UC San Diego
UC San Diego Electronic Theses and Dissertations

Title
Optimization of the BEOL Interconnect Stack for Advanced Semiconductor Technology Nodes

Permalink
https://escholarship.org/uc/item/9sf895pj

Author
Shah, Pooja Pradeep

Publication Date
2015

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Optimization of the BEOL Interconnect Stack for Advanced Semiconductor Technology Nodes

A thesis submitted in partial satisfaction of the
requirements for the degree
Master of Science

in

Electrical Engineering (Computer Engineering)

by

Pooja Pradeep Shah

Committee in charge:

Professor Andrew B. Kahng, Chair
Professor Chung-Kuan Cheng
Professor Bill Lin

2015

Copyright
Pooja Pradeep Shah, 2015
All rights reserved.

The thesis of Pooja Pradeep Shah is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2015

DEDICATION

I dedicate this thesis to my loving husband Rakesh, my mother and my sister. This thesis would not have been possible without their encouragement and support.

TABLE OF CONTENTS

Signature Page iii

Dedication iv

Table of Contents v

List of Figures vii

List of Tables ix

Acknowledgments x

Vita xi

Abstract of the Thesis xii

Chapter 1 Introduction 1
- 1.1 Motivation 2
- 1.2 Our Contributions 3
- 1.3 Limitations 5
- 1.4 Outline of the Thesis 5

Chapter 2 Previous Works 7
- 2.1 Landmark Works 8
- 2.2 Wire Optimization at System Level 11
- 2.3 Wire Optimization Considering Layer Assignment 13
- 2.4 Wire Sizing and Repeater Insertion 14
- 2.5 Other Works 16
- 2.6 Predictive Technology Models 18
- 2.7 Reliability 19
- 2.8 The ITRS BEOL Structure and Prediction 20
- 2.9 Pitch Prediction 22
- 2.10 Summary of Previous Works 23

Chapter 3 Experiments for Interconnect Dimension Optimization and Validation 26
- 3.1 Our Flow 26
  - 3.1.1 Optimization of Interconnect Dimensions 27
  - 3.1.2 Validation 28
- 3.2 Our Interconnect Structure 29
- 3.3 Interconnect Optimization: Single-Stage Analysis 30
  - 3.3.1 Methodology 30

3.3.2 Results 35
3.4 Single-Stage Validation: Predictive Technology Models . . . 38
3.4.1 Methodology 39
3.4.2 Results 41
3.5 Multistage Validation Enablement: Library Scaling 42
3.5.1 Methodology: Device Scaling 43
3.5.2 Results 46
3.5.3 Methodology: Interconnect Scaling 47
3.5.4 Results 48
3.6 Multistage Validation: Design-Level Performance Analysis . 48
3.6.1 Methodology 48
3.6.2 Results 49
3.7 Multistage Validation: Random Path Model 51
3.7.1 Methodology 51
3.7.2 Results 58

Chapter 4 Conclusion and Future Work 59
4.1 Conclusion 59
4.2 Future Work 61

Bibliography 62

LIST OF FIGURES

Figure 1.1: Interconnect stack comparison of Intel $22nm$ and $14nm$ processes [54]. 2

Figure 2.1: Elmore delay model for an RC network. 8
Figure 2.2: Pamanuwa delay model for an RC network with coupling capacitance [4]. 10
Figure 2.3: FinFET parasitic capacitances (adapted from [68]). 19
Figure 2.4: Misaligned via reduces interconnect spacing and enhances electric field [43]. 20
Figure 2.5: The ITRS model [56]. 21
Figure 2.6: ITRS prediction versus real chip data [56]. 21
Figure 2.7: Metal-1 layer pitch comparison [50]. 22
Figure 2.8: Comparison of interconnect width predictions made by previous works and real chips in the same technology [6], [9], [26], [19], [17], [29]. 23
Figure 2.9: Comparison of interconnect spacing predictions made by previous works and real chips in the same technology [6], [9], [26], [19], [17], [29]. 23
Figure 2.10: Comparison of interconnect AR predictions made by previous works and real chips in the same technology [6], [9], [26], [19], [17], [29], [56]. 24

Figure 3.1: Flow of our experiment. 26
Figure 3.2: Overview of the single-stage flow. 27
Figure 3.3: System-level validation flow. 28
Figure 3.4: Cross-section of our model. 29
Figure 3.5: Interconnect optimization using single-stage analysis. 30
Figure 3.6: Circuits for simulation in HSPICE [62]. 31
Figure 3.7: Slew-bounded delay (unit: sec) results. 32
Figure 3.8: Slew-bounded energy-delay product (unit: $J \cdot sec$) results. 33
Figure 3.9: Slew-bounded current density (unit: MA/cm^2) results. 33
Figure 3.10: Slew-bounded energy-delay product (unit: $J \cdot sec$) results with process variation: (ΔT = $\pm 10\%$, ΔCD = $\pm 10\%$). 35
Figure 3.11: Circuits for simulation in HSPICE [62]. 40
Figure 3.12: Performance variation across gate sizes and fanouts. 41
Figure 3.13: Scaling of front-end and back-end of the $28nm$ Bulk library to IDM $14nm$. 43
Figure 3.14: Intel $14nm$ fin information [60]. 44
Figure 3.15: Design-level validation flow. 49
Figure 3.16: Random path flow used for performance comparison. 51
Figure 3.17: Random extracted paths. 53

Figure 3.18: Random generated paths. 53
Figure 3.19: Generated gate-level netlists. 56
Figure 3.20: Generated SPEF files. 57

Figure 4.1: Comparison of AR and DC between the ITRS predicted interconnect dimensions and our obtained interconnect dimensions. 61

LIST OF TABLES

Table 2.1: Parameters of PTM-MG FinFET model. 18
Table 2.2: Summary of previous works and categorization of our work. 25

Table 3.1: Parameters in our optimization studies. 30
Table 3.2: Results comparison of EDP, $Delay$, EDP with variation and J_{RMS} between the ITRS and our determined dimensions. 36
Table 3.3: Library scaling factors for Bulk 28nm to IDM 14nm. 47
Table 3.4: RC scaling factors for local layers using Raphael results. 48
Table 3.5: Power and timing analysis results for design-level validation. 50
Table 3.6: Power and timing analysis results (normal and critical paths) for the random path model. 58

Table 4.1: Suggested optimal dimension range. 59

ACKNOWLEDGMENTS

I would like to thank my parents (my father Pradeep Shah and my mother Vibha Shah), my sister Riddhi Shah and my husband Rakesh Mallem for their endless love and patience. Their support has been of utmost importance prior to and during my graduate program.

I would like to express my sincere gratitude to my advisor Professor Andrew B. Kahng for his invaluable advice and continuous support. His drive, attitude and principles on research have provided a precious lesson before and throughout my M.S. program. His immense knowledge and guidance helped me in conducting research.

Besides my advisor, I would like to thank my thesis committee members, Professor Chung-Kuan Cheng and Professor Bill Lin, for their time to review my research and for their valuable comments.

I would like to thank Hyein Lee for her immense help and support in conducting research, reviewing this thesis and providing insightful comments. I would especially like to thank Yaping Sun for being my partner in the initial stages of this work and working tirelessly towards co-authoring an unpublished paper.

Last, but not least, I would like to thank my other lab-mates in the UCSD VLSI CAD Laboratory (Jiajia Li, Kwangsoo Han, Wei-Ting Chan, Siddhartha Nath, Mulong Luo) and former lab member (Dr. Tuck-Boon Chan) for their enlightening discussions and insights.

VITA

1987	Born, Bangalore, India
2009	B.E., Electrical Engineering, M. V. Jayaraman College of Engineering (Affiliated to Visvesvaraya Technological University), Bangalore, India
2012	Corporate Applications Engineer, Synopsys Inc., Hyderabad, India
2015	M.S., Electrical Engineering (Computer Engineering), University of California, San Diego

ABSTRACT OF THE THESIS

Optimization of the BEOL Interconnect Stack for Advanced Semiconductor Technology Nodes

by

Pooja Pradeep Shah

Master of Science in Electrical Engineering (Computer Engineering)

University of California, San Diego, 2015

Professor Andrew B. Kahng, Chair

Particularly in advanced technology nodes, interconnects significantly affect the power, performance, area and reliability of integrated circuits. Requirements of high integration density, performance, complex patterning technology and cost implications make it imperative to determine optimal *back-end-of-line* (BEOL) stack dimensions for sub-22nm technology nodes. This thesis studies copper interconnect scaling strategies for high-performance IC designs. It focuses on determining optimal dimensions of the BEOL interconnect stack to achieve least possible delay while maintaining low power consumption. Degrees of freedom of the optimization technique in this work are *wire as-*

pect ratio (AR) and *wire line-space duty cycle* (DC).[1] Our study targets the interconnect scaling strategy for 28nm through 7nm *integrated device manufacturer* (IDM) nodes, which have a more rigorous scaling trend than what is seen in the pure-play foundry "node" taxonomy. For example, the $14nm$ IDM node has a minimum Metal-1 pitch of 52nm [31], while the corresponding pitch for the pure-play foundry $14nm$ technology node is 64nm [39]. The studies in this thesis also indicate the advantages of using a low wire aspect ratio and a high line-space duty cycle for sub-22nm technology nodes.

[1] We define AR by T/W = metal thickness / metal width, and DC using $L/(L+S)$ = line-width / pitch.

Chapter 1

Introduction

In recent years, BEOL interconnect technology strategy for advanced VLSI processes has been challenging to determine. Interconnect stack properties such as wire geometry, proximity to other interconnects, and material selection affect design optimization solutions such as repeater insertion, power delivery, and design margin for performance and reliability. Thus, utilization of optimal interconnects is essential. Prediction of interconnect scaling strategy years in advance of a generation helps provide a path to process engineers, circuit designers, and CAD developers.

With shrinking technologies, the interconnect stack starts to play a major role in circuit performance. The interconnect structure affects circuit performance by impacting power, delay and signal integrity. Interconnect dimension and performance scaling is now a complex function of various process options and their respective costs and benefits. Issues to consider while optimizing BEOL stack dimensions include the following.

1. Wide range of parameters.

   Interconnect dimension parameters include metal pitch, width, spacing, thickness and dielectric height of each layer. For a given technology on a given layer, the metal pitch, wire thickness and dielectric height are usually constant due to manufacturing constraints.

2. Manufacturing process.

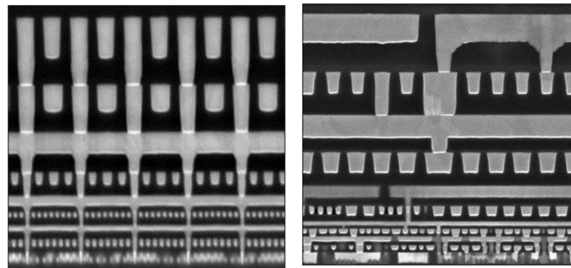
   Lithography technique limitations lead to strict rules for interconnect pitch in advanced nodes. *Chemical-mechanical polishing* (CMP) [45] can cause defects due

to metal dishing and erosion. The variations in *critical dimension* (CD) and metal thickness introduced during the manufacturing process result in variability of interconnect electrical characteristics.

3. Reliability concerns.

   Shrinking interconnect dimensions make reliability issues more challenging. For example, the current density of wires increases due to a smaller cross-sectional area, worsening the *electromigration* (EM) lifetime. The impact on chip lifetime because of *time dependent dielectric breakdown* (TDDB) also worsens with increased interconnect density due to high electric fields between two wires [43], which are exacerbated by overlay error in multi-patterning nodes. These factors have raised chip lifetime concerns due to reliability issues.

1.1 Motivation



(a) Intel 22nm process. (b) Intel 14nm process.

Figure 1.1: Interconnect stack comparison of Intel 22nm and 14nm processes [54].

Three parameters which affect interconnect electrical characteristics are metal pitch, AR and DC. The minimum metal pitch and thickness of each metal layer are defined by the foundry, while designers can vary wire width and spacing to achieve the desired performance. In recent technology nodes, the industry has used a combination of AR $= 2$ and constant DC $= 0.5$. An example is Intel's 22nm process, shown in Figure 1.1(a) [54]. However, in 2014, Intel announced their 14nm process with an unprecedented set of wire dimensions. Figure 1.1(b) [54] shows a lateral cross-section of a

chip built with this technology. Figure 1.1(b) indicates that this technology uses a minimum Metal-1 pitch of $52nm$ with interconnects that are flatter and closer as compared to those in Figure 1.1(a). That is, we see a combination of lower AR and higher DC in the $14nm$ node as compared to the $22nm$ node.

The sudden change in AR and DC choices motivates us to explore the rationale behind revised BEOL stack scaling and to optimize interconnect structures for future technology nodes. This thesis is a step towards the exploration of reasons for this drastic shift. In this thesis, we create an interconnect structure similar to the one in Figure 1.1(b) and analyze impacts of AR and DC choices on the performance of critical-path models and benchmark designs.

1.2 Our Contributions

In this thesis, we start by analyzing the impact of interconnect dimensions on the performance of inverter circuits for various technology nodes, according to criteria such as slew-bounded delay and slew-bounded *energy-delay product* (EDP). We measure the circuit performance for various combinations of interconnect dimensions. By applying this technique to designs to explicitly improve the performance, we show that the combination of low AR and high DC is a likely direction for the interconnect technology roadmap, in contradiction to the predictions in the current ITRS Interconnect Chapter [56]. While we focus on determination of optimal interconnect dimensions for sub-$22nm$ technology nodes, we also analyze the performance of interconnects for the recent $45nm$ and $28nm$ technology nodes to retrospectively confirm previous industry choices for BEOL dimensional scaling.

To validate our findings, we first apply the determined optimal interconnect dimensions to predictive technology libraries (sub-$20nm$ technologies) [46] in order to measure the impact of scaled devices and voltage on the performance of simple circuits. These predictive libraries consider the 3D structure of FinFETs in their models. Lacking standard libraries for advanced technology nodes, we also scale an existing foundry library to perform multistage and multi-layer power and timing analysis of benchmark designs. For this, we scale a foundry library after consultations with experts in industry

[32] [33] and based on projections given in the *process integration, devices, and structures* (PIDS) chapter of the ITRS [57]. Our scaled libraries include multiple configurations for the ITRS dimensions and our dimensions, each with two and three fins (i.e., in a FinFET device). We further use our scaled libraries to evaluate the performance of random paths in benchmark designs.

The contributions of this thesis are as follows.

1. We study the impact of different interconnect configurations on circuit performance for 45nm, 28nm, 22nm, 14nm, 10nm and 7nm technology nodes. Our initial optimization studies consider the effects of advanced process technologies, such as new barrier and dielectric materials by using the predicted values from the ITRS [56]. This helps us maintain the industrial standard of predictions and assumptions for manufacturing technologies.

2. We consider the impact of process variations when determining optimal interconnect dimensions.

3. To ensure the robustness of our determined optimal interconnect dimensions, we perform design-level validation.

4. Based on our results, we suggest that the combination of lower AR (< 2) and higher DC (> 0.5) can achieve a better overall performance for advanced nodes.

A key conclusion we draw from our experiments is that performance-optimized interconnect dimensions will not only lower AR and increase DC, but have the welcome side effect of extending copper-based BEOL to the IDM $5nm$ node. Specifically, a traditional DC $= 0.5$ implies that for the ITRS $21nm$ local-metal layer pitch, wire width will be 10.5nm. This unfortunately runs into two critical showstoppers: the inability to pattern inlaid (damascene) copper below 12-14nm trench CD [39] [35], and the skyrocketing increase in *resistive-capacitive* (RC) delays with shrinking wire dimensions [42]. Our optimal AR and DC values imply feasible local metal CDs into the IDM $5nm$ technology node.

1.3 Limitations

For initial interconnect configuration generation, this thesis considers the ITRS predictions for dielectric permittivity. This work does not consider the existence and impact of air gaps, which are reappearing in the latest technology nodes.

Our initial optimization studies set a slew time upper bound of $50ps$ for $28nm$ implementations. While this constraint is somewhat arbitrary, it is reasonable given that we constrain wirelengths for zero slew degradation for advanced technology nodes.

The circuit simulations we perform to determine optimal interconnect dimensions utilize devices from the *Synopsys 32/28nm Generic Library* [52]. We further validate our interconnect dimensions using predictive technology models and scaled libraries. However, our initial optimization studies do not consider device and voltage scaling.

The thesis does not directly address reliability issues such as EM and TDDB. While we understand that a low RMS current density represents reduced susceptibility to EM, we do not have measurable design-level EM results to support the claim.

This thesis determines optimal dimensions only for local metal layers. While scaling libraries for the design-level/multi-layer experiments, we only apply global scaling factors for interconnect parasitics. We also do not consider the impact of via size, shape or parasitics when determining optimal interconnect dimensions. While we scale the parasitics of a library to estimate design-level performance, we do not scale the dimensions of the devices and interconnects in the library, thus ignoring routing issues associated with advanced technology nodes.

1.4 Outline of the Thesis

The rest of the thesis is organized as follows.

In Chapter 2, we review techniques used by previous works to optimize interconnect performances and reliability over the existing technology nodes. We also review the methodology used by the ITRS *Interconnect International Technology Working Group* (ITWG) [56] to predict interconnect dimensions for future technology nodes, and analyze the ITRS predicted interconnect scaling trend.

In Chapter 3, we describe the experimental flow used to obtain optimal interconnect dimension structure. The methodologies used span interconnect structure creation to multiple validations of the calculated dimensional scaling.

