Statistical Modeling of the Effects of Process Variations on Silicon Photonics

by

Sally Ibrahim El-Henawy

B.Sc., Electrical Engineering, Ain shams University (2012)
M.Sc., Electrical engineering, Ain Shams University (2016)

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2022

Author .....
Department of Electrical Engineering and Computer Science
September 15, 2021

Certified by .....
Duane S. Boning
Clarence J. LeBel Professor
Electrical Engineering and Computer Science
Thesis Supervisor

Accepted by .....
Leslie A. Kolodziejski
Professor of Electrical Engineering and Computer Science
Chair, Department Committee on Graduate Students

Statistical Modeling of the Effects of Process Variations on Silicon Photonics

by

Sally Ibrahim El-Henawy

Submitted to the Department of Electrical Engineering and Computer Science
on September 15, 2021, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Electrical Engineering and Computer Science

Abstract

The rapidly growing field of silicon photonics is an attractive research and manufacturing platform, due to its ability to enable novel functionalities. Silicon photonics leverages existing CMOS processes and fabrication infrastructure, making its components suffer from the process variations present in CMOS technology. Long and repetitive simulations are required to understand the effect of these variations, largely due to the lack of variation-aware models.

This thesis explores methodologies for the development and application of process variation-aware compact models for silicon photonics components to enable photonics design for manufacturability. We consider the effect of a number of common unavoidable process variations, including both systematic and random variations, on the behavior of key optical building blocks. We examine the effect of line edge roughness as a random process variation on different components including Y-branches and coupled resonator optical waveguides. For the Y-branch, we use ensemble simulations to develop behavioral statistical models that can predict the behavior in the presence of different line edge roughness parameters. In the case of coupled resonator optical waveguides, to predict the behavior in the presence of different line edge roughness parameters, we develop an S-parameter based model that can be used directly in circuit simulation. Also, we present methods to develop S-parameter based compact models against systematic variations (geometric variations) in rings for both silicon and silicon nitride waveguides. The models are capable of predicting the behavior much faster than by full wave simulations, and give insight on resulting performance variation to enable yield prediction and optimization. We use the developed compact model to simulate photonic integrated circuits and compare the time required with the case of traditional simulations loops. We also present methods for extraction of spatial variations using variation test chip design and measurement. The spatial variations are decomposed into die-to-die and within-die variations.

We examine modulation (electrical and thermal) as a conventional approach to account for the effect of process variations. For electrical modulation, we study typical operating condition variations it can experience and find that their effect is not as

severe as typical process variations. Moreover, the power budget required to correct for process variations is calculated.

Together, these methods are key components toward design for manufacturability approaches and serve as a basis for extended PDKs for silicon photonics. Such models and methods help increase the speed of the simulation process required in photonics integrated circuit design, and inform designers of potential design modifications to correct for process variations for high yield and performance.

Thesis Supervisor: Duane S. Boning
Title: Clarence J. LeBel Professor
Electrical Engineering and Computer Science

Acknowledgments

Praise be to Allah first and foremost for giving me the strength and ability to complete this work. I would like to express my deep gratitude and appreciation for my advisor, Professor Duane Boning, for his encouragement, care, patience and guidance both technically and personally. The growth, in many aspects of life and research, I have achieved while working with him helped me shape with confidence my next steps. I could not reach what I have but for his unlimited and continuous support and intellectual conversations. I would also like to thank my committee members, Professor Lionel Kimerling and Professor Jelena Notaros, for their advice and comments on my research and this thesis. The experience wouldn't have been complete without the guidance and help of Professor Luca Daniel and my academic advisor Professor Jeffery Lang.

A big thank you to all the current and previous members of the great Boning group. I have learned from each and every person in the group: Germain, Daniel, Hongge, Fan-Keng, Damien, David, and my life saviors and collaborators Chris and Zhengxing; our discussions have always been interesting and fruitful.

I would like to send my thanks and appreciation to the people who helped me have a complete fruitful experience during my PhD. Ryan Miller, working with you especially in the first years helped me a lot and our discussions were always interesting and inspired me. Atashi Basu for her guidance and amazing management during my internship and her continuous mentorship to me. Dan Simon for the interesting and fruitful discussions we had and the continuous support and help afterward; your advice and input was always boosting my confidence.

Also, I am grateful to Prof. Amr Safwat and Prof. Hadia El-Hennawy; without your support I wouldn't have been here at MIT.

The journey of my PhD wouldn't have been possible without all the great friends I knew at MIT. Malik and Amira, my parents and family here; you have always made me feel a member in your family and hosted me always and forever. I feel comfortable and safe knowing that you have my back and that you are here for me, your advice

and input is among the few, if not the only, I trust blindly. Aya and Mahmoud, you never failed to cheer me up whenever needed and our fun moments always helped me. Shimaa, my friend and sister who helped me keep up my sanity during PhD; she always absorbs willingly my different personalities and moods, provides support and encouragement and pushes me in the right direction whenever needed. Mustafa, Mohamed, and all the great Egyptians I met here, you have been my good friends, provided encouragement and support and made my time here fun, which I will never forget.

Finally and most importantly, I would like to thank my great family, their unconditional love, support, confidence and encouragement has always kept me moving. Making my mother Nefissa and father Ibrahim proud and happy has always been my main motivation in order to pay them back a very small part of what they have been, and are still, doing for me. My sisters, Yasmine and Nancy, the fun night calls we have and keeping me always updated, you made it easier for me and helped me be less home sick. The great kids, Yassine, Hana and Zeina, your tiny small love paper notes and gifts you send, always cheered me up in my lowest days, also, the little fights you have, and the ones I have with you as well, and asking for my judgments, kept me attached and always made me feel alive and young. My late grandma Thoria, till now it is your encouragement and belief in me that gave me the confidence that I can do whatever I want!

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In our first chapter, we emphasis on the motivation and purpose of the work in this thesis. Section 1.1 introduces the novel applications and functions enabled by using silicon photonics. In Section 1.2, we present the challenges that the silicon photonics design platform may experience during fabrication. Section 1.3 highlights the process variations that we will examine in the thesis. In Section 1.4, we present the thesis objective and the approach that we use to address the challenges in silicon photonics. Finally, Section 1.5 shows the thesis flow and the main output from the coming chapters.

1.1 Motivation

Silicon photonics, where non-interacting photons (light) rather than electrons are used to transmit data, is the subject of a substantial active research effort in both academic and industrial settings. Many are looking forward to using silicon photonics to help overcome the bottleneck of Moore's law. In addition to providing a means to enable novel applications, silicon photonics also has exciting potential to revolutionize existing technologies [4, 5].

Among the applications and enhancements offered by silicon photonics are higher data rates, faster connections, and higher bandwidth communications, which can be achieved by using optical links and photonic integrated circuit (PIC) based transceivers

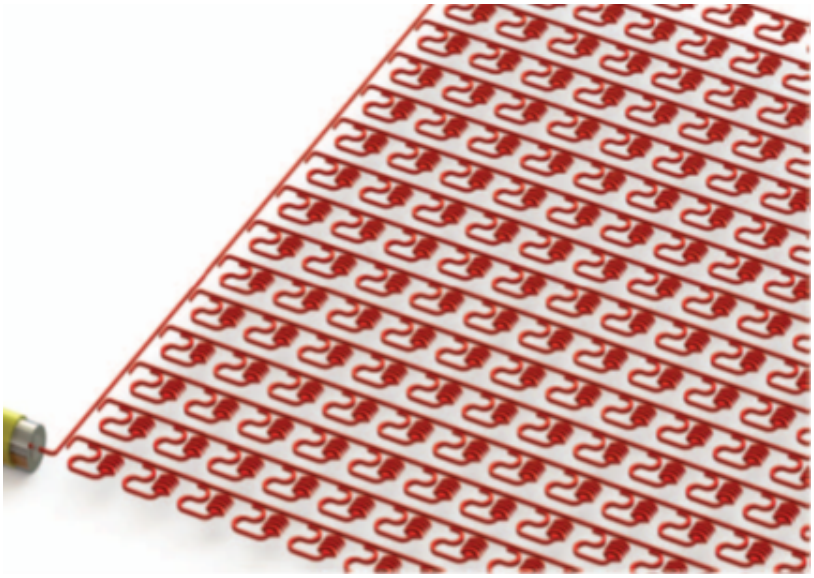


Figure 1-1: 64 × 64 nanophotonic phased array [1].

to replace conventional circuitry for both data center and on-chip communications [6]. Other interesting applications being developed that use silicon photonics technology are biomedical sensing and Lab-On-A-Chip devices, where the whole system, including the optical sensors and the CMOS processing unit, can be integrated in the same chip, providing a compact and fast sensing system with real time measurements [7, 8]. Another novel optically based application is wave front engineering, such as in the Nanophotonic Phased Array (NPA) shown in Fig. 1-1 as presented in [1], where a compact set of 64×64 NPA elements are integrated on a silicon chip, enabling beam steering and focusing. In addition is the promising use of silicon photonics in LIDAR [9, 10].

These novel applications demonstrate the interest in silicon photonics as a new design platform and draw attention to this enabling technology; but as a new technology, silicon photonics has a number of obstacles that need to be cleared first, in order to be widely adopted.

1.2 Challenges

In silicon photonics, both electrical and photonic components are integrated on the same chip. An important aspect and one of the most compelling features of silicon photonics, is its relatively seamless integration with existing Complementary Metal Oxide Semiconductor (CMOS) fabrication infrastructure such that it leverages existing CMOS fabrication processes and infrastructure, making it a cost-effective technology. That means, however, that the process variations present in CMOS technology will be inherently transferred to silicon photonic components.

In CMOS technology, the effect of process variations has been deeply studied and modeled. Moreover, process design kits (PDKs) that include compact models and variation-aware models, describing and predicting the effects of process variations, are readily available for designers to use. This facilitates the design process for CMOS circuits [11], reducing design time, and increasing the yield, as the expected range and margins in performance can be predicted and evaluated in simulation by the designer.

By contrast, when it comes to silicon photonics, we lack variation-aware compact models. Accordingly, the effect of these inherently transferred manufacturing variations on the performance of photonic components are still under examination, with limited number of studies and analyses reported [12, 13, 14, 15, 16, 17, 18, 19] trying to understood their effect. As a result, the design of a silicon photonics based system becomes a time consuming and slow process, particularly if variation and yield are to be taken into account. The designer has to go through the design steps starting from the component level and in each step, they have to check the device physics and performance, which even in the nominal case can require several iterations. Accounting for variations multiplies the design challenge, or may not even be possible if models of likely variations are not available.

So, a key challenge for the emerging silicon photonics industry is the lack of mature process, device, and circuit variation-aware models for the existing IC and photonic process steps. This lack is a road block that must be overcome for this technology to make it into mainstream manufacturing.

In particular, process-variation-aware compact models for silicon photonics is an increasingly important need in order to enhance yield and expedite the design process. Understanding and modeling the various process variations (both systematic and random) at the wafer, chip, and feature scales, and predicting their performance impact on both the device and circuit levels, is a key element of a robust DFM methodology for silicon photonics that would help designers compensate for process variations and help move silicon photonics from the research into high yield manufacturing.

1.3 Process Variations

Process variations result in the deviation of the intended device parameters due to limitations and imperfections during the fabrication process. In CMOS process technology, a large range of variations can occur during the fabrication process. These variations affect both device and circuit levels; the variations can occur at the wafer, chip, or feature scales, and can be systematic, random, or a combination of both.

1.3.1 Systematic Variations

Systematic variations are geometric or material properties that vary based on spatial device location within the die or wafer. These variations can be a change in silicon or other layer and structure width, thickness or partial etch thickness as seen in Fig. 1-2, where ΔW_{SE} is the variation in silicon width, ΔW_{RE} is the width variation in rib waveguide width, and ΔT_{RE} is the variation in silicon thickness for the rib waveguide partial etch. Such variations cause deviation from the intended device design parameters, resulting in skewed or undesirable performance. Systematic variations have a repeatable component that can in some cases be modeled as a function of chip or layout design parameters (e.g., layout pattern density) or spatial location (e.g., wafers with a known center to edge ring non-uniformity).

