation of the current distribution with frequency. In addition to the skin effect mentioned in Subsection 2.3.1, the current distribution inside a conductor also changes with frequency due to the proximity effect [30]. The proximity effect in two parallel interconnects is illustrated in Fig. 2.10. If the current in these two wires flows in opposite directions, the currents concentrate towards each other, as shown in Fig. 2.10(a); otherwise, the two currents shift away from each other, as shown in Fig. 2.10(b). Both the skin effect and the proximity effect are essentially due to the same mechanism — the current

tends to concentrate closer to the current return path in order to minimize the inductance [30]. Note that at high frequencies, the resistance of a conductor also depends on the surrounding signal activities due to the proximity effect.



Figure 2.10: Current distribution in the cross section of two parallel wires at high frequencies due to the proximity effect.

Another effect of frequency on the inductance is due to multi-path current redistribution [49]. In an integrated circuit, there are many possible current return paths, *e.g.*, the power/ground network, nearby signal lines, and the substrate. The distribution of the return current among these possible paths is determined by the impedance of the individual paths. At different frequencies, the relationship among the impedances of different paths will change, as well as the distribution of the return current, as shown in Fig. 2.11. The return current is distributed in those paths so as to minimize the total impedance at a specific frequency [30].

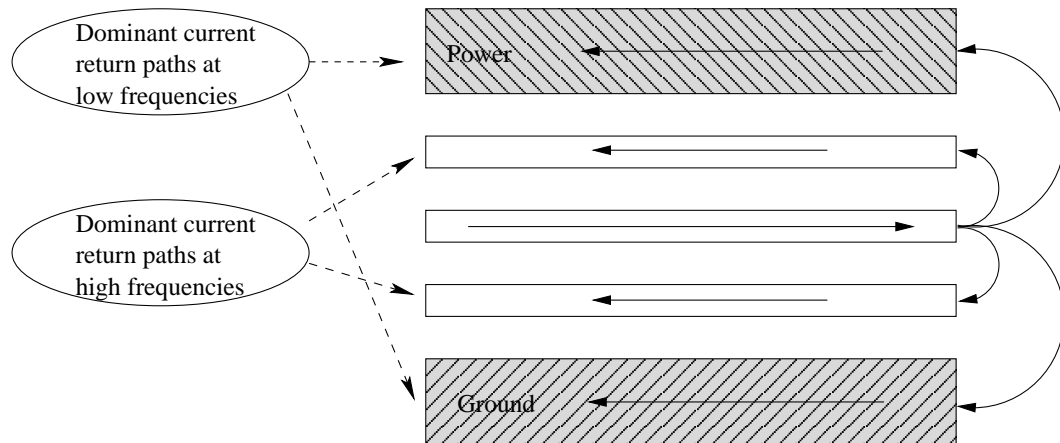


Figure 2.11: Return current distribution at different frequencies.

2.4 Interconnect Models

Interconnect modeling is critical in both the circuit design and verification processes. An efficient and accurate interconnect model can significantly enhance these processes. In Subsections 2.4.1 and 2.4.2, models of single interconnect and coupled interconnects are described, respectively. Model order reduction techniques are reviewed in Subsection 2.4.3.

2.4.1 Single Interconnect

The single interconnect model is the basis for many interconnect network simulation tools. Various on-chip interconnect models have been presented over the past several decades, from lumped $C/RC/RLC$ models to distributed transmission lines. A tradeoff between efficiency and accuracy is required in selecting the appropriate model.

a) Lumped models

For local interconnects with a length of tens of micrometers and below, the circuit behavior is typically dominated by the capacitance and effective resistance of the gates. Modeling the interconnect as a lumped capacitance or lumped RC structure is generally sufficiently accurate. Commonly used lumped models include L, T, and π shaped structures, as depicted in Fig. 2.12.

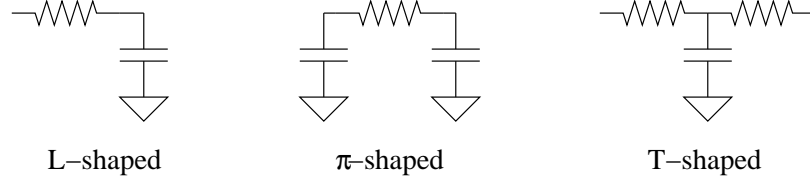


Figure 2.12: Lumped interconnect models.

b) Distributed models

For long intermediate and global interconnects, the signal propagation delay along the interconnect is larger than the gate delay. In this case, the distributed characteristics of the interconnect should be considered. Distributed interconnect can be characterized by the Telegrapher's equations in transmission line theory [50],

$$\frac{\partial V}{\partial x} = -(R + sL)I, \quad (2.13)$$

$$\frac{\partial I}{\partial x} = -sCV, \quad (2.14)$$

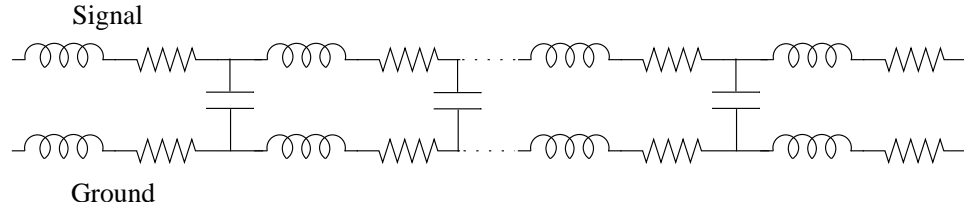
where R , L , and C are the interconnect impedance parameters per unit length, x is the distance along the interconnect, and s is the complex frequency. The conductance between the signal line and ground can typically be ignored in on-chip structures. If the interconnect is non-uniform, these parameters are a function of x . If frequency dependent effects need to be considered, these interconnect parameters are also a function of s . Besides the difficulties in inductance extraction, including inductance in the model also makes circuit analysis more complicated due to inductance induced signal reflection, ringing, and coupling effects. A figure of merit to characterize the condition when on-chip inductance should be considered is presented in [51],

$$\frac{t_r}{2\sqrt{LC}} < l < \frac{2}{R}\sqrt{\frac{L}{C}}, \quad (2.15)$$

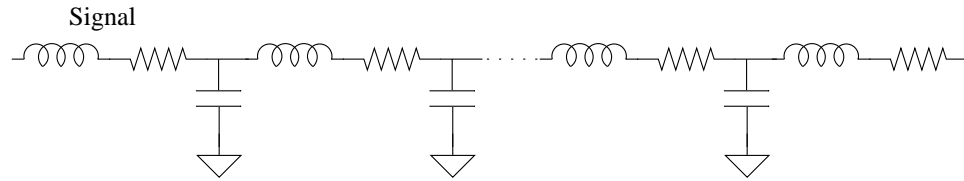
where t_r is the signal transition time and l is the interconnect length.

Transmission line models are based on transverse electro-magnetic (TEM) mode or quasi-TEM mode wave propagation. The TEM or quasi-TEM mode assumption is valid when the line cross-sectional dimension is much smaller than the wavelength [52]. This requirement can be generally satisfied in on-chip structures. For example, the wavelength of a 100 GHz frequency component is on the order of 1 mm, which is several orders greater than the cross-sectional dimension of interconnects in DSM technologies. When using a transmission line model, both the resistance and the inductance should be extracted from the loop formed by the signal line and the

ground return path. Since the resistance of the ground return path is generally much smaller than that of the signal line, the resistance of the ground can be ignored. In a typical circuit representation of a transmission line, the loop inductance is assigned to the signal line as shown in Fig. 2.13(b), rather than the original structure as shown in Fig. 2.13(a). The voltage V in (2.13) is actually the differential voltage between the signal line and the ground line.



(a) Ground line is modeled explicitly.



(b) Ground line is modeled implicitly.

Figure 2.13: Circuit models of transmission lines.

In the capacitance extraction process, assuming the adjacent orthogonal layers as ground also implicitly assumes the voltage on the orthogonal layers follows the voltage on the current return path, which is a reasonable assumption when the voltage on the current return path is small. A more accurate orthogonal layer capacitance model is presented in [53]. In this model, the orthogonal layer is treated as a *supernode*, as

shown in Fig. 2.14. Both the signal line and the ground line in the same layer experience capacitances to the supernode, which are denoted as C_{so} and C_{go} in Fig. 2.14, respectively. This supernode is assumed to float if the node is sufficiently distant from the driver and load. By eliminating this supernode, an equivalent capacitance to ground can be obtained. Applying this method to multiple parallel interconnects, however, results in coupling capacitances between non-adjacent wires [53]. As shown in Fig. 2.14, simply treating the orthogonal layer as ground for capacitance extraction implicitly assumes an infinite C_{go} .

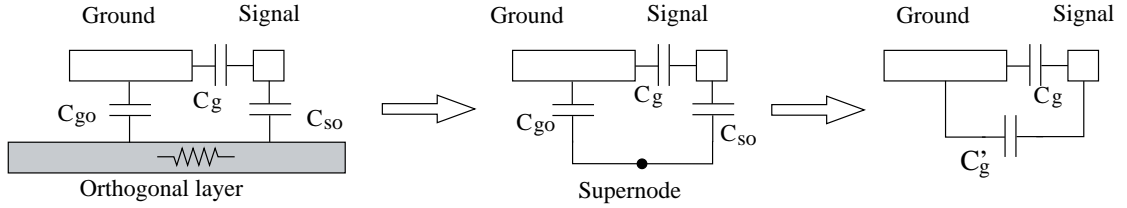


Figure 2.14: Model of orthogonal layer.

c) Lumped representation of distributed interconnects

A transient time domain simulation of a transmission line can be grouped into two categories: impulse response convolution and lumped equivalent circuits [54]. In the first method, the transmission line is initially analyzed in the frequency domain. Next, a time domain impulse response (called a Green's function [55]) is obtained based on the frequency domain solution. Finally, the time domain solution is determined by convolving the Green's function with the voltages at the line ports [55].

Accurate results can be provided with the penalty of long simulation times and excessive memory requirements due to the convolution procedure. Furthermore, this method is not compatible with general circuit simulators, such as SPICE. The second method is to partition the transmission line into a number of segments and model each segment as a lumped structure. Additional segments provide more accurate results, but consume more computational resources. The key issue in this method, therefore, is to determine the appropriate number of segments.

Using lumped models to represent a distributed transmission line introduces inaccuracy when evaluating circuits that operate at high frequencies. The highest frequency of interest, therefore, should be determined in order to evaluate the maximum error induced by using lumped models. The frequency domain representation of a normalized saturated ramp signal with rise time t_r is

$$V_r(s) = \frac{1}{t_r s^2} (1 - e^{-st_r}). \quad (2.16)$$

Inserting $s = j\omega$ into (2.16), the amplitude of the frequency spectrum can be obtained, after some simplification, as

$$|V_r(\omega)| = \frac{t_r}{2} \frac{|\sin x|}{x^2}, \quad (2.17)$$

where

$$x = \omega t_r / 2. \quad (2.18)$$

The normalized amplitude of the frequency spectrum is shown in Fig. 2.15. As shown in Fig. 2.15, the amplitude of the frequency spectrum is infinite at DC and decreases rapidly with increasing frequency.

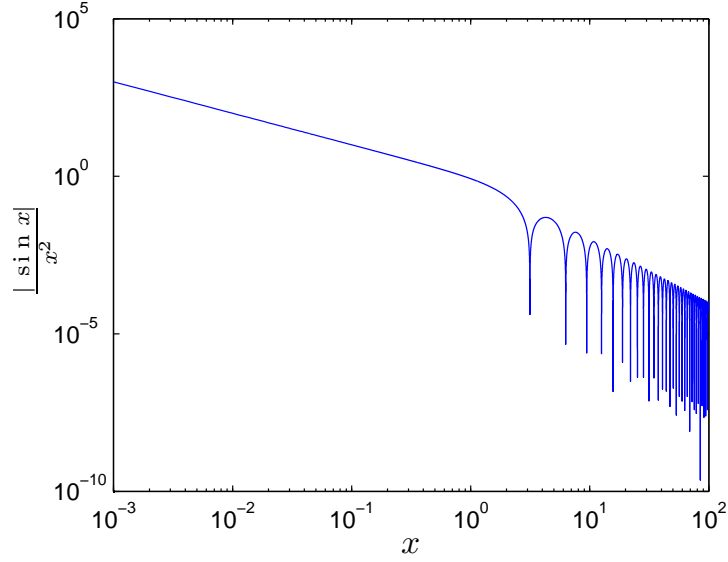


Figure 2.15: Normalized frequency spectrum of a saturated ramp signal.

The normalized integral of the frequency spectrum is defined as

$$S(x) \equiv \frac{\int_0^x \frac{|\sin z|}{z^2} dz}{\int_0^\infty \frac{|\sin z|}{z^2} dz}. \quad (2.19)$$

$S(x)$ is plotted as a function of x in Fig. 2.16. A practical relationship between the maximum frequency f_{max} and the rise time t_r is [54, 56, 52]

$$f_{max} = \frac{0.35}{t_r}. \quad (2.20)$$

From (2.18), this relationship corresponds to $x = 1.1$. Note in Fig. 2.16 that only 9% of the frequency component of a rising edge is at frequencies higher than f_{max} . This percentage increases to about 15% for trapezoidal pulses [53]. Given an error budget of 2.5% on the characteristic impedance, a frequently used rule of thumb to determine the number of lumped segments is theoretically derived in [54] based on the definition of f_{max} given by (2.20): *the propagation delay caused by a segment should be smaller than one fifth of the shortest rise time*. This rule can be mathematically characterized as [54]

$$n \geq \frac{5l\sqrt{LC}}{t_r}, \quad (2.21)$$

where n is the number of segments.

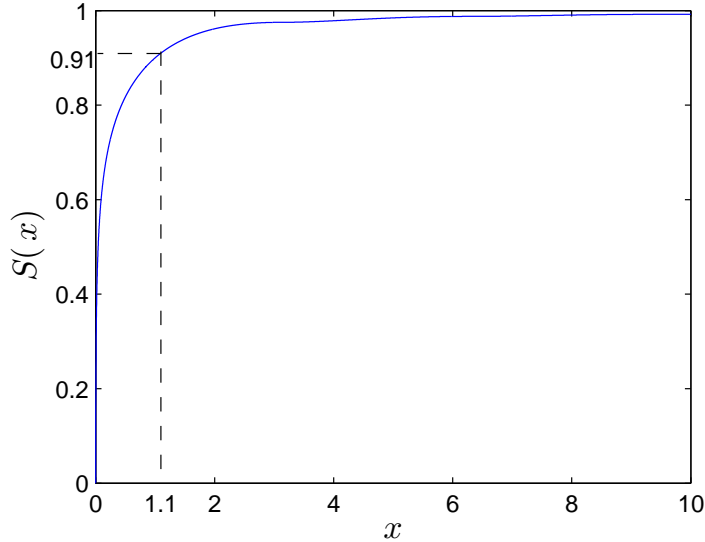


Figure 2.16: Normalized integral of the frequency spectrum of a saturated ramp signal.

d) Modeling frequency dependent effects

After partitioning a distributed line into lumped segments, frequency dependent effects can be modeled in each segment by a ladder structure of frequency independent lumped RL elements, as shown in Fig. 2.17. Additional ladder stages provide higher accuracy when operating at high frequencies. Two stages are used in [45] and three stages are used in [46, 57]. The value of the circuit elements can be obtained by matching the impedance of the model to the extracted impedance at different frequencies.

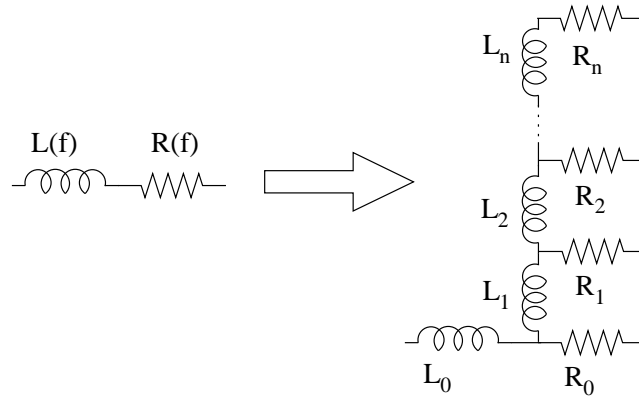


Figure 2.17: Modeling a frequency dependent impedance with lumped elements.

e) Closed-form solutions

By approximating the driver as a voltage source followed by a resistance R_d and the load as a single capacitance C_L , closed-form formulae for the 50% delay of distributed

RC and RLC lines are derived in [58] and [59], respectively,

$$T_{d_{RC}} = 0.377R_tC_t + 0.693(R_dC_L + R_dC_t + R_tC_L), \quad (2.22)$$

$$T_{d_{RLC}} = \frac{e^{-2.9\zeta^{1.35}} + 1.48\zeta}{\omega_n}, \quad (2.23)$$

where

$$\zeta = \frac{R_t}{2} \sqrt{\frac{C_t}{L_t}} \frac{R_T + C_T + R_T C_T + 0.5}{\sqrt{1 + C_T}}, \quad (2.24)$$

$$\omega_n = \frac{1}{\sqrt{L_t(C_t + C_L)}}. \quad (2.25)$$

$R_T = R_d/R_t$, and $C_T = C_L/C_t$. R_t , C_t , and L_t are the total resistance, capacitance, and inductance of the interconnect, respectively. These closed-form expressions play an important role in the interconnect synthesis and optimization phases of the design process.

2.4.2 Parallel Coupled Interconnects

Modeling parallel coupled interconnects draws special attention in the circuit design process due to the commonly used bus structure [60, 61]. A general solution for coupled multiconductor systems is composed of two steps, decoupling the systems into independent interconnects, followed by applying single line models to each of these interconnects. The decoupling procedure is illustrated in Fig. 2.18

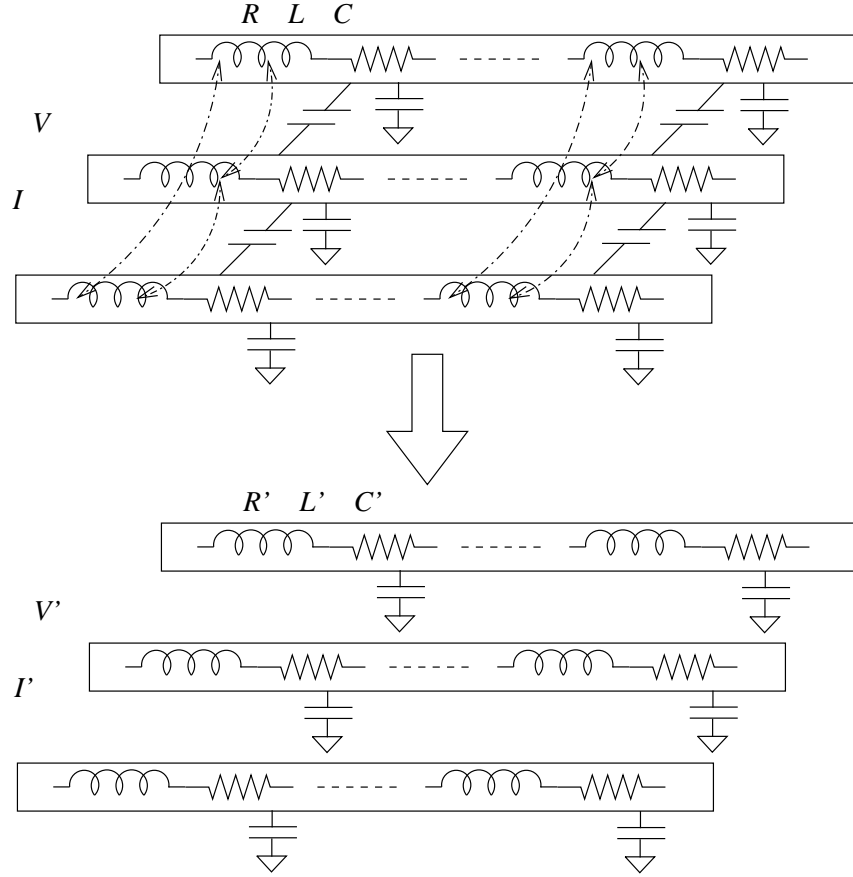


Figure 2.18: Decoupling multiple parallel coupled interconnects.

The Telegrapher's equation describing a coupled multiple interconnect system becomes

$$\frac{\partial \mathbf{V}}{\partial x} = -(\mathbf{R} + s\mathbf{L})\mathbf{I}, \quad (2.26)$$

$$\frac{\partial \mathbf{I}}{\partial x} = -s\mathbf{C}\mathbf{V}, \quad (2.27)$$

where \mathbf{V} and \mathbf{I} are vectors of voltage and current along N coupled interconnects. \mathbf{R} , \mathbf{L} , and \mathbf{C} are the matrices characterizing the impedance parameters per unit length.

The use of (2.26) and (2.27) assumes that the capacitive and inductive coupling among interconnects is restricted in the direction perpendicular to the direction of the signal propagation, *i.e.*, forward coupling [62] is ignored. For well designed circuits, this simplification is often valid [62]. By applying a modal analysis [60, 63], a coupled multiconductor system can be decoupled, *i.e.*, the impedance matrices $\mathbf{R} + s\mathbf{L}$ and $s\mathbf{C}$ in (2.26) and (2.27) can be converted into (much simpler) diagonal matrices. The modal decoupling method, however, generally is not analytically tractable, except for certain special cases, such as two identical interconnects [64], multiple lossless wires [65, 66], wires in a homogeneous dielectric [61], and wires only coupled to direct neighbors [67]. In general, the computational complexity required to decouple a large number of coupled lossy interconnects with a modal analysis is high.

Another commonly used decoupling method is the switch factor based method [68, 69]. Due to the Miller effect, a coupling capacitance C_c between two wires can be modeled as an effective ground capacitance ηC_c , where η depends on the signal switch patterns on the lines and generally ranges between 0 and 2. In [68], the authors demonstrate that the effective capacitance also depends on the slew rates and delay offset of the signals on the two wires, and the range of η is changed to $(-1, 3)$. This switch factor based decoupling method has been extended in [69] to model the effective loop inductance in parallel coupled interconnects. Although not as accurate as the modal analysis, this switch factor based decoupling model is significantly more

computationally efficient and can be used to estimate the delay or delay bounds during the design of global coupled interconnects.

2.4.3 Model Order Reduction

Due to the large number and complex nature of on-chip interconnects, it is impractical to run SPICE-like accurate simulations on an entire IC. A practical approach is to use look-up tables (or fitting parameter based closed-form formulae) to model the gate delay and use model order reduction techniques to simplify the interconnect networks. In the following subsections, generally used model order reduction methods are reviewed, including Elmore delay [70], moment matching [71], and Krylov-subspace based techniques [72, 73, 74].

a) Elmore delay

The Elmore delay was first presented in 1948 [70], where the impulse response is treated as a probability distribution function. As shown in Fig. 2.19, the 50% delay with a step input is the median of the impulse response, *i.e.*, the integral of the impulse response (step response) is divided evenly into two parts by the median. The Elmore delay is the mean of the impulse response and is used to approximate

the median of the impulse response in [70],

$$T_{Elmore} = \int_0^{\infty} h(t)t \, dt. \quad (2.28)$$

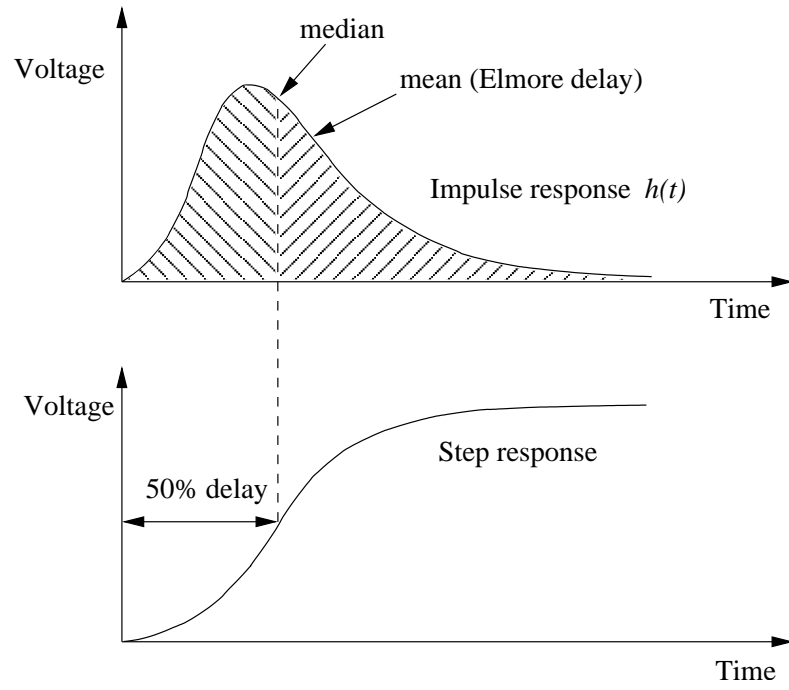


Figure 2.19: Impulse and step responses of RC trees.

Expanding the transfer function into a Taylor series around zero results in

$$\begin{aligned} H(s) &= \int_0^{\infty} h(t)e^{-st} \, dt = \int_0^{\infty} h(t)\left(1 - st + \frac{s^2 t^2}{2!} - \dots\right) \, dt \\ &= 1 - s \int_0^{\infty} h(t)t \, dt + \frac{s^2}{2!} \int_0^{\infty} h(t)t^2 \, dt - \dots \end{aligned} \quad (2.29)$$

The coefficient of different powers of s are referred to as moments of the transfer function. By comparing (2.28) with (2.29), it can be observed that the Elmore delay is the absolute value of the first moment of the transfer function. A derivative form of the Elmore delay is $0.693T_{Elmore}$, which is called the scaled Elmore delay [75]. The scaled Elmore delay can be obtained by approximating the circuit as a one pole system while matching the first moment. The Elmore delay is shown in [76] to be the upper bound of the 50% delay of an RC tree with inputs exhibiting a unimodal derivative. A function $x(t)$ is called unimodal if and only if there exists at least one value t_m such that $x(t)$ is non-decreasing for $t < t_m$ and non-increasing for $t > t_m$ [76].

The Elmore delay is widely used in interconnect synthesis due to the simple closed-form expression, additive property [77], and high fidelity [78], *i.e.*, an optimal solution obtained based on the Elmore delay is also nearly optimal according to the actual delay. The primary disadvantage of Elmore delay is the low accuracy. The resistive shielding effect and effects of inductance can not be captured by the Elmore delay, making it unsuitable for accurate circuit simulation. In [79], an equivalent Elmore delay has been developed that includes the effects of inductance, where the first moment is matched and the second moment is approximated.

b) Moment matching

The moment matching method is generalized in [71], where a q -pole system is obtained by matching the first $2q$ moments (including the zeroth moment). This method is referred to as *asymptotic waveform evaluation* (AWE). By utilizing additional moments, AWE is significantly more accurate than the Elmore delay. The moments at different nodes in an interconnect tree structure can be recursively determined with closed-form expressions [80, 81].

There are three limiting factors preventing AWE from achieving arbitrary accuracy: first, unstable poles may be generated and have to be discarded; second, the computational process becomes unstable with increasing order number [82]; and third, since the moments are based on a Taylor series expansion around zero, the accuracy of the Pade approximation decreases as the frequency increases [52]. Due to these reasons, the number of poles approximated by AWE is typically less than eight [83]. Significant effort has been made to improve AWE with respect to stability and accuracy. In [84], the *complex frequency hopping* (CFH) method is presented, where the moments are matched at multiple expansion points. The *multinode moment matching* (MMM) method is presented in [82], where the spatial information of the moments is utilized, and moments at different nodes are simultaneously matched. In [83], the *Direct Truncation of the Transfer function* (DTT) method is described, where the order of an *RLC* tree is reduced by directly truncating the exact transfer function.

c) Krylov-subspace techniques

In order to achieve higher order, stable, and passive approximations of an interconnect network, another class of model order reduction techniques based on *Krylov-subspace* have been developed in the last decade, such as *Pade via Lanczos* (PVL) [72], *Arnoldi algorithm* [73], and *passive reduced order interconnect macromodeling algorithm* (PRIMA) [74]. In these methods, the moments of the systems are implicitly matched. The reduced model is based on extracting the leading eigenvalues (those with the largest magnitude) of the system rather than extracting the dominant poles in AWE. In [52], it is demonstrated that the poles of a system are the reciprocal of the eigenvalues of the coefficient matrix in the modified nodal analysis (MNA) equations [52]. The essential idea in these Krylov-subspace based methods is to construct a smaller matrix whose eigenvalues are a reasonable approximation of the leading eigenvalues of the original matrix characterizing the system.

2.5 Design Methodologies for Interconnect

Since interconnect plays an important role in ICs, interconnect design methodologies have been developed at different levels to satisfy specific performance requirements. In Subsection 2.5.1, interconnect topology optimization methods are discussed, where interconnect trees are constructed. Wire geometry optimization

methods are reviewed in Subsection 2.5.2. Circuit level interconnect design methodologies are described in Subsections 2.5.3, 2.5.4, and 2.5.5, including buffer insertion, shielding techniques, and net-ordering/wire swizzling, respectively.

2.5.1 Constructing an Interconnect Tree

An interconnect tree network is a commonly used structure in ICs. Signals are transmitted from the root of a tree to each leaf of the tree. When the circuit is dominated by the gates, the interconnects can be modeled as a lumped capacitance. A minimum Steiner tree (MST) is generally constructed in this case such that the total wire length required to connect the source and sinks is minimized. The capacitance of the tree, therefore, is also minimized, as well as the circuit delay and dynamic power.

With the circuit now dominated by the interconnect, both the interconnect resistance and inductance need to be considered during the tree construction process. In this case, the delay at different sinks is different. The required arrival time at each sink is also different. The slack at a node is defined as

$$T_{slack} \equiv T_{rat} - T_{delay}, \quad (2.30)$$

where T_{rat} is the required arrival time at that node and T_{delay} is the delay from the source to that node. In a properly designed tree, the slack at the source should be maximized for high performance while minimizing the area and power overhead.

Some examples of tree constructions are *A-tree* [85], *P-tree* [86], and *C-tree* [87]. In an *A-tree*, the Manhattan distance from the source to each sink is minimized. Subject to this constraint, the total wire length is also minimized. An example of an *A-tree* is illustrated in Fig. 2.20. During constructing of a *P-tree*, the solution space is limited to a set of topologies induced by a permutation on the sinks. From this solution space, the optimal solution is chosen based on the delay or delay-area product. In the *C-tree*, the sinks are first clustered according to the spatial, temporal, and polarity properties. After the clustering procedure, tree structures are built within and among these clusters.

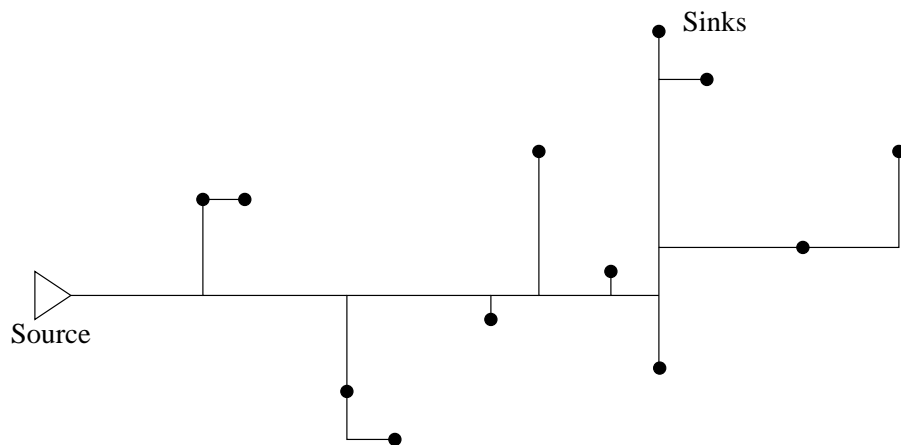


Figure 2.20: An example of an *A-tree*.

2.5.2 Wire Sizing, Shaping, and Spacing

Given a metal layer in a specified technology, the thickness of the wires and inter-layer dielectric (ILD) is fixed. The wire width and space, however, can be

varied to satisfy different design criteria. By explicitly characterizing the relationship between the interconnect impedance and wire geometries, tradeoffs among the delay, bandwidth, and power of the global interconnect can be made [19, 21, 88]. In [89], the effects of inductance are included during the wire width optimization process to lower the power dissipation.

It is shown in [90] that the optimal shape of an RC interconnect that minimizes the Elmore delay is an exponential taper, as shown in Fig. 2.21. Wire tapering increases the wire width near the driver and decreases the wire width near the load. Since the near end resistance sees more downstream capacitance than the far end resistance, assigning less resistance to the near end than to the far end will reduce the total RC delay. In [91], the optimal shape of an RLC interconnect is also shown to be exponential. Exponential shaping, however, is more difficult to implement than uniformly sized wires.

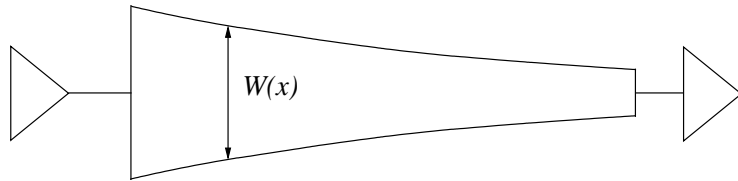


Figure 2.21: Shaping interconnect to minimize delay.

2.5.3 Repeater Insertion

The delay of an RC interconnect is $0.377RCl^2$ [58], which is proportional to the square of the wire length l . By splitting the interconnect into k segments with repeaters, the interconnect delay term is reduced to $0.377RCl^2/k$. These repeaters, however, introduce additional gate delay. The optimal number and size of the repeaters can be determined to achieve the minimum delay [12, 59]. As signals propagate along the interconnect, sharper transition edges are regenerated by the repeaters, increasing the bandwidth of the interconnect. By dividing the interconnect into segments, the coupling between interconnects is also reduced due to the shorter length of coupling between neighboring lines. Inserting repeaters in long interconnects, however, introduces an area and power penalty. A tradeoff among different design criteria is, therefore, required for an efficient repeater insertion methodology. This topic is discussed in greater detail in Chapter 6.

In [92], a repeater staggering technique is proposed to reduce the worst case delay and crosstalk noise in bus structures. As shown in Fig. 2.22, the repeaters in adjacent wires are interleaved. By placing a repeater in the middle of two repeaters in adjacent wires, a potential worst case capacitive coupling only persists for half the wire length. For the other half length, the capacitive coupling is the best case. The worst case delay as well as the delay uncertainty can therefore be reduced. One of the advantages of this technique is that no additional area overhead is required. By staggering the

repeaters, the inductive coupling among the wires can also be averaged. As shown in Fig. 2.22, for two simultaneously switching adjacent wires, the direction of current is the same for half the wire length and opposite for the other half length. Inductive coupling due to the current flowing in different directions in the neighboring wire can be cancelled. In [93], the optimum position of staggered repeaters is determined for *RC* interconnect to achieve the minimum worst case delay.

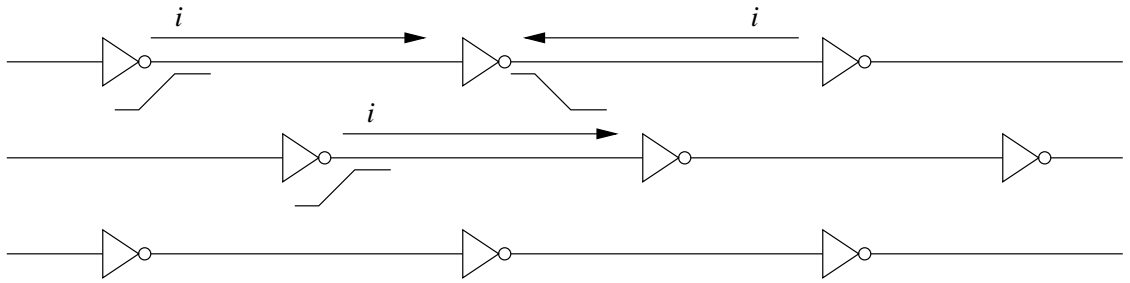


Figure 2.22: Staggering repeaters to reduce the worst case delay and crosstalk noise.

Another significant application of repeater insertion is the buffered tree. The repeaters inserted in an interconnect tree are also called buffers. Buffer insertion in tree structures is an important design tool for interconnect optimization. In [13], van Ginneken presented a dynamic programming algorithm to insert buffers in a Steiner tree to minimize the Elmore delay. van Ginneken's algorithm is composed of two phases. The first phase is a bottom-up process, where all of the possible buffer insertion candidates are determined for each node in the tree. In this process, those suboptimal candidates are eliminated such that the number of candidates does not increase exponentially. After the candidates at the root are determined, the candidate with the

maximum slack is chosen. The second phase traces back the computations in the first phase from this candidate and places buffers at the appropriate locations. Various extensions to this algorithm have been presented in the last decade which consider low power [94], blockage constraints [95], and more accurate delay models [96]. In a properly designed buffered tree, as shown in Fig. 2.23, the buffers should be inserted in the following situations:

- Splitting long interconnect (buffers 1 and 2);
- Isolating large capacitances from the critical path (buffer 3);
- Cascading buffers to drive large capacitances (buffers 4, 5, and 6);
- Reversing the signal polarity if necessary (inverter 7).

Note that interconnect tree construction, buffer insertion, and wire sizing can be performed simultaneously in order to achieve an optimal solution.

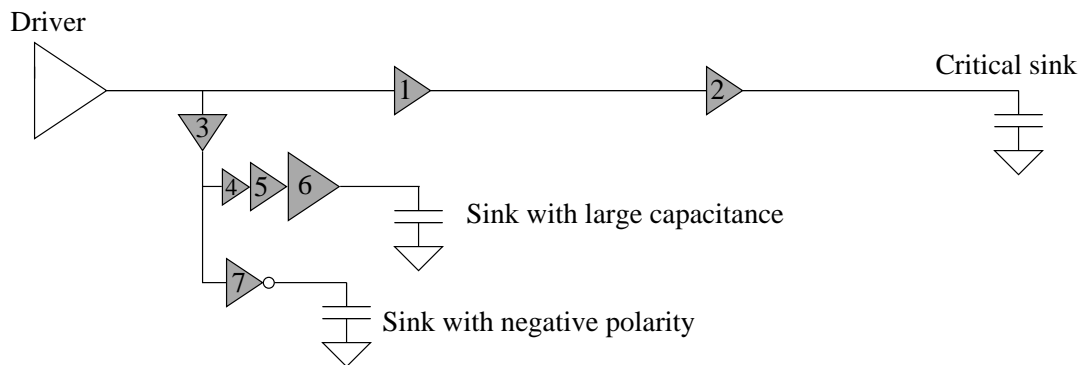


Figure 2.23: Buffered interconnect tree.

2.5.4 Shielding Techniques

Shielding techniques are widely used in ICs to reduce capacitive and inductive coupling. By inserting a shield line (generally connected to the power or ground grid) between signal lines, the effective capacitance of the interconnect is almost fixed and no longer depends upon the signal switching activity. With shielding, the normalized peak crosstalk noise can be reduced to less than 5% of V_{dd} for RC interconnect with lengths ranging up to 2 mm [97].

Inductive coupling can also be reduced by inserting a shield line, though not as efficiently as reducing capacitive coupling due to the long range magnetic coupling property. The shield line provides a nearby current return path, reducing the self and mutual inductance of the signal lines. Due to the importance of the on-chip clock signal, the clock distribution network in a high speed circuit is generally shielded on both sides in the same layer [98]. Additional parallel shielding in the N-2 layer has been reported in [99] to further prevent inductive coupling from the lower layers. The primary drawback of the shielding technique is the overhead of the metal resources.

2.5.5 Net-Ordering and Wire Swizzling

Interconnect coupling is closely related to the signal switching activity. For example, simultaneously opposite switching on two adjacent RC lines produces the worst case delay [100]. By ordering the nets such that the sensitive nets are not placed

adjacent to each other, the total capacitive coupling among the nets can be minimized [101]. Examples of net-ordering and wire swizzling are shown in Fig. 2.24. The net-ordering technique, however, is less efficient in reducing long range inductive coupling. In [102], the net-ordering and shield insertion techniques are simultaneously performed to minimize both capacitive and inductive coupling.

In wire swizzling, the wires are split into several segments, and the wire sequences in each segment are changed, such that the capacitive coupling among the wires averages out for each wire, reducing both the worst case delay and the delay uncertainty [103]. For a group of k wires, the number of permutations required to realize all possible adjacencies is $k/2$. For the example shown in Fig. 2.24, $k = 4$ and two permutations are required: 1234 and 2413. In [104], it is also shown that the mutual inductance in a bus structure can be reduced by wire swizzling.

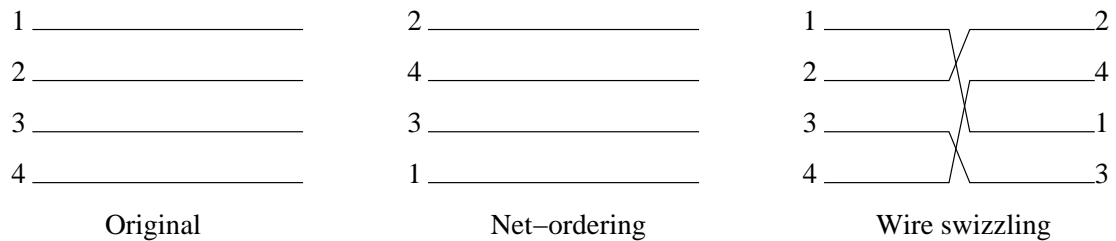


Figure 2.24: Examples of net-ordering and wire swizzling.

2.6 Conclusions

A brief review of electrical on-chip interconnect is presented in this chapter. Design constraints on different criteria have become more stringent in the DSM regime. With higher operating frequencies and smaller wire dimensions, the interconnect parasitic extraction process is also more complicated due to various factors. Distributed *RLC* interconnect models and model order reduction techniques are necessary to analyze circuit performance. Design methodologies at different levels are needed to optimize the interconnect, from wire geometries to layout topologies. Although tremendous effort has been expended over the past two decades, the analysis and design of on-chip interconnect remains an increasingly challenging task in present and future IC technologies.

Chapter 3

An *RLC* Interconnect Model Based on Fourier Analysis

3.1 Introduction

In DSM ICs, interconnect delay dominates gate delay. Furthermore, wire inductance can no longer be ignored due to higher signal frequencies and longer wire lengths [59]. Accurate and efficient *RLC* interconnect models are therefore critical in the design of high performance ICs.

Based on modified Bessel functions, expressions characterizing the transient response of an *RLC* interconnect have been rigorously developed in [105] and [106]. These results, however, are highly complicated and not suitable for an exploratory design process. In order to produce a more efficient solution, the transfer function of the interconnect is truncated and approximated with a few dominant poles, for example, one or two poles in [107], [108], and four poles in [109]. Four pole expressions

are highly accurate, however, no closed form solution has been developed in [109]. In all of these models, a step or ramp input is assumed and no initial conditions are considered. For a periodic signal, however, the initial conditions can have a significant effect on the output waveform.

The performance of a synchronous circuit is heavily dependent on the design of a clock distribution network. *RLC* interconnect trees are common structures in clock networks. An accurate model of an *RLC* interconnect tree, therefore, is critical in modern digital circuit design. In [79] and [107], second order models are used to analyze *RLC* trees. The accuracy of these models, however, is limited. In order to obtain a more accurate result, model order reduction techniques can be adopted at the expense of additional computational complexity, as described in Chapter 2.

With the scaling of semiconductor technologies, interconnect crosstalk has become another important issue. Crosstalk can be caused by either (or both) capacitive coupling and inductive coupling. Capacitive coupling is a short range effect, where typically only adjacent lines need be considered. On the contrary, inductive coupling is a long range effect and is significantly more difficult to analyze. For multiconductor transmission lines, modal analysis [66], [60] is a widely used decoupling method. This decoupling method is extended to drivers and loads in [64] and [61] for two and more interconnects. The extensions, however, are only valid for identical lines with identical drivers and loads.

In this chapter, a new interconnect timing model is presented. The model is based on a Fourier series analysis of a periodic input signal. No approximation is made to the transfer function of the interconnect. The far end response is approximated by the summation of several sinusoids. Since the solution is the steady state response to a periodic signal, the initial conditions are considered. The model is verified by SPICE simulations and successfully extended to *RLC* trees and multiple transmission lines. The rest of this chapter is organized as follows. In Section 3.2, the Fourier series based interconnect model for a single line is described. In Section 3.3, the model is applied to *RLC* trees, and a tree model with linear computational complexity is obtained. Combined with the modal analysis, the proposed model is extended to multiple interconnect lines in Section 3.4 to analyze crosstalk noise. Finally, some conclusions are offered in Section 3.5.

3.2 Single Interconnect Model

The exact transfer function of a widely used interconnect circuit model is described in Subsection 3.2.1, and compared with the transfer functions of some approximate models. A Fourier series analysis of a typical on-chip signal is presented in Subsection 3.2.2. Based on this analysis, an expression for the time domain response at the far end of an interconnect is presented in Subsection 3.2.3. Closed form solutions for the 50% delay and overshoot/undershoots are presented in Subsection 3.2.4. In

Subsection 3.2.5, results from this model are compared with SPICE. A maximum error of about 11% is exhibited.

3.2.1 Interconnect Transfer Function

A classical interconnect circuit model is shown in Fig. 3.1. The interconnect is represented by a distributed RLC transmission line, where l is the interconnect length, and R , L , C are the resistance, inductance, and capacitance per unit length, respectively. The driver is linearized as a voltage source V_{in} serially connected with a driver resistance R_d . The load of the interconnect is modeled as a capacitor C_l .

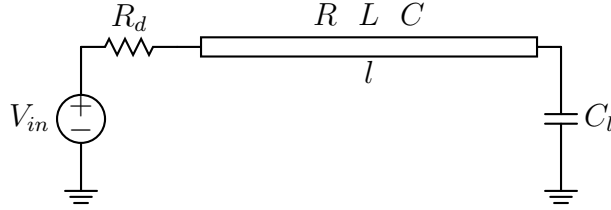


Figure 3.1: Equivalent circuit model of a distributed RLC interconnect.

This equivalent circuit is a linear time-invariant (LTI) system. For LTI systems, the time domain response can be solved by an inverse Laplace transform. From the $ABCD$ parameters [50] of a transmission line, the transfer function from the input to the far end of a line is

$$H(s) = \frac{1}{(1 + R_d C_l s) \cosh(\gamma l) + (R_d / Z_c + Z_c C_l s) \sinh(\gamma l)}, \quad (3.1)$$

where $\gamma = \sqrt{(R + sL)sC}$ and $Z_c = \sqrt{(R + sL)/sC}$ are the propagation coefficient and characteristic impedance, respectively. Since (3.1) includes hyperbolic functions of the complex frequency s , the inverse Laplace transform is difficult to derive directly. In order to simplify the problem, the denominator of the transfer function is expanded into an infinite series. By truncating this series, the transfer function is approximated by a few dominant poles [107], [109]. A distributed RLC line can also be modeled by lumped elements through moment matching [110].

In Fig. 3.2, the transfer function of some existing models [107], [109], [110] are compared with the exact transfer function described in (3.1). In this example, the interconnect parameters are $l = 2$ mm, $R = 8.829$ m $\Omega/\mu\text{m}$, $L = 1.538$ pH/ μm , and $C = 0.18$ fF/ μm . The per unit length parameters are calculated with FastHenry [44] and FastCap [32] for the top layer metal interconnect in a standard $0.18\mu\text{m}$ CMOS technology. The interconnect has a width $w = 2\mu\text{m}$ and a height $h = 1\mu\text{m}$. Partial inductance is used here to emphasize the inductive effect, where the current return path is assumed at infinity. As described in Subsection 2.3.3, the partial inductance overestimate the inductive effect, since nearby current return paths reduce the effective inductance. The driver resistance and load capacitance are $R_d = 30\Omega$ and $C_l = 50$ fF, respectively. The interconnect parameters from this example are used in the rest of this chapter. As illustrated in Fig. 3.2, for this example, a simple L-type lumped model produces the poorest approximation. The two pole model can

be accurate up to 5 GHz. A non-uniform two stage L-type lumped model is a fourth order approximation and has a similar accuracy range as the four pole model, which is accurate up to 9 GHz; however, no closed form solutions for these two models have been reported.

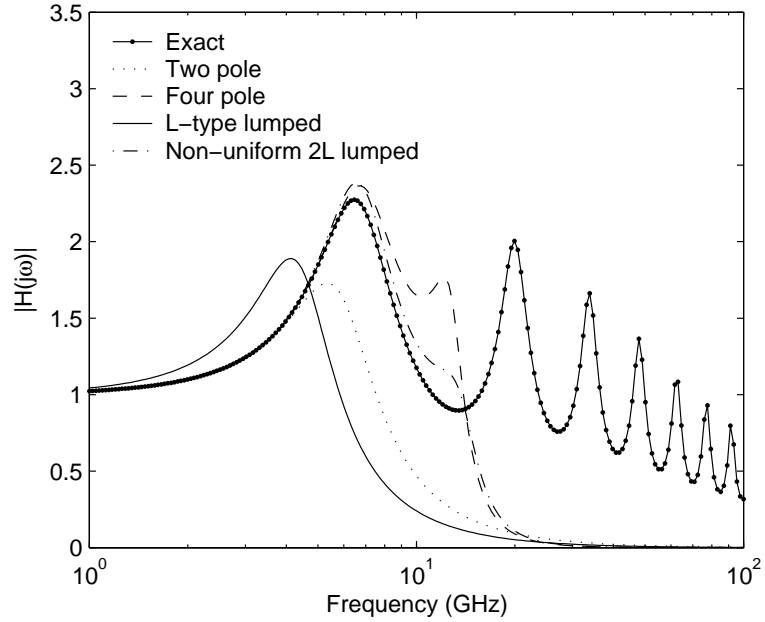


Figure 3.2: The amplitude transfer function of different models of an *RLC* interconnect.

The resonance frequencies (where the peaks occur in the exact transfer function) of the system are related to the poles of the transfer function. A non-uniform 2L model and a four pole model can track the first peak of the exact transfer function, which means these two models can accurately model two poles of the system (the other pole is in the negative frequency domain). The resonance frequencies are due to the reflection of the signal at the terminals, therefore, the resonance frequencies are

approximately multiples of $1/4t_f$, where $t_f = l\sqrt{LC}$ is the time-of-flight. The high peaks in the transfer function indicate strong inductive effects. If the interconnect is RC dominant, the amplitude transfer function has no overshoots and decreases quickly with increasing frequency. In Fig. 3.3, the transfer functions with different inductive effects are shown. The inductance effects are characterized by a parameter ζ [59], and in this example, ζ is varied by changing R_d . A small ζ implies significant inductive effects. As shown in Fig. 3.3, when $\zeta = 0.59$ which corresponds to $R_d = \sqrt{L/C}$ (the character impedance of the interconnect at high frequencies), the reflection coefficient Γ_s at the source is zero, thus no resonance effects occur. When R_d is greater than $\sqrt{L/C}$, ζ is greater than 0.59, Γ_s is positive, and the basic resonance frequency is about $1/2t_f$. Alternatively, when R_d is less than $\sqrt{L/C}$, ζ is less than 0.59, Γ_s is negative, and the basic resonance frequency is about $1/4t_f$.