Finally, in Chapter 4, we discuss the relevance of achieved results to the path of technology scaling, along with directions for future work.

Chapter 2

Previous Works

This chapter reviews previous works which use different interconnect optimization techniques. On the basis of these techniques, we can broadly classify the works into five categories.

1. Landmark works
2. Wire optimization at system level
3. Wire optimization considering layer assignment
4. Wire sizing and repeater insertion
5. Other works

Works in the system-level optimization category optimize interconnect dimensions on the basis of multistage and multi-level (multiple levels within a tier of local, semi-global or global) designs. Their optimization constraint includes the performance of the entire design. Although our optimization technique is based solely upon a single-stage circuit, we include system-level performance analysis to validate the robustness of our obtained optimal dimensions.

It is essential to consider the assignment of nets to each of the layers because each interconnect layer within the stack has different properties such as size, pitch, resistivity, etc. Our work, however, does not aim to optimize interconnect dimensions of

all the layers. We only determine optimal dimensions for local interconnect layers and measure their impact on multi-level designs.

While the structure of a wire determines its performance, the driving capability and parasitics of the devices are also instrumental in determining the overall circuit performance. In our work, we measure the robustness of our obtained optimal dimensions using both wire and gate sizing. Repeater insertion is a technique used by designers to meet timing constraints. Our work does not directly involve detailed repeater size and placement optimization techniques.

2.1 Landmark Works

These works define models and methods which are used by several other works for the optimization of circuit performance.

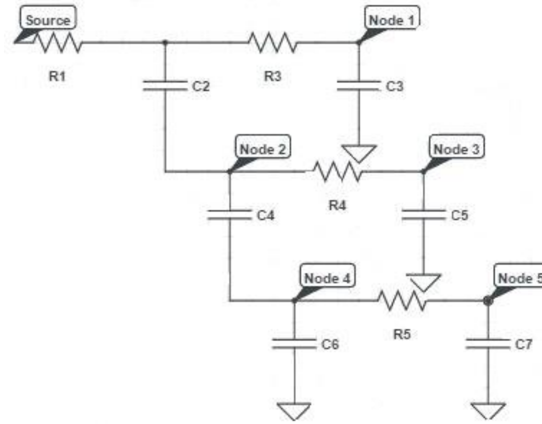


Figure 2.1: Elmore delay model for an RC network.

The Elmore delay model [1] is an integral part of delay modeling for *resistive-capacitive* (RC) networks. In this work, the author models the transient behavior of a linear RC network using the first moment of impulse response of the system. For an RC network with N nodes as shown in Figure 2.1, the delay is given by Equation 2.1.

$$\tau = \sum_{k=1}^{N} C_k R_{ik} \tag{2.1}$$

where k is an index for each capacitor in the circuit and R_{ik} is the common resistance from the input to nodes i and k [25]. The Rent's rule power law model by Landman et al. [2] demonstrates optimal partitioning in a circuit. The model gives an empirical relationship between the average number of pins per net, the number of external pins, and the number of gates in a given block. The Rent's rule is defined by Equation 2.2.

$$T = k \times G^p \tag{2.2}$$

where T denotes the average number of pins per net, k is the number of external pins, G is the average number of gates in the given block, and p is Rent's parameter. This model is helpful in the optimization of placement, layout parameter estimations and wirelength estimation [3]. The power-law model of Rent's rule aids in prediction of the number of terminals required by a group of devices/gates to cross a predefined boundary and communicate with the rest of the circuit.

Minute variations in the Rent's parameter can lead to varied results and interpretations of the rule. There are several works which predict and approximate the Rent's exponent depending on the application used to generate it. The only way to accurately know the Rent's exponent for a particular design is to generate it. However, owing to the associated computational difficulties, the Rent's exponent is often extracted on the basis of sample netlists [3]. The sample netlists are generated on the basis of a partitioning algorithm, which leads to the conclusion that the accuracy of the Rent's exponent derived is based on the quality of the partitioning algorithm. Lower values of the Rent's parameter indicate a better placement optimization scheme [3].

In shrinking technologies, propagation delay of nets is a growing concern. In this regard, another important work is by Bakoglu et al. [12]. The authors propose a model for interconnect delay calculation and optimization using a repeater sizing and placement technique. The authors consider the spacing between the source, sinks and other repeaters. Their expression for interconnect propagation delay with repeater insertion is given in Equation 2.3.

$$T = k \times (2.3\frac{R_0}{h}(\frac{C_{int}}{k} + hC_0) + \frac{R_{int}}{k}(\frac{C_{int}}{k} + 2.3hC_0)) \tag{2.3}$$

where C_0 and R_0 are the input capacitance and output resistance of the minimum size inverter, and C_{int} and R_{int} are the total interconnect resistance and capacitance.

The optimal values for k (number of repeaters) and h (repeater distance) are obtained by the Equations 2.4 and 2.5 respectively.

$$k = \sqrt{\frac{R_{int}C_{int}}{2.3R_0C_0}} \tag{2.4}$$

$$h = \sqrt{\frac{R_0C_{int}}{R_{int}C_0}} \tag{2.5}$$

Ismail et al. [25] present an interconnect delay model with consideration of resistance, capacitance and inductance. The authors obtain a *resistive-inductive-capacitive* (RLC) tree and draw a correlation of accuracies with respect to the RC tree from Elmore delay. The derived model is analytical and can be used in practice instead of the RC tree used for Elmore delay.

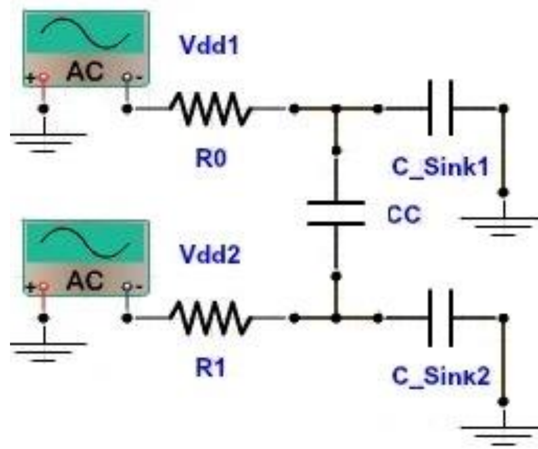


Figure 2.2: Pamanuwa delay model for an RC network with coupling capacitance [4].

Pamanuwa et al. [4] demonstrate a delay model for interconnects with consideration of crosstalk between wires. The authors extend work from [12] to include the impact of coupling capacitance in the delay modeling when repeaters are inserted. The modified delay model for a lumped RC network with coupling capacitance represented

in the Figure 2.2 is given by Equation 2.6.

$$\tau = k[0.7\frac{R_0}{h}(\frac{C_s}{k} + hC_0 + 2.2\frac{2C_c}{k}) + \frac{R}{k}(0.4\frac{C_s}{k} + 0.58\frac{C_c}{k} + 0.7hC_0)] \tag{2.6}$$

where C_0 and R_0 are the output capacitance and resistance of the minimum size inverter, C_c is the coupling capacitance and the constant 2.2 represents the *Miller coupling factor* (MCF). The Pamanuwa delay model leads to the optimal values for k (number of repeaters) and h (repeater distance) obtained by Equations 2.7 and 2.8 respectively.

$$k = \sqrt{\frac{0.4RC_s + 0.58RC_c}{0.7R_0C_0}} \tag{2.7}$$

$$h = \sqrt{\frac{0.7R_0C_s + 3.1R_0C_c}{0.7RC_0}} \tag{2.8}$$

When C_c is set to 0 in Equations 2.7 and 2.8, they simplify to Bakoglu's delay model [12].

2.2 Wire Optimization at System Level

Many researchers apply wire optimization techniques to an entire system/ design/ IC in order to get robust results. In this approach, authors optimize interconnect dimensions on the basis of the performance of the entire design.

Anand et al. [5] [6] present a tool which generalizes the interconnect optimization problem. Using this tool, the authors suggest techniques for interconnect dimension optimization. The demonstrated model aims at achieving high speed with low chip area and power utilization. The authors achieve this goal by reducing the aspect ratio of local wires, effectively minimizing capacitance. In addition, they also suggest using thick, low-κ dielectrics. This optimization technique first minimizes the number of metal layers and then maximizes the achievable speed of the design. The authors use a system-level critical-path model with an assumption that on an average, all nets are the same and critical. The most critical net in their model is a corner-to-corner path consisting of multiple stages. They consider Rent's rule [2] to calculate the average wirelength. The authors suggest utilization of low wire aspect ratio and duty cycle to minimize coupling

capacitance and meet integration density requirements. The authors also suggest the use of a thick interlayer dielectric to minimize the coupling capacitance.

Takahashi et al. [26] optimize the block cycle time on local layers and global interconnect delay at the same time. This work uses the Davis model [10] and Rent's rule [2] to estimate the average wirelength. The authors demonstrate the use of low wire aspect ratio to improve speed in local interconnects. This work also suggests the use of high DC (approximately 0.7) for a high-performance global interconnect.

Zhang et al. [28] present an interconnect model which allows the designer to make a tradeoff between optimum performance and tolerance to manufacturing variance. The authors present an algorithm which makes a tradeoff between interconnect capacitance and RC delay variation in a system. The authors implement a technique called *normal boundary intersection* (NBI). This algorithm finds evenly spaced points on a curve, each of which are optimal solutions.

The work by Mui et al. [19] essentially compares the interconnects with DC > 0.5 with those having DC $= 0.5$. In this work, the authors measure performance on the basis of design frequency. This work suggests that the use of high DC reduces the optimal delay per unit length, the total buffer area and power consumption.

In Cobb et al. [9], we see the emergence of new metrics for measurement and validation of interconnect wire dimensions. The authors consider maximum bandwidth and via resistance, which can be supported by different configurations of the wire dimensions with or without a power constraint. In this work, the authors consider average wirelength of nets on different layers to optimize interconnect dimensions.

Zhu et al. [29] optimize interconnect wire width by constraining scattering impacted wire resistance. The authors optimize interconnects on the basis of an analytical model of optimal DC. The performance metric used in this work is a function of interconnect length, coupling and fringe capacitances and delay. The authors also validate the importance of scattering in their experimental results.

2.3 Wire Optimization Considering Layer Assignment

Interlayer dependencies play a crucial role in achieving good chip level performance. Layer assignment affects both wire resistance and capacitance. Kahng et al. [15] explore wire sizing and repeater insertion constrained by delay to optimize layer assignment during routing. This work considers the impact of vias and the area of repeaters on design performance. The authors suggest a greedy optimization algorithm. Along with an optimum wirelength estimation model, this algorithm helps in achieving low cost placement and routing solutions. The authors differentiate between uniform and non-uniform layer stacks with respect to performance. They observe that non-uniformity allows shorter nets to move to lower metal layers to benefit from low wire widths. The authors infer that the maximum wire width of a layer is independent of delay and constrained by layer stack parameters.

Dasgupta et al. [13] use a metric of $rank$ of interconnect architecture. This work uses a dynamic programming-based algorithm to evaluate interconnect architectures. The performance metric is constrained by conditions such as:

1. Assignment of long wires to higher layers
2. Priority of repeater insertion to long wires until the target delay is met
3. Use of equal sized repeaters

Given a wirelength distribution, this work determines the quality of optimal assignment of wires subject to constraints on the basis of the first wire assignment which does not meet the timing requirement. This work implements incremental wire assignment from bottom up for the stack. The knobs to vary clock frequency in this work are MCF and maximum repeater area. The authors use the Rent's parameter [2] for an accurate wirelength distribution model. A drawback of this model is the assumption that *wirelength distribution* (WLD) is linear in wirelength when wire delay is actually quadratic with respect to length.

Li et al. [17] demonstrate the advantages of implementing buffer sizing and layer assignment concurrently to maximize speed and efficiency. This work compares

the feasibilities of tapered and uniform wire sizing along with simultaneous buffer insertion. The authors implement two techniques for simultaneous buffer insertion and layer assignment. The first technique is used for critical path optimization by delay improvement. Here, the main idea is to improve the slack by moving subnets to higher metal layers. The second technique fixes capacitance and slew violations to perform slew recovery. The idea behind this algorithm is that firstly, slew is a local constraint and secondly, that not all nets have the same criticality. Here, there is a possibility of blockages in placement due to the existence of macros and intellectual property blocks (IPs) which can lead to more slew violations than expected. In the absence of layer assignment, these slew violations can be fixed by buffer insertions. The authors state that layer assignment has better performance over wire sizing because it reduces the interconnect delay and the number of buffers needed. They also minimize the usage of thick metal layers. By assigning each subnet to the same layer, the authors achieve both high performance and low via cost.

More recent work by Ao et al. [7] optimizes delay using layer assignment for *three-dimensional* (3D) routing. The authors assign nets to layers to achieve minimum delay. They then use a greedy algorithm to reduce wire congestion. In this technique, the authors reduce the via count to lower delay. This algorithm is applicable to designs where interconnect layers have drastically different electrical characteristics.

2.4 Wire Sizing and Repeater Insertion

Wire sizing and repeater insertion techniques allow designers to make changes as needed by the design. This section reviews previous works based on these two techniques.

Menezes et al. [18] perform simultaneous gate and interconnect sizing to optimize path delay and circuit area. The authors use a Π model for interconnect delay simulation and demonstrate the impact of both wire and gate sizing in the performance of a design. Their gate sizing technique assumes that the ratio of transistor widths remains constant. The authors optimize performance using delay-based sensitivity techniques for both devices and wires. In addition, the authors observe that a reduction in delay

aids in minimizing the area of the circuit. They observe that the technique of simultaneous gate and wire sizing is advantageous as compared to the technique involving only gate sizing.

Davis et al. [10] propose a model to scale interconnects in a three-tier (local, semi-global and global), multi-level (multiple layers within a tier) network. This model is based on Rent's rule [2]. It optimizes multi-layer wiring by maximizing the clock frequency and minimizing the chip area. The model assumes that an ideal critical-path model has only one net with wirelength greater than the average wirelength of the design. The authors use a simple ratio of total wirelength to number of interconnects to achieve the average wirelength distribution. Using an interconnect density function, the authors calculate the dynamic power dissipation of the signal wires in the circuit. The authors then generate analytical models for the dynamic power dissipation in multi-layered two-tier (local and global) networks. The authors extend this work to n tiers and measure the variation in performance.

Rahman et al. [23] use Rent's rule [2] and the aforementioned Davis model [10] to evaluate system-level performance of 3D circuits. Using Rent's rule, the authors calculate the optimal wirelength distribution for 3D systems. The work follows a non-hierarchical method to extend the wirelength distribution to 3D. The authors use the critical-path model to compare system performances between 2D and 3D. The interconnect delay model used in this work uses Equation 2.9.

$$T_{50\%} = 0.4 R_{int} C_{int} + 0.7 \times (R_g C_{int}.FO + R_g C_L.FO + R_{int} C_L) \tag{2.9}$$

where R_{int} is the interconnect resistance, C_{int} is the interconnect capacitance, R_g is the output resistance of the gate, C_L is the load capacitance and FO is the fanout. This work assumes that the chip area is constrained by wiring between the logic gates. On the basis of their experiments, the authors infer that an increase in number of layers degrades the chip performance.

Cao et al. [8] optimize the local and global interconnect layers to reduce delay. This work suggests that the AR should be 0.5 and DC should be 0.7 for local wires. The proposed model considers both self and mutual inductances in interconnects along with capacitance and resistance to analyze the design. In addition, the second-order

Taylor's expansion is used to analytically measure the impact of interconnects on circuit performance.

Venkatesan et al. [27] optimize a multi-level architecture using Rent's rule [2]. This work determines the wire pitches for metal layers to optimize a design constrained by area, frequency, power dissipation, minimum number of metal layers and minimum repeater area. To create multi-level network designs, the authors fix timing on one level and shift a particular net to the next level if timing requirement is still not met. The authors minimize the macro cell area by using low spacing and high packaging density. In order to improve the frequency of a given design with constant levels and macro cell area, they increase the wiring pitch to maximize the unused area. This work uses Equation 2.3 to model the interconnect delay between equally spaced repeaters.

Gupta et al. [11] analyze energy, delay and bandwidth for various interconnect stacks for two technologies. Here, the authors consider via blockage and repeater insertion in designs. The work also considers stochastic wirelength distribution models proposed by Davis et al. [10] and Rent's rule [2]. From the study, the authors infer that a large pitch provides a high bandwidth at the expense of high power, owing to the high number of repeaters inserted.

2.5 Other Works

Elfadel et al. [21] introduce a tool for modeling, analysis and optimization of BEOL interconnect structures. This work discusses the challenges faced by a CAD tool in modeling interconnect performance. The flow of this work includes both RC and RLC interconnect modeling to ensure accuracy of circuit simulations. The work also compares the similarities between signal wires and transmission lines within a power bay.

Cao et al. [22] model a comprehensive technology extrapolation system. This work considers the rapid advancement in both technology and prediction methodologies. The authors implement an interconnect modeling tool which considers constraints such as the impact of inductance on noise and delay of critical paths, wire sizing, re-

peater optimization, and signal shielding techniques. This system models the impact of shielding, using both RC and RLC interconnect delay models.

Chen et al. [20] perform simulations for an inverter circuit to evaluate the performance of local metal layers in the $7nm$ node. This work shows that a reduction in the liner thickness lowers both wire and via resistances. The authors also analyze the sensitivity of the circuit speed towards high wire resistance when long wires are used at low dimensions. The work uses the tool described by Elfadel et al. [21] for all of their simulations and experiments.