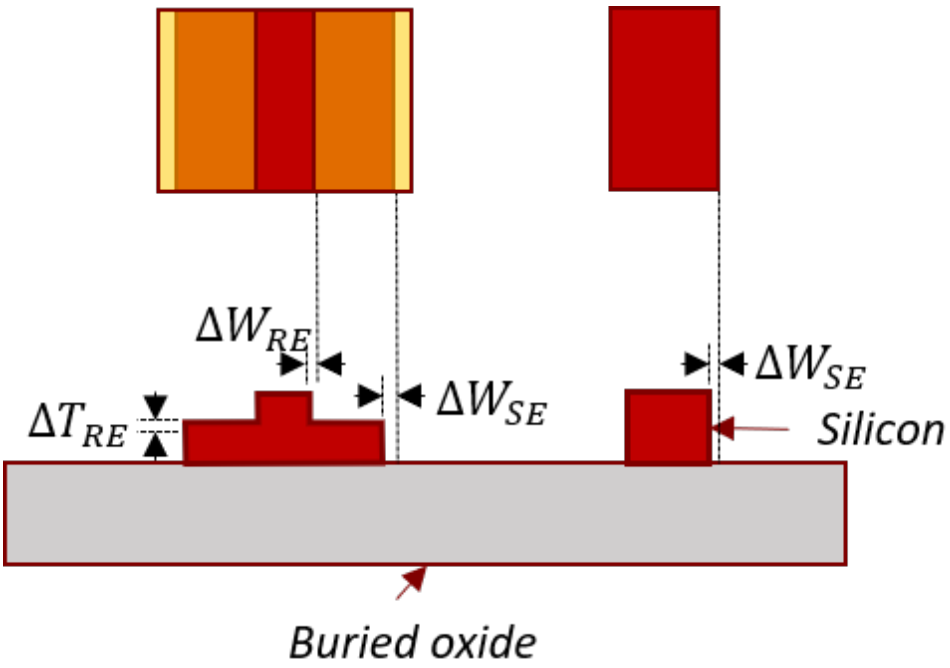


Figure 1-2: Potential systematic geometric variations that photonic components can experience.

1.3.2 Random Variations

A random process variation means that there is not a fixed offset or deviation in some parameter that is known based on location within wafer; rather, the parameter varies randomly from wafer-to-wafer, die-to-die, device-to-device, or even within a given component. Accordingly, these random variations are usually described with their statistics or a function that varies stochastically depending on location.

Line edge roughness (LER) [20] is an example of a random variation that occurs during patterning due to fluctuations in either tools, materials, processes or a combination of them. LER can be modelled by generating noise using a Fourier synthesis technique [20, 21], where LER deviations are commonly described by spectrum of the Gaussian auto-correlation function defined by the LER amplitude (A) and spatial correlation length (L_c). Figure 1-3 shows a waveguide experiencing LER on its side walls.

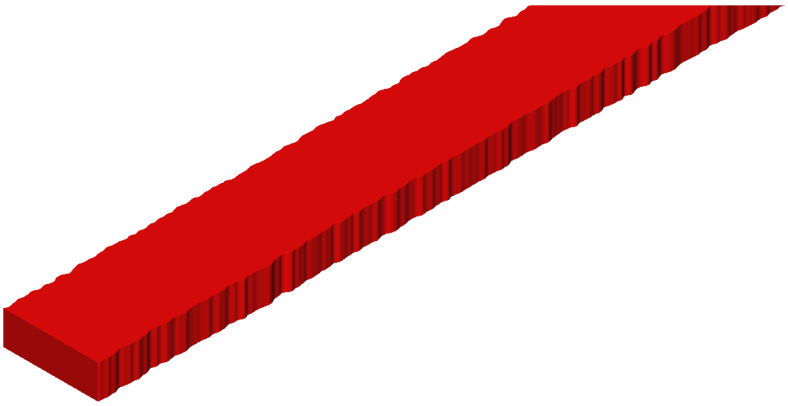


Figure 1-3: Line edge roughness applied to the side walls of a waveguide.

1.4 Objective

These inherently transferred process variations are unavoidable and affect the light-guiding capabilities of the optical components, which degrades the overall circuit performance, and decreases the yield. Process variation-aware compact models as part of variation-aware design methodology are needed to enable modeling, analyzing, and optimizing photonic devices and circuits. Such methods will help designers to predict performance and enable the design of both high yielding and high performing photonic integrated circuits (PICs).

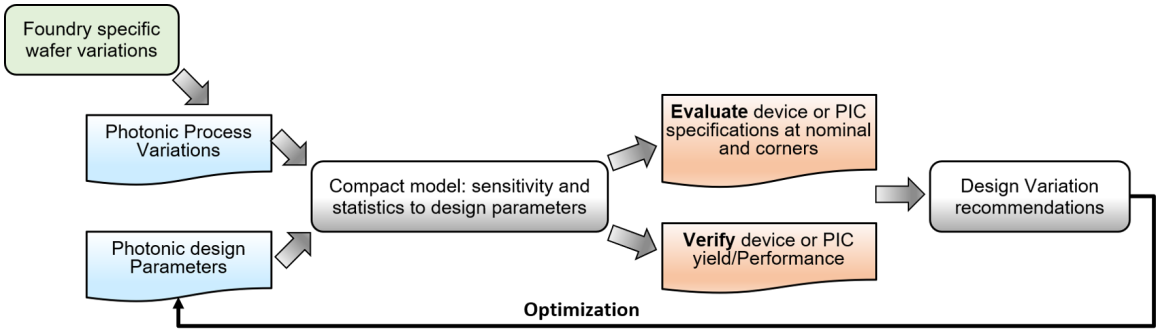


Figure 1-4: General flow of DFM for silicon photonics.

To deal with these variations, we propose developing compact models that describe an input-output relationship between process variations and the resulting performance. The flow of a design enabled by design for manufacturability (DFM) meth-

ods is envisioned in Fig. 1-4. Process variations, which are foundry based variations, along with component design parameters, are fed to the compact models which have information about sensitivity and statistics to design parameter variations. These compact models are capable of evaluating and predicting the device behavior, performance, and yield. Based on these results the designer can perform design explorations and consider modifications to optimize the behavior. A further goal is to do this faster than with traditional simulations in the absence of the compact model.

These compact models can be created by adopting the statistical modeling techniques used for CMOS design, which have proved to be very successful in the CMOS electronic integrated circuit industry, and apply these concepts to the modeling of silicon photonic components to develop compact models. In this work, we seek to develop and follow a methodology that will enable photonic DFM, as highlighted in Figure 1-5. For passive devices, we use FDTD [22] and MODE [23] simulations as the basis for compact models. For active photonic components, we also use CHARGE [24] to generate ranges of behavior for compact models.

1.5 Thesis Flow

In Chapter 2, we explore the effect of LER on Y-branch behavior, and seek to understand and provide a predictive model for the behavior across wavelength as a function of LER parameters. This serves as an example and demonstration of compact models for a key random process variation. Chapter 3 explores a larger component, coupled resonator optical waveguides (CROWs), where we model the effect of systematic spatial variations in thickness and width as well as the LER, and demonstrate the use of such models to predict yield and resonance frequency. Chapter 4 proposes methods for extraction of compact models developed for large silicon nitride based rings, enabling performance prediction and optimization. In Chapter 5, we design test structures and present analysis methods to extract wafer variations from measurements, and to provide design rule guidelines. Then in Chapter 6, we examine the modulator as an active optical device, and model the effect of process variations on

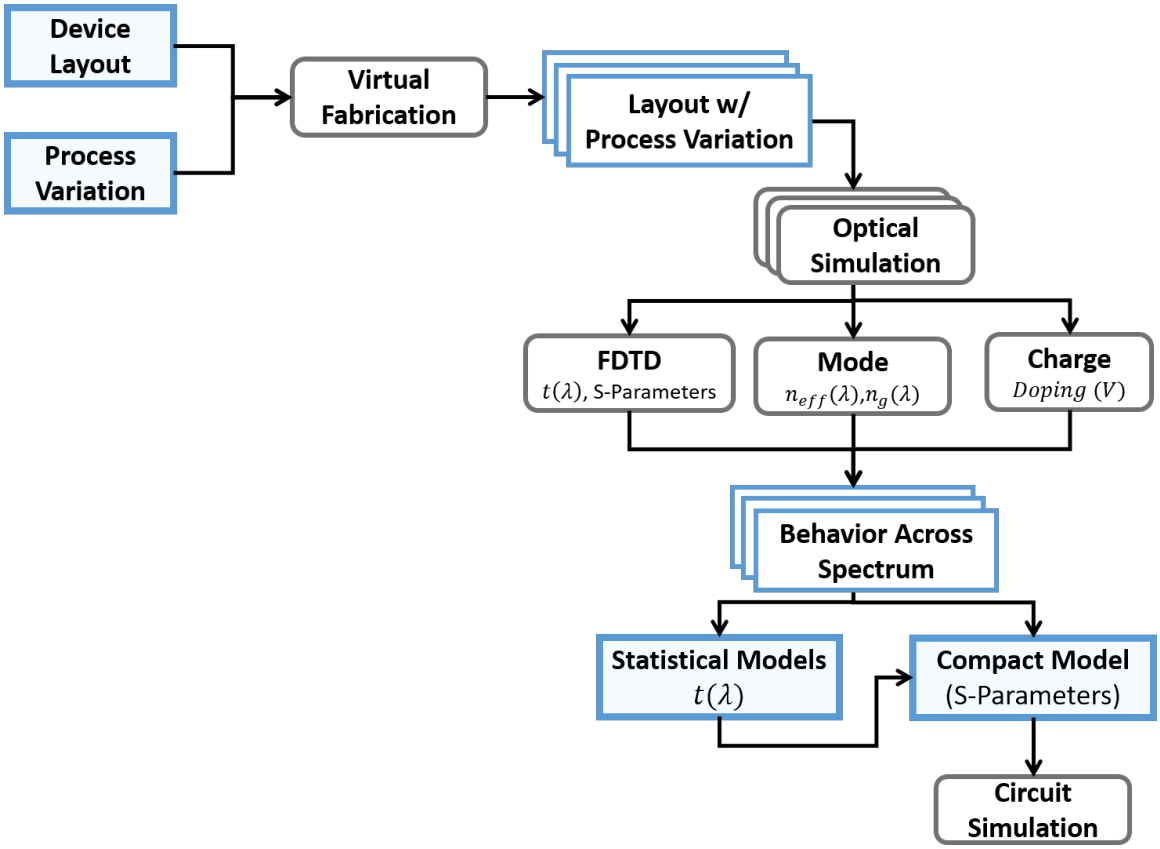


Figure 1-5: Simulation-based methodology for analysis and modeling of process variations in silicon photonics.

the active section, specifically, the effect of mask exposure and doping variations on modulator behavior. In Chapter 7, we combine models developed in earlier chapters together to demonstrate simulation of variation impact on a PIC system. Finally, conclusions and suggestions for future work are presented in Chapter 8

Chapter 2

Y-branch Statistical Modeling

Process variation is the deviation of device geometry from its intended design due to non-idealities in the fabrication process. One common source of process variation is line edge roughness (LER) [20], which occurs due to random fluctuations in the lithography tools, materials, and processes, as well as variations in the plasma etching process. Since LER affects the fabricated component geometry by introducing random perturbations to its sidewalls, as illustrated in Fig. 2-1, it can have a significant impact on the light-guiding abilities of waveguides, and can be the dominant contribution to propagation loss [25, 26].

