

**HIGH PERFORMANCE CLOCK DISTRIBUTION
FOR HIGH-SPEED VLSI SYSTEMS**

A Dissertation

Submitted to

Graduate School of Information Sciences

of

Tohoku University

by

Xu Zhang

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

February 2008



TOHOKU
UNIVERSITY

To my family and my parents

ACKNOWLEDGMENTS

My college years at Tohoku University have exposed me to a variety of challenging, invigorating and enjoyable experiences and I would like to take this opportunity to thank all the wonderful teachers, colleagues, staff, family and friends whom I have been fortunate to interact with during my lifetime.

I first and foremost wish to thank my dissertation advisor, Professor Susumu Horiguchi, who has given me constant supports and provided a wonderful, stimulating environment in which I've completed this work. Professor Horiguchi has always had the great wisdom to know when to guide me and perhaps more importantly, when to let me find my way. Any future success will owe much to the example he has shown me.

I would like to express my gratitude to Assoc. Prof. Dr. Xiaohong Jiang, who has given me lots of valuable discussion and warm encouragements. Dr. Jiang is a patient teacher and I've always admired his serious attitude towards research.

I want to sincerely thank Dr. Masaru Fukushi for his continuous encouragement and help. I really enjoy talking (and drinking?) with him for his humor and generosity.

It is my pleasure to acknowledge Professor Michitaka Kameyama at Intelligent Integrated Systems Laboratory and Professor Takahiro Hanyu at Laboratory for Brainware Systems (RIEC) for serving on my Doctoral Advisory Committee. They have also given me many valuable comments.

I would like to give special thanks to Professor Ken-ichi Itakura at Muroran Institute of Technology for his enthusiastic and amiable encouragement and help.

I greatly appreciate all the members of Horiguchi Laboratory, who have graciously put up with me much and given me lots of help.

There have been a few people along the course of my career that deserve thanks for making me think, and/or giving me many supports as well as a chance at things I was willing to jump into: President Qisan Zhao (Golden-Monkey Group Co. Ltd., China), Mr. Yunhang Xu (the Chief Financial Officer of Golden-Monkey Group Co. Ltd.), President Changxiu Fang (Shanghai Chineway Pharmaceutical Technology Co. Ltd., China), Mr. Noriwa Ikeda (A-TiC Co. Ltd., Japan).

I should also mention Dr. Fangming Gong, Dr. Bianli Lu, and Ms. Rongli Gao for their encouragements and help.

I've had the privilege of meeting a few people, who have kindly helped and encouraged me as much as they could. These people are: Professor Yuqing Wang (Henan Polytechnic University); Dr. Shaohe Luo (Henan Polytechnic University); Dr. Haibin Yuan (Henan Polytechnic University).

In addition to people mentioned above, I owe to my friends and fellows at both Golden-Monkey Group Co. Ltd. (China) and A-TiC Co. Ltd. (Japan) who always helped me to their best and made my life easier.

Finally, I would thank my family for their support and encouragement throughout the 3 academic years. Without their support and devotion, I don't think I could've made it all the way to the end. This dissertation is also dedicated to my parents, who showed me that the odds are never too large and it never too late to take on new challenges.

TABLE OF CONTENTS

	Page
LIST OF TABLES	ix
LIST OF FIGURES	x
SYMBOLS	xiii
ABBREVIATIONS	xv
ABSTRACT	xvi
1 Introduction	1
1.1 Motivations	1
1.2 Research Purposes	2
1.3 Outline of the Dissertation	5
2 Clock Distribution and Clock Distribution Networks	7
2.1 Synchronous Digital Systems and Clocks	7
2.2 Clock Distribution in VLSI Systems	11
2.2.1 Timing Parameters of Clock Signal	12
2.2.2 Performance Metrics of Clock Distribution Networks	12
2.2.3 Conventional Clock Distribution Networks	16
2.2.4 Hierarchy of Clock Distribution	20
2.3 Performance Evaluation of Clock Distribution Network	22
2.3.1 Interconnect Delay Modeling	22
2.3.2 Extraction of Interconnects Parasitics Parameters	24
2.3.3 Statistical Timing Analysis	25
3 H-Tree Clock Distribution Networks in Presence of Process Variations and Inductance Effects	28
3.1 Introduction	28
3.2 Process Variations and Inductance Effects	30

	Page
3.3 Interconnect Delay Calculation	32
3.3.1 RC Delay Model	33
3.3.2 RLC Delay Model	33
3.3.3 Parasitics Extraction of Interconnects without considering coupling effects	35
3.3.4 Parasitics Extraction Considering Coupling Effects	35
3.4 Simulation Methodology	37
3.4.1 H-Tree CDNs	37
3.4.2 Simulation Setup	38
3.4.3 Parameters Setting	40
3.5 Verification of the simulation program	41
3.6 Performance Evaluation Results	43
3.6.1 Effects of Inductance and Process Variations on H-Tree CDNs	43
3.6.2 Coupling Effects on Performance of H-Tree CDNs	46
3.6.3 Impact of Spatial Dependence of Process Variations	48
3.6.4 Performance Sensitivity to the Magnitude of Process Variations	50
3.7 Conclusions	51
4 Variant X-Tree Clock Distribution Network	53
4.1 Introduction	53
4.2 X Architecture	55
4.3 Variant X-Tree CDN	57
4.3.1 Basic Unit of Variant X-Tree CDN	58
4.3.2 Construction Features of Variant X-Tree CDN	59
4.4 Statistical Performance Analysis Model	63
4.4.1 Statistical Performance Evaluation Model	63
4.4.2 Variance Estimation for a Branch in CDN	65
4.5 Methodology of Performance Evaluation	66
4.5.1 Process Variations	66

	Page
4.5.2 Interconnect Delay Calculation	67
4.5.3 Simulation Setup	68
4.5.4 Parasitics Extraction	69
4.5.5 Parameters Setting	70
4.6 Simulation Results	71
4.6.1 Performance Improvement of Variant X-Tree CDN	71
4.6.2 Statistical Performance Evaluation	73
4.7 Conclusions	75
5 Clock Distribution and its Performance Enhancement in 3D ICs	77
5.1 Introduction	77
5.2 3D ICs and Vias Placement/Insertion	81
5.3 Problem Formulation and Optimization	83
5.3.1 Delay Modeling	84
5.3.2 Impedance Matching	85
5.3.3 Optimization of Delay and Reflection Coefficient	87
5.4 Simulation Methodology and Setup	89
5.4.1 Delay Calculation	89
5.4.2 Parasitic Extraction of Vias	90
5.4.3 Parameters Settings	91
5.5 Simulation Results	91
5.5.1 Impact of vias sizing on Interconnect Delay	92
5.5.2 Overall Impacts of Vias Sizing/Placement on Interconnect Delay and Reflection Coefficient	93
5.5.3 Comparison Between Vias Sizing and Vias Placement in Delay Improvement	97
5.5.4 Impact of Vias Sizing and Buffer Insertion on Delay Improvement	99
5.6 Conclusions	101
6 Conclusion	102

	Page
6.1 Summary	102
6.2 Recommendations for Future Works	103
LIST OF REFERENCES	105
A Publications List	109

LIST OF TABLES

Table	Page
3.1 Mean values and standard deviations of major process parameters . . .	40
3.2 Simulation results of the level-2 H-Tree CDN	42
4.1 Mean values and standard deviations of major process parameters . . .	71
5.1 Delay results of interconnects when considering only the vias sizing . .	92
5.2 Delay results of interconnects	94
5.3 Results of reflection coefficient	95
5.4 The impedance characteristics of in a 6-Tires 3D IC	96
5.5 Optimum routing results for the 3mm example wire	96
5.6 Delay results of a 2.5mm wire in presence of different impedance characteristics	97
5.7 Delay results for vias sizing/placement and buffer insertion	100

LIST OF FIGURES

Figure	Page
2.1 The concept of finite-state machine.	8
2.2 State changes in the finite-state machine.	8
2.3 A sequential logic circuit with an edge-triggered flip-flop.	10
2.4 Timing diagram for an edge-triggered flip-flop	10
2.5 Clock timing parameters	12
2.6 Clock skew and jitter	14
2.7 An H-Tree CDN with level=8	18
2.8 A Variant X-Tree CDN with level=8	19
2.9 A mesh CDN	19
2.10 A trunk-based CDN	20
3.1 Equivalent circuit of driver-interconnect-load structure	33
3.2 Coupling capacitance between two adjacent wires	36
3.3 An H-Tree CDN for 64 PEs	37
3.4 Routing structure of clock wire and P/G wires	38
3.5 An H-Tree CDN with level=2	41
3.6 A branch circuit in an H-Tree CDN	42
3.7 Mean value of the maximum delay and skew in <i>CDN-G</i> (RC vs. RLC)	43
3.8 Standard deviations of the maximum delay and skew in <i>CDN-G</i> (RC vs. RLC)	44
3.9 Mean value of the maximum delay and skew in <i>CDN-L</i> (RC vs. RLC) .	44
3.10 Standard deviations of the maximum delay and skew in <i>CDN-L</i> (RC vs. RLC)	45
3.11 Mean value of the maximum delay and skew in <i>CDN-L</i> (RLC model) .	47
3.12 Standard deviations of the maximum delay and skew in <i>CDN-L</i> (RLC model)	47

Figure	Page
3.13 Mean value of the maximum clock delay and skew of <i>CDN-L</i> when considering spatial dependence or not (RLC model)	49
3.14 Standard deviations of the maximum delay and skew in <i>CDN-L</i> when considering spatial dependence or not (RLC model)	49
3.15 Mean value and standard deviation of the maximum delay and skew when process variations change in a Level-6 <i>CDN-L</i> (RLC model)	51
4.1 (a) Preferred direction in X Architecture (b) Non-preferred direction support in M4 and M5	56
4.2 Contrasting Manhattan (left) and Liquid Routing (right)	57
4.3 The basic unit of Variant X-Tree	58
4.4 A Variant X-Tree CDN with 64 sinks (level=6)	59
4.5 Wire intersection occurs in one layer when more levels are constructed	61
4.6 A Level-7 Variant X-Tree CDN	62
4.7 A hybrid CDN with 2 recursive levels Variant X-Tree and 1-level H-Tree link	62
4.8 Illustration of basic unit of variant X-Tree for statistical performance analysis	64
4.9 Equivalent circuit of driver-interconnect-load structure	67
4.10 Routing structure of clock wire and P/G wires	68
4.11 The maximum and minimum delay of CDNs (Variant X-Tree CDN vs. H-Tree CDN)	72
4.12 The clock skew of CDNs (Variant X-Tree CDN vs. H-Tree CDN)	73
4.13 Mean values of the maximum clock delay and clock skew of CDN-X	74
4.14 Standard deviation of the maximum clock delay and clock skew of CDN-X	74
5.1 A possible SoC design based on 3D IC.	78
5.2 The structure schematic of a 3-tiers' 3D IC fabricated with a face-to-back process (The thick vertical lines stand for through-hole vias for inter-tiers interconnections).	79
5.3 Redundant via insertion (line end extension)	82
5.4 Via sizing by redundant via insertion.	83

5.5 Equivalent circuit of driver-interconnect-load structure of inter-tier interconnect in 3D IC. 84

SYMBOLS

c_0	input capacitance of a minimum-sized repeater
c_p	parasitic capacitance of a minimum-sized repeater
C	capacitance
C_{int}	total interconnect capacitance
C_L	load capacitance
C_P	parasitics capacitance of driver (inverter/repeater)
d	delay
f	clock (signal) frequency
f_{ox}	interconnect oxide layer thickness
k	repeater size
L	inductance
L_{eff}	the effective channel length of gate
L_{int}	total interconnect inductance
M	mutual inductance
$P_{dynamic}$	dynamic power
r_0	output resistance of a minimum-sized repeater
R	resistance
R_s	the output resistance of driver (inverter/repeater)
R_{int}	total interconnect resistance
t	time
t_f	time-of-flight delay
t_{rise}	signal rise time
$t_{50\%}$	50% V_{DD} delay
t_{ox}	gate oxide thickness
T	clock (signal) period
V	voltage
V_{DD}	power supply voltage
V_T	threshold voltage

χ	clock skew (SSTA)
η	the maximum delay (SSTA)
ξ	the minimum delay (SSTA)
μ	charge carrier mobility

ABBREVIATIONS

AWE	asymptotic waveform evaluation
BPTM	Berkeley Predictive Technology Model
CDN	clock distribution network
CNT	carbon nanocube
DLL	delay-locked loop
ILD	interlayer dielectric
PE	processor element
PLL	phase-locked loop
SD	standard deviation
SSTA	statistical static timing analysis
STA	static timing analysis
FSM	finite-state machine

ABSTRACT

Zhang, Xu Ph.D., Tohoku University, February, 2008. High Performance Clock Distribution for High-speed VLSI systems. Major Professor: Susumu Horiguchi Professor.

In a synchronous digital system, the clock signal is used to define a time reference for the movement of data within that system. Since this function is vital to the operation of a synchronous system, much attention has been given to the characteristics of these clock signals and the networks used in their distribution, which can affect the overall performance of a VLSI system significantly. Therefore, this research is mainly devoted to studying high performance clock distribution for high-speed VLSI systems.

At first, we study the performance of H-Tree CDNs when both the process variations and inductance effects are jointly taken into account. The impacts of possible coupling effects between adjacent wires upon the performance of an H-Tree CDN are also be addressed. The simulation results show that: 1) the inductance effect can not be neglected for high-speed nanoscale VLSI systems; 2) process variations affect the performance of clock distribution network in different way. It implies that we should consider the different effects caused by process variations carefully.

Then, we propose a new clock distribution network (CDN), namely Variant X-Tree, based on the idea of X-Architecture proposed recently for efficient wiring within VLSI chips. The Variant X-Tree CDN keeps the nice properties of equal-clock-path and symmetric structure of the typical H-Tree CDN, but results in both a lower maximal clock delay and a lower clock skew than its HTree counterpart, as verified by an extensive simulation study that incorporates simultaneously the effects of process variations and on-chip inductance. We also propose a closed-form statistical models for evaluating the skew and delay of the Variant X-Tree CDN. The comparison between the theoretical results and the simulation results indicates that the proposed statistical models can be used to efficiently evaluate the performance of the variant X-Tree CDNs.

Finally, we extend the idea of redundant via insertion of conventional 2D ICs and propose an approach for vias insertion/placement in 3D ICs to minimize the propagation delay of interconnects with the consideration of signal integrity. The simulation results based on a 65nm CMOS technology demonstrate that our approach in general can result in a 9% improvement in average delay and a 26% decrease in reflection coefficient. It is also shown that the proposed approach can be more effective for interconnects delay improvement when it is integrated with the buffer insertion in 3D ICs.

Chapter 1

INTRODUCTION

1.1 Motivations

In a synchronous digital system, the clock signal is used to define a time reference for the movement of data within that system. Since this function is vital to the operation of a synchronous system, much attention has been given to the characteristics of these clock signals and the networks used in their distribution. Clock signals are often regarded as simple control signals; however, these signals have some very special characteristics and attributes. Clock signals are typically loaded with the greatest fanout, travel over the longest distances, and operate at the highest speeds of any signal, either control or data, within the entire system. Since the data signals are provided with a temporal reference by the clock signals, the clock waveforms must be particularly clean and sharp. Furthermore, these clock signals are particularly affected by technology scaling, in that long global interconnect lines become much more highly resistive as line dimensions are decreased. This increased line resistance is one of the primary reasons for the growing importance of clock distribution on synchronous performance. Finally, the control of any differences in the delay of the clock signals can severely limit the maximum performance of the entire system as well as create catastrophic race conditions in which an incorrect data signal may latch within a register.

The clock distribution networks (CDNs) are used to distribute clock signal to synchronize the data flows among different data paths. The most common strategy for distributing clock signals in VLSI systems is to insert buffers either at the clock source and/or along a clock path, forming a tree-based structure called as clock distribution network. Thus, the unique clock source is usually considered as the root of the tree, the initial portion of the tree as the trunk, individual paths driving each

register as the branches, and the registers being driven as the leaves. Additionally, a mesh version of tree structure or hybrid structure of tree and mesh can also be adopted in practical cases.

Since the performance of a VLSI system is usually determined by its clock frequency, the CDN can significantly affect the overall system performance and reliability [27]. Unfortunately, the performance of a clock distribution network is vulnerable to many factors, such as the systematic or random process variations, the interconnect layout-dependent coupling effects, on-chip inductance effects, IR-drop, etc.

The clock period of a CDN is in general determined by both the clock skew and the maximal clock delay of the network. To evaluate the performance of a CDN, we usually need to study both the maximum clock delay and clock skew. Here the clock skew is defined as the difference between the maximum clock delay and the minimum clock delay among all clock paths (interconnects) in the CDN. The clock skew arises mainly from unbalanced delays due to the unequal clock path lengths between clock source and different modules as well as from process variations that cause clock path delay variations [27].

To deal with the clock skew problem, a common way is to adopt the well-balanced H-Tree [5] or mesh CDNs. For a well-balanced CDN, the uncontrollable clock skew mainly comes from the variations in process parameters that affect the interconnect impedance/capacitance and, in particular, any distributed buffer amplifiers [5, 34]. Extensive research efforts have been devoted to studying and modeling the impacts of process variations upon the clock skew of an H-Tree CDN, see, for example, [21, 33, 34].

The evolution of VLSI chips towards larger die size and faster clock speed makes the clock distribution an increasingly important issue. In particular, the performance of CDNs suffers from some unavoidable reasons such as process variation much more nowadays. Therefore, this research is mainly devoted to exploring the real performance of a CDN in presence of different factors, designing new clock distribution networks under available technologies, and studying new methods for clock distribution oriented to high-speed nanoscale VLSI systems.

1.2 Research Purposes

In this research, we focus mainly on three important issues related to clock distribution networks. They are listed as follows.

H-Tree Clock Distribution Networks in Presence of Process Variations and Inductance Effects

With the dramatic increase in clock speed and a significant decrease in the feature size in current high-speed VLSI circuits, the on-chip line inductive effects (in addition to the resistance and capacitance) will significantly affect the interconnect delay and cannot be neglected anymore in the performance evaluation of CDNs, especially for the CDNs with long interconnect lengths and a small signal rise time [12]. Furthermore, adopting the new materials with a low-resistance (e.g. copper) in VLSI fabrication makes inductive effects more significant for a global clock distribution network, where many long and wide wires are usually required. Therefore, the inductive effects of interconnects should be carefully addressed in the performance evaluation of modern high performance CDNs.

In the manufacturing process of a VLSI system, some uncertainties (process variations) may arise due to the parameter fluctuations of devices or environment, which make the overall performance of the system varies with these inherent and unavoidable fluctuations. In nanoscale process or deep sub-micron (DSM) process, the parameter fluctuations impose a growing threat to the system performance, especially for the gigascale interconnection systems where the polysilicon gate length has decreased below the wavelength of light used in the optical lithography process. It is predicted that in a 130nm technology, the variation magnitude in gate length of n-MOS and p-MOS can be as high as 35% [49].

Although many papers about inductance effects or process variations have been published, to the best of our knowledge; however, no results have been reported about the performance evaluation in terms of maximum clock delay and clock skew of an H-Tree CDN when both the inductance effects and process variations are simultaneously considered. Thus, this research is devoted to the real performance evaluation of a high-speed H-Tree CDN when both the process variations and inductance effects are jointly considered. In particular, we will study the impact of possible coupling effects (including inductance coupling effect and capacitance coupling effect) between adjacent interconnects upon the performance of an H-Tree CDN. We will also investigate how the spatial dependent components and the magnitude of process variations affect the performance of an H-Tree CDN.

Variant X-Tree Clock Distribution Network

The critical issues concerning with the design of clock distribution network are to achieve a low clock delay and the minimum or a useful skew in most cases with the minimum buffer size and wire length. The well-balanced H-Tree CDN has been widely adopted to eliminate the skew caused by unequal clock path lengths [21], where the uncontrollable clock skew mainly comes from the variations in process parameters that affect the interconnect impedance/capacitance.

Although H-Tree is attractive for clock distribution due to its small clock skew and a relatively simple implementation, it usually results in a long clock path from the clock source to each sink (clock terminal). Thus, an H-Tree CDN usually causes a higher clock delay.

Mesh or grid is also a popular architecture for distributing clock signals on a chip. It uses inherent redundant interconnects created by loops to smooth out undesirable variations between signal nodes spatially distributed over the chip, and thus results in a lower clock skew. However, the mesh/grid CDN usually occupies a larger wiring area, and consumes more power. Such a condition is becoming worse with the increase of modern VLSI chips' area moreover.

Recently, X Architecture was proposed to wire in a VLSI chip with considerably shorter wiring length than that of traditional Manhattan wiring architecture [4]. It has been demonstrated in [4] that the X Architecture, which supports 45- and 135-degree wires as well as the vertical and horizontal wires, can reduce as high as 29% of the wire length required by the simple Manhattan wiring architecture. As a result, the X Architecture becomes promising to considerably reduce the delay and improve the overall performance of on-chip interconnects.

Consequently, in this research, we extend the X Architecture to clock distribution and intend to propose a novel non-orthogonal clock distribution network, which is able to achieve both a lower maximal clock delay and a lower clock skew than its H-Tree counterpart.

Redundant Vias Insertion for Performance Enhancement in 3D ICs

Three dimensional (3D) integrated circuits (ICs), which comprise multiple tiers of active devices, have the potential to enhance VLSI chip performance, functionality

and device packing density [55]. For example, 3D ICs offer an attractive alternative to conventional 2D planar ICs: they can combine different technologies such as analog and digital circuits within the single chip cube to construct a multi-tier (multi-plane) system. Thus, using 3D ICs allows for integrating the best technology for a particular portion of an application into one chip package.

By expanding vertically rather than spreading in 2D planar area, obviously the delay of signal propagation in interconnects can be decreased due to the decreased length of interconnects in 3D ICs. Thus, the drawback of long interconnects in conventional 2D ICs can be alleviated. Additionally, 3D ICs technology can also result in a reduction of total active power, noise improvement and a greater logical span.

In 3D ICs, signal paths like wires for global clock distribution consist of multiple-segment interconnects routed in different tiers and some vertical inter-tier interconnects implemented by vertical through-hole vias (abbreviated as vias hereafter). Since each tier in 3D ICs may be fabricated with different technologies or processes, the impedance characteristics of different segments of the global interconnects may be disparate [55]. Furthermore, the impedance characteristics of vias may also be different from that of horizontal wires.

Interconnects delay is crucial to the performance of modern digital VLSI systems, since it is a big fraction of total delay and it is in general increasing with technology scaling. To maximize performance improvement in 3D ICs, some approaches [44, 58] related to vias placement and wire routing have been proposed in presence of non-uniform impedance characteristics of interconnects. However, the important signal integrity issues due to the non-uniform impedance characteristics were not addressed. In this research, we extend the redundant via insertion in [38] to 3D ICs and propose an approach of redundant via placement/insertion to minimize the total delay for inter-tier or global interconnects. The issue of signal integrity due to the non-uniform impedance characteristics of interconnects is also carefully addressed.

1.3 Outline of the Dissertation

This dissertation is organized as follows.

We review the techniques of clock distribution in digital VLSI systems in Chapter 2, clock distribution networks, the delay modeling of interconnects and performance evaluation of CDN are also introduced in this chapter. In Chapter 3, we explore

the real performance of nanoscale high-speed H-Tree CDNs in presence of process variations and inductance effect. A novel X Architecture based Variant X-Tree CDN will be introduced in Chapter 4, where its performance simulation results are also given. We propose an approach about redundant vias insertion to for performance enhancement in 3D ICs in Chapter 5. The corresponding simulation results are provided in this chapter as well. Finally, we summarize this research in chapter 6. We also suggest the future works of this research in Chapter 7.

Chapter 2

CLOCK DISTRIBUTION AND CLOCK DISTRIBUTION NETWORKS

This chapter provides an overview of on-chip electrical clock distribution networks. It begins with a review of synchronous, digital systems and the role of the clock in these systems. Next, the performance metrics of a clock distribution network are introduced. The conventional clock distribution styles used for most microprocessors and digital ASICs are discussed. The performance evaluation of a clock distribution network is also provided in this chapter.

2.1 Synchronous Digital Systems and Clocks

The notion of clock and clocking is essential for the concept of synchronous design of digital systems. The synchronous system assumes the presence of the storage elements and combinational logic, which together make up a finite-state machine (FSM). The changes in the FSM are in general the result of two events: clock and input signal changes, as illustrated in Fig.2.1.

The next state, S_{n+1} , is a function of the present state, S_n , and the logic value of the input signals: $S_{n+1} = S_{n+1}(S_n, X_n)$. The remaining question is: When in time will FSM change to the next state, S_{n+1} . This change is determined by the type of clocked storage elements used and the clock signal. The function of the clock signal is to provide a reference point in time when the FSM changes from the present, S_n , to the next state, S_{n+1} . This process is illustrated in Fig. 2.2.

In Fig.2.2, we have implicitly assumed that the moment when the state changes from S_n to S_{n+1} is determined by the change in the clock signal from logic “0” to logic “1.” In fact, this change is determined by the type of clock storage element and its

In a synchronous digital system, the clock signal is used to define a time reference for the movement of data within that system as mentioned above. Since this function is vital to the operation of a synchronous system, much attention has been given to the characteristics of these clock signals and the networks used in their distribution. Clock signals are often regarded as simple control signals; however, these signals have some very special characteristics and attributes. Clock signals are typically loaded with the greatest fanout, travel over the longest distances, and operate at the highest speeds of any signal, either control or data, within the entire system. Since the data signals are provided with a temporal reference by the clock signals, the clock waveforms must be particularly clean and sharp. Furthermore, these clock signals are particularly affected by technology scaling, in that long global interconnect lines become much more highly resistive as line dimensions are decreased. This increased line resistance is one of the primary reasons for the growing importance of clock distribution on synchronous performance. Finally, the control of any differences in the delay of the clock signals can severely limit the maximum performance of the entire system as well as create catastrophic race conditions in which an incorrect data signal may latch within a register.

Most synchronous digital systems consist of cascaded banks of sequential registers with combinatorial logic between each set of registers. The functional requirements of the digital system are satisfied by the logic stages. The global performance and local timing requirements are satisfied by the careful insertion of pipeline registers into equally spaced time windows to satisfy critical worst case timing constraints. The proper design of the clock distribution network further ensures that these critical timing requirements are satisfied and that no race conditions exist [8, 26]. With the careful design of the clock distribution network, system-level synchronous performance can actually increase, surpassing the performance advantages of asynchronous systems by permitting synchronous performance to be based on average path delays rather than worst case path delays, without incurring the handshaking protocol delay penalties required in most asynchronous systems.

One example of how the clock is used to control data within a microprocessor is the sequential logic circuit shown in Fig.2.3. The edge-triggered flip-flop used in this circuit samples the input data on a rising clock edge and holds it steady at the output. The timing diagram for an edge-triggered flip-flop is shown in Fig.2.4. The

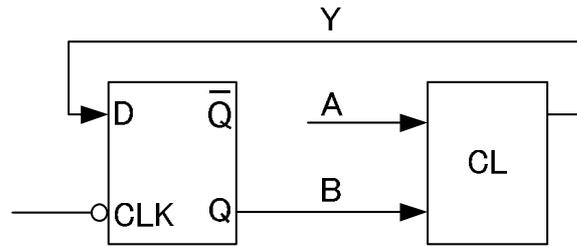


Fig. 2.3. A sequential logic circuit with an edge-triggered flip-flop.

data must be held constant at the input for a setup time, t_s , prior the rising clock edge and a hold time, t_h , after the clock edge. The previous data value, x , remains at the output of the flip-flop for at least a contamination delay, t_c , after the rising clock edge. The sampled data, y , appears at the output after a propagation delay, t_d , at most.