Scheffer et al. [24] describe the shortcomings of the majority of the then-existing works. Among the main issues that the authors illustrate is the verification of predictions/models using a small number of testcases, which does not confirm their robustness. The authors also explain the difficulty of assuming a fixed wirelength distribution in a circuit because, in practice, there exists at least one path which always has a longer route than others. The authors also explain that many models ignore limiting factors such as access of pins in hierarchical designs, via blockages, via congestion issues, etc.

Song et al. [38] investigate the impact of RC scaling for the $7nm$ node. This work analyzes the feasibility of using fin de-population to reduce the input capacitance C_{pin} of the logic cell. In this work, the authors model design performance in terms of wire and device parasitics. The delay and power models are shown in Equations 2.10 and 2.11 respectively.

$$Delay = R_{wire} \times C_{pin} + R_{tr} \times C_{pin} + R_{tr} \times C_{wire} + 1/2 \times R_{wire} \times C_{wire} \tag{2.10}$$

$$Power = P_{dyn} + P_{lkg} = (C_{pin} + C_{wire}) \times {V_{dd}}^2 \times (1/delay) + I_{lkg} \times V_{dd} \tag{2.11}$$

where R_{wire} and C_{wire} are the interconnect resistance and capacitance, C_{pin} is the input pin capacitance of a gate, R_{tr} is the transistor resistance, P_{dyn} and P_{lkg} are the dynamic and leakage components of power, I_{lkg} is the leakage current and V_{dd} is the supply voltage. According to this work, the conventional delay scaling trend degrades with power because of the high wire RC. The authors observe that although fin pitch scaling reduces parasitic capacitances, it trades off with manufacturing difficulties. The work

concludes from experiments that at the $7nm$ technology node, wire resistance R_{wire} can be mitigated by controlling the input pin capacitance C_{pin}.

2.6 Predictive Technology Models

With the rapid advances in technology, the need for predictive libraries and technology files is high. For researchers, the existence of predictive models indicates the ability to continue work on other aspects of technology advancement. *Moore's law* states that the number of transistors on a chip will double every two years. feasible, nor is their performance satisfactory. In the IDM $22nm$ node, Intel introduced 3D FinFETs to replace planar MOSFETs [55]. As a part of our extensive validation process in this work, we use the ASU Predictive Technology Models (PTM) [47] for performance analysis using FinFETs.

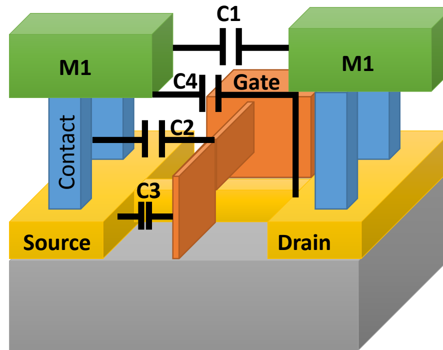
Table 2.1: Parameters of PTM-MG FinFET model.

Node	$20nm$	$14nm$	$10nm$	$7nm$
V_{dd} (V)	0.9	0.8	0.75	0.7
H_{fin} (nm)	28	23	21	18
T_fin (nm)	15	10	9	7
W_{eff} (nm)	71	56	51	43
L_g (nm)	24	18	14	11
Fin-pitch (nm)	60	32	28	22

Sinha et al. [47] model multi-gate FinFET devices for sub-$20nm$ technologies. In this work, the authors create bulk devices with initial parameters from the ITRS [41]. Table 2.1 lists the geometric parameters of their models across four advanced technology nodes. We use these generated models as a part of our validation studies.

Martins et al. [37] model a generic Open Cell Library at the $15nm$ technology node. This work contains various timing models and allows users to utilize multiple corners for design synthesis. The authors also characterize power for the devices in their library. We use this library for partial validation of our results. However, due to

certain missing information in the library, we only use it for reference and do not base our validation studies on results obtained using this library.



C1: S contact to D contact
C2: Gate to S/D contact
C3: Gate to S/D diffusion
C4: S contact to D diffusion

Figure 2.3: FinFET parasitic capacitances (adapted from [68]).

In the semiconductor industry, FinFET adoption has been steadily growing. FinFETs are less vulnerable to leakage power and have lesser area as compared to planar MOSFETs. However, FinFETs have a high gate capacitance due to their structure. Figure 2.3 shows the capacitance extraction model of FinFETs.

2.7 Reliability

In recent years, there has been significant work to analyze and reduce the impact of interconnects on reliability issues. However, the combination of low-κ dielectrics with copper interconnects is reducing the mean time-to-failure in advanced technologies.

The negative impact of TDDB increases with high DC because smaller spacing between wires enhances the electric field. Xia et al. [44] state that TDDB is dominated by via-wire spacing rather than wire-wire spacing. The authors state that in multi-layer interconnect systems, via-to-line is the limiting case for low-κ dielectric TDDB. The authors explain that this causes via bulge-out and via line mis-registration, leading to higher electric fields and thus a low mean time-to-failure. This work also indicates

that the *self-aligned via* (SAV) process has led to significant improvement in via-to-line space for Cu/low-κ TDDB reliability. The authors infer that rigorous process control and validation are key to avoid reliability issues.

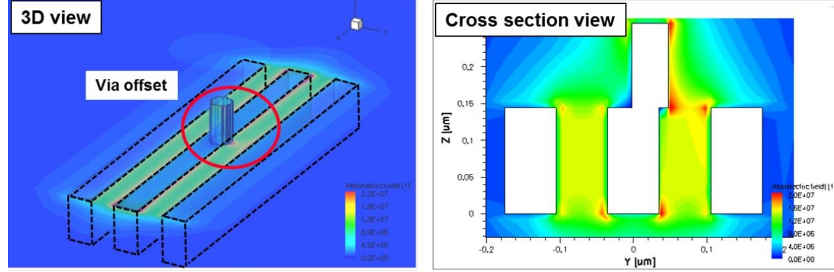


Figure 2.4: Misaligned via reduces interconnect spacing and enhances electric field [43].

As shown in Figure 2.4, Chan et al. [43] demonstrate that via misalignment due to the imperfection in pattering results in higher electric field around the via as compared to the average electric field between wires. In this work, the authors explore the advantages of signal-aware TDDB reliability estimation and post-detailed routing layout estimation to reduce the design margin used to avoid TDDB. This work indicates that the post-routing estimation method has the least impact on design timing and existing *design for manufacturing* (DFM) flows.

2.8 The ITRS BEOL Structure and Prediction

In this section, we review the experimental flow used by the ITRS ITWG [56] for its interconnect predictions.

The ITRS Interconnect ITWG uses the structure shown in Figure 2.5 to model interconnect parasitics. The structure has a central wire with two adjacent wires surrounded by two ground planes on top and bottom. They assume that the wire width is

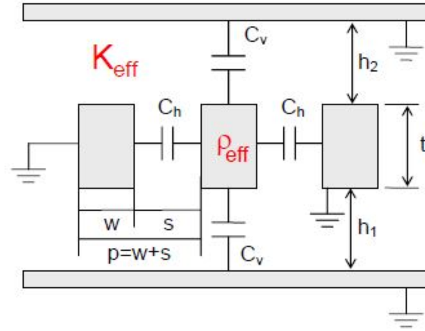


Figure 2.5: The ITRS model [56].

equal to the metal half-pitch. The ITRS model [56] extracts capacitance using the *Synopsys Raphael-2D* field solver tool [66] and does not consider the Miller coupling effect [34]. The resistance per unit length is calculated using Equation 2.12.

$$\frac{R}{L} = \frac{\rho_{eff}}{W \times T} \tag{2.12}$$

where L is the conductor length, W is the wire width, T is the wire thickness and ρ_{eff} is the effective resistivity from the ITRS Interconnect chapter [56].

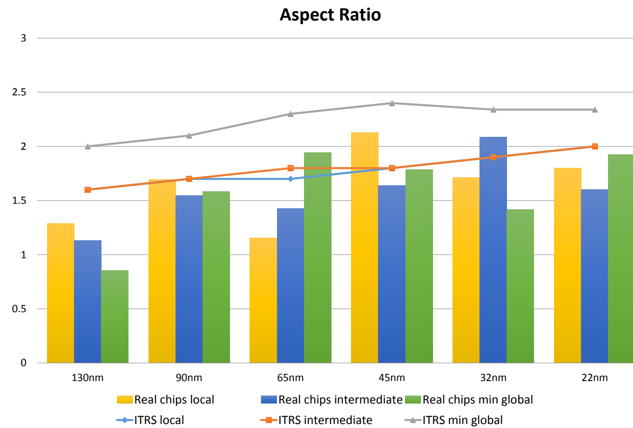


Figure 2.6: ITRS prediction versus real chip data [56].

We observe from the ITRS reports that in the early 2000's, the ITRS predictions for AR and dielectrics were aggressive. These predictions consider low ARs to counter crosstalk of clock and signal wiring levels. Figure 2.6 shows the differences between

the ITRS predictions and actual chip data up to the $22nm$ technology node which were provided to the ITRS Design ITWG for *microprocessor* (MPU) and *system on chip - consumer portable* (SoC-CP) system driver modeling work in 2013 by a company called Chipworks. After 2005, a major driving factor for the ITRS interconnect predictions has been the overall cost of the chip design cycle [56]. The ITRS now gives more importance to factors such as reliability, manufacturability and defect management. For sub-$14nm$ technologies, the ITRS [41] predicts that AR will be higher than 2.1.

2.9 Pitch Prediction

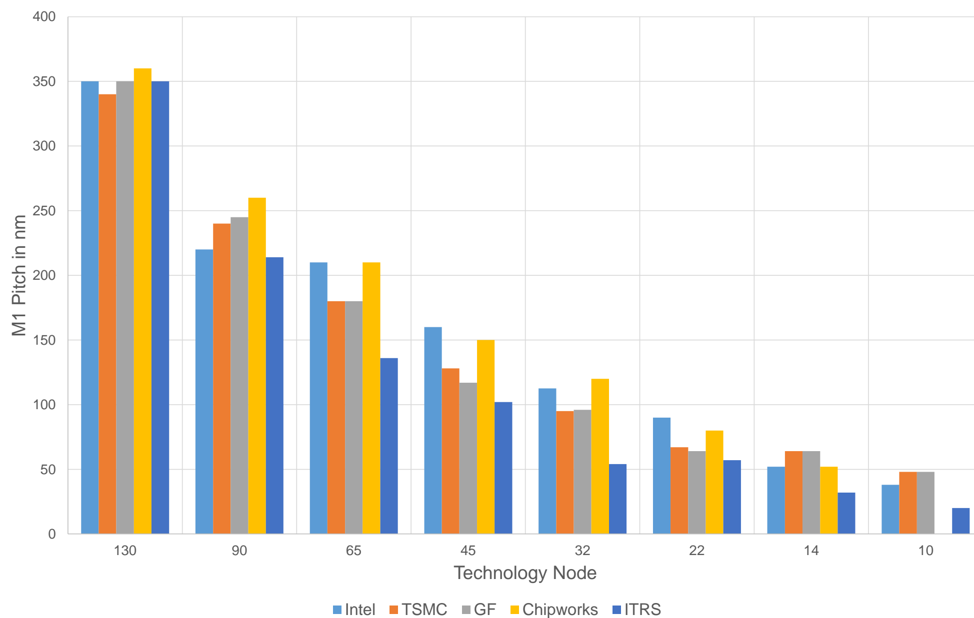


Figure 2.7: Metal-1 layer pitch comparison [50].

There are many conflicts about standardization of pitch for a technology node. After initial research of data, we found that the pitch is dependent on manufacturing techniques and processes. For a given technology node, the pitch value from a foundry can be different as compared to that in an IDM. Figure 2.7 shows the variation of pitch for Metal-1 layer across foundry and IDM technology nodes.

2.10 Summary of Previous Works

In this section, we summarize key works discussed in the chapter and compare them with industrial data.

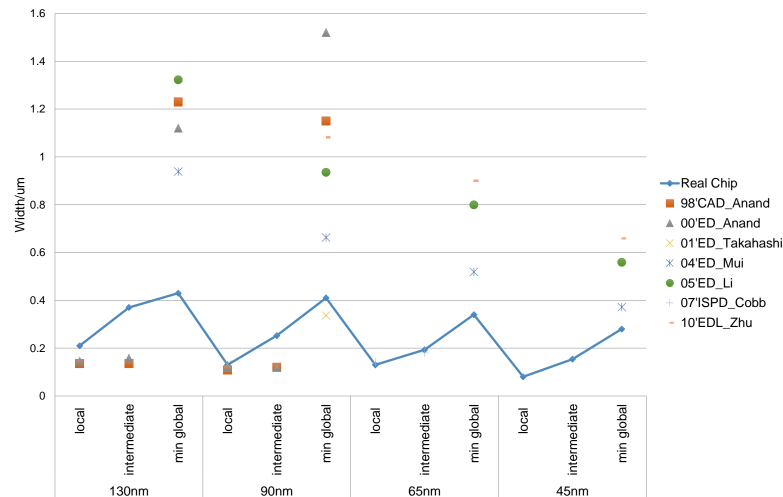


Figure 2.8: Comparison of interconnect width predictions made by previous works and real chips in the same technology [6], [9], [26], [19], [17], [29].

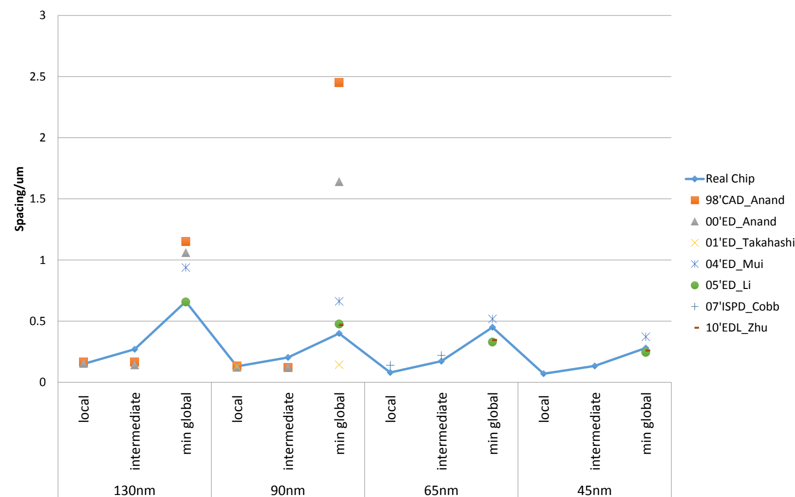


Figure 2.9: Comparison of interconnect spacing predictions made by previous works and real chips in the same technology [6], [9], [26], [19], [17], [29].

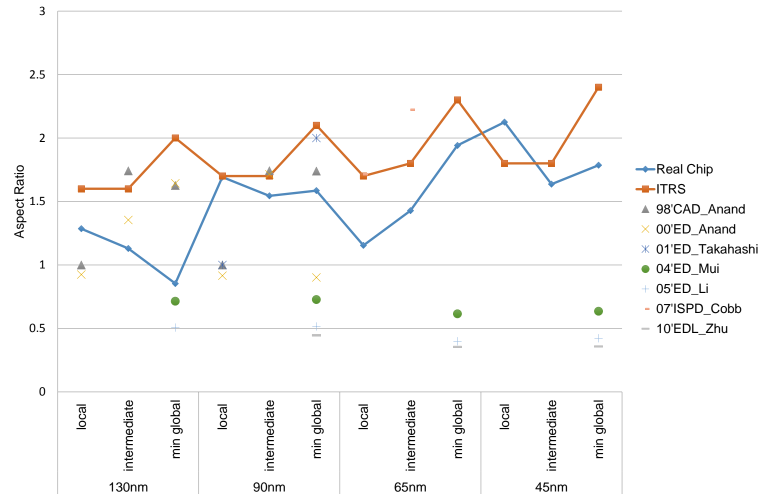


Figure 2.10: Comparison of interconnect AR predictions made by previous works and real chips in the same technology [6], [9], [26], [19], [17], [29], [56].

The works in Section 2.2 make wire width and spacing predictions at the system-level. The comparisons between previous works and manufactured chips are seen in Figures 2.8 (comparison of wire width prediction) and 2.9 (comparison of wire spacing prediction). Figure 2.10 compares the interconnect AR predictions between previous works and real chips.

Table 2.2 shows a summary of the key contributions and methodologies used by various previous works.

Table 2.2: Summary of previous works and categorization of our work.

Category	Author	Significance
Modeling	[1]	RC network delay model
	[2]	Rent's Rule
	[12]	Repeater Insertion
	[25]	RLC network delay model
	[4]	RC network delay model (with coupling)
Sys. level	[6], [5]	Low AR and DC for local interconnects
	[26]	High DC for global interconnects
	[28]	Designer friendly interconnect modeling
	[19]	Bandwidth considered for multi-layered interconnect modeling
	[29]	Scattering in metal considered for interconnect modeling
	Our Work	**Low AR and high DC for local layers with system-level verification**
Layer Assign.	[15]	Maximum wire width independent of delay
	[13]	Longer wires assigned to higher layers
	[17]	Low AR for the stack
	[7]	3D routing using dynamic programming
WLD	[18]	Simultaneous gate and wire sizing
	[10]	Multi-level networks with better area and frequency
	[23]	Multi-level networks in 3D systems
	[8]	Low AR and high DC for local layers
	[27]	Minimum number of layers in the stack and minimum repeater area
	[11]	Comparison of interconnect stacks
	Our work	**Low AR and high DC with random path model**
Others	[21]	CAD tool for interconnect modeling
	[22]	Technology extrapolation system with easy adaptation of new models
	[20]	Reduction of line thickness at $7nm$
	[24]	Shortcomings of interconnect predictive works
	[56]	Roadmap for interconnect scaling in advanced nodes
PTM	[47]	Device models for sub-$20nm$ technologies
	[37]	Device library for $15nm$ technology
Rel.	[43]	Estimation methods to lower design pessimism
	[44]	Importance of process control and positive effects of SAV

Chapter 3

Experiments for Interconnect Dimension Optimization and Validation

3.1 Our Flow

This chapter explains the flows that we use to determine and validate optimal interconnect dimensions for advanced technology nodes.