These spatial perturbations can be viewed as random noise and are characterized by two parameters: the root mean square amplitude (A) and the correlation length (L_c). A is the standard deviation of the displacements from and normal to the smooth (no roughness) surface at each point along the surface. The L_c specifies the longitudinal spatial frequency content for the roughness along the surface; large correlation lengths will cut off high frequency components of the roughness, resulting in a "smoother" profile with fewer oscillations per unit length, and small correlation lengths will retain the higher-frequency parts of the roughness. This noise is described by Eq. 2.1:

$$N = A^2 e^{-(x/L_c)^{2\alpha}} \tag{2.1}$$

where x is the length along the structure, and α is either 0.5 for exponential noise

or 1 for Gaussian noise. In our study, α is always taken to be 1. This value of α is chosen because typically there are extra processing steps after lithography, such as oxidation, that tend to smooth the LER and reduce the high-frequency edges [27]. This leads to sidewall noise characteristics that are better described by a Gaussian correlation function, which suppresses high-frequency noise components more than the exponential correlation function, as shown in Fig. 2-2.

Figure 2-1: Line edge roughness imposed on a Y-branch and waveguide sidewalls.

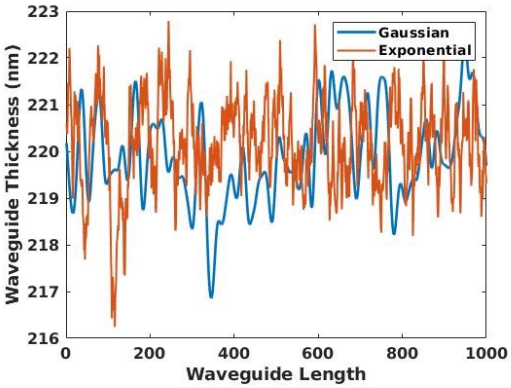


Figure 2-2: Example Gaussian and exponential LER with the same amplitude and correlation length imposed on a waveguide.

To understand the impact of LER on PIC components, it is crucial to be able to model these effects. Previous work has considered LER impact on loss in simple silicon waveguides [14]. In [28], we report the first study to consider the performance impact of LER on a more complicated fundamental photonic structure, a Y-branch [2], through virtual fabrication and optical simulation. The Y-branch serves as either a means to split one stream of light into two or combine two streams of light into one, which makes it a common element of a PIC (for example, two Y-branches can be used in a Mach-Zehnder interferometer) [5]. Three critical performance factors of the Y-branch are considered: transmission imbalance at the two output ports, excess loss, and fractional reflection back into the input port (back reflection).

In this chapter, we present a method for behavioral statistical modeling for the effect of line edge roughness (LER) on Y-branch optical splitters. The model predicts the performance as a function of the operating wavelength and the LER parameters,

which vary within the same process from one step to another or from one foundry to another.

In Section 2.1, the simulation methodology is described, with ensemble simulation results presented in Section 2.2. Section 2.3 describes the compact modeling approach used to capture these variations. Applications of these models for yield analysis is demonstrated in Section 2.4. Comparisons to other splitter geometries are presented in Section 2.5. Finally, Section 2.6 summarizes the chapter.

2.1 Simulation Details

We focus on a Y-branch silicon-on-insulator (SOI) waveguide splitter that was previously optimized to minimize back-scatter at 1550 nm [2]. The silicon waveguiding regions have widths of 500 nm, thicknesses of 220 nm, and the overall component length is 15 μm as shown in Fig. 2-3. The Y-branch is encased in SiO_2 cladding extending 1 μm below the bottom silicon surface, and 1 μm above the bottom silicon surface (2 μm in total). The fundamental TE mode is launched into the input port and the output power is measured at the input port (to measure back reflection) and at the two output ports, referred to as the "upper" and "lower" ports.

The frequency range is chosen to span a wavelength window around the common operating wavelength of $\lambda = 1550\ nm$ in the C-band; this window ranges from 1.5 μm to 1.6 μm into the S-band and L-band, providing insight into the common photonic operating regions. All process modelling (including line edge roughness) is performed in Coventor's SEMulator3D® [29] software. The structures with the LER imposed on the sidewalls are then exported from SEMulator3D via a surface mesh and imported into Lumerical FDTD [22] for optical simulation.

LER can be modelled by generating noise using a Fourier synthesis technique [21, 20], and applying that noise as geometric width perturbations to the sides of the structure. This technique generates roughness N that corresponds to the power spectrum of the Gaussian autocorrelation function defined by the LER root mean squared amplitude A and correlation length L_c, as defined in Eq. 2.1. Various combinations

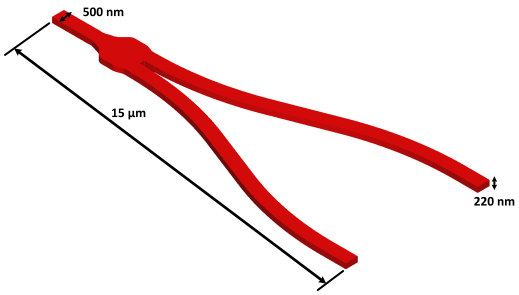


Figure 2-3: Geometry of the Y-branch used for analysis [2].

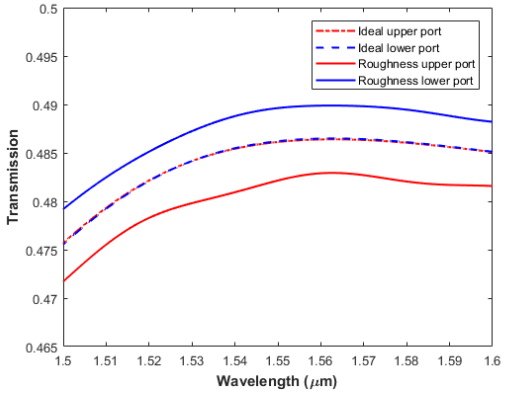


Figure 2-4: Y-branch transmission spectrum when no LER is applied (ideal case) and the imbalanced case when LER is applied.

of A and L_c values are simulated. Amplitudes between 1 nm and 15 nm and correlation lengths between 10 nm and 60 nm are chosen to span realistic observed values reported in previous works [21, 30, 20, 31]. For each A and L_c combination that is simulated, many structures (referred to as instantiations) are generated with different random seeds for the edge noise, in order to capture the statistical variation of the optical response.

The model development flow is shown in Fig. 2-5. In Stage 1, we start by defining the process steps with the LER added to the lithography step with prescribed A and L_c, then this process is used in Coventors's SEMulator3D ®. A 3D mesh representing the final structure (with LER added) is generated by SEMulator3D and exported for use in the optical simulations. In Stage 2, the 3D mesh is imported into FDTD for broadband optical simulations, where the normalized transmission power at each output port is measured for wavelengths in the range of $1500 - 1600$ nm, as shown in Fig. 2-6. To capture the statistical variation in the output as seen in Fig. 2-7, the previous two stages are repeated for 50 different instantiations (a Y-branch with a unique sidewall noise profile) for each amplitude and correlation length combination. This is repeated for 10 different combinations of A and L_c, leading to 500 Stage 1/Stage 2 iterations in total. We choose the 10 DOE values for A and L_c using Latin hypercube sampling for the ranges we care about (based on literature values) with

two extra experiments to consider the effect of changing the A and keeping the L_c constant and vice versa.

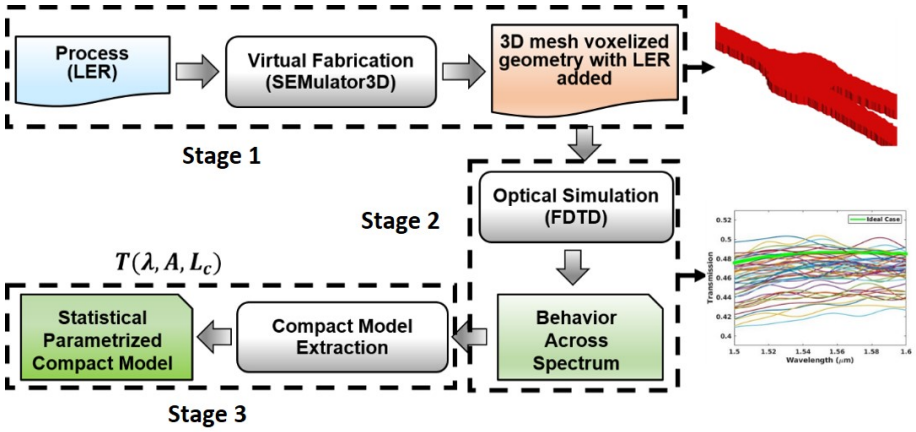


Figure 2-5: The model development flow.

2.2 Simulation Results

2.2.1 Transmission

With no LER applied, the ideal Y-branch transmits power equally between its two output ports, as seen in the simulation (dashed lines) in Fig. 2-4. In this ideal case, we also see the expected wavelength dependence of transmission from $\lambda = 1.5 - 1.6\ \mu m$, with maximum transmission of 0.4865 near 1550 nm, corresponding to excess loss (discussed further in Section 2.2.2) of about 3%. However, when LER is present, imbalanced transmission between the upper and lower ports is observed (solid lines in Fig. 2-4). The amount of the imbalance changes for different instantiations of the same LER parameters, where the power generally favors one port over the other. For the relatively small A and L_c values of 3 nm and 10 nm, respectively, the variations for each port are Gaussian as seen in Fig. 2-7(a), with a mean that is almost equal to the ideal transmission value of 0.4862 and with a standard deviation of 0.004. Therefore, the effect of small LER is a modest degree of statistical variation of the transmission response that fluctuates about the ideal (smooth) result.

In contrast, higher values of A and L_c can lead to more pronounced device degradation, where the mean for both upper and lower ports is shifted from the ideal device mean. For A of 10 nm and L_c of 40 nm, the results in Fig. 2-7(b) show a much larger spread in the upper and lower branch transmissions at $\lambda = 1550$ nm, with mean of upper port of 0.4727 and mean of 0.4783 for the lower port. In this case, the correlated transmissions are below the nominal, indicating substantial excess loss.