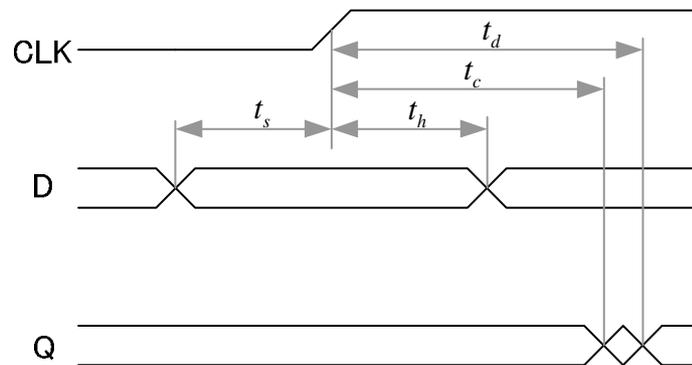


Fig. 2.4. Timing diagram for an edge-triggered flip-flop

The operation of the sequential logic circuit can be understood by following the data for one clock cycle. When the clock goes high, data (B) is sampled by the flip-flop and reaches the input of the combinational logic (CL) block along with timed data from elsewhere in the circuit (A). After the logic delay, t_{dBY} , the result of the combinational logic (Y) wraps around to the input of the flip-flop and is ready to be sampled again at the next rising clock edge.

The performance and reliability of a synchronous microprocessor depend on the long and short-paths. The long-path determines the maximum achievable clock rate for a microprocessor. In practice, long-path errors can be remedied by simply slowing

down the clock. Short-path errors, however, cannot be fixed in this way and generally require delay buffering in the path. Because fixing short-path errors can require modifying of the circuitry, they are particularly costly during the design cycle.

The correct operation of a synchronous circuit depends on the accuracy of the clock signal that is delivered to each of the timing circuits. Variations in the clock period at a single flip-flop (such as in Fig.2.1) or in the arrival time between two flip-flops that share data can result in long- and short-path errors. Therefore, accurately distributing the clock signal to the timing circuits is of vital importance to the operation of the microprocessor.

2.2 Clock Distribution in VLSI Systems

To distribute clock signal, many different approaches have been developed for designing clock distribution networks (CDNs) in synchronous digital integrated circuits [27]. The requirement of distributing a tightly controlled clock signal to each synchronous register on a large non-redundant hierarchically structured integrated circuit within specific temporal bounds is difficult and problematic. Furthermore, the tradeoffs that exist among system speed, physical die area, and power dissipation are greatly affected by the clock distribution network.

The most common strategy for distributing clock signals in VLSI systems is to insert buffers either at the clock source and/or along a clock path, forming a tree-based structure called as clock distribution network. Thus, the unique clock source is usually considered as the root of the tree, the initial portion of the tree as the trunk, individual paths driving each register as the branches, and the registers being driven as the leaves. Additionally, a mesh version of tree structure or hybrid structure of tree and mesh can also be adopted in practical cases.

In this section, we first show the basic timing parameters of clock signal, and then present the performance metrics of clock distribution networks. We also introduce the conventional topologies of clock distribution networks in Section 2.2.3. The hierarchy of clock distribution is discussed in this section.

2.2.1 Timing Parameters of Clock Signal

The clock signal is characterized by its period, T , which is inversely proportional to the clock frequency, f . The time during which the clock is active (assuming logic 1 value) is defined as clock width, W . The clock duty cycle w is defined as the ratio of clock width and clock period, i.e.,

$$w = \frac{W}{T} \quad (2.1)$$

Usually, the clock signal has a symmetric shape, which implies a 50% duty cycle. This is also the best we can expect, especially when distributing a high-frequency clock. Another important point is the ability to precisely control the duty cycle. This point is of special importance when each phase of the clock is used for logic evaluation, or when we trigger the clock storage elements on each edge of the clock. Some recently reported work demonstrates the ability to control the duty cycle to within $\pm 0.5\%$.

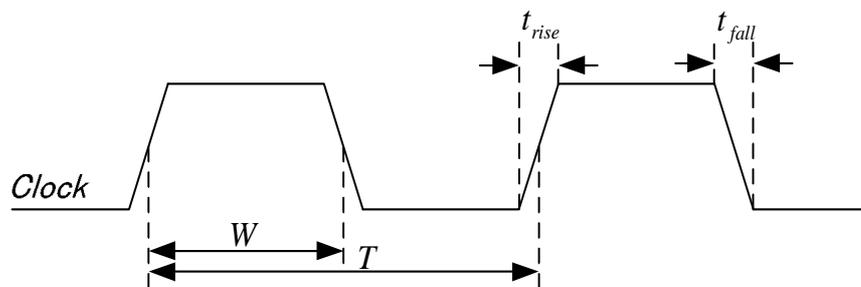


Fig. 2.5. Clock timing parameters

The timing parameters of clock signal are illustrated in Fig.2.5 where t_{rise} (t_{fall}) is called the signal rise (fall) time.

2.2.2 Performance Metrics of Clock Distribution Networks

As mentioned in the previous sections, clock signals are typically loaded with the greatest fanout, travel over the longest distances, and operate at the highest speeds of any signal within the entire system. It is very important to evaluate the performance of clock distribution networks. Generally, the main performance metrics of a clock distribution network are the maximum delay and timing uncertainties (i.e., skew and jitter), which are our crucial concerns when designing a clock distribution network.

In addition, the power dissipation of clock distribution networks becomes important with the increase in low-power requirements of VLSI devices and systems. All of these performance metrics have a significant impact on the design of VLSI systems and its eventual performance and reliability.

Delay and the Maximum Delay

Clock signal is usually generated externally (e.g. by PLL). By means of a clock distribution network, clock signal starts from the clock source, then propagates along different clock paths, and finally arrives at each sink (terminal) that consumes clock signal. Such a procedure naturally results in certain propagation time d which is called delay.

Thus, the maximum delay d_{\max} is defined as the maximum value among all the delay of propagation paths in a clock distribution network, i.e.,

$$d_{\max} = \max\{d_i | i = 1, \dots, n\} \quad (2.2)$$

where n is the total clock sinks, number in the clock distribution network. Similarly, the minimum delay d_{\min} of a clock distribution network is defined as:

$$d_{\min} = \min\{d_i | i = 1, \dots, n\} \quad (2.3)$$

The maximum delay is critical to the traditional clocking mode (CDN) VLSI systems, because it determines the maximum available running frequency of the VLSI systems. Since the signal period should be greater than the maximum delay, and it is a main contributor to determine clock period (see the next section for detail), it is definite that a bigger d_{\max} means a slower available running frequency.

Timing Uncertainties

In a synchronous digital VLSI system, the clock signal regulates the flow of data through the VLSI systems. Therefore, the system reliability and stability depend significantly on the ability to accurately relay a clock signal to millions of individual circuit cells. Any timing error introduced by the clock distribution has the potential to cause a functional error, particularly if it comes in the long- or short-path. Therefore,

the timing uncertainty of the clock signal must be estimated and taken into account in the design stage.

The timing uncertainties in clock distribution networks are categorized as clock skew and clock jitter [13]. The former one refers to the static difference in clock arrival time between two clock sinks in the clock network. The later one refers to dynamic timing uncertainties at a single clock load.

Clock skew is generally caused by unbalanced clock paths or parameters variations in either devices or interconnects within the clock distribution networks or by static temperature or voltage variations around the die. It is a *spatial variation* of clock signal as distributed through the system [43]. An ideal clock distribution would have zero skew, although in practice it is sometimes beneficial to intentionally skew the clock to speed-up specific paths in the design [25,39]. For the purposes of this work, clock skew refers to the unintended differences in clock timing.

The *temporal variation* of a signal edge at a given point on a chip is called jitter [13,43]. The key measure of jitter for a VLSI system is the period or cycle-to-cycle jitter, which is the difference between the nominal cycle time and the actual cycle time. The total clock jitter is the sum of the jitter from the clock source and from the clock distribution. Power supply noise causes jitter in both the clock source and the distribution. Crosstalk noise also adds jitter to the clock distribution. Although clock skew can be helpful or harmful, clock jitter is always bad.

The illustration of clock skew and jitter are portrayed in Fig.2.6.

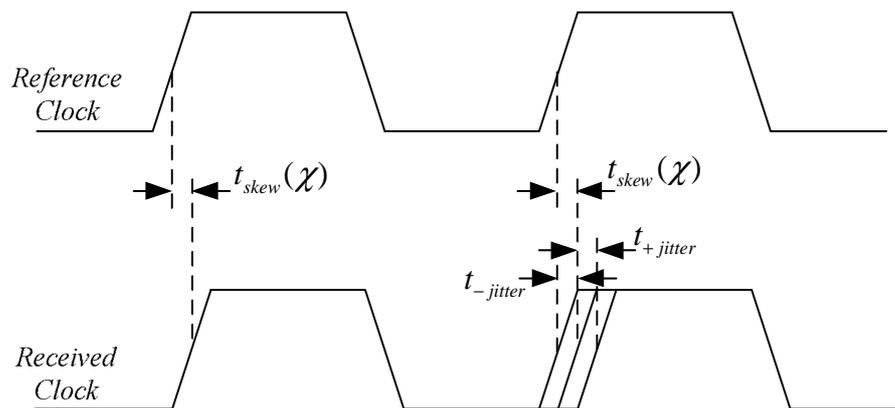


Fig. 2.6. Clock skew and jitter

The clock skew of CDN is the main concern in this work, which is the maximum difference in the arrival time of a clock signal at any two different sinks of a CDN. In other word, the clock skew χ of a CDN is also referred to as the difference between the maximum clock delay and the minimum clock delay among all clock paths in the CDN, i.e.,

$$\chi = d_{\max} - d_{\min} \quad (2.4)$$

The clock skew can also affect the performance of a VLSI system. For example, the clock period T for a traditional clocking mode H-Tree CDN should satisfy the following inequality

$$T \geq d_{\max} + \chi + t_{su} + t_{ds} \quad (2.5)$$

where t_{su} is setup time of the synchronizing elements, t_{ds} is propagation delay within the synchronizing element. While for a pipelined clocking mode H-Tree CDN, the clock period T will be

$$T = \max\{2 \times t_{segments}, 10 \times \chi\} \quad (2.6)$$

by the 10% rule of thumb relating the skew to clock period [42] where $t_{segments}$ is the delay of one stage.

Power Dissipation

In a modern VLSI system, the clock distribution network may drive thousands of registers, creating a large capacitive load that must be efficiently sourced and running at the highest speed. Furthermore, each transition of the clock signal changes the state of each capacitive node within the clock distribution network, in contrast with the switching activity in combinational logic blocks, where the change of logic state is dependent on the logic function. The combination of large capacitive loads and a continuous demand for higher clock frequencies has led to an increasingly larger proportion of the total power of a system dissipated within the clock distribution network, in some applications much greater than 25% of the total power [23, 24].

The battery life in portable devices is proportional to their energy consumption. In high-performance designs, energy consumption has a large impact on the design and may limit performance. It is therefore imperative to design the digital circuits, used in consumer products, that consume the minimum amount of energy for a given task.

The primary component of power dissipation in most CMOS-based digital circuits is dynamic power $P_{dynamic}$, which can be expressed as

$$P_{dynamic} \propto CV_{DD}^2 f_{clk} \quad (2.7)$$

where C is load capacitance, V_{DD} is the voltage of power supply, and f_{clk} is the clock frequency, respectively. So it is possible to reduce dynamic power $P_{dynamic}$ by lowering the clock frequency, the power supply, and/or the capacitive load of the clock distribution network. Lowering the clock frequency, however, conflicts with the primary goal of developing high-speed VLSI systems. Therefore, for a given circuit implementation low dynamic power dissipation is best achieved by employing certain design techniques that either minimize the power supply and/or the capacitive load.

2.2.3 Conventional Clock Distribution Networks

The design methodology and structural topology of the clock distribution network should be considered in the development of a system for distributing the clock signals. Various clock distribution strategies have been developed.

Historical Overview

Usually a clock signal was generated using a quartz-crystal-controlled oscillator that provides an accurate and stable frequency. Given the size limitation of the quartz crystal, the frequency of such a generated clock signal cannot be very high, and frequencies in excess of 30-50 MHz are rarely generated using a quartz crystal. The clock signal is then conditioned and amplified to reach desirable driving strength before it is applied to the outside pins of a VLSI chip, from which it drives an internal PLL or DLL. Before reaching the boundaries of the VLSI chip, adjustments to its shape and form are possible. In contrast, in older computer systems, which consisted of several electronic cabinets distributed over the computer floor, and which contained a number of printed circuit boards, adjustments to the clock signal were made at each level. Thus, the clock signals were distributed over longer distances and over several levels, including the cabinet, printed circuit boards, and internal modules. Those separate entities entered by the clock signal were referred to as “logic islands”, in which some adjusting elements make it possible to control the timing of the leading

as well as the trailing edge of the clock signal, and to produce an early as well as late clock signal with reference to the nominal clock.

With the advent of integration, the systems have shrunk dramatically in size. Today, it is quite common for a processor to have several levels of cache memory contained entirely on a VLSI chip. The chips capacity for hundreds of millions of transistors makes it possible to integrate not only one processor but also a multi-processor system onto a single chip. The inability to introduce tuning elements on the chip further aggravates the problem of distributing the clock signals precisely in time, since it is not possible to make further manual adjustment to the clock signal once it has crossed the boundaries of the VLSI chip. Therefore, careful planning and design of the on-chip clock distribution network is one of the most critical tasks in high-performance processor design.

Clock Distribution Networks in Modern VLSI Systems

Typically, the clock signal has to be distributed to several hundreds of thousands of the clocked storage elements (flip-flops and latches) on a complex processor chip. Clock paths in a CDN can differ in several attributes, such as the length of the path (wire), the physical properties of the material along different paths, the differences in clock buffers on the chip as a consequence of the process variations. The negative effect of these variations on the synchronous design is that different points on the chip will receive the clock signal at different moments, which results in a further increase in both local and global clock uncertainties.

There are several methods for the on-chip clock signal distribution that attempt to minimize the clock skew and to contain the power dissipated by the clock system. They are introduced as follows.

The most common clock distribution network is the tree, where buffers are inserted along the clock distribution path forming a tree structure. An H-Tree CDN is illustrated in Fig.2.7 where the inverters inserted in each branch are not portrayed. In this approach, the primary clock driver is connected to the center of the main “H” structure. The clock signal is transmitted to the four corners of the main “H”. These four close to identical clock signals provide the inputs to the next level of the H-tree hierarchy, represented by the four smaller “H” structures. The distribution process continues through several levels of progressively smaller “H” structures.

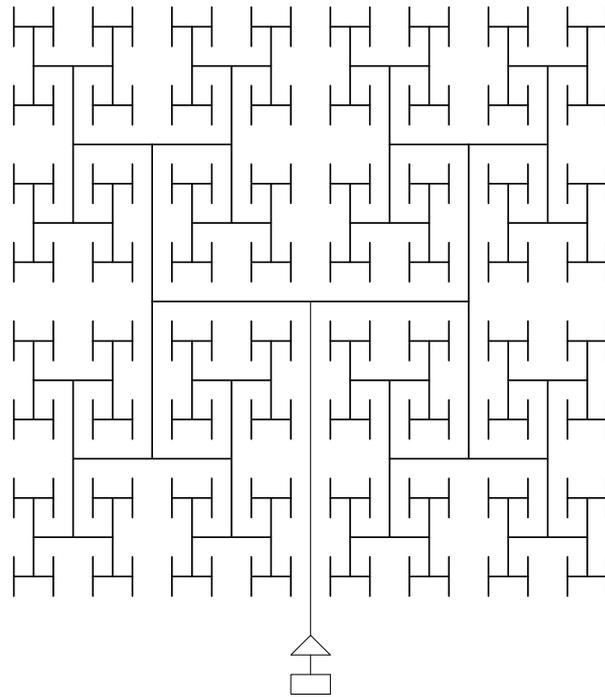


Fig. 2.7. An H-Tree CDN with level=8

Similar to H-Tree CDN, Variant X-Tree CDN (Fig.2.8) is also proposed in this work based on the X Architecture (see Chapter 4 for detail).

Both H-Tree and Variant X-Tree are well-balanced CDNs, all paths from the root (clock source) of tree to all branches have an identical number buffers and the same length, although the buffers size must be adjusted to match different loads. As a result, the propagation delay is almost same, which results in a small clock skew. For a well-balanced tree-based CDN, the uncontrollable clock skew mainly comes from the variations in process parameters that affect the interconnect impedance/capacitance and, in particular, any distributed buffer. However, the longer clock path in tree-based CDN usually causes a bigger delay.

Mesh or grid (see Fig.2.9) is also a popular architecture for distributing clock signals on a chip. It uses inherent redundant interconnects created by loops to smooth out undesirable variations between signal nodes spatially distributed over the chip, and thus results in a lower clock skew. However, the mesh/grid CDN usually occupies a larger wiring area, and consumes more power.

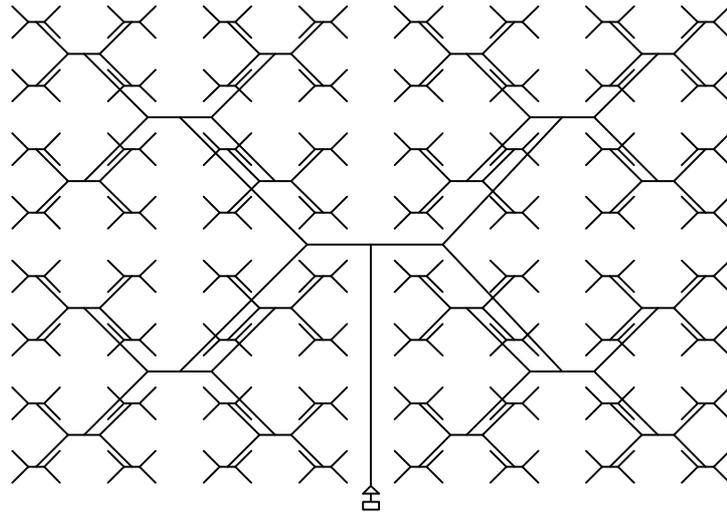


Fig. 2.8. A Variant X-Tree CDN with level=8

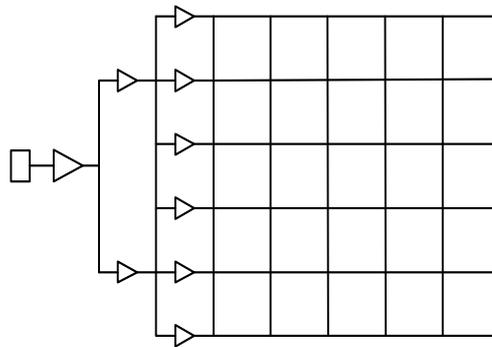


Fig. 2.9. A mesh CDN

In addition, one imposing problem that has limited the applicability of mesh architectures is the difficulty in analyzing them with sufficient accuracy. The main reasons are the huge number of circuit nodes needed to accurately model a fine mesh in a large design and large number of metal loops present in the mesh CDNs. As a result, circuit simulators such as SPICE either require inordinate amount of memory or run-time. In fact, HSPICE (Synopsys) and HSIM (Nassda) failed to analyze even coarse meshes for an industrial design [17].

Note that the H-Tree and X-Tree clock distribution networks are difficult to implement in those VLSI-based systems which are irregular in nature, such as some customized VLSI systems (e.g., ASIC systems). To deal with irregular structure, buffered

tree topologies integrated with structured custom design methodologies should be used in the design of the clock distribution networks in order to maximize system clock frequency, minimize clock delay, and control any deleterious effects of local (particularly, negative) clock skew. Fig.2.10 shows a trunk-based CDN.

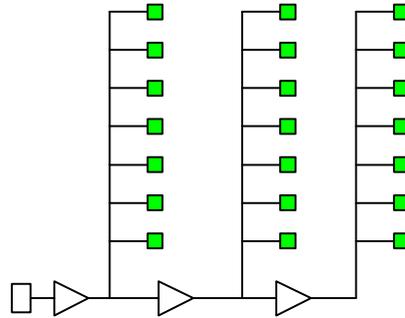


Fig. 2.10. A trunk-based CDN

2.2.4 Hierarchy of Clock Distribution

For the purposes of designing and analyzing clock networks, it is useful to group the network into different levels: global, regional, and local. Because the clock distribution is a fanout network that gets progressively larger in area and complexity, the design issues for each level are distinct. This section highlights the properties of these three levels as they relate to the clock figures-of-merit and explains why different topologies are suited for each level of the clock distribution.

Global Clock Distribution

The global clock distribution spans from the global clock buffer to the inputs of the sector buffers. This level of the distribution tends to be the longest in span because it relays the clock signal from some central point on the die to the sector buffers located throughout the die. The issues in designing the global tree mostly relate to signal integrity maintaining a fast edge rate over long wires while not introducing a large amount of timing uncertainty. Skew and jitter accumulate as the clock signal propagates through the clock network and both tend to accumulate proportional to the latency of the path. Because most of the latency occurs in the global clock

distribution, this is also a primary source of skew and jitter. On the other hand, the global network drives a small amount of the total capacitance in the clock network because it is at the upper portion of the fanout network. Therefore, relatively little power is dissipated at this level of the distribution. From a design point of view, achieving low timing uncertainty is more critical than reducing power dissipation.

Practically all high-performance microprocessors use some form of length-matched tree to distribute the clock at the global level. A tree is just a fanout of interconnects and buffers similar to what is illustrated in Fig.2.5. Often the global level of a clock distribution is completely symmetric to simplify the design and provide nominally low skew (assuming no loading variations). The most commonly used topology is a buffered H-Tree which is shown in Fig.2.7. H-Trees are an extremely popular choice at this level of the clock distribution because they are a simple pattern that efficiently and symmetrically covers large areas. H-Trees minimize the distance from the PLL to the furthest clock pins (without using diagonal traces) and thereby minimize the interconnect latency. The number of buffer levels within the global network which also determines the latency depends on the signal dispersion, loss and on the required power fanout.

Local Clock Distribution

The final level of a clock distribution network is the local level, which is the portion of the network that follows the clock pin. This network drives the final loads of the clock distribution and hence consumes the most power. As a rule of thumb, the power at the local level is about one order of magnitude larger than the power in the global and regional levels combined, with the only notable exceptions being clock networks that use a low-impedance grid at the regional level.

The layout of the local grid is generally included in the design of the macro block and is not the responsibility of global and regional clock network designer. Because of its relatively limited span, it is sufficient to use automatic layout for this portion of the clock network. Due to the irregular nature of most macros, the layout is generally a nonsymmetrical tree which may be length-matched depending on the skew goals for the distribution. Furthermore, the clock distributed to the clocked storage elements is often a derivative clock generated locally from the global clock signal.

2.3 Performance Evaluation of Clock Distribution Network

In this research, we focus mainly on the timing performance of clock distribution networks¹, and introduce timing performance evaluation of clock distribution networks in this section.

2.3.1 Interconnect Delay Modeling

To evaluate the performance of clock distribution network, it is necessary to calculate the total delay that clock signals propagate from clock source to sinks. As the physical dimensions in VLSI technologies scale down, interconnect delay increasingly dominates gate delay in determining circuit performance. In this subsection, we will outline several models for interconnect delay calculation.

Elmore Delay

In the past, many interconnect delay models have been proposed by analyzing the moments of the impulse response. Asymptotic waveform evaluation (AWE) is a generalized approach to response approximation by moment matching. It is very accurate but computationally very expensive. Hence, many moment-matching variants using the first two to four moments have been proposed. Those variants are relatively much more efficient but less accurate. Nevertheless, they may still be too expensive to be used within the tight optimization loops of design synthesis and layout tools. Moreover, for all the models above, the delay is either computed by an iterative procedure or expressed as a sophisticated implicit function of the design parameters. Sensitivity information cannot be easily calculated. Therefore, these models provide little insight into determining the design parameters during design or optimization.

As a result, the Elmore delay model, which is the first moment of the impulse response, is the most widely used interconnect delay model during design synthesis and layout. It can be written as a simple, closed-form expression in terms of design parameters. It is extremely efficient to compute and it provides useful insight for optimization algorithms.

¹Low power dissipation is an important requirement to clock distribution network.

Some literatures have proven that the Elmore delay is the upper bound on the actual 50% V_{DD} delay of an RC tree with a ramp input applied and it always overestimates the delay [28]. To further improve the accuracy of the Elmore delay model as well as to extend prediction to include more characteristics of switching, several extended version of Elmore delay model has been proposed. For example, the 50% delay ($t_{50\%}$) of a distributed RC line can be expressed as follows [47]:

$$t_{50\%} = R_s(0.693C_{int} + 0.693C_L) + R_{int}(0.377C_{int} + 0.693C_L) \quad (2.8)$$

where R_s is the output resistance of driver, C_L is the load capacitance, R_{int} and C_{int} are the total resistance and capacitance of the interconnect, respectively.

Distributed RLC Delay Modeling

Although RC delay model is simple and computation-efficient for timing analysis, it is only applicable to highly resistive nets at local and intermediate layers for low-speed clock signals. As the interconnect length and operating speed entered the nanoscale regime and gigascale regime, respectively, inductance component becomes comparable to resistance component in VLSI circuits (especially for Cu-based interconnect technology since the copper has a low resistance). Consequently, interconnect wires on high-speed VLSI chips exhibit more transmission line effects, so we should study the timing characteristics with transmission line theory for high-speed interconnects. To design these wires properly in high-performance circuits, RLC models and techniques for characterizing the signal transportation are required.

According to [11], transmission line behavior becomes significant when the signal rise time t_{rise} is less than or comparable to the transmission line time-of-flight delay t_f . Here the signal rise time is defined as the time required for the signal to move from 10% to 90% of its final value, and the time-of-flight is expressed as

$$t_f = \frac{l}{v} \quad (2.9)$$

where l is the line length, and v is the propagation speed.

In contrast to RC interconnects, RLC interconnects behave differently during the propagation of voltage switching; they have an increased delay, faster slew rate, and ringing as well as overshoot. These undesirable characteristics may cause undesirable effects: ringing affects the signal stability of clock since large oscillation can sensed

erroneously as a transition, thus causing a logic fault; while voltage overshoots may increase power consumption and degrade the reliability of the gate oxide as well as of the overall devices. To snap these possible effects, the inclusion of inductance in timing and noise analyses is necessary, which increases complexity of delay model significantly.

Similar to the RC delay model, RLC timing can be performed using either general numerical techniques or analytical solutions. In [12], a distributed RLC delay model with the consideration of parasitic capacitance of driver has been proposed. The 50% delay can be obtained using numerical method (Newton-Raphson method). While in [30], the authors proposed a closed-form solution of 50% delay based on curve-fitting.

2.3.2 Extraction of Interconnects Parastics Parameters

Parasitic effects are becoming more critical with the increase in performance, density, complexity, and levels of integration in deep-submicrometer (DSM) or nanoscale VLSI system designs. Also, for high-speed chip designs, faster on-chip rise times, longer wire lengths, plus the use of lower resistivity conductors, e.g., Cu and low- k dielectric insulators (with dielectric constants < 3) require that inductive effects also be considered.