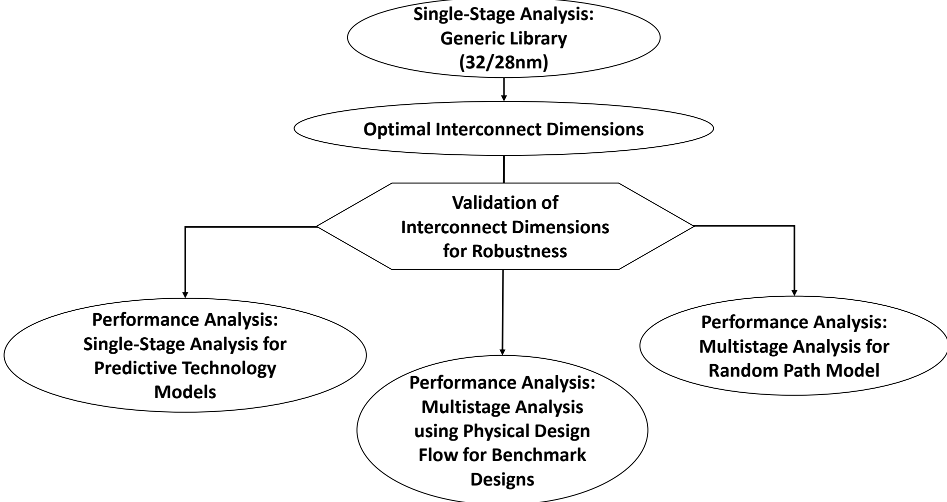


Figure 3.1: Flow of our experiment.

Figure 3.1 shows the steps used to optimize and validate interconnect dimensions in this work. We use single-stage analysis to find the optimal interconnect dimensions. We also perform validation to ensure robustness of the obtained optimal dimensions.

3.1.1 Optimization of Interconnect Dimensions

We start our experiment by constructing a representative interconnect structure using the *Synopsys Raphael-3D vG-2012.06* field solver tool [66]. We determine the optimal dimensions of the local metal layers of the interconnect stack by analyzing their performance in single-stage circuits. We use the flow shown in Figure 3.2 to determine

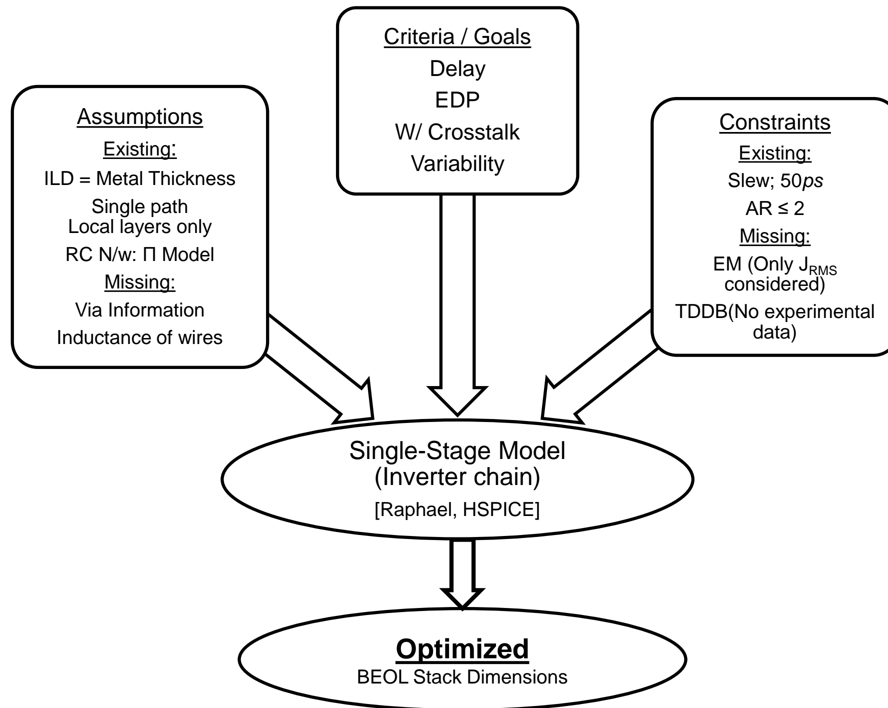


Figure 3.2: Overview of the single-stage flow.

interconnect dimensions for optimal circuit performance.

3.1.2 Validation

We validate the optimality of our determined dimensions using the following techniques.

1. Predictive technology models.

   Here, we use sub-20nm libraries developed by a research group at Arizona State University [46]. We perform single-stage analysis which considers device and voltage scaling.

2. System-level Analysis.

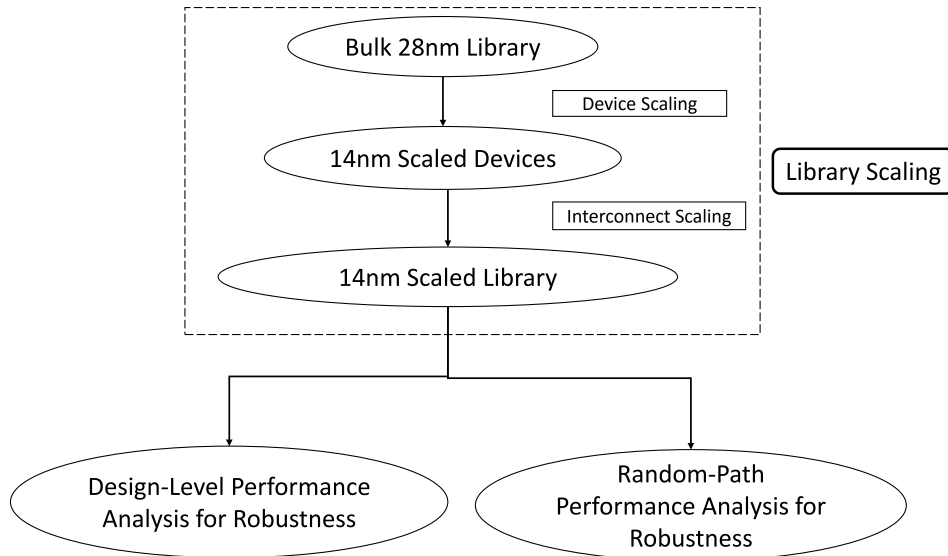


Figure 3.3: System-level validation flow.

   This analysis is performed using a library in which device and interconnect scaling have been incorporated. We use the following techniques to analyze the robustness of the determined interconnect dimensions at the system-level. Figure 3.3 illustrates this flow.

   (a) Design-level performance analysis.

   Here, we perform the entire physical design flow on benchmark designs to measure the impact of our optimized interconnect dimensions on the power and timing of a design.

(b) Random path model.

In this technique, we measure the power and timing of random paths generated based on random extracted paths from a routed benchmark design.

3.2 Our Interconnect Structure

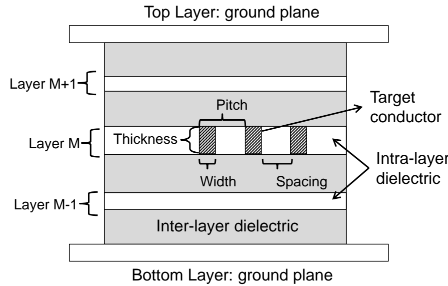


Figure 3.4: Cross-section of our model.

Our experiments use the interconnect structure shown in Figure 3.4, which is similar to the ITRS structure for interconnect modeling. The structure consists of three metal layers. The central metal layer has one target conductor in the middle and two neighboring conductors, one on either side. The length of each conductor is fixed at 40μm. The top and bottom metal layers have ground planes on the outside. The height of interlayer dielectric regions is the same as the metal wire thickness [56]. We use the same *dielectric permittivity* (κ) for interlayer and intralayer dielectrics. To retain assumptions followed by the other researchers, we use the *interlayer dielectric permittivity* (ILD) and *resistivity* (ρ) from the ITRS 2013 interconnect report [56]. The default metal thickness (T_{def}) is selected on the basis of AR and DC provided by the ITRS. In our experiment, we sweep a range of metal thickness values (T) to observe the trend of performance metrics such as slew-bounded delay and EDP. For each of the target technology nodes, the Metal-1 pitch, κ, resistivity (ρ), T_{def} and T are shown in Table 3.1. We then generate interconnect dimension configurations with a DC range of 0.4 to 0.775. The ground and coupling capacitance values are extracted using *Synopsys Raphael-3D* field solver tool [66]. The parasitic resistance per unit length is calculated using Equation 2.12.

Table 3.1: Parameters in our optimization studies.

Node	45nm	28nm	22nm	14nm	10nm	7nm
Metal-1 Pitch (nm)	150	120	80	52	42	36
κ [56]	2.75	2.775	2.59	2.19	1.87	1.8
ρ ($\mu\Omega/cm$)	4.08	4.03	4.77	6.84	9.38	15.02
T_{def} (nm)	135	114	80	55	42	36
T (nm)	135 - 155	110 - 130	70 - 90	40 - 60	30 - 50	25 - 45

3.3 Interconnect Optimization: Single-Stage Analysis

This is the first experiment in the optimization and validation of flow of interconnects in this work. Figure 3.5 illustrates the flow of our experiment to find optimal

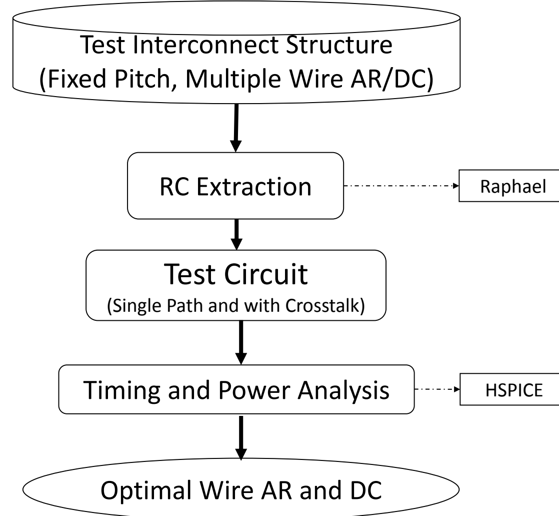
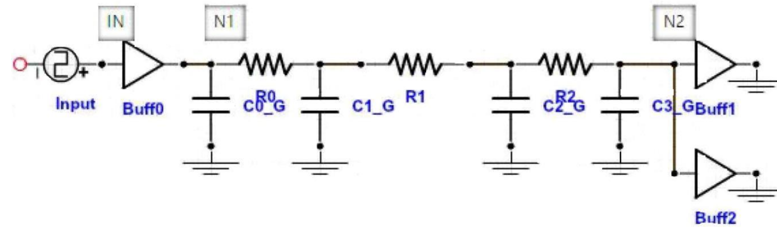


Figure 3.5: Interconnect optimization using single-stage analysis.

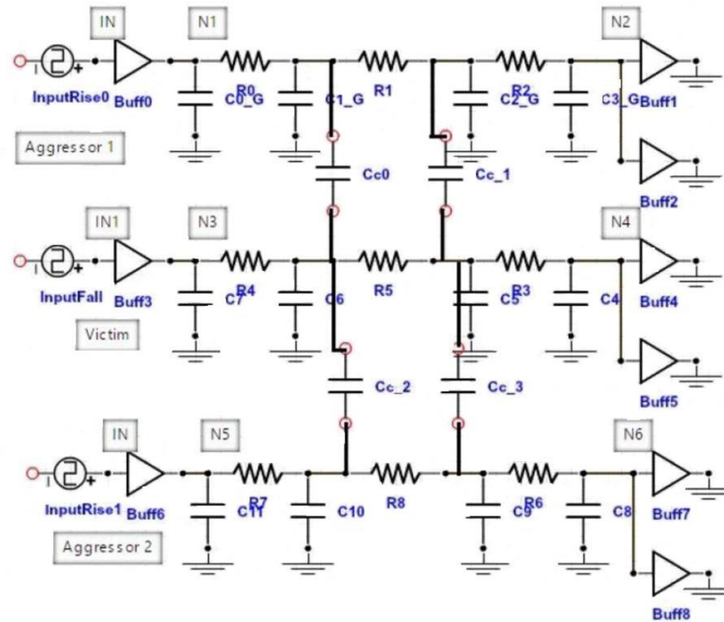
interconnect dimensions.

3.3.1 Methodology

We implement a single-stage inverter circuit, as shown in Figure 3.6(a), to measure the performance for all the wire dimension configurations using $Synopsys$ *HSPICE vG-2012.06-SP1* [62]. We substitute three wire segments in the circuit with Π models



(a) Structure of single path.



(b) Structure for crosstalk analysis.

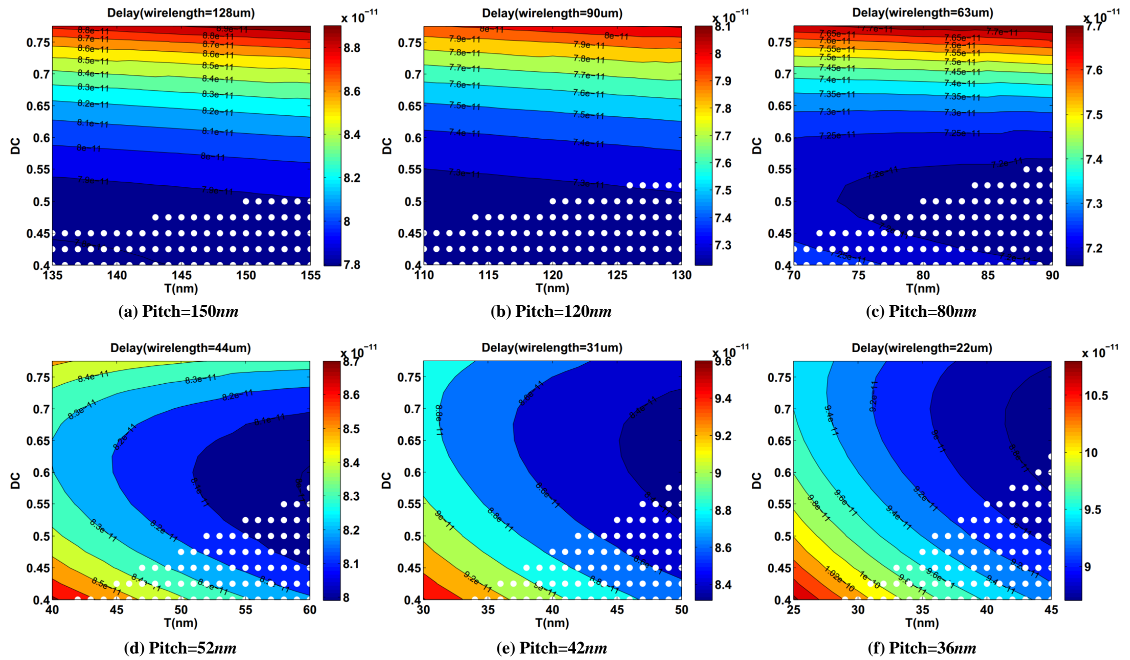
Figure 3.6: Circuits for simulation in HSPICE [62].

of the obtained wire resistance and capacitance values. The circuit has a driving buffer of size X16 from the *Synopsys 32/28nm Generic Library* [52]. We instantiate two additional buffers of the same size to act as loads. To consider coupling issues in our circuit simulations, we use an MCF of 1.4. The supply voltage of the circuit is $1V$. Listed below is the sequence of experiments we perform to determine the optimal interconnect dimensions.

Experiment 1.1: Slew and wirelength measurement.

Since we use devices from the *Synopsys 32/28nm Generic Library* [52] for optimization of interconnect dimensions, we start our experiment with the $28nm$ technology node. We set an upper bound of $50ps$ for the input and output slews at nodes IN and $N2$ respectively. We then calculate the wirelength which is slew-bounded for the described experimental settings. For the other technology nodes, we consider a wirelength scaling factor of $0.7x$ and calculate the corresponding slew upper bounds for all the wire dimension configurations. We then perform the following series of analyses.

Experiment 1.2: Slew-bounded delay measurement.



(a) Pitch=150*nm* (b) Pitch=120*nm* (c) Pitch=80*nm*

(d) Pitch=52*nm* (e) Pitch=42*nm* (f) Pitch=36*nm*

Figure 3.7: Slew-bounded delay (unit: sec) results.

We simulate the circuit for all the wire dimension configurations and measure the slew-bounded circuit delay as shown in Figure 3.7.

Experiment 1.3: Slew-bounded energy-delay product (EDP) measurement.

We measure the slew-bounded EDP for all the interconnect dimension configurations for the calculated wirelengths and illustrate its trend across technology nodes in Figure 3.8.