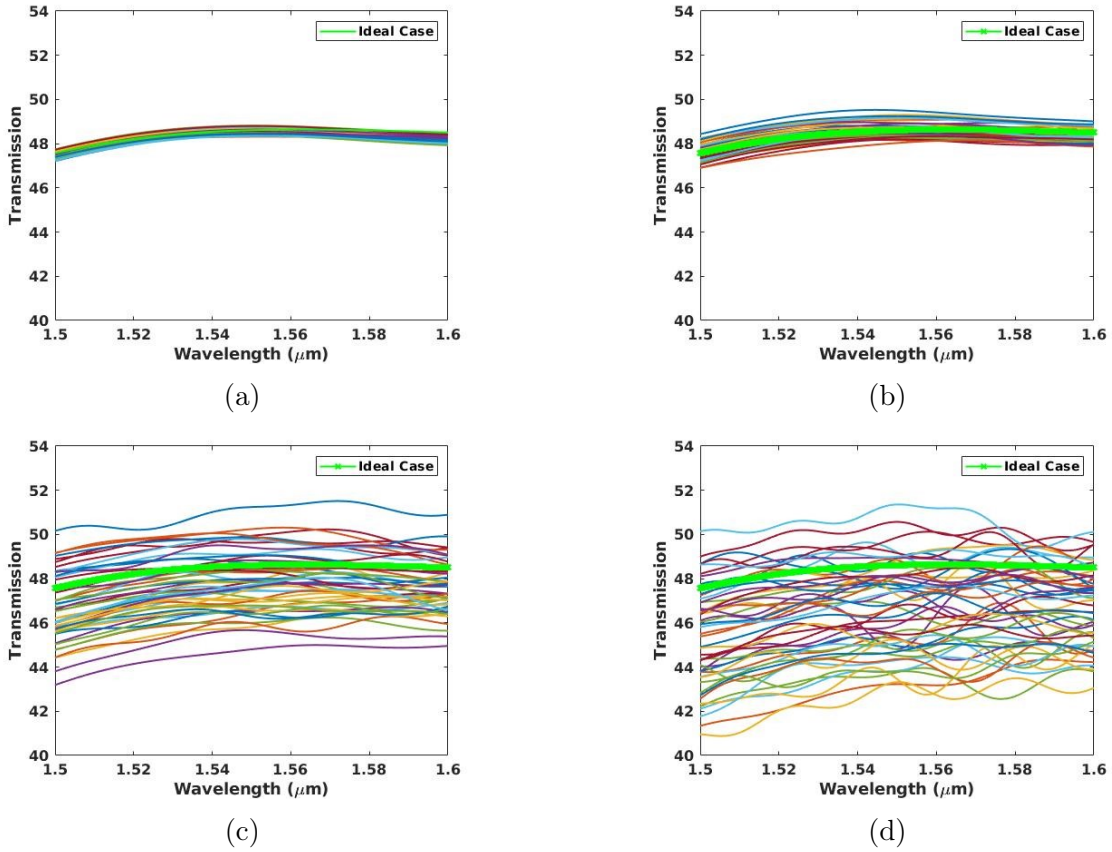
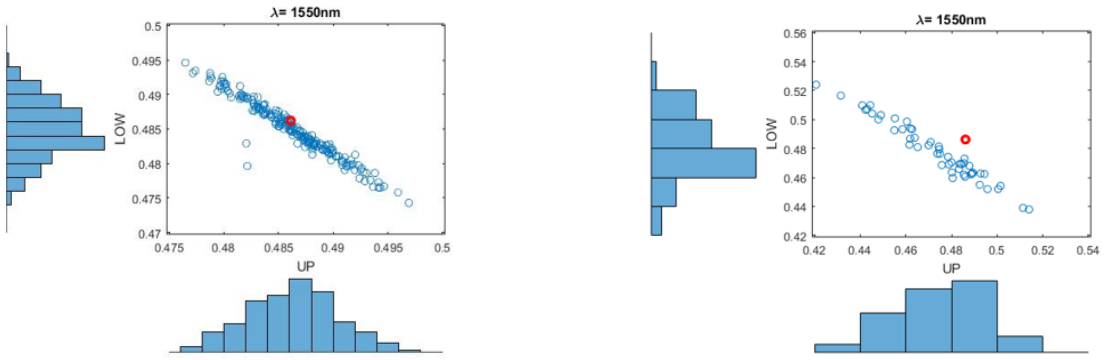


Figure 2-6: The upper port transmission power spectrum of 50 instantiations for different A and L_c combinations. (a) $A = 1$ nm and $L_c = 25$ nm, (b) $A = 3$ nm and $L_c = 20$ nm, (c) $A = 6$ nm and $L_c = 60$ nm, (d) $A = 12$ nm and $L_c = 50$ nm.

To further study the effect of A and L_c on imbalanced transmission at the two output ports, 50 instantiations for different A and L_c combinations are generated and analyzed. The resulting transmission for the upper ports at 1550 nm, shown in Fig. 2-6, indicates that as A or L_c increases, the deviation of the transmission value from the ideal (smooth) case increases, which in turn means that the imbalance between



(a) (b)

Figure 2-7: Joint distribution plot of the transmitted power for the upper and lower ports at $\lambda = 1550\ nm$. (a) 200 instantiations for $A = 3\ nm$ and $L_c = 10\ nm$; red point is nominal (no LER) case. (b) 50 instantiations for $A = 10\ nm$ and $L_c = 40\ nm$; red point is nominal (no LER) case.

the two output ports increases. The effect of A on this imbalance is larger than the effect of L_c, as shown in Table 2.1.

$A\ (nm)$	$L_c\ (nm)$	Imbalance Mean (%)	Imbalance Variance (%)	Maximum Imbalance (%)
1	25	-0.03	0.25	0.49
3	10	-0.16	0.45	1.89
3	20	-0.22	0.64	2.05
3	40	-0.30	0.9	2.44
6	60	0.22	2.71	6.58
7	30	0.05	2.33	4.99
10	10	0.78	1.86	7.95
10	40	1.35	3.85	9.99
12	50	-0.18	4.04	12.46
15	60	-0.37	5.65	16.84

Table 2.1: Mean, standard deviation and maximum values for imbalance in transmission between Y-branch upper and lower ports for 50 instantiations, for different amplitude and correlation length combinations at 1550 nm wavelength.

Figure 2-8 shows the relative imbalance between the transmitted power between the Y-branch two output ports as a function of the LER amplitude and correlation length, for ten different combinations of A and L_c. The maximum imbalance increases more along the amplitude axis than along the correlation length axis, ranging from $\sim$ 1% for the smallest case at $A = 1\ nm$ and $L_c = 25\ nm$ (left-most bubble), to 17%

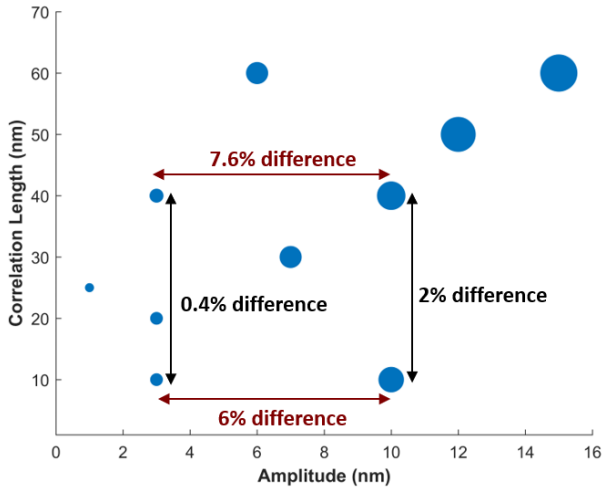


Figure 2-8: Relative imbalance in Y-branch transmission as a function of LER amplitude and correlation length at 1550 nm, the radius of the bubble indicates the magnitude of the imbalance.

for the largest case of $A = 15\ nm$ and $L_c = 60\ nm$ (right-most bubble). This agrees well with the results shown in Fig. 2-6 which clearly shows that increasing the A (moving from the top row to the bottom row) has a greater impact on the imbalance than increasing the L_c (moving from the left column to the right column).

While Fig. 2-8 and Table 2.1 report the maximum imbalance percentage between the two output ports across the 50 instantiations for a single wavelength ($\lambda = 1550\ nm$), Fig. 2-9 shows the average imbalance (averaged for all the 50 instantiations) across the whole wavelength range. The imbalance increases over all wavelengths sampled in both the increasing-A and increasing-L_c directions. At $\lambda = 1550\ nm$ specifically, the imbalance increases by ~4x [3.864] when A increases from 3 nm to 10 nm for $L_c = 10\ nm$, and by ~4x [4.2868] in the $L_c = 40\ nm$ case. For increasing L_c from 10 nm to 40 nm, the imbalance increases by ~1.5x [1.3689] for $A = 3\ nm$ and by ~1.5x [1.5186] in the $A = 10\ nm$ cases. Thus, we see that increasing the LER amplitude and correlation lengths both lead to increased imbalance in the output ports. These effects can be attributed to the distortion of the junction region where the two output branches split. The ideal device and mode source are fully symmetric, hence the equal transmission in the two output ports for the ideal case shown in Fig. 2-4 and in the electric field profile in Fig. 2-10(a). However, LER introduces

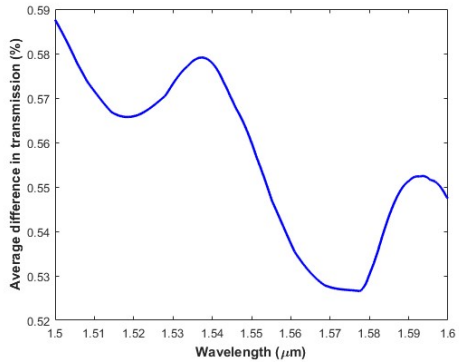
defects to the center region that break the symmetry of the device, causing more light to enter one branch than the other, as seen in the perturbed electric field profile in Fig. 2-10(b). As was mentioned in [2], for short values of L_c (as compared to the effective wavelength in the waveguide core, which is $\sim$440 nm here), the propagating mode does not interact with the rapidly-oscillating sidewall roughness significantly. However, as L_c increases, there are fewer spatial oscillations per unit length, and the mode will begin to interact with the sidewall perturbations more, until L_c becomes much larger than the wavelength and the mode effectively sees no oscillations at all. Increasing A has a more obvious effect on the junction-region distortion. Higher values of A introduce larger peaks and valleys on the device sidewall surfaces, which cause larger deviations from the ideal symmetric junction shape, and more power will inevitably end up in one output branch than the other.

2.2.2 Excess losses

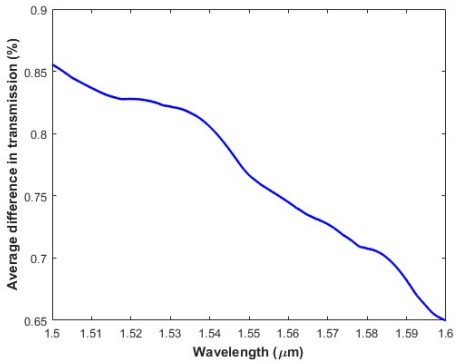
Figure 2-7(b) indicates that a significant amount of incident power is being lost for the large A and L_c case, as the mean power for both ports have deviated significantly from the nominal value, and the highly anticorrelated upper and lower port transmissions are shifted downward from the nominal balanced transmission case. Several studies have been reported to characterize and minimize loss due to LER on straight SOI waveguides [26, 31, 32]. The perturbations introduced by LER to the Y-branch sidewalls not only affect the power imbalance between the two outputs, but also contribute to increased propagation loss the light experiences as it travels down the Y-branch. Excess loss is defined here as ratio of the sum of the power in the two output ports relative to the power in the input port:

$$EL = \frac{P_{upper} + P_{lower}}{P_{input}} \tag{2.2}$$

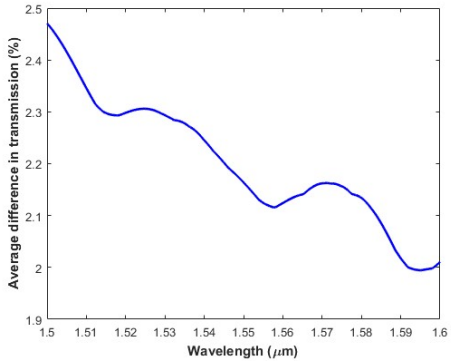
This serves as a metric to determine how much power is lost as the light travels the length of the structure (excess loss does not include the coupling loss from the source). Ideally, in the case of no losses, the EL ratio is 1 (0 dB). However, in real



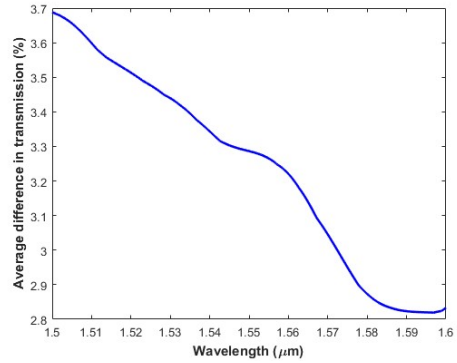
(a)



(b)

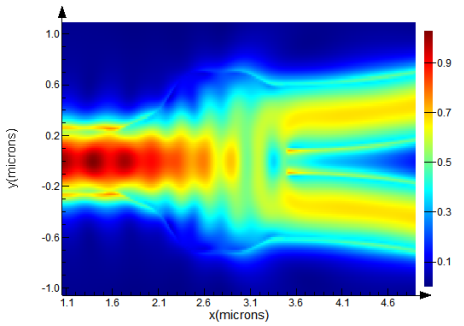


(c)

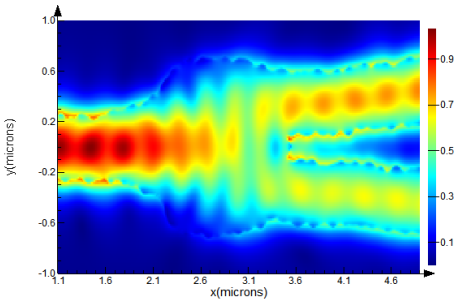


(d)

Figure 2-9: Average imbalance between two output ports for 50 instantiations at: (a) $A = 3\ nm$ and $L_c = 10\ nm$, (b) $A = 3\ nm$ and $L_c = 40\ nm$, (c) $A = 10\ nm$ and $L_c = 10\ nm$, (d) $A = 10\ nm$ and $L_c = 40\ nm$. Notice difference in vertical scales.