It is relatively simple to extract the resistance of interconnect. Sheet resistance is widely used in resistance extraction for VLSI on-chip simulation. To extract resistance for on-chip interconnects in high frequencies system, however, skin and proximity effects should be considered. These effects are frequency dependent, and will increase wire resistance at high frequencies.

Capacitance extraction is challengeable. To model interconnect capacitance, both capacitance to the ground or substrate as well as capacitance to the neighboring wires need to be taken into account. Many literatures have been published for capacitance extraction [19, 48].

Based on the definition of inductance, identification of current loops is necessary to calculate inductance. Because of multiple wires around the current-carrying wire need to be included for any possible current return paths. The substrate may also offer return paths for signals, therefore, its effects need to be included in the simulation. Skin and proximity effects also affect current return paths and decrease wire

inductance at high frequencies. The frequency dependence of inductance is important, especially when there is a ground plane, substrate, or other conductive grids nearby the interconnects.

2.3.3 Statistical Timing Analysis

Under true operating conditions, the parameters chosen by the circuit designer are perturbed from their nominal values due to various types of variations. As a consequence, a single SPICE-level transistor or interconnect model (or an abstraction thereof) is seldom an adequate predictor of the exact behavior of a circuit. These sources of variation can broadly be categorized into two classes [15, 49]:

(1) Process variations result from perturbations in the fabrication process, due to which the nominal values of parameters such as the effective channel length (L_{eff}), the oxide thickness (t_{ox}), the dopant concentration (N_a), the transistor width (w), the interlayer dielectric (ILD) thickness (t_{ILD}), and the interconnect height and width (h_{int} and w_{int} , respectively).

(2) Environmental variations arise due to changes in the operating environment of the circuit, such as the temperature or variations in the supply voltage (V_{DD} and ground) levels. There is a wide body of work on analysis techniques to determine environmental variations, both for thermal issues, and for supply net analysis.

Both of these types of variations can result in changes in the timing and power characteristics of a circuit.

Process variations can also be classified into the following categories [15, 49]:

(1) Inter-die variations: are the variations from die to die, and affect all the devices on same chip in the same way, e.g., they may cause all of the transistor gate lengths of devices on the same chip to be larger or all of them to be smaller.

(2) Intra-die variations: correspond to variability within a single chip, and may affect different devices differently on the same chip, e.g., they may result in some devices having smaller oxide thicknesses than the nominal, while others may have larger oxide thicknesses.

Inter-die variations have been a longstanding design issue, and for several decades, designers have striven to make their circuits robust under the unpredictability of such variations. This has typically been achieved by simulating the design at not just one design point, but at multiple “corners”. These corners are chosen to encapsulate the

behavior of the circuit under worst-case variations, and have served designers well in the past. In nanometer technologies, designs are increasingly subjected to numerous sources of variation, and these variations are too complex to capture within a small set of process corners.

In nanometer technologies, intra-die variations have become significant and can no longer be ignored. As a result, a process corner based methodology, which would simulate the entire chip at a small number of design corners, is no longer sustainable. A true picture of the variations would use one process corner in each region of the chip, but it is clear that the number of simulations would increase exponentially with the number of such regions. This implies that if a small number of process corners are to be chosen, they must be very conservative and pessimistic. For true accuracy, this can be overcome by using a larger number of process corners, but this number may be too large to permit computational efficiency. Furthermore, traditional analysis is limited to verifying the functional correctness by simulating the design at a number of process corners.

However, worst case conditions in a circuit may not always occur with all parameters at their worst or best process conditions. As an example, the worst case for a pipeline stage when the wires within the logic are at their slowest process corner and the wires responsible for clock delay or skew between two stages is at the best case corner. However, a single corner file cannot simultaneously model best-case and worst-case process parameters for different interconnects in a single simulation. Hence, a traditional analysis requires that two parts of the design are simulated separately, resulting in a less unified, more cumbersome and less reliable analysis approach. The strength of statistical analysis is that the impact of parameter variation on all portions of a design are simultaneously captured in a single comprehensive analysis, allowing correlations and impact on yield to be properly understood.

As the magnitude of process variations have grown, there has been an increasing realization that traditional design methodologies (for both analysis and optimization) are no longer acceptable. For example, the magnitude of process variations in gate length (L_{eff}) are predicted to increase from 35% in a 130nm technology to almost 60% in 70nm technology. These variations are generally specified as the fraction $3\sigma/\mu$ where the σ and μ are the standard deviation and mean of the process parameter, respectively. Thus a 60% variation in 70nm technology implies that the standard

deviation of the distribution of gate length across large number of samples is $14nm$. With variations as large as these, it become extremely important that designers treat these variation in statistical manner rather than using guard-bands in deterministic analysis.

Many works related to statistical timing analysis have been published in recent years, see, for example, [6, 7, 16, 21, 34, 49].

Chapter 3

H-TREE CLOCK DISTRIBUTION NETWORKS IN PRESENCE OF PROCESS VARIATIONS AND INDUCTANCE EFFECTS

The evolution of VLSI chips towards larger die size and faster clock speed makes the clock distribution an increasingly important problem. H-Tree clock distribution networks (CDNs) are widely used to reduce the clock skew, but they may suffer from the process variation problem. With the increase of clock speed, the impact of on-chip inductance upon path delay has also become a significant issue. In this research, we study the performance of H-Tree CDNs when both the process variations and inductance effects are jointly taken into account. The impacts of possible coupling effects between adjacent wires upon the performance of an H-Tree CDN will also be addressed in this research. The results presented here will be helpful for assessing the real performance of a high speed H-Tree clock distribution network.

3.1 Introduction

The evolution of VLSI chips towards larger die size and faster clock speed makes the clock distribution an increasingly important issue. The clock distribution networks (CDNs), which are used to synchronize the data flows among different data paths in a digital synchronous system, can significantly affect the overall system performance and reliability [27]. Unfortunately, the performance of a clock distribution network is vulnerable to many factors, such as the systematic or random process variations, the interconnect layout-dependent coupling effects, on-chip inductance effects, IR-drop, etc.

To evaluate the performance of a CDN, we usually need to study the maximum clock delay and clock skew of it, where the clock skew is defined as the difference between the maximum clock delay and the minimum clock delay among all clock paths (interconnects) in the CDN. The clock skew arises mainly from unbalanced delays due to the unequal clock path lengths between clock source and different modules as well as from process variations that cause clock path delay variations [27]. To deal with the clock skew problem, a common way is to adopt the well-balanced H-tree CDNs [5]. For a well-balanced H-tree CDN, the uncontrollable clock skew mainly comes from the variations in process parameters that affect the interconnect impedance/capacitance and, in particular, any distributed buffer amplifiers [5, 34]. Extensive research efforts have been devoted to studying and modeling the impacts of process variations upon the clock skew of a CDN, see, for example, [21, 33, 34].

With the dramatic increase in clock speed and a significant decrease in the feature size in current high-speed VLSI circuits, the on-chip line inductive effects (in addition to the resistance and capacitance) will significantly affect the interconnect delay and cannot be neglected anymore in the performance evaluation of CDNs, especially for the CDNs with long interconnect lengths and a small signal rise time [12, 18, 31]. Furthermore, adopting the new materials with a low-resistance (e.g. copper) in VLSI fabrication makes inductive effects more significant for a global clock distribution network, where many long and wide wires are usually required. Therefore, the inductive effects of interconnects should be carefully addressed in the performance evaluation of modern high performance CDNs.

Although many papers about inductance effects or process variations have been published, to the best of our knowledge, however, no results have been reported about the performance evaluation in terms of maximum clock delay and clock skew of an H-Tree CDN when both the inductance effects and process variations are simultaneously considered. Thus, this research is devoted to the real performance evaluation of a high-speed H-Tree CDN when both the process variations and inductance effects are jointly considered. In particular, we will study the impact of possible coupling effects (including inductance coupling effect and capacitance coupling effect) between adjacent interconnects upon the performance of an H-Tree CDN. We will also investigate how the spatial dependent components and the magnitude of process variations affect the performance of an H-Tree CDN.

The rest of this research is organized as follows. Some preliminaries about process variations and inductance effects are introduced in Section 3.2. The delay evaluation models are described in Section 3.3. In Section 3.4, the simulation methodology is presented. The simulation results and discussion are provided in Section 3.6. Finally, in Section 3.7, we conclude this research.

3.2 Process Variations and Inductance Effects

In the manufacturing process of a VLSI system, some uncertainties (process variations) may arise due to the parameter fluctuations of devices or environment, which make the overall performance of the system vary with these inherent and unavoidable fluctuations. In general, the parameter fluctuation consists of inter-die (die-to-die) parameter fluctuation and intra-die (within-die) parameter fluctuation.

Integrated circuits have always been vulnerable to inherent inter-die and intra-die parameter fluctuations in the manufacturing process. Die-to-die parameter fluctuations resulting from lot-to-lot, wafer-to-wafer, and a portion of the within-wafer variations affect every element on a chip equally. Conversely, within-die parameter fluctuations consisting of both random and systematic components produce a non-uniformity of electrical characteristics cross the chip.

Examples of the lot-to-lot and wafer-to-wafer variations include processing temperatures, equipment properties, wafer polishing, and wafer placement. The within-wafer variations have contributions to both die-to-die and within-die fluctuations. An example of the within-wafer variations that impact the die-to-die fluctuations is the resist thickness across the wafer, which is random from wafer to wafer, but deterministic within the wafer. The aberrations in the stepper lens are an example of systematic within-die variations. As an example of random within-die fluctuations, the placement of dopant atoms in the device channel region, which is an intrinsic effect since it cannot be eliminated by external control of conventional manufacturing processes, varies randomly and independently from device to device.

Traditionally, die-to-die fluctuations have been the main concern in CMOS digital circuit designs, and the within-die fluctuations have been neglected. As polysilicon gate lengths have decreased below the wavelength of light used in the optical lithography process, however, the systematic and random within-die fluctuations of channel length have exceeded the die-to-die fluctuations. Thus, within-die fluctuations are a

growing threat to the performance and functionality of future gigascale integration (GSI) [15].

The importance of accurately estimating the impact of parameter fluctuations on circuit performance is directly related to a company's overall revenue. An overestimation increases the design complexity, possibly leading to an increase in design time, an increase in die size, rejection of otherwise good designs, and even missed market windows. Conversely, an underestimation can compromise the product's performance and overall yield as well as increase the silicon debug time. In summary, overestimating fluctuations impacts the design effort, and underestimating fluctuations impacts the manufacturing effort.

The process variations may affect both the geometry parameters of devices (e.g. inverter) and the geometry parameters of interconnects (such as length, width and thickness) in VLSI systems. In nanoscale process or deep sub-micromicron (DSM) process, the parameter fluctuations impose a growing threat to the system performance, especially for the gigascale interconnection systems where the polysilicon gate length has decreased below the wavelength of light used in the optical lithography process [15]. It is predicted that in a 130nm technology [49], the variation magnitude in gate length of n-MOS and p-MOS can be as high as 35% (specified by the fraction $3\delta/\mu$, where δ and μ are the standard derivation and mean of gate length, respectively).

In this research, we consider both the interconnect parameters variation and device parameters variation in our analysis. For a parameter, its variation σ can be generally modeled as:

$$\sigma = \sigma_{Inter-die} + \sigma_{Intra-die,global} + \sigma_{Intra-die,local} + \varepsilon \quad (3.1)$$

where $\sigma_{Inter-die}$, $\sigma_{Intra-die,global}$, $\sigma_{Intra-die,local}$ are its inter-die variation, location-dependent global intra-die variation and local intra-die variation, respectively, and ε is a random component. For a die, its global intra-die variation can be modeled by a simple radial distribution [15, 21, 40], where the variation at the location (x, y) (normalized coordinates across the die) is expressed as

$$\sigma_{Intra-die,global} = \sigma_0 + \sigma_x \cdot x + \sigma_y \cdot y \quad (3.2)$$

here σ_0 is the location-independent component of variation, σ_x and σ_y are the gradients of parameter indicating the spatial variations of parameter along the x and y

directions, respectively. Therefore, the actual value p of a parameter can be formally expressed as:

$$p = p_0(1 + \sigma) \quad (3.3)$$

where p_0 is the nominal design value and σ is the overall variation of the parameter. It is notable that the overall variation of a parameter under process variations can be modeled by the Gaussian Distribution [37], so we can simply use the traditional Monte Carlo simulation to evaluate the clock skew and the maximum/minimum clock delay of a CDN when the parameter variations are fully considered.

With the increase of clock speed, the impact of line inductance upon path delay has also become a significant issue [31, 45]. Based on the transmission line theory, the line inductance can affect the circuit performance in two distinct ways [12]. First, it can affect the rise/fall time and the signal delay through an interconnect. The inductance usually results in a faster slew rate due to the fact that when a positive voltage is applied to an inductor, it takes some time to build up the charging current. Once the current is established, however, this current supply will continue for some time, resulting in an overall faster transition time and a reduced gate delay. Second, the inductance may cause the reflection and thus the overshoot/undershoots effects, because the line inductance is a part of characteristic impedance. It is notable that such overshoot/undershoots effects may cause the reliability problem to circuits [18]. Therefore, the line inductance can significantly affect circuit behavior when it becomes comparable with resistance component and thus can not be neglected anymore in the circuit behavior analysis.

3.3 Interconnect Delay Calculation

In this section, we introduce the interconnect delay models with or without the consideration of inductance effect. The models for the coupling capacitance and coupling inductance between adjacent wires (lines) are also introduced in this section.

3.3.1 RC Delay Model

For the comparison with the RLC delay model considering the inductance effect, here we first introduce the the traditional simple RC delay model [47], where the possible inductance effect is ignored. According to the RC model, the 50% signal delay (including both the gate delay and interconnect delay) can be estimated as:

$$\tau = R_s(0.693C_{int} + 0.693C_L) + R_{int}(0.377C_{int} + 0.693C_L) \quad (3.4)$$

where R_s is the output resistance of driver, C_L is the load capacitance, R_{int} and C_{int} are the total resistance and capacitance of the interconnect, respectively. It is notable that although the formula (4) is simple and efficient for low speed circuits, it can not be used to efficiently model the interconnection delay of modern high-speed VLSI systems, where the inductance effect becomes non-negligible.

3.3.2 RLC Delay Model

As the interconnect length and operating speed entered the nanoscale regime and gigascale regime, respectively, the inductance component becomes comparable to resistance component in circuits of VLSI (specially for Cu-based interconnect technology with a low resistance) [31]. Thus, the more advanced RLC model should be adopted to fully analyze the real performance of modern CDNs.

Let's consider the typical driver-interconnect-load structure used for delay analysis [12] (Fig.3.1), where the line considered is an uniform line with resistance r , capacitance c , and inductance l per unit length respectively, driven by an inverter of series resistance R_s and output parasitic capacitance C_p , and driving an identical inverter with load capacitance C_L . If we assume that the output resistance, output

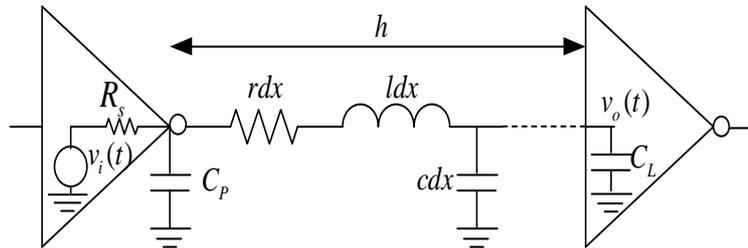


Fig. 3.1. Equivalent circuit of driver-interconnect-load structure

parasitic capacitance, and input capacitance of a minimum sized inverter are r_s , c_p , and c_0 , respectively, then for a repeater that is k times of the minimum sized repeater, its parameters R_s , C_p and C_0 are given by $R_s = r_s/k$, $C_p = c_p k$ and $C_0 = c_0 k$. Based on the delay analysis structure in Fig.?? and the parameters defined above, we can calculate the interconnect delay as follows.

According to the ABCD parameter matrix, the input-output transfer function of the circuit in Fig.3.1 is given by [22]

$$H(s) = \frac{V_o(s)}{V_i(s)} = \frac{1}{[1 + sR_s(C_p + C_L)] \cos(\theta h) + [\frac{R_s}{Z_0} + sC_L Z_0 + s^2 R_s C_p C_L Z_0] \sin(\theta h)} \quad (3.5)$$

Since the step-response of this circuit is given by $V_o(s) = \frac{1}{s}H(s)$ in the Laplace domain, so a second order Padé expansion of the transfer function is

$$H(s) \approx \frac{1}{1 + sb_1 + s^2 b_2}$$

where

$$\begin{aligned} b_1 &= R_s(C_p + C_L) + \frac{rch^2}{2!} + R_s ch + C_L rh \\ b_2 &= \frac{lch^2}{2!} + \frac{r^2 c^2 h^4}{4!} + R_s(C_p + C_L) \frac{rch^2}{2!} + (R_s ch \\ &\quad + C_L rh) \frac{rch^2}{3!} + (C_L lh + R_s C_p C_L rh) \end{aligned} \quad (3.6)$$

The two poles of the above transfer function are

$$s_{1,2} = \frac{-b_1 \pm \sqrt{b_1^2 - 4b_2}}{2b_2}$$

By taking the inverse Laplace transform of $\frac{1}{s}H(s)$, the output voltage of the circuit in Fig. 3.1 is given by

$$v(t) = V_0 \left[1 - \frac{s_2}{s_2 - s_1} e^{s_1 t} + \frac{s_1}{s_2 - s_1} e^{s_2 t} \right] \quad (3.7)$$

Finally, the $f \times 100\%$ (where $0 \leq f < 1$) delay of the circuit in Figure 3.1 can be calculated based on the Equation (3.7) with the Newton-Raphson method when a step input is applied.

3.3.3 Parasitics Extraction of Interconnects without considering coupling effects

The Equations (4)-(7) indicate that we need to extract the parasitic parameters (resistance, capacitance and inductance) for the evaluation of interconnect delay. Here, we first introduce the method of parasitics extraction when the coupling capacitance and coupling inductance between adjacent wires are ignored.

The interconnect resistances R_{int} can be simply evaluated as $R_{int} = r \cdot l/w$, where r , l and w are the resistance of unit length interconnect, the interconnect length and interconnect width, respectively.

For the evaluation of capacitance, we adopt a quasi-3D on-chip capacitance model proposed in [48] to calculate the self component of interconnect capacitance. The main idea of this capacitance model is to decompose a 3D wire structure into a series of 2D segments to achieve an efficient and accurate capacitance extraction for the 3D wire.

Finally, for the inductance used in the RLC delay model, we adopt the following model proposed in [46] to calculate the self inductance of a wire¹:

$$L_{self} = \frac{\mu_0}{2\pi} \left[l \ln\left(\frac{2l}{w+t}\right) + \frac{l}{2} + 0.2235(w+t) \right] \quad (3.8)$$

where l is the length of the wire, w and t are the width and thickness of the rectangular cross section of wire, respectively.

3.3.4 Parasitics Extraction Considering Coupling Effects

With the development of technology, the on-chip coupling capacitance and coupling inductance² are becoming dominating factors for complex layout [41,46]. When coupling effects are considered in delay calculation, we need to additionally consider only the coupling capacitance in the RC model, but we need to take into account both the coupling capacitance and coupling inductance in the RLC model.

Typically, interconnects on one layer of metal are routed in one direction and interconnects on the neighboring metal layers are routed in the orthogonal direction,

¹The coupling inductance (mutual inductance) will be discussed in next subsection.

²We refer to the coupling inductance and coupling capacitance between neighboring wires on a chip as the coupling effects in this research.

so the capacitive and inductive coupling effects between different metal layers are negligible compared to corresponding coupling effects within a layer. Notice that the coupling effects between a wire and its adjacent wire(s) are much significant than the coupling effects between the wire and other non-adjacent wires [12], so we will consider here only the coupling effects between a wire and its adjacent wire(s) within a layer.

For two adjacent wires depicted in Fig.3.2, the coupling capacitance c_{ij} between wires i with length l and j with length m can be calculated as [32]:

$$c_{ij} = \frac{\hat{f}_{ij} l_{ij}}{d_{ij}} \left(1 - \frac{x_i + x_j}{2d_{ij}}\right) \quad (3.9)$$

where \hat{f}_{ij} is unit-length fringing capacitance between wires i and j , x_i and x_j are the

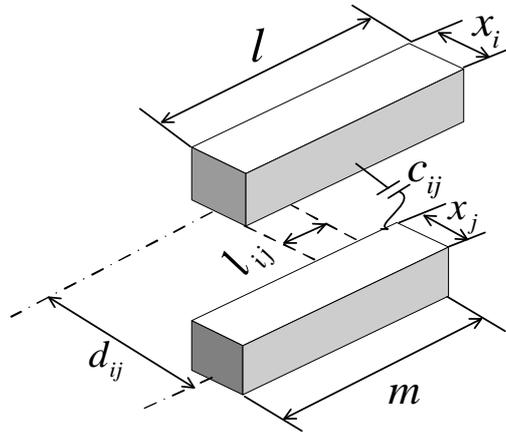


Fig. 3.2. Coupling capacitance between two adjacent wires

width of i and j respectively. We can see from (3.9) that the coupling capacitance c_{ij} between two neighboring wires i and j is proportional to the overlapping length l_{ij} but inversely proportional to the center-to-center distance d_{ij} .

Notice that it is usually formidable to extract the accurate interconnect inductance, because the current return path there is very complicated in a real chip. Thus, to make the evaluation of interconnect inductance tractable, we adopt here the for-

mulas proposed in [46] to extract the inductance. For the coupling scenario illustrated in Fig. 3.2, the mutual inductance between two wires is given by [46]:

$$M = \frac{\mu_0}{4\pi} \left[\begin{aligned} &(l - l_{ij}) \ln\left(\frac{l+m-l_{ij}}{l-l_{ij}}\right) + m \ln\left(\frac{l+m-l_{ij}}{m-l_{ij}}\right) \\ &+ l_{ij} \ln\left(\frac{4l_{ij}(m-l_{ij})}{d_{ij}^2}\right) - 2l_{ij} \end{aligned} \right] \quad (3.10)$$

3.4 Simulation Methodology

In this section, we first describe the H-Tree routing structure, we then introduce the simulation set up and also the parameters setting issue.

3.4.1 H-Tree CDNs

The H-Tree CDNs are widely used to reduce clock skew due to their well-balanced structure and equal clock paths. An H-Tree CDN with level=6 is illustrated in the Fig. 3.3, which is used to distribute the clock signal to $2^6 = 64$ processor elements (PEs).

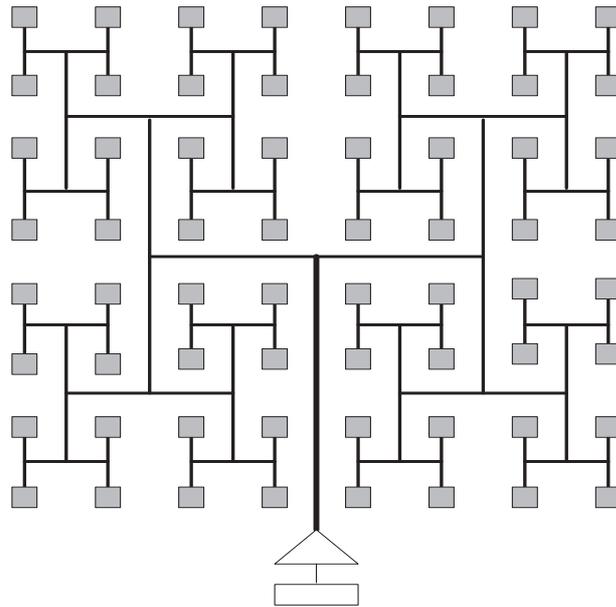


Fig. 3.3. An H-Tree CDN for 64 PEs

An H-Tree CDN can be represented by a binary tree, and the total delay of a clock path from source to a PE can be calculated by summarizing the delay of all

branches along this path. For an H-tree CDN, an inverter is usually inserted into each branch to drive the downstream interconnect (Inverters in different branches are not illustrated in Fig.3.3). Thus, we can apply the Equation (3.7) for RLC delay calculation and apply the Equation (3.4) for RC delay calculation for all branches of the H-Tree CDN.

Since the wires with different directions are usually routed in different metal layers to make the routing problem less complex, so the H-Tree CDNs used for performance evaluation in this research will be routed in two metal layers (each for one direction). For each clock signal wire, a ground line and a power line will be placed on either side of it as illustrated in Fig. 3.4. Based on the above placement, the coupling capacitance between a clock signal wire and its adjacent power/ground wires can be computed by the Equation (3.9), while the loop inductance (L_{loop}) of the clock signal wire interconnects is given by following Equation (3.11):

$$L_{loop} = L_{self_clock} + L_{self_power} - 2M_{clock_power} \quad (3.11)$$

where L_{self_clock} and L_{self_power} are the self inductance of clock and power wire, respectively, and they can be evaluated by the Equation (??). The M_{clock_power} is the mutual inductance between the clock and power wires, and it can be calculated using the formulas of mutual inductance proposed in [46].

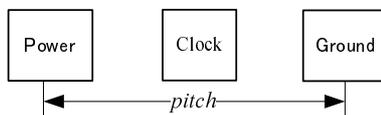


Fig. 3.4. Routing structure of clock wire and P/G wires

3.4.2 Simulation Setup

To fully investigate the performance of H-Tree CDNs in presence of both process variations and inductive effects, we consider here two types of CDNs: the CDNs used for global clock distribution in a SoC/NoC chip and the CDNs used for local clock distribution within a module of the chip.

For the simulation of an H-Tree CDN used for global clock distribution to different PEs in a SoC/NoC chip (we refer to it as *CDN-G* hereafter), we assume that the

effective area of each PE is $0.2\text{mm} \times 0.2\text{mm}$ while its tile area is $0.25\text{mm} \times 0.25\text{mm}$. The input capacitance (load capacitance) of each PE is assumed to be 0.1pF [34], and the width of power/ground wire is assumed to be 2 times as that of the clock signal wire. The pitch between power lines and ground lines is determined by the approach proposed in [18] to obtain the optimal inductance.

For the performance simulation of an H-Tree CDN used for local clock distribution within a functional unit (abbreviated as *CDN-L* hereafter), the distance between clock sinks is set to be $25\mu\text{m}$, and the input capacitance of each clock sink is also assumed to be 0.1pF . The geometry size of power/ground lines is set to be the same as that of the clock signal wires. The pitch between power lines and ground lines is set to $1.2\mu\text{m}$.

Since the main target of this research is to investigate the effects of process variations and inductance upon the H-Tree CDNs, so we assume in our simulation that the power grid is ideal (i.e., it is free of IR-drop, voltage fluctuation, etc).

To understand how the spatial dependence (i.e. spatial correlation) of intra-die components in process variations impacts the performance of CDNs,³ we consider two scenarios in our simulation: in the first scenario, we regard the intra-die components of process variations as mutually independent (i.e., we don't consider the spatial correlation), while in the second scenario, we assume that the intra-die component of process variations is systematic and obeys the radial distribution given by Equation (3.2).

Finally, to investigate the sensitivity of CDNs performance to the process variations, we let the standard deviation of parameters vary from 3% to 21% in our simulations and check the effect of this standard deviation's variation upon the overall CDNs performance. It is notable that the above range of variance variation corresponds to a range 9%–63% for the magnitude of process variation.⁴

The main steps for simulation are summarized as the follows.