3.2.2 Fourier Series Representation of Input Signal

In previous analytical timing models, the excitation signal is modeled as a step or ramp function, and most of the effort is focused on the transfer function. In this chapter, however, a different approach is presented which focuses on the input signal.

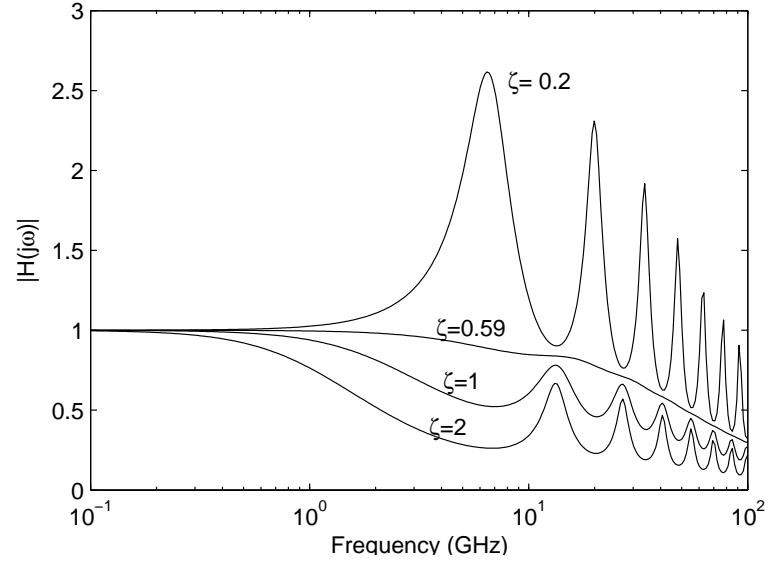


Figure 3.3: The amplitude transfer functions of an *RLC* interconnect with different inductive effects.

The input signal is approximated by a periodic ramp signal [111],

$$V_{in}(t) = \begin{cases} \frac{t - nT}{\tau} V_{dd} & nT \leq t < nT + \tau, \\ V_{dd} & nT + \tau \leq t < (n + \frac{1}{2})T, \\ \left(1 - \frac{t - nT}{\tau} + \frac{T}{2\tau}\right) V_{dd} & (n + \frac{1}{2})T \leq t < (n + \frac{1}{2})T + \tau, \\ 0 & (n + \frac{1}{2})T + \tau \leq t < (n + 1)T, \end{cases} \quad (3.2)$$

where T is the period of $V_{in}(t)$, n is an integer, and τ is the transition time. As is well known, a periodic signal can be represented as a summation of a Fourier series.

The Fourier series representation of $V_{in}(t)$ is

$$V_{in}(t) = \frac{V_{dd}}{2} + \sum_{m=1,3,\dots} A_m \sin(m\omega_0 t + \phi_m), \quad (3.3)$$

$$\phi_m = -\frac{m\omega_0\tau}{2}, \quad (3.4)$$

$$A_m = \frac{2TV_{dd}}{\tau m^2 \pi^2} |\sin \phi_m|, \quad (3.5)$$

where $\omega_0 = 2\pi/T$ is the basis angular frequency, and A_m and ϕ_m are the amplitude and phase of the m^{th} order harmonic, respectively. From (3.3), $V_{in}(t)$ is composed of the DC component and odd order harmonics. Since A_m decreases quadratically with m , $V_{in}(t)$ can be approximated by the first several harmonics [111]. The normalized amplitude of the odd order harmonics is shown in Fig. 3.4 for different τ/T . Note in Fig. 3.4 that the decrease in A_m slows with decreasing τ/T . In the limiting case, $\tau/T = 0$, and $A_m = 2V_{dd}/(m\pi)$, which is reciprocally proportional to m .

3.2.3 Far End Time Domain Response

Since the circuit shown in Fig. 3.1 is linear and the input signal can be represented by a summation of harmonics, the superposition rule can be used to determine the output signal. The transfer function at each angular frequency ω can be represented as

$$H(j\omega) = H(s)|_{s=j\omega} = A(\omega)e^{j\beta(\omega)}. \quad (3.6)$$

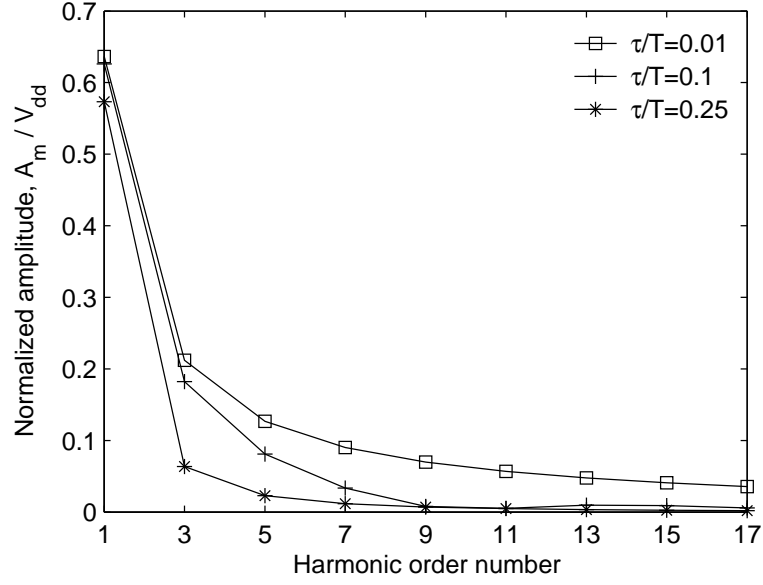


Figure 3.4: Normalized amplitude of odd order harmonics.

From (3.1), the gain of the DC component is $H(0) = 1$. The output, therefore, is

$$V_{out}(t) = \frac{V_{dd}}{2} + \sum_{m=1,3,\dots} A'_m \sin(m\omega_0 t + \phi'_m), \quad (3.7)$$

$$A'_m = A_m A(m\omega_0), \quad (3.8)$$

$$\phi'_m = \phi_m + \beta(m\omega_0). \quad (3.9)$$

$V_{out}(t)$ can also be approximated by the first several lower order harmonics. In this chapter, the Fourier series based models are referred to as Fb3 and Fb5, with the largest harmonic order number of three and five, respectively. The results from Fb3 and Fb5 are compared with SPICE in Fig. 3.5. The input signal parameters are $T = 500$ ps, $\tau/T = 0.1$, and $V_{dd} = 1.5$ volts. In the SPICE simulation, the interconnect

line is divided into 200 segments and each segment is represented by an L-type lumped model. As shown in Fig. 3.5, two harmonics (Fb3) are sufficient to provide a good approximation of the output voltage waveform for this example.

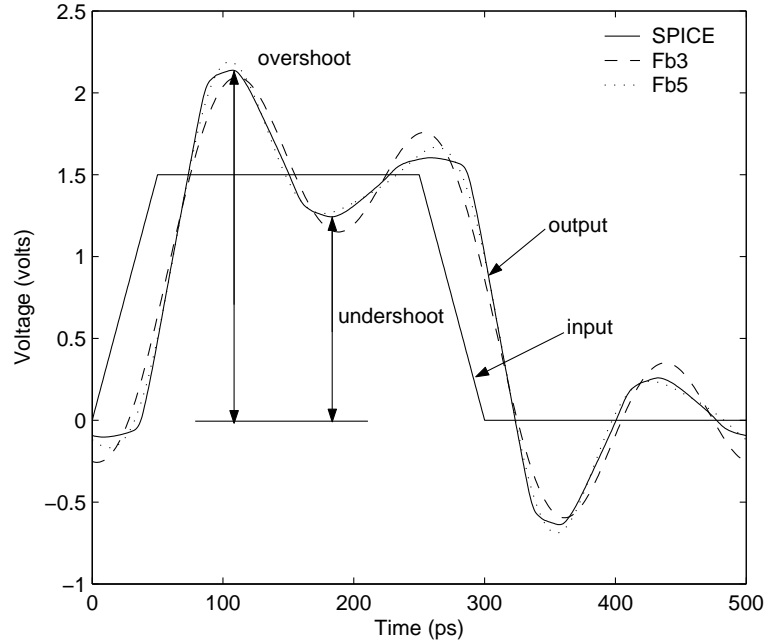


Figure 3.5: Comparison of the time domain response of Fb3 and Fb5 with SPICE.

3.2.4 The 50% Delay and Overshoots/Undershoots

The 50% delay and overshoots/undershoots can be solved numerically from (3.7). In this chapter, the 50% delay is assumed to be less than $T/2 - \tau/2$ (valid in most practical cases), and the overshoots/undershoots caused by the rising edge are measured between the waveform and ground, as shown in Fig. 3.5. For Fb3, since only

two harmonics are considered, a closed form solution is available. In this case,

$$V_{out}(t) \approx \frac{V_{dd}}{2} + A'_1 \sin(\omega_0 t + \phi'_1) + A'_3 \sin(3\omega_0 t + \phi'_3). \quad (3.10)$$

To determine the 50% delay, (3.10) is set to $V_{dd}/2$. By applying the multiple-angle formulae [112], a third order trigonometric expression can be obtained,

$$a_3 x^3 + a_2 x^2 + a_1 x + a_0 = 0, \quad (3.11)$$

where $x = \tan(\omega_0 t)$ and

$$a_0 = A'_1 \sin \phi'_1 + A'_3 \sin \phi'_3, \quad (3.12)$$

$$a_1 = A'_1 \cos \phi'_1 + 3A'_3 \cos \phi'_3, \quad (3.13)$$

$$a_2 = A'_1 \sin \phi'_1 - 3A'_3 \sin \phi'_3, \quad (3.14)$$

$$a_3 = A'_1 \cos \phi'_1 - A'_3 \cos \phi'_3. \quad (3.15)$$

A third order expression has either one or three real roots, and a closed form solution exists [64]. If (3.11) has only one real root x_0 , the output waveform crosses $V_{dd}/2$ only once from low-to-high during the first half of a period, therefore the undershoot

is greater than $V_{dd}/2$. From this real root, the 50% delay can be expressed as

$$t_d = \frac{\arctan x_0}{\omega_0} - \frac{\tau}{2}. \quad (3.16)$$

The value of $\arctan x_0$ is in the range of $[0, \pi]$. If (3.11) has three real roots, the output waveform crosses $V_{dd}/2$ three times during the first half of the period, therefore the undershoot is less than $V_{dd}/2$. In this case, the output waveform is not shaped like a square wave and can no longer represent logic values.

The process for determining the overshoots/undershoots is similar to that of the delay. From (3.10), the derivative of V_{out} is

$$\frac{dV_{out}}{dt} \approx A'_1 \omega_0 \cos(\omega_0 t + \phi'_1) + 3A'_3 \omega_0 \cos(3\omega_0 t + \phi'_3). \quad (3.17)$$

Setting (3.17) to zero and applying the multiple-angle formulae, the following third order trigonometric expression is obtained,

$$b_3 y^3 + b_2 y^2 + b_1 y + b_0 = 0, \quad (3.18)$$

where $y = \tan(\omega_0 t)$ and

$$b_0 = A'_1 \omega_0 \cos \phi'_1 + 3A'_3 \omega_0 \cos \phi'_3, \quad (3.19)$$

$$b_1 = -A'_1 \omega_0 \sin \phi'_1 - 9A'_3 \omega_0 \sin \phi'_3, \quad (3.20)$$

$$b_2 = A'_1 \omega_0 \cos \phi'_1 - 9A'_3 \omega_0 \cos \phi'_3, \quad (3.21)$$

$$b_3 = -A'_1 \omega_0 \sin \phi'_1 + 3A'_3 \omega_0 \sin \phi'_3. \quad (3.22)$$

The time when the extremum occurs can be obtained from the real roots of (3.18). Note that the time obtained can be less than t_f . This behavior occurs because the voltage response described by (3.7) is a steady state response. The extremum which occurs before t_f is the response to the previous period of the input signal. For the response to the current period, the time when the extremum occurs should be

$$t_p = \begin{cases} \arctan y_0 & \arctan y_0 > t_f, \\ \arctan y_0 + \frac{T}{2} & \arctan y_0 \leq t_f, \end{cases} \quad (3.23)$$

where y_0 is a real root of (3.18). The corresponding extremum can be determined by inserting t_p into (3.10),

$$V_{ex} \approx \frac{V_{dd}}{2} + A'_1 \sin(\omega_0 t_p + \phi'_1) + A'_3 \sin(3\omega_0 t_p + \phi'_3). \quad (3.24)$$

The overshoot and undershoot are chosen as the maximum and minimum of the results obtained in (3.24), respectively.

If higher accuracy is required, more harmonics should be included in the model, and higher order (fifth, seventh,...) equations should be solved. Since only real roots are of interest, some efficient root-finding algorithms can be used, such as the Newton-Raphson method. The complexity, however, increases.

Since the output waveform is approximated by a summation of sinusoids, some of the undershoots obtained are not real undershoots (called *pseudo-undershoots* in this chapter) and should be discarded. By comparing the waveforms obtained by the model with SPICE simulations, three such cases are found:

1. There is only one extremum;
2. The last extremum (according to the time index) is the largest;
3. All extremum values are greater than V_{dd} .

In these cases, either the output is overdamped or the period is too short for the waveform to achieve the undershoot within half a period. Examples of different cases are shown in Fig. 3.6.

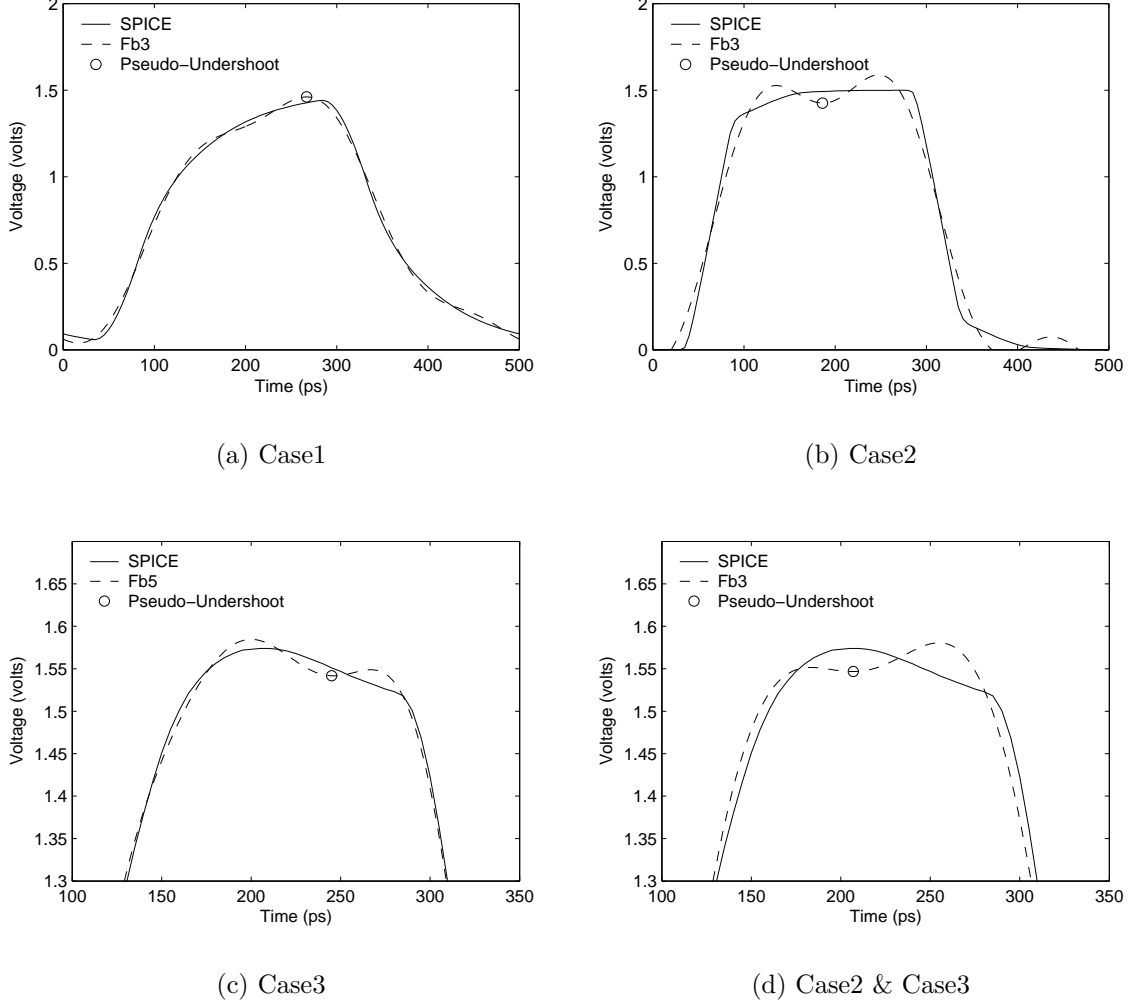


Figure 3.6: Examples of pseudo-undershoots. The input signal parameters are $T = 500$ ps, $\tau/T = 0.1$, and $V_{dd} = 1.5$ volts. The driver and load parameters are (a) $R_d = 100 \Omega$ and $C_l = 500$ fF, (b) $R_d = 100 \Omega$ and $C_l = 50$ fF, (c) $R_d = 60 \Omega$ and $C_l = 500$ fF, and (d) $R_d = 60 \Omega$ and $C_l = 500$ fF.

3.2.5 Model Verification and Discussion

The 50% delay calculated with the proposed model is compared with SPICE in Table 3.1. The interconnect parameters for $w = 6 \mu\text{m}$ are $R = 3.35 \text{ m}\Omega/\mu\text{m}$, $L =$

1.36 pH/ μm , and $C = 0.33 \text{ fF}/\mu\text{m}$. The interconnect parameters for $w = 10 \mu\text{m}$ are $R = 2.2 \text{ m}\Omega/\mu\text{m}$, $L = 1.26 \text{ pH}/\mu\text{m}$, and $C = 0.49 \text{ fF}/\mu\text{m}$. Results from a single pole model with a ramp input [108] are also listed. As expected, the single pole model is accurate only when the circuit is dominated by the RC impedance. When the circuit is dominated by the LC impedance, the error is large. However, the proposed Fourier series based method provides accurate delay estimates for both RC -dominated and LC -dominated circuits. The average error of Fb5 is only 0.6% over a wide range of circuit parameters (the parameters are selected so that the bandwidth requirement is satisfied). The overshoots/undershoots for underdamped responses resulting from Fb3 and Fb5 are compared with SPICE in Table 3.2. As listed in Tables 3.1 and 3.2, the model becomes more accurate with additional harmonics.

In Tables 3.1 and 3.2, the delay and overshoots/undershoots obtained from Fb3, Fb5, and SPICE characterize the steady state response. If the response to a rising edge (or falling edge) cannot converge to V_{dd} (or 0) within half a period, the charge and current at the end of a period become the initial conditions of the following period. These initial conditions, however, can have a significant effect on propagating high frequency signals along long interconnects. The far end response to a single ramp input and a periodic ramp input are compared in Fig. 3.7. As shown in Fig. 3.7, the position and value of the overshoot are quite different for the two responses. For periodic signals, the method presented here is more suitable than other models.

Table 3.1: Comparison of the 50% delay of Fb3 and Fb5 with SPICE and a single pole model. The input signal parameters are $T = 500$ ps, $\tau = 50$ ps, and $V_{dd} = 1.5$ volts. The interconnect parameters are $l = 2$ mm and $h = 1\mu\text{m}$.

w (μm)	R_d (Ω)	C_l (fF)	SPICE (ps)	1-pole (ps)	Fb3 (ps)	Fb5 (ps)
2	20	50	28.9	11.5	25.9	29.0
2	60	100	40.1	25.7	39.1	40.0
2	100	500	73.5	69.0	77.0	74.2
6	20	50	41.8	14.8	40.7	42.0
6	40	100	45.4	26.1	44.4	45.4
6	60	500	68.6	53.5	71.3	69.5
10	20	50	47.2	19.0	45.9	46.9
10	40	100	53.1	34.0	52.6	52.9
10	60	500	74.5	65.7	77.1	75.3
Maximum Error				64.6%	10.4%	1.3%
Average Error				37.7%	3.7%	0.6%

Table 3.2: Comparison of overshoots/undershoots of Fb3 and Fb5 with SPICE simulations. The input signal parameters are $T = 500$ ps, $\tau = 50$ ps, and $V_{dd} = 1.5$ volts. The interconnect parameters are $l = 2$ mm and $h = 1\mu\text{m}$.

w (μm)	R_d (Ω)	C_l (fF)	Overshoot (volts)			Undershoot (volts)		
			SPICE	Fb3	Fb5	SPICE	Fb3	Fb5
2	20	10	2.30	2.11	2.29	1.08	1.08	1.08
2	30	50	2.14	2.10	2.19	1.24	1.15	1.26
2	60	100	1.71	1.77	1.75	1.47	1.40	1.44
6	20	10	2.29	2.35	2.40	1.13	1.00	1.12
6	30	50	2.00	2.08	2.08	1.34	1.25	1.31
6	40	100	1.80	1.89	1.85	1.44	1.38	1.40
10	20	10	2.11	2.21	2.17	1.26	1.21	1.19
10	30	50	1.85	1.96	1.91	1.42	1.38	1.36
10	40	100	1.65	1.75	1.69	1.49	1.44	1.44
Maximum Error			8.3%	4.8%	—	11.5%	5.6%	—
Average Error			4.7%	2.8%	—	4.9%	2.5%	—

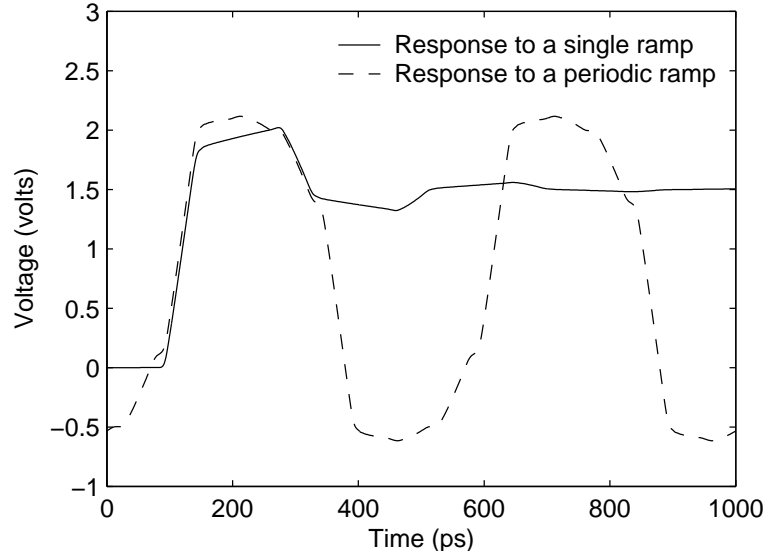


Figure 3.7: The effect of initial conditions on the periodic signals. $l = 5$ mm.

In Fig. 3.8, the delay model is examined for various interconnect lengths. The interconnect inductance is determined for each interconnect length, since the partial inductance does not increase linearly with line length. Another advantage of the proposed model is that frequency dependent effects of the interconnect can be directly included, since the transfer function is calculated at each individual frequency.

The accuracy of the proposed model depends upon the frequency spectrum of the far end response. If most of the signal energy is allocated in the lower order harmonics, neglecting those higher order harmonics will cause little error and the model is accurate; otherwise, the accuracy of the model will decrease. From Figs. 3.3 and 3.4, it can be concluded that the accuracy becomes worse for signals with small τ/T propagating along highly inductive interconnects. The 50% delay with different τ/T is

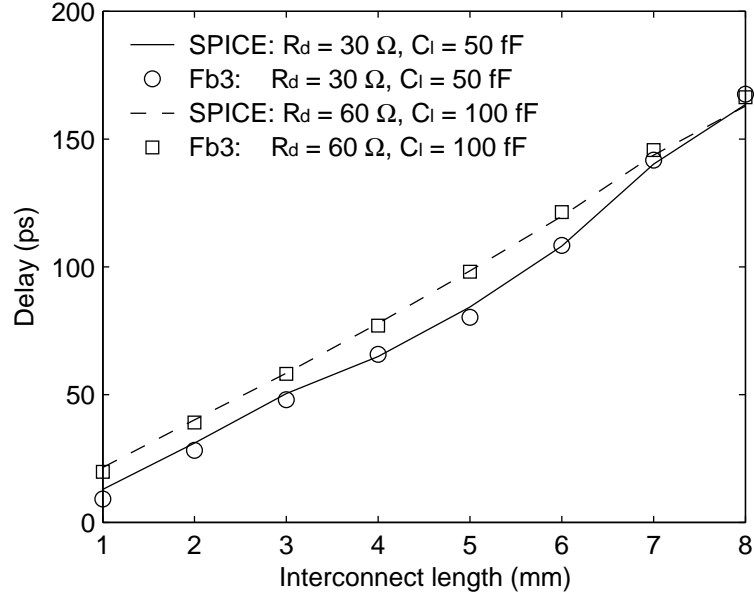


Figure 3.8: The 50% delay versus interconnect length. $w = 2 \mu\text{m}$, $T = 500 \text{ ps}$, and $\tau = 50 \text{ ps}$.

shown in Fig 3.9. Note that the accuracy of the model increases when τ/T increases from zero. From (3.5), when m is large, A_m no longer decreases monotonically with τ/T , as shown in Fig. 3.10, since the term $|\sin \phi_m|$ also depends on τ/T . This effect is demonstrated in Fig. 3.9. Note that when τ/T is greater than 0.2, the results from Fb3 and Fb5 start to deviate from the SPICE simulations and the best accuracy of Fb3 and Fb5 occurs when τ/T is between 0.1 and 0.2.

For highly LC dominant interconnects, the accuracy of the model also depends on the frequency of the signal. The 50% delay and overshoots for different signal frequencies are shown in Fig. 3.11. For a fixed τ/T , changing the frequency of the input signal corresponds to stretching the Fourier series in the frequency domain.

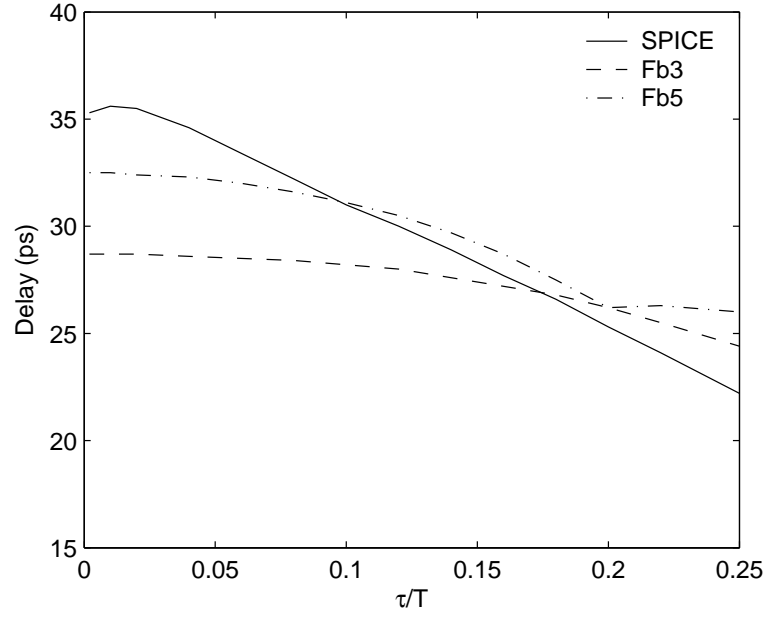


Figure 3.9: The 50% delay for different τ/T . $w = 2\mu\text{m}$, $l = 2\text{mm}$, $T = 500\text{ps}$, $V_{dd} = 1.5\text{volts}$, $R_d = 30\Omega$, and $C_l = 50\text{fF}$.

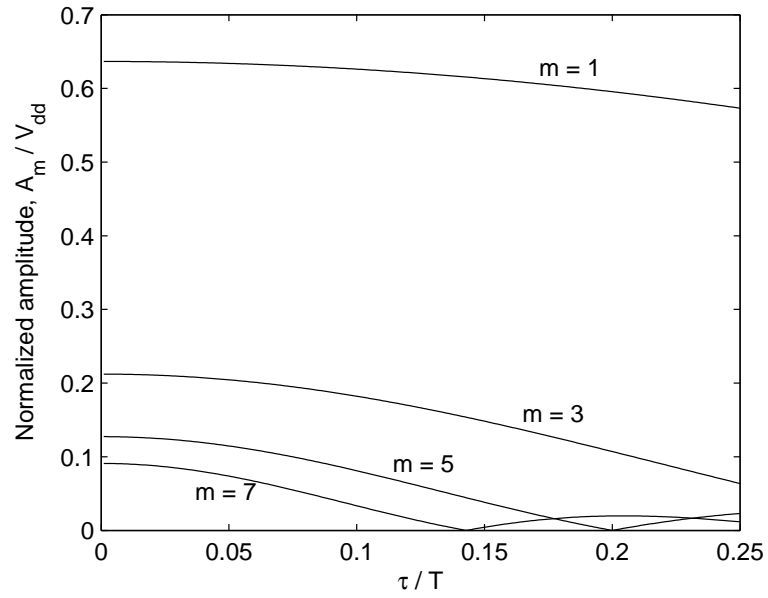
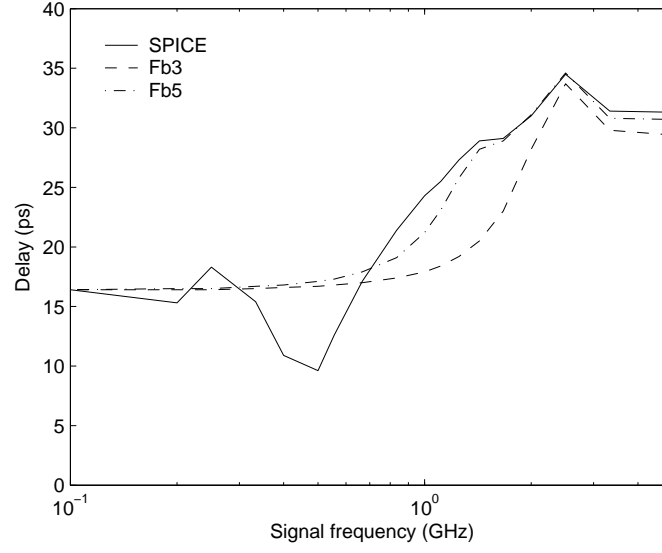
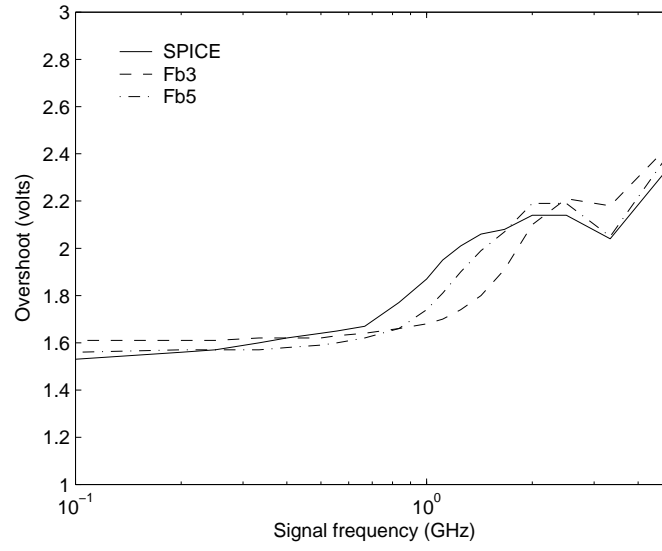


Figure 3.10: Normalized amplitude of harmonics with different τ/T .

When the signal frequency is much less than the resonance frequencies, all of the primary harmonics are located in the flat region of the transfer function curve. Those harmonics which are close to the resonance frequencies are sufficiently small that they can be safely neglected. The interconnect line behaves as a pure delay segment and the proposed model exhibits good accuracy. With the frequency increasing, the first two or three harmonics remain in the flat region in the amplitude transfer function. The other harmonics, however, approach those resonance frequencies and are amplified. Neglecting these harmonics will produce significant error. As shown in Fig. 3.11(a), the maximum error of the 50% delay for this example occurs at 500 MHz. With the signal frequency continuously increasing, the first several harmonics also approach the resonance frequencies and are amplified, therefore, the ratio between the harmonics which are included in the model and the harmonics which are neglected increases, making the proposed model more accurate. Since the resonance frequency is related to t_f , when the interconnect length increases, the resonance frequency decreases. With technology scaling, the global interconnect becomes longer and the clock frequency becomes higher. The proposed model is expected to become more accurate with higher speed circuits.



(a) The 50% delay



(b) Overshoot

Figure 3.11: The effects of signal frequency on the accuracy of the proposed model. $w = 2 \mu\text{m}$, $l = 2 \text{ mm}$, $\tau/T = 0.1$, $V_{dd} = 1.5 \text{ volts}$, $R_d = 30 \Omega$, and $C_l = 50 \text{ fF}$. (a) the 50% delay, (b) Overshoot.

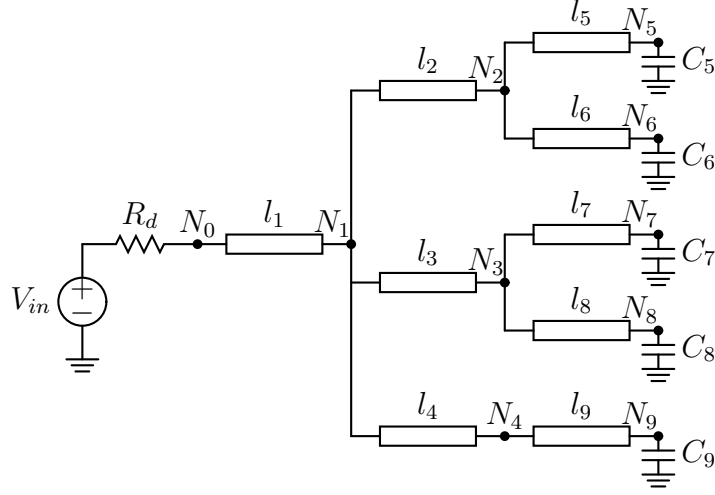
3.3 Distributed *RLC* Trees

Interconnect trees are widely used in clock distribution networks. In this section, the proposed Fourier series based model is extended to tree structures. Arbitrarily accurate results can be obtained by including a different number of harmonics. The computational complexity is linear with the size of the tree and the number of harmonics. In Subsection 3.3.1, the transfer function of a distributed *RLC* tree is developed. In Subsection 3.3.2, a tree example is analyzed with the Fourier series based model.

3.3.1 Transfer Function of Distributed *RLC* Trees

An example of a distributed *RLC* tree is shown in Fig. 3.12. In this example, a driver with an output resistance R_d is connected to the root of the tree N_0 . All of the output nodes ($N_5 \cdots N_9$) are called leaves and connected with load buffers which can be used to drive the *RLC* trees in the next level. The load buffers are modeled by capacitors. All of the branches in the tree are represented by distributed *RLC* lines. The tree can be balanced or unbalanced; however, unbalanced trees exhibit more complex characteristics than balanced trees [79].

The transfer function from N_0 to a certain node N_i is the product of the transfer function of all of the branches along the unique path from N_0 to N_i . For a transmission line of length l with load Z_L at the far end, the input impedance seen from the near

Figure 3.12: A distributed *RLC* tree.

end is

$$Z_{in} = Z_c \frac{Z_L + Z_c \tanh(\gamma l)}{Z_c + Z_L \tanh(\gamma l)}, \quad (3.25)$$

where γ and Z_c are defined in Subsection 3.2.1. For a node with multiple fanout, the load impedance seen at this node is the parallel combination of the input impedance of the downstream branches which are connected to this node. The computational complexity of computing the input impedance at the nodes in the tree is $O(n_{tr})$, where n_{tr} is the number of branches in the entire tree. The transfer function of a single branch can be obtained by replacing R_d by 0 and $C_l s$ by $1/Z_L$ in (3.1),

$$H(s) = \frac{1}{\cosh(\gamma l) + (Z_c/Z_L) \sinh(\gamma l)}. \quad (3.26)$$

The transfer function from the voltage source to a certain node N_i , therefore, is

$$H_i(s) = \frac{Z_{L,0}}{R_d + Z_{L,0}} \prod_k \frac{1}{\cosh(\gamma_k l_k) + (Z_{c,k}/Z_{L,k}) \sinh(\gamma_k l_k)}, \quad (3.27)$$

where $Z_{L,0}$ is the input impedance seen from N_0 , and k is the index covering each branch in the path from N_0 to N_i . From (3.27), the computational complexity of computing the transfer function at node N_i for one frequency is $O(n_i)$, where n_i is the number of branches along the path from N_0 to N_i . Upon obtaining $H_i(s)$, the Fourier series based model can be applied. The total computation complexity to determine the time domain response at node N_i is

$$\Theta(n_f, n_{tr}, n_i) = n_f \cdot O(n_{tr}) + n_f \cdot O(n_i), \quad (3.28)$$

where n_f is the number of harmonics included in the model. Note that the first item in (3.28) is related to calculating the input impedances of the branches, which are calculated only once for a specific tree. To determine the response at another node, the additional computational complexity is the second term in (3.28).

3.3.2 Examples

The tree structure shown in Fig. 3.12 is evaluated in this section. The branches in the tree can have different parasitic interconnect impedances. For simplicity, the

branches are assumed to have the same width of $6\ \mu\text{m}$. In high speed clock networks, ground wires are often placed at each side of the signal line as shields [98, 113], as shown in Fig 3.13.

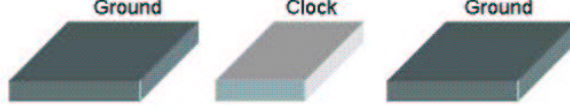


Figure 3.13: An example of a shielded clock wire structure.

Since these ground wires provide a nearby current return path, the effective inductance of the signal wire is greatly reduced. The width of the shield wire is assumed to be $10\ \mu\text{m}$ and the space between the shield and the clock line is $6\ \mu\text{m}$. The interconnect parameters of such a structure are $R = 3.9\ \text{m}\Omega/\mu\text{m}$, $L = 0.43\ \text{pH}/\mu\text{m}$, and $C = 0.36\ \text{fF}/\mu\text{m}$. An effective conductivity of $2.2\ \mu\Omega\cdot\text{cm}$ is used to determine the resistance and inductance. The normalized wire length and load capacitance shown in Fig. 3.12 are listed in Tables 3.3 and 3.4, where l_x and C_x are normalized reference length and capacitance, respectively.

Table 3.3: Interconnect lengths shown in Fig. 3.12 normalized to l_x .

Index	l_1	l_2	l_3	l_4	l_5	l_6	l_7	l_8	l_9
Length	0.5	1	2	0.5	1	0.5	2	1	1

A 2 GHz clock signal with $\tau = 50\ \text{ps}$ is applied at the input of the tree. The 50% delay at node $N5$ and $N7$ are listed in Table 3.5 for a range of circuit parameters. Results from the two pole model [107] and the equivalent Elmore delay model [79] are

Table 3.4: Load capacitances shown in Fig. 3.12 normalized to C_x .

Index	C_5	C_6	C_7	C_8	C_9
Capacitance	2	0.5	2	5	1

Table 3.5: The 50% delays at nodes $N5$ and $N7$ as shown in Fig. 3.12 with different circuit parameters.

l_x (mm)	R_d (Ω)	C_x (fF)	node	SPICE (ps)	[107] (ps)	[79] (ps)	Fb3 (ps)	Fb5 (ps)
0.2	10	20	$N5$	13.1	10.3	10.2	9.0	10.3
0.2	10	20	$N7$	11.6	14.0	13.8	10.6	11.2
0.2	10	500	$N5$	35.5	50.0	49.1	49.0	37.1
0.2	10	500	$N7$	59.9	60.8	58.6	60.1	60.0
0.2	30	20	$N5$	23.9	23.6	23.5	25.1	24.1
0.2	30	20	$N7$	23.5	25.6	25.4	25.7	24.5
0.2	30	100	$N5$	42.7	40.0	39.8	43.2	42.1
0.2	30	100	$N7$	39.4	43.1	42.5	44.1	41.2
1	10	20	$N5$	41.3	54.3	52.2	37.0	40.8
1	10	20	$N7$	75.8	79.7	75.1	77.5	75.3
1	10	100	$N5$	51.3	63.3	60.4	49.9	52.4
1	10	100	$N7$	89.6	93.0	86.9	89.6	89.2
1	20	20	$N5$	49.2	71.3	68.6	48.4	48.3
1	20	20	$N7$	86.4	96.1	89.4	90.6	88.1
1	20	100	$N5$	57.4	85.1	81.7	57.7	58.8
1	20	100	$N7$	104.1	114.0	105.4	103.4	102.9
Maximum Error					48.3%	42.3%	38.0%	21.4%
Average Error					18.0%	15.0%	7.6%	3.3%
Standard Deviation of Error					15.8%	14.7%	10.3%	5.0%

also listed for comparison. Since no closed form solution for a ramp input signal is available with these methods, the values listed in Table 3.5 are obtained through curve measurement. The branches in the tree are represented by lumped L-type models in [79] and [107].

The methods presented in [79] and [107] have similar accuracy and complexity, since both of these models are based on second order approximations. For a response which is low frequency dominant, such as the response at node N_7 , these second order methods can produce accurate delay estimates. For a response which exhibits more high frequency effects, such as the response at node N_5 , the error caused by these second order methods becomes large. These high frequency effects originate not only from the complexity of the tree structure, but also from the distributed properties of the interconnect, which cannot be modeled by lumped elements. As listed in Table 3.5, Fb3 and Fb5 produce higher accuracy than those second order approximations, particularly for node N_5 . The average error of Fb5 is only 3.3%. The time domain response obtained by different models at nodes N_5 and N_7 are compared with SPICE simulations in Fig. 3.14. As shown in Fig. 3.14, although the equivalent Elmore delay model provides satisfactory estimation of the 50% delay at node N_7 , other information characterizing the waveform is lost, such as the overshoots and transition times.

The 10% to 90% transition times from different models are compared with SPICE in Table 3.6. As listed in Table 3.6, all of the models exhibit worse accuracy for the transition time as compared with the 50% delay, since additional high frequency harmonics are required to characterize the signal transition times. Among the models, Fb5 is the most accurate with an average error of 5.8%.

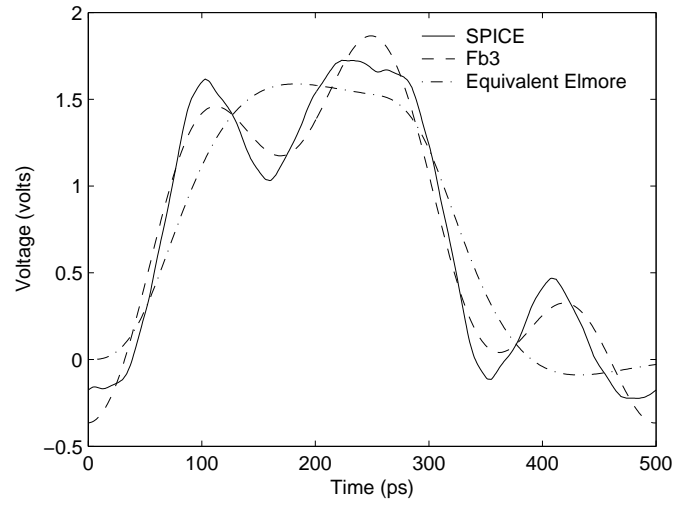
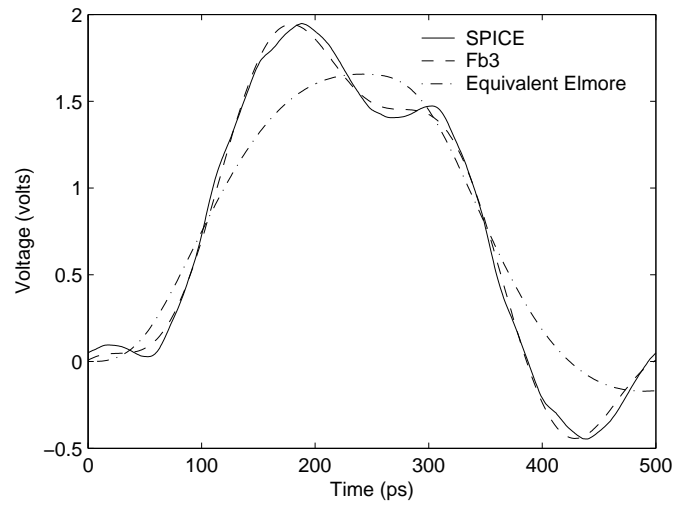
(a) Node $N5$ (b) Node $N7$

Figure 3.14: Time domain response at the leaves of the tree shown in Fig. 3.12. $l_x = 1$ mm, $\tau = 50$ ps, $T = 500$ ps, $V_{dd} = 1.5$ volts, $R_d = 10 \Omega$, and $C_x = 20$ fF. (a) Node $N5$, (b) Node $N7$.

Table 3.6: The transition times at nodes as $N5$ and $N7$ shown in Fig. 3.12 with different circuit parameters.

lx (mm)	R_d (Ω)	C_x (fF)	Node	SPICE (ps)	[107] (ps)	[79] (ps)	Fb3 (ps)	Fb5 (ps)
0.2	10	20	$N5$	40.0	38.9	39.0	60.4	46.3
0.2	10	20	$N7$	32.1	36.3	36.4	56.5	39.6
0.2	10	500	$N5$	130.5	127.2	130.4	159.3	132.8
0.2	10	500	$N7$	97.7	119.8	124.9	98.9	99.9
0.2	30	20	$N5$	69.3	68.3	68.9	75.8	68.9
0.2	30	20	$N7$	60.1	63.2	64.3	71.3	61.0
0.2	30	100	$N5$	132.9	118.5	119.4	126.8	124.2
0.2	30	100	$N7$	106.3	113.7	115.9	111.6	117.0
1	10	20	$N5$	39.1	81.4	81.3	53.8	39.7
1	10	20	$N7$	60.5	107.1	104.9	65.6	58.8
1	10	100	$N5$	45.2	97.1	97.6	52.1	52.4
1	10	100	$N7$	79.2	128.2	126.2	81.7	80.5
1	20	20	$N5$	157.9	149.5	156.9	176.0	156.1
1	20	20	$N7$	90.6	158.4	165.4	88.4	88.0
1	20	100	$N5$	188.3	188.5	198.6	195.3	193.3
1	20	100	$N7$	135.3	198.6	212.3	138.7	139.1
Maximum Error					114.8%	115.9%	76.0%	23.4%
Average Error					34.6%	35.8%	17.0%	5.8%
Standard Deviation of Error					40.1%	40.6%	21.0%	6.8%

The accuracy of the Fourier series based model can be enhanced to capture the fine details of the waveform by including additional harmonics. As shown in (3.28), the complexity of higher order Fourier series based models is linear with the number of harmonics. Furthermore, there are no stability and numerical problems such as suffered by AWE [71]. In Fig. 3.15, the Fourier series based model with a different number of harmonics is compared with SPICE simulations. τ is reduced to 5 ps to emphasize the high frequency effects. As shown in Fig. 3.15, results from the tenth

order Fourier series based model are sufficiently close to the SPICE simulations. These experiments are performed on a SunBlade1000 workstation. In SPICE simulations, each branch is represented by 100 lumped elements. The time required by SPICE to simulate one clock period (500 ps) is 9.6 seconds. The run time of the tenth order model (implemented by Matlab) is about 6.6 ms.

Clock distribution networks are typically hierarchically structured. A high level interconnect tree distributes the clock signal to several buffers which drive the lower level interconnect trees. The buffers in the clock distribution networks are nonlinear devices, and the clock signal is reshaped by these buffers. The proposed model, therefore, is limited to a single tree structure. By combining the signal waveform at the output node of a tree and the buffer model, the effective input signal for the tree of the next level can be obtained. The proposed model, therefore, can be applied to each individual tree in a specific clock network and the entire network can be analyzed.

3.4 Multiple Coupled Interconnect Lines

The solution for a single distributed RLC line can also be extended to multiple coupled transmission lines. In Subsection 3.4.1, the modal analysis based decoupling method is reviewed. A general solution for multiple transmission lines is presented in Subsection 3.4.2. The model is verified by SPICE in Subsection 3.4.3.

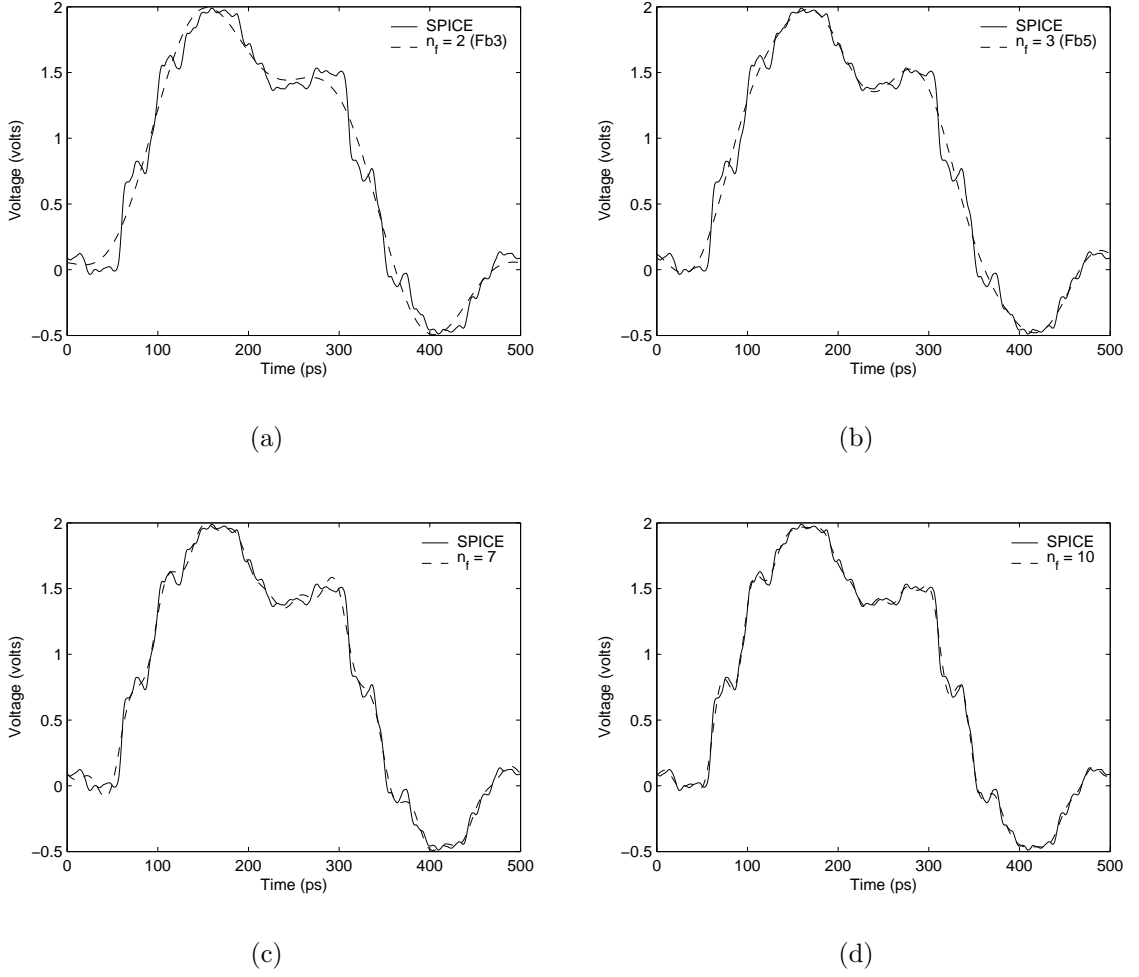


Figure 3.15: Time domain response at node $N7$ in Fig. 3.12 evaluated by the Fourier series based model with different n_f as compared with SPICE simulations. $l_x = 1$ mm, $\tau = 5$ ps, $T = 500$ ps, $V_{dd} = 1.5$ volts, $R_d = 10 \Omega$, and $C_x = 20$ fF.