(a)



(b)

Figure 2-10: Electric field profiles at the Y-branch junction for the (a) smooth case and (b) LER case with $A = 15\ nm$ and $L_c = 60\ nm$, where the roughness breaks the device symmetry and causes more power to go into the upper port for this instantiation.

devices there are inevitable back reflections and scattering losses that make this ratio less than 1 (some negative value on a dB scale).

To analyze the effect of LER on excess loss for the Y-branch, the excess loss is analyzed for the 50 instantiations of the Y-branch with different A and L_c. The results for the 50 instantiations for $\lambda = 1550\ nm$ are shown in Fig. 2-11 for selected A and L_c combinations, along with the results for the nominal (smooth) case. We see that the LER causes an increase in the excess loss values in the majority of the instantiations, and the fraction of instantiations with loss values greater than the ideal case increases with increasing values of A and L_c. At large enough values (for example, $A = 15\ nm$ and $L_c = 60\ nm$), all of the instantiations have higher excess loss than the ideal case. The maximum loss observed over all instantiations increases as well with increasing A and L_c. Figure 2-11 shows that excess losses can reach -0.8dB (16.8%) when $A = 15\ nm$ and $L_c = 60\ nm$, while the excess loss is only about -0.12dB (3%) for the smooth case. The smooth case still suffers loss due to the junction region where the mode splits into the two paths [2], which also occurs in the LER cases. This means that losses greater than this nominal case can be attributed to the LER. Thus, in addition to causing imbalanced transmission between the two output ports of the Y-branch, LER also increases the excess loss that the light experiences as it propagates.

2.2.3 Back reflection

In sourced waveguiding structures, it is also instructive to analyze the back reflection, which is the fractional power returned to the input port. Since the excess loss defined in Section 2.2.2 is the ratio of the total output power (upper port plus lower port) to the input port, the power lost to back reflection is included in the excess loss. However, back reflection can be measured separately in FDTD simulation, which will decouple its effect on excess loss from that of propagation loss. Figure 2-12 shows the back reflection, as a percentage of total incident power, for four combinations of A and L_c. The LER has a lesser effect on the back reflection than on the total excess loss reported in Section 2.2.2. Although the back reflection can reach higher values

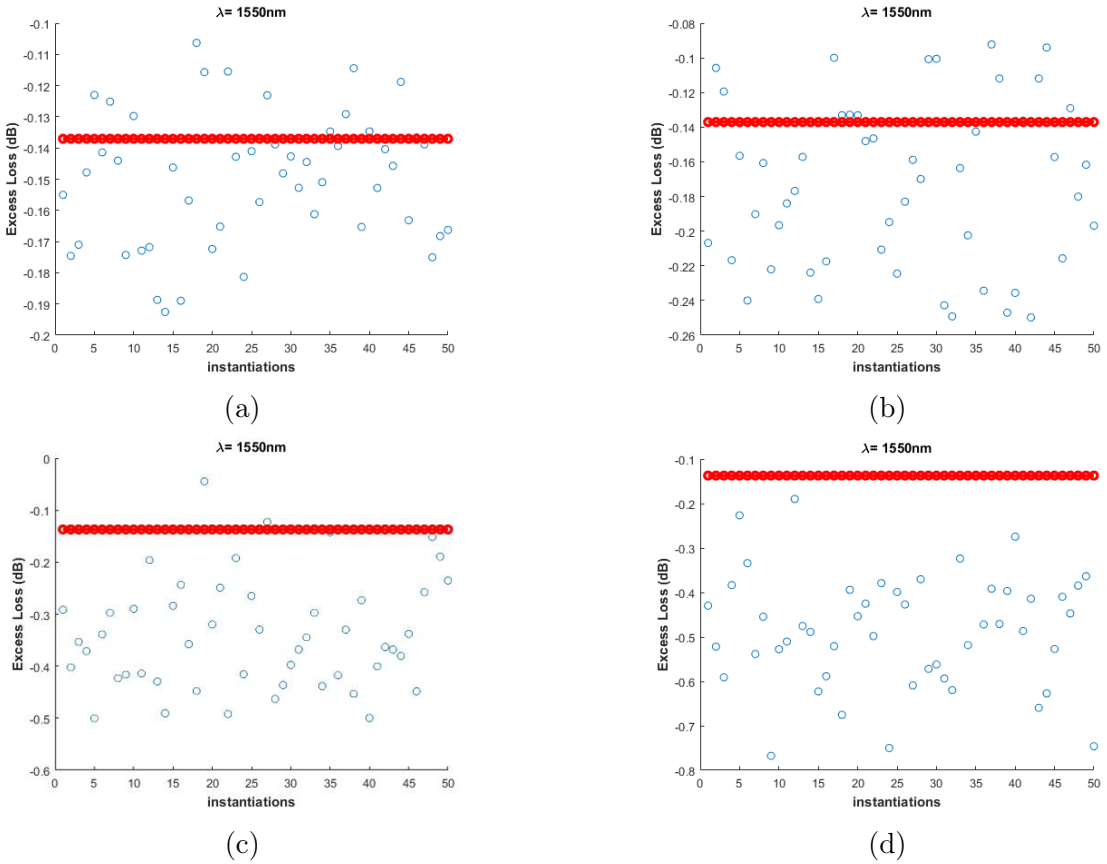


Figure 2-11: The Y-branch excess loss variation from the nominal case (red line) with LER for 50 instantiations at the center wavelength of 1550 nm with LER. (a) $A = 3\ nm$ and $L_c = 10\ nm$, (b) $A = 7\ nm$ and $L_c = 30\ nm$, (c) $A = 12\ nm$ and $L_c = 50\ nm$, (d) $A = 15\ nm$ and $L_c = 60\ nm$. Note differences in vertical scales.

as the A and L_c increase, as shown in Fig. 2-12, and the fraction of instantiations in which the back reflection goes above the nominal case also increases with A and L_c, the loss values even in the worst-case scenario never exceed 1%. Despite the fact that the back reflection losses decrease in almost half of the instantiations, the overall losses still increase with A and L_c because the back reflection is very small compared to the excess loss.

2.2.4 Sensitivity

To further understand the Y-branch, we examine the sensitivity of various sections of the Y-branch geometry to LER, seeking to locate the part that is most sensitive to

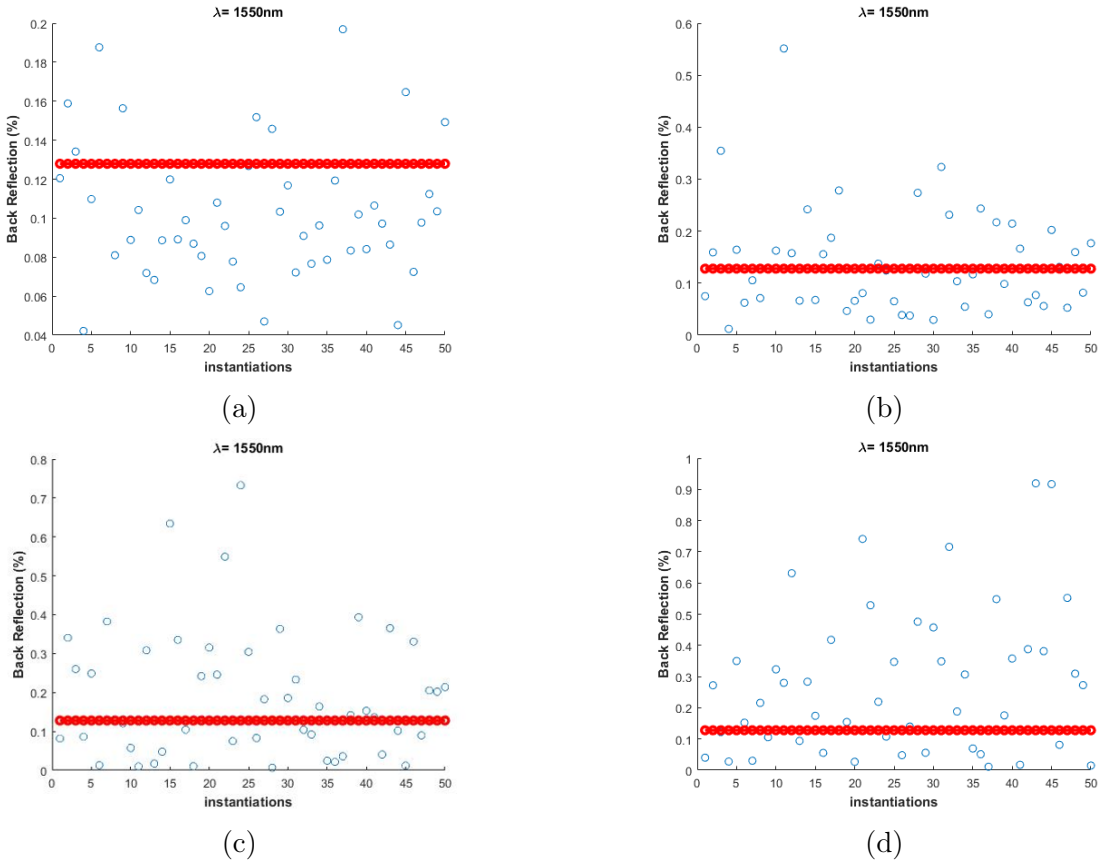


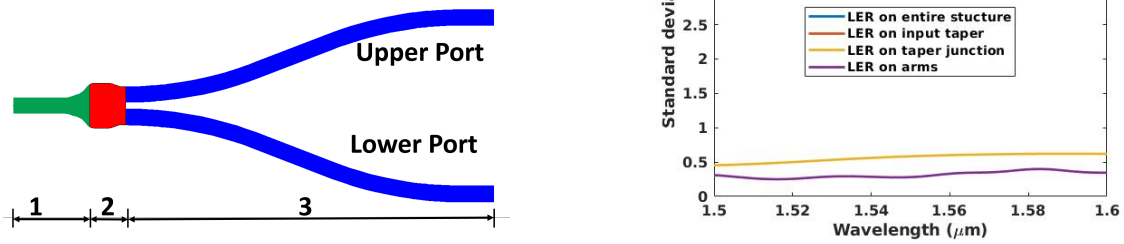
Figure 2-12: The Y-branch back reflection variation from the nominal case (red line) with LER for 50 instantiations at the center wavelength of 1550 nm with LER. (a) $A = 3\ nm$ and $L_c = 10\ nm$, (b) $A = 7\ nm$ and $L_c = 30\ nm$, (c) $A = 12\ nm$ and $L_c = 50\ nm$, (d) $A = 15\ nm$ and $L_c = 60\ nm$. Note differences in vertical scales.

LER. We divide the Y-branch into three main regions: the input taper (region 1), the junction (region 2), and the two Y-branch arms (region 3) as shown in Fig. 2-13(a). LER is applied to each of these regions separately and for each case, we generate 50 different instantiations for each A and L_c combination considered to capture the transmission statistics with respect to the spatial LER. We compare the transmission imbalance caused by each section separately to the case when LER is applied to the entire structure, Fig. 2-13(b).