- (1) *Determine the layout of H-Tree CDN.*
- (2) *Generate independent random data set that follows the Gaussian distribution.*

³The intra-die variation becomes more significant than the inter-die variation for modern deep sub-micrometer technology [15, 21], so we mainly investigate the effects of the first variation here.

⁴The magnitude of process variation (please refer to the Section 2 for its definition) is predicted as high as 60% in a 70nm technology [49].

- (3) *Map the random data to fluctuations of physical dimensions of interconnects and device parameters.*
- (4) *Compute electrical parameters.*
- (5) *Compute the delay of each clock path in the clock distribution network.*
- (6) *Find the minimum/maximum delay and skew.*
- (7) *Evaluate the mean values of the maximum delay and clock skew, and their standard deviations.*

To get a stable estimations of the maximum clock delay and clock skew, all simulations are conducted one million times.

3.4.3 Parameters Setting

Our simulation will be conducted based on a 70nm Cu CMOS technology under the Berkeley Predictive Technology Model (BPTM) [1]. We suppose that a 100X size inverter is inserted into each branch of an H-Tree CDN, and the mean values and standard deviations of some key parameters are summarized in Table 3.1. Where

Table 3.1
Mean values and standard deviations of major process parameters

Parameter	$V_{TN}(\text{V})$	$V_{TP}(\text{V})$	$t_{ox}(\text{\AA})$	$\mu_N(\text{cm}^2/\text{V} \cdot \text{s})$
Mean	0.2	-0.22	25	600
SD	0.01	0.011	12.5	30.0
Parameter	$t(\text{nm})$	$w(\text{nm})$	$f_{ox}(\text{nm})$	$\mu_P(\text{cm}^2/\text{V} \cdot \text{s})$
Mean	600	450	800	140
SD	30	22	40	7.0

t_{ox} is the gate oxide thickness, μ is the charge carrier mobility, V_T is the threshold voltage, the interconnection line is with width w and thickness t on an oxide layer of thickness f_{ox} . The calculation of output resistance R_s of inverter is the same as that of [34], and the evaluation of parasitics capacitance C_p is based on the equations proposed in [56].

The fluctuations of geometry parameters of inverter and wire are all considered as the normal distribution and the magnitude of process variations is set to a conservative value 15%, i.e., the standard deviation of a parameter is 5% of its nominal value.

3.5 Verification of the simulation program

To verify our simulation program based on the distributed RLC delay model in this chapter, we first conducted several simulations on a small scale level-2 H-Tree CDN illustrated in Fig.3.5. Notice that an inverter is inserted at the beginning of each branch in the CDN and it is also portrayed explicitly in the figure. Since each branch can be modeled as a distributed RLC circuit in which a parent inverter drives two sub stream inverters as illustrated in Fig.3.6, we can calculate the propagation delay of each branch segment in the clock paths using the distributed RLC delay model described in this chapter. The total propagation delay from clock source driver to a PE is the sum of propagation delay of all segments along the clock path. In detail, we first calculated the RLC delay of all clock paths from the clock driver to each PE with our simulation program, and then compared the results with the delay calculated by Synopsys HSPICE.

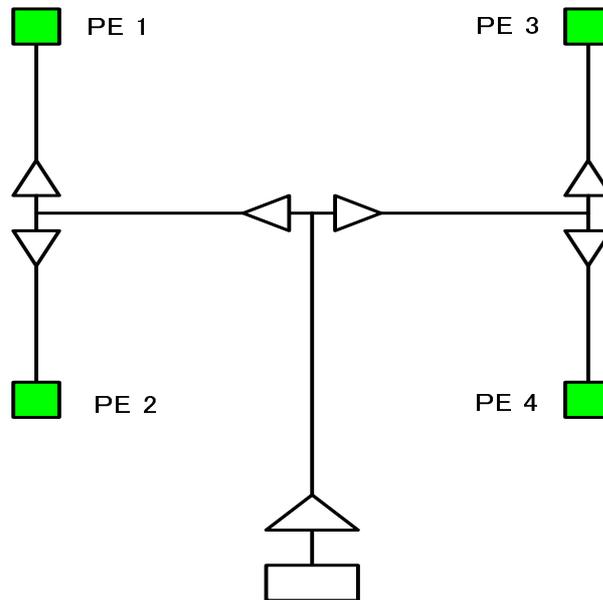


Fig. 3.5. An H-Tree CDN with level=2

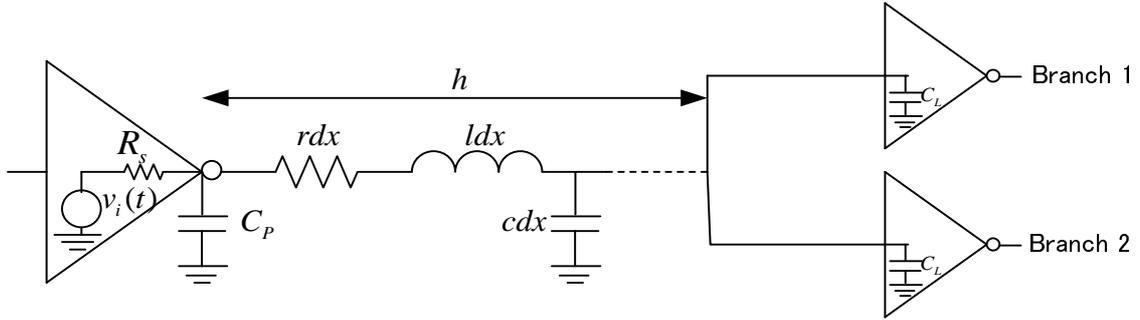


Fig. 3.6. A branch circuit in an H-Tree CDN

The test simulations are based on the 130nm parameter provided by BPTM. The distance between PEs is set to be $100\mu m$ in the H-Tree CDN. Here, we assume that the size of inverter is 50X than the minimum size inverter. The other parameters are the same as those listed in Section 3.4.2 and Section 3.4.3.

We conducted the simulations 1000 times. The simulation results are listed in Table 3.2 where both mean value and standard deviations of delay are calculated.

Table 3.2
Simulation results of the level-2 H-Tree CDN

Clock Sinks	Simulation Program (ps)		HSPICE (ps)		Differences [ps (%)]	
	Mean	SD	Mean	SD	Mean	SD
PE 1	22.4	1.2	21.8	0.9	0.6 (2.7)	0.3 (25)
PE 2	24.6	1.1	23.8	1.2	0.8 (3.3)	0.1 (9.1)
PE 3	23.7	1.3	22.5	1.1	1.2 (5.1)	0.2 (15.3)
PE 4	23.9	1.4	22.3	1.2	1.6 (6.7)	0.2 (14.3)
Average	23.65	1.25	22.6	1.1	1.05 (4.4)	0.2 (15.9)

Based on the simulation results, we can observe that the average delay difference between our simulation program and HSPICE is 4.4% and the corresponding result of their SDs is about 15.9%. Therefore, we confirm that the accuracy of our simulation program is enough to estimate the performance of H-Tree CDN in this research.

3.6 Performance Evaluation Results

In this section, we first investigate how the inductance effect and process variations affect the performance of a CDN, then we explore the effects of coupling on the performance of CDNs. In the last two subsections, we will study how the spatial dependence of process variations affect the performance of CDNs and analyze the sensitivity of CDNs performance to the process variations.

3.6.1 Effects of Inductance and Process Variations on H-Tree CDNs

To study how the inductance effect, in addition to the process variations, affect the performance of a CDN, we simulated the mean value and standard deviations of the maximum clock delay and clock skew for both *CDN-G* and *CDN-L* when the RC delay model and the RLC delay model were adopted, respectively.

For *CDN-G* with different size (determined by the number of PEs), the simulation results for the mean values of the maximum clock delay and the clock skew are summarized in Fig.3.7 and the results of their standard deviations are illustrated in Fig.3.8. The corresponding results of *CDN-L* are shown in Fig.3.9 and Fig.3.10, respectively.

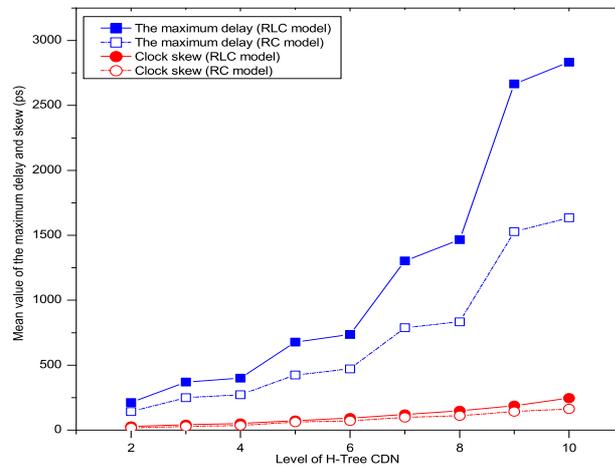


Fig. 3.7. Mean value of the maximum delay and skew in *CDN-G* (RC vs. RLC)

The Fig.3.7~Fig.3.10 all indicate that for either the *CDN-G* or the *CDN-L*, the simulation results of both clock skew and clock delay based on the RLC model are

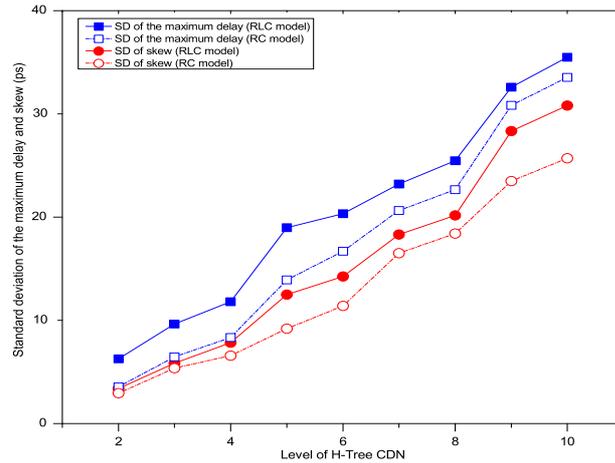


Fig. 3.8. Standard deviations of the maximum delay and skew in *CDN-G* (RC vs. RLC)

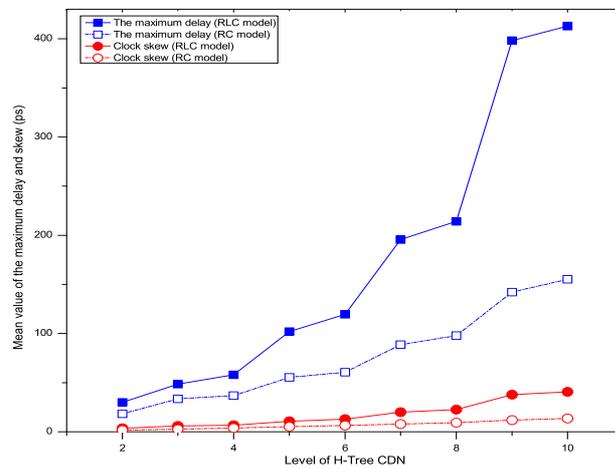


Fig. 3.9. Mean value of the maximum delay and skew in *CDN-L* (RC vs. RLC)

very different from that of the RC model. For example, for the *CDN-G* with 64 PEs (Level=6), we can see from the Fig.3.7 that the mean value of the maximum delay and skew are 471.7ps and 72.2ps when the RC model is adopted, while their corresponding values for RLC model are 735.9ps and 91.1ps, respectively. Thus, for this example, the RC delay model and the RLC delay model can cause about 56% differences in the maximum delay estimation and about 26% difference in the clock skew estimation. Similarly, we can find from the Fig.6 that for this example, the corresponding standard deviation of the maximum delay and clock skew estimation

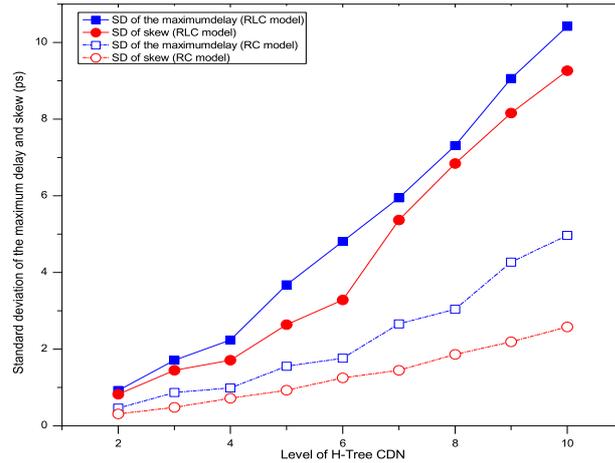


Fig. 3.10. Standard deviations of the maximum delay and skew in *CDN-L* (RC vs. RLC)

are about 16.7ps and 11.4ps for the RC model, 20.3ps and 14.2ps for the RLC model. So both models can cause about 25% difference in the standard deviation estimation of the maximum delay and clock skew. Similar estimation behaviors of RC and RLC model can also be observed in Fig.3.9 and Fig.3.10 for the small scale *CDN-L* networks. All the results in Fig.3.7~Fig.3.10 clearly show that the inductance has a significant impact on the timing issues and thus the performance of both *CDN-G* and *CDN-L* networks, especially for the deep-submeter or nanoscale VLSI chips, where the lower resistance lines exhibit more inductance effects so the reactive component of lines impedance is comparable to the resistive component.

A further observation of Fig.3.7 and Fig.3.9 reveals that for both the clock delay and clock skew, although the estimation differences between the RC model and the RLC model in either *CDN-G* or *CDN-L* always grow as their size increases⁵, the estimation difference for *CDN-L* increases more significantly than its *CDN-G* counterpart. For example, when the CDN level grows from 4 to 10, the results in Fig.3.7 show that the estimation difference for the maximum delay of a *CDN-G* network increases 56% (from 118% to 174%) and the estimation difference of its clock skew increases 42% (from 121% to 163%), while the results in Fig.3.9 for the *CDN-L* network indicate that the estimation difference of its maximum delay increases as high

⁵The estimation difference is defined by the fraction $(t_{RLC} - t_{RC})/t_{RC}$, where the t_{RLC} , t_{RC} are the clock delay (or skew) based on the RLC model and the RC model, respectively.

as 76% (from 190% to 266%) and the estimation difference of its clock skew significantly increases 70% (from 227% to 297%). The above changes in the increase rate of estimation difference between the *CDN-G* and *CDN-L* are mainly due to the following reasons. First, the geometry size and layout (e.g., the width and pitch of power/ground lines) of *CDN-G* are different from that of its *CDN-L* counterpart, so their parasitics capacitance and inductance are very different. Second, in *CDN-G* we adopted an design proposed in [18] to obtain the optimum inductance while we did not do so in *CDN-L*, because we don't have much freedom in *CDN-L* to conduct inductance optimization (e.g. wire sizing and spacing etc.) due to the more strict design constraints there (such as the physical design limit).

The results in both Fig.3.8 and Fig.3.10 reveal that due to the same reasons above, the standard deviation estimations of the maximum delay and clock skew also follow the similar trends as that of their mean values. That is to say, the estimation differences between the RC model and the RLC model in both the *CDN-G* and *CDN-L* networks always grow as their size increases, but the estimation difference for *CDN-L* increases more significantly than its *CDN-G* counterpart.

The results in this subsection indicate clearly that the delay and skew estimations based the RC delay model are much more optimistic than that of the RLC delay model, especially for the *CDN-L* networks, so the RC delay model may significantly overestimate the performance of H-Tree CDNs. Thus, the RLC model should be adopted in the performance evaluation of modern high performance CDNs to capture the increasing inductive effects. It is also notable that neglecting the inductance effect in performance verification of modern VLSI systems can even cause catastrophic logic failure or reliability problem, see, for example, [12].

3.6.2 Coupling Effects on Performance of H-Tree CDNs

To demonstrate the effect of coupling effects (coupling capacitance and coupling inductance) upon the performance of H-Tree CDNs, we further simulated the performance of H-Tree CDNs for both the cases when the coupling effects are considered and not considered, respectively. Since the performance trends of the *CDN-G* and *CDN-L* are very similar, so we include only the performance evaluation results of the *CDN-L* for discussion in this section and also in the following sections. Notice that the inductance effect is crucial to the performance evaluation of H-Tree CDNs as

discussed in the last section, so we adopted the RLC delay model in our simulation for the interconnect delay evaluation.

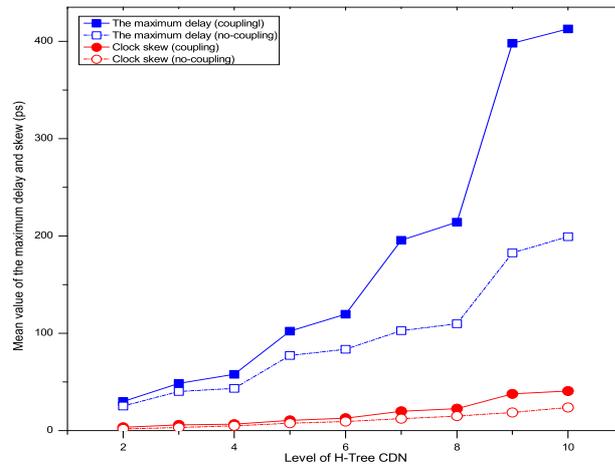


Fig. 3.11. Mean value of the maximum delay and skew in *CDN-L* (RLC model)

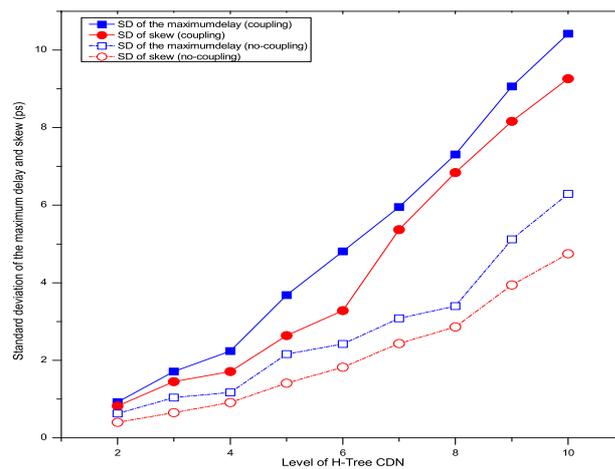


Fig. 3.12. Standard deviations of the maximum delay and skew in *CDN-L* (RLC model)

For both the cases considering and not considering the coupling effects, the simulation results of the mean values of the maximum delay and clock skew are summarized in Fig.3.11, while the corresponding results of their standard deviations are summarized in Fig.3.12. We can observe from both Fig.3.11 and Fig.3.12 that the coupling effects, in addition to the inductance effect and process variations, can also signifi-

cantly affect the performance of H-Tree CDNs. For the H-Tree CDN driving 64 clock sinks, the maximum delay considering the coupling effects is 120ps, which is almost 1.5 times as that of the 84ps maximum clock delay obtained without considering the coupling effects. Similarly, the clock skew of the CDN is 13ps when the coupling effects are considered, which is also significantly higher than the skew 9ps when the coupling effects are neglected. The results in Fig.3.12 indicate that the standard deviations of both the maximum delay and clock skew also become greater when the coupling effects are taken into account in the delay evaluation.

It is interesting to notice from Fig.3.11 and Fig.3.12 that although in general the mean value and standard deviations of the maximum delay and clock skew increase as the network size grows for both cases considering and not considering coupling effects, the estimation differences of the mean values between these two cases (defined similarly as that in Section 3.6.1) vary differently from the corresponding estimation differences of standard deviation. For example, when the level of $CDN-L$ grows from 4 to 10, the results in Fig.3.11 show that the estimation difference of the mean value of the maximum delay increases about 70% (from 33% to 107%) and the estimation difference of clock skew increases only slightly (from 70% to 71%), while the results in Fig.3.12 indicate that the estimation difference of standard deviation for clock skew increases about 7% (from 87% to 94%) but the corresponding results for the maximum delay decreases about 26% (from 91% to 65%).

3.6.3 Impact of Spatial Dependence of Process Variations

In order to investigate how the spatially dependent components of process variations affect the performance of CDNs, we also conducted the simulations when the process variations are determined by a simple radial distribution (see Equation (3.2)) to mimic the effects of spatial dependence of process variations. Based on the results in [15], we assumed that spatial variation of parameter is less than 10% in our simulation. Following the same arguments as that of section 3.6.1 and section 3.6.2, we only conducted the simulations for the $CDN-L$ networks when both the inductance effect (by using RLC model) and the coupling effects are considered. For both the maximum delay and clock skew, the corresponding simulation results of their mean values and standard deviations are summarized in Fig.3.13 and Fig.3.14, respectively. For comparison, we also include in Fig.3.13 and Fig.3.14 the simulation results when

spatially dependence of process variations is neglected (i.e., the process variations are assumed to be completely random).

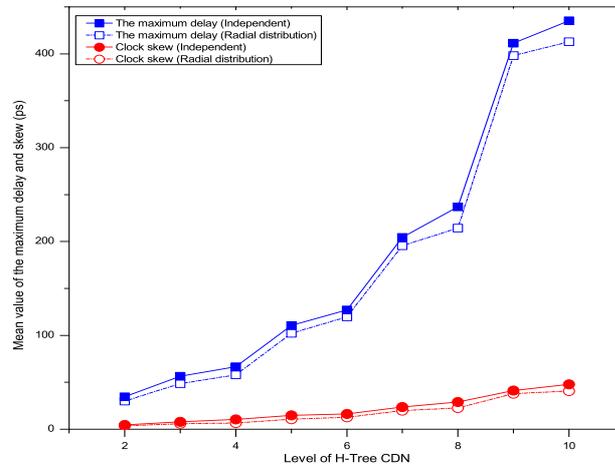


Fig. 3.13. Mean value of the maximum clock delay and skew of $CDN-L$ when considering spatial dependence or not (RLC model)

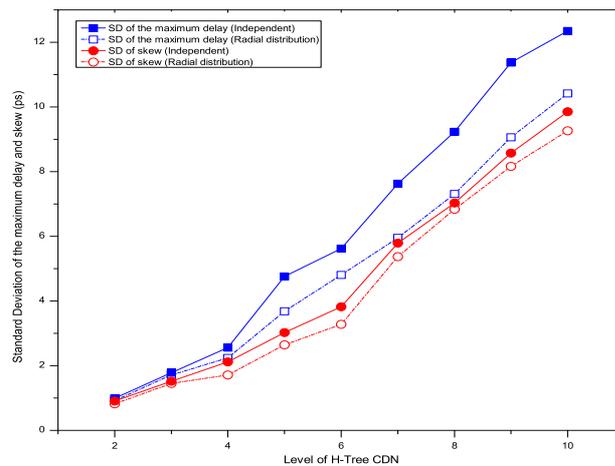


Fig. 3.14. Standard deviations of the maximum delay and skew in $CDN-L$ when considering spatial dependence or not (RLC model)

The results in the Fig.3.13 show that there is no big difference between the clock delay estimation from the spatial variation model (Equation (3.2)) and the corresponding estimation from the complete random model. For example, even for the large $CDN-L$ network with level 10, the estimations of the maximum delay are 412ps

and 435ps for the radial distribution model and random model, respectively, so the estimation difference between the two models is less than 4%. The above results indicate that if the spatial parameters variation is limited within 10% [15], we will not see a big difference between these two variation models when they are adopted for the maximum clock delay estimation.

The skew estimation results in Fig.3.13 show, however, that the skew estimation difference (defined similarly as that in Section 3.6.1) between the two variation models is much significant than the corresponding delay estimation differences. E.g., for the same *CDN-L* network with level 10, the skew estimation difference is 14% (41ps for the spatial variation model and 47ps for the random model), which is significantly higher than the 4% difference of the corresponding delay estimation.

On the other hand, the estimation results of standard deviation in Fig.?? show that for the *CDN-L* network, the estimation difference of the maximum delay between the two variation models is much significant than that of the clock skew between them. For example, to the *CDN-L* with level=10, the estimation differences of the maximum delay there is about 20% (10.4ps for the spatial variation model and 12.4ps for the random model) while the corresponding difference for clock skew is just 6%.

The results in this section indicate that the spatial dependence of process variations may also significantly affect the timing analysis for high-speed CDNs, in particular for the pipeline-based CDNs where the clock speed is dominated by the clock skew rather than the maximum clock delay.

3.6.4 Performance Sensitivity to the Magnitude of Process Variations

To explore the performance sensitivity of a CDN to the magnitude change of process variations, we provide in Fig.3.15 the simulation results of skew and the maximum clock delay for the level-6 *CDN-L* network. In our simulation, the RLC model was adopted and both the coupling effects and spatial dependence of process variations are jointly considered (same as that of the section 2 and section 3). As discussed in Section 3.4.2, we assumed a range of 9% – 63% for the magnitude of process variation.

We can observe from Fig.3.15 that both the maximum clock delay and clock skew of a *CDN-L* increase slightly as the magnitude of process variations grows. For example, when the magnitude of process variations increases from 9% to 54%, the

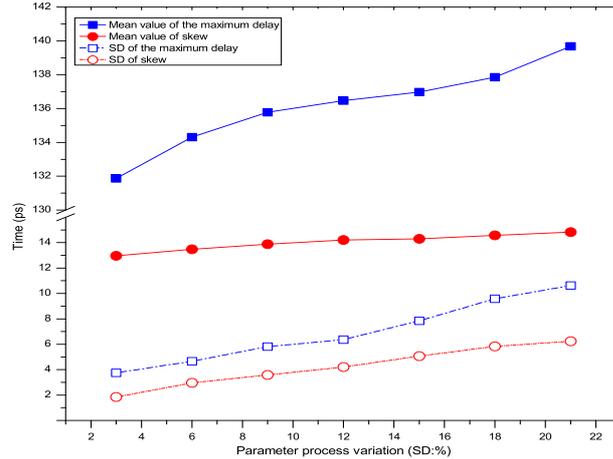


Fig. 3.15. Mean value and standard deviation of the maximum delay and skew when process variations change in a Level-6 *CDN-L* (RLC model)

mean value of the maximum clock delay increases only slightly by 3% (from 134.3ps to 137.8ps) while the mean value of clock skew increases about 14% (from 12.96ps to 14.5ps).

However, a further observation of the results in Fig.3.15 reveals that as the magnitude of process variations grows, the standard deviations of both the maximum delay and clock skew increase more significantly than their mean values. For example, as the magnitude of process variations changes from 9% to 54%, the standard deviation of the maximum clock delay increase 186% (from 3.7ps to 10.6ps) while the standard deviation of clock skew increases 240% (from 1.8ps to 6.2ps), respectively. Thus, the standard deviations of the maximum delay and skew are much more sensitive to the magnitude of process variation than their mean values.

3.7 Conclusions

In this research, we conducted extensive simulation to investigate the possible impact upon H-Tree CDNs' the delay and skew of the following effects: process variations, inductance, and coupling (coupling capacitance and coupling inductance between adjacent wires). Our results in this research indicate that

- (1) A significant overestimation of the H-Tree CDNs' performance may be introduced if the inductance effect is neglected in the delay evaluation, especially for the modern VLSI systems, where the low-resistance material is adopted and the inductance component of line impedance becomes comparable to the resistive component.
- (2) The performance of H-Tree CDNs can also be significantly overestimated if the coupling effects are not taken into consideration, since the coupling capacitance and the coupling inductance (mutual inductance) are becoming dominant over their self components with the continuous increase of the wire routing density.
- (3) Although the spatial dependence of process variations does not significantly affect the maximum clock delay of H-Tree CDNs, it may remarkably affect the clock skew and the standard deviation of the maximum delay.
- (4) The standard deviations of the maximum delay and clock skew vary significantly with the increase of the magnitude of process variation, although their mean values are not very sensitive to the magnitude of process variation. Thus, the yields of the maximum clock delay and clock skew may be significantly degraded by magnitude of process variation.