3.4.1 Decoupling Multiconductor Systems

For multiple transmission lines, the interconnect parameters per unit length are the resistance matrix \mathbf{R} , inductance matrix \mathbf{L} , and capacitance matrix \mathbf{C} . All of these matrices are symmetric with the dimension of $N \times N$, where N is the number of lines.

From the Telegrapher equations of N coupled transmission lines, the voltage vector \mathbf{V} and current vector \mathbf{I} have the following relationship in the frequency domain,

$$\frac{\partial}{\partial x} \begin{pmatrix} \mathbf{V} \\ \mathbf{I} \end{pmatrix} = - \begin{pmatrix} \mathbf{0} & \mathbf{Z} \\ \mathbf{Y} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{V} \\ \mathbf{I} \end{pmatrix}, \quad (3.29)$$

where $\mathbf{Z} = \mathbf{R} + s\mathbf{L}$ and $\mathbf{Y} = s\mathbf{C}$.

Decoupling (3.29) can be achieved by applying a modal analysis [66], [60]. The matrix \mathbf{ZY} for a practical system is always diagonalizable [60],

$$\mathbf{ZY} = \mathbf{MQM}^{-1}, \quad (3.30)$$

where \mathbf{Q} is a diagonal matrix with eigenvalues of \mathbf{ZY} as the diagonal elements, and matrix \mathbf{M} has the corresponding eigenvectors of \mathbf{ZY} as the columns. Performing a linear transformation of \mathbf{V} and \mathbf{I} ,

$$\mathbf{V} = \mathbf{M}\hat{\mathbf{V}}, \quad (3.31)$$

$$\mathbf{I} = (\mathbf{M}^T)^{-1}\hat{\mathbf{I}}, \quad (3.32)$$

and substituting (3.31) and (3.32) into (3.29), (3.29) becomes

$$\frac{\partial}{\partial x} \begin{pmatrix} \hat{\mathbf{V}} \\ \hat{\mathbf{I}} \end{pmatrix} = - \begin{pmatrix} \mathbf{0} & \hat{\mathbf{Z}} \\ \hat{\mathbf{Y}} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \hat{\mathbf{V}} \\ \hat{\mathbf{I}} \end{pmatrix}, \quad (3.33)$$

where $\hat{\mathbf{Z}} = \mathbf{M}^{-1} \mathbf{Z} (\mathbf{M}^T)^{-1}$ and $\hat{\mathbf{Y}} = \mathbf{M}^T \mathbf{Y} \mathbf{M}$. Since \mathbf{Z} and \mathbf{Y} are symmetric, $\hat{\mathbf{Z}}$ and $\hat{\mathbf{Y}}$ are both diagonal [66]. The N coupled interconnect lines are, therefore, decoupled into N independent lines. The characteristic impedance matrix $\hat{\mathbf{Z}}_c$ and the propagation coefficient matrix $\hat{\gamma}$ of the decoupled system are

$$\hat{\mathbf{Z}}_c = \sqrt{\hat{\mathbf{Z}} \hat{\mathbf{Y}}^{-1}} = \text{diag}(\hat{Z}_{c,1}, \hat{Z}_{c,2}, \dots, \hat{Z}_{c,N}), \quad (3.34)$$

$$\hat{\gamma} = \sqrt{\hat{\mathbf{Z}} \hat{\mathbf{Y}}} = \sqrt{\mathbf{Q}} = \text{diag}(\hat{\gamma}_1, \hat{\gamma}_2, \dots, \hat{\gamma}_N). \quad (3.35)$$

This decoupling method has been extended to drivers and loads in [64] and [61] for two and more interconnects. These extensions, however, are only suitable for identical lines with identical drivers and loads. Furthermore, the inductance matrix in [61] is obtained as $\mathbf{L} = \mathbf{C}^{-1}/v^2$, where v is the speed of light in a dielectric. This expression is only valid for a homogeneous dielectric environment with an ideal ground for the current return path. Due to these constraints, the practical generality of these models is greatly limited.

3.4.2 Far End Response

Applying the $ABCD$ parameter concept to (3.33), the voltage and current vectors at the boundary have the following relationship,

$$\begin{pmatrix} \hat{\mathbf{V}}_d \\ \hat{\mathbf{I}}_d \end{pmatrix} = \begin{pmatrix} \hat{\mathbf{A}}_p & \hat{\mathbf{B}}_p \\ \hat{\mathbf{C}}_p & \hat{\mathbf{D}}_p \end{pmatrix} \begin{pmatrix} \hat{\mathbf{V}}_r \\ \hat{\mathbf{I}}_r \end{pmatrix}, \quad (3.36)$$

where the subscript d and r represent the driver side (or near end) and receiver side (or far end), respectively. The $ABCD$ [50] matrices of the decoupled transmission lines are

$$\hat{\mathbf{A}}_p = \text{diag} (\cosh(\hat{\gamma}_k l), k = 1, 2, \dots, N), \quad (3.37)$$

$$\hat{\mathbf{B}}_p = \text{diag} (\hat{Z}_{c,k} \sinh(\hat{\gamma}_k l), k = 1, 2, \dots, N), \quad (3.38)$$

$$\hat{\mathbf{C}}_p = \text{diag} \left(\frac{\sinh(\hat{\gamma}_k l)}{\hat{Z}_{c,k}}, k = 1, 2, \dots, N, \right), \quad (3.39)$$

$$\hat{\mathbf{D}}_p = \text{diag} (\cosh(\hat{\gamma}_k l), k = 1, 2, \dots, N). \quad (3.40)$$

The boundary conditions of the N coupled interconnects are

$$\mathbf{V}_d = \mathbf{V}_{in} - \mathbf{R}_d \mathbf{I}_d, \quad (3.41)$$

$$\mathbf{I}_r = s\mathbf{C}_l \mathbf{V}_r, \quad (3.42)$$

where \mathbf{R}_d and \mathbf{C}_l are the driver resistance matrix and load capacitance matrix, respectively. Both of these matrices are diagonal. Combining (3.36) with (3.31), (3.32), (3.41), and (3.42), the voltage vector at the far end of the interconnects is

$$\mathbf{V}_r = \mathbf{H}\mathbf{V}_{in}, \quad (3.43)$$

$$\mathbf{H} = (\mathbf{R}_d\mathbf{C}_p + s\mathbf{R}_d\mathbf{D}_p\mathbf{C}_l + \mathbf{A}_p + s\mathbf{B}_p\mathbf{C}_l)^{-1}, \quad (3.44)$$

where \mathbf{H} is the transfer function matrix. The $ABCD$ matrices of the coupled transmission lines are

$$\mathbf{A}_p = \mathbf{M}\hat{\mathbf{A}}_p\mathbf{M}^{-1}, \quad (3.45)$$

$$\mathbf{B}_p = \mathbf{M}\hat{\mathbf{B}}_p\mathbf{M}^T, \quad (3.46)$$

$$\mathbf{C}_p = (\mathbf{M}^T)^{-1}\hat{\mathbf{C}}_p\mathbf{M}^{-1}, \quad (3.47)$$

$$\mathbf{D}_p = (\mathbf{M}^T)^{-1}\hat{\mathbf{A}}_p\mathbf{M}^T. \quad (3.48)$$

In general, \mathbf{M} is a matrix function of s and cannot be expressed in closed form [60]. Furthermore, the matrix inverse operation in (3.44) does not permit an analytic expression (or an analytic low order approximation) of the transfer function to be obtained. Conventional inverse Laplace transform based methods [105]–[109], which assume a step or ramp input, can no longer be used. The proposed model, which assumes a periodic input signal, remains valid, since the solution of (3.44) is only

required at certain discrete frequencies (*e.g.*, the harmonic frequencies of the input signal), and can be solved numerically at each frequency. When N is less than five, closed form solutions exist [61] to calculate \mathbf{M} and \mathbf{Q} . For a larger N , numerical methods have to be used, and the computing complexity increases. When $s = 0$, \mathbf{H} becomes an identity matrix. Since no approximation is made in this derivation, (3.44) is the exact transfer function of a coupled multiconductor system. For the structure shown in Fig. 3.16, the accuracy of (3.44) is illustrated in Fig. 3.17. The interconnect parameters are

$$\mathbf{R} = \text{diag} (11.6, 19.9, 11.0, 19.9, 11.6) \text{ m}\Omega/\mu\text{m}, \quad (3.49)$$

$$\mathbf{L} = \begin{bmatrix} 0.52 & 0.32 & 0.21 & 0.14 & 0.08 \\ 0.32 & 0.70 & 0.38 & 0.25 & 0.14 \\ 0.21 & 0.38 & 0.65 & 0.38 & 0.21 \\ 0.14 & 0.25 & 0.38 & 0.70 & 0.32 \\ 0.08 & 0.14 & 0.21 & 0.32 & 0.52 \end{bmatrix} \text{ pH}/\mu\text{m}, \quad (3.50)$$

$$\mathbf{C} = \begin{bmatrix} 232 & -57 & 0 & 0 & 0 \\ -57 & 187 & -56 & 0 & 0 \\ 0 & -56 & 231 & -56 & 0 \\ 0 & 0 & -56 & 187 & -57 \\ 0 & 0 & 0 & -57 & 232 \end{bmatrix} \text{ aF}/\mu\text{m}. \quad (3.51)$$

For simplicity, capacitive coupling is assumed to exist only between adjacent lines. The other circuit parameters are $l = 2 \text{ mm}$, $\mathbf{R}_d = \text{diag}(50, 30, 40, 50, 30) \Omega$, and $\mathbf{C}_l = \text{diag}(50, 100, 80, 80, 50) \text{ fF}$. In Fig. 3.17, $H_{i,k}$ represents the amplitude transfer function from the input of line k to the far end of line i . Upon obtaining the transfer coefficient at each harmonic frequency, the output signal can be determined in the same manner as in (3.7). The time-of-flight t_f of a multiconductor system is the minimum t_f of all of the wave modes. In this multiconductor model, no constraints are made on the interconnect parameters, making the solution of general use. With a periodic ramp signal applied to line 1, the far end response from Fb3 and Fb5 are compared with SPICE in Fig. 3.18. The model is implemented by Matlab and the run time for this example is about 17 ms on a SunBlade1000 workstation. The time required by SPICE to simulate one clock period (500 ps) is 7.7 seconds.

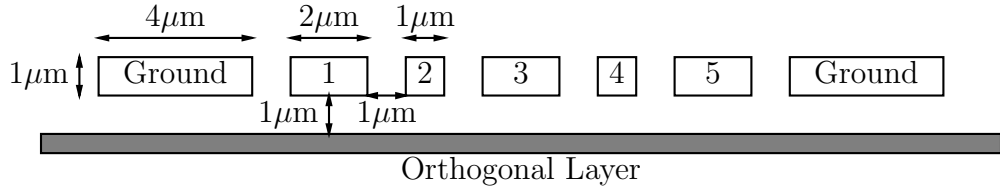


Figure 3.16: Geometric characteristics of five parallel interconnect lines

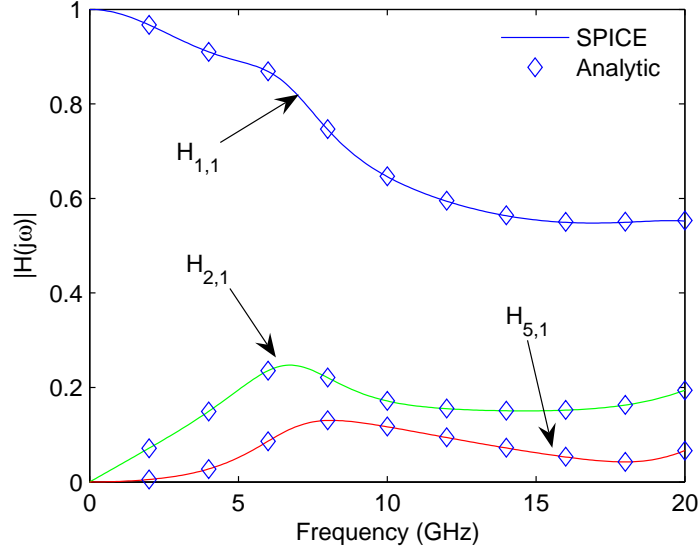
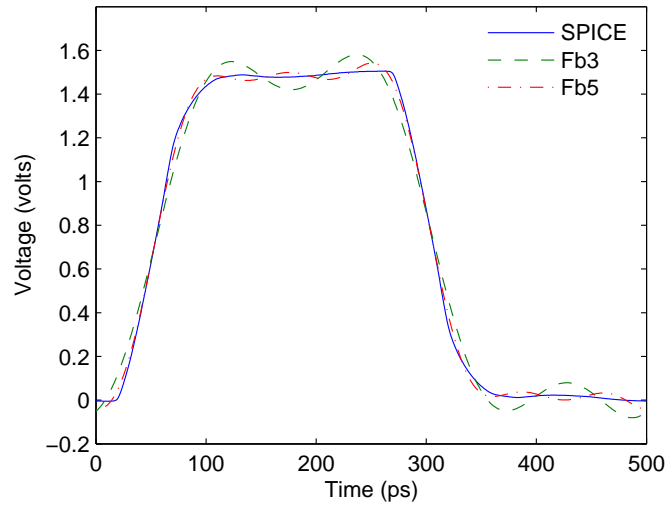


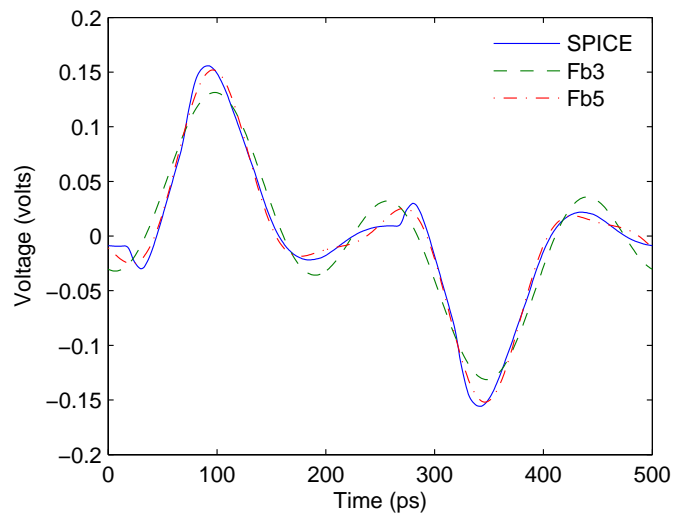
Figure 3.17: Amplitude transfer functions of a five line system.

3.4.3 Model Verification and Discussion

For simplicity, only one aggressor is considered in the following examples. Multi-aggressor circuits can be solved by applying superposition. The maximum crosstalk noise determined by the Fourier series-based model is compared with SPICE in Table 3.7. In the experiments, line 1 is the aggressor, and all of the other lines are quiet victims (represented as V2 to V5 in Table 3.7). As shown in Table 3.7, Fb3 provides limited accuracy in multiconductor systems. Note that the error reaches 40.7% as compared with SPICE. This result is not surprising since, as shown in Fig. 3.17, the magnitude of the transfer function of the victims is small at low frequencies (zero at DC). Thus, the high frequency components are comparable or greater than the low frequency components at the output, and are therefore not negligible. By including



(a) Line 1



(b) Line 2

Figure 3.18: Comparison of the far end response from Fb3 and Fb5 with SPICE in a five line coupled system. The input signal parameters are $T = 500$ ps, $\tau = 50$ ps, and $V_{dd} = 1.5$ volts.

Table 3.7: Comparison of the maximum crosstalk noise of Fb3 and Fb5 with SPICE simulations. The input signal parameters are $T = 500$ ps, $\tau = 50$ ps, and $V_{dd} = 1.5$ volts.

l (mm)	Victim	SPICE (mV)	Fb3		Fb5	
			(mV)	% Error	(mV)	% Error
2	V2	155.9	131.4	15.7	151.9	2.6
	V3	67.6	48.9	27.7	69.8	3.3
	V4	54.6	39.3	28.0	57.5	5.3
	V5	40.6	26.8	34.0	40.9	0.7
4	V2	190.5	197.0	3.4	195.2	2.5
	V3	68.8	73.4	6.7	62.0	9.9
	V4	60.3	54.4	9.8	54.0	10.4
	V5	48.2	38.4	20.3	34.7	28.0
6	V2	188.8	201.8	6.9	192.4	1.9
	V3	110.6	79.6	28.0	99.0	10.5
	V4	95.0	66.7	29.8	87.4	8.0
	V5	74.0	43.9	40.7	60.9	17.7
Maximum Error			—	40.7%	—	28.0%
Average Error			—	20.9%	—	8.4%

one more frequency component, Fb5 is significantly more accurate with an average error of 8.4%. Also note that, for the nearby victim 2 which suffers larger crosstalk noise, the analytical model has higher accuracy than for the far victims.

In integrated circuits, since logic gates have a low pass filtering property [114], the sharp spikes in the time domain waveforms normally cannot cause a circuit to fail. The Fourier series based model, which ignores these high frequency effects, is therefore an effective method to analyze crosstalk noise.

3.5 Conclusions

By exploiting a Fourier series representation of a typical on-chip signal, an analytic time-domain solution for an RLC interconnect is shown to be an effective modeling strategy which can be used in early circuit level design stages to estimate the time characteristics of periodic signals. Expressions for the 50% delay and the overshoots/undershoots are also provided and are within 11% of SPICE over a wide range of circuit parameters. The single line model is applied to tree structures and a tree model with linear computational complexity is obtained, which is shown to be an effective analysis tool for clock distribution networks. Combined with the modal analysis based decoupling method, the proposed model is also extended to coupled interconnect systems to analyze crosstalk noise. For three harmonic frequencies (Fb5), the average error is 8.4%.

Chapter 4

Transient Response of a Distributed RLC Interconnect Based on Direct Pole Extraction

4.1 Introduction

On-chip global interconnects exhibit significant transmission line behavior. An efficient solution to analyze transmission lines is therefore highly desirable. Sakurai presented in [58, 115] an accurate closed-form solution for distributed RC interconnect based on a single pole approximation. By truncating the transfer function, multi-pole models have been proposed in the last decade to capture the effect of inductance [107, 109]. In [105], the solution for an open-ended interconnect with a step input signal is rigorously developed. This solution however is highly complicated and not suitable for an exploratory design process. In [116], a traveling wave analysis (TWA) model has

been presented, where the key points of the waveform are determined with a three-pole model and linear or RC approximations are used to connect those key points to construct the waveform. This method is improved in [117], where the key points and slopes are more accurately determined with the model described in [105], and straight lines are used to construct the signal waveforms in different time regions. In both of these papers, the output response is divided into a number of time regions where the waveform expressions for each of the regions are different, making the models less compact. Furthermore, none of these aforementioned papers consider frequency dependent effects. With higher on-chip frequencies, frequency dependent effects in wider interconnect can no longer be ignored. In Chapter 3, a Fourier analysis based interconnect model is proposed, where the far end response is approximated by the first several harmonics. Frequency dependent effects can be included in this model; however, the model is only suitable for periodic signals.

In this chapter, a novel method for computing the far end response of a transmission line is proposed. The proposed model is based on a direct pole extraction of the exact transfer function of a transmission line, rather than approximating the poles by truncating the transfer function [107, 109] or matching moments [71]. Closed-form waveform expressions are developed, permitting flexible tradeoffs between accuracy and efficiency. The rest of the chapter is organized as follows. In Section 4.2, the exact poles of two special case interconnect systems are determined. Based on these poles,

the step and ramp responses are developed. In Section 4.3, an interconnect system with general circuit parameters is solved. The Newton-Raphson method is used to determine the exact poles of the system. Frequency dependent effects are successfully included in Section 4.4. Finally, some conclusions are offered in Section 4.5.

4.2 Special Cases of a Single Interconnect System

For a distributed RLC interconnect driven by a voltage source with a driver resistance R_d and loaded with a lumped capacitance C_L , as shown in Fig. 4.1, the transfer function is [107, 118]

$$H(s) = \frac{1}{(1 + R_d C_L s) \cosh(\theta) + (R_d / Z_c + Z_c C_L s) \sinh(\theta)}, \quad (4.1)$$

where $\theta = \sqrt{(R + Ls)Cs}$ and $Z_c = \sqrt{(R + Ls)/Cs} = \theta/Cs$. R , L , and C are the resistance, inductance, and capacitance of the interconnect. The poles of (4.1) are difficult to solve directly except for two special cases: an RC interconnect and an RLC interconnect with a zero driver resistance. In Section 4.2.1, the poles of an RC interconnect system are solved. In Section 4.2.2, the poles of an RLC interconnect with a zero driver resistance are solved. Step and ramp responses are developed in Section 4.2.3.

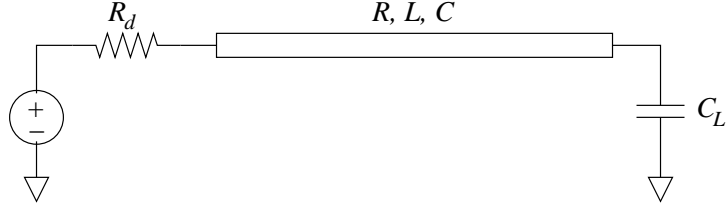


Figure 4.1: Distributed interconnect with a lumped capacitive load and driver resistance.

4.2.1 RC interconnect

For RC interconnect, $L = 0$. The transfer function (4.1) can be rewritten as

$$H(s) = \frac{1}{(1 + A\theta^2) \cosh(\theta) + B\theta \sinh(\theta)}, \quad (4.2)$$

where $A = R_T C_T$, $B = R_T + C_T$, $R_T = R_d/R$, $C_T = C_L/C$, and $\theta = \sqrt{RCs}$. Let $F(s) = 1/H(s)$. The poles of $H(s)$ are zeros of $F(s)$ and satisfy $F(s) = 0$. Observe that θ needs to be an imaginary number to make $F(s)$ zero. Assume $\theta = jx$, where x is a real number. Expression $F(s) = 0$ can be transformed to

$$(1 - Ax^2) \cos x - Bx \sin x = 0, \quad (4.3)$$

or

$$\tan x = \frac{1 - Ax^2}{Bx}. \quad (4.4)$$

The roots of (4.4) are the crossing points of the functions of $y = \tan x$ and $y = (1 - Ax^2)/(Bx)$, as shown in Fig. 4.2.

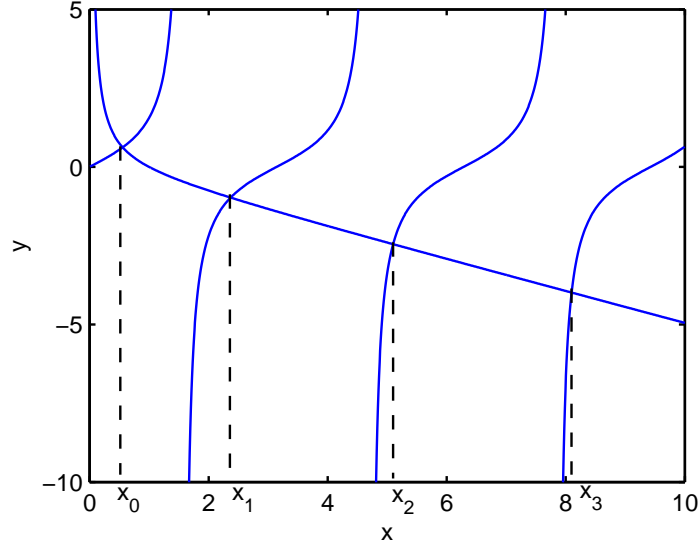


Figure 4.2: Graphic view of the roots of (4.4), $R_T = C_T = 1$.

Applying Taylor series expansions of $\cos x \approx 1 - x^2/2 + x^4/24$ and $\sin x \approx x - x^3/6$ to (4.3), and ignoring those terms with an order higher than x^4 results in the following expression,

$$\left(\frac{1}{2}A + \frac{1}{6}B + \frac{1}{24}\right)x^4 - \left(\frac{1}{2} + A + B\right)x^2 + 1 = 0. \quad (4.5)$$

Solving (4.5) for the smaller x^2 yields

$$x_0^2 = \frac{\frac{1}{2} + A + B - \sqrt{(A + B)^2 - A + \frac{1}{3}B + \frac{1}{12}}}{A + \frac{1}{3}B + \frac{1}{12}}. \quad (4.6)$$

When $R_T = C_T = 0$, the exact value of x_0^2 is $\pi^2/4$. In order to capture this trend,

(4.7) is revised to

$$x_0^2 = \frac{\frac{1}{2} + A + B - \sqrt{(A + B)^2 - A + \frac{1}{3}B + \frac{1}{11.54}}}{A + \frac{1}{3}B + \frac{1}{12}}. \quad (4.7)$$

Note that if the term x^4 in (4.5) is ignored, the solution simplifies to

$$x_0^2 = \frac{1}{0.5 + A + B} = \frac{1}{0.5 + R_T + C_T + R_T C_T}, \quad (4.8)$$

which is similar to the solution provided in [58].

Since the Taylor series approximations used in (4.5) are expanded around zero, the solution shown in (4.6) corresponds to the root x_0 which is most close to zero, as show in Fig. 4.2. In order to obtain other high order solutions, Taylor series approximations expanded at $n\pi$ ($n = 1, 2, \dots$) are used. Since the negative roots of (4.3) have the same absolute value as the positive roots, only positive roots are considered in this chapter. Let $\Delta x = x - n\pi$, $\cos x \approx (-1)^n[1 - (\Delta x)^2/2]$, and $\sin x \approx (-1)^n \Delta x$. Substituting these Taylor series approximations into (4.3) and ignoring those terms with an order higher than $(\Delta x)^2$ results in

$$\left(A + B - \frac{1}{2}E\right) (\Delta x)^2 + (2A + B)n\pi \Delta x + E = 0, \quad (4.9)$$

where $E = An^2\pi^2 - 1$. Solving (4.9) for x_n results in

$$x_n = \frac{-(2A + B)n\pi + \sqrt{(n\pi B)^2 + 4(A + B) + 2E^2}}{2(A + B) - E} + n\pi. \quad (4.10)$$

The accuracy of (4.6) and (4.10) is illustrated in Fig. 4.3 for different values of R_T and C_T . As shown in Fig. 4.3, the error of the higher order solutions is larger for greater values of R_T and C_T . In these cases, the effect of the higher order solutions however is negligible.

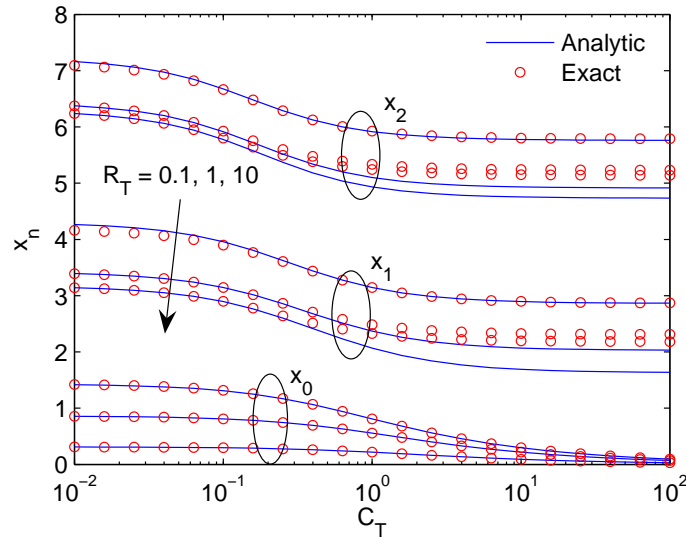


Figure 4.3: Analytic solution of (4.3) as compared with the exact solution for different values of R_T and C_T .

After solving x_n , the poles of an RC interconnect system can be obtained,

$$p_n = \frac{\theta^2}{RC} = \frac{-x_n^2}{RC}, \quad n = 0, 1, 2, \dots \quad (4.11)$$

The residue of the corresponding poles is

$$k_n = \lim_{s \rightarrow p_n} \frac{s - p_n}{F(s)} = \frac{1}{F'(p_n)} = \frac{2x_n/(RC)}{(1 + B - Ax_n^2) \sin x_n + (2A + B)x_n \cos x_n}, \quad (4.12)$$

where $F'(p_n)$ is the derivative of $F(s)$ at p_n .

4.2.2 *RLC* interconnect with a zero R_d

If R_d is zero, (4.1) simplifies to

$$H(s) = \frac{1}{\cosh(\theta) + C_T \theta \sinh(\theta)}. \quad (4.13)$$

Note that θ also needs to be an imaginary number to make $F(s)$ zero. Similar to the approach for the *RC* case, assume $\theta = jx$, where x is a real number. The poles of the transfer function should satisfy

$$\cos x - C_T x \sin x = 0, \quad (4.14)$$

or

$$x = \frac{\cot x}{C_T}. \quad (4.15)$$

The roots of (4.15) are the crossing point of the curves of functions $y = x$ and $y = \cot x/C_T$, as shown in Fig. 4.4.

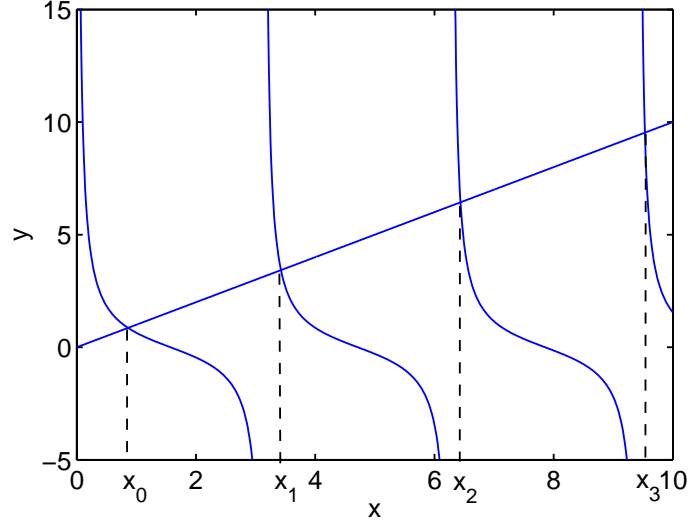


Figure 4.4: Graphic view of the roots of (4.15), $C_T = 1$.

By applying Taylor series approximations, x can be solved as

$$x_n = \begin{cases} \sqrt{\frac{\frac{1}{2} + C_T - \sqrt{C_T^2 + \frac{1}{3}C_T + \frac{1}{12}}}{\frac{1}{3}C_T + \frac{1}{12}}}, & n = 0 \\ \frac{(1 + C_T)n\pi + \sqrt{(C_T n\pi)^2 + 2 + 4C_T}}{1 + 2C_T}, & n \geq 1. \end{cases} \quad (4.16)$$

Note that when C_T approaches zero, (4.14) becomes $\cos x = 0$, and the solution x_n approaches $(n + 1/2)\pi$, where $n = 0, 1, 2, \dots$. In order to capture this trend, (4.16) is revised as

$$x_n = \begin{cases} \sqrt{\frac{\frac{1}{2} + C_T - \sqrt{C_T^2 + \frac{1}{3}C_T + \frac{1}{11.54}}}{\frac{1}{3}C_T + \frac{1}{12}}}, & n = 0 \\ \frac{(1 + C_T)n\pi + \sqrt{(C_T n\pi)^2 + \frac{\pi^2}{4} + 4C_T}}{1 + 2C_T}, & n \geq 1. \end{cases} \quad (4.17)$$

The accuracy of (4.17) is illustrated in Fig. 4.5 for different values of C_T . As shown in Fig. 4.5, when C_T increases from zero to infinity, x_n decreases from $(n + 1/2)\pi$ to $n\pi$.

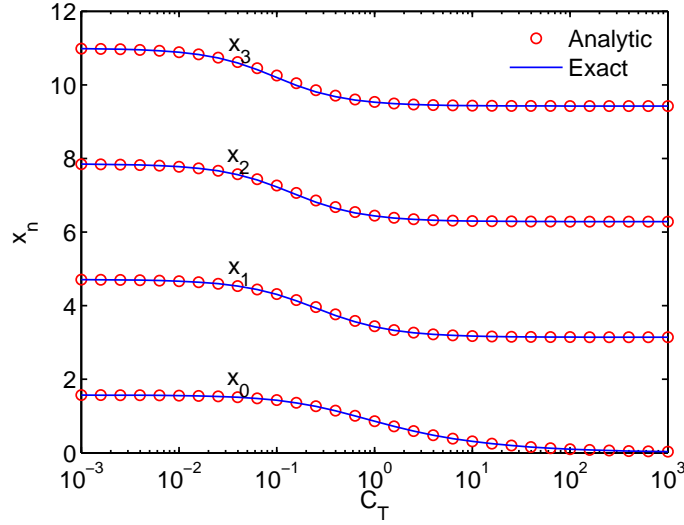


Figure 4.5: Analytic solution of (4.14) as compared with the exact solution for different values of C_T .

The poles of the transfer function can be obtained from the following expression,

$$LCs^2 + RCs = \theta^2 = -x_n^2, n = 0, 1, 2, \dots \quad (4.18)$$

Each x_n corresponds to a pair of poles,

$$p_{n,\pm} = \frac{-RC \pm \sqrt{R^2C^2 - 4LCx_n^2}}{2LC}. \quad (4.19)$$

The residue of the corresponding poles $k_{n,\pm}$ can be solved as

$$k_{n,\pm} = \lim_{s \rightarrow p_{n,\pm}} \frac{s - p_{n,\pm}}{F(s)} = \frac{1}{F'(p_{n,\pm})} = \frac{\pm 2x_n}{D[(1 + C_T) \sin x_n + C_T x_n \cos x_n]}, \quad (4.20)$$

where $D = \sqrt{R^2 C^2 - 4LCx_n^2}$.

4.2.3 Step and Ramp Response

From the poles and corresponding residues, the transfer function can be represented as

$$H(s) = \sum_i \frac{k_i}{s - p_i}, \quad (4.21)$$

where i is the index covering all of the poles. Consider a wire structure example as shown in Fig. 4.6. The interconnect parameters per unit length are $R_{int} = 12.24 \text{ m}\Omega/\mu\text{m}$, $L_{int} = 0.74 \text{ pH}/\mu\text{m}$, and $C_{int} = 0.266 \text{ fF}/\mu\text{m}$, which are extracted from FastHenry [44] and FastCap [32] with a signal frequency of 2 GHz.

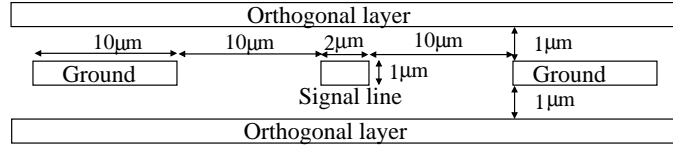


Figure 4.6: Wire geometry of an example circuit, where the signal wire is shielded by two ground lines.

The amplitude of the transfer function obtained from (4.21) is compared with the exact transfer function for the RC case in 4.7(a) and RLC case with a zero R_d

in 4.7(b), respectively. In Fig. 4.7(a), m is the number of poles considered in the model. In Fig. 4.7(b), m is the number of pole pairs, since the poles in this case are in pairs. As shown in the figure, the analytic transfer function converges to the exact transfer function with increasing m . As compared with the RC case, more poles are required for the RLC case to obtain an accurate result.

From (4.21), the normalized step response $V_s(t)/V_{dd}$ and ramp response $V_r(t)/V_{dd}$ are, respectively,

$$\frac{V_s(t)}{V_{dd}} = u(t) \left[1 + \sum_i \left(\frac{k_i}{p_i} e^{p_i t} \right) \right], \quad (4.22)$$

$$\frac{V_r(t)}{V_{dd}} = V_1(t) - V_1(t - t_r), \quad (4.23)$$

$$V_1(t) = \frac{u(t)}{t_r} \left[t + m_1 + \sum_i \left(\frac{k_i}{p_i^2} e^{p_i t} \right) \right], \quad (4.24)$$

where $u(t)$ is the step function. The following moment information is used,

$$-\sum_i \frac{k_i}{p_i} = m_0 = 1, \quad (4.25)$$

$$-\sum_i \frac{k_i}{p_i^2} = m_1. \quad (4.26)$$

For an RC interconnect, the first moment $m_1 = -R_d(C + C_L) - R(0.5C + C_L)$, and for an RLC interconnect with a zero driver resistance, the first moment $m_1 = -R(0.5C + C_L)$.

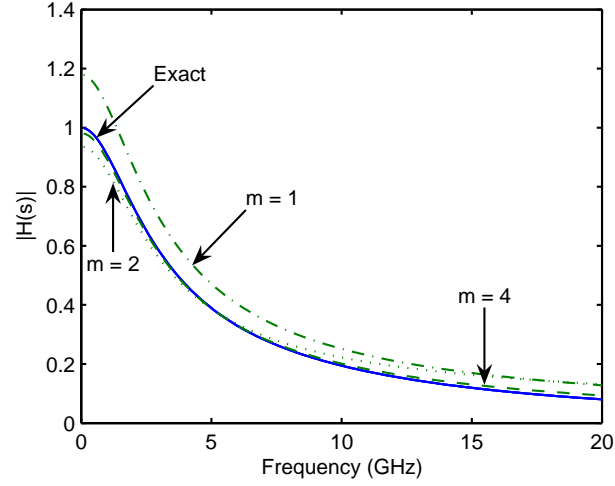
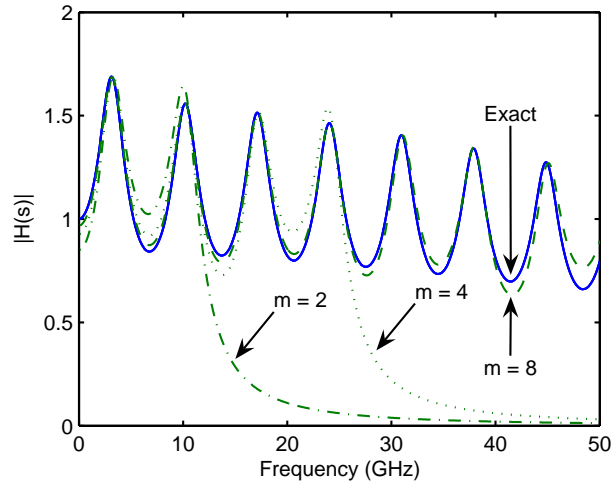
(a) *RC* interconnect(b) *RLC* interconnect with $R_d = 0$

Figure 4.7: Comparison between the analytic expression (4.21) and the exact transfer function. The wire length is 5 mm and the load capacitance is $C_L = 50$ fF. a) *RC* interconnect case, $R_d = 30 \Omega$. b) *RLC* interconnect with a zero R_d .

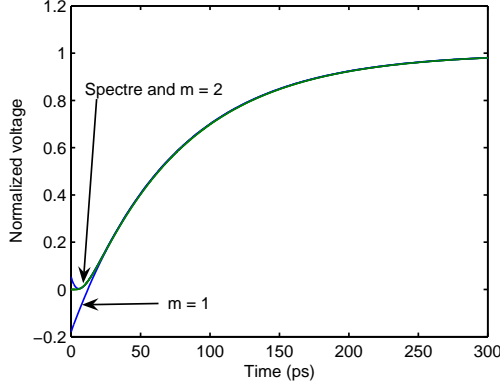
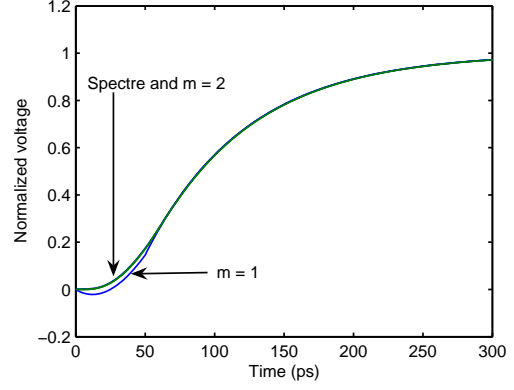
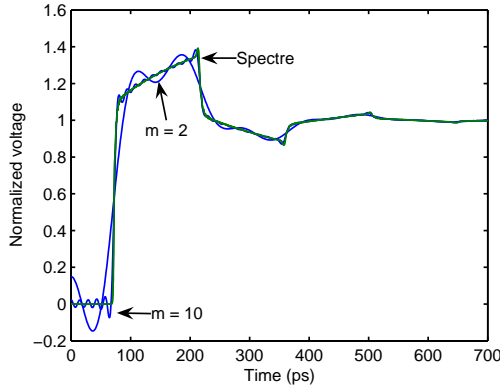
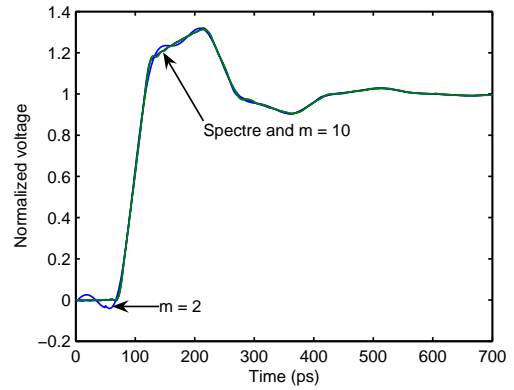
(a) Step response, RC (b) Ramp response, RC , $t_r=50$ ps(c) Step response, RLC (d) Ramp response, RLC , $t_r=50$ ps

Figure 4.8: Step and ramp response obtained analytically as compared with Spectre simulations. (a) Step response, RC , (b) Ramp response, RC , (c) Step response, RLC , and (d) Ramp response, RLC .

The step and ramp responses obtained from (4.22) and (4.23) are compared with Spectre simulations in Fig. 4.8. In the Spectre simulation, the transmission line is modeled as a series of π -shaped RC or RLC segments. Each segment is $10\ \mu\text{m}$ long. Good agreement between the analytic solution and Spectre simulations is observed.

The accuracy of the ramp response is much higher than that of the step response since a ramp signal consists of fewer high frequency components.

4.3 Distributed *RLC* Interconnect with Driver Resistance

For an interconnect driven by a gate, there are primarily two kinds of approaches for timing analysis. In the first approach, the driver and the interconnect are separated. The voltage waveform at the gate output is obtained through pre-characterized delay and transition time information characterizing the gate [119]. This waveform is applied at the input of the interconnect to obtain the far end response. With increasing inductive effects, more complicated driver output models are required to characterize the reflection behavior of the propagating signals, such as the two-ramp model described in [120] and the three-piece model in [121]. Recently, several current source models (CSM) have been developed [122, 123, 124], where the nonlinear behavior of the gate is characterized, making the driver output response more accurate. In the second approach, the driver and interconnect are analyzed as a single system, where the Thevenin model is generally used [107, 105, 118, 59, 125], as shown in Fig. 4.1. In this approach, the interaction between the driver and the interconnect is modeled as a single system.

For the first approach, the analytic solution proposed in Section 4.2.2 can be applied directly by representing the driver output voltage response as a piecewise-linear waveform. For the second voltage-approach, the method proposed in Section 4.2.2 needs to be improved to include the effect of the driver resistance. With a system transform, the poles of a general *RLC* interconnect system are solved in Section 4.3.1. The accuracy of the poles are further improved with the Newton-Raphson method as described in Section 4.3.2. The accuracy and efficiency of the proposed model are discussed in Section 4.3.3.

4.3.1 System Transform

In [117], the circuit model as shown in Fig. 4.1 is mapped into an open-ended interconnect system by matching the moments. Similarly, the interconnect system with a driver resistance can also be mapped into a system without a driver resistance. Consider a step signal at the input of the circuit shown in Fig. 4.1. The height of the initial step at the driver output is $V_{dd}Z_0/(R_d + Z_0)$, where $Z_0 = \sqrt{L/C}$ is the characteristic impedance of a lossless line. As described in [51], the attenuation coefficient of a transmission line saturates with increasing frequency to the asymptotic value $R/(2Z_0)$. Assume the total interconnect resistance of the new system (without a driver resistance) is R' and the load capacitance is C'_L . By matching the amplitude

of the initial propagating wave,

$$V_{dd} \frac{Z_0}{R_d + Z_0} e^{-\frac{R}{2Z_0}} = V_{dd} e^{-\frac{R'}{2Z_0}}, \quad (4.27)$$

R' can be obtained as

$$R' = R + 2Z_0 \log\left(1 + \frac{R_d}{Z_0}\right). \quad (4.28)$$

By matching the first moments of the two systems,

$$-m_1 = R_d(C_L + C) + R(0.5C + C_L) = R'(0.5C + C'_L), \quad (4.29)$$

C'_L can be obtained as

$$C'_L = \frac{-m_1}{R'} - 0.5C. \quad (4.30)$$

After this conversion, the method proposed in Section 4.2.2 can be applied. In Fig. 4.9, the waveform obtained from the proposed model is compared with Spectre simulations and another four-pole model described in [109]. This four-pole model is obtained by truncating the denominator of the transfer function to the fourth order; however, no closed-form solution is available for solving the four poles.

Note that although both the proposed model and the four-pole model are based on an approximation of the four poles of the system, the proposed model is much more accurate than the four-pole model when inductive effects are important (a system with

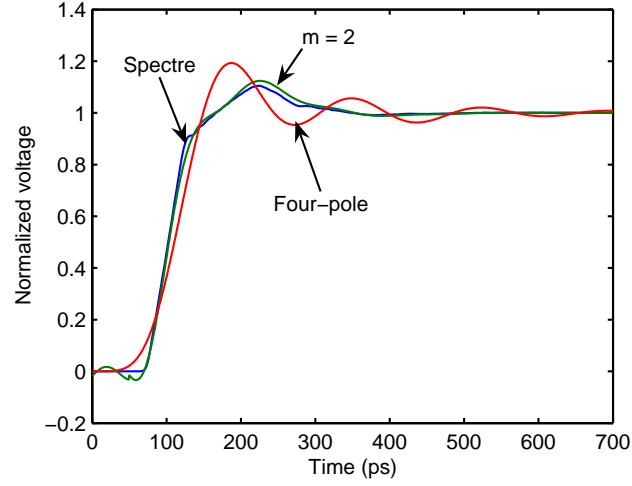
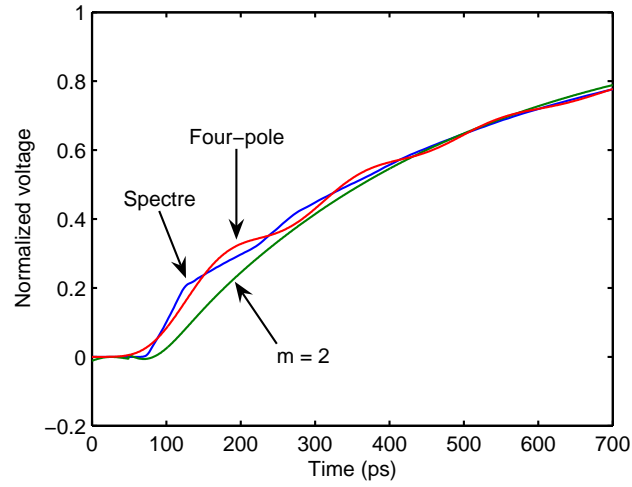
(a) $R_d = 20 \, \Omega$ (b) $R_d = 300 \, \Omega$

Figure 4.9: Transient response of a transmission line obtained with the proposed model, four-pole model, and Spectre simulations. $t_r = 50 \text{ ps}$ and $C_L = 50 \text{ fF}$. (a) $R_d = 20 \, \Omega$. (b) $R_d = 300 \, \Omega$.

a small driver resistance), as shown in Fig. 4.9(a). When the system is dominated by the driver resistance, the proposed model is less accurate, particularly at the beginning period of the waveform, as shown in Fig. 4.9(b).

4.3.2 Improve the Accuracy of the Poles

The location of the low order poles obtained analytically is compared with the location of the exact poles in Fig. 4.10. From the figure, note that there is a one-to-one mapping between the approximated poles and the exact poles. The real pole without an arrow in Fig. 4.10 should be mapped to a real pole which is out of the range of the figure. From these approximated poles, the exact poles can be obtained through the Newton-Raphson method, permitting the accuracy of the model to be significantly improved. In general, the number of iterations required for convergence is less than five.

Special attention needs to be paid to those real poles when applying the Newton-Raphson method. For example, the Newton-Raphson process starting from the approximated pole -3.892×10^{10} (the left real pole as shown in Fig. 4.10) incorrectly converges to the exact pole -6.396×10^9 rather than converges to the exact pole outside the range of the figure. In order to distinguish this case from the double real pole case, the following condition needs to be evaluated. If p is a double real pole of

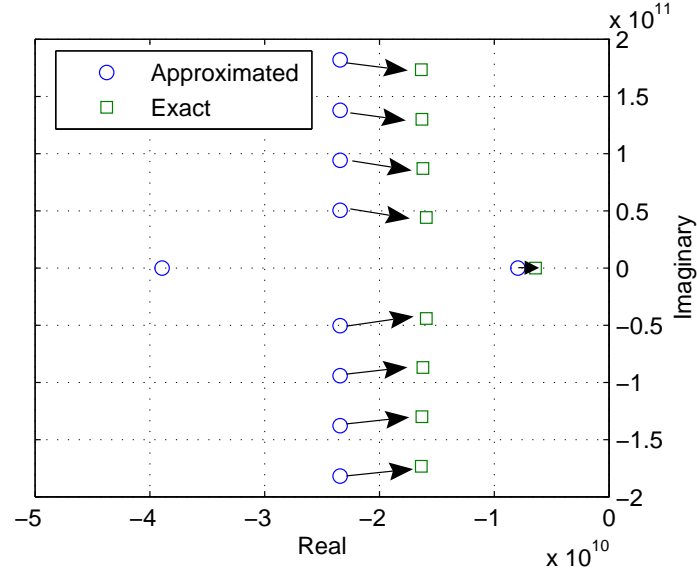


Figure 4.10: Mapping between the approximated poles and the exact poles, $R_d = 100\Omega$.

the system, p satisfies the following expression,

$$\lim_{s \rightarrow p} \frac{F(s)}{s - p} = F'(p) = 0. \quad (4.31)$$

For systems with multiple real poles, the system is dominated by the real pole with the smallest magnitude and the effect of the other real poles can be ignored, unless these poles are close to the dominant pole. The distance between the other real poles and the dominant real pole is related to the value of $F'(p_d)$, where p_d is the dominant pole. If there is another pole p_x which is close to p_d , $F'(p_d)$ should be small. When p_x approaches p_d , the value of $F'(p_d)$ approaches zero. In the limit, $p_x = p_d$, p_d is a double pole, and $F'(p_d) = 0$, as expressed in (4.31).

```

Input:  $R, L, C, R_d, C_L$ 
Output:  $m$  pairs of low-order poles
Find_poles( $R, L, C, R_d, C_L$ )
{
    Calculate  $R'$  and  $C'_L$ 
    Calculate  $m$  pairs of approximated poles  $p_{n,\pm}$ 

    over_damped=0
    for  $n = 0 : m - 1$ 
    {
         $p_{n,\pm}$ =Newton_Raphson( $p_{n,\pm}$ )
        if  $p_{n,+}$  is real
            if over_damped==1
                Discard poles  $p_{n,\pm}$ 
            else
                over_damped=1
                if  $F'(p_{n,+}) < F_{th}$ 
                     $p_{n,-}$ =Newton_Raphson( $2p_{n,+}$ )
                    if  $p_{n,-} == p_{n,+}$ 
                        output message: double-pole case
                        return
                    end
                else
                     $p_{n,-}$ =Newton_Raphson( $5p_{n,+}$ )
                    if  $p_{n,-} == p_{n,+}$  or  $p_{n,-}$  does not converge
                        Discard  $p_{n,-}$ 
                    end
                end
            end
        end
    }

    return  $p_{n,\pm}, n = 0 : m - 1$ 
}

```

Figure 4.11: Pseudo-code for computing the exact poles. Function *Newton_Raphson()* is the Newton-Raphson converging process starting with the input argument.