Table 2.2 shows the standard deviation for the imbalance between the two output ports when LER is applied to these different sections of the Y-branch. From the results in Table 2.2 and Fig. 2-13, we find that the input taper (region 1) in the Y-branch, where the power starts to split between the two output ports, is the most

A (nm)	L_c (nm)	Entire structure	Input taper	Junction	Arms
1	25	$0.25\binom{0.31}{0.21}$	$0.25\binom{0.31}{0.21}$	$0.04\binom{0.05}{0.04}$	$0.04\binom{0.05}{0.04}$
3	10	$0.45\binom{0.56}{0.38}$	$0.45\binom{0.56}{0.38}$	$0.08\binom{0.10}{0.07}$	$0.07\binom{0.09}{0.06}$
3	20	$0.64\binom{0.79}{0.53}$	$0.63\binom{0.79}{0.53}$	$0.12\binom{0.15}{0.10}$	$0.09\binom{0.11}{0.08}$
3	40	$0.90\binom{1.12}{0.75}$	$0.90\binom{1.12}{0.75}$	$0.14\binom{0.18}{0.12}$	$0.11\binom{0.14}{0.09}$
6	60	$2.71\binom{3.38}{2.26}$	$2.59\binom{3.24}{2.18}$	$0.29\binom{0.36}{0.24}$	$0.25\binom{0.31}{0.21}$
7	30	$2.33\binom{2.90}{1.94}$	$2.28\binom{2.85}{1.92}$	$0.32\binom{0.40}{0.27}$	$0.21\binom{0.26}{0.18}$
10	10	$1.86\binom{2.31}{1.51}$	$1.86\binom{2.31}{1.51}$	$0.32\binom{0.40}{0.27}$	$0.18\binom{0.23}{0.15}$
10	40	$3.85\binom{4.80}{3.23}$	$3.83\binom{4.79}{3.22}$	$0.59\binom{0.74}{0.50}$	$0.29\binom{0.36}{0.24}$
12	50	$4.04\binom{5.05}{3.39}$	$4.09\binom{5.11}{3.43}$	$0.64\binom{0.80}{0.54}$	$0.41\binom{0.51}{0.34}$
15	60	$5.65\binom{7.06}{4.75}$	$5.83\binom{7.29}{4.9}$	$0.88\binom{1.1}{0.74}$	$0.64\binom{0.80}{0.54}$

Table 2.2: Summary of imbalanced transmission standard deviation (in percent) for LER applied to different sections of a Y-branch at a wavelength of 1550 nm. The values in the brackets are the 95% lower and upper bound confidence intervals calculated for the ensemble of 50 simulations.



(a) (b)

Figure 2-13: (a) Y-branch geometry where region 1 (green) is the input taper, region 2 (red) is the junction and region 3 (blue) contains the output arms. (b) Standard deviation σ (in percent) of the transmission imbalance caused by applying LER for different sections of the Y-branch for $A = 10\ nm$ and $L_c = 40\ nm$.

sensitive section to LER, and regions 2 and 3 are not as sensitive.

2.3 Modeling

In previous sections, we have demonstrated that when the Y-branch is subjected to LER, there is an asymmetric power distribution at the output ports, as shown in Fig. 2-4. This imbalance is considered to be a random device variation, and is highly dependent on the statistical nature of the LER.

Figure 2-6 shows that the output imbalance is not unique for a given set of A, L_c and λ, since any particular instantiation (which corresponds to one Y-branch on a wafer) can have very different output characteristics than others. Therefore, any model aiming to quantify the performance of a Y-branch subjected to LER should be constructed for statistical performance values, namely the transmission mean (μ) and variance (σ^2), as a function of the LER parameters and wavelength of operation.

For each A and L_c combination, the μ and σ^2 are calculated using results from its 50 instantiations, as shown for selected combinations in Fig. 2-14. The 2σ (95%) confidence intervals (CI) are also plotted. We note that the mean decreases with increasing A and L_c. This is likely due to the increased levels of losses introduced by

the LER peaks, as the peaks act as scattering locations for the guided fundamental mode to leak out of the main waveguide [32]. Also, the variance increases with increasing A and/or L_c, which is likely due to these regimes presenting the greatest deviation of the device geometry over a given length from its nominal (noise-free) structure, as the LER becomes more impactful on the Y-branch. In Stage 3, shown in the workflow in Fig. 2-5, the data from the above ensemble virtual fabrication and optical simulations can be used to develop statistical compact models.

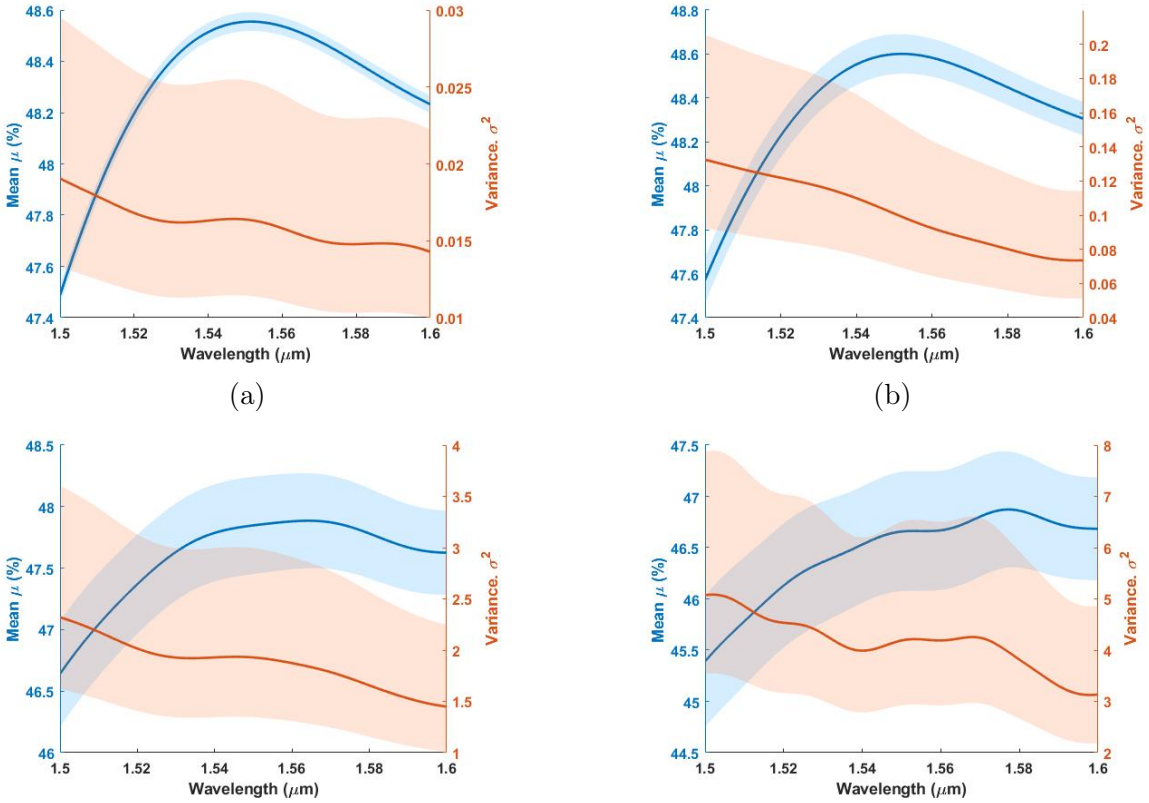


Figure 2-14: Mean and variance for the upper port transmission calculated from an ensemble of 50 runs for different amplitude and correlation length combinations. The shaded area is the 95% confidence interval. (a) $A = 1\ nm$ and $L_c = 25\ nm$, (b) $A = 3\ nm$ and $L_c = 20\ nm$, (c) $A = 6\ nm$ and $L_c = 60\ nm$, (d) $A = 12\ nm$ and $L_c = 50\ nm$. Note differences in vertical scales.

This section details the methods used to develop such statistical compact models. Two different machine learning regression methods are used: polynomial linear regression [33] and Gaussian process (GP) regression [34, 35, 36]. For each method,

models are developed to predict the mean and variance as a function of the LER parameters (A and L_c) and the operating wavelength (λ), and the performance of each model is compared. The models are developed using data from the ten sets of A and L_c combinations, each containing 50 unique instantiations of the structure, previously described. For each set, the mean and variance are calculated. At the end, we use another data set to test the model, this test set data for $A = 4\ nm$ and $L_c = 50\ nm$ is never used for model fitting, only for final model testing. Since there are only ten A and L_c sets, there is not enough data to perform validation analysis on the training data, i.e., by dividing the data into two sets (training and validation). Instead, leave one out cross validation (LOOCV) [33] is used on the training data. LOOCV interchangeably leaves out one set of data to be used as a validation set, the model is fit with the remaining data, and a validation error is calculated for that case. Then all the validation errors are averaged, so that the two methods can be compared.

2.3.1 Mean

The mean of the transmission not only gives an insight on the average port transmission but also reflects the losses experienced due to LER. For modeling the transmission mean, we start by employing polynomial regression. A linear regression model order that prevents overfitting as well as accurately explains most of the data is a second-order polynomial model in the three parameters (A, L_c, and λ):

$$\mu(A, L_c, \lambda) = \beta_0+\beta_1 A+\beta_2 L_c+\beta_3\lambda+\beta_4 A^2+\beta_5 L_c^2+\beta_6\lambda^2+\beta_7 A L_c+\beta_8 L_c\lambda+\beta_9 A\lambda \quad (2.3)$$

The coefficient of determination, R^2, which says how well the model predicts the observed data compared to an overall mean, is 0.966 for this model. This indicates that the model explains 96.6% of the data, while the R_A^2 (which takes into account the number of terms forming the model) value is 0.9654. The mean squared error (MSE) of the the transmission power model is 0.04. Figure 2-15 shows the model performance over different training data sets, i.e., different A and L_c combinations.

The model fails to predict the mean within the 95% CI of the simulated mean in the low-amplitude cases. This is because the variance is very low in these cases (as shown in Fig. 2-14) leading to small CIs.