Chapter 4

VARIANT X-TREE CLOCK DISTRIBUTION NETWORK

The evolution of VLSI chips towards larger die size, smaller feature size and faster clock speed makes the clock distribution an increasingly important issue. In this research, we propose a new clock distribution network (CDN), namely Variant X-Tree, based on the idea of X-Architecture proposed recently for efficient wiring within VLSI chips. The Variant X-Tree CDN keeps the nice properties of equal-clock-path and symmetric structure of the typical H-Tree CDN, but results in both a lower maximal clock delay and a lower clock skew than its H-Tree counterpart, as verified by an extensive simulation study that incorporates simultaneously the effects of process variations and on-chip inductance. We also propose a closed-form statistical models for evaluating the skew and delay of the Variant X-Tree CDN. The comparison between the theoretical results and the simulation results indicates that the proposed statistical models can be used to efficiently and rapidly evaluate the performance of the variant X-Tree CDNs.

4.1 Introduction

Clock signals that operate at the highest speed of any signals within a VLSI chip play a central role in the design of modern digital synchronous systems. The evolution of VLSI chips towards larger die size, smaller feature size and faster clock speed makes the clock distribution an increasingly important issue [27]. The clock distribution networks (CDNs), which are used to distribute clock signals to synchronize the data flows among different data paths, can significantly affect the overall system performance and reliability.

To evaluate the performance of a CDN, we usually need to study the maximum clock delay and clock skew of it, where the clock skew is defined as the difference between the maximum clock delay and the minimum clock delay among all clock paths (interconnects) in the CDN. The clock skew arises mainly from unbalanced delays due to the unequal clock path lengths between clock source and different modules as well as from process variations that cause clock path delay variations [27].

The critical issues concerning with the design of clock distribution network are to achieve a low clock delay and the minimum or a useful skew in most cases with the minimum buffer size and wire length. The well-balanced H-Tree CDN has been widely adopted to eliminate the skew caused by unequal clock path lengths [21], where the uncontrollable clock skew mainly comes from the variations in process parameters that affect the interconnect impedance/capacitance and, in particular, any distributed buffers or amplifiers [5, 21]. Extensive research efforts have been devoted to studying and modeling the impacts of process variations upon the clock skew of a CDN, see, for example, [21, 33, 34].

Although H-Tree is attractive for clock distribution due to its small clock skew and a relatively simple implementation, it usually results in a long clock path from the clock source to each sink (clock terminal). Thus, an H-Tree CDN usually causes a higher clock delay.

Mesh or grid is also a popular architecture for distributing clock signals on a chip. It uses inherent redundant interconnects created by loops to smooth out undesirable variations between signal nodes spatially distributed over the chip, and thus results in a lower clock skew. However, the mesh/grid CDN usually occupies a larger wiring area, and consumes more power. Such a condition is becoming worse with the increase of modern VLSI chips' area moreover.

Recently, X Architecture was proposed to wire in a VLSI chip with considerably shorter wiring length than that of traditional Manhattan wiring architecture [4]. It has been demonstrated in [53] that the X Architecture, which supports 45- and 135-degree wires as well as the vertical and horizontal wires, can reduce as high as 29% of the wire length required by the simple Manhattan wiring architecture. As a result, the X Architecture becomes promising to considerably reduce the delay and improve the overall performance of on-chip interconnects.

In this research, we extend the X Architecture to clock distribution and propose a novel non-orthogonal clock distribution network, namely Variant X-Tree, which preserves the nice properties of equal-clock-path and symmetric structure of the typical H-Tree CDN. We will analyze the detail layout and construction features of the Variant X-Tree CDN. The simulation results show that the Variant X-Tree CDN is able to achieve both a lower maximal clock delay and a lower clock skew than its H-Tree counterpart. We also introduce a statistical performance evaluation model that is able to estimate its performance rapidly with statistical analysis method while considering the process variations. Experiment results indicate that the proposed model is effective to serve as an upper bound of performance. Moreover, it can also be integrated into design flow expediently for a set of closed-form equations are derived based on this model.

The rest of this research is organized as follows. Some preliminaries about X Architecture are introduced in Section 4.2. The layout and construction features of Variant X-Tree CDN are presented in Section 4.3, and the statistical performance model is described in Section 4.4. The experiment methodology of performance evaluation is introduced in Section 4.5. Section 4.6 provides the simulation results and discussions, and finally, the Section ?? concludes this research.

4.2 X Architecture

Within traditional VLSI chips, interconnects have been routed using so-called Manhattan architecture, namely, only the vertical wires and horizontal wires are permitted in a chip. The X Architecture that belongs to non-Manhattan architecture [35] was proposed recently for efficient integrated circuit wiring based on the pervasive use of diagonal wires [4]. Compared with the traditional, currently ubiquitous, Manhattan architecture, the X Architecture demonstrates a wire length reduction of more than 20% and a via reduction of more 30%. Because of the rapidly increasing percentage of delay due to interconnect and the manufacturing challenges due to vias in the nanometer realm, these length and via reductions result simultaneously in a chip performance improvement of 10%, a power reduction of 20%, and a die cost reduction of 30%. Furthermore, the reduction in both wire length and parallel runs on different layers often both reduces die size and improves signal integrity. Remarkably, on

virtually every important measure of chip quality, the X Architecture is superior to the Manhattan architecture.

The X Architecture (Fig.4.1(a)), which in its full incarnation applies to chips with at least five layers of metal, begins with the rotation of the fourth and fifth metal layers (i.e., M4 and M5) by 45 and -45 degrees, respectively, with respect to their Manhattan equivalents (Fig.4.1(a)). On the large scale, M1–M3 remain Manhattan, which simplifies the use of most standard cells, module and memory compilers, and hard IP today in conjunction with diagonal wiring.

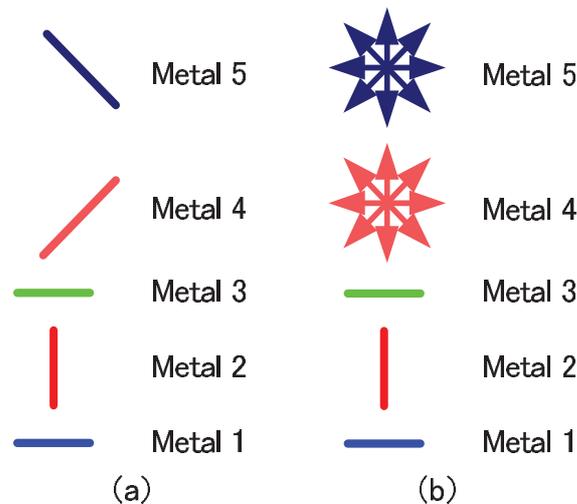


Fig. 4.1. (a) Preferred direction in X Architecture (b) Non-preferred direction support in M4 and M5

In general, a preferred-direction implementation of the X Architecture is likely to increase the number of vias and can be worse for delay despite the reduction in wire length. For wire placement, only a router that exploits all eight compass directions on all layers (see Fig.4.1(b)) without the artificial constraint of a preferred direction provides the benefits of diagonal wiring without the penalty of introducing extra vias; we call this technology *liquid routing*. Therefore, it was suggested recently to combine the techniques of X-aware placement and liquid routing (i.e., wiring with all eight compass directions in all layers) to take the full advantages of X Architecture [4]. A realistic example on the left side of Fig.4.2 shows a portion of a commercial chip, routed with a commercial Manhattan router. On the right side of Fig.4.2 is a liquid

router's re-routing of this region, without re-placement. The wire length has been reduced by about 14%, and the via count has been reduced by over 40%.

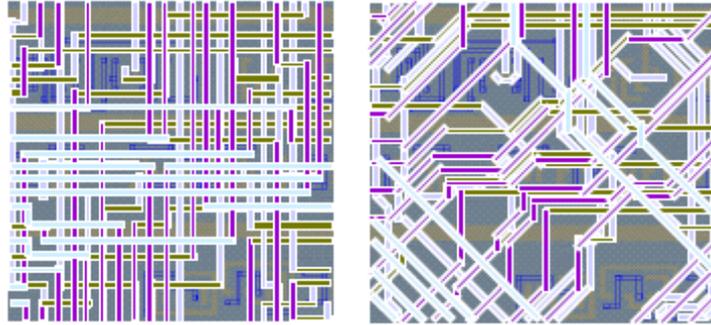


Fig. 4.2. Contrasting Manhattan (left) and Liquid Routing (right)

The concept of X Architecture is simple and its benefits are clear. Compared to the traditional Manhattan architecture, it can shorten wiring by up to 17% (the maximum reduction of 29% theoretically) across a die [9] in a average case. The utilization of X Architecture is becoming popular, some VLSI chips based on X Architecture have been released (e.g., a GPU chip by ATI in 2005 and a 10Gb Ethernet chip by Teranetics in 2006) according to [4]. Design rules and EDA tools that support X Architecture are also available.

According to [53], EDA tools can handle X-aware placement and liquid routing with enhanced graphics interface and database. It is also shown that X architecture is manufacturable and scalable [4].

4.3 Variant X-Tree CDN

Since the X Architecture is promising in reducing the wire length and thus the interconnect delay, we expect that the performance of clock distribution can be improved if the X Architecture is applied to the design of a clock distribution network.

To support our new CDN, in this research, we make an assumption that all the eight directions wiring (including preferred directions and non-preferred directions) can be permitted on VLSI chip wiring. An example that non-preferred directions are also supported in M4 and M5 is illustrated as Fig.4.1(b).

In this section, we first introduce the basic unit of Variant X-Tree based on X Architecture. We then study the detail layout features of Variant X-Tree CDN.

4.3.1 Basic Unit of Variant X-Tree CDN

We apply the building fashion of the well-known H-Tree CDN to the construction of new CDN based on X Architecture. It is notable, however, that we can not directly apply the X Architecture for wiring to construct an X-Tree in a recursive way as that of H-Tree, because the overlapping among the 45-degree and 135-degree wires increase sharply with the increase of supported clock sinks such that we can not construct a conventional X-Tree CDN recursively which exceeds 2 levels in one metal layer.

Here, we propose a scalable approach to constructing large scale CDNs based on an extension of X Architecture (we refer to this new CDN as Variant X-Tree CDN hereafter).

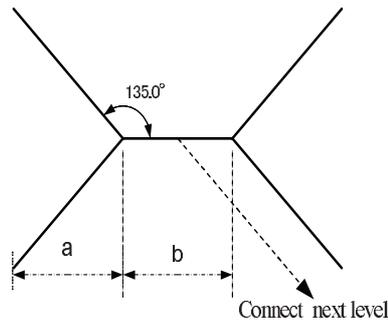


Fig. 4.3. The basic unit of Variant X-Tree

The main idea of Variant X-Tree CDN is to define a basic unit (the minimum unit) for clock distribution of 4 sinks first, as portrayed in Fig.4.3. The parameter a in Fig.4.3 is just a half of the distance between two sinks and b is the offset distance relative to sinks horizontally, where the parameter b must satisfy the following inequality (4.1) to guarantee the connectivity between two adjacent network levels.

$$b \geq 2\sqrt{2} \cdot P_{min} \quad (4.1)$$

where P_{min} is the minimum permitted pitch of interconnects within a VLSI chip.

Based on the basic unit illustrated in Fig.4.3, we can construct a large scale Variant X-Tree CDN recursively based on the similarity of H-Tree CDNs. Fig.4.4 shows a level 6 Variant X-Tree CDN which distributes clock signal to $2^6 = 64$ clock sinks.

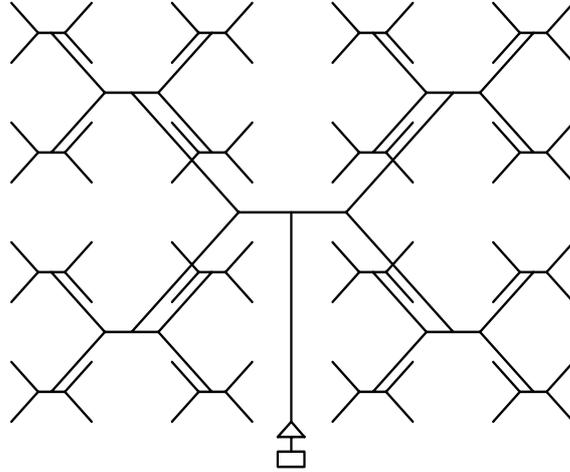


Fig. 4.4. A Variant X-Tree CDN with 64 sinks (level=6)

According to the Fig.4.3 and the Fig.4.4, the length of clock path that starts from clock source to each clock sink is equal obviously like H-Tree CDN. The demerit of Variant X-Tree CDN is that the distance between sinks is not always equal and it is axis-symmetry. But the isometry of Variant X-Tree CDN can be achieved approximately when the distance between sinks (geometric parameter a) becomes enough far and b is set to the minimum permitted value.

4.3.2 Construction Features of Variant X-Tree CDN

It is necessary to study the detail construction features in order to build a large scale CDN based on the basic unit of Variant X-Tree. In this subsection we describe some important layout properties of Variant X-Tree CDN.

Observation 1: *To generate a Variant X-Tree with a shorter clock path length than its H-Tree counterpart, the following inequality must be satisfied.*

$$b < 2(2 - \sqrt{2}) \cdot a \quad (4.2)$$

Proof According to the construction rule of H-Tree CDN and Variant X-Tree CDN, we can see easily that the clock path length ($L_{H-Tree}^{(n)}$) of a n -level H-Tree CDN ¹ is given by (4.3), while the clock path length ($L_{X-Tree}^{(n)}$) of a Variant X-Tree CDN is given by (4.4), respectively.

$$L_{H-Tree}^{(n)} = (2^{1+n/2} - 2) \cdot a \quad (4.3)$$

$$L_{X-Tree}^{(n)} = \frac{1}{2}(2^{n/2} - 1) \cdot (2\sqrt{2}a + b) \quad (4.4)$$

Obviously, to obtain a shorter clock path than that of H-Tree, we need

$$L_{X-Tree}^{(n)} < L_{H-Tree}^{(n)}$$

that yields Inequality (4.2). ■

Observation 2: *The recursive construction levels that can be achieved in one metal layer do not exceed 2 in Variant X-Tree CDN.*

Proof Consider that two recursive levels Variant X-Tree CDN has been constructed in a layer as portrayed in Fig.4.5, where h_1 is the distance between the branch connecting next level and right-bottom branch of level-2, h_2 is the distance between the left-bottom branch of level-1 and the right-bottom branch of level-2. $h_1^{(n)}$ and $h_2^{(n)}$ of n -level CDN can be computed with geometry constraints,

$$h_1^{(n)} = 2^{\frac{n}{2}} \cdot \frac{\sqrt{2}}{4}b \quad (4.5)$$

$$h_2^{(n)} = 2^{\frac{n}{2}}\sqrt{2}a + 2^{\frac{n}{2}} \cdot \frac{\sqrt{2}}{4}b \quad (4.6)$$

Clearly, here $h_1^{(n)}$ is smaller than $h_2^{(n)}$. Therefore, it results in a intersection when a complete level-2 Variant X-Tree CDN connects the next level in the same metal layer. It demonstrates that the next level Variant X-Tree can not be routed in the same metal layer. Thus the maximum levels of complete Variant X-Tree that can be achieved in one layer should be 2 at most. Additionally, Observation 2 can be proved intuitively from Fig.4.5. ■

¹H-Tree CDN can be considered as a Variant X-Tree CDN where the distance of clock sinks is $2 \cdot a$, $b=0$ and no diagonal wires are permitted. Note the vertical distance of sinks in Variant X-Tree CDN is also $2 \cdot a$.

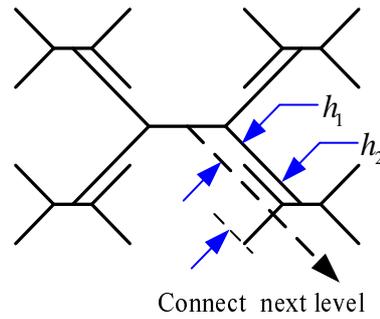


Fig. 4.5. Wire intersection occurs in one layer when more levels are constructed

Since a basic unit of Variant X-Tree can connect 4 clock sinks, so its level will be 2 ($4 = 2^2$) according to the traditional level definition of H-Tree. Thus, two recursive levels Variant X-Tree CDN will correspond to a 4 levels CDN, and so on. Therefore, Observation 2 indicates that there are $2^4 = 16$ clock sinks (terminals) in one metal layer at most, so we can determine the maximum number of supported clock sinks if the number of routing metal layers of CDN is specified. For examples, the sinks can be as many as 256 when Variant X-Tree CDN is wired in 2 metals, 1024 sinks for 3 metal layers and 4096 for 4 metal layers. Therefore we consider that it is enough for most of CDNs and suitable to the semi-global or global clock distribution on a VLSI chip.

We can also connect two even $(n - 1)$ -level sub-Variant X-Tree CDNs to construct an odd n -level Variant X-Tree CDN using 45 or 135 degree wires (see Fig.4.6, where a level-7 Variant X-Tree CDN is built on two level-6 CDNs).

Furthermore, it is notable that we can connect two 2 recursive levels Variant X-Tree with horizontal/vertical wires rather than orthogonal wires so as to obtain 64 terminals in one metal layer without the occurrence of wire intersection. In other word, we can construct a Variant X-Tree CDN with 512 clock sinks in two metal layers by connecting two level-6 CDN using a horizontal/vertical wire. Moreover, with this idea, we can derive a hybrid CDN that combines the topology of Variant X-Tree and H-Tree and maximize the number of clock sinks of a CDN. Fig.4.7 shows such a CDN that adds one H-Tree level after constructing two recursive levels of Variant X-Tree CDN. Thus we have the following corollary:

Observation 3: *A hybrid CDN with 2 recursive levels Variant X-Tree and 1-level H-Tree link supports 64 terminals in one metal layer.*

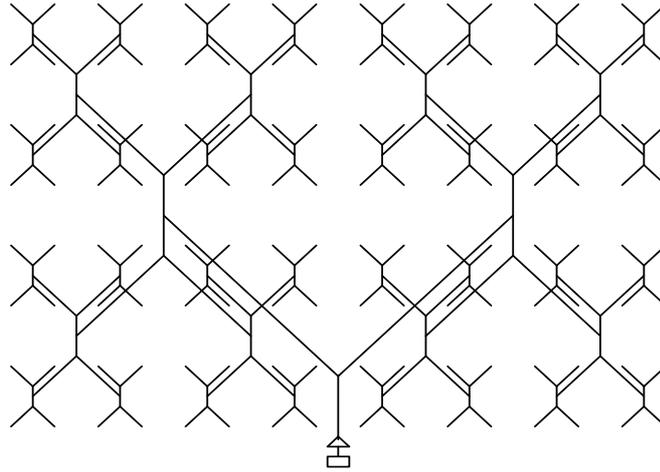


Fig. 4.6. A Level-7 Variant X-Tree CDN

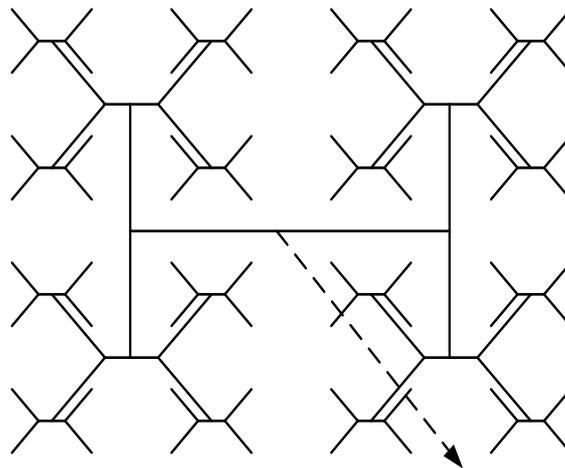


Fig. 4.7. A hybrid CDN with 2 recursive levels Variant X-Tree and 1-level H-Tree link

So we can construct a larger size CDN with this hybrid Variant X+H Tree structure while making the number of metal layers routed small ($2^6 = 64$ clock sinks in one metal layers).

Notice that the isometry can not be reserved in Variant X-Tree CDN, namely, the pitch of clock sinks (the distance of clock sinks) is non-uniform within whole Variant X-Tree CDN. However, the isometry of Variant X-Tree CDN can be preserved approximately with a big a and a small b when the application scenario is not so sensitive or strict to the isometry of CDN.

Additionally, the proportion of length and width of routing area is $(2a + b)/2a$ for an even level Variant X-Tree CDN. Also, the proportion will approach 1 when a big a and a small b are set in Variant X-Tree CDN. Clearly, the hybrid Variant X+H Tree can also be used to obtain a square routing area approximately. It is notable that the routing area is not square for both H-Tree CDN and Variant X-Tree CDN with odd levels.

The rules mentioned above enables to choose proper type of CDN depending upon the number of clock sinks and the available wiring metal layers.

4.4 Statistical Performance Analysis Model

As the CMOS technology advances into the nanometer feature size and multi-gigahertz regime and with the adoption of Cu-based on-chip interconnects, the performance of VLSI circuits is getting more sensitive to the process variations. Process variations can significantly impact both devices and interconnect performance so as to affect the circuit performance (especially for clock distribution network). Traditional corner-based analysis could be conservative [16]. The statistical timing analysis becomes important in the last few years. In this section, we propose a statistical performance evaluation approach for Variant X-Tree CDN in order to give designers a guideline for estimating the performance of CDN in initial design stage.

4.4.1 Statistical Performance Evaluation Model

The authors in [34] proposed a statistical skew modeling for general clock distribution network in a bottom-up manner. Specially, a closed-form model of clock skew and maximum clock delay is also presented for a well-balanced H-Tree CDN.

For the Variant X-Tree CDN (even the hybrid CDN—Variant X+H Tree proposed in this research), the statistical performance model in [34] can also be applicable because the Variant X-Tree CDN is a binary clock tree essentially and well-balanced like that of H-Tree. We thus adopt this statistical performance model to estimate the performance of Variant X-Tree CDN. The statistical performance model is concluded as follows.

In Fig.4.8, ξ , η and χ are defined as the maximum delay, the minimum delay and clock skew at the intersection node of two branches respectively. d is the delay of

clock paths (branches of CDN) that connects different nodes. Then for a N level

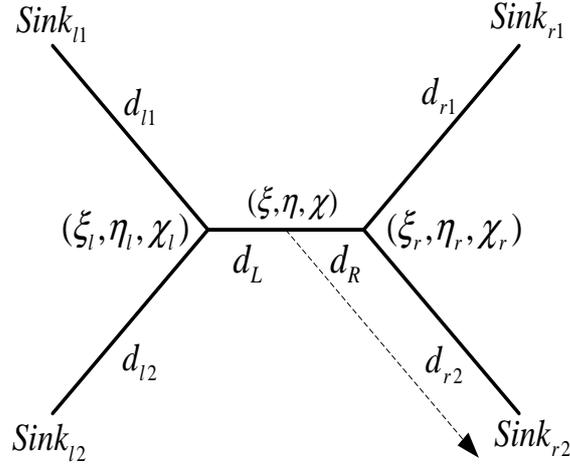


Fig. 4.8. Illustration of basic unit of variant X-Tree for statistical performance analysis

Variant X-Tree, let d_i , $i = 0, \dots, N$ be the actual delay of branch i of a clock path. The mean values and the variances of the maximal clock delay, ξ , and the minimal clock delay, η , of the Variant X-Tree are then given by following equations:

$$\begin{aligned}
 E(\xi) &= \sum_{i=0}^N E(d_i) \\
 &+ \frac{1}{\sqrt{\pi}} \sum_{i=1}^N \sqrt{\sum_{k=1}^i \left(\frac{\pi-1}{\pi}\right)^{k-1} \cdot D(d_{N-i+k})} \quad (4.7)
 \end{aligned}$$

$$\begin{aligned}
 E(\eta) &= \sum_{i=0}^N E(d_i) \\
 &- \frac{1}{\sqrt{\pi}} \sum_{i=1}^N \sqrt{\sum_{k=1}^i \left(\frac{\pi-1}{\pi}\right)^{k-1} \cdot D(d_{N-i+k})} \quad (4.8)
 \end{aligned}$$

$$D(\xi) = D(\eta) = \sum_{i=0}^N \left(\frac{\pi-1}{\pi}\right)^i \cdot D(d_i) \quad (4.9)$$

The expected clock skew $E(\chi)$ and skew variance $D(\chi)$ of the N level Variant X-Tree CDN are given by:

$$\begin{aligned} E(\chi) &= E(\xi) - E(\eta) \\ &= \frac{2}{\sqrt{\pi}} \sum_{i=1}^N \sqrt{\sum_{k=1}^i \left(\frac{\pi-1}{\pi}\right)^{k-1} \cdot D(d_{N-i+k})} \end{aligned} \quad (4.10)$$

$$\begin{aligned} D(\chi) &= D(\xi) + D(\eta) - 2 \cdot \rho \cdot \sqrt{D(\xi) \cdot D(\eta)} \\ &= 2 \cdot (1 - \rho) \cdot \sum_{i=0}^N \left(\frac{\pi-1}{\pi}\right)^i \cdot D(d_i) \end{aligned} \quad (4.11)$$

Where $E(\cdot)$ and $D(\cdot)$ represent the mean value and the variance of a random variable, respectively, ρ is the correlation coefficient of ξ and η , and ρ can be recursively evaluated for a network. The closed-form expressions (4.7)-(4.11) indicate clearly how the clock skew and the maximum clock delay are accumulated along the clock paths and with the increase of Variant X-Tree CDN's size.

4.4.2 Variance Estimation for a Branch in CDN

The delay variance of a clock path can be determined in term of variances of these independent random variables. To make the problem practicable, we assume that process variations of devices and interconnect are independent. Thus we can calculate the delay variance as described in [34].

To analyze the delay in the presence of process variations rapidly, the authors in [6] proposed a statistical model that discards higher order terms while does not result in a loss of accuracy. The proposed models enable closed-form computation of means and variances of interconnect delay for given magnitudes of relevant process variations. Therefore we can also adopt this model for efficient analysis (means and variances) of interconnect delay.

Note that considering the correlation of parameters can lead to more accurate result, but this correlation is neglected here for simplified computation. According to [7, 34] the correlation will lead the path delay in the same chip tending to be a positive dependent. Neglecting it will guarantee that the expected values of skew and maximal clock delay are still kept as a upper bound. It will be shown that the model is effective by the experiment results in Section 4.6 under such a case.

4.5 Methodology of Performance Evaluation

To verify the performance improvement of Variant X-Tree CDN and the proposed statistical performance model, we conducted some simulations in presence of process variations. In this section, we first introduce the considerations about process variations and the delay calculation method, we then describe the simulation setups and parasitics extraction method related to delay calculations. Finally, we show the parameters setting issue.