Pseudo-code for generating the exact poles of a single interconnect system is shown in Fig. 4.11. In Fig. 4.11, the variable *over_damped* is used to indicate whether the system is overdamped or not. For overdamped systems, the higher order real poles (with $n > 0$) are ignored. A threshold value F_{th} is set for $F'(p)$, which is used to indicate the distance between other high order real poles and the dominant real pole.

After the dominant real pole (if the system has real poles, the dominant real pole is always $p_{0,+}$) is found, $F'(p_{0,+})$ is evaluated. $F(s)$ can be represented by the poles as

$$F(s) = \prod_{n=0}^{\infty} \left(1 - \frac{s}{p_{n,+}}\right) \left(1 - \frac{s}{p_{n,-}}\right). \quad (4.32)$$

From (4.32),

$$F'(p_{0,+}) = \frac{-1}{p_{0,+}} \left(1 - \frac{p_{0,+}}{p_{0,-}}\right) \prod_{n=1}^{\infty} \left(1 - \frac{p_{0,+}}{p_{n,+}}\right) \left(1 - \frac{p_{0,+}}{p_{n,-}}\right) < \frac{-1}{p_{0,+}} \left(1 - \frac{p_{0,+}}{p_{0,-}}\right). \quad (4.33)$$

If $|p_{0,-}| > 2|p_{0,+}|$, $F'(p_{0,+}) < -0.5/p_{0,+}$. With some overhead, F_{th} is determined as $-0.3/p_{0,+}$. If $F'(p_{0,+}) < F_{th}$, which means pole $p_{0,-}$ is close to $p_{0,+}$, a *Newton-Raphson* process is launched from point $2p_{0,+}$ to determine $p_{0,-}$. Otherwise, the *Newton-Raphson* process is launched from point $5p_{0,+}$ to find $p_{0,-}$. If the process does not converge or incorrectly converges to $p_{0,+}$, which means the true value of $|p_{0,-}|$ is greater than $5|p_{0,+}|$, the effect of $p_{0,-}$ can be ignored. For the double pole case, the process of solving the residue requires the second order derivative of $F(s)$, which is complicated. The code produces an output message if a double pole occurs. In this case, a small change in the circuit parameters can avoid a double pole, while the effect on the output signal waveform caused by this parameter change cannot be distinguished. After the exact poles are extracted, a step or ramp response is constructed from (4.22) or (4.23). In order to eliminate the artificial discontinuity of

the waveform at the end of the input rising edge, the first moment m_1 in (4.24) is calculated from the truncated summation, as shown in the left side of (4.26), rather than the exact value of $0.5R'C + R'C'_L$.

For the same circuit examples used in Fig. 4.9, the waveform obtained from the improved method is re-plotted in Fig. 4.12. From Fig. 4.12, the difference between the analytic waveforms and Spectre simulations is difficult to distinguish except for the period of the initial time-of-flight.

4.3.3 Model Accuracy and Efficiency

The 50% delay, 10%-to-90% rise time, and the normalized overshoot obtained from the proposed model are compared in Fig. 4.13 with Spectre simulations for different input rise times (the input rise time is determined from 0 to V_{dd}). Since the signal delay is generally determined by the low frequency components, two pairs of poles provide a sufficiently accurate delay estimation. The average error is 1% for different input rise times. For the output rise time and overshoot, the error is larger for smaller input rise times. The error decreases with increasing input rise time, since the output rise time and overshoot are closely related to the high frequency components (a signal with a shorter rise time consists of additional high frequency components). The average error with two pairs of poles is 9.5% for the output rise time and 5.5% for the overshoot. When the number of pole pairs increases to ten,

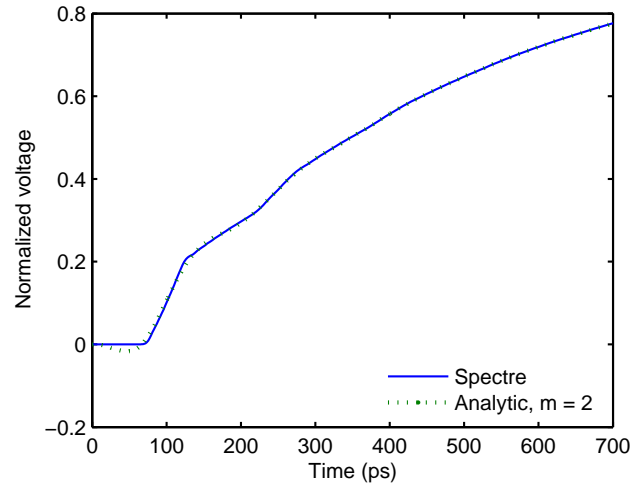
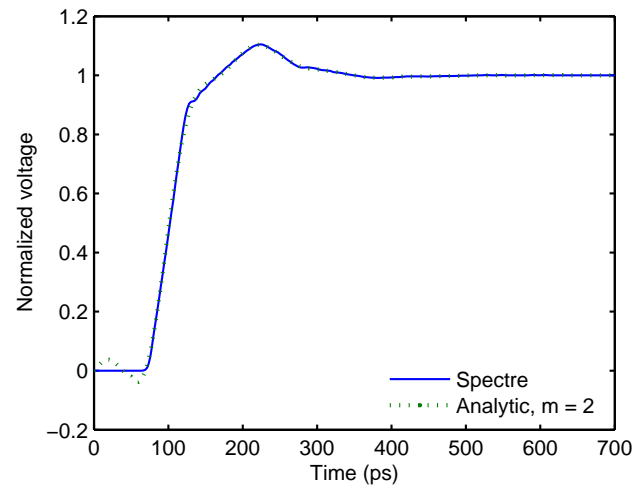
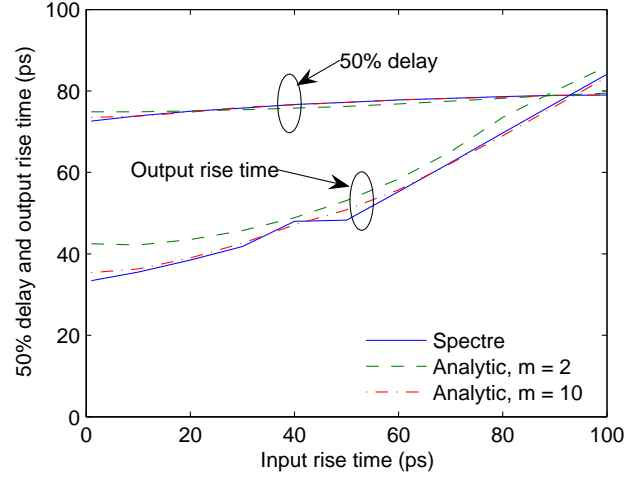
(a) $R_d = 20 \, \Omega$ (b) $R_d = 300 \, \Omega$

Figure 4.12: Transient response of transmission line obtained with the improved analytic method as compared with Spectre simulations. $m = 2$. (a) $R_d = 20 \, \Omega$. (b) $R_d = 300 \, \Omega$.

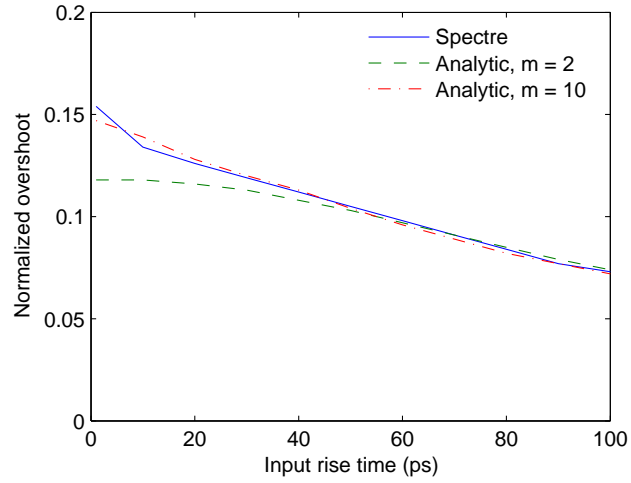
these two average errors decrease to 2.0% and 1.9%, respectively. The computational complexity of the proposed method is approximately proportional to the number of pole pairs. These experiments have been performed on a SunBlade1500 workstation. The time required for Spectre to perform a 700 ps transient simulation (250 time steps) is 1.8 s. The proposed model is implemented with Matlab. The run time is 3.1 ms for $m = 2$ and 10.9 ms for $m = 10$. To achieve a similar accuracy of the proposed model ($m = 2$), more than 12 poles are required in the traditional moment matching method. Since there are no closed-form solutions for solving the poles from the moments, the computational complexity of the moment matching method is high as compared with the proposed method. Specifically, the run time for the moment matching method with 12 poles is 13.5 ms as compared to 3.1 ms for the proposed method with $m = 2$. Furthermore, the moment matching method suffers the numerical stable problem for high order approximations. The accuracy of the proposed model is also verified for different interconnect lengths and is illustrated in Fig. 4.14.

4.4 Frequency Dependent Effects

Both interconnect inductance and resistance are a function of frequency. This frequency dependent interconnect impedance affects the signal waveform, particularly for those signals containing a greater number of high frequency components.

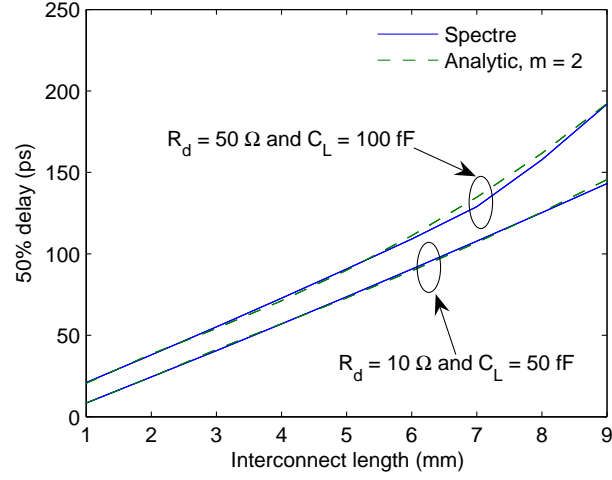


(a) Delay and output rise time

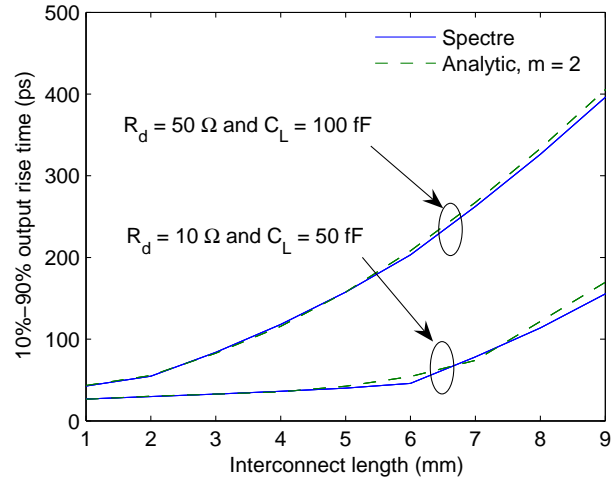


(b) Overshoot

Figure 4.13: Comparison of the 50% delay, 10%-to-90% output rise time, and the normalized overshoot obtained from the proposed model and Spectre simulations. $R_d = 20 \Omega$, $C_L = 50 \text{ fF}$, and $l = 5 \text{ mm}$. (a) Delay and rise time. (b) Overshoot.



(a) 50% delay



(b) 10%-to-90% output rise time

Figure 4.14: Comparison of the 50% delay and 10%-to-90% output rise time obtained from the proposed model and Spectre simulations. $t_r = 50$ ps. (a) 50% delay. (b) 10%-to-90% output rise time.

From (4.28), the contribution of the driver resistance to the effective interconnect resistance is $R_{d_eff} = 2Z_0 \log(1 + R_d/Z_0)$, which is frequency independent (the frequency dependence due to Z_0 is ignored, and the Z_0 used here is determined at DC). The effective load capacitance is also determined at DC, as shown in (4.30). Considering the effect of the driver resistance and the frequency dependence of R and L of the interconnect, the effective propagation coefficient θ becomes

$$\theta = \sqrt{[R_{d_eff} + R(s) + L(s)s]Cs}. \quad (4.34)$$

For different functional forms of $R(s)$ and $L(s)$, the poles of the transfer function of an interconnect can be obtained by solving (4.34). Closed-form solutions may also be available depending upon the expressions of $R(s)$ and $L(s)$.

The frequency dependent impedance can be modeled by ladder structures of frequency-independent elements [45, 57]. These ladder structures are particularly suitable to capture the skin effects. A two stage ladder structure [45] is adopted in this chapter for simplicity, as shown in Fig 4.15.

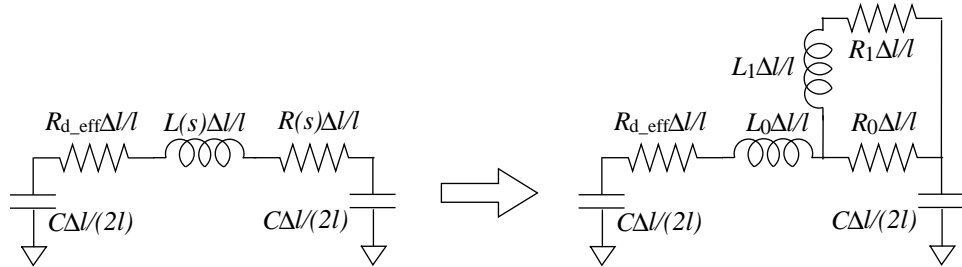


Figure 4.15: A segment of interconnect with length Δl .

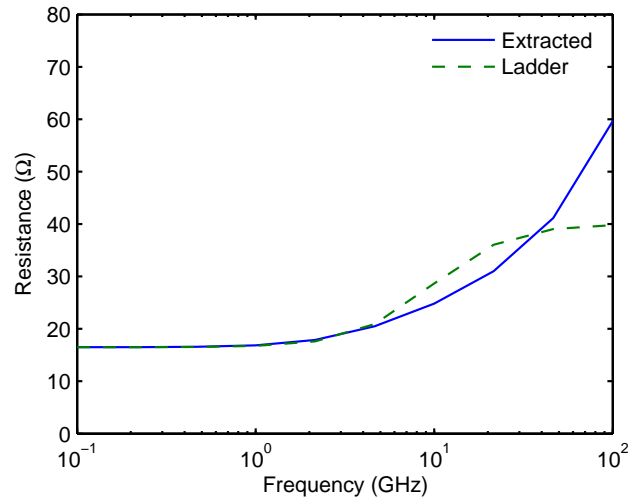
Since the frequency dependent effect is naturally more significant at high frequencies, a wider interconnect is adopted here as an example so that additional high frequency components can propagate across the interconnect, making the frequency dependent effect distinguishable. The signal wire width is $10\text{ }\mu\text{m}$, the space between the signal line and ground is $5\text{ }\mu\text{m}$, and the remaining geometric parameters are the same as depicted in Fig. 4.6. The parameters in the ladder structure are calculated by matching the DC and high frequency resistance and inductance of the ladder structure with the extracted values. Since the resistance of the interconnect does not saturate at high frequencies, a value of $40\text{ }\Omega$ is assumed as the high frequency resistance in this example, resulting in the following parameters: $R_0 = 40\text{ }\Omega$, $R_1 = 28.1\text{ }\Omega$, $L_0 = 1.9\text{ nH}$, and $L_1 = 1.12\text{ nH}$. The DC impedance is $R_{dc} = 16.5\text{ }\Omega$, $L_{dc} = 2.287\text{ nH}$, and $C = 4.18\text{ pF}$. The resistance and inductance of the ladder approximation are compared with the extracted values in Fig. 4.16.

With this ladder approximation, the expression used to solve the poles of the system becomes

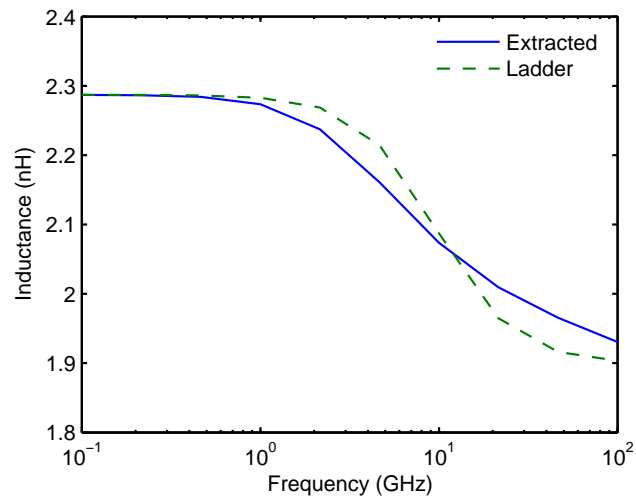
$$\left[R_{d_eff} + L_0 s + \frac{R_0(R_1 + L_1 s)}{R_0 + R_1 + L_1 s} \right] C s = \theta^2 = -x_n^2. \quad (4.35)$$

The poles can be analytically solved as

$$p_{n,\pm} = -\frac{a_2}{3} - \frac{X}{2} \pm \frac{\sqrt{3}}{2} i \sqrt{X^2 + 4Q}, \quad (4.36)$$



(a) Resistance



(b) Inductance

Figure 4.16: Frequency dependent impedance of an interconnect with a length of 5 mm. (a) Resistance. (b) Inductance.

where

$$Q = \frac{3a_1 - a_2^2}{9}, \quad (4.37)$$

$$P = \frac{9a_1a_2 - 27a_0 - 2a_2^3}{54}, \quad (4.38)$$

$$X = \sqrt[3]{P + \sqrt{Q^3 + P^2}} + \sqrt[3]{P - \sqrt{Q^3 + P^2}}, \quad (4.39)$$

and

$$a_2 = \frac{L_0R_0 + L_0R_1 + R_0L_1 + L_1R_{d_eff}}{L_0L_1}, \quad (4.40)$$

$$a_1 = \frac{R_0R_1C + (R_0 + R_1)R_{d_eff}C + x_n^2L_1}{L_0L_1C}, \quad (4.41)$$

$$a_0 = \frac{(R_0 + R_1)x_n^2}{L_0L_1C}. \quad (4.42)$$

From (4.36), the Newton-Raphson method can be applied to solve the exact poles as illustrated in Section 4.3.2. In Fig. 4.17, the output signal waveforms are compared for the DC impedance case and the frequency dependent (FD) impedance case. As shown in Fig. 4.17, by considering the FD effect, additional high frequency components are suppressed, making the waveform smoother since the high frequency components experience much greater attenuation due to the increasing interconnect resistance, as shown in Fig. 4.18. For the high frequency related waveform properties, such as the rise time and overshoot, the FD effect should be considered. For low frequency related waveform properties, such as delay, the FD effect can be neglected.

Similar results are also described in [126]. The run time of the Spectre simulation (700 ps, 225 time steps) is 2.45 s and the run time for the proposed analytic method ($m = 2$) is 3.8 ms (a three orders of magnitude improvement in computational time).

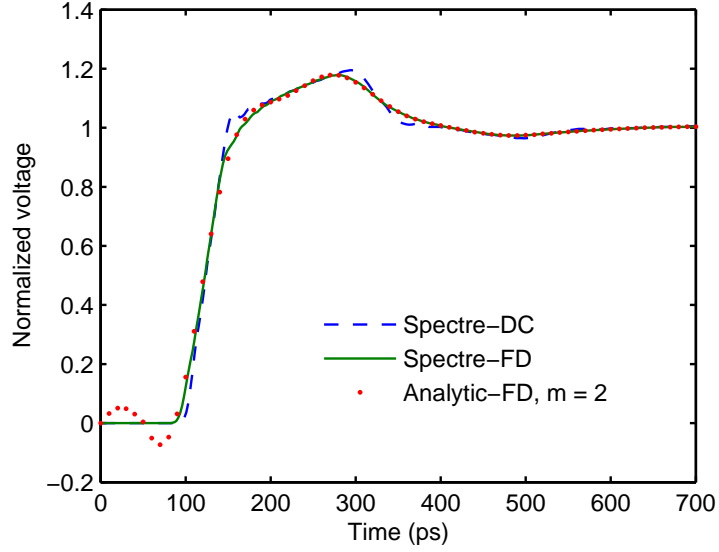


Figure 4.17: Comparison of the output signal waveforms with and without the frequency dependent effect, $R_d = 10 \Omega$, $C_L = 50 \text{ pF}$, and $t_r = 50 \text{ ps}$.

4.5 Conclusions

By extracting the exact poles, an efficient method has been proposed in this chapter for determining the transient output response of a distributed RLC interconnect. As demonstrated in this chapter, two pairs of poles can provide an accurate delay estimate exhibiting an average error of 1% as compared with Spectre simulations. For

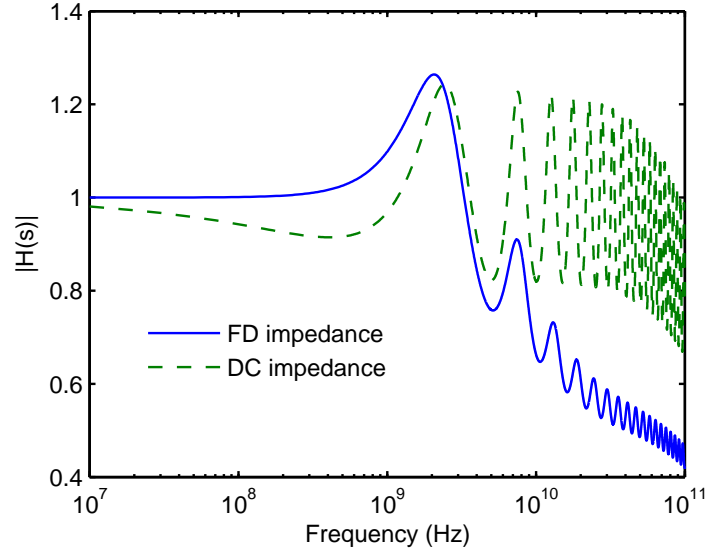


Figure 4.18: Comparison of transfer functions with and without the frequency dependent effect, $R_d = 10 \Omega$ and $C_L = 50 \text{ pF}$.

high frequency related waveform properties, such as the rise time and overshoot, an average error of less than 2% can be obtained with ten pairs of poles. The computational complexity of the proposed method is proportional to the number of pole pairs. By using a ladder structure, frequency dependent effects can also be included in the method. Excellent agreement is observed between the proposed model and Spectre simulations.

Chapter 5

Effective Capacitance of RLC Loads for Estimating Short-Circuit Power

5.1 Introduction

Since power has become an important design criterion in integrated circuits, accurate and efficient power estimation is required in the circuit design process. As compared with dynamic power which is well characterized, short-circuit power is more difficult to model due to the complicated transient behavior of the short-circuit current. In [127], Veendrick developed a closed form expression for short-circuit power dissipation in an unloaded CMOS inverter. More accurate analyses have been presented recently by including short-channel effects [128, 129] and output overshoot effects [130, 131]. In these analyses, a lumped capacitor is assumed as the load. An L shaped RC model is used in [132] and a π shaped RC model is used in [133] and

[134] to characterize the shielding effect of the interconnect resistance. An effective capacitance of the RC π structure for short-circuit power estimation is described in [125] and [135] to maintain compatibility with popular look-up table or k -factor based power models. With increasing on-chip frequencies and longer interconnects, the interconnect inductance also needs to be considered. As described in [136], the interconnect inductance also exhibits a shielding effect on the load capacitance, increasing the short-circuit power dissipated by the driver.

In this chapter, an effective capacitance of an RLC load is developed to accurately estimate short-circuit power. Both resistive and inductive shielding effects are considered. The rest of this chapter is organized as follows. In Section 5.2, a distributed RLC network is reduced into a π model. From this π model, the effective capacitance is determined. In Section 5.3, the proposed effective capacitance model is verified by Cadence Spectre simulations. Finally, some conclusions are offered in Section 5.4.

5.2 Effective Capacitance of an RLC Load

Model order reduction techniques are commonly used to analyze the timing and power of interconnects to improve simulation efficiency. In Section 5.2.1, a π model is generated from a distributed RLC tree through moment matching. In Section 5.2.2, this π model is further reduced into an effective capacitance.

5.2.1 π -Model Representation of RLC Interconnects

By matching the first three moments (y_1 , y_2 , and y_3) of the admittance at the driving point, an RC network can be reduced into an RC π model [137]. In the same way, an RLC network can be reduced into an RLC π model by matching the first four moments, as shown in Fig. 5.1. This reduction, however, can be unrealizable (the value of the circuit element is not positive real). In order to obtain a realizable RLC π model, a coefficient y_3^* is introduced in [138], which is the third order admittance moment (without considering inductance). By matching y_1 , y_2 , y_3 , and y_3^* , the π model parameters can be obtained as [138]

$$C_f = y_2^2 / y_3^*, \quad (5.1)$$

$$C_n = y_1 - C_f, \quad (5.2)$$

$$R_\pi = -y_2 / C_f^2, \quad (5.3)$$

$$L_\pi = (y_3^* - y_3) / C_f^2, \quad (5.4)$$

where C_n and C_f denote the near end and far end capacitance, respectively.

The input admittance of a distributed RLC interconnect with a load admittance Y_l is [118]

$$Y(s) = \frac{Z_c Y_l + \tanh \theta}{Z_c (1 + Z_c Y_l \tanh \theta)}, \quad (5.5)$$

where $\theta = \sqrt{(R_t + sL_t)sC_t}$ and $Z_c = \theta / (C_t s)$. R_t , C_t , and L_t are the total resistance,

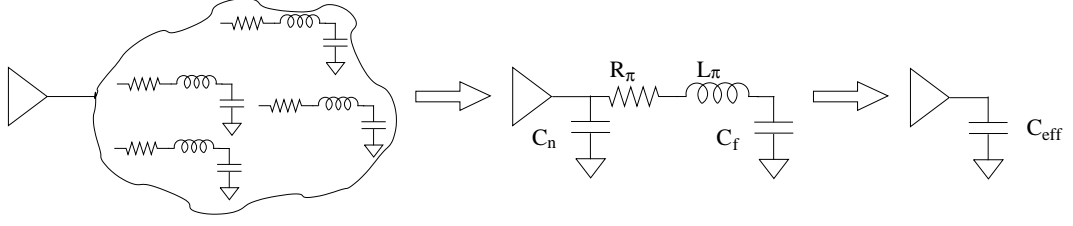


Figure 5.1: Reduction of an *RLC* interconnect network.

capacitance, and inductance of the interconnect, respectively. By expanding $Y(s)$ into a Taylor series of s around zero, the moments at the input of the *RLC* interconnect can be obtained as

$$y_1 = y_{l,1} + C_t, \quad (5.6)$$

$$y_2 = y_{l,2} - R_t(y_{l,1}^2 + y_{l,1}C_t + \frac{1}{3}C_t^2), \quad (5.7)$$

$$y_3 = y_{l,3} - R_t(2y_{l,1}y_{l,2} + y_{l,2}C_t) + R_t^2(y_{l,1}^3 + \frac{4}{3}y_{l,1}^2C_t + \frac{2}{3}y_{l,1}C_t^2 + \frac{2}{15}C_t^3) - L_t(y_{l,1}^2 + y_{l,1}C_t + \frac{1}{3}C_t^2), \quad (5.8)$$

$$y_3^* = y_{l,3}^* - R_t(2y_{l,1}y_{l,2} + y_{l,2}C_t) + R_t^2(y_{l,1}^3 + \frac{4}{3}y_{l,1}^2C_t + \frac{2}{3}y_{l,1}C_t^2 + \frac{2}{15}C_t^3). \quad (5.9)$$

The input admittance moments of a distributed *RLC* tree can be determined by recursively applying (5.6)-(5.9). From these moments and (5.1)-(5.4), the corresponding π structure can be obtained. If the *RLC* network includes resistance loops, the generalized Y- Δ transformation technique [139, 140] can be used to eliminate all of the nodes except the driver so as to determine the admittance moments.

5.2.2 Effective Capacitance for Short-Circuit Power

Although a π model is highly accurate, four coefficients are required in this model, making it incompatible with k -factor expressions or look-up table based power models. An effective capacitance greatly simplifies the model with little penalty in accuracy, as shown in Fig. 5.1.

The shielding effect of the interconnect resistance is well known and the effective capacitance of RC interconnects has been developed for estimating delay and short-circuit power in [125, 119]. The interconnect inductance however also has a shielding effect [136]. This inductive shielding effect is illustrated with an example (shown in Fig. 5.2). In Fig. 5.2, a distributed RLC tree is driven by a $0.18\ \mu\text{m}$ CMOS inverter. The size of the transistors in the inverter is $W_n = 10\ \mu\text{m}$ and $W_p = 25\ \mu\text{m}$. The impedance parameters of the interconnect are $R_{int} = 12.23\ \text{m}\Omega/\mu\text{m}$ and $C_{int} = 0.245\ \text{fF}/\mu\text{m}$. The load capacitance is $C_L = 100\ \text{fF}$. The short-circuit current of the inverter is illustrated in Fig. 5.3 for different values of interconnect inductance. When the interconnect inductance becomes larger, greater far end capacitance is shielded. Less effective capacitance is therefore seen at the inverter output, permitting the output voltage to change faster at the beginning of the signal transition, thereby producing a larger short-circuit current. The currents are measured at the source of the PMOS transistor with a rising edge input as shown in Fig. 5.4. The discontinuity of the waveform is due to the discontinuity of the transistor capacitance model used

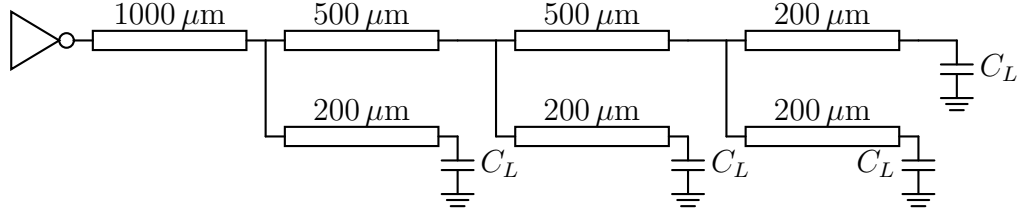


Figure 5.2: An example of a distributed RLC tree.

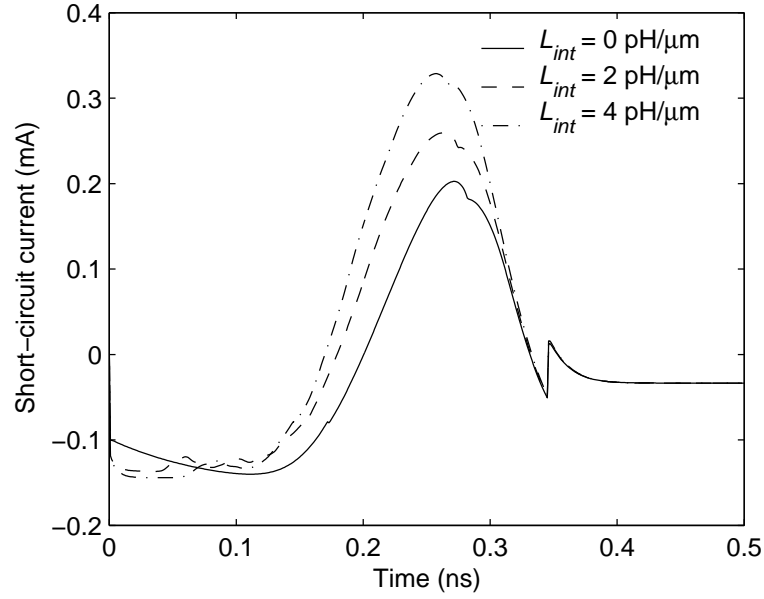


Figure 5.3: Effect of inductance on short-circuit current. $t_r = 0.5$ ns.

in the simulation. Strictly speaking, the currents shown in Fig. 5.3 include two non-short-circuit current components. The first component is the current flowing through the capacitance C_{gs} , as shown in Fig. 5.4. This component can be determined as $I_{gs} = C_{gs}V_{dd}/t_r$ and is independent of the load. The second component is the current I_{ov} flowing from the output to V_{dd} due to the overshoot at the output at the beginning of the signal transition. I_{ov} returns a small amount of charge stored in the output node back to V_{dd} , slightly reducing the dynamic power.

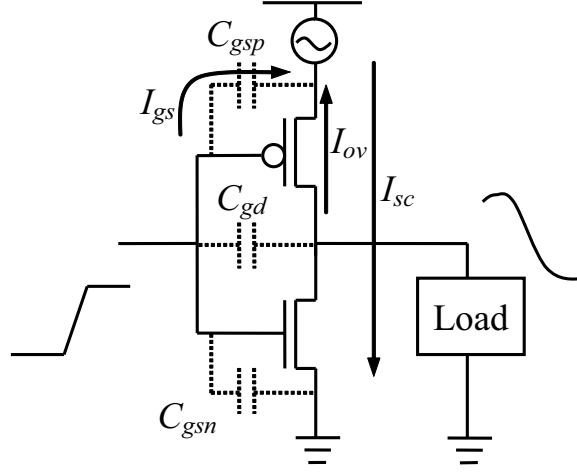


Figure 5.4: Current components in a CMOS inverter.

In [119], the output waveform of a CMOS gate is approximated by a quadratic function followed by a linear function. In this chapter, the output waveform of a gate is modeled as a quadratic function during the input transition. Assuming the output waveform is $v(t) = at^2$ for a rising edge, the current drawn from the gate by an RLC π structure is

$$I_{\pi}(s) = \frac{2a}{s^3} \left(\frac{C_f s}{1 + R_{\pi} C_f s + L_{\pi} C_f s^2} + C_n s \right). \quad (5.10)$$

Applying an inverse Laplace transformation on (5.10), the current in the time domain is

$$i_{\pi} = 2aC_f(-R_{\pi}C_f + t + k_1e^{s_1t} + k_2e^{s_2t}) + 2aC_nt, \quad (5.11)$$

where

$$s_{1,2} = \frac{-R_\pi \pm \sqrt{R_\pi^2 - \frac{4L_\pi}{C_f}}}{2L_\pi}, \quad (5.12)$$

$$k_1 = \frac{1}{s_1^2(s_1 - s_2)L_\pi C_f}, \quad (5.13)$$

$$k_2 = \frac{1}{s_2^2(s_2 - s_1)L_\pi C_f}. \quad (5.14)$$

The current drawn from the gate by an effective capacitance is

$$i_{ceff} = 2aC_{eff}t. \quad (5.15)$$

Equating the average of i_π and i_{ceff} during a period from 0 to an evaluation time t_{ev} , C_{eff} can be obtained as

$$C_{eff} = C_n + C_f \left[1 - \frac{2R_\pi C_f}{t_{ev}} + \frac{2k_1}{t_{ev}^2 s_1} (e^{s_1 t_{ev}} - 1) + \frac{2k_2}{t_{ev}^2 s_2} (e^{s_2 t_{ev}} - 1) \right]. \quad (5.16)$$

Similarly, C_{eff} for an RC π structure is

$$C_{eff} = C_n + C_f \left[1 - \frac{2R_\pi C_f}{t_{ev}} + \frac{2R_\pi^2 C_f^2}{t_{ev}^2} (1 - e^{-\frac{t_{ev}}{R_\pi C_f}}) \right]. \quad (5.17)$$

As expected, C_{eff} is between C_n and $C_n + C_f$. From (5.16), C_{eff} is a function of t_{ev} as shown in Fig. 5.5. The π model parameters are obtained from the tree

structure as shown in Fig. 5.2 with an inductance per unit length $L_{int} = 0.74 \text{ pH}/\mu\text{m}$. With increasing t_{ev} , C_{eff} increases from C_n and approaches $C_n + C_f$. In [119], t_{ev} is the time when the driver output achieves 50% of V_{dd} , which is the objective and is not known *a priori*. Several iterations are therefore required to determine C_{eff} . In [125], t_{ev} is determined as the end point of the short-circuit period. Since the short-circuit current exists when the input is between V_{thn} and $V_{dd} + V_{thp}$, the appropriate evaluation time t_x is in the range from 0 to $t_r(1 - \frac{|V_{thp}|}{V_{dd}} - \frac{V_{thn}}{V_{dd}})$. Note that $t = 0$ corresponds to the time when the input reaches V_{thn} for a rising edge ($V_{dd} + V_{thp}$ for a falling edge). As shown in Fig. 5.5, for the time period $0 < t < t_x$, the effective capacitance is overestimated and the short-circuit current is underestimated. For the period $t_x < t < t_r(1 - \frac{|V_{thp}|}{V_{dd}} - \frac{V_{thn}}{V_{dd}})$, the effective capacitance is underestimated and the short-circuit current is overestimated. By properly adjusting t_x , the estimation error of the short-circuit current in different time regions can be canceled. By comparing Spectre simulations, a fitting parameter is adopted to determine t_x ,

$$t_x = 0.46t_r(1 - \frac{|V_{thp}|}{V_{dd}} - \frac{V_{thn}}{V_{dd}}). \quad (5.18)$$

The short-circuit current waveforms in an inverter with different load models are compared in Fig. 5.6. For this inverter, $V_{thn} = 0.5 \text{ V}$ and $V_{thp} = -0.5 \text{ V}$. As shown in Fig. 5.6, the π model can accurately characterize a tree structure. The waveform obtained with a π model is indistinguishable from the waveform with the original

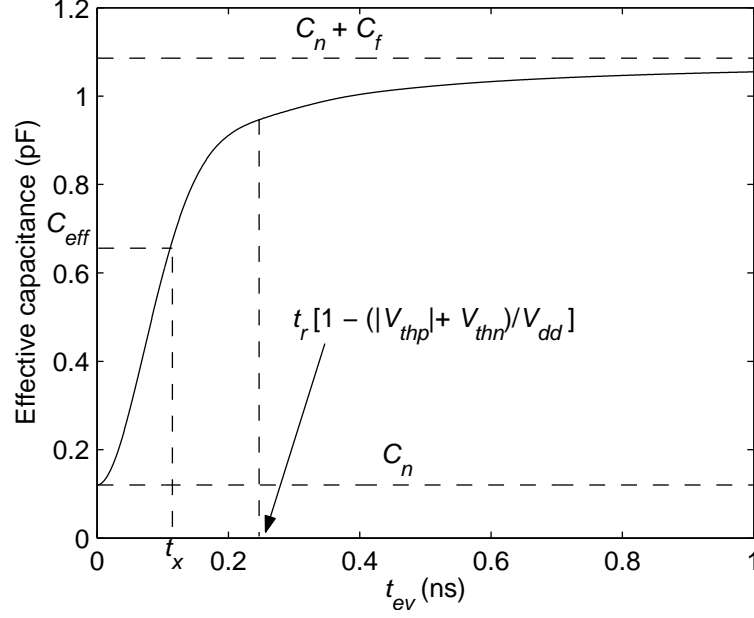


Figure 5.5: Effective capacitance as a function of t_{ev} . $C_n = 120.1$ fF, $C_f = 965.9$ fF, $R_\pi = 15.9 \Omega$, and $L_\pi = 0.96$ nH.

RLC tree (shown in Fig. 5.2). Using the total capacitance $C_{tot} = C_n + C_f$ as the load significantly underestimates the short-circuit current. As previously mentioned, the short-circuit current with the effective capacitance is first underestimated and then overestimated. No iterations are required to determine C_{eff} . Since an RLC π model is commonly required in timing analysis, the additional computational expense required to determine C_{eff} is small. Note that unlike C_{eff} for estimating the delay [119], C_{eff} for short-circuit power estimation is independent of the transistor size.

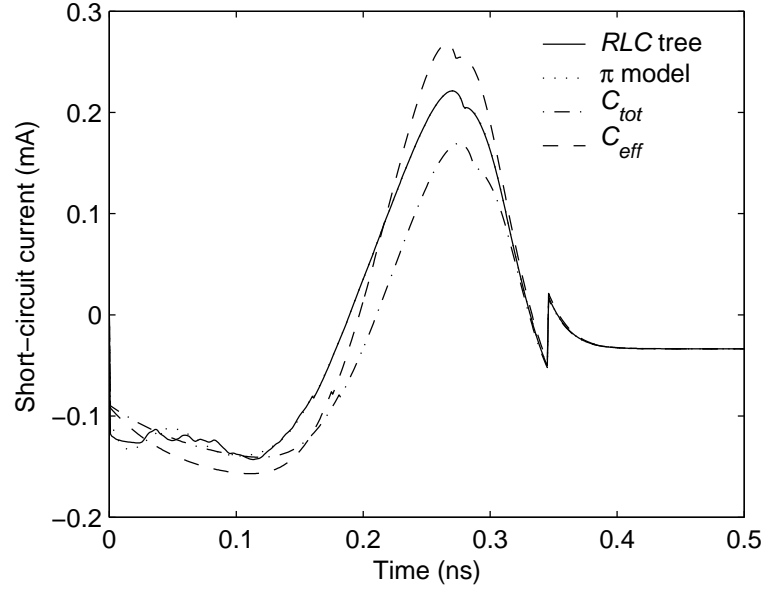


Figure 5.6: Short-circuit current with different output loads.

5.3 Model Verification

For the circuit shown in Fig. 5.2, the short-circuit energy dissipated over a full signal transition with different loads is compared in Fig. 5.7. As shown in Fig. 5.7, the total capacitance always underestimates the short-circuit energy as compared with a distributed RLC tree. For example, the error for $t_r = 0.5$ ns is 28.1%. More accurate estimations can be obtained with C_{eff_RC} (only considering the resistive shielding effect) and C_{eff} (considering both resistive and inductive shielding effects).

The inductive shielding effect is most important in the range from $t_r = 0.2$ ns to $t_r = 0.8$ ns for this example. The inductive shielding effect can be evaluated by the ratio of C_{eff} to C_{eff_RC} . In Fig. 5.8, this ratio is plotted with different input

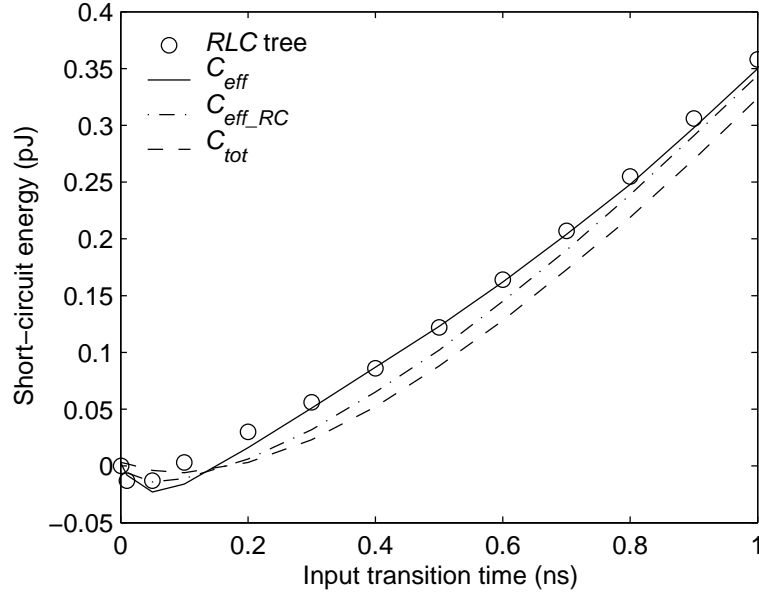


Figure 5.7: Short-circuit energy with different loads. $L_{int} = 0.74 \text{ pH}/\mu\text{m}$.

transition times and wire inductance values for the example circuit shown in Fig. 5.2. As shown in Fig. 5.8, C_{eff} decreases with increasing interconnect inductance. The ratio of C_{eff} to C_{eff_RC} can be smaller than 0.3. When t_r approaches zero, the driver only sees the near-end capacitance, both C_{eff} and C_{eff_RC} approach C_n , and the ratio C_{eff}/C_{eff_RC} approaches one. When t_r is sufficiently large, the driver has sufficient time to charge and discharge the far end capacitance, both C_{eff} and C_{eff_RC} approach C_{tot} , and the ratio also approaches one.

Since the dynamic power is usually determined as $\alpha f V_{dd}^2 C_{load}$ (where α is the switching factor), the reduction P_{red} in dynamic power due to I_{ov} is considered part of the short-circuit power such that the summation of the two power components (dynamic and short-circuit) can represent the total transient power. With fast inputs,

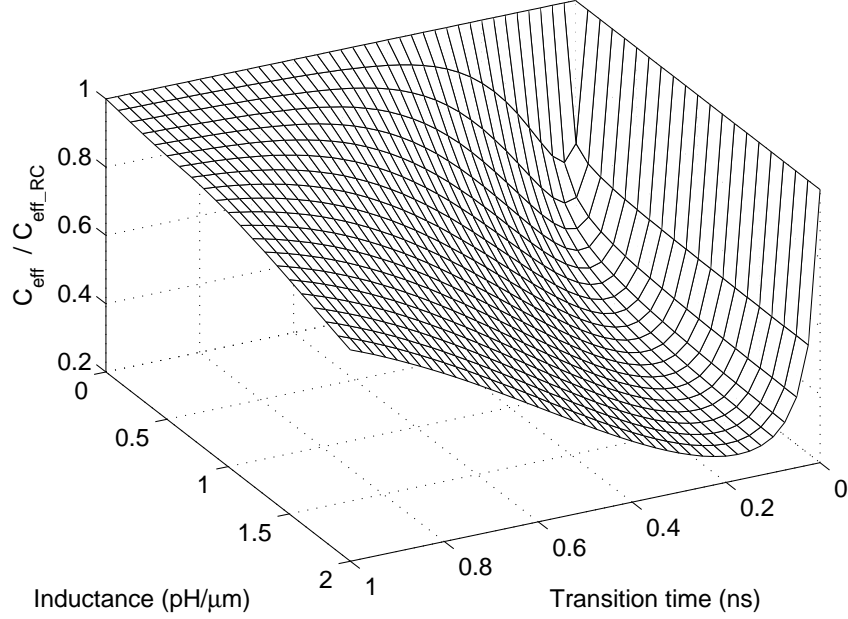


Figure 5.8: Effect of inductance on the effective capacitance.

P_{red} can dominate the short-circuit power, producing a negative short-circuit power, as shown in Fig. 5.7. Since P_{red} cannot be characterized by C_{eff} , the error of the power estimation is greater for fast inputs.

The effective capacitance concept can also be applied to other logic gates, such as NAND and NOR gates. The short-circuit energy consumed by an inverter and a two input NAND gate during a full signal transition is listed in Table 5.1 for different inputs and loads. The two inputs of the NAND gate are denoted as the upper input and the lower input according to the relative location of the input terminal. Three switching patterns are considered: only the upper input is switched (the lower input is tied to V_{dd}), only the lower input is switched, and both of the two inputs are

Table 5.1: Short-circuit energy dissipation during a full signal switch.

t_r (ns)	$R_\pi/L_\pi/C_n/C_f$ ($\Omega/\text{nH}/\text{fF}/\text{fF}$)	C_{eff} (fF)	Short-circuit energy (pJ)											
			Inverter			NAND(upper)			NAND(lower)			NAND(both)		
			π	C_{eff}	C_{tot}	π	C_{eff}	C_{tot}	π	C_{eff}	C_{tot}	π	C_{eff}	C_{tot}
0.5	100/2/200/600	369.4	0.16	0.16	0.11	0.12	0.12	0.08	0.11	0.10	0.07	0.06	0.05	0.02
1	100/2/200/600	517.3	0.45	0.44	0.38	0.35	0.35	0.30	0.36	0.35	0.29	0.30	0.29	0.24
2	100/2/200/600	641.3	1.13	1.13	1.06	0.93	0.93	0.87	0.98	0.97	0.91	0.88	0.86	0.81
0.5	200/3/100/800	205.2	0.20	0.20	0.10	0.15	0.15	0.08	0.14	0.13	0.06	0.09	0.07	0.02
1	200/3/100/800	322.6	0.51	0.51	0.36	0.41	0.41	0.28	0.41	0.41	0.28	0.35	0.34	0.23
2	200/3/100/800	483.3	1.21	1.20	1.01	0.99	1.00	0.83	1.05	1.05	0.87	0.94	0.93	0.78
0.5	300/4/100/300	167.9	0.21	0.21	0.15	0.16	0.16	0.12	0.15	0.14	0.10	0.09	0.08	0.05
1	300/4/100/300	228.8	0.56	0.56	0.48	0.45	0.45	0.38	0.45	0.45	0.38	0.38	0.37	0.32
2	300/4/100/300	292.7	1.34	1.33	1.25	1.11	1.11	1.04	1.16	1.16	1.09	1.03	1.02	0.96
Average % Error			—	0.9	22.1	—	1.2	22.2	—	1.6	24.4	—	7.2	33.5

connected and simultaneously switched. The size of the transistors in the NAND gate is $W_n = 10 \mu\text{m}$ and $W_p = 25 \mu\text{m}$. As listed in Table 5.1, the effective capacitance can accurately capture the shielding effect of the resistance and inductance in determining the short-circuit power. For a single switching input, the average error of the short-circuit power is less than 2% as compared with the π model. The average error with the total capacitance, however, is more than 20% for these examples. As compared to the single switching input, the error with C_{eff} for the connected inputs is greater, exhibiting an average error of 7.2%. With very slow inputs, although the output voltage waveform deviates from a quadratic behavior, the proposed method remains accurate as listed in Table 5.1. In these cases, the shielding effects are small, and the effective capacitance approaches C_{tot} . This behavior is well captured by (5.16).

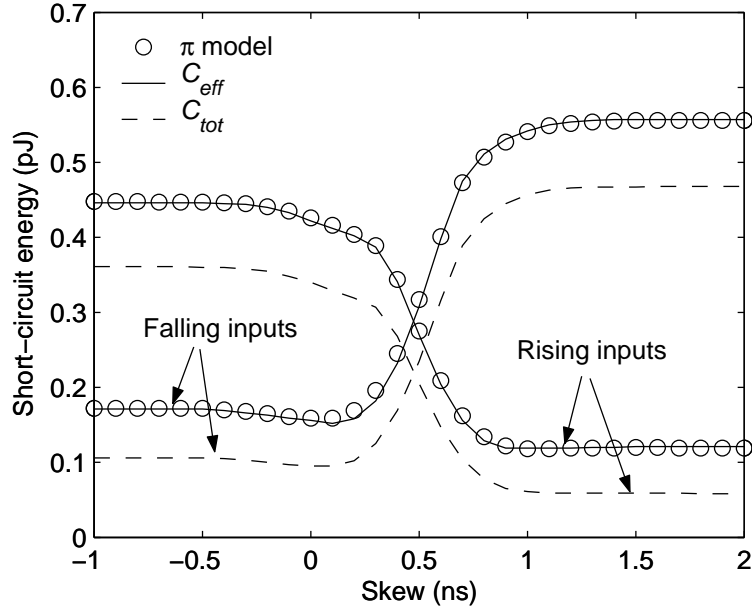


Figure 5.9: Short-circuit energy with multiple switching inputs. $C_n = 100$ fF, $C_f = 800$ fF, $R_\pi = 200 \Omega$, and $L_\pi = 3$ nH.