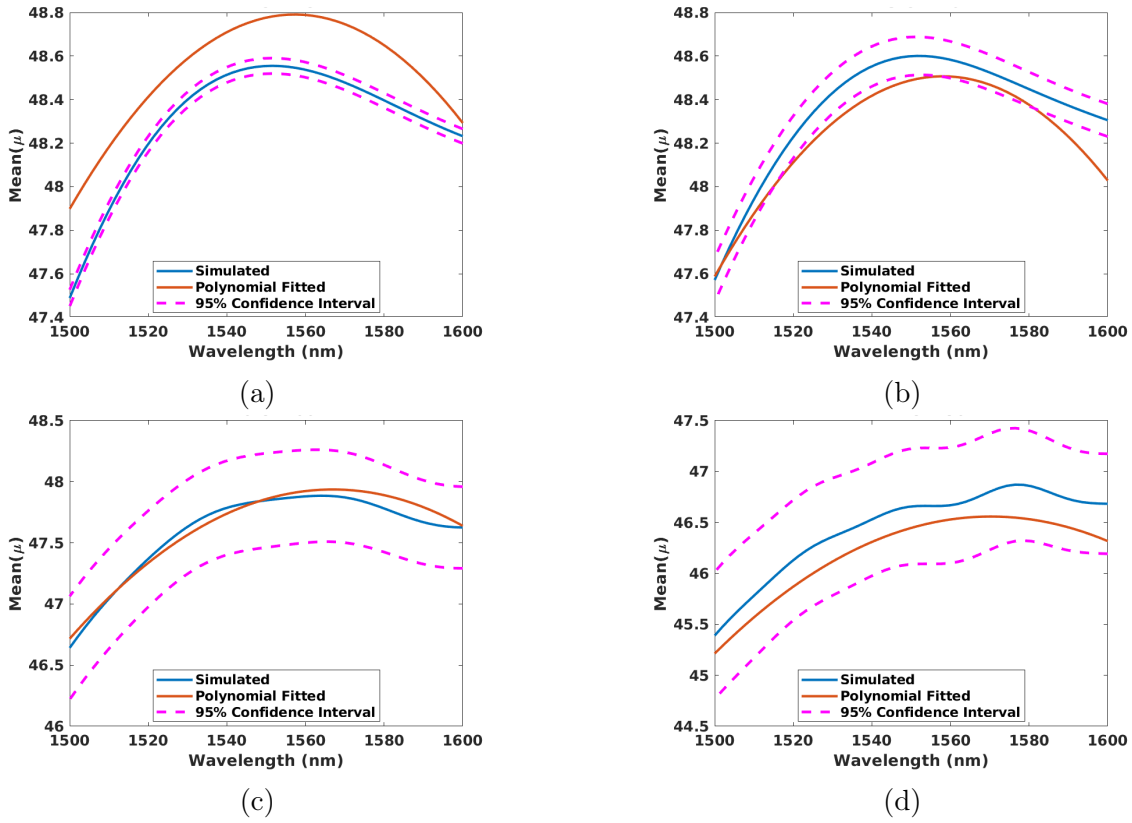


Figure 2-15: The mean values predicted from the polynomial model vs. the values obtained from simulation for different A and L_c combinations. (a) $A = 1\ nm$ and $L_c = 25\ nm$, (b) $A = 3\ nm$ and $L_c = 20\ nm$, (c) $A = 6\ nm$ and $L_c = 60\ nm$, (d) $A = 12\ nm$ and $L_c = 50\ nm$. Note differences in vertical scales.

The next statistical modeling technique, Gaussian process (GP) regression, is a powerful non-linear prediction tool when there is a small number of data points, as is the case in this study, whereas neural networks can fail to generalize to unseen data and often overfit. GP is a nonparametric kernel-based probabilistic framework that can predict a response (Y) as a multivariate normal distribution using a certain covariance (kernel) function, $k(x_i, x_j)$, that correlates different inputs (x). The kernel function determines how close the responses (Y) are to each other based on the relations between the inputs (x), where points with equal or close input values are

likely to have equal or close responses (Y). Accordingly, using the kernel function on new data points (x_{new}) can predict the response (Y_{new}) at those points. In this study, the squared exponential kernel function is used:

$$k(x_i, x_j|\theta) = \sigma_f \exp[-\frac{1}{2}\sum_{m=1}^{d}\frac{(x_{im} - x_{jm})^2}{\sigma_m^2}] \tag{2.4}$$

where θ represents the hyperparameters (σ_f, σ_m) that parametrize the kernel function; σ_f is the signal standard deviation, σ_m is the characteristic length scale which dictates how far apart two points (x) need to be in order for the response (Y) to be uncorrelated, and d is the dimension of the input. In this study, d is third order (A, L_c and λ).

The GP regression (GPR) for mean is developed using the three training features (A, L_c and λ), where the model's calculated hyperparameters are σ_f=1.0245; σ_A=0.3552, which is the LER amplitude characteristic length scale; σ_{L_c}=3.2078, which is the LER correlation length characteristic length scale; and σ_λ=0.4152, which is the wavelength characteristic length scale. The amplitude's characteristic length scale has a low value, which means that responses for different amplitudes are not very correlated to one another. This agrees with the findings presented in Section 2.2 and in [28], which showed that the LER amplitude is the most influential parameter in causing the imbalanced transmission, i.e., variations in A causes larger changes in the response compared to L_c. The GPR developed here has an R^2 value of 0.999 and an MSE of 8.3×10^{-6}. As seen in Fig. 2-16, this model accurately predicts both high and low amplitude values, unlike the linear regression model.

The MSE for the LOOCV for both models are shown in Table 2.3. The GP regression model has lower MSE values than the polynomial regression. Moreover, both models are used to predict the performance of a hold out test set ($A = 4\ nm$ and $L_c = 50\ nm$), for visualization of typical fits. The results are shown in Fig. 2-17. The GP and linear regression models perform similarly in predicting the test data set. Since the linear regression model performs poorly for small amplitude values, we conclude that the GP regression model outperforms the linear regression model for

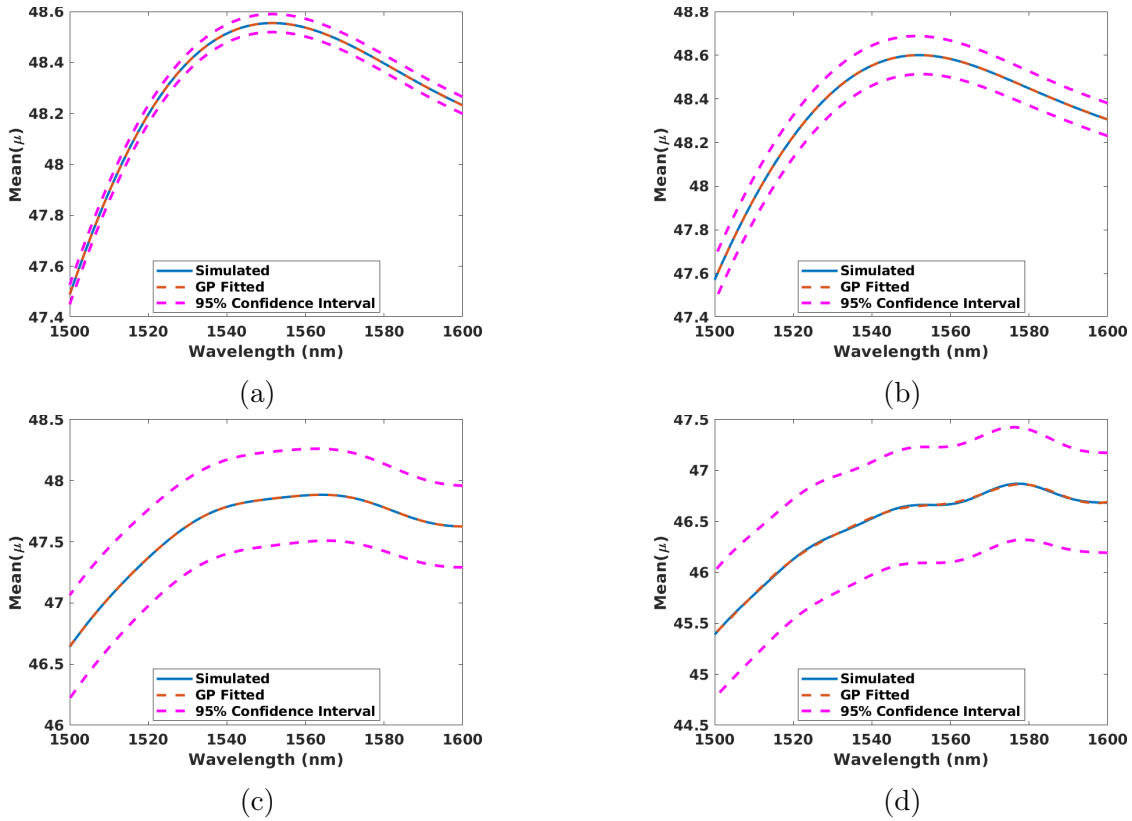


Figure 2-16: The Gaussian Process model predicted mean values vs. the virtual fabrication simulation values for different A and L_c combinations. (a) $A = 1\ nm$ and $L_c = 25\ nm$, (b) $A = 3\ nm$ and $L_c = 20\ nm$, (c) $A = 6\ nm$ and $L_c = 60\ nm$, (d) $A = 12\ nm$ and $L_c = 50\ nm$. Note differences in vertical scales.

mean prediction in this study.

Model	**Across whole spectrum**	**At** 1550 nm
Linear regression (2^{nd} order)	0.165	1.578
GP regression	0.093	0.862

Table 2.3: Comparison between the LOOCV mean squared error in transmission output port for different models of the mean.

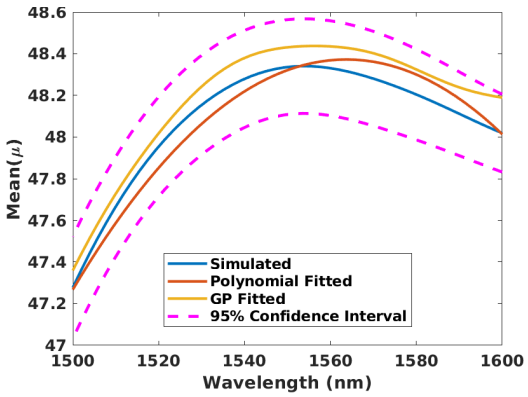


Figure 2-17: The model predicted mean values vs. the virtual fabrication simulation values, for the hold out test data set with $A = 4\ nm$ and $L_c = 50\ nm$. The 95% confidence interval corresponds to the original ensemble simulations shown in blue.

2.3.2 Variance

The above procedure for mean prediction is also used for the variance. The following third-order regression model is used:

$$\begin{aligned}\sigma^2(A, L_c, \lambda) = {} & \beta_0 + \beta_1 A + \beta_2 L_c + \beta_3 \lambda + \beta_4 A^2 + \beta_5 L_c^2 + \beta_6 \lambda^2 + \beta_7 A L_c + \beta_8 L_c \lambda \\ & + \beta_9 A\lambda + \beta_{10} A^3 + \beta_{11} L_c^3 + \beta_{12} \lambda^3 + \beta_{13} A^2 L_c + \beta_{14} A^2 \lambda + \beta_{15} A L_c^2 \\ & + \beta_{16} L_c^2 \lambda + \beta_{17} A \lambda^2 + \beta_{18} L_c \lambda^2 + \beta_{10 A L_c \lambda}\end{aligned} \tag{2.5}$$

The model has the same R^2 and R_A^2 values (0.9994), and an MSE value of 4.43×10^{-4}. The GPR model for variance has calculated hyperparameters: σ_f=0.6844; σ_A=0.3617, which is the LER amplitude characteristic length scale; σ_{L_c}=1.4667, which is the LER correlation length characteristic length scale; and σ_λ=0.7263, is the wavelength characteristic length scale. The GP regression model has an R^2 value of 0.999 and an MSE of 2.2×10^{-5}. The performance of both models on the training set is shown in Fig. 2-18. Unlike the mean prediction, both models, polynomial and GP regression, perform well for all of the A and L_c combinations.

To further compare and choose between the models, LOOCV is performed. The computed MSE is shown in Table 2.4. The linear regression model has MSE that is lower by an order of magnitude. Both models are then run with the hold out test

data set ($A = 4\ nm$ and $L_c = 50\ nm$). The results are shown in Fig.2-19. Visually, one sees that the polynomial regression model more closely adheres to the test data. Given that the linear regression model has a lower LOOCV MSE, and that it is computationally cheaper than GP (especially in the presence of more training data), we conclude that the linear regression model outperforms the GP model for variance prediction.