4.5.1 Process Variations

In the manufacturing process of a VLSI system, some uncertainties (process variations) may arise due to the parameter fluctuations of devices or environment, which make the overall performance of the system varies with these inherent and unavoidable fluctuations. In general, the parameter fluctuation consists of inter-die parameter fluctuation and intra-die parameter fluctuation. The former one is the results of lot-to-lot and wafer-to-wafer variations of parameters related to equipment properties, wafer polishing and wafer placement, and it usually affects every element on a chip equally. On the other hand, the intra-die parameter fluctuation, such as the resist thickness fluctuation across wafer and the aberrations in stepper lens, usually affects the elements of a chip unequally and it produces a non-uniformity of electrical characteristics across the chip.

The process variations may affect both the geometry parameters of devices (e.g. inverter) and the geometry parameters of interconnects (such as length, width and thickness) in VLSI systems. In nanoscale process or deep sub-micron (DSM) process, the parameter fluctuations impose a growing threat to the system performance, especially for the gigascale interconnection systems where the polysilicon gate length has decreased below the wavelength of light used in the optical lithography process [15]. It is predicted that in a 130nm technology [49], the variation magnitude in gate length of n-MOS and p-MOS can be as high as 35% (specified by the fraction $3\delta/\mu$, where δ and μ are the standard deviation and mean of gate length, respectively).

In this research, we consider both the interconnect parameters variation and device parameters variation in our analysis. For a parameter, its variation σ can be generally modeled as:

$$\begin{aligned} \sigma = & \sigma_{Inter-die} + \sigma_{Intra-die,global} \\ & + \sigma_{Intra-die,local} + \varepsilon \end{aligned} \quad (4.12)$$

where $\sigma_{Inter-die}$, $\sigma_{Intra-die,global}$, $\sigma_{Intra-die,local}$ are its inter-die variation, location-dependent global intra-die variation and local intra-die variation, respectively, and ε is a random component.

4.5.2 Interconnect Delay Calculation

As the interconnect length and operating speed entered the nanoscale regime and gigascale regime, respectively, the inductance component becomes comparable to resistance component in circuits of VLSI (specially for Cu-based interconnect technology with a low resistance) [30]. Thus, the more advanced RLC model should be adopted to fully analyze the real performance of modern CDNs.

In this research, we calculate the delay of clock path in CDN with a distributed RLC model which circuit structure is portrayed in Fig.4.9. In [30], an empirical RLC delay equation based on curve-fitting was derived as Equation (4.13):

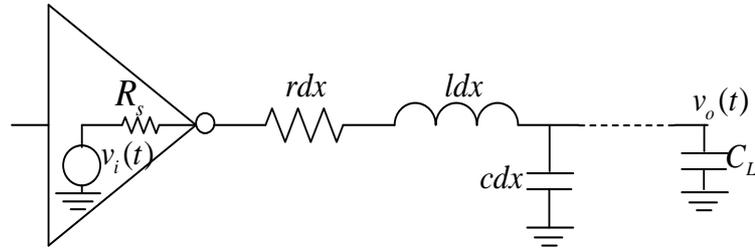


Fig. 4.9. Equivalent circuit of driver-interconnect-load structure

$$t_{50\%} = (e^{-2.9\zeta^{1.35}} + 1.48\zeta) / \varpi_n \quad (4.13)$$

where

$$\varpi_n = \frac{1}{\sqrt{L_{int}(C_{int} + C_L)}} \quad (4.14)$$

$$\zeta = \frac{R_{int}}{2} \sqrt{\frac{C_{int}}{L_{int}}} \frac{R_T + C_T + R_T C_T + 0.5}{\sqrt{1 + C_T}} \quad (4.15)$$

$$R_T = R_s/R_{int} \quad (4.16)$$

$$C_T = C_L/C_{int} \quad (4.17)$$

C_L is the load capacitance and R_{int} , C_{int} , and L_{int} are the total line resistance, capacitance, and inductance, respectively. R_s is the output resistance of driver (inverter).

Similar to H-Tree CDN, a Variant X-Tree CDN can also be represented by a binary tree, and the total delay of a clock path from source to a sink can be calculated by summarizing the delay of all branches along this path. For a Variant X-Tree CDN, we make a suppose that an inverter is inserted into each branch to drive the downstream interconnect (see an example illustrated in Fig.4.4, inverters in different branches are not illustrated). Thus, we can apply the Equation (4.13) for RLC delay calculation.

4.5.3 Simulation Setup

To fully investigate the performance improvements of Variant X-Tree CDN compared to H-Tree CDN in presence of process variations, we consider here two types of CDNs: a Variant X-Tree CDN (we refer to it as CDN-X hereafter) and an H-Tree CDN (abbreviated as CDN-H hereafter). Both of them are oriented global clock distribution and used to distribute clock signals to different processor elements (PEs) in a SoC/NoC chip.

For each clock signal wire in both CDN-X and CDN-H, a ground line and a power line will be placed on either side of it as illustrated in Fig.4.10.

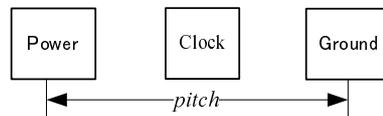


Fig. 4.10. Routing structure of clock wire and P/G wires

Since the main target of this research is to investigate the delay and clock skew for both H-Tree CDN and Variant X-Tree CDN in the presence of process variations,

so we assume in our simulation that the power grid is ideal (i.e., it is free of IR-drop, voltage fluctuation, etc).

The main steps for simulation are summarized as the follows.

- (1) *Determine the layout of CDN.*
- (2) *Generate independent random data set that follows the Gaussian distribution.*
- (3) *Map the random data to fluctuations of physical dimensions of interconnects and device parameters.*
- (4) *Compute electrical parameters.*
- (5) *Compute the delay of each clock path in the clock distribution network.*
- (6) *Find the minimum/maximum delay and skew.*
- (7) *Evaluate the mean values of the maximum delay and clock skew, and their standard deviations.*

To get a stable estimations of the maximum clock delay and clock skew, all simulations are conducted one million times.

4.5.4 Parasitics Extraction

The Equations (4.14)–(4.17) indicate that we need to extract the parasitic parameters (resistance, capacitance and inductance) for the evaluation of interconnect delay.

The interconnect resistances R_{int} can be simply evaluated as $R_{int} = r \cdot l/w$, where r , l and w are the resistance of unit length interconnect, the interconnect length and interconnect width, respectively.

For the evaluation of capacitance, we adopt a quasi-3D on-chip capacitance model proposed in [48] to calculate interconnect capacitance. The main idea of this capacitance model is to decompose a 3D wire structure into a series of 2D segments to achieve an efficient and accurate capacitance extraction for the 3D wire. The capacitance between crossover wires (even for non-orthogonal wires) can be calculated using Equation (4.18) with the proposed concept *effective width* (W_{eff}):

$$C_{cross} = W_{eff}(90^\circ) \csc(\phi) C_{self} \quad (4.18)$$

where ϕ is the rotation angle of crossover wires. The authors have shown that an excellent agreement exists between their results and that of the 3D capacitance solver.

Finally, we need calculate the inductance used in the RLC delay model. Notice that it is usually formidable to extract the accurate interconnect inductance, because the current return path is very complicated in a real chip. To make the evaluation of interconnect inductance tractable, we adopt here the formulas proposed in [46] to extract the inductance. Thus, the loop inductance (L_{loop}) of the clock signal wire interconnects is given by following Equation (4.19) depending upon their routing structure:

$$L_{loop} = L_{self_clock} + L_{self_power} - 2M_{clock_power} \quad (4.19)$$

where L_{self_clock} and L_{self_power} are the self inductance of clock and power wire, respectively, and the M_{clock_power} is the mutual inductance between the clock and power wires. Both self inductance and mutual inductance can be calculated using the formulas of mutual inductance proposed in [46].

4.5.5 Parameters Setting

For the simulation of CDN-X and CDN-H used for global clock distribution to different processor elements (PEs), we assume the geometric parameters of CDN-X as $a=0.12mm$ and $b=10\mu m$, the distance between PEs in CDN-H is set to $0.24mm^2$. The input capacitance (load capacitance) of each PE is assumed to be $0.1pF$ [34], and the width of power/ground wire is assumed to be the same as that of the clock signal wire. The pitch of power line and ground line is set to be $1.5\mu m$.

Our simulation will be conducted based on a $70nm$ Cu CMOS technology under the Berkeley Predictive Technology Model (BPTM) [1]. We suppose that a 100X size inverter whose size is 100 times bigger than that of the minimum sized inverter is inserted into each branch of an H-Tree CDN, and the mean values and standard deviations of some key parameters are summarized in Table 4.1. Where t_{ox} is the gate oxide thickness, μ is the charge carrier mobility, V_T is the threshold voltage, the interconnection line is with width w and thickness t on an oxide layer of thickness f_{ox} . The calculation of output resistance R_s of inverter is the same as that of [34],

²Note that the distance between PEs in CDN-H is equal to the vertical distance between PEs in CDN-X.

Table 4.1
Mean values and standard deviations of major process parameters

Parameter	$V_{TN}(V)$	$V_{TP}(V)$	$t_{ox}(\text{\AA})$	$\mu_N(cm^2/V \cdot s)$
Mean	0.2	-0.22	25	600
SD	0.01	0.011	12.5	30.0
Parameter	$t(nm)$	$w(nm)$	$f_{ox}(nm)$	$\mu_P(cm^2/V \cdot s)$
Mean	600	450	800	140
SD	30	22	40	7.0

and the evaluation of input capacitance (including gate capacitance and parasitic capacitance) is based on the equations proposed in [56].

The fluctuations of geometry parameters of inverter and wire are all considered as the normal distribution and the magnitude of process variations is set to a conservative value 15%³, i.e., the standard deviation of a parameter is 5% of its nominal value.

4.6 Simulation Results

In this section, we first investigate the performance improvements of Variant X-Tree CDN compared to H-Tree CDN, then we present the results data to validate the proposed statistical performance analysis model.

4.6.1 Performance Improvement of Variant X-Tree CDN

To study the performance improvement of Variant X-Tree compared to H-Tree CDN in presence of process variations, we simulated the mean values of the maximum/minimum clock delay and clock skew for both CDN-X and CDN-H when the RLC model is adopted.

For two types of CDN with different size (determined by the number of PEs), the mean values of maximum clock delay and minimum clock delay are illustrated in Fig.4.11, and the corresponding results of clock skew are summarized in Fig.4.12.

³The magnitude of process variation (please refer to the Section 4.5.1 for its definition) is predicted as high as 60% in [49].

The Fig.4.11 and Fig.4.12 indicate clearly that the simulation results of both clock delay and clock skew of Variant X-Tree (CDN-X) are very different from that of H-Tree CDN (CDN-H). For example, we can see from the Fig.4.11 and Fig.4.12 that the mean value of the maximum delay and skew are 141.6ps and 12.2ps for CDN-X, while their corresponding values for CDN-H are 152.7ps and 13.6ps, respectively. Thus, for this example, the Variant X-Tree CDN-X can cause about 7% differences in the maximum delay estimation and about 12% difference in the clock skew estimation.

Similarly, the difference in the maximum delay is 8% and an improvement of 17% in clock skew can be observed when the number of sinks reaches 1024 (level=10).

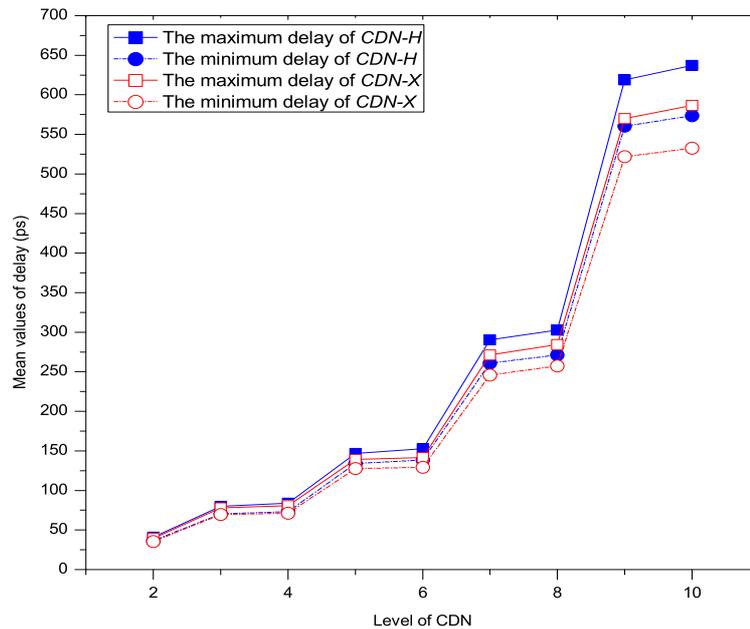


Fig. 4.11. The maximum and minimum delay of CDNs (Variant X-Tree CDN vs. H-Tree CDN)

We contribute this as the decrease of clock path length of Variant X-Tree CDN comparing with the H-Tree CDN. It can be sure that performance of CDN which adopts X-Architecture is improved. On the other hand, that delay of clock signal and clock skew almost have no change or a small alteration between two types of CDN when the number of sinks in CDN keeping a small level can also be seen from above illustrations.

It is notable that the clock path in a level n (n is even) Variant X-Tree CDN is as long as that in an odd $(n - 1)$ -level Variant X-Tree CDN, so the maximum delay

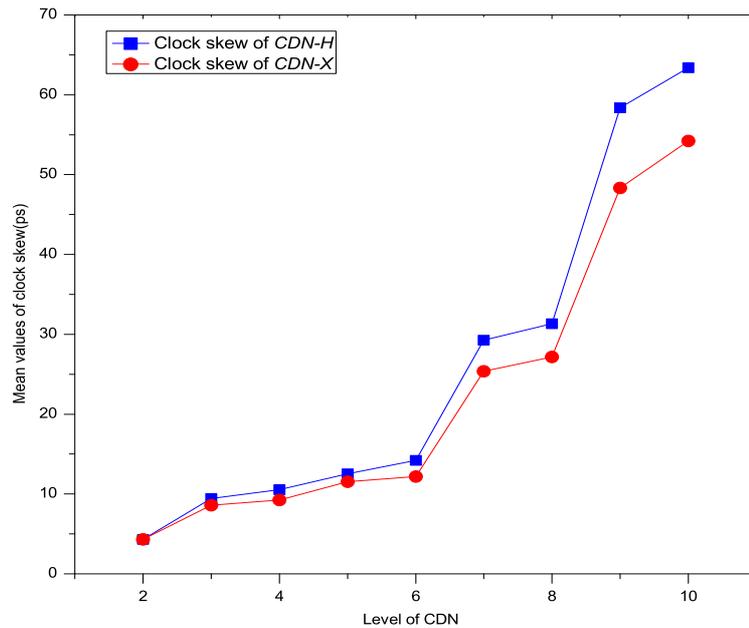


Fig. 4.12. The clock skew of CDNs (Variant X-Tree CDN vs. H-Tree CDN)

and clock skew of a Variant X-Tree CDN with even $n - level$ will not be significantly different from that of its odd level $(n - 1)$ counterpart.

4.6.2 Statistical Performance Evaluation

To verify the proposed statistical performance analysis model of Variant X-Tree CDN, we also perform some experiment simulations about CDN-X. In particular, we simulated the mean values and standard deviations (SDs) of both the maximum clock delay and clock skew and compared the simulation results with the theoretical results calculated from the proposed statistical model. Here, we performed the variance estimation with respect to the process variations with the method proposed in [6] for efficient calculation.

For CDN-X with different size (determined by the number of PEs), the results of mean values of the maximum clock delay and clock skew based on the proposed theoretical model and the Monte Carlo simulations are summarized in Fig.4.13. We provide in Fig.4.14 the corresponding results of standard deviations of the maximum clock delay and clock.

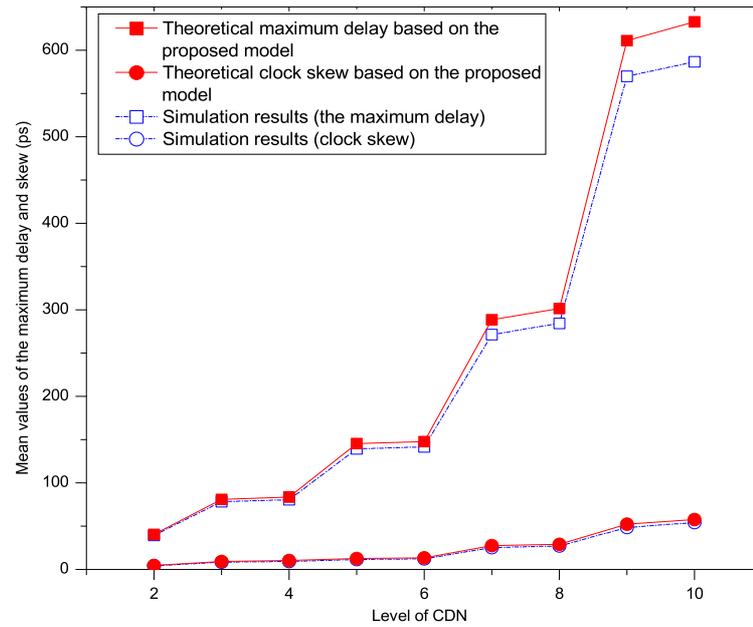


Fig. 4.13. Mean values of the maximum clock delay and clock skew of CDN-X

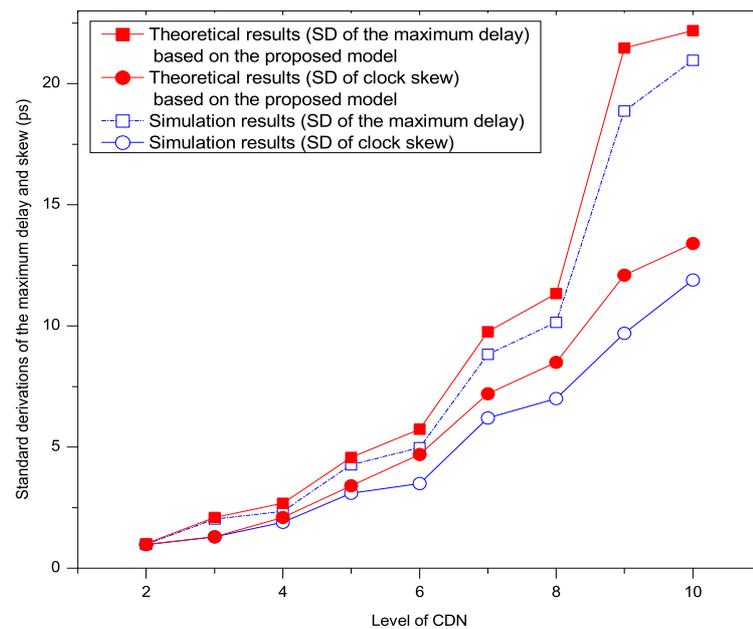


Fig. 4.14. Standard deviation of the maximum clock delay and clock skew of CDN-X

The results in Fig.4.13 show that there is no big difference between the maximum clock delay estimation and clock skew between the simulation results and theoretical

results based on the proposed statistical model. For example, even for the large CDN-X network with level 10, the estimations of the maximum clock delay are 586.6ps and 632.7ps for simulation results and theoretical results respectively, while the corresponding estimations of the clock skew are 54.2ps and 57.8ps, so the estimation differences of both the maximum clock delay and clock skew between simulation results and theoretical results are less than 7%. The above results indicate that the statistical model proposed in this research is suitable for estimation of the maximum clock delay and clock skew.

On the other hand, the estimation results of standard deviations of the maximum clock delay and clock skew in Fig.4.14 show that for CDN-X network, the estimation differences of standard deviations of the maximum clock delay and clock skew are much significant than that of mean values of them. For example, the maximum estimation difference of the maximum clock delay's SD is 17% in a level 9 CDN-X network, and the corresponding results of clock skew's SD is 13%. But it is notable that the simulation results do not exceed the theoretical results, namely, the theoretical results (model) can be considered as an upper bound at least.

It is notable that as the size of a Variant X-Tree CDN increases, its achievable clock frequency tends to decrease, for both the traditional clocking mode and pipeline clocking mode. For a CDN based on traditional clocking mode, its clock frequency is usually determined by the maximum clock delay, which in general increases as the CDN size grows. For a pipeline-based CDN, on the other hand, its clock frequency is mainly determined by the clock skew [42], which also in general increases as the CDN size grows.

The results in this section show that although there are higher estimation differences of standard deviations with the statistical performance model proposed in this research, it remains effective and can be applicable in performance evaluation of Variant X-Tree CDN, especially in the pre-design stage of a CDN.

4.7 Conclusions

In this research, we present a novel non-orthogonal CDN based on X Architecture for on-chip wiring. We also conduct simulation to validate its performance when both process variations and inductance effects are taken into account. Our simulation results indicate that comparing to the traditional H-Tree CDN, the proposed new

CDN has the potential to improve the overall clock distribution performance in terms of maximal clock delay and clock skew.

We also study the layout features of Variant X-Tree in detail. It enables to determine the proper size Variant X-Tree clock distribution network with these rules. A statistical performance evaluation model is proposed as well. Experiment simulation results show it is suitable to estimate the performance of a Variant X-Tree in design stage can also be integrated into design flow easily for its closed-form.

Chapter 5

CLOCK DISTRIBUTION AND ITS PERFORMANCE ENHANCEMENT IN 3D ICs

Three dimensional (3D) integrated circuits (ICs) have the potential to significantly enhance VLSI chip performance, functionality and device packing density. High performance clock distributions play an important role in these VLSI systems based on 3D ICs. In particular, interconnects delay and signal integrity issues are critical for clock distribution in chip design. In this research, we extend the idea of redundant via insertion of conventional 2D ICs and propose an approach for vias insertion/placement in 3D ICs to minimize the propagation delay of interconnects with the consideration of signal integrity. The simulation results based on a $65nm$ CMOS technology demonstrate that our approach in general can result in a 9% improvement in average delay and a 26% decrease in reflection coefficient. It is also shown that the proposed approach can be more effective for interconnects delay improvement when it is integrated with the buffer insertion in 3D ICs.

5.1 Introduction

Semiconductor chips have been facing constant pressure to improve their performance with a reduced power and cost. As technology node of VLSI scales, the chip area and wire length continue to increase, causing increased interconnect delays. To overcome this problem, some methods such as repeater/buffer insertion and wire sizing have been developed. However, these methods will result in other problems like augmented power consumption and thermal integrity. All of the above problems can have deleterious implications on chip performance, reliability and design effort.

Furthermore, such phenomena will be sustained or aggravated as clock frequencies of VLSI systems increase.

Three dimensional (3D) integrated circuits (ICs), which comprise multiple tiers of active devices, have the potential to enhance VLSI chip performance, functionality and device packaging density [55]. For example, the 3D ICs offer an attractive alternative to conventional 2D planar ICs: they can combine different technologies such as analog and digital circuits within the single chip cube to construct a multi-tier (multi-plane) system. Thus, using 3D ICs allows for integrating the best technology for a particular portion of an application into one chip package, for example, a possible SoC design based on 3D IC is illustrated in Fig.5.1.

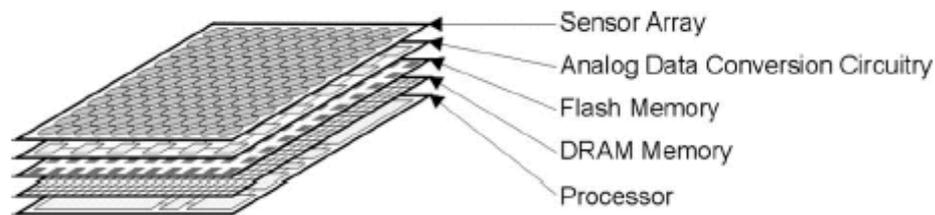


Fig. 5.1. A possible SoC design based on 3D IC.

One of several promising solutions to 3D ICs is vertical integration, in which multiple layers of active devices are stacked with vertical interconnects between tiers (planes) to form 3D integrated circuits. Depending on the fabrication technology, 3D ICs can be accomplished by different processes or schematics. For example, chip-level and wafer-level schematics based on face-to-face or face-to-back stacking have been proposed, where the stacked tiers are bonded with metal pads or blanket dielectric fusion bonding (or an adhesive interlayer) [55] to achieve interconnection between the neighboring tiers (wafers or chips). Fig.5.2 shows a schematic diagram of 3D IC structure implemented with a face-to-back process.

3D ICs offer a promising solution, reducing both footprint and interconnect length without shrinking the transistors at all. By expanding vertically rather than spreading in 2D planar area, obviously the delay of signal propagation in interconnects can be decreased due to the decreased length of interconnects in 3D ICs. Thus, the drawback of long interconnects in conventional 2D ICs can be alleviated. Additionally, the 3D

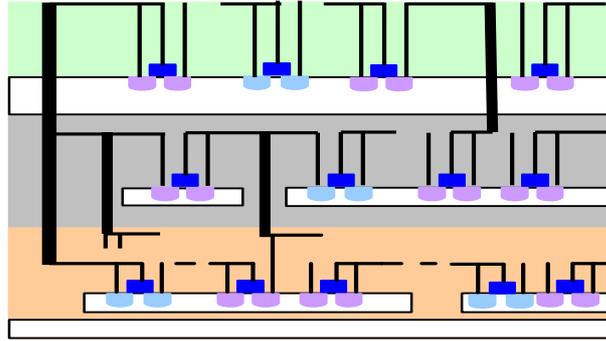


Fig. 5.2. The structure schematic of a 3-tiers' 3D IC fabricated with a face-to-back process (The thick vertical lines stand for through-hole vias for inter-tiers interconnections).

ICs technology can also result in a reduction of total active power, noise improvement and a greater logical span.

In 3D ICs, signal paths like wires for global clock distribution consist of multiple-segment interconnects routed in different tiers and some vertical inter-tier interconnects implemented by vertical through-hole vias (abbreviated as vias hereafter). Since each tier in 3D ICs may be fabricated with different technologies or processes, the impedance characteristics of different segments of the global interconnects may be disparate [55]. Furthermore, the impedance characteristics of vias may also be different from that of horizontal wires. Hence, the impedance characteristic of wires for inter-tier and global interconnects is actually not uniform [44] in 3D ICs.

Interconnects delay is crucial to the performance of modern digital VLSI systems, since it is a big fraction of total delay and it is in general increasing with technology scaling. To maximize performance improvement in 3D ICs, some approaches related to vias placement and wire routing have been proposed in presence of non-uniform impedance characteristics of interconnects. The authors in [58] assumed that the vertical wire (via) is in the middle of the net (wire) regardless of the length and impedance of the wire, and estimated the delay distribution in their 3D integration schematic with a RC delay model. Based on a geometric programming tool, Pavlidis *et al* [44] proposed an approach for vias placement to minimize the total delay of inter-tier interconnects with the consideration of the non-uniform impedance characteristics of interconnects. Their results showed an average performance improvement

of 16% can be achieved, compared to the approach where vias are equally spaced in a wire. However, for high-speed or ultra high-speed 3D IC systems, the important signal integrity issues due to the non-uniform impedance characteristics were not addressed in both [58] and [44]. It is notable that signal integrity issues such as reflection caused by impedance discontinuities can be very deleterious to digital VLSI systems, since it may significantly affect the logical operation and reliability of a VLSI system.

Recently, Lee *et al* [38] proposed a novel approach for redundant via insertion to improve yield/reliability and manufacturability in traditional 2D ICs under via density constraints. This approach is an extension of conventional via insertion/placement, since it can help to decrease partial or complete via failure due to various reasons (like manufacturing process). According to their approach, the yield/reliability may be significantly improved by inserting redundant via properly without violating via density constraints.