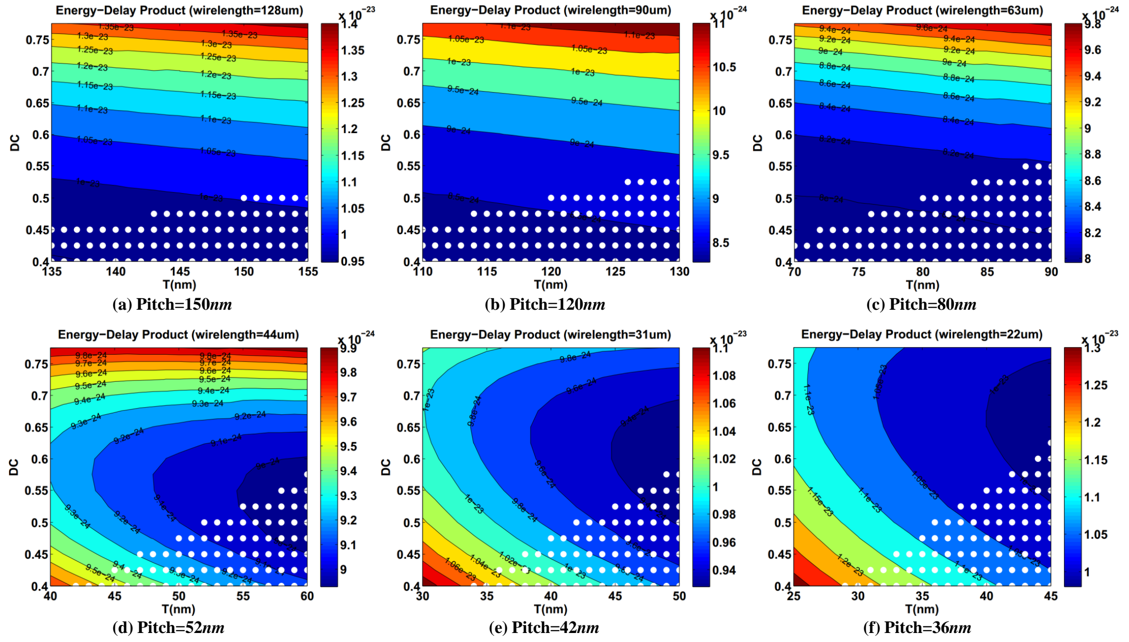


Figure 3.8: Slew-bounded energy-delay product (unit: $J \cdot sec$) results.

Experiment 1.4: Slew-bounded current density measurement.

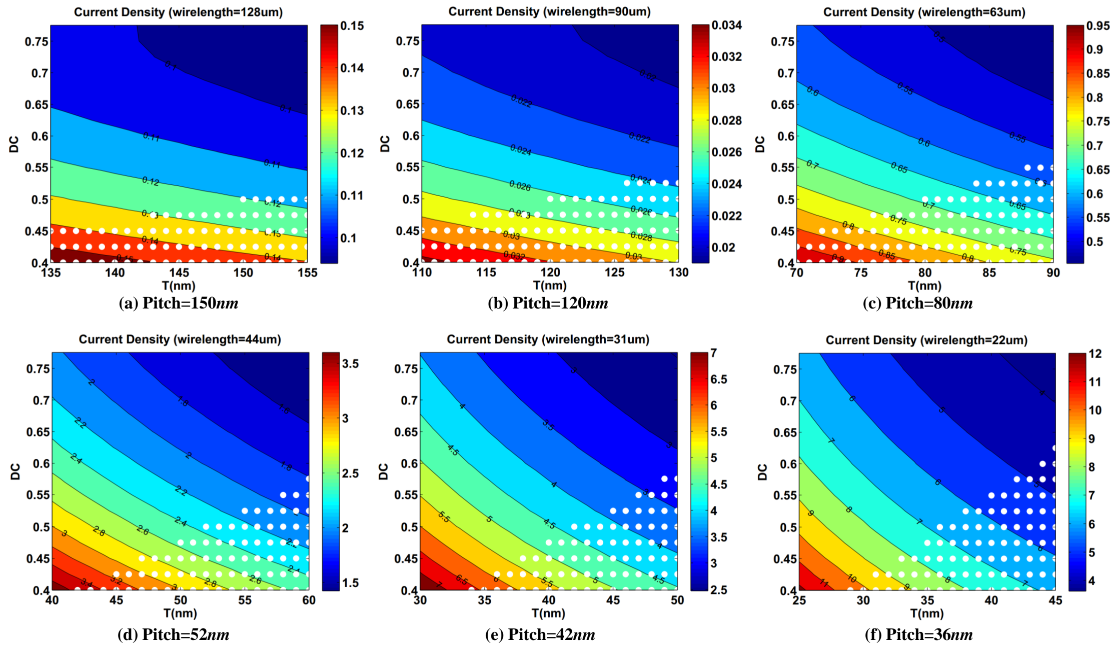


Figure 3.9: Slew-bounded current density (unit: MA/cm^2) results.

To gauge the susceptibility of EM in the interconnects, we measure the *RMS current density*, J_{RMS}, for all the interconnect dimension configurations and illustrate the results in Figure 3.9. J_{RMS} for our experiments is calculated by using *RMS current*, I_{RMS}, in the Equation 3.1.

$$J_{RMS} = \frac{I_{RMS}}{Pitch \times DC \times T} \tag{3.1}$$

where I_{RMS} is the RMS current.

Experiment 1.5: Measurement of the impact of process variation.

To check for resilience of the interconnect structure towards litho-etch variations, we measure the EDP for all the wire dimension configurations by accounting for the change in effective dimensions. Considering the dimension variations, the *effective wire width* W_{eff} can be formulated using the CD variation (ΔCD) and the *nominal width* (W_{nom}) by Equation 3.2.

$$W_{eff} = (1 \pm \Delta CD) \times W_{nom} \tag{3.2}$$

Similarly, the effective wire thickness T_{eff} due to process variation can be calculated using Equation 3.3.

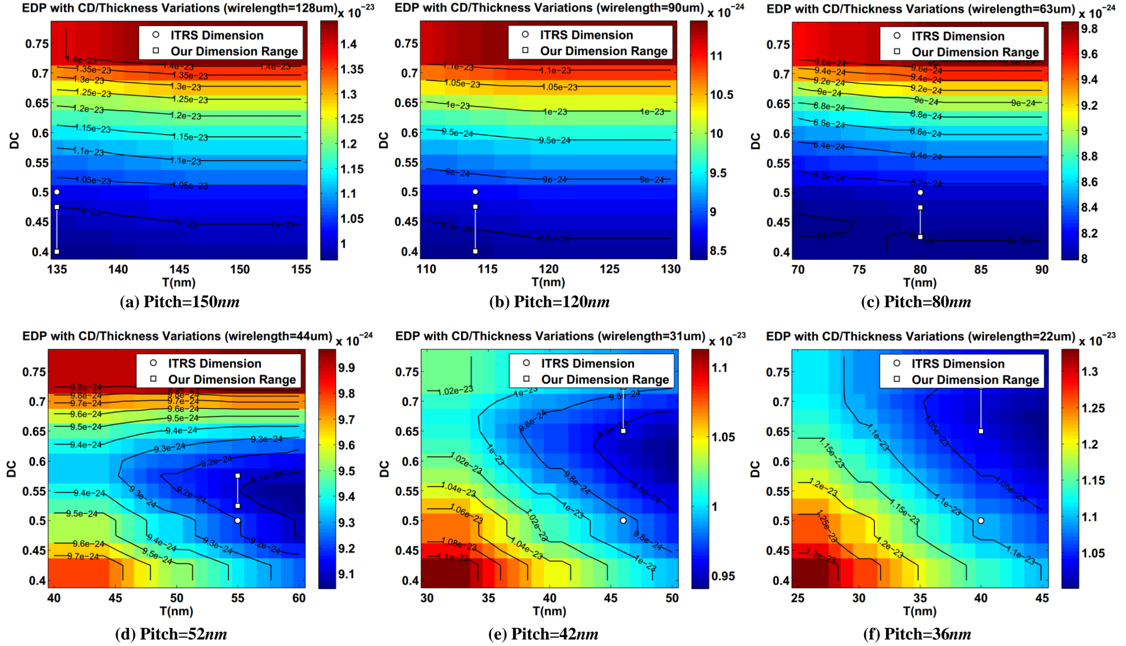
$$T_{eff} = (1 \pm \Delta T) \times T_{nom} \tag{3.3}$$

where ΔT is the variation in wire thickness and T_{nom} is the nominal wire thickness.

In our experiment we calculate the worst EDP for each wire dimension configuration considering $\pm 10\%$ CD variation and $\pm 10\%$ wire thickness variation and illustrate the trend in Figure 3.10.

Experiment 1.6: Crosstalk evaluation.

We evaluate the performance of our interconnects in the presence of neighboring wires. As shown in Figure 3.6(b), the circuit has a target victim wire with one aggressor wire on each side. The structure of each wire is similar to that of the single path circuit 3.6(a). The supply voltage is $1V$. The coupling capacitance values are obtained from the previous *Synopsys Raphael* [66] simulation results in Section 3.1.1. We measure



(a) Pitch=150nm (b) Pitch=120nm (c) Pitch=80nm

(d) Pitch=52nm (e) Pitch=42nm (f) Pitch=36nm

Figure 3.10: Slew-bounded energy-delay product (unit: $J \cdot sec$) results with process variation: ($\Delta T = \pm 10\%$, $\Delta CD = \pm 10\%$).

the impact of crosstalk on delay of the victim interconnect for all the wire dimension configurations between nodes $N3$ and $N4$.

3.3.2 Results

In our experiment, we study the impact of various interconnect dimensions on circuit performance. We set the wirelength and slew constraint for each node in the circuits shown in Figures 3.6(a) and 3.6(b) to be constant for all T and DC pairs.[2] We determine optimal interconnect dimensions based on the best EDP results at T_{def} for all nodes.

To evaluate the benefits of our interconnect dimensions, we compare the obtained results with those obtained using the ITRS predictions. Table 3.2 shows the performance metric comparison, where positive values indicate an improvement of our result as compared to the results generated by the ITRS dimensions.

[2]In the all result plots, the x-axis corresponds to wire thickness T and y-axis corresponds to DC. The white dotted section indicates an interconnect with AR ≥ 2.

Table 3.2: Results comparison of EDP, $Delay$, EDP with variation and J_{RMS} between the ITRS and our determined dimensions.

Node		$45nm$	$28nm$	$22nm$	$14nm$	$10nm$	$7nm$
Metal-1 Pitch (nm)		150	120	80	52	42	36
T_{def} (nm)		135	114	80	55	46	40
Wirelength (μm)		128	90	63	44	31	22
Slew (ps)		60	50	45	70	90	95
ITRS	AR	1.8	1.9	2	2.1	2.2	2.2
	DC	0.5	0.5	0.5	0.5	0.5	0.5
Our	AR	2.25	2.375	2.35	1.9	1.54	1.53
	DC	0.4	0.4	0.425	0.55	0.65	0.725
	Δ *EDP* (%)	+3.50	+2.52	+0.65	+0.21	+1.26	+3.37
	ΔJ_{RMS} (%)	-20.29	-21.40	-15.72	+8.38	+21.74	+29.78
	Δ *EDP*$\pm Var$ (%)	+4.68	+3.51	+1.30	+1.13	+2.94	+5.03
	$\Delta Delay$ (%)	+0.74	+0.36	-0.48	+0.44	+1.34	+2.28
$XTalk$	$\Delta Delay$ (%)	-10.75	-9.54	-5.10	+1.26	+2.60	+4.15

1. Delay results and comparison.

   Figure 3.7 shows the slew-bounded delay results for nodes $45nm$, $28nm$, $22nm$, $14nm$, $10nm$ and $7nm$. These results show a trend of reduced delay as DC surpasses 0.5 for advanced nodes. While the *dark blue* region with larger metal thickness has best results, the trend is the same even when we consider a T_{def} closer to the metal pitch value. However, for the existing nodes $45nm$, $28nm$ and $22nm$, it is clear that DC $\leq$ 0.5 is a better choice. We define the delay improvement, $\Delta Delay$, by Equation 3.4.

$$\Delta Delay = \frac{Delay_{ITRS} - Delay_{Ours}}{Delay_{ITRS}} \tag{3.4}$$

   We observe the delay improvement trend in Table 3.2.

2. Δ EDP results and comparison.

   Figure 3.8 illustrates the EDP results for all nodes. We observe an improvement trend for the interconnect dimension configurations with DC $>$ 0.5 (high DC) and AR $<$ 2 (low AR). Since EDP involves energy and delay, we use it to determine our range of optimal dimensions for advanced technology nodes. We define the EDP improvement Δ ***EDP*** by Equation 3.5.

$$\Delta EDP = \frac{(EDP_{ITRS} - EDP_{Ours})}{EDP_{ITRS}} \tag{3.5}$$

   Based on the EDP results at T_{def}, we determine optimal AR and DC values which give the smallest EDP. From the results in Table 3.2, we observe that for the 14nm node, our interconnect dimensions of AR $=$ 1.9 and DC $=$ 0.55 have an EDP improvement of 0.15% over that of the ITRS dimensions. The EDP improvement increases for smaller nodes as we increase DC and reduce AR. Unlike advanced nodes, the existing technologies have better EDP as compared to the ITRS values only when DC $<$ 0.5 and AR $>$ 2.2 for a given T_{def}. We verify that our dimensions can improve EDP by $0.21\% \sim 3.5\%$ as compared to that of the ITRS predictions for the given technology nodes. For advanced technology nodes, we also observe that for the given T_{def}, dimension combinations with AR $>$ 2 (high AR) and DC $<$ 0.5 (low DC) do not have a comparatively good performance. These combinations have EDP in the *lighter blue* to *red* regions, depending on the technology node.

3. EDP results and comparison with process variation.

   Figure 3.10 shows the EDP results considering $\pm 10\%$ CD variation and $\pm 10\%$ wire thickness variation. The trend of results confirm our previous inference of using DC $>$ 0.5 (high DC) values to achieve smaller EDP for advanced nodes. We use Equation 3.5 to calculate the relative improvement in EDP with process variation. Table 3.2 shows that when we consider a process variation of $\pm 10\%$, our dimensions can achieve $1.13\% \sim 5.03\%$ EDP improvement over the ITRS predictions. These results indicate that interconnect structures with AR $<$ 2 (low

AR) and DC > 0.5 (high DC) for advanced nodes are more resilient to process variations.

4. Current density results and comparison.

   Figure 3.9 shows the J_{RMS} results for all the test technology nodes. Since the increase of wire width reduces the interconnect current density, large DC results in small current density as shown in the upper right region of each result plot. We observe that for a constant DC, J_{RMS} increases with shrinking geometry. Hence, we suggest utilization of higher DC especially for sub-10nm nodes. We define the relative improvement of the J_{RMS} using Equation 3.6.

$$\Delta J_{RMS} = \frac{J_{ITRS} - J_{Ours}}{J_{ITRS}} \tag{3.6}$$

   We also see that the improvements in J_{RMS} for 10nm and $7nm$ technologies are more than 20%. Since current density is a primary influencing factor for EM, interconnects with our dimensions can perform better in mitigating EM reliability issues in future nodes.

5. Delay results and comparison with crosstalk.

   In Table 3.2, we show the relative impact of crosstalk on wire delay (referred to as $XTalk$). We use Equation 3.4 to calculate the improvement in delay. This initial study shows that as the geometry shrinks, the positive effect of reduced AR on crosstalk noise dominates the adverse impact of high DC.

3.4 Single-Stage Validation: Predictive Technology Models

This is the second experiment in the optimization and validation flow of interconnects in this work. In this work, we use the ASU PTM library [46] as a sanity check for our previously obtained results from Section 3.3.2. The ASU PTM models are created with reference to the ITRS reports [41]. Another reason for using the ASU PTM

models is to observe the behavior of our interconnect dimensions in the presence of FinFETs instead of planar MOSFETs. This experiment analyzes the performance of our interconnect dimensions for 20nm, 14nm, 10nm and 7nm technology nodes.

3.4.1 Methodology

We implement inverter circuits with four fanout options (FO1, FO2, FO3 and FO4), as shown in Figure 3.11, to measure the performance for all the wire dimension configurations using $Synopsys$ *HSPICE vG-2012.06-SP1* [62]. We only calculate the Δ *EDP* values to compare the performance of our interconnect dimensions over that of the ITRS.

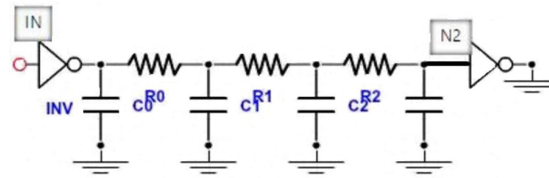
The circuit has a driving inverter which is built based on PTM-MG HP device models. We substitute three wire segments in the circuit with Π models of the obtained wire resistance and capacitance values. The inverters at the load are of the same size as the driver. To consider the impact of crosstalk in our simulations, we use an MCF of 1.4. The supply voltage of the circuit is consistent with V_{dd} in Table 2.1 [47]. Given below is the sequence of experiments we perform to validate the performance of our previously obtained results in Table 3.2.

Experiment 2.1: Wirelength measurement.

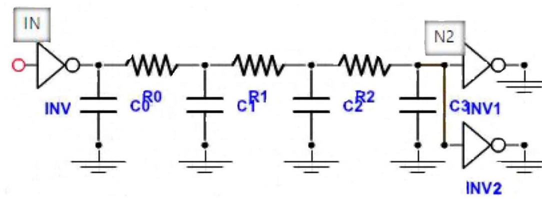
We set an upper bound of 50ps for the input and output slews as shown in Figure 3.11 at nodes IN and $N2$ respectively. We then calculate the slew-bounded wirelength for all the wire dimension configurations. We select the minimum wirelength for the following performance metrics measurement. The input signal is a pulse source with a period of $2ns$ and a peak voltage of $1V$.