For multiple switching inputs with offsets in delay (non-simultaneous input signals), an equivalent input signal has been developed in [141] for estimating the short-circuit power. From this equivalent input signal, an effective capacitance can be obtained from (5.16) and (5.18). The short-circuit energy dissipated in a NAND gate is shown in Fig. 5.9 for different skew between the two inputs. The transition time of the upper input and lower input are 1 ns and 2 ns, respectively. The skew is determined as $t_{upper} - t_{lower}$, where t_{upper} and t_{lower} are the starting time of the transitions at the upper input and lower input, respectively. From Fig. 5.9, it can be seen that the effective capacitance model is also valid for multiple switching inputs with delay offsets.

Look-up tables or k -factor expressions are commonly used to model short-circuit power as a function of t_r and C_L . The total transient power in a CMOS gate driving an interconnect network, therefore, can be represented as

$$P_{total} = P_{sc}(t_r, C_{eff}) + \alpha f C_{load} V_{dd}^2, \quad (5.19)$$

where C_{load} includes the total interconnect capacitance and the parasitic capacitance of the transistors. If glitches occur at the output, both dynamic power and short-circuit power depend upon the transient voltage waveform at the output. Expression (5.19) is no longer valid in this case and a more complicated analysis is required.

5.4 Conclusions

In deep submicrometer integrated circuits, interconnects not only dominate the gate delay, but also greatly affect the power dissipation. In this chapter, an effective capacitance of a distributed RLC load is described to accurately estimate the short-circuit power. For a single switching input, the average error of the short-circuit power obtained with C_{eff} is less than 2% as compared with an RLC π model. This effective capacitance can be used in look-up tables or k -factor expressions to estimate short-circuit power as well as in analytic models to simplify the interconnect analysis process.

Chapter 6

Low Power Repeaters Driving RC and RLC Interconnects with Delay and Bandwidth Constraints

6.1 Introduction

Repeater insertion is an efficient method for driving long interconnects in integrated circuits as described in Subsection 2.5.3. The size of a delay-optimal repeater is typically much larger than a minimum sized repeater. Since millions of repeaters will be inserted to drive global interconnects in future high complexity circuits [142], significant power will be consumed by these repeaters, particularly if delay-optimal repeaters are used. A power-delay tradeoff is, therefore, necessary to support efficient repeater insertion design methodologies [143].

The number and size of the repeaters to minimize the dynamic power and area of an interconnect while satisfying a target delay constraint have been described by

Nalamalpu and Burleson in [144]. Burleson *et al.* further compared repeaters with boosters in [145]. The input transition time of a repeater is generally greater than the output transition time. In this case, the short-circuit power can be comparable or even greater than the dynamic power [146]. With CMOS technology scaling, leakage power is increasing rapidly, and is expected in future technologies to reach the same magnitude as the dynamic power [7]. By including both short-circuit and leakage power, a low power repeater design methodology is presented in [147]. The power is minimized with a 5% delay penalty. Closed form solutions, however, are not provided. In these papers, inductance effects are also not included. In upper metal layers, wide interconnects are frequently used, which have low resistance, making inductance effects non-negligible in high speed circuits.

With on-chip signal frequencies continuously increasing, bandwidth requirement also needs to be considered during the repeater insertion procedure. In this chapter, a new repeater insertion methodology is proposed for achieving the minimum power while satisfying delay and bandwidth constraints. The rest of this chapter is organized as follows. In Section 6.2, the timing and power models of RC interconnects are reviewed. Based on these models, analytic methods are presented for achieving the minimum power while satisfying delay and bandwidth constraints. In Section 6.3, the effects of inductance on this repeater design methodology are analyzed. Finally, some conclusions are offered in Section 6.4.

6.2 Power Dissipation in an RC Interconnect with Delay and Bandwidth Constraints

By including the effects of the input transition time, a new timing model of an RC interconnect with repeaters is presented in Subsection 6.2.1. The three primary power dissipation sources in interconnects are reviewed in Subsection 6.2.2. Given a delay or a bandwidth constraint, a design space for a repeater system can be determined. The minimum achievable power in the design space is described in Subsections 6.2.3 and 6.2.4 for delay and bandwidth constraints, respectively. Multiple constraints are analyzed in Subsection 6.2.5.

6.2.1 Delay and Transition Time Model of RC Interconnects

As shown in Fig. 6.1, a distributed RC interconnect is evenly divided into k segments by repeaters. R_t and C_t are the total resistance and capacitance, respectively, of the interconnect. The repeaters are h times as large as a minimum sized repeater, with the output resistance R_{tr0}/h , output capacitance hC_{d0} , and input capacitance hC_{g0} , where R_{tr0} , C_{d0} , and C_{g0} are the output resistance, output capacitance, and input capacitance, respectively, of a minimum sized repeater.

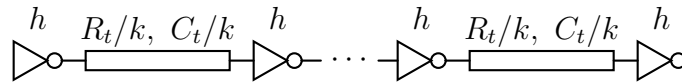


Figure 6.1: Repeater insertion in a long RC interconnect line.

Table 6.1: Device parameters of BPTM 45 nm model. $V_{dd} = 1.1$ volts.

Device	Temp (°C)	I_{dsat}/W ($\mu\text{A}/\mu\text{m}$)	I_{sub}/W ($\text{nA}/\mu\text{m}$)	$ V_t $ (volts)	$ V_{dsat} $ (volts)	α
NMOS	25	1064	42.7	0.330	0.468	0.91
NMOS	100	1035	712.0	0.257	0.516	0.88
PMOS	25	534	25.3	0.325	0.600	1.05
PMOS	100	514	374.1	0.243	0.672	0.97

The repeater is assumed, in this chapter, to be implemented as a CMOS inverter. The inverter is also assumed to be symmetric such that the effective output resistance is the same for both rising and falling signal transitions. The Berkeley predictive technology model (BPTM) [148, 149] for a 45 nm printed channel length is used, corresponding to the 80 nm technology node described in the ITRS [7]. Some model parameters are modified to capture the trends of the saturated drain current and subthreshold current predicted by the ITRS. The device parameters used in this chapter are listed in Table 6.1, where α is the velocity saturation index and is determined with the method described in [150]. The PMOS transistor is 2.2 times as large as the NMOS transistor in the inverter, and the minimum gate width is assumed to be 45 nm.

C_{g0} and C_{d0} can be obtained with SPICE by measuring the charge stored on the input and output of the minimum sized repeater during signal transitions. In this

chapter, $C_{g0} = 0.455$ fF and $C_{d0} = 0.413$ fF. R_{tr0} can be approximated as

$$R_{tr0} = K \frac{V_{dd}}{I_{dn0}}, \quad (6.1)$$

where K is a fitting parameter, and I_{dn0} is the saturated drain current of a minimum sized NMOS transistor with both V_{gs} and V_{ds} equal to V_{dd} . K can be determined by matching the 50% delay or transition time of the step response of an RC equivalent circuit to SPICE simulations. Note that the K obtained by matching the 50% delay and the K obtained by matching the transition time are different and are denoted as K_d and K_r , respectively. In this chapter, K_d is 0.78 and K_r is 0.55. The corresponding output resistances are R_{d0} and R_{r0} .

The delay t_{ds} and transition time t_{rs} of a single interconnect stage for a step input can be obtained from [58],

$$t_{ds} = 0.377 \frac{R_t C_t}{k^2} + 0.693 \left(R_{d0} C_0 + \frac{R_{d0} C_t}{hk} + \frac{R_t C_{g0} h}{k} \right), \quad (6.2)$$

$$t_{rs} = \frac{t_{90\%} - t_{10\%}}{0.8} = 1.1 \frac{R_t C_t}{k^2} + 2.75 \left(R_{r0} C_0 + \frac{R_{r0} C_t}{hk} + \frac{R_t C_{g0} h}{k} \right), \quad (6.3)$$

where $C_0 = C_{d0} + C_{g0}$. With a finite input slew rate, both the repeater delay [150] and repeater output transition time [151] depend linearly on the input transition time t_{r_in} . The contribution of t_{r_in} to the repeater delay can be represented by γt_{r_in} . For

a rising input, γ is determined as [150]

$$\gamma_r = \frac{1}{2} - \frac{1 - v_{tn}}{1 + \alpha_n}, \quad (6.4)$$

where $v_{tn} = V_{tn}/V_{dd}$. By changing the suffix n to p in (6.4), the coefficient γ_f for a falling input can be obtained. An average of γ_r and γ_f is used as γ in the rest of this chapter to determine the interconnect delay.

The linear dependence of the repeater output transition time on t_{r_in} is only valid for slow input signals with small load capacitances [151]. Furthermore, the signal is degraded by the interconnect impedance before reaching the far end, decreasing the sensitivity of the far end transition time to the input slew rate. The effect of the input slew rate on the far end transition time is, therefore, ignored in this chapter. The signal transition times determine the highest switching speed an on-chip signal can achieve, *i.e.*, the bandwidth of the circuit. The total delay of the interconnect is

$$T_{total} = k(t_{ds} + \gamma t_{rs}) = a_1 \frac{R_t C_t}{k} + a_2 (R_0 C_0 k + \frac{R_0 C_t}{h} + R_t C_{g0} h), \quad (6.5)$$

where

$$a_1 = 0.377 + 1.1\gamma, \quad (6.6)$$

$$a_2 = 0.693 + 2.75\gamma, \quad (6.7)$$

$$R_0 = \frac{0.693R_{d0} + 2.75\gamma R_{r0}}{a_2}. \quad (6.8)$$

The total delay obtained from (6.5) as well as the model neglecting input transition time effects are compared with SPICE in Fig. 6.2 for different numbers of repeaters.

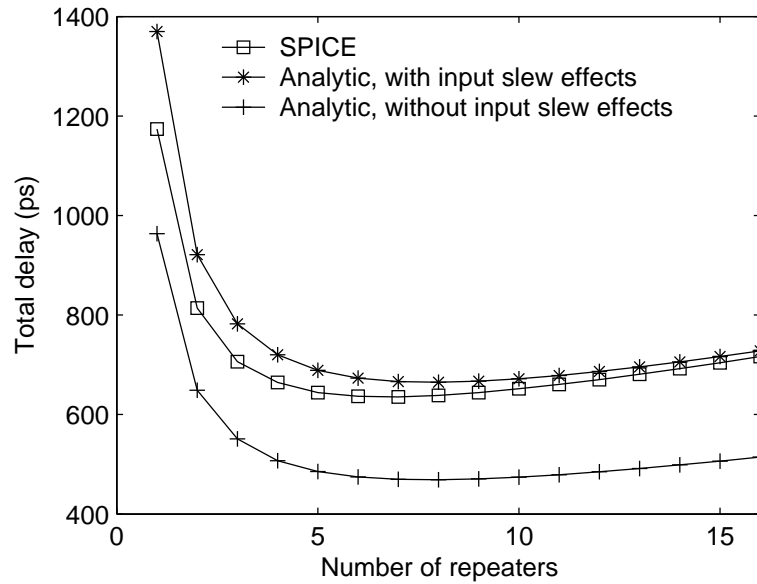


Figure 6.2: Total delay for an RC interconnect driven by repeaters. $R = 0.31 \Omega/\mu\text{m}$, $C = 0.223 \text{ fF}/\mu\text{m}$, $l = 5 \text{ mm}$, and $h = 50$.

In the SPICE simulation, the first driver and the far end load are assumed to be implemented by the same sized repeater. The input signal slew of the first driver is assumed to be the same as the signal slew at the end of the interconnect. The

interconnect parameters are extracted for a minimum sized global interconnect at the 80 nm technology node [7].

As shown in Fig. 6.2, neglecting the effects of the input transition time significantly underestimates the total delay. When the number of repeaters is small, each repeater drives a long interconnect. The signal transition time at the input of each repeater (the output of the previous stage) is sufficiently large; therefore, the assumption made in [150] ($t_{r_in} < 3t_{r_out}$) is no longer valid. The gate delay does not increase linearly with the input transition time and (6.5) becomes less accurate. This situation will normally not occur in practical circuits due to the slow transition time. In this example, when more than four repeaters are inserted, the error of (6.5) is within 7% of SPICE.

By setting $\partial T_{total}/\partial k$ and $\partial T_{total}/\partial h$ to zero, the optimal k and h to minimize T_{total} can be obtained,

$$k_{opt} = \sqrt{\frac{a_1 R_t C_t}{a_2 R_0 C_0}}, \quad (6.9)$$

$$h_{opt} = \sqrt{\frac{R_0 C_t}{R_t C_{g0}}}. \quad (6.10)$$

The corresponding minimum delay is

$$T_{min} = 2\sqrt{a_1 a_2 R_t C_t R_0 C_0} \left(1 + \sqrt{\frac{a_2 C_{g0}}{a_1 C_0}} \right). \quad (6.11)$$

This delay-minimal repeater design methodology is not necessarily an appropriate strategy in practical circuits. First, the delay is not sensitive to the size of the repeaters near the optimal point, therefore, significant power and area are wasted to achieve only a small improvement in speed when approaching the optimal point (for minimum delay). Second, with increasing on-chip signal frequencies, it is possible that a delay-minimal design methodology will not satisfy the specific bandwidth requirement.

6.2.2 Power Dissipation Components in Interconnects with Repeaters

Power dissipation is a primary criterion in VLSI circuits due to high integration densities and high speeds. There are three significant power dissipation mechanisms in digital CMOS circuits: dynamic power, short-circuit power, and leakage power.

a) Dynamic power

Dynamic power is the power consumption due to charging and discharging the load capacitance. Dynamic power has been well studied and is characterized by the following well known expression,

$$P_d = \alpha_s f C_L V_{dd}^2, \quad (6.12)$$

where f is the clock frequency and α_s is the switching factor (assumed here as 0.15 [147]). For an RC interconnect with repeaters, the load capacitance C_L includes both the interconnect capacitance and the parasitic capacitance of the repeaters. The total dynamic power in a RC line with repeaters is

$$P_d = \alpha_s f C_L V_{dd}^2 = \alpha_s f (C_t + khC_0) V_{dd}^2. \quad (6.13)$$

b) Short-circuit power

If the signal applied at the input of a CMOS inverter has a finite slew rate, a direct current path exists between V_{dd} and ground when the input signal switches between V_{tn} and $V_{dd} + V_{tp}$. The power consumed in this way is called short-circuit power [127]. The short-circuit power is a function of the input transition time, output load capacitance, and the size of the transistor. A closed form model of the short-circuit power [129] is adopted here because the model provides a clear relationship between the short-circuit power and related circuit parameters. From this model, the short-circuit energy dissipated in a CMOS inverter during a full signal switch ($0 \rightarrow 1 \rightarrow 0$) can be approximated as

$$E_s = \frac{4I_{dsat}^2 t_r^2 V_{dd}}{V_{dsat} G C_{out} + 2H I_{dsat} t_r}. \quad (6.14)$$

In this expression, C_{out} is the output load capacitance. $V_{dsat} = (V_{dsatn} + V_{dsatp})/2$ and $I_{dsat} = (I_{dsatn} + I_{dsatp})/2$. $G = (G_r + G_f)/2$ and $H = (H_r + H_f)/2$. G_r and H_r are the coefficients associated with rising inputs, and can be determined as [129]

$$G_r = \frac{(\alpha_n + 1)(1 - v_{tn})^{\alpha_n}(1 - v_{tp})^{\alpha_p/2}}{F_r(1 - v_{tn} - v_{tp})^{\alpha_p/2 + \alpha_n + 2}}, \quad (6.15)$$

$$H_r = \frac{2^{\alpha_p}(\alpha_p + 1)(1 - v_{tp})^{\alpha_p}}{(1 - v_{tn} - v_{tp})^{\alpha_p + 1}}, \quad (6.16)$$

where

$$F_r = \frac{1}{\alpha_n + 2} - \frac{\alpha_p}{2(\alpha_n + 3)} + \frac{\alpha_p(\alpha_p/2 - 1)}{\alpha_n + 4}. \quad (6.17)$$

G_f and H_f can be obtained by exchanging n and p suffixes in (6.15)-(6.17). The total capacitance in a single interconnect stage includes the output parasitic capacitance of the repeater, the interconnect capacitance, and the input capacitance of the following repeater,

$$C_{stage} = C_0h + \frac{C_t}{k}. \quad (6.18)$$

Due to the shielding effect of the interconnect resistance, the load capacitance seen by the repeater is less than C_{stage} during an input signal transition. An effective capacitance C_{eff} , therefore, is determined with the method described in Chapter 4 to estimate the short-circuit power. The total short-circuit power of the inserted

repeaters, therefore, is

$$P_s = \frac{4\alpha_s f I_{d0}^2 t_r^2 V_{dd} k h^2}{V_{dsat} G C_{eff} + 2H I_{d0} t_r h}, \quad (6.19)$$

where I_{d0} is the average saturated drain current of the NMOS and PMOS transistors in a minimum sized repeater, and C_{eff} is the effective capacitance of each interconnect stage.

c) Leakage power

In deep submicrometer CMOS technologies, the dominant leakage current source is composed of subthreshold current and gate leakage current [152]. The total leakage power dissipated in the repeaters is

$$P_l = h k V_{dd} (I_{sub0} + I_{g0}), \quad (6.20)$$

where I_{sub0} is the average subthreshold current of the NMOS and PMOS transistors in a minimum sized repeater. I_{g0} is the average gate leakage current of a minimum sized repeater with low and high inputs. The leakage power is expected to dominate dynamic power and short-circuit power in future technologies. Since the subthreshold current increases rapidly with increasing temperature, a worst case temperature of 100 °C is assumed in this chapter to emphasize the leakage power. In this case, $I_{sub0} = 34.5 \text{ nA}$ and $I_{g0} = 1.4 \text{ nA}$.

6.2.3 Power Dissipation with Delay Constraints

For a delay constraint T_{req} greater than T_{min} , the design space of the repeaters is the area inside the closed curves shown in Fig. 6.3.

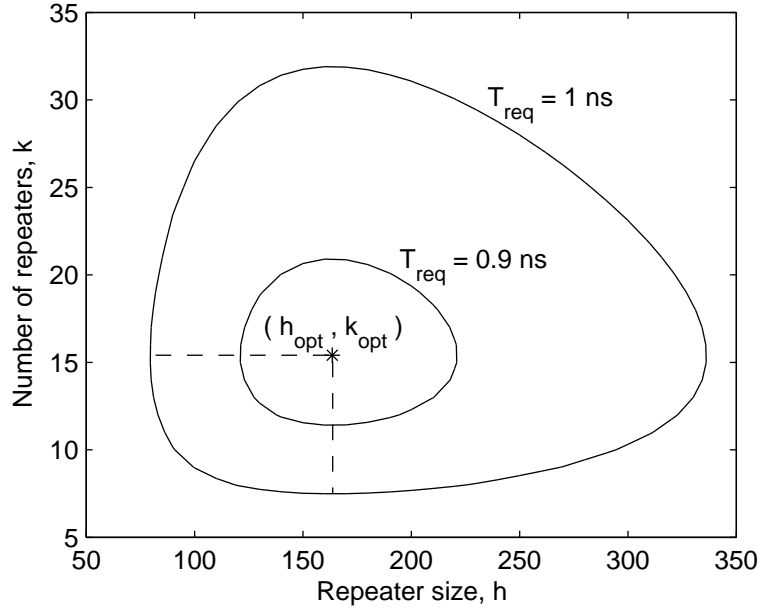


Figure 6.3: Repeater design space with delay constraint. $R = 0.31 \Omega/\mu\text{m}$, $C = 0.223 \text{ fF}/\mu\text{m}$, and $l = 10 \text{ mm}$.

An expression characterizing the edge of the design space is

$$T_{req} = a_1 \frac{R_t C_t}{k} + a_2 \left(R_0 C_0 k + \frac{R_0 C_t}{h} + R_t C_{g0} h \right). \quad (6.21)$$

With T_{req} approaching T_{min} , the design space converges to the minimum delay point (h_{opt}, k_{opt}) . The minimum h that can satisfy the delay requirement occurs when $k = k_{opt}$, and the minimum k that can satisfy the delay requirement occurs when $h = h_{opt}$.

The total power dissipated by an RC interconnect with repeaters is the summation of the three primary power dissipation components,

$$P_{total} = P_d + P_s + P_l. \quad (6.22)$$

In Fig. 6.4, P_{total} is plotted as a function of k and h . For each h , an optimal k exists to achieve the minimum power. If k is too small, the signal transition time will be large, and the total power is dominated by the short-circuit power. If k is too large, the total power is dominated by the dynamic power and leakage power. P_d and P_l increase linearly with increasing h for a fixed k . P_s , however, is more complicated.

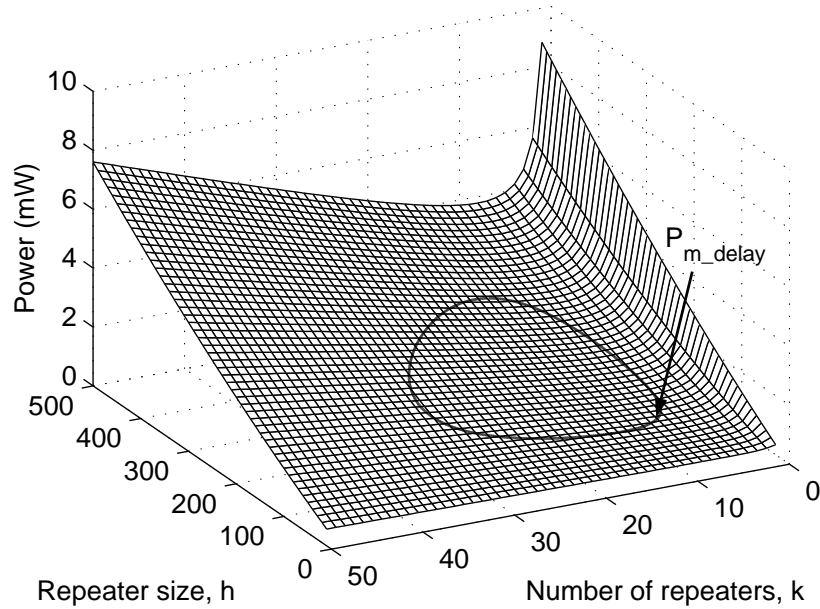


Figure 6.4: Total power dissipation in an interconnect with repeaters as a function of h and k . $f = 1$ GHz, $R = 0.31 \Omega/\mu\text{m}$, $C = 0.223$ fF/ μm , and $l = 10$ mm.

In order to obtain an analytic solution, some approximations are made to the C_{eff} in (6.19). The effective capacitance C_{eff} in one stage is a function of h and k . The ratio between C_{eff} and C_{stage} is plotted in Fig. 6.5. In most cases, this ratio is varied in the range from 0.5 to 1. An average ratio of 0.75 is used in (6.19) to evaluate the short-circuit power. With this approximation, $\partial P_s / \partial h$ is always positive, which means that P_s increases monotonically with increasing h for a fixed k . The total power P_{total} , therefore, also increases monotonically with increasing h for a fixed k . This behavior is illustrated in Fig. 6.4. For the design space determined by the delay constraint as shown in Fig. 6.3, the minimum power can only be reached on the left edge of the design space.

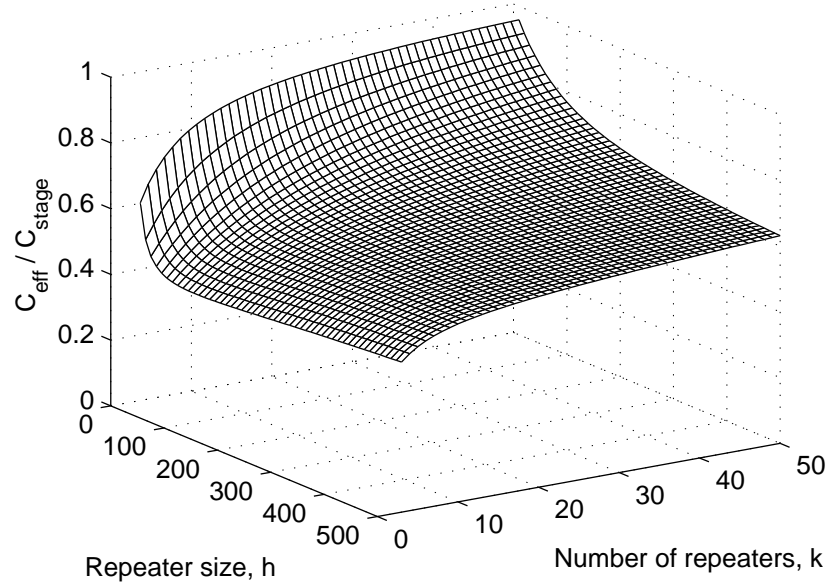


Figure 6.5: The ratio of C_{eff} to C_{stage} . $R = 0.31 \Omega/\mu\text{m}$, $C = 0.223 \text{ fF}/\mu\text{m}$, and $l = 10 \text{ mm}$.

The power dissipation at the edge of the design space is plotted as a function of h in Fig. 6.6(a). The dynamic and leakage power is plotted together since both of these power components depend linearly on kh . The minimum total power with delay constraints P_{m_delay} can be obtained by solving $dP_{total}/dh = 0$. Note that at the edge of the design space, k is a function of h .

In order to provide a closed form solution for P_{m_delay} , the curve of $P_d + P_l$ around the power-optimal point is approximated by a part of an ellipse, as shown in Fig. 6.6(b). The optimal design parameters (h_0, k_0) for minimizing $P_d + P_l$ with a delay constraint T_{req} can be solved by the Lagrange method [144],

$$k_0 = \frac{-b - \sqrt{b^2 - T_{req}^2 a_1 a_2 R_0 C_0 R_t C_t}}{T_{req} a_2 R_0 C_0}, \quad (6.23)$$

$$h_0 = \frac{T_{req} k_0 - 2a_1 R_t C_t}{2a_2 R_t C_{g0} k_0}, \quad (6.24)$$

$$b = a_2 R_0 R_t C_t (a_2 C_{g0} - a_1 C_0) - \frac{T_{req}^2}{4}. \quad (6.25)$$

In Fig. 6.6(b), h_1 is the minimum repeater size that can satisfy a target delay constraint, which can be obtained by inserting $k_1 = k_{opt}$ into (6.21). P_0 and P_1 are the corresponding values of $P_d + P_l$ at (h_0, k_0) and (h_1, k_1) , respectively. The curve of P_s is approximated by a linear function. With these approximations, the power-optimal

repeater size h_p with a delay constraint is

$$h_p = h_0 - \frac{x_0(h_0 - h_1)^2}{\sqrt{x_0^2(h_0 - h_1)^2 + (P_1 - P_0)^2}}, \quad (6.26)$$

where

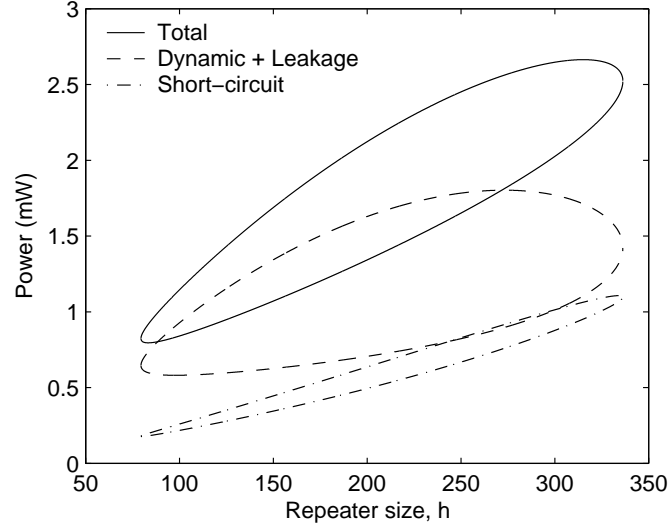
$$x_0 = \frac{4\alpha_s f I_{d0}^2 V_{dd} k_0^2 t_{r0}^2 [1.5V_{dsat}G(C_0 h_0 k_0 + C_t)h_0 + 2H I_{d0} k_0 t_{r0} h_0^2]}{[0.75V_{dsat}G(C_0 h_0 k_0 + C_t) + 2H I_{d0} k_0 t_{r0} h_0]^2}, \quad (6.27)$$

$$t_{r0} = t_r(h_0, k_0). \quad (6.28)$$

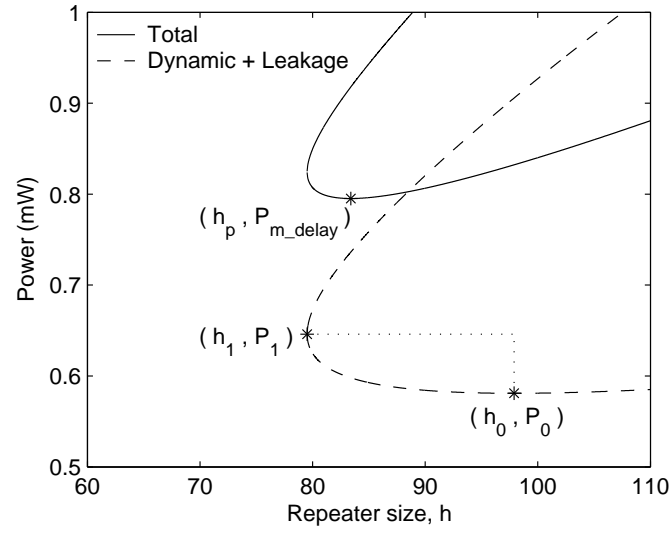
A detailed derivation is provided in Appendix A. The corresponding k_p can be solved by inserting h_p into (6.21). Upon obtaining h_p and k_p , P_{m_delay} can be obtained directly from (6.22). If k_p is not an integer, the nearest two integers are used to determine the minimum power (h_p will need to be re-calculated).

For different interconnect loads and delay constraints, results from the proposed method are compared with SPICE simulations as listed in Table 6.2. The average error of the analytically obtained minimum power is 7%. In these experiments, the total power does not include the power consumed by the load buffer.

In Table 6.3, different power components from the analytic model are listed separately for delay-optimal circuits and power-optimal circuits with delay constraints. The dynamic power listed in the table is only due to the parasitic capacitance of the repeaters, since the dynamic power due to the interconnect capacitance is a constant



(a)



(b)

Figure 6.6: Power dissipation with constant delay. $f = 1 \text{ GHz}$, $T_{req} = 1 \text{ ns}$, $R = 0.31 \Omega/\mu\text{m}$, $C = 0.223 \text{ fF}/\mu\text{m}$, and $l = 10 \text{ mm}$.

Table 6.2: Minimum power with delay constraints obtained analytically as compared with SPICE simulations. $f = 1$ GHz.

R_t (k Ω)	C_t (pF)	T_{req} (ps)	SPICE			Analytic			
			k_p	h_p	$P_{m, delay}$ (μ W)	k_p	h_p	$P_{m, delay}$	
								(μ W)	% Error
1	1	400	4	90	315.3	4	88.9	335.2	6.3
1	1	500	3	59	267.8	4	54.6	283.0	5.7
2	2	800	8	90	624.5	9	85.1	669.7	7.2
2	2	900	7	69	558.5	8	67.1	602.8	7.9
2	2	1000	7	56.5	528.3	7	56.7	565.8	7.1
3	1	700	7	47.5	300.6	7	49.7	331.1	10.1
3	1	800	7	36	274.9	7	36.6	296.2	7.7
3	1	900	6	30.5	260.7	6	30.6	277.5	6.4
2	3	1000	9	112	935.2	10	102.2	982.3	5.0
2	3	1200	9	71.5	801.6	9	71.1	857.4	7.0
2	3	1400	9	55	753.7	8	55.9	799.2	6.0

for a specific interconnect. As compared with the power dissipation in a delay-optimal circuit, significant power savings is achieved by adopting a power-optimal circuit with delay constraints. For a power-optimal circuit, all of the three power components decrease with increasing delay targets. Note that for a delay-optimal circuit, the short-circuit power is slightly less than the dynamic power of the repeaters. For a power-optimal circuit with delay constraints, the short-circuit power, however, can be greater than the dynamic power of the repeaters. The short-circuit power grows in significance with increasing delay targets as compared with other power components. The leakage power is less than a quarter of the repeater dynamic power for the examples listed in Table 6.3.

Table 6.3: Different power components dissipated in the repeaters. $f = 1$ GHz.

R_t (k Ω)	C_t (pF)	T_{req} (ps)	Delay-optimal			Power-optimal		
			$P_{d,rep}$ (μ W)	P_s (μ W)	P_l (μ W)	$P_{d,rep}$ (μ W)	P_s (μ W)	P_l (μ W)
1	1	400	182.2	171.7	45.7	56.0	73.8	13.6
1	1	500	182.2	171.7	45.7	34.4	50.8	8.4
2	2	800	364.3	343.3	91.3	111.9	147.5	27.2
2	2	900	364.3	343.3	91.3	84.5	117.2	20.6
2	2	1000	364.3	343.3	91.3	62.5	108.8	15.2
3	1	700	175.3	172.9	43.9	54.7	71.4	13.2
3	1	800	175.3	172.9	43.9	40.3	56.1	9.8
3	1	900	175.3	172.9	43.9	28.9	52.2	7.0
2	3	1000	520.6	519.6	130.5	160.9	208.0	39.1
2	3	1200	520.6	519.6	130.5	100.8	163.1	24.5
2	3	1400	520.6	519.6	130.5	70.4	145.5	17.1

The effect of switching factor α_s on the solution of the power-optimal design is illustrated in Fig. 6.7. The curve is step like since the corresponding k_p is an integer. As shown in Fig. 6.7, under a delay constraint, h_p decreases with increasing α_s . At the limiting case, $\alpha_s = 0$, only leakage power exists, and the optimal repeater size is $h_p = h_0$.

As shown in Fig. 6.6, the short-circuit power at the edge of the design space increases with increasing repeater size, while dynamic power and leakage power decrease with increasing repeater size around the power optimal solution. A larger repeater size with fewer number of repeaters is, therefore, preferable for dynamic and leakage power dominant cases, and a smaller repeater size with a greater number of repeaters

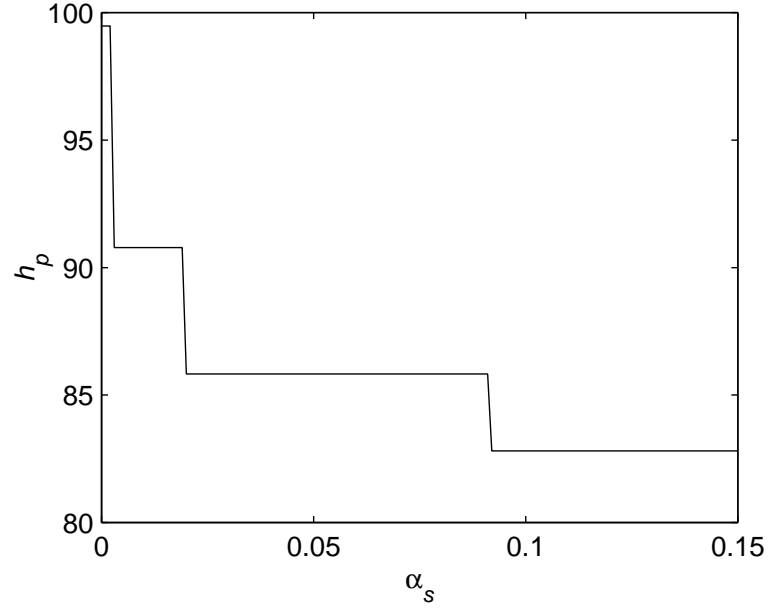


Figure 6.7: The effect of α_s on the optimal repeater size h_p . $R = 0.31 \Omega/\mu\text{m}$, $C = 0.223 \text{ fF}/\mu\text{m}$, $l = 10 \text{ mm}$, $f = 1 \text{ GHz}$, and $T_{req} = 1 \text{ ns}$.

is preferable for short-circuit power dominant cases. For circuits with low power supplies, the threshold voltage is normally reduced to maintain performance. The leakage power, therefore, is a more dominant component of the total power consumption. In this case, the optimal repeater size h_p is closer to h_0 .

6.2.4 Power Dissipation with Bandwidth Constraints

The bandwidth of an interconnect is assumed in this chapter to be limited solely by the output signal transition time. Faster signal transition times support a shorter signal bit period, therefore, a higher bandwidth. For a bandwidth constraint B_{req} , the signal transition time is assumed to be less than or equal to half the bit period,

i.e., $t_r \leq 1/2B_{req}$. The design space for different bandwidth constraints is shown in

Fig. 6.8. The design space is the area in the upper-right side of the curve.

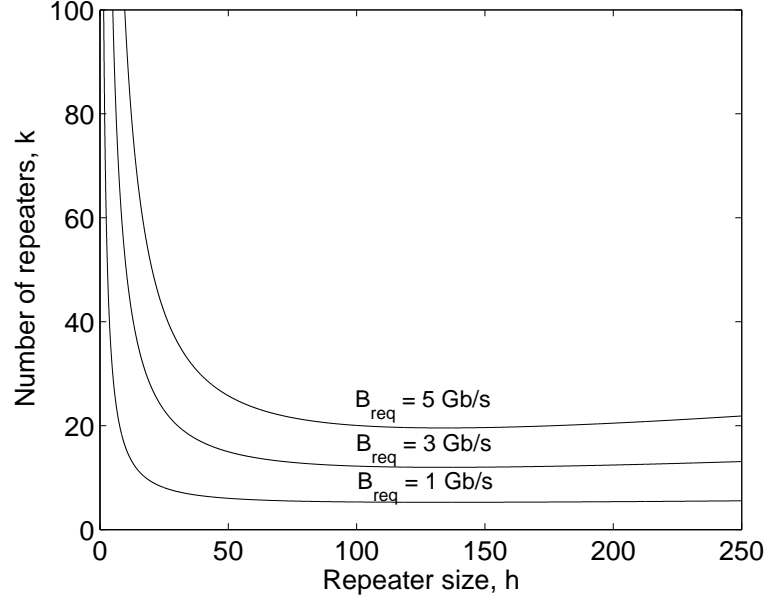


Figure 6.8: Repeater design space with bandwidth constraints. $R = 0.31 \Omega/\mu\text{m}$, $C = 0.223 \text{ fF}/\mu\text{m}$, and $l = 10 \text{ mm}$.

An expression for the design space edge is

$$t_r = \frac{1}{2B_{req}}. \quad (6.29)$$

From (6.29), k can be solved as a function of h ,

$$k(h) = \frac{\sqrt{\tau_2^2 - 4.4\tau_1 R_t C_t} + \tau_2}{-2\tau_1}, \quad (6.30)$$

where

$$\tau_1 = 2.75R_{r0}C_0 - \frac{1}{2B_{req}}, \quad (6.31)$$

$$\tau_2 = 2.75\left(\frac{R_{r0}C_t}{h} + R_tC_{g0}h\right). \quad (6.32)$$

In order for k to be a positive real number, τ_1 should be negative. An upper limit, therefore, is placed on the bandwidth by the process technology,

$$B_{req} \leq \frac{1}{5.5R_{r0}C_0}. \quad (6.33)$$

Similar to the delay-constraint case, the minimum power with a bandwidth constraint can only be reached at the edge of the design space. P_s in (6.19) can be rewritten as

$$P_s = \frac{4\alpha_s f I_{d0}^2 t_r^2 V_{dd} kh}{0.75V_{dsat}G(C_0 + \frac{C_t}{kh}) + 2HI_{d0}t_r}. \quad (6.34)$$

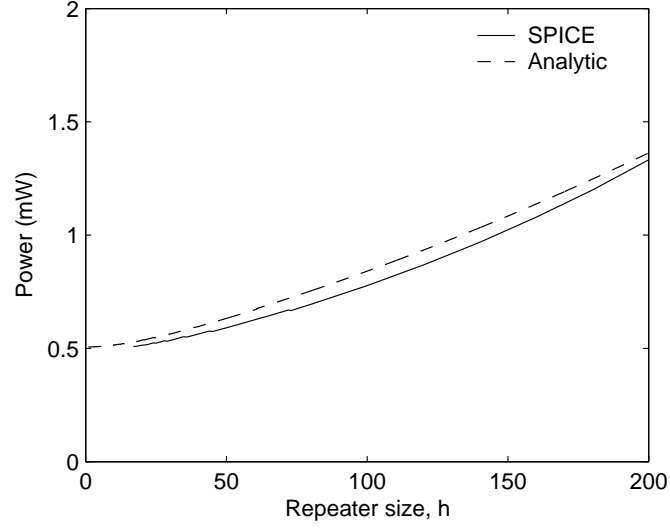
For a fixed t_r , P_s increases monotonically with increasing kh . This relationship is also valid for P_d and P_l . At the edge of the design space, $t_r = 1/2B_{req}$; therefore, kh can be obtained from (6.30),

$$kh = \frac{\sqrt{(\tau_2 h)^2 - 4.4\tau_1 R_t C_t h^2} + \tau_2 h}{-2\tau_1}. \quad (6.35)$$

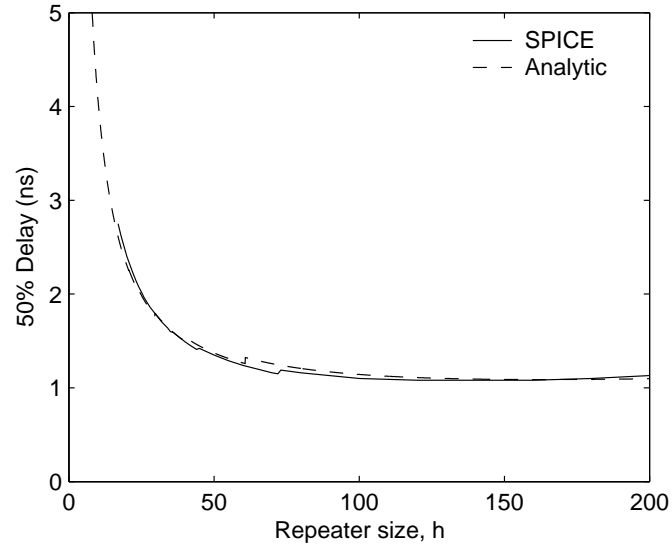
From (6.35), kh increases monotonically with h (note that τ_1 is negative). The total power at the edge of the design space, therefore, increases monotonically with increasing h , as shown in Fig. 6.9(a). The minimum power satisfying the bandwidth constraint can be achieved with minimum sized repeaters. For minimum sized repeaters, the corresponding k and total delay, however, are unpractically large as shown in Figs. 6.8 and 6.9(b). In order to produce an effective repeater system, delay and area constraints should also be considered.

6.2.5 Power Dissipation with both Delay and Bandwidth Constraints

The design space under both delay and bandwidth constraints is the intersection of the design spaces described in Sections 6.2.3 and 6.2.4, as shown in Fig. 6.10(a). The minimum power is also achieved at the edge of the design space. As described in Section 6.2.3, the minimum power satisfying the delay constraint occurs at (h_p, k_p) . If these design parameters satisfy the bandwidth requirement, (h_p, k_p) is the optimal design point for minimizing power while satisfying both the delay and bandwidth constraints. If (h_p, k_p) can not satisfy the bandwidth requirement, the minimum power occurs at the left intersection of the two design space edges, as shown in Figs. 6.10(a) and 6.10(b). The coordinates of the intersection are obtained by solving (6.21) and (6.29). If no intersection exists between the two design spaces, the two



(a) Power dissipation



(b) 50% delay

Figure 6.9: Power dissipation and 50% delay at the edge of the design space with bandwidth constraint. $B_{req} = 1 \text{ Gb/s}$, $R = 0.31 \text{ } \Omega/\mu\text{m}$, $C = 0.223 \text{ fF}/\mu\text{m}$, and $l = 10 \text{ mm}$.

constraints cannot be simultaneously satisfied, and one or both of the constraints have to be released. Other constraints, such as the number and size of the repeaters, can be handled similarly.

6.3 Effects of Inductance on the Repeater Insertion

Methodology

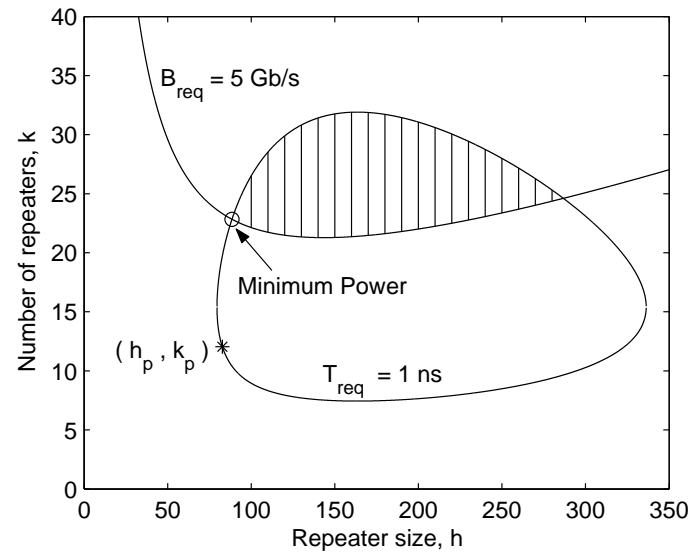
For wide global interconnects, the inductance is not negligible and has to be considered in repeater design methodologies. In Section 6.3.1, a timing model of an *RLC* interconnect is reviewed. In Section 6.3.2, the effects of inductance on the repeater design space are analyzed. The minimum power consumption while satisfying delay and bandwidth constraints is described in Section 6.3.3.

6.3.1 Timing Model of *RLC* Interconnects

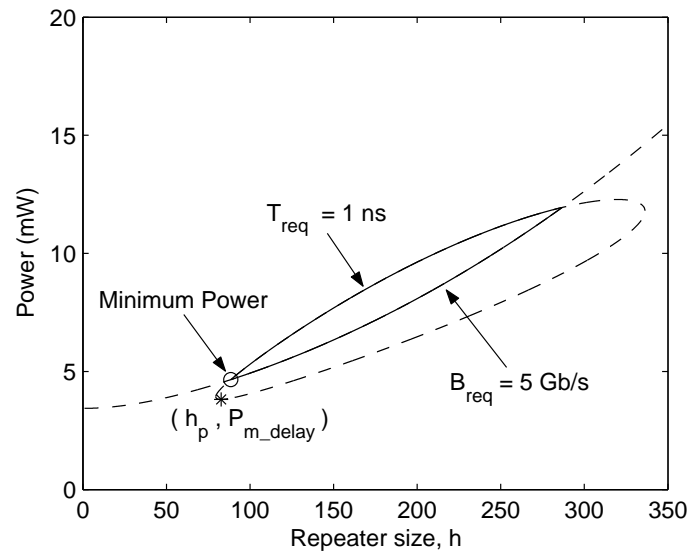
In [59], a variable ζ is introduced to characterize the effects of inductance. By including the repeater output capacitance, ζ becomes

$$\zeta = \frac{Rl}{2k} \sqrt{\frac{C}{L}} \cdot \frac{R_T C_T (1 + \frac{C_{d0}}{C_{g0}}) + C_T + R_T + 0.5}{\sqrt{1 + C_T}}, \quad (6.36)$$

where $R_T = kR_{tr0}/(hRl)$ and $C_T = hkC_{g0}/(Cl)$. The corresponding ζ with R_{d0} and R_{r0} is denoted as ζ_d and ζ_r , respectively. The delay model of an *RLC* interconnect is



(a) Design space



(b) Power dissipation

Figure 6.10: The design space and power dissipation at the edge of the design space with both delay and bandwidth constraints.

an extension of the result from [59] where the repeater output capacitance and input slew effects are included. The delay of a single stage interconnect for a step input can be obtained by curve fitting,

$$t_{ds} = \frac{e^{-2.3\zeta_d^{1.5}} + 1.48\zeta_d}{w_n}, \quad (6.37)$$

where $w_n = k/\sqrt{Ll(Cl + C_{g0}hk)}$. The coefficients in (6.37) are slightly different from those in [59] due to the effects of the repeater output capacitance. In [153], an accurate estimate of the rise time in an RLC interconnect is also obtained by curve fitting. The expressions, however, are analytically complicated. In this chapter, a simplified piecewise approximation of the rise time is used,

$$t_r = \frac{t_{90\%} - t_{10\%}}{0.8} = \begin{cases} \frac{4.4\zeta_r - 1.8}{0.8w_n} & \zeta_r > 0.41, \\ 0 & \text{otherwise.} \end{cases} \quad (6.38)$$

When $\zeta_r < 0.5$, the interconnect is highly inductance dominant, and (6.38) can introduce a large error. In Fig. 6.11, ζ_r is plotted for different repeater sizes and interconnect lengths. The driver size is normalized to the size of a minimum inverter. The size of the load gate is the same as the driver. W_{min} is the minimum wire width specified in the ITRS [7]. The space between adjacent interconnects is assumed equal to the interconnect width. As shown in Fig. 6.11, inductance effects become more

significant with larger drivers. Note that for a fixed driver size, a minimum ζ_r can be achieved in this example when the interconnect length is approximately 1 mm. When the wire length is too short or too long, the interconnect is dominated either by the repeater resistance or the wire resistance, respectively. In Fig. 6.12, the inductance per unit length is plotted as a function of the space between the signal line and the current return path. The wire thickness is $0.4\ \mu\text{m}$. Three wire widths are examined, $0.36\ \mu\text{m}$, $1.8\ \mu\text{m}$, and $9\ \mu\text{m}$. The width of the reference line for the current return path is assumed to be the same as the signal line width. The inductance values are obtained with FastHenry [44] for a wire length of 10 mm. As shown in Fig. 6.12, the interconnect inductance increases slowly with increasing space between the signal line and the current return path. With the same line space, wider wires exhibit smaller inductance. When the return paths are within $10\ \mu\text{m}$, the inductance ranges from $0.5\ \text{pH}/\mu\text{m}$ to $1.5\ \text{pH}/\mu\text{m}$.

6.3.2 Effects of Inductance on the Repeater Design Space

By including interconnect inductance, both the delay and signal transition time of an interconnect are affected. The repeater design space satisfying delay or bandwidth constraints is also changed, which is described in the following two subsections, respectively.

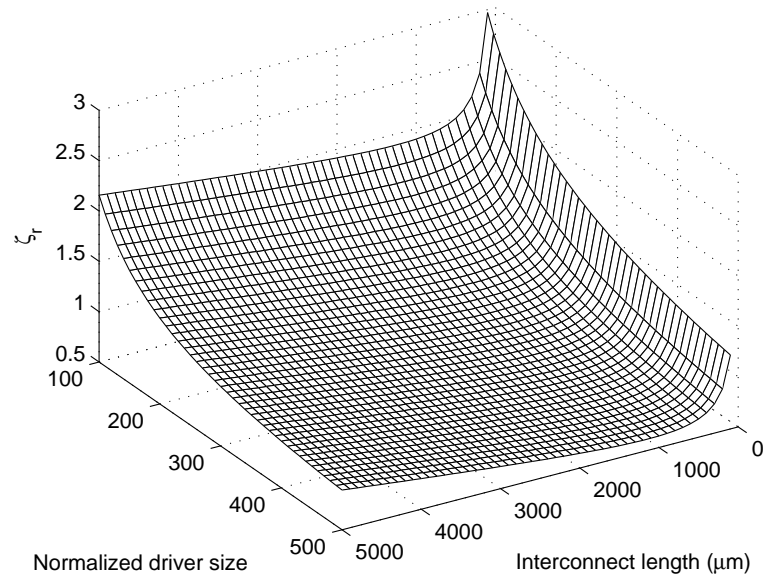


Figure 6.11: Inductance effect for different driver size and interconnect length. $W = 20W_{min}$ and $L = 1 \text{ pH}/\mu\text{m}$.