In this research, we extend the redundant via insertion in [38] to 3D ICs and propose an approach for redundant via placement/insertion to minimize the total delay of inter-tier or global interconnects. The issue of signal integrity due to the non-uniform impedance characteristics of interconnects is also carefully addressed. The main contributions of our work include:

- We extend the idea for redundant via insertion to 3D ICs, and propose an approach to minimizing the total interconnect delay based on the redundant via insertion/placement.
- We address signal integrity issues of interconnects with the consideration of non-uniform impedance characteristics in 3D ICs. In particular, we consider the signal reflection issues that may result in deleterious effects such as ring and undershoot, and propose impedance matching to decrease the reflection coefficient based on the redundant via insertion.
- We formulate the above considerations to a multi-objective optimization and propose an efficient algorithm to solve it. In addition, we prove that the optimization related to our approach is practical.

The rest of this research is organized as follows. Some preliminaries about 3D ICs and vias placement/insertion are introduced in Section 5.2. The problem formulation is described in Section 5.3. In Section 5.4, the simulation methodology is presented.

The simulation results and discussion are provided in Section 5.5. Finally, in Section 5.6, we conclude this research.

5.2 3D ICs and Vias Placement/Insertion

Three-dimensional (3D) integrated circuits (ICs), which contain multiple layers of active devices, have the potential to dramatically enhance chip performance, functionality, and device packing density. They also provide for microchip architecture and may facilitate the integration of heterogeneous materials, devices, and signals.

Compared to the conventional planar ICs, the 3D ICs are more suitable for the integration of heterogeneous materials, devices, signals and technologies. Usually, multiple tiers (also called planes) are included in a 3D IC, and these physical tiers are closely and vertically stacked using bonding process. In 3D ICs, interconnects routed in different tiers are connected with bonding medium (adhesive medium or metal pads) [55] to construct whole signal paths and achieve inter-tier communication. Therefore, the impedance characteristics of interconnects and vias may be very different from each other, namely, the impedance characteristics of inter-tier signal paths are non-uniform. Some approaches have been proposed in [58] and [44] to evaluate the impact of vias placement/insertion upon total delay in the presence of non-uniform impedance.

In the traditional 2D ICs, via discontinuities have a negligible effect on the propagated edge rate and the near end or far end coupled noise, and its effect on delay is also insignificant [20]. In the 3D IC systems, however, these effects can be significant. Although the modern techniques can thin a 200mm silicon wafer to $\sim 20\mu m$ in 3D ICs manufacturing [44], but the height of a single via in a real 3D IC may be much higher than $20\mu m$, because the vias there usually need to across several tiers. Even for single tier vias, their height can still be larger than $20\mu m$, which is the minimum achievable wafer height now. Thus, the vias height in a 3D IC is usually much longer than the vias height in its 2D counterpart. As a result, the vias in the 3D ICs not only affect the interconnect delay more significantly but also result in a more serious impedance discontinuity between wires and vias. It is notable that the impedance discontinuity will cause signal reflection and thus a signal integrity degradation, because the signal reflection occurs whenever the impedance characteristic changes along the signal propagation path [14]. Therefore, it is necessary to consider simultaneously

the impacts of vias on delay and signal integrity to evaluate their real performance in 3D ICs.

Lee *et al* [38] proposed a redundant via insertion approach in 2D ICs using the methods of end-line extension and redundant vias insertion adjacent to a single via (we call these methods as redundant via insertion for simplicity hereafter). Fig.5.3 shows the structure of redundant via insertion when the line end extension is applied. In an 2D IC layout, a via provides the connection between net segments from neighboring metal layers. Partial or complete vias failure may occur due to different reasons, such as cut misalignment and line-end shortening during a fabrication process, electromigration and thermal stress. As a consequence, these via failures usually result in an increased contact resistance and parasitic capacitance, or leave an open net in a circuit and invalidate the functionality of overall design. By redundant via insertion, both partial and complete vias failure can be alleviated. Exploiting the methods and algorithms proposed in [38], via density constraint will not be violated, while the yield/reliability and manufacturability can significantly be improved.

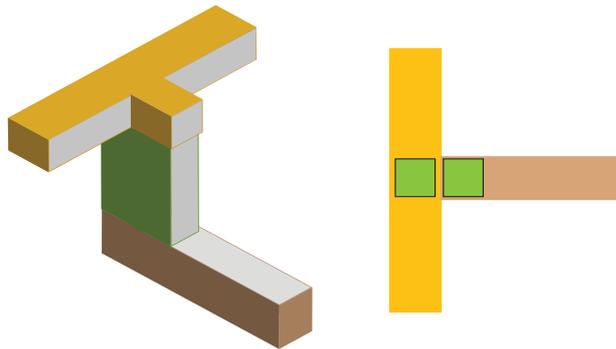


Fig. 5.3. Redundant via insertion (line end extension)

In vertical integration-based 3D ICs, the reliable formation of high-aspect-ratio (AR) vias are required to connect different wafers or chips to achieve communication among them [55]. However, all metallization techniques have their specific limitations on the maximum available aspect ratio of vias, which will result in an additional design constraint with respect to the layout of 3D ICs. Considering such a condition, it is necessary to pay attention to the issues of reliability and limitation of vias fabrication in 3D ICs, compared to vias in conventional planar VLSI systems. By the application of redundant vias insertion in 3D ICs, it is possible to decrease the

aspect ratio of vias. Also, it can relax design constraint and improve the reliability and manufacturability/yield of 3D VLSI systems like that in 2D VLSI systems.

Moreover, by inserting a redundant via near the original via, the current handling capacity of via can be increased, and the impedance characteristic and parasitics of via may also be altered. Therefore, the idea of redundant via insertion introduced in [38] offers us an opportunity to conduct via sizing or vias insertion to minimize the delay and to consider signal integrity simultaneously without violating via density constraint. For example, we can extend via size along the wire routing direction or the direction perpendicular to the wire routing direction in available routing area. It is similar to inserting an additional via at the side of original via (see Fig.5.4). Consequently, in this research, we extend this idea to 3D ICs, and propose a redundant via insertion-based approach to minimizing the delay of inter-tier interconnect while addressing also the signal integrity issues.

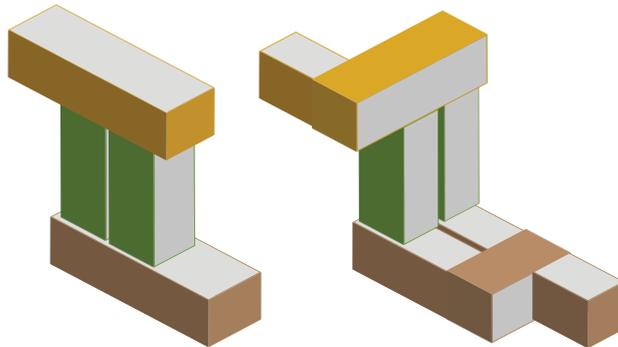


Fig. 5.4. Via sizing by redundant via insertion.

5.3 Problem Formulation and Optimization

In this section, we first describe the delay calculation of inter-tier interconnects, then we introduce the impedance matching that is helpful for improving signal integrity. We summarize the problem formulation and propose the optimization method in Section 5.3.3.

5.3.1 Delay Modeling

Since the delay of interconnects is crucial to the performance of VLSI systems, it is important to find an optimum scheme of via placement for an interconnect in 3D ICs to minimize the total delay of the interconnect with the consideration of non-uniform impedance.

As the interconnect length and operating speed entered the nanoscale regime and gigascale regime, respectively, the inductance component becomes comparable to resistance component in circuits of VLSI (specially for Cu-based interconnect technology with a low resistance) [30]. Thus, the more advanced RLC model should be adopted to fully analyze the timing feature of interconnects.

In this research, our primary goal is to minimize total signal propagation delay for given inter-tier interconnects. Without loss of generality, we consider a global inter-tier distributed RLC interconnect going through n tiers, as illustrated in Fig.5.5. Since buffer insertion is an efficient method to satisfy delay constraint for long wire, we assume that one inverter is inserted in each tier (please refer to Fig.5.5).

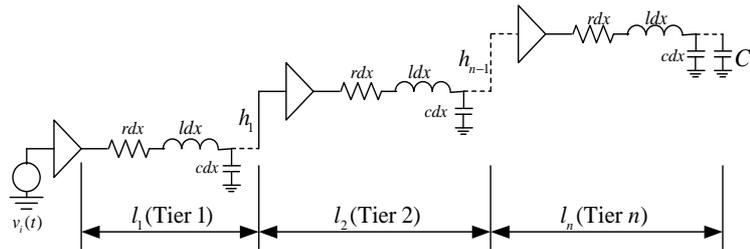


Fig. 5.5. Equivalent circuit of driver-interconnect-load structure of inter-tier interconnect in 3D IC.

We use l_i ($i = 1, \dots, n$) to denote the length of segment that is routed in Tier i , and use h_i ($i = 1, \dots, n - 1$) to denote the height of via that connects Tier i and Tier $(i + 1)$. The sum of length of all the segments and the height of vias should be the total given length L , i.e.,

$$\sum_{i=1}^n l_i + \sum_{i=1}^{n-1} h_i = L \quad (5.1)$$

The total delay D of such an interconnect is the sum of all the segments delay d_{l_i} and vias delay d_{h_i} , which can be expressed as:

$$D = \sum_i^n d_{l_i} + \sum_i^{n-1} d_{h_i} \quad (5.2)$$

Since each tier i ($i = 1, \dots, n$) may be of different impedance characteristic, we can alter the position of each via to minimize the total delay of the interconnect. The length of segments l_i should satisfy [44]:

$$l_{\min_i} \leq l_i \leq L - \sum_{j=1, j \neq i}^{n-1} l_{\min_j} - \sum_j^{n-1} h_j \quad (5.3)$$

Here l_{\min_i} is the minimum permitted length of segment in Tier i , and it is determined by design rule or design constraints.

5.3.2 Impedance Matching

If one signal is traveling down an interconnect and the instantaneous impedance the signal encounters at each step ever changes, some of the signal will be reflected and the remaining distorted signal will continue down the line. These reflections and distortions lead to degradation in signal quality and cause signal integrity issues, such as ring and undershoot [14].

It is notable that the signal rise time is also an important factor that affects reflections [14, 29]. The author in [14] suggests that when the impedance changes at vias and corners, the signal rise time should be considered carefully in the analysis of signal reflections. The work in [29] further indicates that the signal rise time begins to have a significant effect on the signal wave shape when it is smaller than the two times of the transmission line delay. Thus, it yields a signal frequency range on which the reflection resulted from via discontinuities should be considered carefully in 3D ICs. For example, for a $20\mu m$ ($40\mu m$) interconnect wire/via driven by an ideal source with a 0.05pF load, its RLC propagation delay based on transmission line theory [30] can be as high as 4.4ps (6.4ps) based on the process parameters in [12]. If the signal rise time is assumed to be $1/6$ of the signal period (i.e., $t_{rise} = 1/6T$ under 50% duty cycle)¹,

¹Notice that the system designers usually try to make the signal rise time as small as possible to provide an enough bandwidth [14].

the above results indicate that a lower bound of signal frequency is about 18GHz (14GHz) for considering the reflections. Based on the process parameters in [30], the corresponding RLC propagation delay will be about 0.9ps (1.2ps) for 20 μm (40 μm) wire/via, which results in a lower bound of signal frequency of 90GHz (70GHz) for the consideration of signal reflection issue resulted from via discontinuities in 3D ICs.

Due to the inspiring progress on the developments of terahertz devices and interconnect systems, the sub-terahertz and terahertz VLSI technologies have been proposed recently [3, 52]. For such technologies, the signal rise time will be much smaller than that of current gigahertz technology. Therefore, the consideration of impedance matching becomes very important for reflection reduction in the design of 3D terahertz or sub-terahertz VLSI systems.

In this section, we consider the issue of impedance matching to alleviate signal reflections in 3D ICs based on the redundant vias insertion technique. The basic idea is to adjust the characteristic impedance of vias by vias sizing, such that the reflection coefficient between vertical vias and horizontal interconnects can be minimized and thus the impedance matching can be achieved. Therefore, we regard the impedance matching (i.e., the minimization of reflection coefficient) as the our second goal in the performance enhancement of 3D vias insertion (remember our first goal is to minimize the total delay of interconnects).

For impedance matching, we need to first determine the reflection coefficient. According to the transmission line theory, the amount of signal that is reflected depends on the magnitude of the change in the instantaneous impedance. The reflection coefficient ρ is used to measure the amount of reflected signal and it is expressed by following Equation (5.4) when signal enters segment 2 from segment 1 [14]:

$$\rho = \frac{Z_2 - Z_1}{Z_2 + Z_1} \quad (5.4)$$

Here Z_1 is the instantaneous impedance of the segment 1 from which the signal initially enters, Z_2 is the instantaneous impedance of the segment 2 where the signal just enters.

The instantaneous impedance of wire depends on the cross section of the wire and the material properties [14], and it should be equal to the characteristic impedance of the wire for an impedance-controlled interconnect. The characteristic impedance Z_0

of transmission line with distributed R , L , C and G components can be determined by the following Equation (5.5) [54]:

$$Z_0 = \sqrt{\frac{R + j\omega L}{G + j\omega C}} \quad (5.5)$$

In the cases when the frequency is high enough, Equation (5.5) can be reduced to [54]:

$$Z_0 = \sqrt{\frac{R + j\omega L}{G + j\omega C}} \approx \sqrt{\frac{L}{C}} \quad (5.6)$$

Thus, based on Equation (5.4) and Equation (5.6), we can calculate the reflection coefficient ρ_i ($i = 1, \dots, n-1$) when signal enters via h_i from a segment l_i of interconnect in Tier i as:

$$\rho_i = \frac{Z_{0l_i} - Z_{0h_i}}{Z_{0l_i} + Z_{0h_i}} \quad (5.7)$$

Then we can minimize the reflection coefficient ρ_i using vias sizing based on the idea of redundant vias insertion introduced in Section 5.2.

5.3.3 Optimization of Delay and Reflection Coefficient

Based on the above analysis, we summarize our problem as follows.

Given a total length of L for inter-tier interconnect through n tiers and the permitted maximum/minimum via size $V_{via-max}/V_{via-min}$ (the height of via is fixed and can not be changed), to find a proper size and position of each via ($n-1$) to minimize the total delay D of the interconnect and the reflection coefficient of each tier ρ_i :

Minimize

$$\begin{aligned} D &= \sum_i^n d_{l_i} + \sum_i^{n-1} d_{h_i} \\ \rho_i &= \frac{Z_{0l_i} - Z_{0h_i}}{Z_{0l_i} + Z_{0h_i}}, i = 1, \dots, n-1 \end{aligned}$$

subject to

$$\begin{aligned} \sum_{i=1}^n l_i + \sum_{i=1}^{n-1} l_i &= L \\ l_{\min_i} \leq l_i &\leq L - \sum_{j=1, j \neq i}^{n-1} l_{\min_j} - \sum_j^{n-1} h_j \\ V_{via-min} \leq V_{via_{h,i}} &\leq V_{via-max}, i = 1, \dots, n-1 \end{aligned}$$

This is a multi-objective optimization problem, and it can be decomposed into two equivalent sub-optimization problems: via-sizing and vias insertion/placement. Based on an iteration procedure, we can solve this optimization problem.

We first set the size of each via to an initial value (for example, the permitted minimum size) and perform vias insertion/placement to minimize the total delay with the consideration of non-uniform impedance. For each vias insertion/placement corresponding to the minimum total delay, we then alter the size of each via to minimize reflection coefficient ρ_i ($i = 1, \dots, n - 1$). Since the total delay of the interconnect may be changed for each new vias size, we need to verify the total delay based on the new vias size. If a greater total delay is obtained, a new vias insertion/placement will be conducted again based on the new vias size. Based on the above iteration, we will finally find the minimum total delay for the interconnect and the corresponding reflection coefficient. We summarize the above procedure as the following algorithm:

1. Initialize $V_{via_i} = V_{via-\min}$, $i = 1, \dots, n - 1$
2. Perform vias insertion/placement to minimize the total delay of whole interconnect
3. Repeat
 - (a) Resize each via V_{via_i} ($V_{via_i} \leq V_{via-\max}$), compute and find the minimum ρ_i , $i = 1, \dots, n - 1$
 - (b) Check the total delay
 - (c) Perform vias insertion/placement for a greater total delay
 - (d) Update vias' geometric size, position and the total delay of interconnect
4. Return the optimum results

Since the expressions of segment and via delay can not be expressed as a simple linear form, the Lagrangian relaxation can be utilized to pre-process them in order to obtain the solution more efficiently.

5.4 Simulation Methodology and Setup

To verify our approach proposed in this research, we conducted some simulations. The simulation methodologies are described in this section.

5.4.1 Delay Calculation

Since we adopt a distributed RLC interconnect structure to model the interconnects in 3D ICs, we calculate the delay of interconnects based on a distributed RLC model proposed in [30]. According to [30], an empirical RLC delay equation based on curve-fitting was derived as:

$$t_{50\%} = (e^{-2.9\zeta^{1.35}} + 1.48\zeta)/\varpi_n \quad (5.8)$$

where

$$\varpi_n = \frac{1}{\sqrt{L_{int}(C_{int} + C_L)}} \quad (5.9)$$

$$\zeta = \frac{R_{int}}{2} \sqrt{\frac{C_{int}}{L_{int}} \frac{R_T + C_T + R_T C_T + 0.5}{\sqrt{1 + C_T}}} \quad (5.10)$$

$$R_T = R_s/R_{int} \quad (5.11)$$

$$C_T = C_L/C_{int} \quad (5.12)$$

here C_L is the load capacitance, R_{int} , C_{int} and L_{int} are the total line resistance, capacitance, and inductance, respectively, and R_s is the output resistance of driver (inverter).

Remark 1: *The objective function D (Equation (5.2)) based on the delay Equation (8) (Equation (5.1)) is convex under constraint L .*

Proof First we observe that the delay of a wire can be considered as the function of its length, since the electrical parameters in Equation (5.8) like interconnect resistance (R_{int}), capacitance (C_{int}) and inductance (L_{int}) can be written as the function of interconnect length. Also, the ζ (Equation (5.10)) is non-negative under interconnect length constraint L (Equation (5.1)). Furthermore, since the function $f(x) = e^x$ is strictly convex in the interval $(-\infty, +\infty)$, so the monomial $e^{-2.9\zeta^{1.35}}$ is also convex. Additionally, it is notable that ϖ_n (Equation (5.9)) is non-decreasing function. According to [10], the delay equation (5.8) is hence convex. Similarly, we can know that the objective function D (Equation (5.2)) is also convex. ■

According to Remark 1, the objective function D has a single global minimum, which implies that our approach is practical under the RLC delay model.

It is notable that other delay models/equations can also be used in our approach. For example, we can prove that the objective function D with the traditional or fitted Elmore delay model is also convex. Notice that the objective function D expressed by Equation (5.2) is of a separable non-linear form, so it can also be transformed into a form solved by geometric programming tools efficiently.

5.4.2 Parasitic Extraction of Vias

Since we intend to minimize the reflection coefficient of different segments and the delay resulted from vias, it is necessary to calculate the parasitic parameters such as resistance, capacitance and inductance of vias. Nevertheless, it is impractical to extract accurate parasitic parameters with existing equations rapidly. In this research, we apply the following methods to extract the resistance, capacitance and inductance of vias respectively.

The vias' resistance R_{via} can be simply evaluated as $R_{via} = r \cdot l/w$, where r , l and w are the resistance of unit length interconnect, the interconnect length and interconnect width, respectively.

For the evaluation of capacitance, we adopt a quasi-3D on-chip capacitance model proposed in [48] to calculate interconnect capacitance. Since the detailed layout of vias is unknown, we can only calculate the self component of vias' capacitance, i.e., the capacitance of vias coupling to substrate.

Finally, we need to calculate the inductance used in the RLC delay model. Notice that it is usually formidable to extract the accurate interconnect inductance, because the current return paths are very complicated in a real chip. To make the evaluation of interconnect inductance tractable, we adopt here the formulas proposed in [46] to extract the inductance. Similar to the calculation of capacitance, the mutual inductance between one via and the others can not be computed without the detail layout.

Notice again that we use those parasitic extraction equations described here only for enabling the simulation of our approach, the accurate parasitic parameters can be extracted accurately with 3D field solver by EDA tools when detailed layouts of chip are almost determined.

5.4.3 Parameters Settings

Our simulation are conducted based on a 65nm CMOS technology under the Berkeley Predictive Technology Model (BPTM) [1]. We assume that a 50X size inverter is inserted into each segment of tier. The output resistance R_s and input capacitance of inverter are calculated with SPICE. The load capacitance C_L is $0.1pF$.

To mimic the different impedance characteristics of each tier in 3D ICs, the range of resistance, capacitance of interconnect are the same as the ranges proposed in [44] and they are extracted for several interconnect structures using a commercial impedance extraction tool. The resistance of segments ranges from $5\Omega/mm$ to $25\Omega/mm$, and the capacitance of segments ranges from $100fF/mm$ to $300fF/mm$. The inductance ranges from $650pH/mm$ to $17.9nH/mm$, where the minimum value is the same as that in [30] and the maximum value is the same as that used in [12]. The minimum length routed in each tier is set as $20\mu m$.

We assume that the height of via that connects two tier ranges from $20\mu m$ to $40\mu m$. Depending upon the size of via, the resistance, capacitance and inductance of via are calculated using the methods/equations mentioned in Section 5.4.2. The cross-section of via is supposed to be rectangle and its size ranges from $100nm \times 100nm$ to $300nm \times 600nm$.

We tested 3 inter-tier interconnects whose length are $1.5mm$, $2.5mm$ and $3.5mm$, respectively. These interconnects are supposed to be routed in 3D ICs with 4, 6, 8 and 10 tiers. We also assume that the tiers are bonded with copper metal pads.

Note that as a demonstration in this research, the frequency-dependent components of resistance, capacitance and inductance are not included in our simulation.

5.5 Simulation Results

We accomplished the simulations with MATLAB. In this section, we first verify the impact of vias sizing on interconnect delay, then we demonstrate the improvements of delay and reflection coefficient when both vias sizing and vias placement are considered simultaneously. Then, we compare between vias sizing and via placement their impacts on delay improvement in presence of different impedance characteristics in 3D ICs. Finally, we explore the delay improvement when both the vias sizing/placement and buffer insertion are jointly considered.

5.5.1 Impact of vias sizing on Interconnect Delay

To verify the impact of vias sizing on interconnect delay, we conducted several simulations when the vias sizing is considered while the impedance matching is ignored. Note that these simulations are the special cases of the proposed algorithm where the impedance matching is disabled and vias are placed along the interconnects with equal-space.

The calculated delay results are provided in Table 5.1. Here, D_{equ} is the delay results for interconnects divided equally by vias with the minimum size, D_{opt} is the optimum delay results. The delay improvements are also listed in Table 5.1.

Table 5.1
Delay results of interconnects when considering only the vias sizing

Tiers (n)	Length of wires (L)	D_{equ} (ps)	D_{opt} (ps)	Delay Improvement (%)
4	1.5mm	32.87	30.61	6.88
	2.5mm	40.36	39.02	3.32
	3.5mm	52.54	50.34	4.19
6	1.5mm	34.74	32.58	6.22
	2.5mm	43.41	41.61	4.15
	3.5mm	54.96	52.35	4.75
8	1.5mm	35.79	33.63	6.04
	2.5mm	46.75	45.12	3.49
	3.5mm	56.81	54.01	4.93
10	1.5mm	38.26	35.72	6.64
	2.5mm	49.34	46.74	5.27
	3.5mm	57.49	54.52	5.17
Average				5.1%

The results in Table 1 show that an average delay improvement of 5.1% can be achieved if only the vias sizing is considered. The maximum delay improvement is about 6.9% for a 1.5mm wire routed 4-tiers 3D IC and the maximum delay decrease

of almost $3ps$ can also be observed. Taken together, these results prove that our approach about vias sizing is effective to improve interconnects delay in 3D ICs.

5.5.2 Overall Impacts of Vias Sizing/Placement on Interconnect Delay and Reflection Coefficient

In this section, we first explore the overall impacts of vias sizing and vias placement on interconnect delay, then we show their impacts on improvement of reflection coefficient.

Improvement of Interconnect Delay

For different combinations between tiers and wire length, we summarize in Table 5.2 the optimum interconnect delay (D_{opt}) determined by our optimization framework in Section 5.3.3 and also the interconnect delay (D_{equ}) when the wire is equally divided by vias with the minimum size.

The results in Table 5.2 indicate that the total delay for all 3 inter-tier interconnects based on our approach are smaller than that of interconnects divided by vias equally, where the maximum delay improvement from our approach is as high as about 16% for a $1.5mm$ wire routed in a 10-tiers 3D IC. We can also see from Table 5.2 that an average delay improvement of 8.7% is obtained for all 3 interconnects, compared to the instance where interconnects is divided by vias equally.

A further observation to results in Table 5.2 reveals that the relatively shorter interconnects routed in many tiers have greater delay improvement than longer ones routed in few tiers. It means that the relatively shorter interconnects are more sensitive to different impedance characteristics of tiers in 3D ICs than longer ones.

It is notable that delay improvement in our approach is smaller than that in [44]. The reason is that a distributed RLC delay model is adopted and a inverter is inserted into the segments in each tier in our simulation to imitate the actual scenario, so the overall delay of interconnect has already been significantly reduced, thus leave a small space for further delay reduction.

Table 5.2
Delay results of interconnects

Tiers (n)	Length of wires (L)	D_{equ} (ps)	D_{opt} (ps)	Delay Im- provement (%)
4	1.5mm	32.87	29.63	9.86
	2.5mm	40.36	37.47	7.16
	3.5mm	52.54	48.79	7.14
6	1.5mm	34.74	30.42	12.44
	2.5mm	43.41	39.82	8.27
	3.5mm	54.96	51.87	5.62
8	1.5mm	35.79	31.45	12.13
	2.5mm	46.75	44.08	5.71
	3.5mm	56.81	53.43	5.95
10	1.5mm	38.26	32.22	15.79
	2.5mm	49.34	45.18	8.43
	3.5mm	57.49	54.17	5.77
Average				8.7%

Improvement of Reflection Coefficient

To explore the effects of impedance matching based on our approach, we calculate the average reflection coefficient of all 3 interconnects for both optimum method (ρ_{opt}) proposed in this research and equal-space vias placement (ρ_{equ}). The corresponding results are listed in Table 5.3.

In the Table 5.3, a significant difference can be observed in the reflection coefficient, depending on whether the impedance matching is considered or not. The maximum improvement is up to 37% for 3.5mm wires routed in a 10-ters 3D IC; and the average improvement of reflection coefficient for all 3 interconnects is as high as about 26%. The low relatively improvement for the 1.5mm wire routed in a 8-tiers 3D IC appears (the absolute improvement of reflection coefficient is only 0.01). It is likely that our approach can not handle them to obtain a small reflection coefficient, due to a

Table 5.3
Results of reflection coefficient

Tiers (n)	Length of wire (L)	ρ_{equ}	ρ_{opt}	Improvement (%)
4	1.5mm	0.11	0.09	18.18
	2.5mm	0.14	0.12	14.29
	3.5mm	0.13	0.11	15.38
6	1.5mm	0.14	0.09	35.71
	2.5mm	0.21	0.13	38.1
	3.5mm	0.21	0.14	33.33
8	1.5mm	0.14	0.13	7.14
	2.5mm	0.24	0.21	12.5
	3.5mm	0.23	0.16	30.43
10	1.5mm	0.18	0.12	33.33
	2.5mm	0.22	0.14	36.36
	3.5mm	0.24	0.15	37.5
Average				26

big difference in the characteristic impedance between segments in the wire and vias. Nevertheless, we obtained a considerable average improvement of reflection coefficient in the overall cases.