Experiment 2.2: Slew-bounded performance metrics measurement.

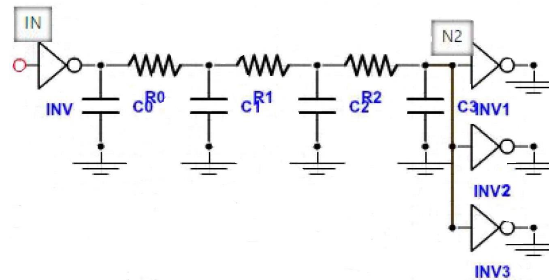
We simulate the circuit for all the wire dimension configurations and measure the circuit delay, EDP and current density. The input signal is a pulse with a period of $2ns$ and a peak voltage of V_{dd}. We measure rise delay and fall delay and use the mean value as the circuit delay. The power is measured during one clock period.



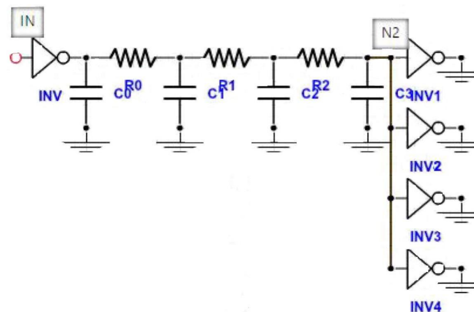
(a) Structure of single path with FO1.



(b) Structure of single path with FO2.



(c) Structure of single path with FO3.



(d) Structure of single path with FO4.

Figure 3.11: Circuits for simulation in HSPICE [62].

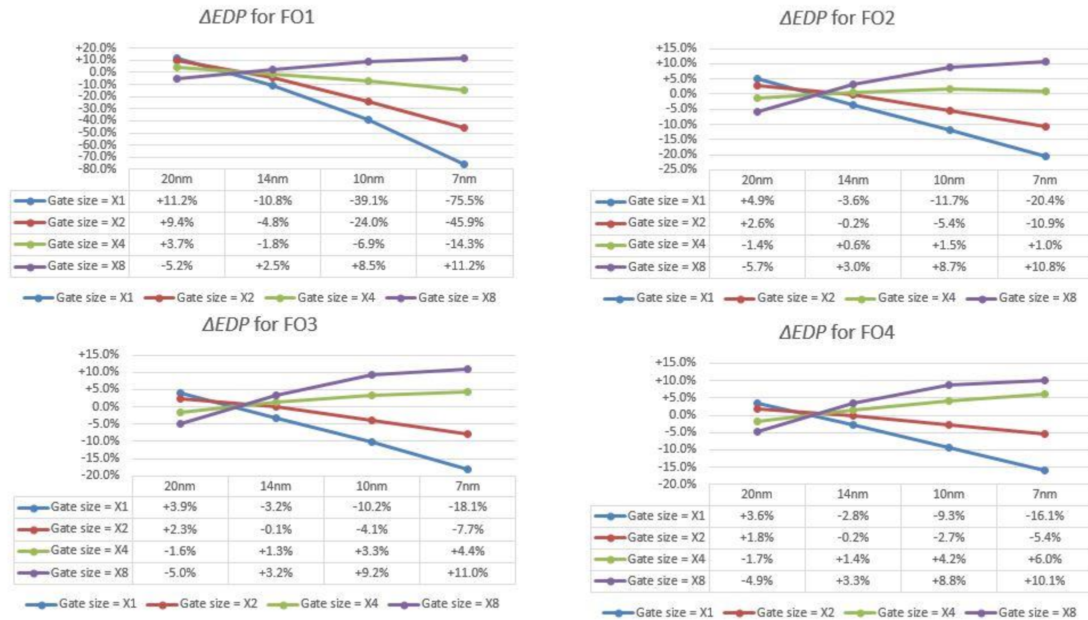


Figure 3.12: Performance variation across gate sizes and fanouts.

3.4.2 Results

Our initial experimental results shown in Section 3.3.2 do not consider the impact of device and voltage scaling on the performance of our suggested interconnect dimensions.

Figure 3.12 shows the variation in performance of our obtained optimal interconnect dimensions in advanced technologies. This performance metric comparison indicates improvement with respect to the ITRS dimension predictions with a positive value. Since our optimization is based on Δ *EDP* results (with respect to the ITRS predictions), we include the comparison of this metric for various circuit, gate size and wirelength combinations. We define the EDP improvement Δ *EDP*, by Equation 3.5.

We see a consistent improvement in EDP values with increasing fanout and gate sizes. In the $7nm$ testcase, we observe that EDP varies from -75.5% (for a gate size of X1 and fanout of 1) to +10.1% (for a gate size of X8 and fanout of 4). This EDP variation trend is due to the following factors.

1. Total capacitance of interconnects.

   Our suggested optimized DC for the $7nm$ technology node is 0.725 whereas the DC predicted by the ITRS is 0.5. From the studies in Section 3.2, we see that the total capacitance for interconnects with our optimized dimensions is 1.68 times higher than that of the ITRS. For smaller fanouts (such as FO1) and the same device, power dissipated by our optimized interconnects will be considerably higher than of the interconnects predicted by the ITRS. This results in a Δ *EDP* drop of 75%. However, for the same gate size, when the fanout increases to four, the Δ *EDP* drop is only 16.1%. This is because the impact of interconnect parasitics reduces with increasing effective load capacitance.

2. Increase in number of fins.

   To increase gate size in the given models, we increase the number of fins. As we increase the gate size, the conductivity of the device improves which in turn reduces the delay, thus increasing Δ *EDP*.

3.5 Multistage Validation Enablement: Library Scaling

This is the third experiment in the optimization and validation flow of interconnects in this work. To perform system-level experiments, we need to scale the library such that the power and timing performances of the devices and interconnects are comparable to those of an advanced node. This section elucidates the methodology that we follow to scale the library. First, we obtain appropriate scaling factors to scale the effective power and delay of devices in a library. The intention is to scale the metrics of a library containing planar MOSFETS (2D) to a library containing FinFETs (3D). We then scale the effective resistance and capacitance of the entire interconnect stack to measure the impact of optimized interconnect dimensions on the design performance.

We illustrate the entire scaling flow (inclusive of device and interconnect scaling) in Figure 3.13. We use this scaled $14nm$ library to perform multistage performance validation of the obtained interconnects.

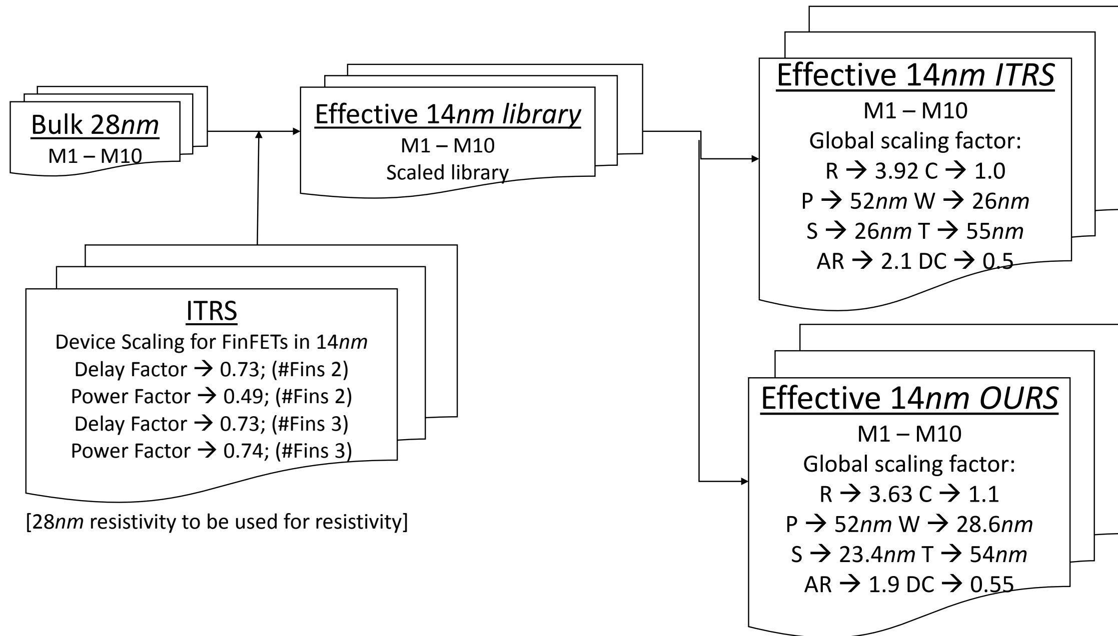


Figure 3.13: Scaling of front-end and back-end of the $28nm$ Bulk library to IDM $14nm$.

3.5.1 Methodology: Device Scaling

In this methodology, we see the steps in which capacitance, delay and power scaling factors are calculated to scale a foundry BULK $28nm$ library to IDM $14nm$. The channel-width based capacitance is calculated for the $28nm$ technology node using the information in the ITRS PIDS and MPU chapters. This capacitance is used to calculate the power and delay of devices.

1. Available data for $28nm$.

   (a) Pitch.

   We know that the Metal-1 pitch for the $28nm$ technology node is $120nm$.

   (b) Gate Capacitance.

   From the ITRS PIDS chapter [59], the value of average total gate capacitance per micron (C_{avg}) for a device in the BULK $28nm$ library is $0.84fF/\mu m$.

   (c) Drain current.

   From the ITRS PIDS chapter [59], I_{dsat} for an nmos is $1.21mA/\mu m$.

(d) Effective width.

From the ITRS MPU model [58], we see that the channel width normalized to the Metal-1 pitch ($W_{normFactor}$) is 5.

2. Available data for $14nm$.

(a) Pitch.

The Metal-1 pitch for the IDM $14nm$ technology node is $52nm$.

(b) Gate capacitance.

From the ITRS PIDS Tables [57], the average total gate capacitance per micron (C_{avg}) for a device in the $14nm$ library is $1.1fF/\mu m$.

(c) Height of FinFETs.

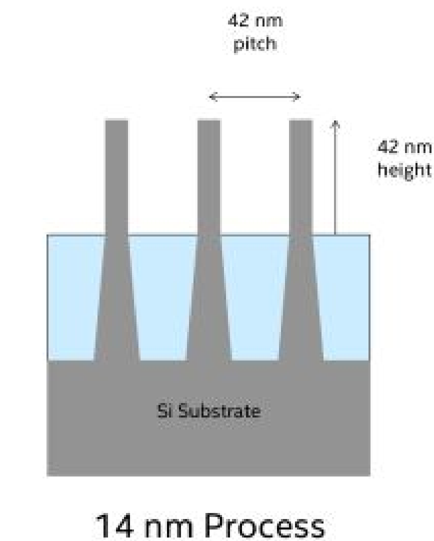


Figure 3.14: Intel $14nm$ fin information [60].

The height of the FinFETs (Fin_{height}) used is $42nm$ for devices with two and three fins in the Intel technology [31] [60] . Figure 3.14 shows the fin height and fin pitch measurement. We test the performance of the IDM $14nm$ library for two and three fins for the FinFETs.

3. Data to be calculated.

   (a) For $14nm$, we need to calculate the following data:

      i. Effective width.

         The effective width of a FinFET is shown in Equation 3.7 [71].

$$W_{eff} = \#Fins \times (2 \times Fin_{height} + W_{fin}) \quad (3.7)$$

         where W_{eff} is the effective width, Fin_{height} is the height of the fin and W_{fin} is the fin width.

      ii. Drain current.

         The drain current is calculated using Equation 3.8 [71].

$$I_{dsat} = \frac{W_{eff} \times \mu \times C_{ox}}{L} \times (V_g - V_{th})^2 \quad (3.8)$$

         where I_{dsat} is the drain current, C_{ox} is the capacitance of per unit gate area, L is the gate length, μ is the mobility, V_g is the gate voltage and V_{th} is the threshold voltage.

         From Equation 3.8, the drain current of FinFETs is proportional to the number of fins. Thus, the driving capability of a triple-fin FinFET is greater than that of a double-fin FinFET. This relationship is also established in [69] and [70].

         A. Double-fin FinFET

            Natarajan et al. [31] state that for IDM $14nm$ technology node, n-type FinFETs with two fins have a drain saturation current of $1.04mA/\mu m$ (used for calibration).

         B. Triple-fin FinFET

            Since drain current is directly proportional to the number of fins, we deduce the drain current for a triple-fin n-type FinFET in Equation 3.9.

$$I_{dsat,3fins} = 1.5 \times I_{dsat,2fins} \quad (3.9)$$

(b) The calculations needed for 28nm and 14nm library are shown as follows.

i. Normalized width.

From the ITRS PIDS chapter [57], for FinFETs, the $W_{normFactor}$ is a function of W_{eff}. To calculate the $W_{normFactor}$, we use Equation 3.10.

$$W_{normFactor} = \frac{W_{eff}}{Pitch \times 0.5} \tag{3.10}$$

ii. Effective gate capacitance.

The average gate capacitance (for FinFETs and planar MOSFETs) is normalized to the half-pitch of the Metal-1 layer using Equation 3.11 because the effective gate capacitance is dependent on the channel width.

$$C_{eff} = C_{avg} \times W_{eff} \times Pitch \times 0.5 \tag{3.11}$$

iii. Delay.

Using the effective gate capacitance values, we calculate the intrinsic delay using Equation 3.12.

$$Delay = \frac{C_{eff} \times V}{I_{dsat}} \tag{3.12}$$

iv. Power.

Using the effective gate capacitance values, we calculate the internal power using the Equation 3.13.

$$Power = C_{eff} \times V^2 \tag{3.13}$$

3.5.2 Results

We use information from Section 3.5.1 to calculate the scaling factors for Bulk 28nm to IDM 14nm, as shown in Table 3.3.

Table 3.3: Library scaling factors for Bulk 28nm to IDM 14nm.

Technology Node	Bulk 28nm	IDM 14nm	
$\#Fins$	nil	2	3
Voltage (V)	1.10	0.86	
Drain Current (I_{dsat}) ($mA/\mu m$)	1.20	1.04	1.56
Local-layer Pitch (μm)	0.12	0.052	
C_{avg} ($fF/\mu m$)	0.84	1.10	
Fin_{height} (μm)	nil	0.042	0.042
W_{Fin} (μm)	nil	0.008	
W_{eff}	5.00	7.08	10.62
C_{eff} ($fF/\mu m$)	0.25	0.20	0.30
Capacitance scaling factor	1.0	0.8	1.2
Delay (ps)	0.23	0.17	0.17
Delay Scaling Factor	1.00	0.73	0.73
Power ($fJ/\mu m$)	0.30	0.15	0.22
Power Scaling Factor	1.0	0.49	0.74
Leakage Current ($nA/\mu m$)	0.34	0.10	0.10
Leakage Power ($fJ/\mu m$)	0.37	0.09	0.09
Leakage Power Factor	1.00	0.23	0.23

Table 3.3 shows that capacitance scaling is dependent on the number of fins. This change in the number of fins also determines the variation in power scaling of the library. This observation is similar to the work in [38].

3.5.3 Methodology: Interconnect Scaling

Apart from device scaling, we also perform appropriate scaling of the resistance and capacitance values for the interconnects. For this, we use the results from our Raphael studies in Section 3.1.1. Since the resistance results from Raphael are incorrect, we use the resistivity of the Metal-1 layer from our 28nm bulk (4.6$\Omega/\mu m$) library to obtain reasonable resistance values (at 14nm).

3.5.4 Results

Table 3.4: RC scaling factors for local layers using Raphael results.

Technology	$28nm$ Bulk	$14nm$ ITRS	RC Factors	$14nm$ Ours	RC Factors
Width (nm)	0.051	0.026		0.0286	
Thickness (nm)	0.11	0.055		0.054	
DC	0.425	0.5		0.55	
AR	2.16	2.1		1.9	
R $(\Omega/\mu m)$	8.2	32.17	3.92	29.79	3.63
$C_c(F/\mu m)$	1.24E-16	1.25E-16		1.36E-16	
C_g $(F/\mu m)$	8.86E-18	5.84E-18		5.94E-18	
C_{total} $(F/\mu m)$	1.33E-16	1.31E-16	0.98	1.42E-16	1.07

From the $28nm$ library we obtain the resistance per micron to be 4.6Ω. Using this value and our Raphael results, we generate the interconnect RC scaling shown in Table 3.4.

3.6 Multistage Validation: Design-Level Performance Analysis

This is the fourth experiment in the optimization and validation flow of interconnects in this work.

This system-level validation of our interconnect dimensions is performed by running the entire physical design flow from synthesis to place-and-route on benchmark designs as shown in Figure 3.15.