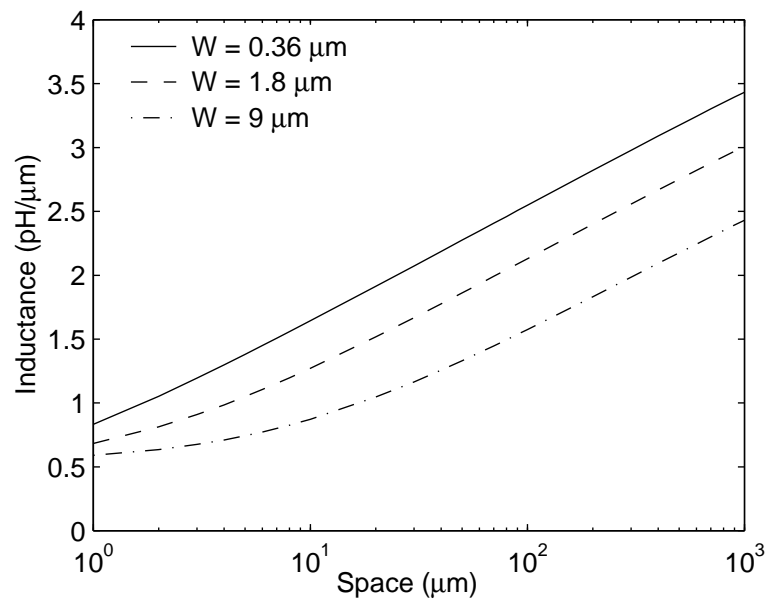


Figure 6.12: Inductance values with difference current return paths.

a) Bandwidth constraints

The signal transition time at the far end of an RLC interconnect decreases with increasing inductance effects [146]. The inductance, therefore, increases the bandwidth of an interconnect. The repeater design space satisfying a bandwidth constraint is plotted in Fig. 6.13 for different values of inductance. With increasing inductance, the number and size of the repeaters can be reduced while maintaining the same signal transition time.

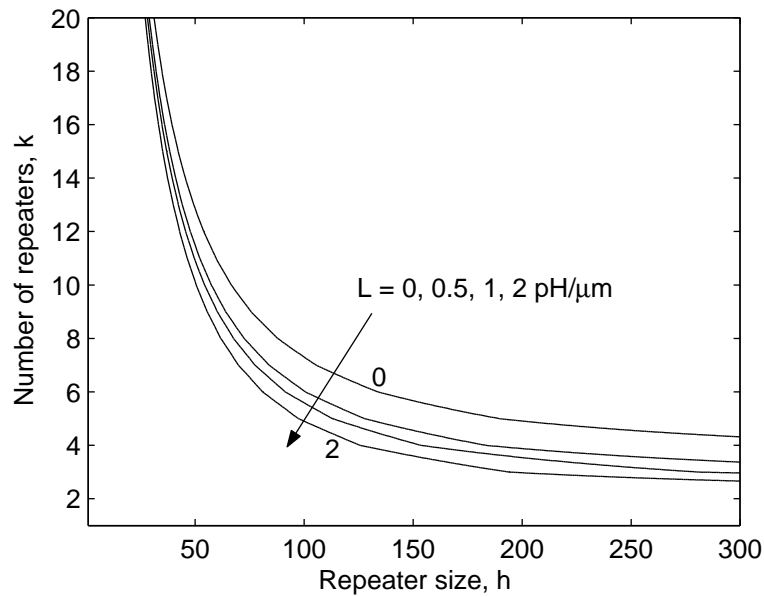


Figure 6.13: Effects of inductance on the repeater design space satisfying bandwidth constraints. $B_{req} = 2 \text{ Gb/s}$, $l = 10 \text{ mm}$, and $W = 10W_{min}$.

b) Delay constraints

The delay of an interconnect with repeaters can be affected by inductance in three ways. First, the propagation delay along the interconnect can increase with increasing inductance [136]. Second, the inductance reduces the signal transition time, decreasing the gate delay due to the input slew effect. Third, due to the inductive shielding effect (described by El-Moursy and Friedman in [136]), both the effective capacitance seen by the driver and the equivalent output resistance of the driver are reduced. The gate delay is, therefore, further reduced. (Since the delay model used in this chapter is based on curve fitting, and a constant driver resistance is assumed, the third inductance effect is not considered in this model.)

As presented above, the interconnect inductance has competing effects on the total delay. The total delay of an interconnect with repeaters is plotted in Fig. 6.14. As shown in Fig. 6.14, with increasing line inductance, the total delay decreases until a minimum delay is achieved. The analytic model overestimates the inductance effects when the inductance is low; however, the trend of the inductance effect is captured. In Fig. 6.15, the repeater design space satisfying a delay constraint is plotted for different values of inductance. Only the portion of the design space with fewer and smaller repeaters is of interest. As shown in Fig. 6.15, the design space first expands and then shrinks with increasing inductance. Larger inductance does not necessarily result in smaller repeaters. When the inductance changes from $2 \text{ pH}/\mu\text{m}$ to $4 \text{ pH}/\mu\text{m}$,

the number and/or size of the repeaters need to be increased to satisfy the same delay constraint.

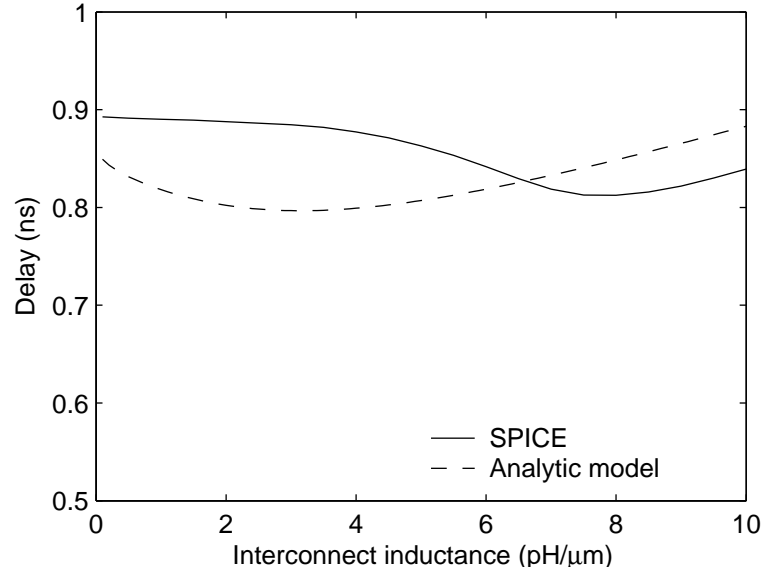


Figure 6.14: Effects of inductance on the interconnect delay with repeaters. $l = 10$ mm, $k = 10$, $h = 100$, and $W = 10W_{min}$.

6.3.3 Power Dissipation with Delay and Bandwidth Constraints

The inductance affects the minimum power with delay and bandwidth constraints in two ways. First, the design space is changed as discussed in Section 6.3.2. Second, the short-circuit power consumed by the repeaters may also be affected by the inductance for a fixed interconnect configuration. As presented in Section 6.3.2, the inductance can produce faster signal transition times, reducing the time during which

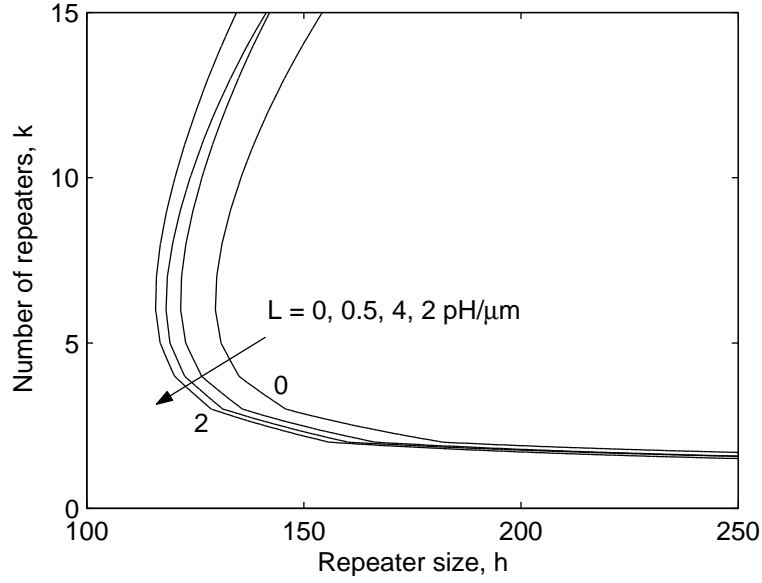


Figure 6.15: Effects of inductance on repeater design space satisfying delay constraints. $T_{req} = 700$ ps, $l = 10$ mm, and $W = 10W_{min}$.

the short-circuit current can flow [146]. The inductance also shields part of the far end capacitance [136], resulting in a smaller effective load capacitance and increasing the peak short-circuit current. The effective capacitance of an RLC interconnect is also determined with the method described in Chapter 4.

In Fig. 6.16, the short-circuit current of a repeater in an interconnect system is illustrated for different inductance values. The short-circuit energy consumed in one signal transition is depicted in Fig. 6.17.

When $h = 100$, the effect of the inductance on the transition time cancels the inductive shielding effect on the load, making the short-circuit power less sensitive to inductance. From this result, it can be seen that the common assumption that

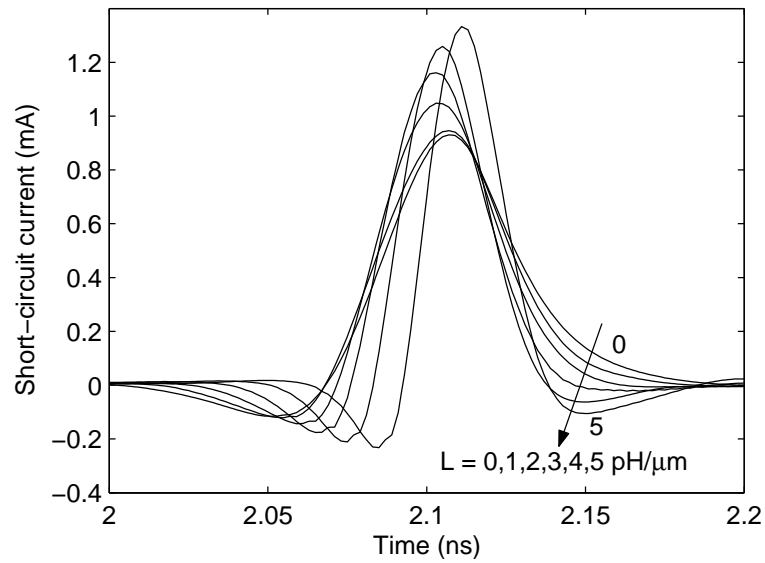


Figure 6.16: Effects of inductance on short-circuit current in repeaters. $l = 10$ mm, $k = 10$, $h = 150$, and $W = 10W_{min}$.

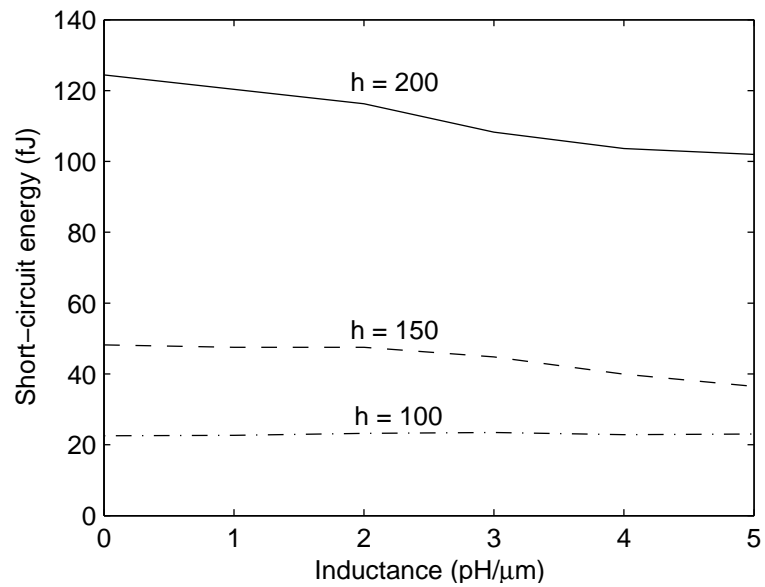


Figure 6.17: Effects of inductance on the short-circuit power in repeaters. $l = 10$ mm, $k = 10$, and $W = 10W_{min}$.

inductance can reduce short-circuit power is not always true. Actually, the short-circuit energy increases slightly with increasing inductance until a maximum energy is achieved. With larger repeater size, the effect of inductance on the transition time increases and starts to dominate the inductive shielding effect on the load for large inductances, decreasing the short-circuit power. For $h = 150$, the short-circuit energy is almost constant when $L < 2 \text{ pH}/\mu\text{m}$, however, both the period and peak value of the short-circuit current vary over this range of inductance, as shown in Fig. 6.16. When $h = 200$, the effect of inductance on the transition time dominates the inductive shielding effect for any value of inductance and the short-circuit power always decreases with increasing inductance.

As described in Section 6.2, the minimum power of an RC interconnect with repeaters can be achieved at the edge of the design space. For practical RLC interconnect structures, this behavior is also valid. Given a design space, the minimum power can be solved numerically by applying the Lagrange method. In Fig. 6.18, the minimum achievable power of an interconnect with inserted repeaters while satisfying a delay constraint is plotted for different values of inductance. The clock frequency is 1 GHz. As shown in Fig. 6.18, by including inductance, the minimum interconnect power under a delay constraint is slightly reduced. This reduction is partially due to the extension of the design space (for low values of inductance) and partially due to the reduction in short-circuit power (for large values of inductance).

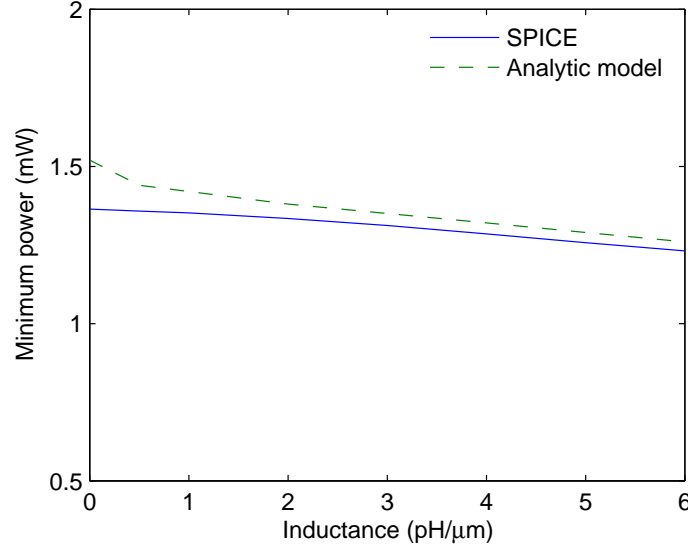


Figure 6.18: Effects of inductance on the minimum interconnect power while satisfying a delay constraint. $l = 15$ mm, $W = 10W_{min}$, and $T_{req} = 1$ ns.

As described in Section 6.2, the minimum power of an RC interconnect with bandwidth constraints can be achieved by using the minimum sized repeater in the design space. This statement, however, is not correct for RLC interconnect. The optimal k for an RLC line to achieve the minimum power is normally unpractically large for small values of inductance. In Fig. 6.19, the minimum achievable power of an RLC interconnect satisfying a bandwidth constraint is plotted for different values of inductance. In this example, k is limited up to 10. As shown in Fig. 6.19, the inductance reduces the minimum power under a bandwidth constraint. Note in Figs. 6.18 and 6.19 that the analytic model overestimates the inductance effect for small values of inductance. The error of the analytic method is less than 11% in Fig. 6.18 and less than 4% in Fig. 6.19.

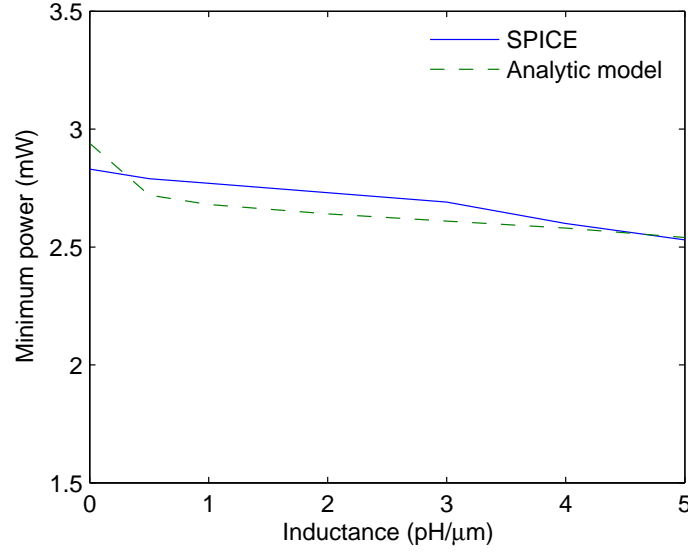


Figure 6.19: Effects of inductance on the minimum interconnect power while satisfying a bandwidth constraint. $l = 15$ mm, $W = 10W_{min}$, and $B_{req} = 2$ Gb/s.

6.4 Conclusions

In this chapter, a repeater insertion design methodology is presented to achieve the minimum power with delay and bandwidth constraints. Input slew effects are considered in the delay model. The minimum power is achieved at the edge of the design space. Closed form solutions for the minimum power in an RC interconnect are developed with delay constraints, where the average error of the model is 7% as compared with SPICE simulations. Satisfying a bandwidth constraint, the minimum power dissipated in an RC interconnect can be achieved with minimum sized repeaters. The effects of inductance on the repeater insertion methodology are also

analyzed. It is shown that the effect of inductance on the interconnect delay (including the delay of the repeaters) and on the short-circuit power is non-monotonic. The overall effects of inductance reduce the minimum achievable power under a delay or bandwidth constraint.

Chapter 7

Predictions of CMOS Compatible On-Chip Optical Interconnect

7.1 Introduction

In deep submicrometer VLSI technologies, multiple design criteria are considered in the interconnect design process, such as delay, power, bandwidth, and noise. With technology scaling, the device dimensions and clock period continuously decrease. The delay uncertainty caused by process and environmental variations consumes a significant part of the clock period, reducing both performance and yield. It has become increasingly difficult for conventional copper based electrical interconnect to satisfy these requirements. One promising candidate to solve this problem is the optical interconnect.

Optical devices are widely used in the telecommunication area, and have been applied as board level interconnects. The concept of on-chip optical interconnect

was first introduced by Goodman in 1984 [154]. Since electrical/optical and optical/electrical conversion is required, optical interconnect is particularly attractive for global interconnects, such as data buses and clock distribution networks, where the required signal conversions can be more easily justified. Recently, several comparisons between on-chip electrical and optical interconnects have been described in [155, 156]. In these papers, the inductive effects of electrical interconnect are ignored, and highly approximate parameters characterizing the optical devices are assumed. The successful realization of on-chip optical interconnect, however, greatly depends upon the development of enhanced CMOS compatible optical devices. Without a reasonable prediction of trends in optical device development, the conclusions presented in [155, 156] are less definitive. Furthermore, delay uncertainty is not addressed in these papers.

Based on a reasonable prediction of optical device development, a more comprehensive comparison between optical and electrical interconnects is performed in this chapter for different technology nodes, considering the design criteria of delay uncertainty, latency, power dissipation, and bandwidth density. This comparison is particularly challenging since optical interconnect is a fast developing technology while electrical interconnect is relatively mature. The rest of this chapter is organized as follows. In Section 7.2, a delay-optimal design of *RLC* interconnect is presented.

In Section 7.3, an on-chip optical data path is introduced. Predictions of the performance characteristics of next generation optical devices are made based on current technology trends. In Section 7.4, electrical and optical interconnects are evaluated for different design criteria. Potential challenges in the development of on-chip optical interconnect are discussed in Section 7.5. Some conclusions are offered in Section 7.6.

7.2 Electrical Interconnect

Repeaters are widely used in global interconnects to reduce interconnect delay, transition times, and crosstalk noise. In Subsection 7.2.1, an *RLC* interconnect with optimal repeaters to achieve the minimum delay is determined for different technology nodes [7]. A model of delay uncertainty is discussed in Subsection 7.2.2.

7.2.1 Delay Optimal Design

A distributed *RLC* interconnect with evenly inserted repeaters is evaluated in this section, as shown in Fig. 7.1. The capacitance and resistance per unit length of the interconnect can be obtained directly from the physical geometry, where the space between adjacent interconnects is assumed to be equal to the minimum interconnect width. The interconnect inductance, however, depends upon the distribution of the current return paths, which are difficult to estimate before the physical design of the circuit is completed. The sensitivity of a signal waveform to errors in the on-chip

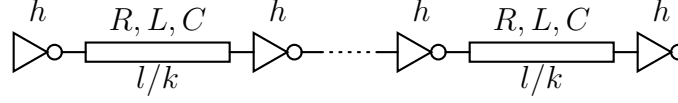


Figure 7.1: Repeater insertion in an RLC interconnect.

inductance, however, is low, and the magnitude of the on-chip inductance is a slowly varying function of the wire geometry [157]. Based on these two characteristics, a fixed value of $0.5 \text{ pH}/\mu\text{m}$ [157] is assumed for all of the technology nodes.

In this analysis, repeaters are implemented as CMOS inverters, and the PMOS transistor is assumed to be twice as large as the NMOS transistor in the inverters. The repeater output capacitance is assumed to be approximately the same as the input gate capacitance [158]. The sensitivity of the timing model to this assumption is relatively low. The delay model used here is the same model as described in Section 6.3.1. The optimal number and size of repeaters along an RLC interconnect can be determined to achieve the minimum delay [59]. This minimum delay can be further decreased by increasing the wire width [159]. Three degrees of freedom are, therefore, explored in electrical interconnect design: the wire width, and the number and size of the repeaters. Various combinations are examined to determine a delay optimal interconnect design. The minimum delay per unit length for different wire widths is illustrated in Fig. 7.2. The interconnect widths are normalized to the minimum wire width W_{min} as predicted in the ITRS [7]. The achievable minimum delay of an RC interconnect at the 90 nm technology node is also plotted in Fig. 7.2

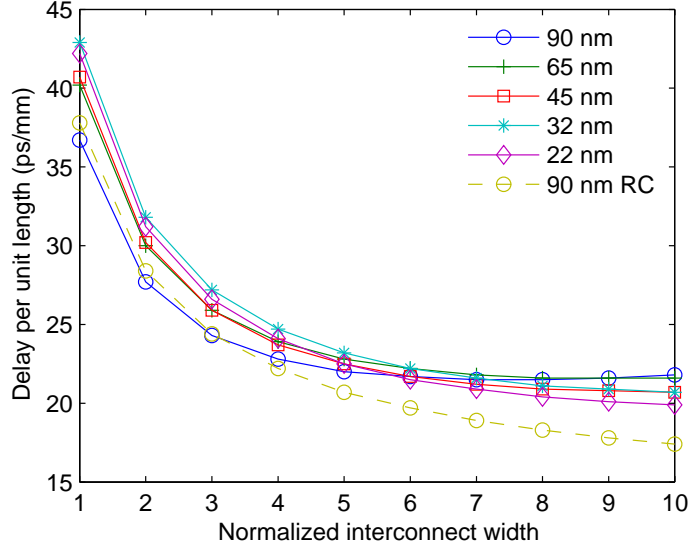


Figure 7.2: Minimum delay per unit length as a function of interconnect width.

to illustrate the effect of inductance on the delay. As shown in Fig. 7.2, at smaller interconnect widths (W_{min} to $4W_{min}$), the interconnect is dominated by the interconnect resistance, making the effect of inductance negligible. For wider interconnects, the effect of inductance on the delay can be significant. It can also be seen from the figure that scaling has only a small effect on the delay of interconnects with repeaters, consistent with the conclusions from [20]. The decrease in delay with increasing wire width slows when the wire width exceeds $3W_{min}$. The minimum achievable delay per unit length is approximately in the range of 20 ps to 22 ps/mm for all technology nodes of interest. Further increasing the wire width greater than $7W_{min}$ only produces small delay differences; the optimal wire width, therefore, is chosen as $7W_{min}$ for each technology node.

7.2.2 Delay Uncertainty Model

Once the delay optimal design is determined, the delay uncertainty caused by process and environmental variations is analyzed through the Monte Carlo method. Since the signal delay is directly related to the power and ground voltage levels, a specific model [160] is used to analyze the effect of power/ground noise on the delay. In this chapter, the 50% delay is based on the effective power/ground voltage levels (rather than the ideal power/ground), which corresponds to the second delay definition among the four types of buffer delays described in [160]. The delay uncertainty caused by power/ground noise consists of two components, one component t_{dif} is due to the differential mode noise (*i.e.*, $\Delta V_{dd} - \Delta V_{ss}$), which affects the delay by changing the effective driver resistance; the other component t_{com} is due to the common mode noise (*i.e.*, $\Delta V_{dd} + \Delta V_{ss}$), which affects the delay by changing the effective switching threshold [160]. Changing the variable V_{dd} in the delay expressions as described in Chapter 5 only addresses the differential mode noise, the effect of the common mode noise, however, needs to be considered explicitly as described in [160],

$$t_{com} = -\frac{\alpha t_r}{2(1 + \alpha)V_{dd_ideal}}(\Delta V_{dd} + \Delta V_{ss}), \quad (7.1)$$

where α is the velocity saturation index of a MOS transistor [150].

7.3 On-Chip Optical Data Path

Introducing optical interconnects into VLSI architectures requires compatibility with CMOS technology. This requirement significantly limits the choice of materials and processes available for fabricating optical components. One of the most significant issues in optical interconnect is the absence of an efficient silicon-based laser that can be monolithically integrated. Only configurations that utilize an external laser as a light source are considered. This type of architecture, however, suffers from increased optical losses due to the coupling of light from the laser to the input waveguide.

A diagram of an optical interconnect system is shown in Fig. 7.3. The system consists of four primary optical elements: an off-chip laser, an optical modulator, a waveguide, and an optical detector. Because the optical modulator and detector in each data path operate at the same wavelength, there is an inherent conflict in the requirements of the optical material. In contrast to a modulator, which requires negligible optical loss, a detector relies on the absorption of light. Considering compatibility with CMOS technology, a practical solution is a $1.5\ \mu\text{m}$ wavelength light source with a silicon modulator and a SiGe or Ge photo-detector [161]. Unlike electrical devices, optical devices are not readily scalable due to the light wavelength constraint [161]. The performance and integration ability of optical devices, however, can be further improved by technological inventions and structural optimizations. Although various device models exist for these optical elements, a specific design is selected to satisfy

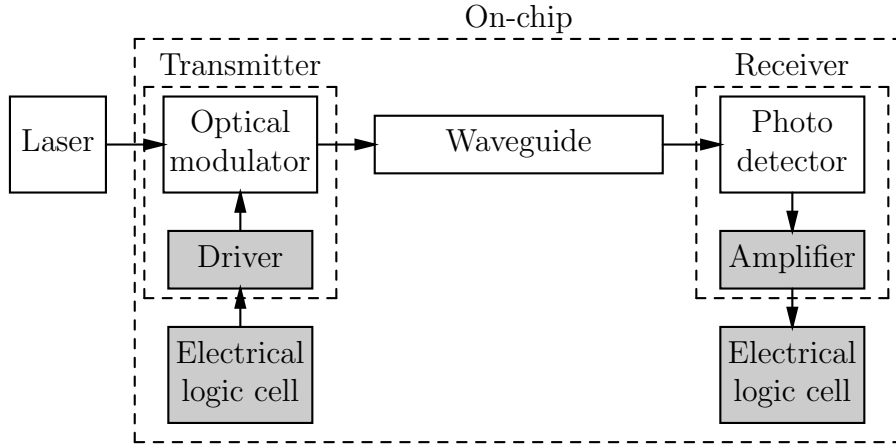


Figure 7.3: An on-chip optical interconnect data path.

the on-chip requirements. Based on this specific design, trends in the development of optical transmitters, waveguides, and optical receivers are described, respectively, in the following subsections.

7.3.1 Transmitters

A transmitter is composed of an electro-optical modulator and a driver circuit. The design of a fast and cost efficient CMOS compatible electro-optical modulator is one of the most challenging tasks on the path towards realizing on-chip optical interconnects. In a modulator, conversion between electrical and optical signals is performed in two steps. First, certain optical properties of the medium, *e.g.*, the refractive index or absorption coefficient, are changed by the electrical signals. Second, the optical signals are modulated, either in amplitude or in phase, by varying the optical properties.

Unstrained bulk crystalline silicon, unfortunately, does not exhibit a linear Pockels effect, and refractive index changes due to the Kerr effect are very weak [162]. One of the few suitable mechanisms for varying the refractive index in pure silicon is the free carrier plasma dispersion effect [162]. There are primarily two electrical structures for changing the carrier concentration in silicon devices. One technique is to inject and extract carriers in the intrinsic region of a p-i-n diode. This structure is described in [163]. With this approach, a substantial change in the carrier concentration can be obtained over a large volume. The speed of this structure, however, is limited by the carrier extraction process. To enhance the speed, a high voltage is needed to extract carriers. An alternative electrical structure is a MOS capacitor. The change in the carrier concentration is achieved by redistribution rather than injection and extraction of carriers, causing a higher modulation speed. The first MOS capacitor based electro-optical modulator was demonstrated by Liu *et al.* [164] and operates at frequencies up to ten gigahertz [165]. By design optimization and technology improvements, such as thinning the gate oxide and using an epitaxial over-growth technique, the bandwidth of the modulator is expected to increase to 30 GHz to 40 GHz by the year 2016.

Because the device structure used in [165] is a Mach-Zehnder interferometer, the modulator has a large footprint (~ 10 mm long), resulting in an excessive capacitance; hence, increasing the delay and power consumption of the driver circuits. Simulations

Table 7.1: Predictive model of future silicon based electro-optical modulators.

		Electrical structure	
		p-i-n diode	MOS capacitor
Optical Structure	Mach-Zehnder	—	Reference [165] High speed (up to 10 GHz) CMOS compatible Large size (10 mm) High power consumption
	Resonator	Reference [163] CMOS compatible Small size (38 μm) Low power consumption	Predictive closed-form model [167] High speed CMOS compatible Small size Low power consumption

and early experiments performed by Barrios *et al.* [166] show that an optical resonator-based structure can drastically decrease the modulator size down to 10 μm to 30 μm while maintaining the same operating principle and speed. Further reductions in the modulator size are possible by using photonic bandgap structures.

In this chapter, a predictive modulator model is used that combines the advantages of the structures used in [163] and [165], as listed in Table 7.1. The performance of a modulator significantly depends on the dimensions and related physical parameters. For example, an increase in extinction coefficient can degrade the modulator bandwidth. To optimize the performance of a modulator, a comprehensive closed-form model [167] is used to choose a proper tradeoff among all of the physical parameters characterizing a MOS modulator.

A series of tapered inverters [168] is used to drive the modulator. If the inverter output capacitance is equal to the input gate capacitance, the optimal size ratio between two neighboring inverters is 3.6 [169]. A minimum sized inverter is used as the first stage. The number of stages can be obtained as $N = \log \frac{C_M}{C_{g0}} / \log 3.6$, where C_M is the modulator capacitance. The delay model of each stage is described in [150].

7.3.2 Waveguides

The performance of optical waveguides is primarily limited by the wavelength of the utilized light and the choice of optical material. Given the operating wavelength of on-chip optical interconnect, there are primarily two candidates for the waveguide material [170]. For applications requiring dense and short waveguide arrays, a silicon-on-insulator (SOI) structure is more beneficial due to the smaller waveguide pitch. For longer paths, optimized for signal propagation delays and smaller losses [171], low-loss polymer waveguides are better suited [172]. Although polymer waveguides require more area than SOI waveguides, polymer waveguides are fabricated on an additional layer and therefore do not consume on-chip silicon resources. In this chapter, a low-refractive index strip polymer waveguide is assumed with a core cross section of $1.5 \mu\text{m} \times 1.5 \mu\text{m}$. The core index and cladding index are 1.6 and 1.1, respectively. The mode effective index of this waveguide is 1.48 for a wavelength of $1.5 \mu\text{m}$.

7.3.3 Receivers

The receiver has two components: a photo-detector that converts light into current followed by an amplifier that converts the analog current signal into a digital voltage signal. A simplified equivalent circuit model is shown in Fig. 7.4.

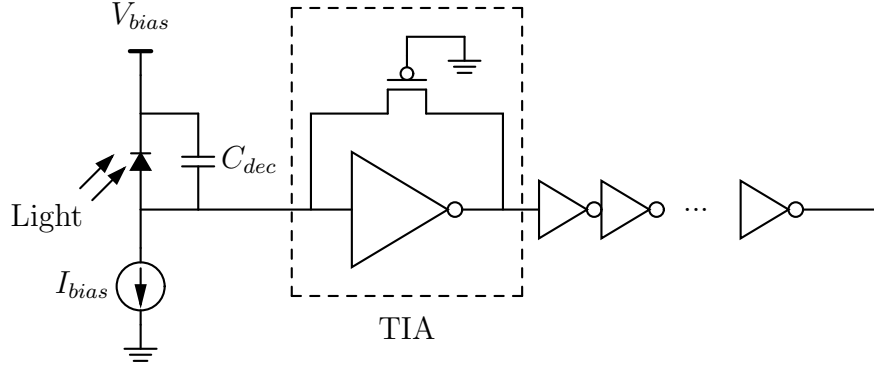


Figure 7.4: Circuit model of an optical receiver.

In this chapter, interdigitated SiGe p-i-n or metal semiconductor metal (MSM) detectors are considered due to the fast response and reasonable quantum efficiency of these structures. The signal rise time (response time) of the detector can be expressed as $T_r = \sqrt{T_{tr}^2 + T_{RC}^2}$, where T_{tr} is the time required for the photo-generated carriers to drift to the electrical contact, and T_{RC} is the RC response time of the detector [173]. The 3 dB bandwidth of a detector is $\Delta f_{dec} = 0.35/T_r$. Based on a one pole approximation, the delay of the photo-detector is related to the rise time as $t_{dec} = 0.315T_r$. In 2002, an interdigitated Ge p-i-n detector fabricated on a Si substrate with a 3 dB bandwidth of 3.8 GHz was demonstrated [174]. Several other papers have been published on SiGe detectors [175, 176]. These detectors exhibit similar performance

levels. The bandwidth and delay of most of these detectors are limited by the carrier transit time, which can be improved through device optimization.

Based on a model proposed by Averine *et al.* [173], the trend in the performance of future detectors is projected. In Fig. 7.5, the MSM detector response time is plotted as a function of electrode width for different detector sizes and compared with some experimental results. Designs A and B are described in [173]. Designs C and D are described in [177] and [175], respectively. The spacing between the electrodes is assumed to be equal to the electrode width. As shown in Fig. 7.5, an optimal electrode width exists that produces a minimum response time. When the electrode is too narrow, the response time is dominated by T_{RC} . When the electrode is too wide, the response time is dominated by T_{tr} .

In this chapter, the electrode width is assumed to be optimized for minimum response time. The minimum response time of a detector decreases with decreasing detector area. This response time is expected in the near future to drop significantly, from tens of ps to a few ps. The reason for this decrease is that present detectors are generally bulky, and a longer time is required for carriers to transit. Effort has been placed on making smaller detectors. Once efficient coupling between the waveguides and detectors is realized, the size of the detector can be significantly reduced, decreasing the response time. This trend, however, is expected to slow and eventually saturate due to fundamental limitations in material properties [178].

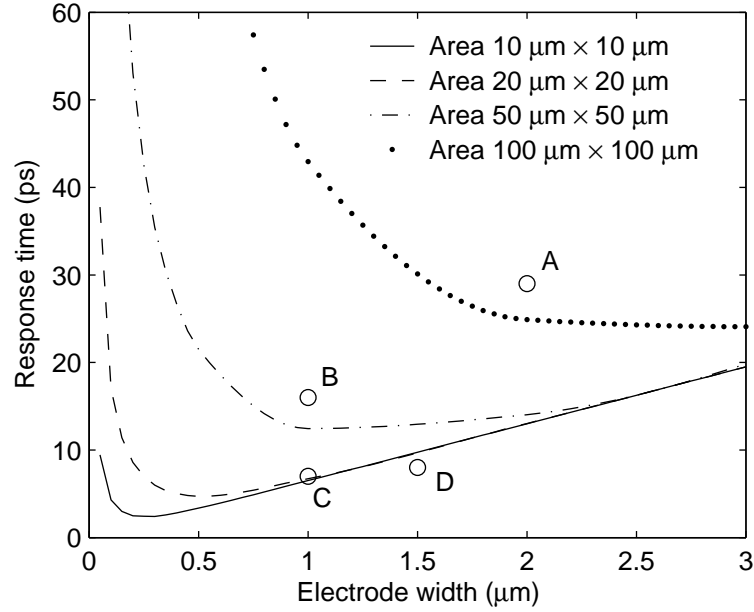


Figure 7.5: Detector response time versus electrode width. A ($100\ \mu\text{m} \times 100\ \mu\text{m}$), B ($50\ \mu\text{m} \times 50\ \mu\text{m}$), C ($20\ \mu\text{m} \times 20\ \mu\text{m}$), and D ($10\ \mu\text{m} \times 10\ \mu\text{m}$).

The photo-current I_{ph} from the photo-detector is amplified by a transimpedance amplifier (TIA). The TIA consists of an inverter and a feedback resistor, implemented as a PMOS transistor. Additional minimum sized inverters are used to amplify the signal to a digital level. A current source I_{bias} is used to bias the input DC current to zero. All of the inverters are assumed to be biased at $V_{dd}/2$. The size of the inverter and feedback transistor in the TIA is determined by bandwidth and noise constraints [158]. The bandwidth requirement of the receiver is assumed to be 0.7 times the bit rate, and the bit error rate (BER) is assumed to be 10^{-15} [158]. For the

receiver circuits, the static power dominates and is

$$P_{rec} = W_{TIA}I_{dsat0}V_{dd} + (I_{bias}V_{dd} + I_{ph}V_{bias})/2 + N_{inv}I_{dsat0}V_{dd}, \quad (7.2)$$

where I_{dsat0} is the saturation drain current of a minimum sized inverter biased at $V_{dd}/2$. W_{TIA} is the size of the TIA normalized to a minimum sized inverter. N_{inv} is the number of additional inverter stages determined by the output swing requirements [158]. The delay of the receiver amplifier t_{amp} is obtained by approximating the circuit as a one pole system, $t_{amp} = 0.7/(2\pi\Delta f_{req})$, where Δf_{req} is the bandwidth requirement. The input optical power is assumed to be 200 μ W [179] with a responsivity of 0.32 A/W.

7.4 Comparison between Electrical and Optical Interconnects

In this section, different criteria used in the design of the two interconnect systems described in Sections 7.2 and 7.3 are compared. The interconnect length is 10 mm. In Subsection 7.4.1, the delay uncertainty of the two kinds of interconnects is analyzed. Delay uncertainty can significantly affect the actual circuit delay due to the pipeline nature of synchronous systems, which is discussed in Subsection 7.4.2. The power and bandwidth density are compared in Subsections 7.4.3 and 7.4.4, respectively.

The critical length above which optical interconnect is beneficial is determined in Subsection 7.4.5.

7.4.1 Delay Uncertainty

Delay uncertainty is caused by process and environmental variations. Variations in the environment include power/ground noise, temperature fluctuations, and crosstalk coupling. In this chapter, all of the variations are assumed to be random with a normal distribution and independent unless explicitly indicated.

Process variations include both die-to-die and within-die variations. Temperature variations also exhibit a similar behavior. The average on-chip temperature is different from die to die due to the ambient environment and local circuit activity. Since different on-chip blocks typically dissipate different amounts of power, the temperature is also non-uniform within a die [180]. The within-die variations at different locations are generally correlated according to the separation distances. Since the focus of this chapter is on comparing electrical and optical global interconnects which cross different regions of an IC, within-die spatial correlation effects need to be considered. The spatial correlation coefficient is modeled as [181]

$$\rho_{cor}(x) = \begin{cases} 1 - \frac{x}{X_L}(1 - \rho_B) & x \leq X_L, \\ \rho_B & x > X_L, \end{cases} \quad (7.3)$$

where x is the separation distance between the variations, ρ_B is the correlation coefficient of the die-to-die variations, and X_L is related to the gradient of the systematic within-die variations. In this chapter, ρ_B is assumed to be 0.5 for process and temperature variations, which means the variations are caused equally by die-to-die and within-die variations. X_L is assumed to be 10 mm for MOS device variations, temperature variations, and power/ground noise, 2 mm for the interconnect height and ILD thickness, and 5 mm for the interconnect width and spacing [182].

The process variations considered in the MOS transistors are: L_{eff} , T_{ox} , N_{ch} , and R_{sd} . By using the analytic models of V_{th} [183, 184], μ [185], and I_{dsat} [186], the variation of the transistor performance can be analyzed. A detailed discription of MOS transistor modeling is provided in Appendix B.

For electrical interconnects, process variations occur in the following primary geometric parameters: wire width W , wire height H , space between wires in the same level S , and thickness of the interlevel dielectric oxide T_{ILD} . The variation in resistance R is primarily determined by geometric parameters and the temperature of the interconnect,

$$R = \frac{\rho_0 l}{WH} [1 + \beta(T - T_0)], \quad (7.4)$$

where β is the temperature coefficient of resistivity, and is 0.0036/°C at 20°C [29]. The interconnect capacitance can be expressed as $C_{gnd} + 2\eta C_c$, where C_{gnd} is the ground capacitance and C_c is the coupling capacitance to the neighboring wires. C_{gnd} and C_c

can be determined from the closed-form expressions provided in [37]. The switching factor η is used to model the Miller effect due to switching activities on neighboring wires [68]. Although the switching patterns are discrete, the signal skews and transition times are continuous and also significantly affect the switching factor. Furthermore, by staggering the repeaters [92], the coupling effect inside a bus structure can be significantly reduced. Based on these considerations, η is assumed to exhibit a continuous normal distribution with $3\sigma = 1$ rather than a discrete distribution, where σ is the standard deviation. Since the sensitivity of the inductance L on the wire geometry is low [187], the effect of process variations on the inductance is ignored. The variation of L is primarily due to neighboring coupling effects, which change the current return paths. In [69], it is shown that, for multiple interconnect systems, the effective wire inductance can be described as

$$L = L_{self} + \sum_i \xi_i M_i, \quad (7.5)$$

where L_{self} is the self partial inductance and M_i is the mutual partial inductance between wire i and the wire of interest. ξ_i is a coefficient which depends upon the signal switching patterns and wire capacitances [69]. To simplify the problem, the inductance is modeled in this chapter as $L = L_0(1 + \xi)$, where L_0 is the typical inductance value $0.5 \text{ pH}/\mu\text{m}$ [157] and ξ is used to model the coupling effect on the effective inductance.

The interconnect is decomposed into a number of segments by repeaters. In each segment, process, temperature, and power/ground variations are assumed to be uniformly distributed, *i.e.*, the ρ_{cor} inside a segment is 1. The ρ_{cor} between different segments is determined by (7.3). η and ξ , however, are assumed to be uniform along the total length of the interconnect, since in a bus structure wires often experience the same neighboring coupling environment over the total length of the line.

Those parameters considered to vary and the corresponding 3σ values are listed in Table 7.2, which are extracted from [7, 188, 189]. Among these parameters, three kinds of correlations are considered. First, power and ground are inversely correlated with a correlation coefficient of -0.5 , since the load current generally causes opposite voltage variations on the power and ground due to the parasitic impedance of power/ground networks. The 3σ value of V_{dd} and V_{ss} is 5.8% of the nominal V_{dd} such that $V_{dd} - V_{ss}$ has a 3σ value of 10%. Second, assuming a fixed wire pitch, the interconnect width and space are inversely correlated with a coefficient of -1 . This behavior is also true for the electrode finger width W_f and space S_f in an interdigitated photo-detector. H and T_{ILD} are assumed to be correlated with a coefficient of -0.5 [188]. Third, ξ is assumed to be correlated to η with a correlation coefficient of -0.3 . This assumption is based on the observation that oppositely switching neighbors result in a greater effective capacitance (*i.e.*, a greater η), and a smaller effective inductance (*i.e.*, a smaller ξ), since neighboring wires provide nearby current return

Table 7.2: Parameters and 3σ variations.

Year		2004	2007	2010	2013	2016
Technology node		90 nm	65 nm	45 nm	32 nm	22 nm
MOS Trans.	L_{eff} (nm)	$37 \pm 10\%$	$25 \pm 10\%$	$18 \pm 10\%$	$13 \pm 10\%$	$9 \pm 10\%$
	T_{ox} (nm)	$2.0 \pm 4\%$	$1.3 \pm 4\%$	$1.1 \pm 4\%$	$1.0 \pm 4\%$	$0.9 \pm 4\%$
	N_{ch} (10^{18}cm^{-3})	$1.55 \pm 5\%$	$2.74 \pm 5\%$	$4.00 \pm 5\%$	$5.85 \pm 5\%$	$10.0 \pm 5\%$
	R_{sd} ($\Omega \cdot \mu\text{m}$)	$180 \pm 10\%$	$162 \pm 10\%$	$135 \pm 10\%$	$107 \pm 10\%$	$79 \pm 10\%$
Electr. Int.	H (μm)	$0.431 \pm 15\%$	$0.319 \pm 15\%$	$0.236 \pm 15\%$	$0.168 \pm 15\%$	$0.125 \pm 15\%$
	T_{ILD} (μm)	$0.431 \pm 15\%$	$0.319 \pm 15\%$	$0.236 \pm 15\%$	$0.168 \pm 15\%$	$0.125 \pm 15\%$
	W (μm)	$1.44 \pm 3\%$	$1.02 \pm 3\%$	$0.72 \pm 3\%$	$0.49 \pm 3\%$	$0.35 \pm 3\%$
	S (μm)	$0.205 \pm 20\%$	$0.145 \pm 20\%$	$0.103 \pm 20\%$	$0.07 \pm 20\%$	$0.05 \pm 20\%$
	η	1 ± 1	1 ± 1	1 ± 1	1 ± 1	1 ± 1
	ξ	0 ± 0.5	0 ± 0.5	0 ± 0.5	0 ± 0.5	0 ± 0.5
Optical Mod.	W_{mod} (μm)	$0.89 \pm 3\%$	$0.89 \pm 2.1\%$	$0.89 \pm 1.5\%$	$0.89 \pm 1.0\%$	$0.89 \pm 0.7\%$
	H_{mod} (μm)	$0.1 \pm 36\%$	$0.1 \pm 26\%$	$0.1 \pm 20\%$	$0.1 \pm 14\%$	$0.1 \pm 10\%$
	N_d (10^{18}cm^{-3})	$7.4 \pm 5\%$	$7.4 \pm 5\%$	$7.4 \pm 5\%$	$7.4 \pm 5\%$	$7.4 \pm 5\%$
	N_a (10^{18}cm^{-3})	$5.4 \pm 5\%$	$5.4 \pm 5\%$	$5.4 \pm 5\%$	$5.4 \pm 5\%$	$5.4 \pm 5\%$
	$T_{ox,mod}$ (nm)	$1.01 \pm 8\%$	$1.01 \pm 5\%$	$1.01 \pm 4.3\%$	$1.01 \pm 4.0\%$	$1.01 \pm 3.6\%$
	V_{bias} (V)	$1.2 \pm 10\%$	$1.2 \pm 9.2\%$	$1.2 \pm 8.3\%$	$1.2 \pm 7.5\%$	$1.2 \pm 6.7\%$
Waveguide	H_{wav} (μm)	$1.5 \pm 4\%$	$1.5 \pm 3\%$	$1.5 \pm 2\%$	$1.5 \pm 2\%$	$1.5 \pm 1\%$
	W_{wav} (μm)	$1.5 \pm 3\%$	$1.5 \pm 2\%$	$1.5 \pm 1\%$	$1.5 \pm 1\%$	$1.5 \pm 1\%$
Detect.	W_f (μm)	$0.505 \pm 8\%$	$0.105 \pm 28\%$	$0.105 \pm 20\%$	$0.095 \pm 15\%$	$0.085 \pm 12\%$
	S_f (μm)	$0.505 \pm 8\%$	$0.105 \pm 28\%$	$0.105 \pm 20\%$	$0.095 \pm 15\%$	$0.085 \pm 12\%$
Env.	V_{dd} (V)	1.2 ± 0.07	1.1 ± 0.064	1.0 ± 0.058	0.9 ± 0.052	0.8 ± 0.046
	V_{ss} (V)	0 ± 0.07	0 ± 0.064	0 ± 0.058	0 ± 0.052	0 ± 0.046
	T ($^{\circ}\text{C}$)	100 ± 50	100 ± 50	100 ± 50	100 ± 50	100 ± 50

paths. Unlike the effective capacitance, which is only related to the immediate neighbors, the effective inductance depends on the neighboring wires over a long distance, making the correlation between η and ξ fairly weak.

For an optical interconnect system, although the waveguide crosses a long distance, geometric variations are assumed to be uniform across the total length. This assumption overestimates the delay uncertainty, since those independent components

of variations in different parts of the waveguide can average out, producing a smaller delay uncertainty. This overestimation, however, does not affect the conclusions of this chapter, since delay uncertainty caused by the waveguide is small as compared with other parts of the system. As described in Section 7.3.3, the receiver amplifier is designed to satisfy the target bandwidth and noise constraints. The design margins of bandwidth and noise are assigned such that target requirements can still be satisfied with process and dynamic environmental variations. The design of the amplifier will be more challenging in future technology nodes due to parameter variations. The input optical power needs to be increased to produce an effective circuit. Although parameter variations in different parts of an optical interconnect may be correlated, the effects of these variations on the delay uncertainty are different due to the different operational mechanisms. In this chapter, the delay uncertainty generated at different parts of the optical data path is assumed to be independent, resulting in the following expression for the standard deviation of the total delay,

$$\sigma_{optical} = \sqrt{\sigma_{drv}^2 + \sigma_{mod}^2 + \sigma_{wav}^2 + \sigma_{dec}^2 + \sigma_{amp}^2}. \quad (7.6)$$

Based on these assumptions, the delay uncertainty of both the electrical and optical interconnect is analyzed. The delay distribution of a 10 mm electrical interconnect at the 45 nm technology node is shown in Fig. 7.6. The delay uncertainty (defined as from -3σ to 3σ) is about one half of the nominal delay. The delay and 3σ values

Table 7.3: Delay (ps) and 3σ value of a 10 mm optical data path.