Generally, the results in this section reveal that our approach is effective for vias insertion/placement in 3D ICs. It can not only reduce the total delay of inter-tier interconnects but also reduce the signal reflection and thus improve the signal integrity.

Notice that we calculate the characteristic impedance Z_0 of all segments and vias approximately with Equation (5.6) here. An accurate calculation of characteristic impedance of segments and vias can be achieved by some EDA tools with field solver. Other accurate characteristic impedance formulas can also be integrated into our approach without affecting the validity of our approach.

Vias Placement and Vias Insertion for an Example Wire in 3D ICs

To illustrate the performance improvement of the proposed approach in this chapter, we give here the detailed routing result for an example wire with specified impedance characteristics.

In this example, we assume that a 3.0mm wire is routed in a 6 tiers 3D IC where their impedance characteristics are listed in Table 5.4. Other parameters are the same as those given in Section 5.4.

Table 5.4
The impedance characteristics of in a 6-Tires 3D IC

Tier	Resistance (Ω/mm)	Inductance (nH/mm)	Capacitance(fF/mm)
1	22	3.2	150
2	15	10.5	110
3	8	15	200
4	20	8	250
5	25	17	180
6	10	6.5	280

Table 5.5
Optimum routing results for the 3mm example wire

Tier	Segment Length in Each Tier (μm)	Size of Vias Between Neighboring Tiers ($nm \times nm$)
1	1250	100X160
2	745	100X220
3	350	240X260
4	215	120X300
5	165	160X280
6	275	—

The routing results for the wire are listed in Table 5.5 where both vias size and segments length are given detailedly. Here, The total propagation delay is 52.3ps and the average reflection coefficient is 0.22 when the interconnect is divided vias with the minimum size specified in Section 5.4. The corresponding optimum results are 50.1ps for propagation delay and 0.15 for reflection coefficient. Thus, we obtained a delay reduction of 4.3% and reflection coefficient improvement of 31.8% in this example.

5.5.3 Comparison Between Vias Sizing and Vias Placement in Delay Improvement

Table 5.6
Delay results of a 2.5mm wire in presence of different impedance characteristics

Scenario	Tiers	D_{equ} (ps)	<i>Simulation A</i> (Vias sizing only)		<i>Simulation B</i> (Vias sizing and vias placement)	
			D_{opt} (ps)	Improvement (%)	D_{opt} (ps)	Improvement (%)
S1	4	41.36	40.02	3.24	39.87	3.6
	6	43.54	41.61	4.43	41.03	5.76
	8	46.61	44.32	4.91	43.78	6.07
	10	48.58	45.74	5.85	45.23	6.9
	Average			4.61	—	5.58
S2	4	42.61	40.82	4.2	39.13	8.17
	6	43.95	41.71	5.1	40.92	6.89
	8	47.06	44.54	5.35	43.38	7.82
	10	50.47	47.26	6.36	45.49	9.87
	Average			5.25	—	8.19

To compare the effects of vias sizing and vias placement on the interconnect delay in presence of different impedance characteristics in 3D ICs, we also performed several extended simulations. Here, we consider two scenarios in our simulations for a 2.5mm

wire routed in 3D ICs with different tiers: in the first scenario (called *S1* hereafter), we assume that the variation range of a parasitic parameter among all tiers is only half of the maximum range discussed in Section 5.4.3, while in the second scenario (named as *S2* hereafter) the variation range of the parameter is just its maximum range. We first simulated the delay when only the vias sizing is considered (*Simulation A*), then evaluated the delay when both vias sizing and vias placement are considered simultaneously (*Simulation B*). The corresponding results are listed in Table 5.6, where the notations D_{equ} and D_{opt} are defined in the same way as that of Section 5.5.2. For comparison, the delay improvements compared with D_{equ} are also included in Table 5.6.

The Table 5.6 indicates clearly that a certain delay improvements can always be achieved in *Simulation A* and *Simulation B* for both Scenario *S1* and Scenario *S2*. Compared with the delay results for equal-space vias placement D_{equ} , the average delay improvements in *Simulation A* and *Simulation B* for Scenario *S1* are 4.61% and 5.58%, respectively; while the corresponding results for Scenario *S2* are 5.25% and 8.19%, respectively. These results show that the delay improvements in *Simulation B* are bigger than those in *Simulation A*, suggesting that our proposed approach based on the combination of via sizing and placement is more effective for delay performance enhancement in 3D ICs.

However, a further observation of Table 5.6 shows that the difference of delay improvement between *Simulation A* and *Simulation B* in Scenario *S1* is much smaller than that of Scenario *S2*. For example, in the Scenario *S1*, the delay improvements are 4.61% for *Simulation A* and 5.58% for *Simulation B*, resulting a delay improvement difference 0.97%; while in the Scenario *S2*, the delay improvements are 5.25% for *Simulation A* and 8.19% for *Simulation B*, so a delay improvement difference 2.94% is achieved between *Simulation A* and *Simulation B*. Notice that the magnitude of parasitic parameters variation in *S2* is two times as that of *S1*, so the the above comparison results indicate that when the variation of impedance characteristics is relatively small among different tiers in a 3D IC, the vias placement technique can not reduce the interconnect delay significantly, while vias sizing can be an effective approach for delay improvement. When impedance variation among different tiers is relatively big, however, our above results imply that the vias placement is more efficient than the via sizing in delay improvement.

Based on the above results, we can conclude that: 1) when variation of parasitic parameters among different tiers is small (e.g., Scenario $S1$), vias sizing is very effective for improving interconnect delay in 3D ICs; 2) in the case that the impedance variation among different tiers is relatively big (e.g., Scenario $S2$), the vias sizing should be integrated with the vias placement for a significant delay improvement.

5.5.4 Impact of Vias Sizing and Buffer Insertion on Delay Improvement

Buffer insertion is a traditional method to reduce the propagation delay of an interconnect by partitioning it into shorter sections. To assess the impact of vias sizing/placement and buffer insertion on delay improvement, we conducted several simulations that integrate buffer insertion and vias insertion together.

In our simulations, the vias insertion/placement is first conducted for an interconnect, then the buffer insertion is performed to further reduce its delay. Considering the fact that tiers in a 3D IC may be manufactured by different process, so the buffer insertion is performed for different tiers separately. We adopted the buffer insertion method proposed in [30], which presents the closed-form solutions for both the optimum section length h_{opt} and buffer size k_{opt} with the consideration of inductance effect. Based on the optimum section length h_{opt} , the buffer insertion will be only applied to these segments which are longer than h_{opt} . Simulations have been performed for a $1.5mm$ wire and a $3.5mm$ wire that are routed through 4, 6, 10 tiers, respectively. The corresponding simulation results are listed in Table 5.7, where D_{equ} is the interconnect delay when the wire is equally divided by vias with the minimum size, $D_{opt-spr}$ is the delay results when both vias sizing/placement and buffer insertion are jointly considered. For comparison, the optimum delay results based only on vias sizing/placement (D_{opt-sp}) and their delay improvements compared with the D_{equ} are also included in Table 5.7.

The results in Table 5.7 show that more delay improvements can be obtained by integrating buffer insertion with vias sizing/placement. For example, the average delay improvement are 13.6% for the $1.5mm$ wire and 11.2% for the $3.5mm$ wire when both vias sizing/placement and buffer insertion are considered, and these average delay improvements are bigger than the corresponding delay improvements 12.7% and 6.2% when only the vias sizing/placement is applied.

Table 5.7
 Delay results for vias sizing/placement and buffer insertion

Wires	Tiers	D_{equ} (ps)	Vias sizing/placement only		Vias sizing/placement and buffer insertion	
			D_{opt-sp} (ps)	Improvement (%)	$D_{opt-spr}$ (ps)	Improvement (%)
1.5mm	4	32.87	29.63	9.86	28.74	12.56
	6	34.74	30.42	12.44	30.42	12.44
	10	38.26	32.22	15.79	32.22	15.79
	Average			12.7	—	13.6
3.5mm	4	52.54	48.79	7.14	45.35	13.68
	6	54.96	51.87	5.62	49.16	10.55
	10	57.49	54.17	5.77	52.1	9.38
	Average			6.2	—	11.2

However, a careful observation of the above results indicates that compared with the 3.5mm wire, the difference of delay improvements between D_{opt-sp} and $D_{opt-spr}$ for the 1.5mm wire is not significant. These delay improvement differences are due to the reasons for the shorter wires, the segments routed in different tiers are too short to adopt buffer insertion for delay improvement, so the buffer insertion there is not a main contributor to delay improvement. For the longer wires, however, it is clear that the buffer insertion can be very effective for delay improvement as expected.

In summary, the above results show that for shorter wires routed in many tiers, vias sizing/placement rather than buffer insertion should be primarily considered for delay improvement. On the other hand, for the longer wires that are routed in less tiers, integrating buffer insertion with vias sizing/placement will be more effective for reducing delay.

5.6 Conclusions

By extending idea of redundant vias insertion in 2D ICs, we propose in this research an approach for vias placement and impedance matching in 3D ICs. The simulation results demonstrated that our approach is effective and can result in a significant improvement in interconnects delay and reflection coefficient for 3D IC systems, especially in the post routing design stage of such systems.

We expect that the idea of redundant vias insertion/placement can also be exploited for the performance improvement of both power and clock distribution networks in 3D ICs with the consideration of process variation issues [33,34]. Another possible extension of this work is the thermal vias insertion/placement for the stability enhancement of 3D VLSI systems.

Chapter 6

CONCLUSION

6.1 Summary

In this research, we focus mainly on the timing issues of clock distribution in high-speed VLSI systems. The main contributions of our research are summarized as follows.

We first conducted several simulations to explore the real performance of nanoscale high-speed clock distribution networks in presence of process variations and inductance effect. The simulation results show that: 1) A significant overestimation of the H-Tree CDNs' performance may be introduced if the inductance effect is neglected in the delay evaluation, especially for the modern VLSI systems, where the low resistance material is adopted and the inductance component of line impedance becomes comparable to the resistive component; 2) The performance of H-Tree CDNs can also be significantly overestimated if the coupling effects are not taken into consideration, since the coupling capacitance and the coupling inductance (mutual inductance) are becoming dominant over their self components with the continuous increase of the wire routing density; 3) Although the spatial dependence of process variations does not significantly affect the maximum clock delay of H-Tree CDNs, it may remarkably affect the clock skew and the standard deviation of the maximum delay; 4) The standard deviations of the maximum delay and clock skew vary significantly with the increase of the magnitude of process variation, although their mean values are not very sensitive to the magnitude of process variation. Thus, the yields of the maximum clock delay and clock skew may be significantly degraded by magnitude of process variation.

Second, we proposed a novel Variant X-Tree CDN based on X Architecture and a statistical performance model used to evaluate its performance rapidly. Simula-

tion results indicate that comparing to the traditional H-Tree CDN, the proposed new CDN has the potential to improve the overall clock distribution performance in terms of maximal clock delay and clock skew. We also study the layout features of Variant XTree in detail. It enables to determine the proper size Variant X-Tree clock distribution network with these rules. A statistical performance evaluation model is proposed as well. Experiment simulation results show it is suitable to estimate the performance of a Variant X-Tree in design stage can also be integrated into design flow easily for its closed-form.

Finally, by extending idea of redundant vias insertion in 2D ICs, we propose in this paper an approach for vias placement and impedance matching in 3D ICs. The simulation results demonstrated that our approach is effective and can result in a significant improvement in interconnects delay and reflection coefficient for 3D IC systems, especially in the post routing design stage of such systems. We expect that the idea of redundant vias insertion/placement can also be exploited for the performance improvement of both power and clock distribution networks in 3D ICs with the consideration of process variation issues. Another possible extension of this work is the thermal vias insertion/placement for the stability enhancement of 3D VLSI systems.

6.2 Recommendations for Future Works

The resistance of copper interconnects, with cross-sectional dimensions of the order of the mean free path of electrons ($\sim 40nm$ in Cu at room temperature) in current and imminent technologies [2], is increasing rapidly under the combined effects of enhanced grain boundary scattering, surface scattering and the presence of the highly resistive diffusion barrier layer [51]. The steep rise in parasitic resistance of copper interconnects poses serious challenges for interconnect delay [50] (especially at the global level where wires traverse long distances) and for interconnect reliability [50], hence it has a significant impact on the performance and reliability of VLSI circuits.

In order to alleviate such problems, changes in the material used for on-chip interconnections have been sought even in earlier technology generations, for example the transition from aluminum to copper some years back. Carbon nanotubes (CNTs) have recently been proposed as a possible replacement for metal interconnects in future technologies [36]. The high resistance associated with an isolated CNT (greater

than $6.45K\Omega$) necessitates the use of a bundle (rope) of CNTs conducting current in parallel to form an interconnection [36].

However, CNT-based interconnect for VLSI systems faces many problems for practical applications. For example, due to the lack of control on chirality, any bundle of CNTs consists of metallic as well as semi-conducting nanotubes (the semi-conducting CNTs do not contribute to current conduction in an interconnect). Based on some literatures, the semi-conducting nanotubes may be as high as 2/3 in a CNT bundle. In addition, perfect contact of CNT bundles is also challengeable. While an imperfect contact for CNT bundles may result in a significant increase in contact resistance ($\sim 300K\Omega$), it is crucial to consider such an imperfect contact for performance analysis of CNT-based interconnects. Although many papers have been proposed to study the timing performance of VLSI circuits with statistical approach (called as SSTA) in modern traditional VLSI systems, no literature focuses on the statistical performance analysis in presence of the above factors for CNT-based nanoscale VLSI systems.

Therefore, as a future work, we will explore the performance of CNT-based clock distribution networks for nanoscale VLSI systems. We also intend to study the statistical performance analysis frame for CNT-based CDN with the considerations of random chirality and imperfect contact, and furthermore develop a probabilistic design method for CNT-based CDNs. Moreover, the fault-tolerant design and reliability for CNT-based CDN will also be investigated in this research.

LIST OF REFERENCES

LIST OF REFERENCES

- [1] <http://www.eas.asu.edu/~ptm/>.
- [2] <http://www.itrs.org>.
- [3] <http://www.semiconductor.net/article/ca267248.html>: Exploiting nanotechnology for terahertz interconnects.
- [4] <http://www.xinitiative.org/>.
- [5] M. Afghahi and C. Svensson. Performance of synchronous and asynchronous schemes for vlsi systems. *IEEE Transactions on Comput.*, 41(7):858–872, July 1992.
- [6] K. Agarwal, M. Agarwal, D. Sylvester, and D. Blaauw. Statistical interconnect metrics for physical-design optimization. *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on*, 25(7):1273–1288, July 2006.
- [7] C. S. Amin, N. Menezes, K. Killpack, F. Dartu, U. Choudhury, N. Hakim, and Y. I. Ismail. Statistical static timing analysis: how simple can we get? In *DAC '05: Proceedings of the 42nd annual conference on Design automation*, pages 652–657, San Diego, California, USA, 2005. ACM Press.
- [8] F. Anceau. A synchronous approach for clocking vlsi systems. *IEEE Journal of Solid-State Circuits*, SC-17:51–56, February 1982.
- [9] N. Arora, L. Song, S. Shah, K. Joshi, K. Thumaty, A. Fujimura, L. Yeh, and P. Yang. Interconnect characterization of x architecture diagonal lines for vlsi design. *Semiconductor Manufacturing, IEEE Transactions on*, 18(2):262–271, May 2005.
- [10] M. Avriel. *Nonlinear Programming: Analysis and Methods*. Dover, 2003.
- [11] H. Bakoglu. *Circuits, Interconnections, and Packaging for VLSI*. Addison Wesley, 1990.
- [12] K. Banerjee and A. Mehrotra. Analysis of on-chip inductance effects for distributed rlc interconnects. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 21(8):904–915, August 2002.
- [13] K. Bernstein, K. M. Carrig, C. Durham, P. R. Hansen, D. Hogenmiller, E. Nowak, and N. Rohrer. *High Speed CMOS Design Styles*. Kluwer Academic Publishers, 1998.
- [14] E. Bogatin. *Signal Integrity: Simplified*. Prentice Hall, 2003.
- [15] K. Bowman, S. Duvall, and J. Meindl. Impact of die-to-die and within-die parameter fluctuations on the maximum clock frequency distribution for gigascale integration. *IEEE Journal of Solid-State Circuits*, 37(2):183–190, February 2002.
- [16] H. Chang and S. S. Sapatnekar. Statistical timing analysis under spatial correlations. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 24(9):1467–1482, Sep. 2005.

- [17] H. Chen, C. Yeh, G. Wilke, S. Reddy, H. Nguyen, W. Walker, and R. Murgai. A sliding window scheme for accurate clock mesh analysis. In *International Conference on Computer Aided Design (ICCAD 2005)*, pages 939–946, 2005.
- [18] S. Das, K. Agarwal, D. Blaauw, and D. Sylvester. Optimal inductance for on-chip rlc interconnections. In *Proceedings. 21st International Conference on Computer Design, 2003.*, pages 264–267, 2003.
- [19] J. Davis and J. D. Meindl, editors. *Interconnect Technology and Design for Gigascale Integration*. Kluwer Academic Publishers, 2003.
- [20] A. Deutsch, P. Coteus, G. Kopcsay, H. Smith, C. Surovic, B. Krauter, D. Edelstein, and P. Restle. On-chip wiring design challenges for gigahertz operation. *Proceedings of the IEEE*, 89(4):529–555, April 2001.
- [21] S. G. Duvall. Statistical circuit modeling and optimization. In *The 5th Intl. Workshop on Statistical Metrology*, pages 56–63, 2000.
- [22] L. Dworsky. *Modern Transmission Line Theory and Application*. Wiley, New York, 1979.
- [23] G. P. E., B. W. J., P. R. P., G. M. K., and A. R. L. High-performance microprocessor design. *IEEE Journal of Solid-state Circuits*, 33:676–686, May 1998.
- [24] D. W. D. et al. A 200-mhz 64-b dual-issue cmos microprocessor. *IEEE Journal of Solid-State Circuits*, 27:1555–1565, November 1992.
- [25] J. P. Fishburn. Clock skew optimization. *IEEE Transactions on Computer*, 39:945–951, July 1990.
- [26] E. Friedman, editor. *High Performance Clock Distribution Networks*. Kluwer Academic Publishers, 1997.
- [27] E. G. Friedman. Clock distribution networks in synchronous digital integrated circuits. *Proc. of IEEE*, 89(5):665–692, May 2001.
- [28] R. Gupta, B. Krauter, B. Tutuianu, J. Willis, and L. T. Pillage. The elmore delay asa bound for rc trees with generalized input signals. In *ACM/IEEE Design Automation Conf.*, pages pp.364–369, 1995.
- [29] S. H. Hall, G. W. Hall, and J. A. McCall. *High-Speed Digital System Design: A Handbook of Interconnect Theory and Design Practices*. John Wiley & Sons, Inc., 2000.
- [30] Y. Ismail and E. Friedman. Effects of inductance on the propagation delay and repeater insertion in vlsi circuits. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 8(2):195–206, April 2000.
- [31] Y. Ismail, E. Friedman, and J. Neve. Exploiting the on-chip inductance in high-speed clock distribution networks. *IEEE Transactions on VLSI systems*, 5(12):963–973, December 2001.
- [32] I. Jiang. Optimal reliable crosstalk-driven interconnect optimization. In *Proceeding of ISPD*, pages 128–140, 2000.
- [33] X. Jiang and S. Horiguchi. Optimization of wafer scale h-tree clock distribution network based on a new statistical skew model. In *DFT '00: Proceedings of the 15th IEEE International Symposium on Defect and Fault-Tolerance in VLSI Systems*, pages 96–104, Washington, DC, USA, 2000. IEEE Computer Society.
- [34] X. Jiang and S. Horiguchi. Statistical skew modeling for general clock distribution networks in presence of process variations. *IEEE Trans. Very Large Scale Integr. Syst.*, 9(5):704–717, 2001.

- [35] C.-K. Koh and P. H. Madden. Manhattan or non-manhattan?: a study of alternative vlsi routing architectures. In *GLSVLSI '00: Proceedings of the 10th Great Lakes symposium on VLSI*, pages 47–52, New York, NY, USA, 2000. ACM Press.
- [36] F. Kreupl, A. P. Graham, G. S. Duesberg, W. Steinhogel, M. Liebau, E. Unger, and W. Honlein. Carbon nanotubes in interconnect applications. *Microelectronic Engineering*, 64(1-4):399–408, October 2002.
- [37] S. Kugelmass and K. Steiglitz. A probabilistic model for clock skew. In *Proceeding of Internal Systolic Arrays, Sandiego*, 1998.
- [38] K.-Y. Lee, T.-C. Wang, and K.-Y. Chao. Post-routing redundant via insertion and line end extension with via density consideration. In *Computer-Aided Design, 2006. ICCAD '06. IEEE/ACM International Conference on*, pages 633–640, Double Tree Hotel, San Jose, CA, USA, 2006.
- [39] I. Lin, J. A. Ludwig, and K. Eng. Analyzing cycle stealing on synchronous circuits with level-sensitive latches. In *ACM/IEEE Design Automation Conf.*, pages 393–98, June 1992.
- [40] Y. Liu, S. R. Nassif, L. T. Pileggi, and A. J. Strojwas. Impact of interconnect variations on the clock skew of a gigahertz microprocessor. In *DAC '00: Proceedings of the 37th conference on Design automation*, pages 168–171, New York, NY, USA, 2000. ACM Press.
- [41] J. Lou and W. Chen. Crosstalk-aware placement. *IEEE Design & Test of Computers*, 4:28–32, Feb 2004.
- [42] M. Nekili, G. Bois, and Y. Savaria. Pipelined h-tree for high-speed clocking of large integrated systems in presence of process variations. *IEEE Transactions on VLSI Systems*, 5(2):161–174, February 1997.
- [43] V. G. Oklobdzija, V. M. Stojanovic, D. M. Markovic, and N. M. Nedovic. *Digital System Clocking: High-Performance and Low-Power Aspects*. Wiley-IEEE Press, 2003.
- [44] V. F. Pavlidis and E. G. Friedman. Via placement for minimum interconnect delay in three-dimensional (3-d) circuits. In *Proceedings of the International Symposium on Circuits and Systems, 2006*, pages 4587–4590, 2006.
- [45] X. Qi, B. Kleveland, Z. Yu, S. Wong, R. Dutton, and T. Young. On-chip inductance modeling of vlsi interconnects. In *Proceeding of IEEE International Solid-State Circuits Conference*, 2000.
- [46] X. Qi, G. Wang, Z. Yu, R. Dutton, T. Young, and N. Chang. On-chip inductance modeling and rlc extraction of vlsi interconnects for circuit simulation. In *Proceedings of the IEEE Custom Integrated Circuits Conference, 2000.*, pages 487 – 490, 2000.
- [47] T. Sakurai. Closed-form expressions for interconnection delay, coupling, and crosstalk in vlsi. *IEEE Transactions on Electron Devices*, 40(1):118 – 124, Jan 1993.
- [48] S.-P. Sim, S. Krishnan, D. Petranovic, N. Arora, and C. Kwyro Lee Yang. A unified rlc model for high-speed on-chip interconnects. *IEEE Transactions on Electron Devices*, 50(6):1501– 1510, June 2003.
- [49] A. Srivastava, D. Sylvester, and D. Blaauw. *Statistical Analysis and Optimization for VLSI: Timing and Power*. Springer, 2005.

- [50] N. Srivastava and K. Banerjee. A comparative scaling analysis of metallic and carbon nanotube interconnections for nanometer scale vlsi technologies. In *Proc. VMIC*, pages 393–398, Sep 2004.
- [51] W. Steinhögl, G. Schindler, G. Steinlesberger, and M. Engelhardt. Size-dependent resistivity of metallic wires in the mesoscopic range. *Phys. Rev. B*, 66(7):075414, Aug 2002.
- [52] K. Takase, T. Ohkubo, F. Sawada, D. Nagayama, J. Kitagawa, and Y. Kadoya. Propagation characteristics of terahertz electrical signals on micro-strip lines made of optically transparent conductors. *Jpn. J. Appl. Phys*, 44(32):L1011–L1014, 2005.
- [53] S. L. Teig. The x architecture: not your father’s diagonal wiring. In *SLIP '02: Proceedings of the 2002 international workshop on System-level interconnect prediction*, pages 33–37, New York, NY, USA, 2002. ACM Press.
- [54] S. C. Thierauf. *High-Speed Circuit Board Signal Integrity*. Artech House, 2006.
- [55] A. W. Topol, J. D. C. La Tulipe, L. Shi, D. J. Frank, K. Bernstein, S. E. Steen, A. Kumar, G. U. Singco, A. M. Young, K. W. Guarini, and M. Jeong. Three-dimensional integrated circuits. *IBM J. Res. Dev.*, 50(4/5):491–506, 2006.
- [56] N. H. Weste and D. Harris. *CMOS VLSI Design, 3rd Edition*. Addison Wesley, 2005.
- [57] J. Woods, P. Day, S. Furber, J. Garside, N. Paver, and S. Temple. Amulet1: An asynchronous arm microprocessor. *IEEE Transactions on Computers*, 46(4):385–398, April 1997.
- [58] R. Zhang, K. Roy, C. Koh, and D. Janes. Stochastic interconnect modeling, power trends, and performance characterization of 3-dimensional circuits. *IEEE Transactions on Electron Devices*, 48(4):638–652, April 2001.

APPENDIX

A. PUBLICATIONS LIST

1. Xu Zhang, Xiaohong Jiang and Susumu Horiguchi, *Performance of H-tree Clock Distribution Networks in Presence of Process Variations and Inductance Effects*, SOIM-COE05, Sendai, Japan.
2. Xu Zhang, Xiaohong Jiang and Susumu Horiguchi, *Performance Analysis of Variant X-tree Clock Distribution Networks*, SOIM-COE06, Sendai, Japan.
3. Xu Zhang, Xiaohong Jiang and Susumu Horiguchi, *Performance of H-tree Clock Distribution Networks in Presence of Process Variations and Inductance Effects*, SASIMI (Synthesis And System Integration of Mixed Information technologies) 2006, pp.166–170, Nagoya, Japan.
4. Xu Zhang, Xiaohong Jiang and Susumu Horiguchi, *A Nonorthogonal Clock Distribution Network and Its Performance Evaluation in Presence of Process Variations and Inductive Effects*, GLSVLSI (ACM Great Lakes Symposium on VLSI) 2006, pp.336–340, Philadelphia, USA.
5. Xu Zhang, Xiaohong Jiang and Susumu Horiguchi, *Variant X-Tree Clock Distribution Network and Its Performance Evaluations*, IEICE Transactions on Electronics, Vol. E90-C No.10, pp.1909–1918, 2007.
6. Xu Zhang, Xiaohong Jiang and Susumu Horiguchi, *Redundant Vias Insertion for Performance Enhancement in 3D ICs*, IEICE Transactions on Electronics, Vol. E91-C No.4 (In press)
7. Xu Zhang, Xiaohong Jiang and Susumu Horiguchi, *Performance of H-Tree Clock Distribution Networks in Presence of Process Variations and Inductance Effects*, submitted to Journal of System and Architecture.