3.6.1 Methodology

We apply the interconnect dimensions from Section 3.5.4 to the benchmark designs *USB, AES* and *DMA*, which are synthesized using a scaled $14nm$ library as shown in Section 3.5.1. After obtaining the scaled libraries for slow and fast process corners,

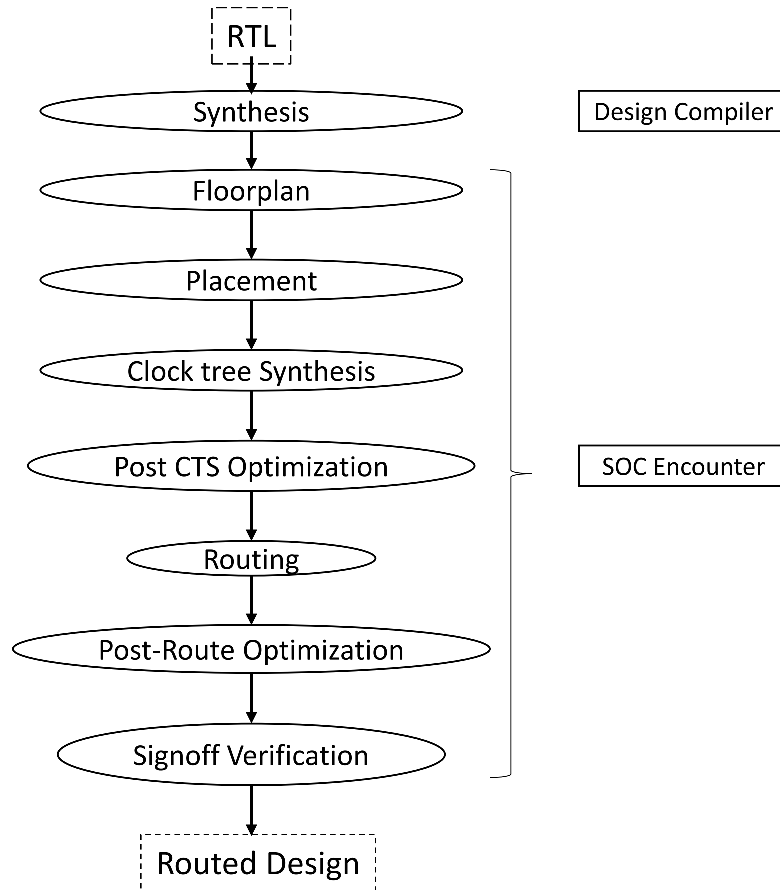


Figure 3.15: Design-level validation flow.

we proceed to perform the physical design flow for power and timing analysis of the routed designs. This flow is implemented using *Synopsys Design Compiler* [61] and *Cadence SOC Encounter* [67]. Although this method does not consider the routing difficulties that are seen in advanced nodes, we obtain an estimate of the performance of designs using timing and power analysis.

3.6.2 Results

In this section, we observe and analyze the results obtained from our multistage simulations. From Table 3.5, we see the power and timing analysis for two scaled libraries ($\#Fins = 2$ and $\#Fins = 3$).

Table 3.5: Power and timing analysis results for design-level validation.

Interconnect Parameters		ITRS		OURS	
Local-layer Pitch (nm)		52			
AR		2.1		1.9	
DC		0.5		0.55	
Design Parameters					
$\#Fins$		2	3	2	3
USB	Clock (ns)	0.7			
	# Inst	358	439	371	439
	Dyn. Power (mW)	9.55E-01	1.36E+00	1.00E+00	1.48E+00
	Avg. Dyn Power/Inst (mW)	2.67E-03	3.09E-03	2.70E-03	3.38E-03
	Leak. Power (mW)	2.00E-03	3.30E-03	2.10E-03	3.50E-03
	Avg. Leak. Power/Inst (mW)	5.59E-06	7.52E-06	5.66E-06	7.97E-06
	Slack (ns)	2.70E-02	2.60E-02	1.00E-02	2.60E-02
	Avg. Slack/Inst (ns)	7.54E-05	5.92E-05	2.70E-05	5.92E-05
AES	Clock (ns)	1.5			
	# Inst	11229	1167	11351	11884
	Dyn. Power (mW)	1.56E+01	2.01E+01	1.71E+01	2.11E+01
	Avg. Dyn Power/Inst (mW)	1.39E-03	1.72E-02	1.50E-03	1.77E-03
	Leak. Power (mW)	3.60E-02	4.22E-02	3.67E-02	4.57E-02
	Avg. Leak. Power/Inst (mW)	3.21E-06	3.61E-05	3.23E-06	3.85E-06
	Slack (ns)	0	1.10E-02	1.40E-02	9.00E-03
	Avg. Slack/Inst (ns)	0	9.43E-06	1.23E-06	7.57E-07
DMA	Clock (ns)	0.8			
	# Inst	1686	1648	1693	1679
	Dyn. Power (mW)	2.87E+00	3.76E+00	3.25E+00	3.78E+00
	Avg. Dyn Power/Inst (mW)	1.70E-03	2.28E-03	1.92E-03	2.25E-03
	Leak. Power (mW)	9.10E-03	9.50E-03	1.05E-02	1.01E-02
	Avg. Leak. Power/Inst (mW)	5.40E-06	5.76E-06	6.20E-06	6.02E-06
	Slack (ns)	2.50E-02	3.00E-03	2.40E-02	2.70E-02
	Avg. Slack/Inst (ns)	1.48E-05	1.82E-06	1.42E-05	1.61E-05

3.7 Multistage Validation: Random Path Model

This is the fifth experiment in the optimization and validation flow of interconnects in this work. In addition to the design-level studies, we verify the performance of our interconnects on random paths in a routed design. We perform this study on two benchmark designs, *USB* and *DMA*.

3.7.1 Methodology

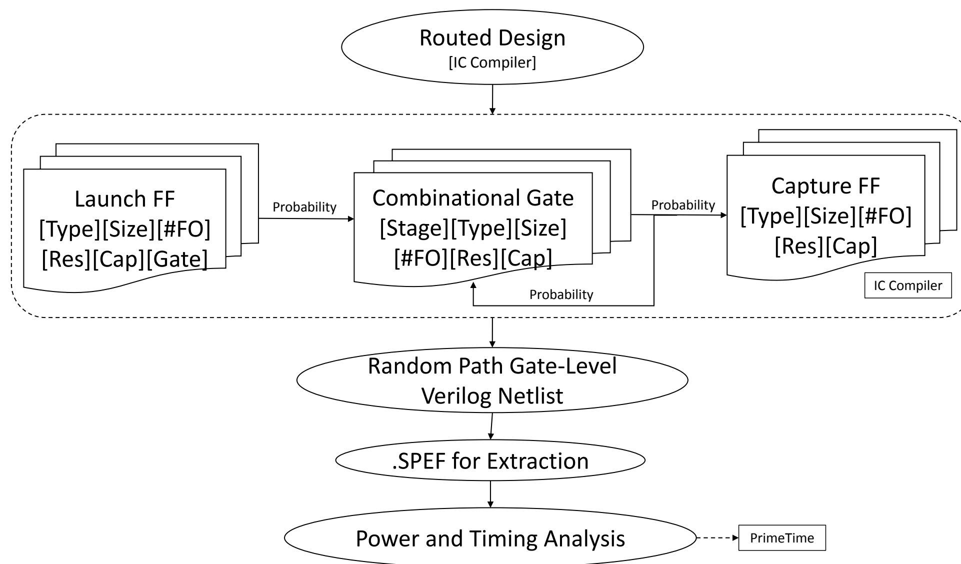


Figure 3.16: Random path flow used for performance comparison.

We illustrate the experimental flow in Figure 3.16.

1. **Random path extraction.**

   We first extract random paths from a routed design, as shown in Algorithm 1. An example of an extracted random path is shown in Figure 3.17.

2. **Random path generation.**

   We then adopt the Markov chain model and statistically generate random paths between any two flip-flops in the design. Algorithm 2 explains this flow. An example of two generated random paths is also shown in Figure 3.18.

Algorithm 1 Extraction of Random paths.

```
Input: Routed design
Output: Extracted path; Path_Ext
for Path counter do
  Generate a random number
  for Each random number do
    Extract pin information for the number
    Get associated pin list
    for Each pin in the pin list do
      List all available flip flops; FF_max
      for FF_Rand do
        List all fanouts FO_max
        Select 1 random fanout FO
        if FO is a FF then
          Break
        end if
        Get associated output pin
        Append to new pin list
      end for
    end for
    for Each pin in the new pin list do
      Get timing paths within a desired slack range
    end for
    for Path do
      Get pin type
      if Pin direction is out then
        Open writable file
        Extract cell name
        Extract cell type
        Extract cell size
        Find all connected cells
        Find capacitance of net
        Find resistance of net
        Get next pin
        Get next object
        Get layer information
        Increment number of stage
        Write collected information in file
      end if
    end for
  end for
end for
```

```
Stage InstName CellType Size #Fanout Cap Res
--------------Path--------------
slack : 0.315099
1 i_rx_phy_dpll_state_reg_1__u0 C12T32_LRHF_SDFPQX4 4 2 0.001938 0.121685
2 U528 C12T32_LR_CB4I1X8 8 1 0.000825 0.049290
3 i_rx_phy_dpll_state_reg_1__u0 C12T32_LRHF_SDFPQX4 4
--------------Path--------------
slack : 0.274203
1 i_rx_phy_fs_state_reg_0__u0 C12T32_LRHF_SDFPQX4 4 2 0.001772 0.156647
2 U442 C12T32_LR_OAI21X5 5 3 0.002849 0.207613
3 U380 C12T32_LR_AOI12X6 6 1 0.001381 0.135132
4 U543 C12T32_LR_OAI22X5 5 1 0.001441 0.106292
5 i_rx_phy_fs_state_reg_2__u0 C12T32_LRHF_SDFPQX4 4
```

Figure 3.17: Random extracted paths.

```
--------------Path--------------
1 C12T32_LRHF_SDFPQX4D4 4 2 0.003 0.192
2 C12T32_LRHF_SDFPQX4D4 4
--------------Path--------------
1 C12T32_LRHF_SDFPQX4D4 4 5 0.004 0.347
2 C12T32_LR_NAND2X3D3 3 2 0.003 0.224
3 C12T32_LR_NOR3AX6D6 6 1 0.001 0.068
4 C12T32_LRHF_SDFPQX4D4 4
```

Figure 3.18: Random generated paths.

Algorithm 2 Generation of random paths.

```
Input: Extracted paths; Path_Ext
Output: Generated paths; Path_Gen
for Each path in Path_Ext do
    Get instance name
    Get associated cell name
    Extract cell type; CurrentType
    Extract cell size; CurrentSize
    Extract cell fanout; CurrentFO
    Extract Cell capacitance; CurrentCap
    Extract Cell resistance; CurrentRes
    Extract Cell stage; CurrentStage
    Append all cell information to one variable; curDetails
    if First stage then
        Set first stage to curDetails
    else
        Set cur to curDetails
        Set prev to prevDetails
    end if
    if First occurrence of cur − prev then
        if First occurennce of prev then
            Set new variable
            Set occurrence of prev
            Append to a list sol
        end if
        Append cur to prev
        Set occurrence of cur − prev to 1
    else
        Increment count cur − prev
    end if
    Increment count of path
end for
for Each path do
    Probability of a cur − prev is occurrences/path
end for
```

Algorithm 3 Generation of individual netlist and .SPEF for analysis in *PrimeTime* [65].

Input: Generated paths; $Path_{Ext}$
Output: Generated Netlist and .SPEF; $NetList_{Path}$ $SPEF_{Path}$
Read $.lib$ file
Extract cell type
Initiate inputPinList to null
Initiate outputPinList to null
Check if cell is a flipflop
if Direction of cell is ”input” **then**
  Append pin to $inputPinList$
else
  Append pin to $outputPinList$
end if
Close $.lib$ file
for Each cell in $path$ **do**
  Get path from $Path_{Gen}$
  Data manipulation of $Path_{Gen}$
  Format to create individual netlist; $Path.v$
  Get pin capacitance information from $.lib$
  Format to create corresponding .SPEF; $Path.spef$
end for

3. **Netlist and .SPEF generation.**

   Since we obtain all the required information during the random-path extraction and generation stages, we can create the gate-level netlists and .SPEF files with little effort. The generated paths are manipulated using scripts to generate these files. Algorithm 3 explains the said flow.

```
module test ( in, clk, out );

input  in, clk;

output out;

C12T32_LR_SDFPQX8 U0 ( .D(in), .CP(clk), .Q(net_1) );

C12T32_LL_BFX8 U1 ( .A(net_1), .Z(net_2) );

C12T32_LR_SDFPQX8 U2 ( .D(net_2), .CP(clk), .Q(out) );

endmodule
```

Figure 3.19: Generated gate-level netlists.

   Examples of generated gate-level netlists and .SPEF file are shown in Figures 3.19 and 3.20 respectively.

4. We then perform power and timing analysis of these randomly generated netlists using *Synopsys PrimeTime* [65].

```
*D_NET net_1 0.001

*CONN

*I U0:Q O

*I U1:A I *L 0.001

*CAP

1 U1:A 0.001

*RES

1 U0:Q U1:A 0.083

*END

*D_NET net_2 0.029

*CONN

*I U1:Z O

*I U2:D I *L 0.029

*CAP

1 U2:D 0.029

*RES

1 U1:Z U2:D 0.132

*END
```

Figure 3.20: Generated SPEF files.

3.7.2 Results

In this section, we measure and compare the power and timing analysis results obtained from the *random path model* (RPM) validation approach.

Table 3.6: Power and timing analysis results (normal and critical paths) for the random path model.

Interconnect Parameters		ITRS		OURS		ITRS		OURS	
Local pitch (nm)		52							
AR		2.1		1.9		2.1		1.9	
DC		0.5		0.55		0.5		0.55	
Paths		Normal Paths				Critical Paths			
$\#Fins$		2	3	2	3	2	3	2	3
USB	Clock (ns)	0.7							
	# Paths	1000	1000	1000	1000	1000	1000	1000	1000
	Dyn. Power (mW)	8.80E-03	1.23E-02	8.85E-03	1.23E-02	9.25E-03	1.29E-02	9.22E-03	1.30E-02
	Avg. Dyn Power/Inst (mW)	8.80E-06	1.23E-05	8.85E-06	1.23E-05	9.25E-06	1.29E-05	9.22E-06	1.30E-05
	Leak. Power (mW)	2.65E-05	3.98E-05	2.90E-05	3.76E-05	5.56E-05	7.18E-05	6.51E-05	6.89E-05
	Avg. Leak. Power/Inst (mW)	2.65E-08	3.98E-08	2.90E-08	3.76E-08	5.56E-08	7.18E-08	6.51E-08	6.89E-08
	Slack (ns)	3.25E+02	3.13E+02	3.26E+02	3.14E+02	2.63E+02	2.54E+02	2.75E+02	2.61E+02
	Avg. Slack/Inst (ns)	3.25E-01	3.13E-01	3.26E-01	3.14E-01	2.63E-01	2.54E-01	2.75E-01	2.61E-01
DMA	Clock (ns)	0.8							
	# Paths	1000	1000	1000	1000	1000	1000	1000	1000
	Dyn. Power (mW)	8.80E-03	1.20E-02	8.90E-03	1.22E-02	8.80E-03	1.20E-02	9.00E-03	1.21E-02
	Avg. Dyn Power/Inst (mW)	8.80E-06	1.20E-05	8.90E-06	1.22E-05	8.80E-06	1.20E-05	9.00E-06	1.21E-05
	Leak. Power (mW)	2.80E-05	2.63E-05	2.61E-05	2.91E-05	3.25E-05	2.62E-05	3.53E-05	2.89E-05
	Avg. Leak. Power/Inst (mW)	2.80E-08	2.63E-08	2.61E-08	2.91E-08	3.25E-08	2.62E-08	3.53E-08	2.89E-08
	Slack (ns)	3.16E+02	3.05E+02	3.26E+02	3.08E+02	3.21E+02	3.37E+02	3.05E+02	3.08E+02
	Avg. Slack/Inst (ns)	3.16E-01	3.05E-01	3.26E-01	3.08E-01	3.21E-01	3.37E-01	3.05E-01	3.08E-01

From Table 3.6 we see that the power and timing results are comparable to that of the ITRS for normal and random critical paths. We also observe performance variation for two different library scaling factors (for $\#Fins = 2$ and $\#Fins = 3$).

Chapter 4

Conclusion and Future Work

In this chapter we review the results of our validation studies. We also discuss an optimized interconnect dimension range for the advanced technologies.

4.1 Conclusion

Our studies throughout this work are aimed at minimizing power while achieving least possible delay. In our first experiment where we design the optimal interconnect structure, we attempt to maximize the ΔEDP as shown in Section 3.4.1. Our studies from Section 3.3.2 indicate that in contradiction to the current interconnect predictions by the ITRS, copper interconnects with lower AR and higher DC will have better *EDP* for IDM $14nm$ node onwards.

Table 4.1: Suggested optimal dimension range.