Year	2004	2007	2010	2013	2016
Tech. node	90 nm	65 nm	45 nm	32 nm	22 nm
Mod. drv.	$37.3 \pm 20.9\%$	$26.5 \pm 20.4\%$	$16.6 \pm 23.5\%$	$10.3 \pm 29.1\%$	$5.2 \pm 40.4\%$
Modulator	$40.0 \pm 67.0\%$	$40.0 \pm 51.0\%$	$40.0 \pm 41.0\%$	$40.0 \pm 32.0\%$	$40.0 \pm 27.0\%$
Waveguide	$49.3 \pm 1.1\%$	$49.3 \pm 0.8\%$	$49.3 \pm 0.5\%$	$49.3 \pm 0.2\%$	$49.3 \pm 0.1\%$
Detector	$2.5 \pm 5.6\%$	$1.1 \pm 21.9\%$	$0.6 \pm 14.1\%$	$0.5 \pm 9.3\%$	$0.4 \pm 7.1\%$
Amplifier	$34.0 \pm 10.6\%$	$13.5 \pm 23.8\%$	$8.7 \pm 17.6\%$	$5.7 \pm 15.8\%$	$3.4 \pm 15.0\%$
Total	$163.1 \pm 17.3\%$	$130.4 \pm 16.4\%$	$115.2 \pm 14.7\%$	$105.8 \pm 12.5\%$	$98.3 \pm 11.2\%$

for different parts of a 10 mm optical data path are listed in Table 7.3. The delay of the transmitter and receiver is determined as explained in Sections 7.3.1 and 7.3.3, respectively. The signal delay in the waveguide is treated as a light propagation delay. The delay uncertainty of the optical interconnect is dominated by the modulator, as listed in Table 7.3. The dimensions of the modulator are not scaled; the manufacture process, however, can be controlled more accurately with technology improvement, reducing the delay uncertainty of the modulator. The total delay uncertainty of the optical interconnect, therefore, is expected to be lower in future technology nodes. The delay uncertainty of the electrical interconnect, in contrast, is expected to slowly increase in future technology nodes due to the larger number of inserted repeaters. A comparison of the standard deviation of the delay of electrical and optical interconnect is shown in Fig. 7.7.

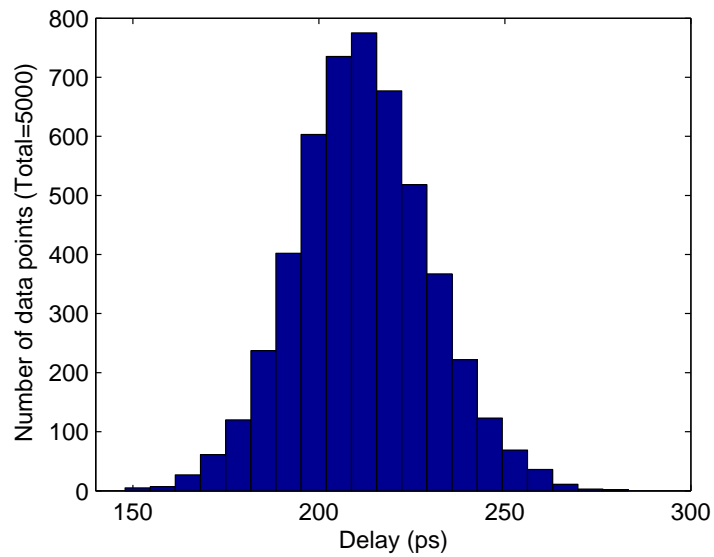


Figure 7.6: Delay distribution of a 10 mm electrical interconnect at the 45 nm technology node.

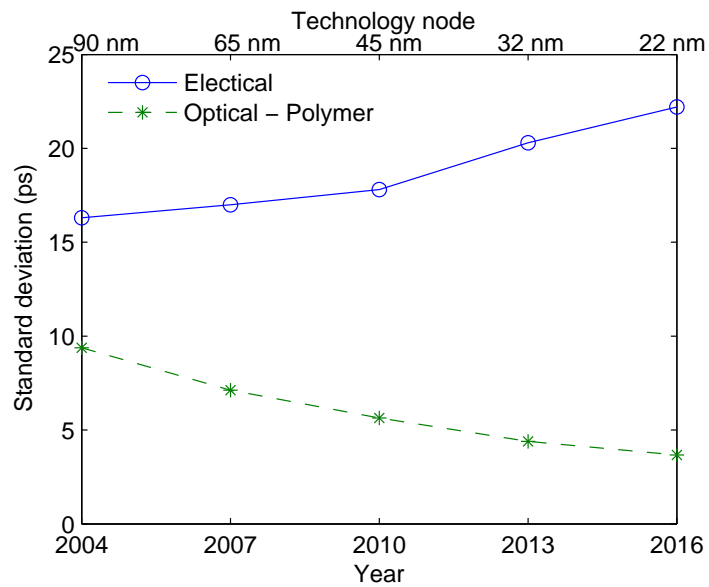


Figure 7.7: Comparison of standard deviation of delays of electrical and optical interconnects.

7.4.2 Delay

As shown in Fig 7.2, the minimum delay of electrical interconnect is about 20 ps to 22 ps/mm. This minimum delay, however, may not be achievable due to the effect of delay uncertainty. A timing diagram of the data and clock waveforms is shown in Fig. 7.8. In the figure, T_{un} is the delay uncertainty, T_{setup} and T_{hold} are the minimum setup and hold requirements at the receiving register, respectively, and T_{bit} is the bit period which is the same as the clock period T_{clk} . The clock signal is assumed to be properly skewed in order for the data to be correctly latched. In this chapter, the timing budget assigned to T_{setup} and T_{hold} is assumed to be 20% of the clock period, *i.e.*, the delay uncertainty cannot exceed 80% of the clock period. Note that this 20% clock period timing budget also includes the delay uncertainty of the register, clock jitter, and the rise time effect (*i.e.*, the rise time of the data cannot be excessively large such that the data waveform can maintain the correct value for a certain period). If this requirement is not satisfied, additional pipeline registers are inserted such that the timing requirements of each stage are satisfied. The delay of the interconnect considering delay uncertainty is

$$T_{total} = m(T_{max} + T_{setup} + T_{C-Q}), \quad (7.7)$$

where m is the number of register stages, T_{C-Q} is the time required for the data to leave the register after the clock signal arrives, and T_{max} is the maximum delay in a stage and can be determined as the summation of the nominal stage delay and 3σ of the delay uncertainty of each stage. $T_{setup} + T_{C-Q}$ is also assumed to be 20% of the clock period.

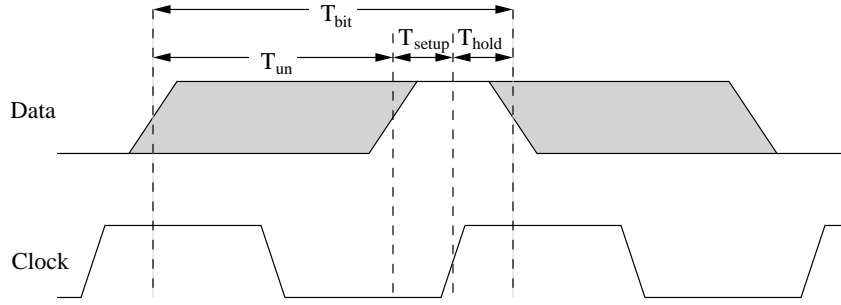


Figure 7.8: A timing diagram of data and clock waveforms.

Expression (7.7) can also be used to calculate the delay of the optical interconnect. Since no register-like device can be inserted into an optical data path, the delay uncertainty determines an upper bound on the channel bandwidth,

$$B_{optical} = \frac{1}{T_{bit}} \leq \frac{0.8}{T_{un}}. \quad (7.8)$$

From Table 7.3, note that the clock frequency, as predicted in the ITRS [7], can be achieved by optical interconnect for each technology node except the 22 nm node. For the 22 nm node, the highest bandwidth determined by (7.8) is 36.3 Gbs. The actual delay of the electrical and optical interconnect is compared in Table 7.4. The delay

Table 7.4: Delay comparison between electrical and optical interconnects.

Year		2004	2007	2010	2013	2016
Technology node		90 nm	65 nm	45 nm	32 nm	22 nm
Electrical	Delay (ps)	311.9	313.2	291.3	312.0	317.8
	# of register stages	1	2	2	4	7
	# of clock cycles	2	3	5	7	12
Optical	Delay (ps)	238.9	173.3	145.4	127.7	114.8
	# of register stages	1	1	1	1	1
	# of clock cycles	1	2	2	3	4

at the 22 nm node is determined with a clock frequency of 36.3 GHz. As listed in Table 7.4, the actual delay of the electrical interconnect remains approximately fixed for all of those technology nodes. The delay of the optical interconnect, however, decreases with future technology nodes due to the increasing performance of the electrical circuits in the modulator driver and the receiver amplifier.

7.4.3 Power

The power dissipated by the electrical interconnect includes dynamic power, short-circuit power, and leakage power. The electrical interconnect power models used in this analysis are the same as those models described in [17]. The power of the registers can be estimated by scaling a typical master-slave D flip-flop and the result is negligible as compared to other interconnect power components.

The power consumed by the optical interconnect is almost independent of the interconnect length, since the length is sufficiently short such that the optical power

Table 7.5: Power consumption (mW) in optical and electrical interconnects.

Year	2004	2007	2010	2013	2016
Technology node	90 nm	65 nm	45 nm	32 nm	22 nm
Transmitter	0.9	1.9	3.4	5.9	11.2
Receiver	0.5	0.5	0.3	0.3	0.3
Total optical	1.4	2.4	3.7	6.2	11.5
Electrical	9.8	16.9	21.7	33.4	45.3

loss in the waveguide is negligible. In this chapter, only electrical power is evaluated for the optical data path, as listed in Table 7.5. The power consumed by the transmitter dominates the power of the receiver, which is in contrast to the assumption made in [155]. The reason for this difference is that the modulator assumed in this analysis is CMOS compatible. The size as well as the capacitance of the modulator is large, requiring a large driver circuit. The power consumed by a 10 mm electrical interconnect is also listed for comparison in Table 7.5. The power consumption of both the electrical and optical interconnects increases due to the higher signal switching frequencies and greater leakage current. Optical interconnect consumes less power than electrical interconnect for all of the technology nodes.

7.4.4 Bandwidth Density

Bandwidth density is an effective criterion for evaluating the ability to transmit data through a unit width. The maximum bit rate for a single interconnect is assumed to be the clock rate (one bit is transmitted per clock period). As described in

Section 7.4.2, the clock frequency at the 22 nm technology node is assumed to be 36.3 GHz. Another parameter related to the bandwidth density is the interconnect pitch. As illustrated in Fig. 7.2, the optimal interconnect width is $7W_{min}$, corresponding to a pitch of $8W_{min}$. For optical interconnects, the waveguide size should be larger than the optical mode size. Based on this limitation, the waveguide pitch is assumed to be $4\text{ }\mu\text{m}$, much larger than the pitch of the electrical interconnect. Single wavelength optical interconnects, therefore, are not beneficial if high bandwidth density is desired. The bandwidth of optical interconnect, however, can be significantly improved by introducing wavelength division multiplexing (WDM) [190]. The bandwidth density of different interconnects is compared in Fig. 7.9.

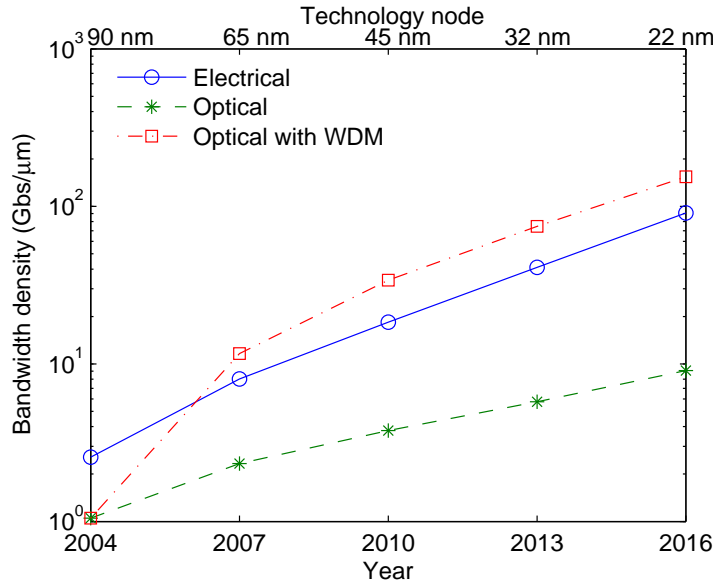


Figure 7.9: Comparison of bandwidth density of electrical and optical interconnects.

For optical interconnect with WDM, the channel number in a waveguide is assumed to be one at the 90 nm technology node, and to increase by four for each new technology node.

7.4.5 Discussion

The critical length beyond which optical interconnect overcomes electrical interconnect is plotted in Fig. 7.10 for different design criteria. The lengths are normalized to the chip edge dimension. As shown in Fig. 7.10, the critical length is approximately one tenth of the chip edge length at the 22 nm technology node.

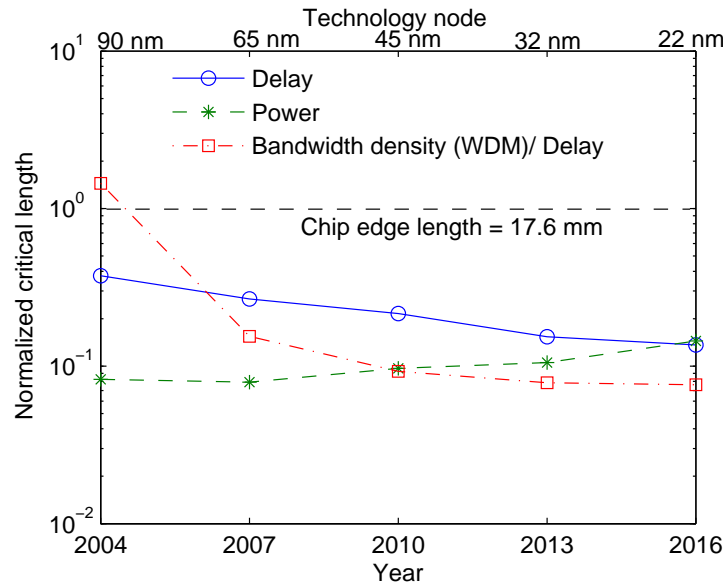


Figure 7.10: Normalized critical length beyond which optical interconnect is advantageous over electrical interconnect.

A direct area comparison of on-chip optical and electrical interconnects might not be legitimate due to the different chip layers used by the two systems. With the

use of a polymer waveguide in optical interconnect, an entire new layer is required. Electrical interconnects, however, are implemented on traditional metal layers. The large optical transmitter and receiver are located at the two ends of the waveguide, in contrast to electrical repeaters, which are distributed along the interconnect. Via congestion issues, therefore, are avoided in optical interconnects.

As compared with [156], the results obtained in this analysis are optimistic for optical interconnect. The reason for this optimism is the choice of device models in this analysis. Rather than a nitride waveguide [156], a polymer waveguide is assumed, increasing the light speed in the waveguide. Furthermore, a more aggressive WDM scheme is applied here, four additional channels per technology node rather than one additional channel per two technology nodes [156]. Another difference in this analysis is that a CMOS-compatible modulator is assumed, which is shown to be one of the most challenging elements in the optical data path. An additional advantage of optical interconnect is the smaller crosstalk noise as compared with electrical interconnect.

7.5 Potential Challenges in Optical Interconnects

Although significant progress had been made towards the development of on-chip optical interconnects over the past decade, there still remains a number of issues that need to be solved [170]. The first problem is the large footprint and power consumption of the optical components and driving circuits, particularly the modulator. This

characteristic permits only a limited number of electrical connections that can be replaced with optics. A solution to this problem can be found by using alternative optical platforms, such as photonic bandgap structures [191] or ring resonators [163], which result in more compact optical components. These enhancements come at the price of stricter fabrication tolerances. A second problem is the generation of sufficient optical power to maintain optical operation. Although a state-of-the-art detector requires only $200\text{ }\mu\text{W}$ of optical power at the input [179], and passive optical loss during the light propagation can be as low as 25%, the number of required detectors can exceed 100 to 1000, even for simple optical interconnect systems. For example, a 64-bit optical interconnect system with 20 point-to-point optical connections requires 0.5 Watts of optical power at the IC input. Generation of this optical power requires multiple off-chip lasers and optical couplers. Using optical interconnects in multiple fan-out applications will further increase the input optical power requirement. Thus, both efficient light sources and detectors are crucial for the development of future on-chip optical interconnects. Finally, a set of integrated silicon-compatible WDM components needs to be developed in order to fully exploit the inherent advantages of optical interconnects.

7.6 Conclusions

A prediction of the performance characteristics of future CMOS compatible optical devices is described in this chapter. Based on this prediction, electrical and optical on-chip interconnects are compared for various design criteria at different technology nodes. Critical lengths beyond which optical interconnect becomes advantageous in terms of delay, power, and bandwidth density/delay are presented. With technology scaling, these lengths are well below expected die size dimensions. Delay uncertainty of both the electrical and optical interconnects is shown to significantly affect the actual signal delay.

Chapter 8

Conclusions

Interconnect now dominates a number of design metrics. As a result of decades of IC design experience, commonly used logic blocks have been highly optimized, making the performance of advanced ICs primarily limited by the global block level interconnects. The focus of the IC design process has therefore shifted from logic optimization to interconnect optimization. Efficient design and modeling of on-chip interconnect are necessary to support this new design perspective.

Various interconnect models have been presented over the last several decades. It is well accepted that global interconnects need to be modeled as transmission lines in GHz applications. In this dissertation, two accurate and efficient solutions for simulating on-chip transmission lines are developed. The first solution is based on a Fourier series analysis of a typical on-chip periodic signal. The far end response is approximated by the summation of several sinusoids. The proposed solution is shown to be an effective strategy which can be used in early circuit level design stages to

estimate the temporal characteristics of periodic signals. Expressions for the 50% delay and overshoots/undershoots have been developed and are shown to be within 11% of SPICE over a wide range of circuit parameters. The single line solution has been applied to tree structures with linear computational complexity, which is shown to be an effective analysis tool for periodic signals such as clock distribution networks. Combined with the modal analysis based decoupling method, the proposed model is also extended to coupled interconnect systems to analyze crosstalk noise.

The second solution is based on a direct pole extraction of the transfer function of a transmission line. Closed-form step and ramp responses are obtained based on these poles. Two pairs of poles can provide an accurate delay estimate, exhibiting an average error of 1% as compared with Spectre simulations. For high frequency related waveform properties, such as the rise time and overshoot, an average error of less than 2% is obtained with ten pairs of poles. The computational complexity of the proposed method is proportional to the number of pole pairs. By using a ladder structure, frequency dependent effects can also be included in the method. Excellent agreement is observed between the proposed model and Spectre simulations.

Interconnects not only dominate the overall circuit delay, but also significantly affect power dissipation. The interconnect resistance and inductance shield part of the load capacitance, resulting in a faster voltage transition at the output of the driver. The short-circuit power dissipated in the driver is therefore larger. In order

to capture the effect of impedance shielding on the short-circuit power, an effective capacitance of a distributed RLC load is developed. This effective capacitance can be used in look-up tables or in empirical k -factor expressions to estimate short-circuit power as well as in analytic analysis to simplify interconnect load models.

In future high complexity circuits, the number of on-chip repeaters will reach millions. The size of a delay-optimal repeater is typically much larger than a minimum sized repeater. These repeaters can consume a significant amount of power. Only a small portion of the on-chip interconnects are in the critical paths, where minimum delay is desired. For non-critical paths, a delay minimal repeater is not suitable and a power-delay tradeoff is needed. Delay or bandwidth constraints on an interconnect determine a design space in terms of the size and number of repeaters. The minimum power can be achieved at the edge of the design space. Closed-form solutions for the minimum power in an RC interconnect with delay constraints are developed. Satisfying a bandwidth constraint, the minimum power dissipated in an RC interconnect can be achieved with minimum sized repeaters. The effects of inductance on the repeater insertion methodology are also analyzed. It is shown that inductance reduces the minimum achievable power under a delay or bandwidth constraint. This repeater insertion methodology provides guidelines for designing repeaters under a strict power budget.

As the requirement of different design criteria becomes more stringent, on-chip optical interconnect has been considered as a promising candidate to replace global electrical interconnect. Based on predictive analytic models of optical devices, including a modulator, waveguide, and detector, an on-chip optical data path is analyzed. Electrical and optical interconnects are compared at different technology nodes for various design criteria, such as delay uncertainty, latency, power, and bandwidth density. Critical lengths beyond which optical interconnect becomes advantageous in terms of different metrics are presented. With technology scaling, these lengths are well below expected die size dimensions. The delay uncertainty of both electrical and optical interconnects is shown to significantly affect the actual signal delay.

Interconnect has become a primary bottleneck in high performance integrated circuits. Both the accuracy and efficiency of interconnect modeling need to be improved due to higher clock frequencies and more complicated interconnect topologies. Interconnect design is no longer a single metric driven process and tradeoffs among different design criteria need to be considered.

Chapter 9

Future Research

As described in previous chapters, on-chip interconnect design is no longer a single criterion process. Tradeoffs among performance, power, and reliability are required. Furthermore, new design challenges are emerging. Existing interconnect design methodologies and analysis tools need to be improved to meet these challenges.

Repeater insertion, as a commonly used interconnect design method, reduces the delay, rise time, and crosstalk noise, but increases power dissipation and area. The effects of repeater insertion on delay uncertainty also need to be analyzed. This future research problem is described in Section 9.1. With on-chip signal frequencies increasing, frequency dependent effects have become more important. A figure of merit to characterize the importance of frequency dependent effects needs to be determined and is discussed in Section 9.2. One possible application of on-chip optical interconnect is the clock distribution network. A design methodology for optical clock distribution networks is discussed in Section 9.3. The combination of 3-D and optical

techniques can significantly reduce the global interconnect delay, which is discussed in Section 9.4. Finally, a summary is provided in Section 9.5.

9.1 Effect of Repeaters on Delay Uncertainty

As discussed in Chapter 7, the delay uncertainty in an interconnect with repeaters is due to several factors such as process variations, temperature variations, power/ground noise, and coupling effects. Repeater insertion affects delay uncertainty in two ways. First, process variations can change the transistor behavior in the inserted repeaters, increasing the delay uncertainty. Second, by segmenting an interconnect with repeaters, the statistical properties of the interconnect variations also change. Interconnect variations include both random components and systematic components. As the interconnect length increases, the random components experience greater averaging, reducing the effective variation [192]. In order to evaluate the effect of repeaters on delay uncertainty, a more accurate model of interconnect variation is needed.

As described in [193], delay uncertainty can be reduced by increasing the size of the repeaters. The effect of the number of repeaters on delay uncertainty also needs to be evaluated. Similar to the delay and bandwidth constraints, a constraint on delay uncertainty also determines a design space for the application of repeaters in global interconnects. By analyzing the effect of repeaters on delay uncertainty, this design

space can be explicitly described and combined with the design space determined by other constraints (see Chapter 6) to optimize the interconnect, as shown in Fig. 9.1, where the shaded area is the design space satisfying different design constraints.

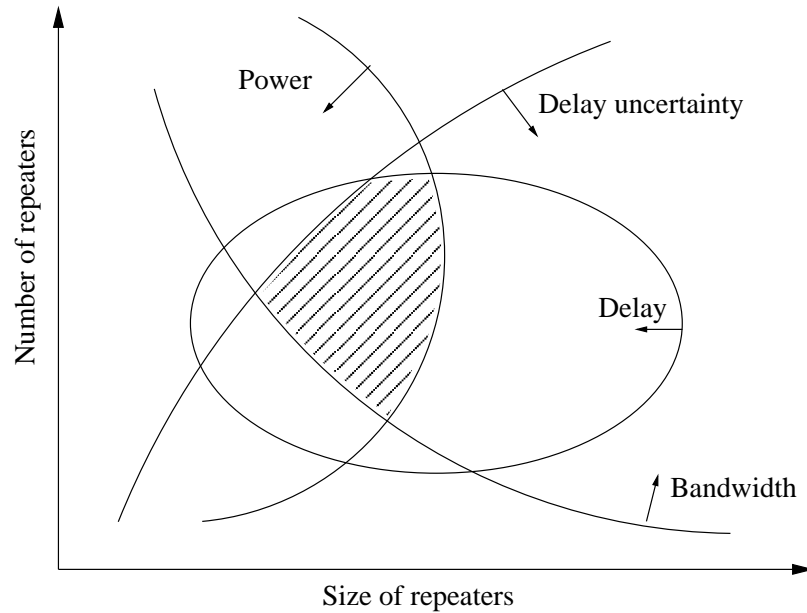


Figure 9.1: Design space for repeaters in global interconnect.

9.2 Figure of Merit to Characterize the Importance of Frequency Dependent Effects

As described in Section 2.3, both the resistance and inductance of an interconnect are a function of frequency due to the skin effect, proximity effect, and multi-path current re-distribution [30]. In order to capture these frequency dependent effects,

more complicated circuit models are required, consuming greater computational resources in the circuit design process. A figure of merit to characterize the importance of frequency dependent effects, therefore, is critical to the IC design process.

Frequency dependent impedance effects on the on-chip signals are determined by the magnitude of the changes in impedance over the frequency range of interest, as shown in Fig. 9.2. At the device level, the relationship between the interconnect impedance and the various physical dimensions needs to be quantified. In certain cases, the impedance changes slowly with increasing frequency, making the frequency dependent effect negligible over the frequency range of interest.

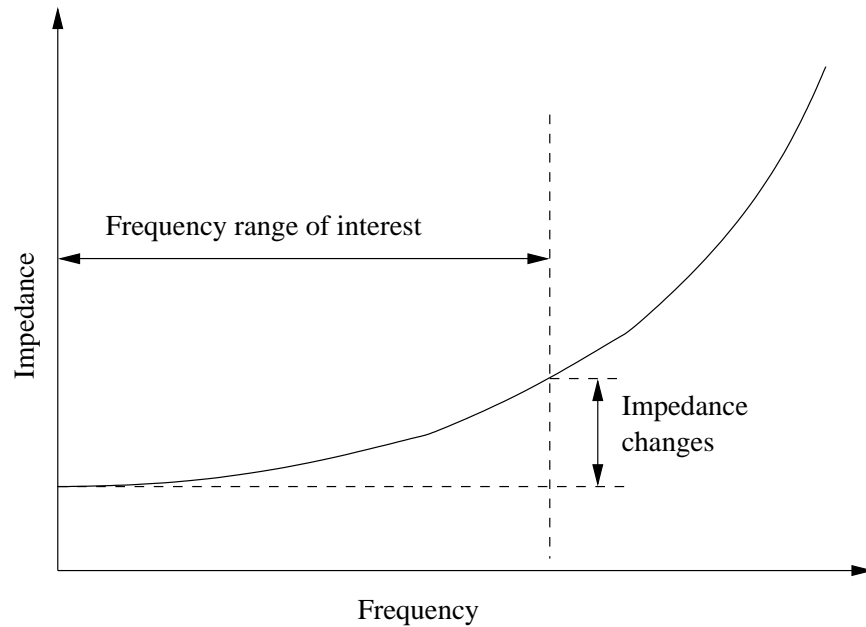


Figure 9.2: Variations in impedance over the frequency range of interest.

At the circuit level, the frequency range of interest is determined by the input signal and the transfer function (assuming a DC impedance). When the input signal

frequency is low or the transfer function exhibits a low pass behavior (due to a small driver, large load, or large interconnect attenuation), frequency dependent effects can also be ignored.

9.3 Design Methodology for Optical Clock Distribution Networks

A clock signal is used to provide a temporal reference for the movement of data in synchronous digital circuits [14]. A low jitter, low power clock distribution network is critical in modern synchronous ICs. As a potential application of on-chip optical interconnect, optical clock distribution networks have been proposed and analyzed [194, 195, 196, 142]. A diagram of an H-tree optical distribution network is shown in Fig. 9.3. As shown in this figure, only the highest levels of the clock distribution network are implemented by optical interconnects. The local clock networks are implemented by electrical interconnects. This topology is due to the limitation placed on the input optical power, bending loss in the waveguides, and the area and power overhead of the optical receivers. A design methodology is therefore necessary to combine global optical interconnect with the local electrical interconnect so as to produce an effective clock distribution network that achieves the objectives of low power, small jitter, and low skew.

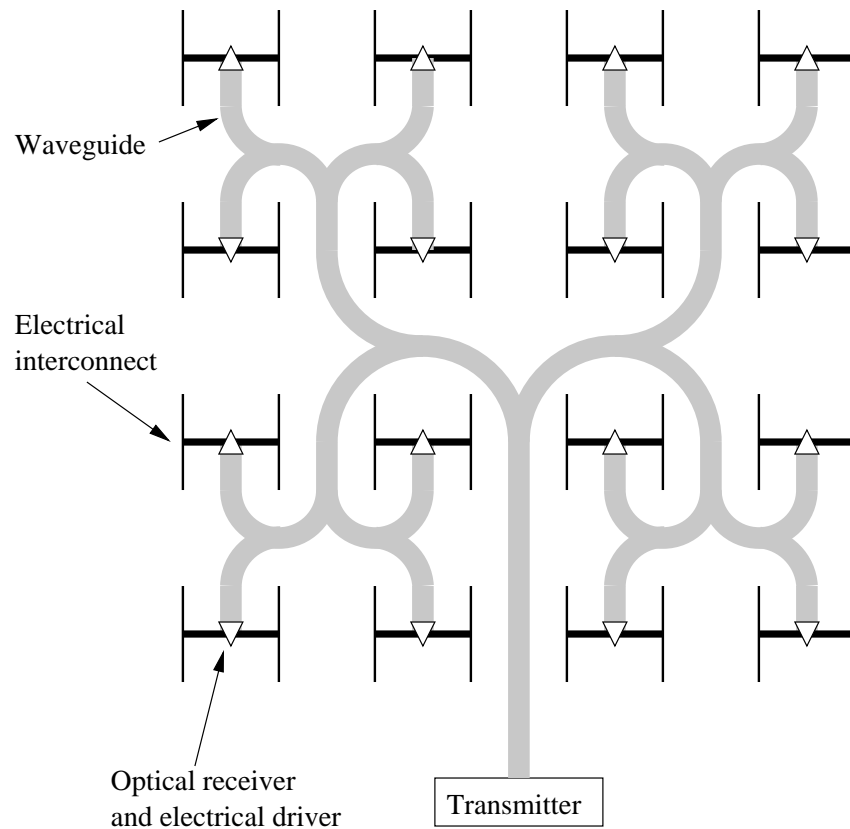


Figure 9.3: Optical-electrical clock distribution network.

9.4 3-D Integration with Optical Interconnects

A primary performance bottleneck in advanced microprocessors is the global interconnect delay. The combination of 3-D technology and on-chip optical interconnect is a novel design paradigm with great potential to overcome this problem. 3-D integration reduces the length of the global interconnects; while optical interconnect enables a faster signal transmitting speed. A design methodology for optimizing the global interconnect delay in a 3-D optical interconnected system is highly desirable. Relevant issues are circuit layer partitioning, vertical via placement, and optical waveguide

routing. A 3-D data path composed of both electrical and optical devices also needs to be properly modeled and analyzed.

9.5 Summary

Several research topics have been described in this chapter to improve existing methodologies for designing conventional electrical interconnects and exploring design issues in optical interconnects. With the scaling of CMOS technology, the on-chip interconnect will continue to be a dominant factor growing in importance. Significant research is necessary to satisfy increasingly challenging interconnect-related design requirements.

Bibliography

- [1] J. M. Rabaey, *Digital Integrated Circuits: A Design Perspective*. NJ: Prentice Hall, 1996.
- [2] G. E. Moore, "Cramming More Components onto Integrated Circuits," *Electronics*, Vol. 38, No. 8, pp. 114-117, April 1965.
- [3] G. E. Moore, "Progress in Digital Integrated Electronics," *Proceedings of the IEEE International Electron Devices Meeting*, pp. 11-13, December 1975.
- [4] Moore's Law. [Online]. Available: http://www.intel.com/museum/archives/history_docs/mooreslaw.htm
- [5] S. D. Naffziger *et al.*, "The Implementation of a 2-Core, Multi-Threaded Itanium Family Processor," *IEEE Journal of Solid-State Circuits*, Vol. 41, No. 1, pp. 197-209, January 2006.
- [6] Technology Trends - Microprocessor Trends. [Online]. Available: <http://www.icknowledge.com/trends/uproc.html>
- [7] *International Technology Roadmap for Semiconductors*. Semiconductor Industry Association, 2003.
- [8] TSMC Unveils Nexsys 90-Nanometer Process Technology. [Online]. Available: http://www.tsmc.com/english/b_technology/b01_platform/b010101_90nm.htm
- [9] J. Cong, "An Interconnect-Centric Design Flow for Nanometer Technologies," *Proceedings of the IEEE*, Vol. 89, No. 4, pp. 505-528, April 2001.
- [10] H. Veendrick, *Deep Submicron CMOS ICs - From Basics to ASICs*. Deventer, Netherlands: Kluwer, 1998.
- [11] Physical Synthesis. [Online]. Available: <http://direct.xilinx.com/bvdocs/whitepapers/wp140.pdf>

- [12] H. B. Bakoglu and J. D. Meindl, "Optimal Interconnection Circuits for VLSI," *IEEE Transactions on Electron Devices*, Vol. ED-32, No. 5, pp. 903-909, May 1985.
- [13] L. P. P. van Ginneken, "Buffer Placement in Distributed RC -tree Network for Minimal Elmore Delay," *Proceedings of the IEEE International Symposium of Circuits and Systems*, pp. 865-868, May 1990.
- [14] E. G. Friedman, "Clock Distribution Networks in Synchronous Digital Integrated Circuits," *Proceedings of the IEEE*, Vol. 89, No. 5, pp. 665-692, May 2001.
- [15] V. Kursun, "Supply and Threshold Voltage Scaling Techniques in CMOS Circuits," Ph.D. Dissertation, University of Rochester, Rochester, New York, 2004.
- [16] N. Magen *et al.*, "Interconnect-Power Dissipation in a Microprocessor," *Proceedings of the ACM International Workshop on System Level Interconnect Prediction*, pp. 7-13, February 2004.
- [17] G. Chen and E. G. Friedman, "Low Power Repeaters Driving RC and RLC Interconnects with Delay and Bandwidth Constraints," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, Vol. 14, No. 2, pp. 161-172, February 2006.
- [18] J. G. Proakis and M. Salehi, *Communication Systems Engineering*. Upper Saddle River, NJ: Prentice Hall, 2002.
- [19] X. Li *et al.*, "Global Interconnect Width and Spacing Optimization for Latency, Bandwidth and Power Dissipation," *IEEE Transactions on Electron Devices*, Vol. 52, No. 10, pp. 2272-2279, October 2005.
- [20] R. Ho, K. W. Mai, and M. A. Horowitz, "The Future of Wires," *Proceedings of the IEEE*, Vol. 89, No. 4, pp. 490-504, April 2001.
- [21] A. Naeemi *et al.*, "Optimal Global Interconnects for GSI," *IEEE Transactions on Electron Devices*, Vol. 50, No. 4, pp. 980-987, April 2003.
- [22] D. Pamunuwa, L. R. Zheng, and H. Tenhunen, "Maximizing Throughput over Parallel Wire Structure in Deep Submicrometer Regime," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, Vol. 11, No. 2, pp. 224-243, April 2003.
- [23] P. Kapur, J. P. McVittie, and K. C. Saraswat, "Technology and Reliability Constrained Future Copper Interconnects—Part I: Resistance Modeling," *IEEE Transactions on Electron Devices*, Vol. 49, No. 4, pp. 590-597, April 2002.

- [24] F. Chen and D. Gardner, "Influence of Line Dimensions on the Resistance of Cu Interconnections," *IEEE Electron Device Letters*, Vol. 19, No. 12, pp. 508-510, December 1998.
- [25] A. H. Ajami *et al.*, "Analysis of *IR*-Drop Scaling with Implications for Deep Submicron P/G Network Designs," *Proceedings of the IEEE International Symposium on Quality Electronic Design*, pp. 35-40, March 2003.
- [26] R. Sarvari and J. D. Meindl, "On the Study of Anomalous Skin Effect for GSI Interconnections," *Proceedings of the IEEE International Interconnect Technology Conference*, pp. 42-44, June 2003.
- [27] W. Wu and K. Maex, "Studies on Size Effect of Copper Interconnect lines," *Proceedings of International Conference on Solid-State and Integrated-Circuit Technology*, pp. 416-418, October 2001.
- [28] S. M. Rossnagel and T. Kuan, "Alteration of Cu Conductivity in the Size Effect Region," *Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures*, Vol. 22, No. 1, pp. 240-247, January 2004.
- [29] C. K. Hu and J. M. E. Harper, "Copper Interconnect: Fabrication and Reliability," *Proceedings of the VLSI Technology, Systems, and Applications*, pp. 18-22, June 1997.
- [30] A. V. Mezhiba and E. G. Friedman, *Power Distribution Networks in High Speed Integrated Circuits*. MA: Kluwer Academic Publishers, 2004.
- [31] R. Sarvari, A. Naeemi, and J. D. Meindl, "General Compact Model for Bit-Rate Limit of Electrical Interconnects Considering DC Resistance, Skin Effect and Surface Scattering," *Proceedings of the IEEE International Interconnect Technology Conference*, pp. 163-165, June 2004.
- [32] K. Nabors and J. White, "FastCap: A Multipole Accelerated 3-D Capacitance Extraction Program," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 10, No. 11, pp. 1447-1459, November 1991.
- [33] W. H. Kao *et al.*, "Parasitic Extraction: Current State of the Art and Future Trends," *Proceedings of the IEEE*, Vol. 89, No. 5, pp. 729-739, May 2001.
- [34] U. Choudhury and A. Sangiovanni-Vincentelli, "Automatic Generation of Analytical Models for Interconnect Capacitances," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 14, No. 4, pp. 470-480, April 1995.

- [35] T. Sakurai and K. Tamaru, "Simple Formulas for Two- and Three-Dimensional Capacitance," *IEEE Transactions on Electron Devices*, Vol. 30, No. 2, pp. 183-185, February 1983.
- [36] J. H. Chern *et al.*, "Multilevel Metal Capacitance Models for CAD Design Synthesis Systems," *IEEE Electron Device Letters*, Vol. 13, No. 1, pp. 32-34, January 1992.
- [37] S. Wong, G. Lee, and D. Ma, "Modeling of Interconnect Capacitance, Delay, and Crosstalk in VLSI," *IEEE Transactions on Semiconductor Manufacturing*, Vol. 13, No. 1, pp. 108-111, February 2000.
- [38] A. Ruehli, "Inductance Calculations in a Complex Integrated Circuit Environment," *IBM Journal of Research and Development*, Vol. 16, No. 5, pp. 470-481, September 1972.
- [39] A. Ruehli, "Equivalent Circuit Models for Three-Dimensional Multiconductor Systems," *IEEE Transactions on Microwave Theory and Techniques*, Vol. 22, No. 3, pp. 216-221, March 1974.
- [40] K. Gala *et al.*, "Inductance 101: Analysis and Design Issues," *Proceedings of the IEEE/ACM Design Automation Conference*, pp. 329-334, June 2001.
- [41] B. Krauter and L. Pileggi, "Generating Sparse Partial Inductance Matrices with Guaranteed Stability," *Proceedings of the IEEE/ACM International Conference on Computer-Aided Design*, pp. 45-52, November 1995.
- [42] K. L. Shepard and Z. Tian, "Return-Limited Inductances: A Practical Approach to On-Chip Inductance Extraction," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 19, No. 4, pp. 425-436, April 2000.
- [43] A. Devgan, H. Ji, and W. Dai, "How to Efficiently Capture On-Chip Inductance Effects: Introducing a New Circuit Element K," *Proceedings of the IEEE/ACM International Conference on Computer-Aided Design*, pp. 150-155, November 2000.
- [44] M. Kamon, M. J. Tsuk, and J. White, "FastHenry: A Multipole Accelerated 3-D Inductance Extraction Program," *IEEE Transactions on Microwave Theory and Techniques*, Vol. 42, No. 9, pp. 1750-1758, September 1994.
- [45] B. Krauter and S. Mehrotra, "Layout Based Frequency Dependent Inductance and Resistance Extraction for On-Chip Interconnect Timing Analysis," *Proceedings of the IEEE/ACM Design Automation Conference*, pp. 303-308, June 1998.

- [46] S. Sim *et al.*, "A Unified *RLC* Model for High-Speed On-Chip Interconnects," *IEEE Transactions on Electron Devices*, Vol. 50, No. 6, pp. 1501-1510, June 2003.
- [47] X. Huang *et al.*, "Loop-Based Interconnect Modeling and Optimization Approach for Multigigahertz Clock Network Design," *IEEE Journal of Solid-State Circuits*, Vol. 38, No. 3, pp. 457-463, March 2003.
- [48] S. Yu *et al.*, "Loop-Based Inductance Extraction and Modeling for Multiconductor On-Chip Interconnects," *IEEE Transactions on Electron Devices*, Vol. 53, No. 1, pp. 135-145, January 2006.
- [49] A. Mezhiba and E. G. Friedman, "Frequency Characteristics of High Speed Power Distribution Networks," *Analog Integrated Circuits and Signal Processing*, Vol. 35, No. 2/3, pp. 207-214, May/June 2003.
- [50] L. N. Dworsky, *Modern Transmission Line Theory and Applications*. New York: John Wiley & Sons, 1979.
- [51] Y. I. Ismail, E. G. Friedman, and J. L. Neves, "Figures of Merit to Characterize the Importance of On-Chip Inductance," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, Vol. 7, No. 4, pp. 442-449, December 1999.
- [52] R. Achar and M. S. Nakhla, "Simulation of High-Speed Interconnects," *Proceedings of the IEEE*, Vol. 89, No. 5, pp. 693-727, February 2001.
- [53] C. K. Cheng *et al.*, *Interconnect Analysis and Synthesis*. New York: John Wiley & Sons, 2000.
- [54] T. Dhaene and D. D. Zutter, "Selection of Lumped Element Models for Coupled Lossy Transmission Lines," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 11, No. 7, pp. 805-815, July 1992.
- [55] A. R. Djordjevic, T. K. Sarkar, and R. F. Harrington, "Analysis of Lossy Transmission Lines with Arbitrary Nonlinear Terminal Networks," *IEEE Transactions on Microwave Theory and Techniques*, Vol. 34, No. 6, pp. 660-666, June 1986.
- [56] S. Y. Kim, N. Gopal, and L. T. Pillage, "Time-Domain Macromodels for VLSI Interconnect Analysis," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 13, No. 10, pp. 1257-1270, October 1994.
- [57] S. Sim, K. Lee, and C. Y. Yang, "High-Frequency On-Chip Inductance Model," *IEEE Electron Device Letters*, Vol. 23, No. 12, pp. 740-742, December 2002.

- [58] T. Sakurai, "Closed-Form Expressions for Interconnection Delay, Coupling, and Crosstalk in VLSI's," *IEEE Transactions on Electron Devices*, Vol. 40, No. 1, pp. 118-124, January 1993.
- [59] Y. I. Ismail and E. G. Friedman, "Effects of Inductance on the Propagation Delay and Repeater Insertion in VLSI Circuits," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, Vol. 8, No. 2, pp. 195-206, April 2000.
- [60] S. Lin and E. Kuh, "Transient Simulation of Lossy Interconnects Based on the Recursive Convolution Formulation," *IEEE Transactions on Circuits and Systems*, Vol. 39, No. 11, pp. 879-892, November 1992.
- [61] J. Chen and L. He, "A Decoupling Method for Analysis of Coupled *RLC* Interconnects," *Proceedings of the ACM Great Lakes Symposium on VLSI*, pp. 41-46, April 2002.
- [62] T. Lin, M. W. Beattie, and L. T. Pileggi, "On the Efficacy of Simplified 2D On-Chip Inductance Models," *Proceedings of the IEEE/ACM Design Automation Conference*, pp. 757-762, June 2002.
- [63] G. Lei, G. Pan, and B. K. Gilbert, "Examination, Clarification, and Simplification of Modal Decoupling Method for Multiconductor Transmission Lines," *IEEE Transactions on Microwave Theory and Techniques*, Vol. 43, No. 9, pp. 2090-2100, September 1995.
- [64] L. Yin and L. He, "An Efficient Analytical Model of Coupled On-Chip *RLC* Interconnects," *Proceedings of the IEEE Design Automation Conference – Asia and South Pacific*, pp. 385-390, January 2001.
- [65] F. Chang, "Transient Analysis of Lossless Coupled Transmission Lines in a Nonhomogeneous Dielectric Medium," *IEEE Transactions on Microwave Theory and Techniques*, Vol. 18, No. 9, pp. 616-626, September 1970.
- [66] F. Romeo and M. Santomauro, "Time-Domain Simulation of n Coupled Transmission Lines," *IEEE Transactions on Microwave Theory and Techniques*, Vol. 35, No. 2, pp. 131-136, February 1987.
- [67] D. Gao, , A. T. Yang, and S. M. Kang, "Modeling and Simulation of Interconnection Delays and Crosstalks in High-Speed Integrated Circuits," *IEEE Transactions on Circuits and Systems*, Vol. 37, No. 1, pp. 1-9, January 1990.
- [68] A. B. Kahng, S. Muddu, and E. Sarto, "On Switch Factor Based Analysis of Coupled *RC* Interconnects," *Proceedings of the IEEE Great Lakes Symposium on VLSI*, pp. 79-84, June 2000.

- [69] Y. Cao *et al.*, “Switch-Factor Based Loop *RLC* Modeling for Efficient Timing Analysis,” *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, Vol. 13, No. 9, pp. 1072-1078, September 2005.
- [70] W. C. Elmore, “The Transient Response of Damped Linear Networks,” *Journal of Applied Physics*, Vol. 19, pp. 55-63, January 1948.
- [71] L. T. Pillage and R. A. Rohrer, “Asymptotic Waveform Evaluation for Timing Analysis,” *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 9, No. 4, pp. 352-366, April 1990.
- [72] P. Feldmann and R. W. Freund, “Efficient Linear Circuit Analysis by Pade Approximation via the Lanczos process,” *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 14, No. 5, pp. 639-649, May 1995.
- [73] M. Silveria, M. Kamon, and J. White, “Efficient Reduced-Order Modeling of Frequency-Dependent Coupling Inductances Associated with 3-D Interconnect Structures,” *IEEE Transactions on Components, Packaging and Manufacturing Technology—Part B: Advanced Packaging*, Vol. 19, No. 2, pp. 283-288, May 1996.
- [74] A. Odabasioglu, M. Celik, and L. T. Pillage, “PRIMA: Passive Reduced-Order Interconnect Macromodeling Algorithm,” *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 17, No. 8, pp. 645-654, August 1998.
- [75] C. J. Alpert *et al.*, “*RC* Delay Metrics for Performance Optimization,” *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 20, No. 5, pp. 571-582, May 2001.
- [76] R. Gupta, B. Tutuianu, and L. T. Pileggi, “The Elmore Delay as a Bound for *RC* Trees with Generalized Input Signals,” *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 16, No. 1, pp. 95-102, January 1997.
- [77] C. V. Kashyap, C. J. Alpert, and A. Devgan, “An Effective Capacitance Based Delay Metric for *RC* Interconnect,” *Proceedings of the IEEE/ACM International Conference on Computer-Aided Design*, pp. 229-234, November 2000.
- [78] J. Cong, “Modeling and Layout Optimization of VLSI Devices and Interconnects in Deep Submicron Design,” *Proceedings of the IEEE Design Automation Conference — Asia and South Pacific*, pp. 121-126, January 1997.

- [79] Y. Ismail, E. G. Friedman, and J. L. Neves, "Equivalent Elmore Delay for *RLC* Trees," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 19, No. 1, pp. 83-97, January 2000.
- [80] Q. Yu and E. S. Kuh, "Exact Moment Matching Model of Transmission Lines and Application to Interconnect Delay Estimation," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, Vol. 3, No. 2, pp. 311-322, April 1995.
- [81] Q. Yu and E. S. Kuh, "Moment Computation of Lumped and Distributed *RC* Trees with Application to Delay and Crosstalk Estimation," *Proceedings of the IEEE*, Vol. 89, No. 5, pp. 772-788, May 2001.
- [82] Y. I. Ismail, "Improved Model-Order Reduction by Using Spatial Information in Moments," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, Vol. 11, No. 5, pp. 900-908, October 2003.
- [83] Y. I. Ismail and E. G. Friedman, "DTT: Direct Truncation of the Transfer Function—An Alternative to Moment Matching for Tree Structured Interconnect," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 21, No. 2, pp. 131-144, February 2002.
- [84] E. Chiprout and M. S. Nakhla, "Analysis of Interconnect Networks Using Complex Frequency Hopping (CFH)," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 14, No. 2, pp. 186-200, February 1995.
- [85] J. Cong, K. Leung, and D. Zhou, "Performance-Driven Interconnect Design Based on Distributed *RC* Delay Model," *Proceedings of the IEEE/ACM Design Automation Conference*, pp. 606-611, June 1993.
- [86] J. Lillis *et al.*, "New Performance Driven Routing Techniques with Explicit Area/Delay Tradeoff and Simultaneous Wire Sizing," *Proceedings of the IEEE/ACM Design Automation Conference*, pp. 395-400, June 1996.
- [87] C. J. Alpert *et al.*, "Buffered Steiner Trees for Difficult Instance," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 21, No. 1, pp. 3-14, January 2002.
- [88] M. L. Mui, K. Banerjee, and A. Mehrotra, "A Global Interconnect Optimization Scheme for Nanometer Scale VLSI with Implications for Latency, Bandwidth, and Power Dissipation," *IEEE Transactions on Electron Devices*, Vol. 51, No. 2, pp. 195-203, February 2004.

- [89] M. A. El-Moursy and E. G. Friedman, "Optimizing Inductive Interconnect for Low Power," *Canadian Journal of Electrical and Computer Engineering*, Vol. 27, No. 4, pp. 183-187, October 2002.
- [90] J. P. Fishburn and C. A. Schevon, "Shaping a Distributed- RC Line to Minimize Elmore Delay," *IEEE Transactions on Circuits and Systems—Part I: Fundamental Theory and Applications*, Vol. 42, No. 12, pp. 1020-1022, December 1995.
- [91] M. A. El-Moursy and E. G. Friedman, "Optimum Wire Shaping of an RLC Interconnect," *Proceedings of the IEEE Midwest Symposium on Circuits and Systems*, December 2003.
- [92] A. B. Kahng *et al.*, "Interconnect Tuning Strategies for High-Performance ICs," *Proceedings of the IEEE Design, Automation and Test in Europe Conference*, pp. 471-478, February 1998.
- [93] M. Ghoneima and Y. Ismail, "Optimum Positioning of Interleaved Repeaters in Bidirectional Buses," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 24, No. 3, pp. 461-469, March 2005.
- [94] J. Lillis, C. Cheng, and T. Y. Lin, "Optimal Wire Sizing and Buffer Insertion for Low Power and a Generalized Delay Model," *IEEE Journal of Solid-State Circuits*, Vol. 31, No. 3, pp. 437-447, March 1996.
- [95] J. Hu *et al.*, "Buffer Insertion with Adaptive Blockage Avoidance," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 22, No. 4, pp. 494-498, April 2003.
- [96] C. J. Alpert, A. Devgan, and S. T. Quay, "Buffer Insertion with Accurate Gate and Interconnect Delay Computation," *Proceedings of the IEEE/ACM Design Automation Conference*, pp. 479-484, June 1999.
- [97] J. Zhang and E. G. Friedman, "Effect of Shield Insertion on Reducing Crosstalk Noise between Coupled Interconnects," *Proceedings of the IEEE International Symposium on Circuits and Systems*, pp. 529-532, May 2004.
- [98] R. Escovar and R. Suaya, "Optimal Design of Clock Trees for Multigigahertz Applications," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 23, No. 3, pp. 352-366, March 2004.
- [99] F. Anderson, J. S. Wells, and E. Z. Berta, "The Core Clock System on the Next Generation Itanium Microprocessor," *Proceedings of the IEEE International Solid-State Circuits Conference*, pp. 110-111, February 2002.