Node	$45nm$	$28nm$	$22nm$	$14nm$	$10nm$	$7nm$
Metal-1 Pitch (nm)	150	120	80	52	42	36
T_{def} (nm)	135	114	80	55	46	40
AR	2.25	$2.375 \sim 2$	$2.35 \sim 2.10$	$1.9 \sim 1.82$	$1.68 \sim 1.51$	$1.71 \sim 1.53$
DC	0.4	$0.4 \sim 0.475$	$4.25 \sim 0.475$	$0.55 \sim 0.575$	$0.65 \sim 0.725$	$0.65 \sim 0.725$

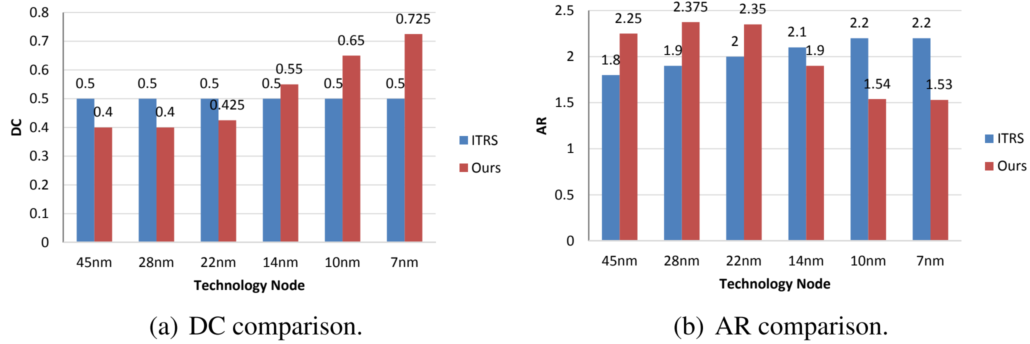
We understand and analyze the behavior of *BEOL* interconnect stack in various operating conditions. Since specific values of bulk resistivity, dielectric permittivity and other material/technology parameters will affect optimal interconnect dimensions, Table

4.1 suggests ranges of AR and DC values which offer EDP improvements over ITRS roadmap values, while also enabling manufacturable wire CD into the IDM 5nm node. Specifically, when DC is 0.65 at the 5nm node (21nm local metal pitch), it implies a wire CD of 13.65nm. On the other hand, when we consider a DC of 0.5 (as suggested by ITRS), the wire CD has to be 10.5nm, which makes manufacturing comparatively difficult. These conclusions are based on our initial studies of interconnect scaling trend in Section 3.3.2.

We observe the optimality of the suggested range of interconnect dimensions in Figure 3.8, where the values corresponding to the *dark blue* region denote the best performance in *EDP*, *delay* and *EDP with process variation*. Our studies also include tests for robustness of the determined interconnect dimensions in the presence of high-capacity drivers from the *Synopsys 32/28nm Generic Library* [52], FinFETs, voltage scaling, library scaling, etc. We know that AR and DC are selected based on reliability, cost, process integration, and a number of other considerations which have not yet been addressed by this work. This being said, our obtained values are shown to be robust to manufacturing variability (litho-etch and CMP variability), driven wirelength, slew time upper bound (Sections 3.4.1 and 3.3.2) and supply voltage (Section 3.4.2).

Despite having design-level power and timing performance which is either similar or slightly degraded as compared to interconnects with ITRS dimensions, we understand that interconnects with low AR and high DC can perform better. If wires are closer than half-pitch (DC $>$ 0.5), it can result in increased crosstalk. However, when combined with wires of low AR (shorter wires), the interaction between neighboring wires is reduced. This helps balance signal integrity and switching power. In addition, wider interconnects have low susceptibility to EM as compared to thinner interconnects because of the lowered current density. Thus, we believe that our suggested interconnect dimension scaling trend is optimal.

An important part of our experiment is to validate perform multistage power and timing analysis without design dependency. When we use the random path model in Section 3.7 to extract timing and power information, we see that the performance of our interconnects is either comparable slightly degraded as compared to the ITRS values. We can conclude that interconnects with our determined dimensions perform



(a) DC comparison.

(b) AR comparison.

Figure 4.1: Comparison of AR and DC between the ITRS predicted interconnect dimensions and our obtained interconnect dimensions.

better than those predicted by the ITRS, owing to better resilience towards manufacturing variablility and lower potential for EM. The comparison between the ITRS predicted interconnect dimensions and our optimized interconnect dimensions for AR and DC can be seen in Figure 4.1.

4.2 Future Work

Our ongoing work includes system-level validation for the $10nm$ and $7nm$ technology nodes. In addition, we also intend to measure the impact of the suggested interconnect dimensions on designs which have devices with higher number of fins. Currently our work comprises of determining only local optimal layers. We intend to explore optimal dimensions for intermediate and global layers. Another important aspect that needs to be explored in the future is the impact of shrinking *BEOL* interconnect dimensions on reliability issues such as EM and TDDB. This will help gauge an overall perspective and a more precise optimization trend.

Bibliography

[1] W. C. Elmore, "The Transient Response of Damped Linear Networks with Particular Regard to Wideband Amplifiers", *Journal of Applied Physics* 19(1) (1948), pp. 55-63.

[2] B. S. Landman and R. L. Russo, "On a Pin Versus Block Relationship for Partitions of Logic Graphs", *IEEE Trans. Computers* 100(12) (1971), pp. 1469-1479.

[3] P. Christie and D. Stroobandt, "The Interpretation and Application of Rent's Rule", *IEEE Trans. VLSI* 8(6) (2000), pp. 639-648.

[4] D. Pamanuwa and H. Tenhunen, "Repeater Insertion to Minimise Delay in Coupled Interconnects", *Proc. VLSI Design*, 2001, pp. 513-517.

[5] M. Anand, H. Shibata and M. Kakumu, "Multiobjective Optimization of VLSI Interconnect Parameters", *IEEE Trans. CAD* 17(12) (1998), pp. 1252-1261.

[6] M. Anand, H. Shibata and M. Kakumu, "Optimization Study of VLSI Interconnect Parameters", *IEEE Trans. Electron Devices* 47(1) (2000), pp. 178-186.

[7] J. Ao, S. Dong, S. Chen and S. Goto, "Delay-Driven Layer Assignment in Global Routing under Multi-tier Interconnect Structure", *Proc. ISPD*, 2013, pp. 101-107.

[8] Y. Cao, T. Sato, M. Orshansky, D. Sylvester and C. Hu, "New Paradigm of Predictive MOSFET and Interconnect Modeling for Early Circuit Simulation", *Proc. CICC*, 2000, pp. 201-204.

[9] J. Cobb, R. Garg and S. P. Khatri, "A Methodology for Interconnect Dimension Determination", *Proc. ISPD*, 2007, pp. 189-195.

[10] J. A. Davis, V. K. De and J. D. Meindl, "A Stochastic Wire-Length Distribution for Gigascale Integration (GSI) - Part II: Applications to Clock Frequency, Power Dissipation and Chip Size Estimation", *IEEE Trans. Electron Devices* 45(3) (1998), pp. 590-597.

[11] P. Gupta, A. B. Kahng, Y. Kim and D. Sylvester, "Investigation of Performance Metrics for Interconnect Stack Architectures", *Proc. SLIP*, 2004, pp. 23-29.

[12] H. B. Bakoglu and J. D. Meindl, "Optimal Interconnection Circuits for VLSI", *IEEE Trans. Electron Devices* 32(5) (1985), pp. 903-909.

[13] P. Dasgupta, A. B. Kahng and S. Muddu, "A Novel Metric for Interconnect Architecture Performance", *IEEE Design, Automation and Test in Europe Conf. and Exhibition*, March 2003, pp. 448-453.

[14] R. Ho, K. W. Mai and M. A. Horowitz, "The Future of Wires", *Proc. IEEE*, 2001, pp. 490-504.

[15] A. B. Kahng and D. Stroobandt, "Wiring Layer Assignments with Consistent Stage Delays", *Proc. ISPD*, 2000, pp. 115-122.

[16] X. C. Li, J. F. Mao, H. F. Huang and Y. Liu, "Global Interconnect Width and Spacing Optimization for Latency, Bandwidth and Power Dissipation", *IEEE Trans. Electron Devices* 52(10) (2005), pp. 2272-2279.

[17] Z. Li, C. J. Alpert, S. Hu, T. Muhmud, S. T. Quay and P. G. Villarrubia, "Fast Interconnect Synthesis with Layer Assignment", *Proc. ISPD*, 2008, pp. 71-77.

[18] N. Menezes, S. Pullela and L. T. Pileggi, "Simultaneous Gate and Interconnect Sizing for Circuit-Level Delay Optimization", *Proc. DAC*, 1995, pp. 690-695.

[19] M. L. Mui, K. Banerjee and A. Mehrotra, "A Global Interconnect Optimization Scheme for Nanometer Scale VLSI with Implications for Latency, Bandwidth and Power Dissipation", *IEEE Trans. Electron Devices* 51(2) (2004), pp. 195-203.

[20] J. H. C. Chen, T. E. Standaert, E. Alptekin, T. A. Spooner and V. Paruchuri, "Interconnect Performance and Scaling Strategy at $7nm$ Node", *Proc. IITC/AMC*, 2014, pp. 93-96.

[21] I. M. Elfadel, M. B. Anand, A. Deutsch, O. Adekanmbi, M. Angyal, H. Smith, B. Rubin and G. Kopcsay, "AQUAIA: A CAD Tool for On-Chip Interconnect Modeling, Analysis, and Optimization", *Proc. EPEP*, 2002, pp. 337-340.

[22] Y. Cao, C. Hu, X. Huang, A. B. Kahng, I. L. Markov, M. Oliver, D. Stroobandt and D. Sylvester, "Improved a Priori Interconnect Predictions and Technology Extrapolation in the GTX System", *IEEE Trans. VLSI Syst.* 11(1) (2003), pp. 3-14.

[23] A. Rahman and R. Reif, "System-Level Performance Evaluation of Three-Dimensional Integrated Circuits", *IEEE Trans. VLSI Syst.* 8(6) (2000), pp. 671-678.

[24] L. Scheffer and E. Nequist, "Why Interconnect Prediction Doesn't Work", *Proc. SLIP*, 2000, pp. 139-144.

[25] Y. I. Ismail, E. G. Friedman and J. L. Neves, “Equivalent Elmore Delay for RLC Trees”, *IEEE Trans. CAD* 19(1) (2000), pp. 83-97.

[26] S. Takahashi, M. Edahiro and Y. Hayashi, “Interconnect Design Strategy: Structures, Repeaters and Materials with Strategic System Performance Analysis (S^2 PAL) Model”, *IEEE Trans. Electron Devices* 48(2) (2001), pp. 239-251.

[27] R. Venkatesan, J. A. Davis, K. A. Bowman and J. D. Meindl, “Optimal N-Tier Multilevel Interconnect Architectures for Gigascale Integration (GSI)”, *IEEE Trans. VLSI Syst.* 9(6) (2001), pp. 899-912.

[28] Q. Zhang, J. J. Liou, J. McMacken, J. Thomson and P. Layman, “Development of Robust Interconnect Model Based on Design of Experiments and Multiobjective Optimization”, *IEEE Trans. Electron Devices* 48(9) (2001), pp. 1885-1891.

[29] Z. Zhu, D. Wan and Y. Yang, “An Interconnect-Line-Size Optimization Model Considering Scattering Effect”, *IEEE Electron Device Lett.* 31(7) (2010), pp. 641-643.

[30] W-T. J. Chan, A. B. Kahng and S. Nath, “Methodology for Electromigration Signoff in the Presence of Adaptive Voltage Scaling”, *Proc. SLIP*, 2014, pp. 1-7.

[31] S. Natarajan, M. Agostnelli, S. Akbar, M. Bost, A. Bowonder, V. Chikamane, S. Chouksey, A. Dasgupta and others, “A $14nm$ Logic Technology Featuring 2nd Generation FinFET Transistor, Air-Gapped Interconnects, Self-Aligned Double Patterning and a $0.0588\mu m^2$ SRAM Cell Size”, *Proc. IEDM*, 2014, pp. 3-7.

[32] Dr T. B. Chan, *Personal communication*, May 2015.

[33] Dr Gargini, *Personal communication*, May 2015.

[34] M. Stucchi, IMEC, *Personal communication*, October 2012.

[35] M. Badaroglu, K. Ng, M. Salmani, S. Kim, G. Klimeck, C. P. Chang, C. Cheung and Y. Fukuzaki, “More Moore Landscape for System Readiness-Itrs2.0 Requirements”, *Proc. ICCD*, 2014, pp. 147-152.

[36] C. M. Fiduccia and R. M. Matthesyses, “A Linear-Time Heuristic for Improving Network Partitions”, *Proc. DAC*, 1982, pp. 175-181.

[37] M. Martins, J. Matos, R. Ribas, A. Reis, G. Schlinker, L. Rech and J. Michelsen, “Open Cell Library in $15nm$ FreePDK Technology”, *Proc. ISPD*, 2015, pp. 171-178.

[38] S. C. Song, J. Xu, N. N. Mojumder, K. Rim, D. Yang, J. Bao, J. Zhu, J. Wang, M. Badaroglu, V. Machkaoutsan, P. Narayanasetti, B. Bucki, J. Fischer and G. Yeap, “Holistic Technology Optimzation and Key Enablers”, *Proc. VLSIT*, 2015.

[39] M. Neisser, SEMATECH, *Personal communication*, September 2014.

[40] *International Technology Roadmap for Semiconductors*, *http://www.itrs.net/*

[41] *International Technology Roadmap for Semiconductors*, Lithography Table 2011. *http://www.itrs.net/*

[42] K. Schuegraf, M. C. Abraham, A. Brand, M. Naik and R. Thakur, "Semiconductor Logic Technology Innovation to Achieve Sub-10nm Manufacturing", *IEEE Trans. Electron Devices Society* 1(3) (2013), pp. 66-75.

[43] T. B. Chan and A. B. Kahng, "Post-Routing Back-End-Of-Line Layout Optimization for Improved Time-Dependent Dielectric Breakdown Reliability", *Proc. SPIE Advanced Lithography*, 2013, pp. 86840L-86840L.

[44] F. Xia, J. He, P. Prabhumirashi, A. Lowrie, J. Hicks, Y. Shusterman and R. Brain, "Characterization and Challenge of TDDB Reliability in Cu/Low K Dielectric Interconnect", *Proc. IRPS*, 2011, pp. 2C.1.1-2C.1.4.

[45] A. B. Kahng and K. Samadi, "CMP Fill Synthesis: A Survey of Recent Studies", *IEEE Trans. CAD* 27(1) (2008), pp. 3-19.

[46] *Predictive Technology Model (PTM)*. *http://ptm.asu.edu/*

[47] S. Sinha, B. Cline, G. Yeric, V. Chandra and Y. Cao, "Design Benchmarking to 7nm with FinFET Predictive Technology Models", *Proc. ISLPED*, 2012, pp. 15-20.

[48] A. Balijepalli, S. Sinha and Y. Cao, "Compact Modeling of Carbon Nanotube Transistor for Early Stage Process-design Exploration", *Proc. ISLPED*, 2007, pp. 2-7.

[49] W. Zhao and Y. Cao, "New Generation of Predictive Technology Model for Sub-45nm Early Design Exploration", *IEEE Trans. Electron Devices* 53(11) (2006), pp. 2816-2823.

[50] Semiwiki Article, *https://www.semiwiki.com/forum/content/3884-who-will-lead-10nm.html*

[51] Synopsys Raphael-2D 3D Field Solver, *https://www.synopsys.com/Tools/TCAD/InterconnectSimulation/Pages/Raphael.aspx*

[52] Synopsys 32/28nm Generic Library, *http://www.synopsys.com/COMMUNITY/UNIVERSITYPROGRAM/Pages/32- 28nm-generic-library.aspx*

[53] Intel 14nm Technolog, *http://www.intel.com/*

[54] Intel 14nm vs 22nm, *http://www.anandtech.com/show/8367/intels-14nm-technology-in-detail*

[55] Intel 22nm FinFETs, *http://download.intel.com/newsroom/kits/22nm/pdfs/22nm-details_presentation.pdf*

[56] *International Technology Roadmap for Semiconductors*, Interconnect Chapter 2013, *http://www.itrs.net/*

[57] *International Technology Roadmap for Semiconductors*, PIDS Tables 2013. *http://www.itrs.net/*

[58] *International Technology Roadmap for Semiconductors*, MPU Tables 2013. *http://www.itrs.net/*

[59] *International Technology Roadmap for Semiconductors*, FOCUS C: PIDS Tables 2013. *http://www.itrs.net/*

[60] Intel 14nm data, *http://www.intel.com/content/dam/www/public/us/en /documents/presentation/advancing-moores-law-in-2014-presentation.pdf*

[61] Synopsys Design Compiler, *http://www.synopsys.com/Tools/Implement ation/RTLSynthesis/DesignCompiler/Pages/default.aspx*

[62] Synopsys HSPICE, *https://www.synopsys.com/tools/Verification/AMSVerification /CircuitSimulation/HSPICE/Pages/default.aspx*

[63] Synopsys IC Compiler, *http://www.synopsys.com/Tools/Implementation/Physical Implementation/Pages/ICCompiler.aspx*

[64] Synopsys Interconnect Technology Format, *http://www.synopsys.com/community/interoperability/pages/tapinitf.aspx*

[65] Synopsys PrimeTime, *http://www.synopsys.com/Tools/Implementation/SignOff /PrimeTime/Pages/default.aspx*

[66] Synopsys Raphael, *http://www.synopsys.com/Tools/TCAD/InterconnectSimulation /Pages/Raphael.aspx*

[67] Cadence SOC Encounter, *http://www.cadence.com/products/di/soc_encounter/pa ges/default.aspx*

[68] Synopsys FinFET Challenges, *https://www.synopsys.com/COMPANY/PUBLICA TIONS/SYNOPSYSINSIGHT/Pages/Art2-finfet-challenges-ip-IssQ3-12.aspx*

[69] “BSIM-MG”, *http://www-device.eecs.berkeley.edu/bsim/page=BSIMCMG_FAQ*

[70] M. Z. Hossain, M. A. Hossain, M. S. Islam, M. M. Rahman and M. H. Chowdhury, “Electrical Characteristics of Trigate Finfet”, *Global Journal of Researches in Engineering* 11(7) (2011).

[71] H. W. Cheng, C. H. Hwang and Y. Li, “Electrical Characteristics of Nanoscale Multi-Fin Field Effect Transistors with Different Fin Aspect Ratio”, *Proc. IEEE Nano*, 2009, pp. 606-612.