- [100] K. T. Tang and E. G. Friedman, "The Effect of Signal Activity on Propagation Delay of CMOS Logic Gates Driving Coupled On-Chip Interconnections," *Integration, the VLSI Journal*, Vol. 31, No. 3, pp. 209-224, June 2002.
- [101] T. Gao and C. L. Liu, "Minimum Crosstalk Channel Routing," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 15, No. 5, pp. 465-474, May 1996.
- [102] L. He and K. M. Lepak, "Simultaneous Shield Insertion and Net ordering for Capacitive and Inductive Coupling Minimization," *Proceedings of the ACM International Symposium on Physical Design*, pp. 56-61, 2000.
- [103] P. Gupta and A. B. Kahng, "Wire Swizzling to Reduce Delay Uncertainty Due to Capacitive Coupling," *Proceedings of the International Conference on VLSI Design*, pp. 431-436, 2004.
- [104] B. Soudan, "The Effects of Swizzling on Inductive and Capacitive Coupling for Wide Signal Busses," *Proceedings of the International Conference on Microelectronics*, pp. 300-303, December 2003.
- [105] J. A. Davis and J. D. Meindl, "Compact Distributed *RLC* Interconnect Models—Part I: Single Line Transient, Time Delay, and Overshoot Expressions," *IEEE Transactions on Electron Devices*, Vol. 47, No. 11, pp. 2068-2077, November 2000.
- [106] R. Venkatesan, J. A. Davis, and J. D. Meindl, "A Physical Model for the Transient Response of Capacitively Loaded Distributed *RLC* Interconnects," *Proceedings of the IEEE/ACM Design Automation Conference*, pp. 763-766, June 2002.
- [107] A. B. Kahng and S. Muddu, "An Analytical Delay Model for *RLC* Interconnects," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 16, No. 12, pp. 1507-1514, December 1997.
- [108] A. B. Kahng, K. Masuko, and S. Muddu, "Analytical Delay Models for VLSI Interconnects under Ramp Input," *Proceedings of the IEEE/ACM International Conference on Computer-Aided Design*, pp. 30-36, November 1996.
- [109] K. Banerjee and A. Mehrotra, "Accurate Analysis of On-Chip Inductance Effects and Implications for Optimal Repeater Insertion and Technology Scaling," *Proceedings of the IEEE Symposium on VLSI Circuits*, pp. 195-198, June 2001.

- [110] A. B. Kahng and S. Muddu, "Optimal Equivalent Circuits for Interconnect Delay Calculations Using Moments," *Proceedings of the European Design Automation Conference*, pp. 164-169, September 1994.
- [111] K. T. Tang and E. G. Friedman, "Lumped Versus Distributed RC and RLC Interconnect Impedance," *Proceedings of the IEEE Midwest Symposium on Circuits and Systems*, pp. 136-139, August 2000.
- [112] T. J. Bromwich and T. M. MacRobert, *An Introduction to the Theory of Infinite Series, 3rd edition*. NY: Chelsea, 1991.
- [113] J. Zhang and E. G. Friedman, "Crosstalk Noise Model for Shielded Interconnects in VLSI-Based Circuits," *Proceedings of the IEEE International SOC Conference*, pp. 243-244, September 2003.
- [114] S. H. Choi, B. C. Paul, and K. Roy, "Dynamic Noise Analysis with Capacitive and Inductive Coupling," *Proceedings of the IEEE International Conference on VLSI Design*, pp. 65-70, January 2002.
- [115] T. Sakurai, "Approximation of Wiring Delay in MOSFET LSI," *IEEE Journal of Solid-State Circuits*, Vol. 18, No. 4, pp. 418-426, August 1983.
- [116] Y. Eo, J. Shim, and W. R. Eisenstadt, "A Traveling-Wave-Based Waveform Approximation Technique for the Timing Verification of Single Transmission Lines," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 21, No. 6, pp. 723-730, June 2002.
- [117] J. Chen and L. He, "Piecewise Linear Model for Transmission Line With Capacitive Loading and Ramp Input," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 24, No. 6, pp. 928-937, June 2005.
- [118] G. Chen and E. G. Friedman, "An RLC Interconnect Model Based on Fourier Analysis," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 24, No. 2, pp. 170-183, February 2005.
- [119] J. Qian, S. Pullela, and L. Pillage, "Modeling the Effective Capacitance for the RC Interconnect of CMOS Gates," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 13, No. 12, pp. 1526-1535, December 1994.
- [120] K. Agarwal, D. Sylvester, and D. Blaauw, "An Effective Capacitance Based Driver Output Model for On-Chip RLC Interconnects," *Proceedings of the IEEE/ACM Design Automation Conference*, pp. 376-381, June 2003.

- [121] L. K. Vakati and J. Wang, "A New Multi-Ramp Driver Model with *RLC* Interconnect Load," *Proceedings of the IEEE International Symposium of Circuits and Systems*, pp. V269-V271, May 2004.
- [122] Open Source ECSM Format Specification Version 1.2. [Online]. Available: <http://www.cadence.com/webforms/ecsm>
- [123] Composite Current Source Modeling. [Online]. Available: http://www.synopsys.com/products/solutions/galaxy/ccs/cc_source.html
- [124] J. F. Croix and D. F. Wong, "Blade and Razor: Cell and Interconnect Delay Analysis Using Current-Based Models," *Proceedings of the IEEE/ACM Design Automation Conference*, pp. 386-389, June 2003.
- [125] F. Dartu, N. Menezes, and L. T. Pileggi, "Performance Computation for Precharacterized CMOS Gates with *RC* Loads," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 15, No. 5, pp. 544-553, May 1996.
- [126] Y. Cao *et al.*, "Impact of On-Chip Interconnect Frequency-Dependent $R(f)L(f)$ on Digital and RF Design," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, Vol. 13, No. 1, pp. 158-162, January 2005.
- [127] H. J. M. Veendrick, "Short-Circuit Dissipation of Static CMOS Circuitry and Its Impact on the Design of Buffer Circuits," *IEEE Journal of Solid-State Circuits*, Vol. SC-19, No. 4, pp. 468-473, August 1984.
- [128] S. R. Vemuru and N. Scheinberg, "Short-Circuit Power Dissipation Estimation for CMOS Logic Gates," *IEEE Transactions on Circuits and Systems—Part I: Fundamental Theory and Applications*, Vol. 41, No. 11, pp. 762-765, November 1994.
- [129] K. Nose and T. Sakurai, "Analysis and Future Trend of Short-Circuit Power," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 19, No. 9, pp. 1023-1030, September 2000.
- [130] L. Bisdounis, S. Nikolaidis, and O. Koufopavlou, "Propagation Delay and Short-Circuit Power Dissipation Modeling of CMOS Inverter," *IEEE Transactions on Circuits and Systems—Part I: Fundamental Theory and Applications*, Vol. 45, No. 3, pp. 259-270, March 1998.
- [131] J. L. Rosselló and J. Segura, "Charge-Based Analytical Model for the Evaluation of Power Consumption in Submicron CMOS Buffers," *IEEE Transactions*

- on *Computer-Aided Design of Integrated Circuits and Systems*, Vol. 21, No. 4, pp. 433-448, April 2002.
- [132] V. Adler and E. G. Friedman, "Delay and Power Expressions for a CMOS Inverter Driving a Resistive-Capacitive Load," *Analog Integrated Circuits and Signal Processing*, Vol. 14, No. 1/2, pp. 29-39, September 1997.
 - [133] A. Hirata, H. Onodera, and K. Tamaru, "Proposal of a Timing Model for CMOS Logic Gates Driving a $CRC\ \pi$ Load," *Proceedings of the IEEE/ACM International Conference on Computer-Aided Design*, pp. 537-544, November 1998.
 - [134] A. Chatzigeorgiou, S. Nikolaidis, and I. Tsoukalas, "Modeling CMOS Gates Driving RC Interconnect Loads," *IEEE Transactions on Circuits and Systems—Part II: Analog and Digital Signal Processing*, Vol. 48, No. 4, pp. 413-418, April 2001.
 - [135] J. L. Rosselló and J. Segura, "A Simple Power Consumption Model of CMOS Buffers Driving RC Interconnect Lines," *Proceedings of the International Workshop on Power and Timing Modeling Optimization and Simulation*, pp. 4.2.1-4.2.10, September 2001.
 - [136] M. A. El-Moursy and E. G. Friedman, "Shielding Effect of On-Chip Interconnect Inductance," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, Vol. 13, No. 3, pp. 396-400, March 2005.
 - [137] P. R. O'Brien and T. L. Savarino, "Modeling the Driving-Point Characteristic of Resistive Interconnect for Accurate Delay Estimation," *Proceedings of the IEEE/ACM International Conference on Computer-Aided Design*, pp. 512-515, April 1989.
 - [138] X. Yang *et al.*, "Hurwitz Stable Reduced Order Modeling for RLC Interconnect Trees," *Proceedings of the IEEE/ACM International Conference on Computer-Aided Design*, pp. 222-228, November 2000.
 - [139] Z. Qin and C. Cheng, "Realizable Parasitic Reduction Using Generalized Y - Δ Transformation," *Proceedings of the IEEE/ACM Design Automation Conference*, pp. 220-225, June 2003.
 - [140] C. S. Amin, M. H. Chowdhury, and Y. I. Ismail, "Realizable Reduction of Interconnect Circuits Including Self and Mutual Inductances," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 24, No. 2, pp. 271-277, February 2005.

- [141] Q. Wang and S. B. K. Vrudhula, "A New Short Circuit Power Model for Complex CMOS Gates," *Proceedings of the IEEE Alessanfro Volta Memorial Workshop on Low-Power Design*, pp. 98-106, March 1999.
- [142] P. Kapur, G. Chandra, and K. C. Saraswat, "Power Estimation in Global Interconnects and Its Reduction Using a Novel Repeater Optimization Methodology," *Proceedings of the IEEE/ACM Design Automation Conference*, pp. 461-466, June 2002.
- [143] V. Adler and E. G. Friedman, "Repeater Design to Reduce Delay and Power in Resistive Interconnect," *IEEE Transactions on Circuits and Systems—Part II: Analog and Digital Signal Processing*, Vol. 45, No. 5, pp. 607-616, May 1998.
- [144] A. Nalamalpu and W. Burleson, "A Practical Approach to DSM Repeater Insertion: Satisfying Delay Constraints while Minimizing Area and Power," *Proceedings of the IEEE ASIC/SOC Conference*, pp. 152-156, September 2001.
- [145] A. Nalamalpu, S. Srinivasan, and W. P. Burleson, "Boosters for Driving Long Onchip Interconnects—Design Issues, Interconnect Synthesis, and Comparison with Repeaters," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 21, No. 1, pp. 50-62, January 2002.
- [146] Y. Ismail, E. G. Friedman, and J. L. Neves, "Exploiting the On-Chip Inductance in High-Speed Clock Distribution Networks," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, Vol. 9, No. 6, pp. 963-973, December 2001.
- [147] K. Banerjee and A. Mehrotra, "A Power-Optimal Repeater Insertion Methodology for Global Interconnects in Nanometer Designs," *IEEE Transactions on Electron Devices*, Vol. 49, No. 11, pp. 2001-2007, November 2002.
- [148] Berkeley Predictive Technology Model. [Online]. Available: <http://www-device.eecs.berkeley.edu/~ptm>
- [149] Y. Cao *et al.*, "New Paradigm of Predictive MOSFET and Interconnect Modeling for Early Circuit Design," *Proceedings of the IEEE Custom Integrated Circuits Conference*, pp. 201-204, May 2000.
- [150] T. Sakurai and A. R. Newton, "Alpha-Power Law MOSFET Model and Its Applications to CMOS Inverter Delay and Other Formulas," *IEEE Journal of Solid-State Circuits*, Vol. 25, No. 2, pp. 584-594, April 1990.
- [151] S. O. Nakagawa *et al.*, "On-Chip Crosstalk Noise Model for Deep-Submicrometer ULSI Interconnect," *The Hewlett-Packard Journal*, pp. 39-45, August 1998.

- [152] A. Ferré and J. Figueras, "Leakage Power Bounds in CMOS Digital Technologies," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 21, No. 6, pp. 731-738, June 2002.
- [153] N. H. Mahmoud and Y. I. Ismail, "Accurate Rise Time and Overshoots Estimation in *RLC* Interconnects," *Proceedings of the IEEE International Symposium on Circuits and Systems*, pp. 477-480, May 2003.
- [154] J. W. Goodman *et al.*, "Optical Interconnects for VLSI Systems," *Proceedings of the IEEE*, Vol. 72, No. 7, pp. 850-866, July 1984.
- [155] P. Kapur and K. C. Saraswat, "Comparisons Between Electrical and Optical Interconnects for On-Chip Signaling," *Proceedings of the IEEE International Interconnect Technology Conference*, pp. 89-91, June 2002.
- [156] M. J. Koblinsky *et al.*, "On-Chip Optical Interconnects," *Intel Technology Journal*, Vol. 8, No. 2, pp. 129-141, May 2004.
- [157] Y. I. Ismail and E. G. Friedman, "Sensitivity of Interconnect Delay to On-Chip Inductance," *Proceedings of the IEEE International Symposium on Circuits and Systems*, pp. 403-406, May 2000.
- [158] P. Kapur, "Scaling Induced Performance Challenges/Limitations of On-Chip Metal Interconnects and Comparisons with Optical Interconnects," Ph.D. Dissertation, Stanford University, Stanford, California, 2002.
- [159] M. A. El-Moursy and E. G. Friedman, "Optimum Wire Sizing of *RLC* Interconnect With Repeaters," *Integration, the VLSI Journal*, Vol. 38, No. 2, pp. 205-225, December 2004.
- [160] L. H. Chen, M. Marek-Sadowska, and F. Brewer, "Buffer Delay Change in the Presence of Power and Ground Noise," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, Vol. 11, No. 3, pp. 461-473, June 2003.
- [161] G. T. Reed and A. P. Knights, *Silicon Photonics: An Introduction*. NJ: John Wiley & Sons, 2004.
- [162] R. A. Soref and B. R. Bennett, "Electrooptical Effects in Silicon," *IEEE Journal of Quantum Electronics*, Vol. 23, No. 1, pp. 123-129, January 1987.
- [163] Q. Xu *et al.*, "Micrometer-Scale Silicon Electro-Optic Modulator," *Nature*, Vol. 435, pp. 325-327, May 2005.
- [164] A. Liu *et al.*, "A High-Speed Silicon Optical Modulator Based on a Metal-Oxide-Semiconductor Capacitor," *Nature*, Vol. 427, pp. 615-618, February 2004.

- [165] L. Liao *et al.*, "High Speed Silicon Mach-Zehnder Modulator," *Optics Express*, Vol. 13, No. 8, pp. 3129-3135, April 2005.
- [166] C. A. Barrios, V. R. Almeida, and M. Lipson, "Low-Power-Consumption Short-Length and High-Modulation-Depth Silicon Electrooptic Modulator," *Journal of Lightwave Technology*, Vol. 21, No. 4, pp. 1089-1098, April 2003.
- [167] M. Haurylau *et al.*, "Closed-Form Model of a Capacitor-Based Electro-Optical Modulator," *IEEE Transactions on Electron Devices*, 2007 (in review).
- [168] B. S. Cherkauer and E. G. Friedman, "A Unified Design Methodology for CMOS Tapered Buffers," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, Vol. 3, No. 1, pp. 99-111, March 1995.
- [169] N. Hedenstierna and K. O. Jeppson, "CMOS Circuit Speed and Buffer Optimization," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 6, No. 2, pp. 270-281, March 1987.
- [170] M. Haurylau *et al.*, "Optical Interconnect Roadmap: Challenges and Critical Directions," *IEEE Journal of Selected Topics in Quantum Electronics*, Vol. 12, No. 6, pp. 1699-1705, November/December 2006.
- [171] Y. A. Vlasov and S. J. McNab, "Losses in Single-Mode Silicon-On-Insulator Strip Waveguides and Bends," *Optical Express*, Vol. 12, No. 8, pp. 1622-1631, April 2004.
- [172] L. Eldada and L. W. Shacklette, "Advances in Polymer Integrated Optics," *IEEE Journal of Selected Topics in Quantum Electronics*, Vol. 6, No. 1, pp. 54-68, January/February 2000.
- [173] S. V. Averine, Y. C. Chan, and Y. Lam, "Geometry Optimization of Interdigitated Schottky-Barrier Metal-Semiconductor-Metal Photodiode Structures," *Solid-State Electronics*, Vol. 45, No. 3, pp. 441-446, March 2001.
- [174] J. Oh *et al.*, "Interdigitated Ge p-i-n Photodetectors Fabricated on a Si Substrate Using Graded SiGe Buffer Layers," *IEEE Journal of Quantum Electronics*, Vol. 38, No. 9, pp. 1238-1241, September 2002.
- [175] D. Buca *et al.*, "Metal-Germanium-Metal Ultrafast Infrared Detectors," *Journal of Applied Physics*, Vol. 92, No. 12, pp. 7599-7605, December 2002.
- [176] J. Oh, S. K. Banerjee, and J. Campbell, "Metal-Germanium-Metal Photodetectors on Heteroepitaxial Ge-on-Si with Amorphous Ge Schottky Barrier Enhancement Layers," *IEEE Photonics Technology Letters*, Vol. 15, No. 5, pp. 745-747, May 2003.

- [177] S. Y. Chou and Y. Liu, "32 GHz Metal-Semiconductor-Metal Photodetector on Silicon," *Applied Physics Letters*, Vol. 61, No. 15, pp. 1760-1762, October 1992.
- [178] L. Pavesi and D. J. Lockwood, *Silicon Photonics*. New York: Springer, 2004.
- [179] B. Yang *et al.*, "10-Gb/s All-Silicon Optical Receiver," *IEEE Photonics Technology Letters*, Vol. 15, No. 5, pp. 745-747, May 2003.
- [180] A. H. Ajami, K. Banerjee, and M. Pedram, "Modeling and Analysis of Nonuniform Substrate Temperature Effects on Global ULSI Interconnects," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 24, No. 6, pp. 849-861, June 2005.
- [181] P. Friedberg *et al.*, "Modeling Within-Die Spatial Correlation Effects for Process-Design Co-Optimization," *Proceedings of the IEEE International Symposium on Quality Electronic Design*, pp. 516-521, March 2005.
- [182] Y. Cao *et al.*, "Effective On-Chip Inductance Modeling for Multiple Signal Lines and Application to Repeater Insertion," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, Vol. 10, No. 6, pp. 799-805, December 2002.
- [183] Z. Liu *et al.*, "Threshold Voltage Model for Deep-Submicrometer MOSFET's," *IEEE Transactions on Electron Devices*, Vol. 40, No. 1, pp. 86-95, January 1993.
- [184] G. F. Niu, G. Ruan, and R. M. M. Chen, "Further Comments on 'Threshold Voltage Model for Deep-Submicrometer MOSFET's' and Its Extension to Subthreshold Operation," *IEEE Transactions on Electron Devices*, Vol. 43, No. 12, pp. 2311-2312, December 1996.
- [185] N. Arora, *MOSFET Models for VLSI Circuit Simulation – Theory and Practice*. NY: Springer-Verlag/Wien, 1993.
- [186] N. G. Einspruch, *VLSI Electronics Microstructure Science – Advanced MOS Device Physics*. San Diego, California: Academic Press, 1989.
- [187] Y. I. Ismail and E. G. Friedman, "On the Extraction of On-Chip Inductance," *Journal of Circuits, Systems and Computers*, Vol. 12, No. 1, pp. 31-40, February 2003.
- [188] Y. Cao *et al.*, "Design Sensitivities to Variability: Extrapolation and Assessments in Nanometer VLSI," *Proceedings of the IEEE ASIC/SOC Conference*, pp. 411-415, September 2002.

- [189] V. Venkatraman and W. Burleson, "Robust Multi-Level Current-Mode On-Chip Interconnect Signaling in the Presence of Process Variations," *Proceedings of the IEEE International Symposium on Quality Electronic Design*, pp. 522-527, March 2005.
- [190] G. P. Agrawal, *Fiber-Optic Communication Systems*. New York: Wiley, 1997.
- [191] M. Soljacic *et al.*, "Photonic-Crystal Slow-Light Enhancement of Nonlinear Phase Sensitivity," *Journal of Optical Society of America B*, Vol. 19, No. 9, pp. 2052-2059, September 2002.
- [192] B. Razavi, *Design of Analog CMOS Integrated Circuits*. Boston, MA: McGraw-Hill, 2001.
- [193] D. Velenis, R. Sundaresha, and E. G. Friedman, "Buffer Sizing for Delay Uncertainty Induced by Process Variations," *Proceedings of the IEEE International Conference on Electronics, Circuits and Systems*, pp. 415-418, December 2004.
- [194] K. Chen *et al.*, "Comparisons of Conventional, 3-D, Optical and RF Interconnects for On-Chip Clock Distribution," *IEEE Transactions on Electron Devices*, Vol. 51, No. 2, pp. 233-239, February 2004.
- [195] G. Tosik *et al.*, "Power Dissipation in Optical and Metallic Clock Distribution Networks in New VLSI Technologies," *Electronics Letters*, Vol. 40, No. 3, pp. 198-200, February 2004.
- [196] A. V. Mule *et al.*, "Electrical and Optical Clock Distribution Networks for Gigascale Microprocessors," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, Vol. 10, No. 5, pp. 582-594, October 2002.
- [197] Lagrange Multiplier. [Online]. Available: <http://mathworld.wolfram.com/LagrangeMultiplier.html>
- [198] W. Shockley, "A Unipolar Field Effect Transistor," *Proceedings of IRE*, Vol. 40, pp. 1365-1376, November 1952.
- [199] Berkeley Short-Channel IGFET Model. [Online]. Available: <http://www-device.eecs.berkeley.edu/~bsim3>
- [200] T. Sakurai and A. R. Newton, "A Simple MOSFET Model for Circuit Analysis," *IEEE Transactions on Electron Devices*, Vol. 38, No. 4, pp. 887-893, April 1991.
- [201] B. Iniguez, "Comments on 'Threshold Voltage Model for Deep-Submicrometer MOSFET's'," *IEEE Transactions on Electron Devices*, Vol. 42, No. 9, p. 1712, September 1995.

- [202] B. Yu *et al.*, "Short-Channel Effect Improved by Lateral Channel-Engineering in Deep-Submicrometer MOSFET's," *IEEE Transactions on Electron Devices*, Vol. 44, No. 4, pp. 627-634, April 1997.
- [203] B. Cheng and J. Woo, "A Temperature-Dependent MOSFET Inversion Layer Carrier Mobility Model for Device and Circuit Simulation," *IEEE Transactions on Electron Devices*, Vol. 44, No. 2, pp. 343-345, February 1997.
- [204] A. G. Sabnis and J. T. Clemens, "Characterization of the electron mobility in the inverted $\langle 100 \rangle$ Si Surface," *Technical Digest of International Electron Devices Meeting*, pp. 18-21, 1979.
- [205] D. Vasileska and D. K. Ferry, "Scaled Silicon MOSFET's: Universal Mobility Behavior," *IEEE Transactions on Electron Devices*, Vol. 44, No. 4, pp. 577-583, April 1997.
- [206] R. F. Pierret, *Advanced Semiconductor Fundamentals*. Upper Saddle River, NJ: Pearson Education, 2003.
- [207] K. Toh, P. Ko, and R. G. Meyer, "An Engineering Model for Short-Channel MOS Devices," *IEEE Journal of Solid-State Circuits*, Vol. 23, No. 4, pp. 950-958, August 1988.

Appendix A

Minimizing P_{total} with a Delay Constraint for RC Interconnects

As described in Subsection 6.2.3, the total power P_{total} in an RC interconnect with repeaters is

$$P_{total} = P_d + P_s + P_l. \quad (\text{A.1})$$

The minimum P_{total} with delay constraints can be obtained by solving $dP_{total}/dh = 0$.

As shown in Fig. 6.6(b), the curve of $P_d + P_l$ around the power-optimal point is approximated as a part of an ellipse,

$$\frac{(h - h_0)^2}{(h_1 - h_0)^2} + \frac{(P_d + P_l - P_0)^2}{(P_1 - P_0)^2} = 1, \quad (\text{A.2})$$

where h_1 is the minimum repeater size that can satisfy a target delay constraint, which can be obtained by inserting $k_1 = k_{opt}$ into (6.21). h_0 is the optimal repeater size for minimizing $P_d + P_l$ (the corresponding optimal repeater number is k_0). P_0 and P_1 are

the corresponding values of $P_d + P_l$ at (h_0, k_0) and (h_1, k_1) , respectively. Since both P_d and P_l linearly depend on kh , the problem of minimizing $P_d + P_l$ can be formulated as: minimizing function $X(h, k) = kh$ subject to the constraint $T_{total}(h, k) \leq T_{req}$. As described in Subsection 6.2.3, the minimum $P_d + P_l$ can only be achieved at the edge of the design space. The constraint $T_{total} \leq T_{req}$, therefore, can be further simplified as

$$T_{total}(h, k) - T_{req} = 0. \quad (\text{A.3})$$

From the Lagrange method [197], the solution should satisfy the following two equations,

$$\frac{\partial X}{\partial h} + \lambda \frac{\partial(T_{total} - T_{req})}{\partial h} = k + \lambda \left(a_2 R_t C_{g0} - \frac{a_2 R_0 C_t}{h^2} \right) = 0, \quad (\text{A.4})$$

$$\frac{\partial X}{\partial k} + \lambda \frac{\partial(T_{total} - T_{req})}{\partial k} = h + \lambda \left(a_2 R_0 C_0 - \frac{a_1 R_t C_t}{k^2} \right) = 0, \quad (\text{A.5})$$

where λ is called the Lagrange multiplier. Similar to the approach in [144], eliminating λ from (A.4) and (A.5) results in

$$a_2 R_0 C_0 k + \frac{a_2 R_0 C_t}{h} = a_2 R_t C_{g0} h + \frac{a_1 R_t C_t}{k}. \quad (\text{A.6})$$

From the constraint expression (6.21), it can be observed that both sides of (A.6) are

equal to $T_{req}/2$,

$$a_2 R_0 C_0 k + \frac{a_2 R_0 C_t}{h} = \frac{T_{req}}{2}, \quad (\text{A.7})$$

$$a_2 R_t C_{g0} h + \frac{a_1 R_t C_t}{k} = \frac{T_{req}}{2}. \quad (\text{A.8})$$

Solving the above two expressions, k_0 and h_0 can be obtained, as presented in (6.23) and (6.24), respectively. From (A.2), $P_d + P_l$ can be determined as

$$P_d + P_l = P_1 + \frac{(P_0 - P_1) \sqrt{(h_1 - h_0)^2 - (h - h_0)^2}}{h_0 - h_1}. \quad (\text{A.9})$$

kh achieves the minimum value at h_0 , therefore,

$$\left. \frac{d(kh)}{dh} \right|_{h_0} = 0. \quad (\text{A.10})$$

By utilizing this result, the derivative $x_0 = \left. \frac{dP_s}{dh} \right|_{h_0}$ can be obtained as shown in (6.27).

The short-circuit power around the power-optimal point is approximated as

$$P_s = P_s(h_0) + x_0(h - h_0). \quad (\text{A.11})$$

From (A.9) and (A.11), the derivative of P_{total} can be determined as

$$\frac{dP_{total}}{dh} = x_0 - \frac{(h - h_0)(P_0 - P_1)}{(h_0 - h_1)\sqrt{(h_1 - h_0)^2 - (h - h_0)^2}}. \quad (\text{A.12})$$

Setting (A.12) to zero, the optimal repeater size h_p for minimizing P_{total} can be obtained as described in (6.26).

Appendix B

Modeling of MOSFET Transistors

Since the invention of the first MOSFET transistor, various models have been developed to characterize the I-V behavior of a MOS transistor, from the long channel Shockley model [198] to the complicated Berkeley Short-Channel IGFET Model (BSIM) [199] series models which require hundreds of parameters. Although the BSIM models are highly accurate, the computational complexity of these models are significant and an expensive parameter extraction procedure is required. In [150] and [200], Sakurai presented an α power law model and an $n - th$ power law model, respectively. These two models have been widely used in analytic timing and power analysis due to the high efficiency and reasonable accuracy. In these two models, the I-V characteristics are determined empirically, therefore less physical insight is provided. In order to predict the circuit performance of future technologies and study the effects of process variations, a physical MOSFET model is required to capture short channel effects with fewer fitting parameters. An NMOS transistor is used as

an example in this appendix. The rest of this appendix is organized as follows. In Section B.1 and B.2, the models of the two critical electrical parameters of a MOS transistor, threshold voltage and mobility, are described, respectively. The effects of process variations and environmental variations on these two parameters are analyzed. In Section B.3, the current-voltage characteristic of a MOS transistor is described. The transconductance and output resistance are modeled in Section B.4.

B.1 Threshold voltage

By solving a quasi two-dimensional Poisson's equation in the depletion region, the threshold voltage of a deep submicrometer MOSFET with zero body bias can be represented as [183, 201, 184]

$$V_{th} = V_{th0} - \Delta V_{th}, \quad (\text{B.1})$$

where V_{th0} is the long channel threshold voltage and ΔV_{th} is the threshold voltage roll-off due to short-channel effects and drain induced barrier lowering (DIBL). V_{th0}

Table B.1: Parameters used in the model of the threshold voltage.

Symbol	Description
V_{FB}	Flat band voltage
N_{sub}	Uniform substrate doping concentration
X_{dep}	Depletion layer thickness
ϕ_f	Fermi potential of substrate
C_{ox}	Gate capacitance per unit area, ε_{ox}/T_{ox} ,
T_{ox}	Gate oxide thickness
Q_{ox}	Equivalent interface charge
Q_I	Threshold adjusting implanted impurities charge
Φ_m	Metal work function
Φ_{si}	Substrate Si work function
V_t	Thermal voltage, $(kT)/q$
q	Charge of an electron, 1.6×10^{-19} C
k	Boltzman constant, 1.38×10^{-23} J/K
n_i	Intrinsic carrier concentration of Si
ε_{si}	Permittivity of Si, 1.04×10^{-10} F/m
ε_{ox}	Permittivity of SiO ₂ , 3.45×10^{-11} F/m

is determined as

$$V_{th0} = V_{FB} + 2\phi_f + \frac{qN_{sub}X_{dep}}{C_{ox}}, \quad (B.2)$$

$$V_{FB} = \Phi_m - \Phi_{si} - \frac{Q_{ox} + Q_I}{C_{ox}}, \quad (B.3)$$

$$\phi_f = \frac{kT}{q} \ln \frac{N_{sub}}{n_i} = V_t \ln \frac{N_{sub}}{n_i}, \quad (B.4)$$

$$X_{dep} = 2\sqrt{\frac{\varepsilon_{si}\phi_f}{qN_{sub}}}. \quad (B.5)$$

The description of individual parameters used in (B.2)–(B.5) are listed in Table B.1.

For an n^+ polysilicon gate with a p substrate, the flat band voltage can be rewritten as

$$V_{FB} = -\phi_{f-polys} - \phi_f - \frac{Q_{ox} + Q_I}{C_{ox}} \approx -\frac{E_g}{2} - \phi_f - \frac{Q_{ox} + Q_I}{C_{ox}}, \quad (B.6)$$

where E_g is the energy gap of Si and depends upon the temperature in the following form [185],

$$E_g = 1.206 - 2.73 \times 10^{-4}T \text{ (eV)}, \text{ for } T \geq 250 \text{ K}. \quad (B.7)$$

Note that $\Phi_{f-polys}$ and ϕ_f are both the absolute value of the Fermi potential. ΔV_{th} is determined as [184]

$$\Delta V_{th} = \frac{a + b \cosh \gamma}{\sinh^2 \gamma}, \quad (B.8)$$

where

$$a = V_{bi} - 2\phi_f + \frac{V_{ds}}{2}, \quad (B.9)$$

$$b = (V_{bi} - 2\phi_f) \sqrt{1 + \frac{V_{ds}}{V_{bi} - 2\phi_f}}, \quad (B.10)$$

$$\gamma = \frac{L}{2l}, \quad (B.11)$$

$$l = \sqrt{\frac{\varepsilon_{si} T_{ox} X_{dep}}{\varepsilon_{ox} \eta_l}}. \quad (B.12)$$

l is the characteristic length. η_l is a fitting parameter introduced to characterize the variation of the lateral field in the depletion layer. X_{dep}/η_l can be treated as an average of the depletion layer thickness along the channel [183]. V_{bi} is the built-in

potential between the source/drain and substrate, and can be determined as

$$V_{bi} = V_t \ln\left(\frac{N_{sd}N_{sub}}{n_i^2}\right), \quad (\text{B.13})$$

where N_{sd} is the the source/drain doping concentration.

There are several primary process or environmental factors which affect the threshold voltage, such as L , T_{ox} , N_{sub} , T , and V_{dd} . The effects of these parameters are individually discussed in the following subsections.

B.1.1 Effect of L variation

The threshold voltage of a MOSFET transistor can be affected by the channel length through the short-channel effect (SCE), as shown in (B.8). When the channel length becomes shorter, the depletion region in the channel due to the gate overlaps the depletion region due to the source/drain junctions [185]. The source and drain further deplete the channel region, lowering V_{th} . This short-channel effect can also be analyzed through a charge-sharing model [185], which is less accurate than the 2-D Poisson's equation based model. For a uniformly doped channel, V_{th} decreases monotonically with decreasing L . During the fabrication process, the doping along the channel can be non-uniform in the lateral direction due to oxidation enhanced diffusion or implant damage enhanced diffusion [202]. As a result, V_{th} increases with

decreasing L . This effect is called the reverse short-channel effect (RSCE). By increasing the doping concentration in the channel near the source and drain junctions, the V_{th} roll-off can be reduced, since as L decreases, the effective channel doping concentration increases, producing a higher V_{th} . RSCE is not considered in this analysis since RSCE strongly depends upon the concentration profile. The effect of L on V_{th} is described as

$$\frac{\partial V_{th}}{\partial L} = -\frac{\partial \Delta V_{th}}{\partial L} = \frac{S}{2l \sinh^3 \gamma}, \quad (\text{B.14})$$

where

$$S = 2a \cosh \gamma + b(1 + \cosh^2 \gamma). \quad (\text{B.15})$$

B.1.2 Effect of T_{ox} variation

T_{ox} affects V_{th0} via C_{ox} and affects ΔV_{th} via l . The sensitivity of V_{th} to T_{ox} can be obtained by solving the partial derivative of V_{th} ,

$$\frac{\partial V_{th}}{\partial T_{ox}} = \frac{\partial V_{th0}}{\partial T_{ox}} - \frac{\partial \Delta V_{th}}{\partial T_{ox}} = \frac{qN_{sub}X_{dep} - Q_{ox} - Q_I}{\varepsilon_{ox}} - \frac{\gamma S}{2T_{ox} \sinh^3 \gamma}. \quad (\text{B.16})$$

B.1.3 Effect of N_{sub} variation

The substrate doping can affect V_{th} via ϕ_f , X_{dep} , V_{bi} , and l . The sensitivity of V_{th} to N_{sub} is

$$\frac{\partial V_{th}}{\partial N_{sub}} = \frac{\partial V_{th0}}{\partial N_{sub}} - \frac{\partial \Delta V_{th}}{\partial N_{sub}}, \quad (\text{B.17})$$

$$\frac{\partial V_{th0}}{\partial N_{sub}} = \frac{V_t}{N_{sub}} + \frac{1}{C_{ox}} \left[qX_{dep} + \frac{2\varepsilon_{si}(V_t - \phi_f)}{N_{sub}X_{dep}} \right], \quad (\text{B.18})$$

$$\frac{\partial \Delta V_{th}}{\partial N_{sub}} = \frac{1}{N_{sub} \sinh^3 \gamma} \left[\frac{\varepsilon_{si}(V_t - \phi_f)\gamma S}{qN_{sub}X_{dep}^2} - V_t \left(1 + \frac{a}{b} \cosh \gamma \right) \sinh \gamma \right]. \quad (\text{B.19})$$

B.1.4 Effect of T variation

The temperature variation affects V_{th} primarily via n_i . n_i is a strong function of T [185],

$$n_i = AT^{3/2} \exp \left(-\frac{E_g(0)}{2kT} \right) \quad (\text{cm}^{-3}), \quad (\text{B.20})$$

where A is a constant which can be determined by equating (B.20) to $1.5 \times 10^{10} \text{ cm}^{-3}$ [7] at $T = 300 \text{ K}$. $E_g(0) = 1.206 \text{ eV}$ is the band gap energy at $T = 0$. From (B.4) and (B.20), the temperature sensitivity of ϕ_f is [185]

$$\frac{\partial \phi_f}{\partial T} = \frac{1}{T} (\phi_f - 0.603 - 1.5V_t). \quad (\text{B.21})$$

The temperature effect on V_{th0} is [185]

$$\frac{\partial V_{th0}}{\partial T} = \frac{1}{T}(\phi_f - \phi_{f-poly}) + \frac{2\varepsilon_{si}}{X_{dep}C_{ox}T}(\phi_f - 0.603 - 1.5V_t). \quad (B.22)$$

$$\frac{\partial \Delta V_{th}}{\partial T} = \frac{1}{T \sinh^3 \gamma} \left[\frac{\gamma(\phi_f - 0.603 - 1.5V_t)S}{4\phi_f} + (V_{bi} - 2\phi_f)\left(1 + \frac{a}{b} \cosh \gamma\right) \sinh \gamma \right]. \quad (B.23)$$

B.1.5 Effect of V_{dd} variation

V_{th} is a function of V_{ds} due to DIBL. The sensitivity of V_{th} to V_{ds} is

$$\frac{\partial V_{th}}{\partial V_{ds}} = -\frac{b + (V_{bi} - 2\phi_f) \cosh \gamma}{2b \sinh^2 \gamma}. \quad (B.24)$$

The saturation current I_{dn0} when $V_{gs} = V_{ds} = V_{dd}$ is often referred as an index of the driving ability of a NMOS transistor. The effect of V_{dd} variations on the saturation threshold voltage can be determined from (B.24) by replacing V_{ds} by V_{dd} .

B.2 Mobility

In a semiconductor, the velocity of a carrier is proportional to the applied electrical field. This relationship is characterized by the mobility μ ,

$$v = \mu E. \quad (B.25)$$

Mobility is an important parameter in MOS transistor modeling since it significantly affects the I-V behavior of the transistor. There are primarily three major scattering mechanisms that affect the carrier mobility in the inversion layer [203]: phonon scattering (μ_{ph}), surface roughness scattering (μ_{sr}), and Coulombic scattering (μ_c). The effective mobility can be determined from

$$\frac{1}{\mu_{eff}} = \frac{1}{\mu_{ph}} + \frac{1}{\mu_{sr}} + \frac{1}{\mu_c}. \quad (\text{B.26})$$

As shown in [204], the effective mobility, when plotted as a function of the effective transverse electric field E_{eff} , follows a universal curve independent of the substrate doping. E_{eff} is determined as

$$E_{eff} = \frac{1}{\varepsilon_{si}}(\xi Q_{dep} + \eta Q_{inv}), \quad (\text{B.27})$$

where Q_{dep} and Q_{inv} are the charge per unit area in the depletion layer and inversion layer, respectively. ξ and η are weighting coefficients. In [205], it is shown that ξ is strongly influenced by the shape of the doping profile, while η is not significantly affected by the doping profile and is approximately 0.5 for an electron. For a uniformly doped substrate, ξ is approximately 1 [205]. Under this assumption, η is a function of T for an electron,

$$\eta(T) = 0.86 - 1.23 \times 10^{-3}T. \quad (\text{B.28})$$

Ignoring the effect of V_{ds} , E_{eff} can be determined as [185]

$$E_{eff} = \frac{C_{ox}}{\varepsilon} [\eta(V_{ds} - V_{th}) + qN_{sub}X_{dep}]. \quad (\text{B.29})$$

An empirical model for the effective mobility is often used in circuit simulations [185],

$$\mu_{eff} = \frac{\mu_0}{1 + \alpha_\theta E_{eff}}, \quad (\text{B.30})$$

where α_θ is called the scattering constant and $\mu_0 = 0.067 \text{ m}^2/\text{V} \cdot \text{sec}$ at 300 K. α_θ is selected so that the μ_{eff} obtained from (B.30) matches the value predicted by the ITRS [7]. Inserting (B.29) into (B.30), the effective mobility is

$$\mu_{eff} = \frac{\mu_0}{1 + \theta [\eta(V_{gs} - V_{th}) + qN_{sub}X_{dep}]}, \quad (\text{B.31})$$

where

$$\theta = \frac{\alpha_\theta C_{ox}}{\varepsilon_{si}} \approx \frac{\alpha_\theta}{3T_{ox}}. \quad (\text{B.32})$$

When the lateral electrical field E_y is sufficient high, the carrier velocity is no longer proportional to the lateral field and tends to saturate. A commonly used

piecewise model for velocity saturation is [185]

$$v = \begin{cases} \frac{\mu_{eff} E_y}{1 + (E_y/E_{sat})} & E_y \leq E_{sat}, \\ \frac{\mu_{eff} E_{sat}}{2} & E_y > E_{sat}, \end{cases} \quad (\text{B.33})$$

where E_{sat} is the saturated lateral electrical field and is determined by $2v_{sat}/\mu_{eff}$. v_{sat} is the carrier saturation velocity and is approximated 1×10^5 m/s for an electron at 300 K. v_{sat} depends on T in the following form [206],

$$v_{sat} = \frac{v_0}{1 + A_s \exp(T/T_0)}, \quad (\text{B.34})$$

where $v_0 = 2.4 \times 10^5$ m/s, $A_s = 0.8$, and $T_0 = 600$ K. μ_0 depends on T as [185]

$$\mu_0(T) = \mu_0(T_0) \left(\frac{T}{T_0}\right)^{-m}. \quad (\text{B.35})$$

For NMOS transistors, m is assumed to be 1.5. The temperature effect on mobility is dominated by μ_0 . By ignoring the temperature effect on E_{eff} , the temperature coefficient of mobility is [185]

$$\frac{\partial \mu_{eff}}{\partial T} = -\frac{1.5 \mu_{eff}}{T}. \quad (\text{B.36})$$

The dependence of mobility on the other parameters can be obtained by solving the corresponding partial derivatives.

B.3 I-V characteristics

The saturation drain current per unit channel width is obtained in the ITRS [7] as

$$I_{dn0} = \frac{I_{d,ideal}}{1 + \frac{R_{sd}I_{d,ideal}}{V_{ov}}}, \quad (\text{B.37})$$

where R_{sd} is the drain/source resistance, and $V_{ov} = V_{gs} - V_{th}$ is the overdrive voltage.

I_{ideal} is the ideal saturation current ignoring the effect of R_{sd} and is given as

$$I_{d,ideal} = \frac{v_{sat}C_{ox}V_{ov}^2}{V_{ov} + LE_{sat}}. \quad (\text{B.38})$$

By including channel length modulation effects, (B.38) can be modified as

$$I_{d,ideal} = \frac{v_{sat}C_{ox}V_{ov}^2}{V_{ov} + (L - \Delta L)E_{sat}}. \quad (\text{B.39})$$

ΔL is the length of the velocity saturation region, which is given by [186]

$$\Delta L = l \ln \frac{\frac{V_{ds} - V_{dsat}}{l} + E_m}{E_{sat}}, \quad (\text{B.40})$$

where E_m is the electrical field at the drain junction,

$$E_m = \sqrt{\left(\frac{V_{ds} - V_{dsat}}{l}\right)^2 + E_{sat}^2}. \quad (\text{B.41})$$

The saturation drain voltage is [207]

$$V_{dsat} = \frac{E_{sat}(L - \Delta L)V_{ov}}{E_{sat}(L - \Delta L) + V_{ov}}. \quad (\text{B.42})$$

l is the characteristic length provided in (B.12). T_{ox} in (B.12) is the equivalent oxide thickness which is given by $T_d/(\varepsilon_k/\varepsilon_{ox})$. T_d and ε_k are the thickness and dielectric constant of the gate dielectric, respectively. Note that V_{dsat} and ΔL depend on each other. Several iterations, therefore, are needed to solve these equations.

B.4 Transconductance and output resistance

Normally, $V_{dsat} \ll V_{ds}$ and $E_{sat} \ll (V_{ds} - V_{dsat})/l$. E_m can be approximated as

$$E_m = \frac{V_{ds}}{l}, \quad (\text{B.43})$$

and (B.40) can be approximated as

$$\Delta L = l \ln \frac{2V_{ds}}{lE_{sat}}. \quad (\text{B.44})$$

With these approximations, the ideal transconductance can be obtained as

$$g_{m_ideal} = \frac{\partial I_{d_ideal}}{\partial V_{gs}} = I_{d_ideal} \left(\frac{2}{V_{ov}} - \frac{1 + (L - \Delta L + l) \frac{2v_{sat}\theta}{\mu_0}}{V_{ov} + E_{sat}(L - \Delta L)} \right). \quad (\text{B.45})$$

The output conductance is [158, 207]

$$g_{ds_ideal} = \frac{\partial I_{d_ideal}}{\partial V_{ds}} = \frac{I_{d_ideal} E_{sat}}{E_m [V_{ov} + (L - \Delta L) E_{sat}]}. \quad (\text{B.46})$$

By including the drain/source resistance R_{ds} , the effective transconductance and output resistance can be obtained [7, 186], respectively, as

$$g_m = \frac{g_{m_ideal}}{1 + 0.5 R_{sd} g_{m_ideal}}, \quad (\text{B.47})$$

$$r_{ds} \approx \frac{1}{g_{ds_ideal}} + 2R_{sd}. \quad (\text{B.48})$$

Appendix C

Publications

Journal Papers

1. G. Chen and E. G. Friedman, "Effective Capacitance of Inductive Interconnects for Short-Circuit Power Analysis," submitted to *IEEE Transactions on Circuits and Systems II*.
2. G. Chen and E. G. Friedman, "Transient Response of a Distributed *RLC* Interconnect Based on Direct Pole Extraction," submitted to *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*.
3. M. Haurylau, J. Zhang, H. Chen, G. Chen, N. A. Nelson, D. H. Albonesi, E. G. Friedman, and P. M. Fauchet, "Closed-Form Model of a Capacitor-Based Electro-Optical Modulator," submitted to *IEEE Transactions on Electronic Devices*.
4. G. Chen, H. Chen, M. Haurylau, N. Nelson, D. Albonesi, P. M. Fauchet, and E. G. Friedman, "Predictions of CMOS Compatible On-Chip Optical Interconnect," *Integration, the VLSI journal* 2007(in press).
5. M. Haurylau, G. Chen, H. Chen, J. Zhang, N. A. Nelson, D. H. Albonesi, E. G. Friedman, and P. M. Fauchet, "On-Chip Optical Interconnect Roadmap: Challenges and Critical Directions," *IEEE Journal of Selected Topics in Quantum Electronics*, Vol. 12, No. 6, pp. 1699-1705, November/December, 2006.

6. G. Chen and E. G. Friedman, "Low Power Repeaters Driving RC and RLC Interconnects with Delay and Bandwidth Constraints," *IEEE Transactions on Very Large Scale Integration (VLSI) systems*, Vol. 14, No. 2, pp. 161-172, February 2006.
7. G. Chen and E. G. Friedman, "An RLC Interconnect Model Based on Fourier Analysis," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 24, No. 2, pp. 170-183, February, 2005.
8. C. Liu, Z. Wang, G. Chen, Y. Li, E. Wu, D. Li, B. Li, W. Dou, Z. Dong, "A DAB Transmitter Prototype with High Flexibility and Low Cost," *IEEE Transactions on Broadcasting*, Vol. 48, No. 3, pp. 173-178, September 2002.

Conference Papers

1. G. Chen, H. Chen, M. Haurylau, N. Nelson, D. Albonesi, P. M. Fauchet, and E. G. Friedman, "On-Chip Copper-Based vs. Optical Interconnects: Delay Uncertainty, Latency, Power, and Bandwidth Density Comparative Predictions," *Proceedings of the IEEE International Interconnect Technology Conference*, pp. 39-41, June 2006.
2. G. Chen and E. G. Friedman, "Effective Capacitance of RLC Loads for Estimating Short-Circuit Power," *Proceeding of IEEE International Symposium on Circuits and Systems*, pp. 2065-2068, May 2006.
3. M. Haurylau, H. Chen, J. Zhang, G. Chen, N. A. Nelson, D. H. Albonesi, E. G. Friedman, and P. M. Fauchet, "On-Chip Optical Interconnect Roadmap: Challenges and Critical Directions," *Proceedings of the IEEE International Conference on Group IV Photonics*, pp. 17-19, September 2005.

4. G. Chen, H. Chen, M. Haurylau, N. Nelson, D. Albonesi, P. M. Fauchet, and E. G. Friedman, "Electrical and Optical On-Chip Interconnects in Scaled Microprocessors," *Proceedings of the IEEE International Symposium on Circuits and Systems*, pp. 2514-2517, May 2005.
5. G. Chen and E. G. Friedman, "Low Power Repeaters Driving RLC Interconnects with Delay and Bandwidth Constraints," *Proceedings of the IEEE International Symposium on Circuits and Systems*, pp. 596-599, May 2005.
6. G. Chen and E. G. Friedman, "A Fourier Series-Based RLC Interconnect model for Periodic Signals," *Proceedings of the IEEE International Symposium on Circuits and Systems*, pp. 4126-4129, May 2005.
7. G. Chen, H. Chen, M. Haurylau, N. Nelson, D. Albonesi, P. M. Fauchet, and E. G. Friedman, "Predictions of CMOS Compatible On-Chip Optical Interconnect," *Proceedings of the International Workshop on System Level Interconnect Prediction*, pp. 13-20, April 2005.
8. N. Nelson, G. Briggs, M. Haurylau, G. Chen, H. Chen, D. H. Albonesi, E. G. Friedman, P. M. Fauchet, "Alleviating Thermal Constraints while Maintaining Performance via Silicon-Based On-Chip Optical Interconnects", *Proceedings of the Workshop on Unique Chips and Systems*, pp. 45-52, March 2005.
9. G. Chen and E. G. Friedman, "Low Power Repeaters Driving RC Interconnects with Delay and Bandwidth Constraints," *Proceedings of the IEEE International SOC Conference*, pp. 335-339, September 2004.
10. M. Margala, R. Alonzo, G. Chen, B. J. Jasionowski, K. Kraft, M. Lay, J. Lindner, M. Popovich, J. Suss, "Low-voltage Power-efficient Adder Design," *Proceedings of IEEE Midwest Symposium on Circuits and Systems*, Vol. 3, pp. 461-464, August 2002 .

Patent Application

1. G. Chen and E. G. Friedman, "Efficient Algorithm for Simulating On-Chip RLC Interconnects" (patent pending).