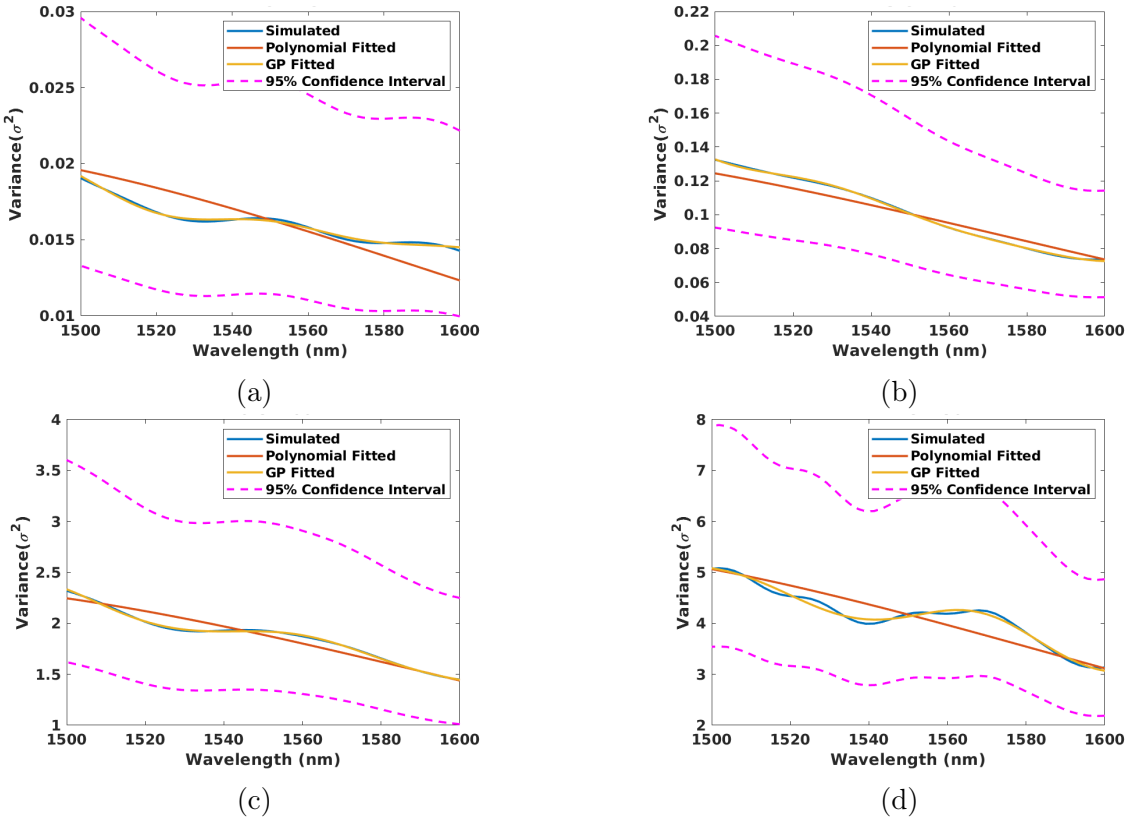


Figure 2-18: The upper port model predicted variance values vs. the virtual fabrication simulation values for different A and L_c values. (a) $A = 1\ nm$ and $L_c = 25\ nm$, (b) $A = 3\ nm$ and $L_c = 20\ nm$, (c) $A = 6\ nm$ and $L_c = 60\ nm$, (d) $A = 12\ nm$ $L_c = 50\ nm$. Note differences in vertical scales.

Model	**Across whole spectrum**	**At 1550** nm
Linear regression (3^{rd} order)	1.71×10^{-3}	0.01475
GP regression	0.0214	0.23198

Table 2.4: Comparison between the LOOCV mean squared error for different models of the port transmission variance.

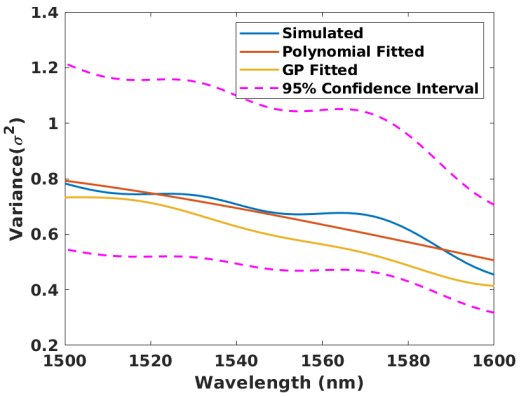


Figure 2-19: The model predicted variance values vs. the virtual fabrication simulation value for the hold out test data set with $A = 4\ nm$ and $L_c = 50\ nm$. The 95% confidence interval corresponds to the original ensemble simulations shown in blue.

2.4 Model Applications

This section aims to demonstrate ways in which the models developed above can be leveraged in the design process. The wavelength dependent GP model for mean and polynomial model for variance are used to generate transmission spectra for many instantiations, for comparison to Fig. 2-6 which shows actual ensemble virtual fabrication simulated data. The generation of a single instantiation using these models takes only few seconds, compared to several hours needed for ensemble generation and simulation using FDTD and SEMlator3D simulations. Figure 2-20(a-b) shows 50 different random instantiations generated for the upper port transmission using the models for selected A and L_c combinations. Comparing to Fig. 2-6, one notices the same general shape of the transmission curves (the peak near the middle of the queried band), and similar ranges of the transmission for each parameter combination.

However, the data generated from the models show far less fluctuations in the queried wavelength range than do the simulation data; this is because the models developed do not account for the within-wavelength variations between instantiations. To address this issue, we further add a within wavelength variance between instantiations (σ_i) as a function of the LER amplitude and correlation length. The final

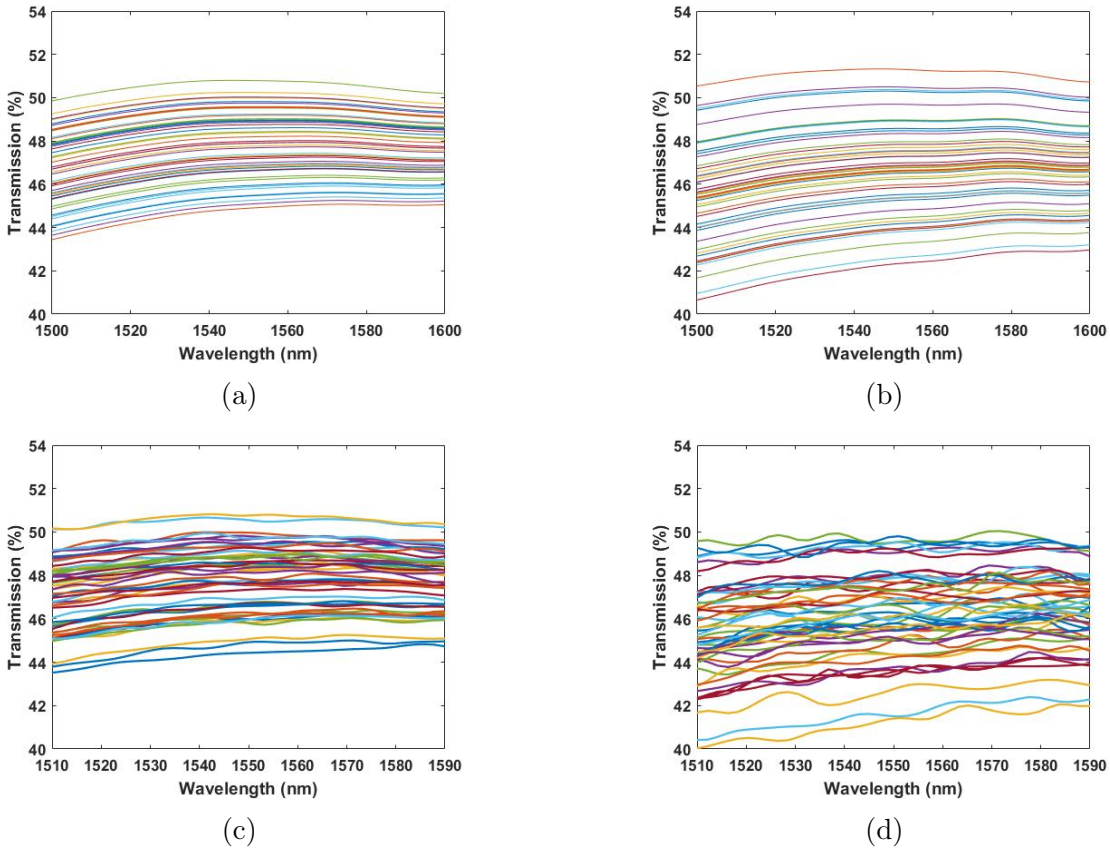


Figure 2-20: Generated random instantiations for upper port transmission using the statistical compact models, for different A and L_c values without considering the variance across wavelength: (a) $A = 6\ nm$ and $L_c = 60\ nm$, (b) $A = 12\ nm$ and $L_c = 50\ nm$. With adding the variance across wavelength: (c) $A = 6\ nm$ and $L_c = 60\ nm$, (d) $A = 12\ nm$ and $L_c = 50\ nm$.

model to generate an instantiation becomes:

$$T(A, L_c, \lambda) = N_s(\mu_s, \sigma_s) + N_i(0, \sigma_i) \tag{2.6}$$

where N_s is the normal distribution of the statistics of the transmission such that $\mu_s = \mu(A, L_c, \lambda)$ and $\sigma_s = \sqrt{\sigma^2(A, L_c, \lambda)}$ are the mean and standard deviations modeled above, while N_i is the normal distribution of the variation within-wavelength between instantiations, and σ_i is its standard deviation. The generated instantiations using the updates model are shown in Fig. 2-20(c-d), which are similar the ensemble simulations shown in Fig. 2-6.

These generated spectra can help in predicting the yield (the percentage of working devices) when the Y-branch is subjected to a specific set of LER parameters. Here, the yield is defined as the percentage of Y-branches having imbalance less than a certain threshold value. Most photonic applications have a very small threshold for imbalance and are not forgiving. Typically a threshold of 1% is considered, which is adopted for our example. Transmission spectra for 500 random instantiations for each different combinations of A and L_c are generated using our developed statistical compact models, and the yield is calculated for each combination. Figure 2-21 shows the yield at λ=1550 nm as a function of LER correlation length (L_c) and amplitude (A). Information such as this can help a designer predict yield, if they have information about the foundry-specific LER parameters their devices will be subjected to.

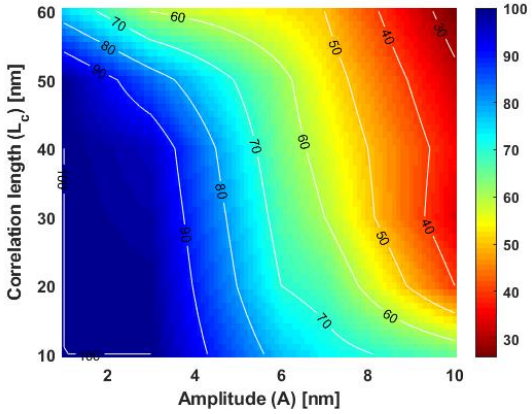


Figure 2-21: Yield (%) predicted at λ=1550 nm for the Y-branch using 500 generated instantiations for each A and L_c, at 1% imbalance threshold.

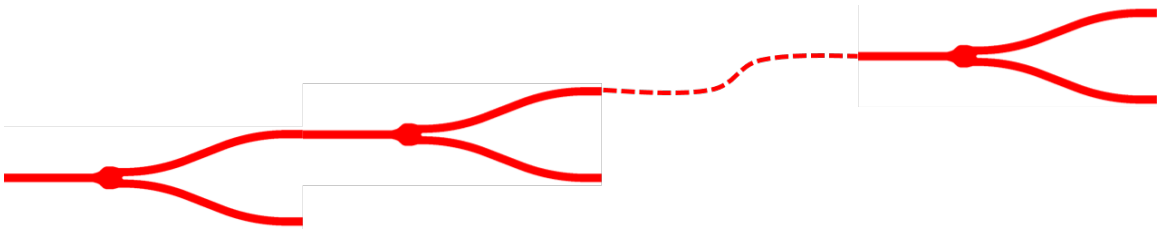


Figure 2-22: Series of Y-branches connected together to form a 1 : 2N split ratio.

A Y-branch often exists within a series of connected Y-branches to form a $1 : 2^N$ splitting network, where N is the number of levels of Y-branches in the network, as pictured in Fig. 2-22. To show this, six Y-branches are generated from the developed

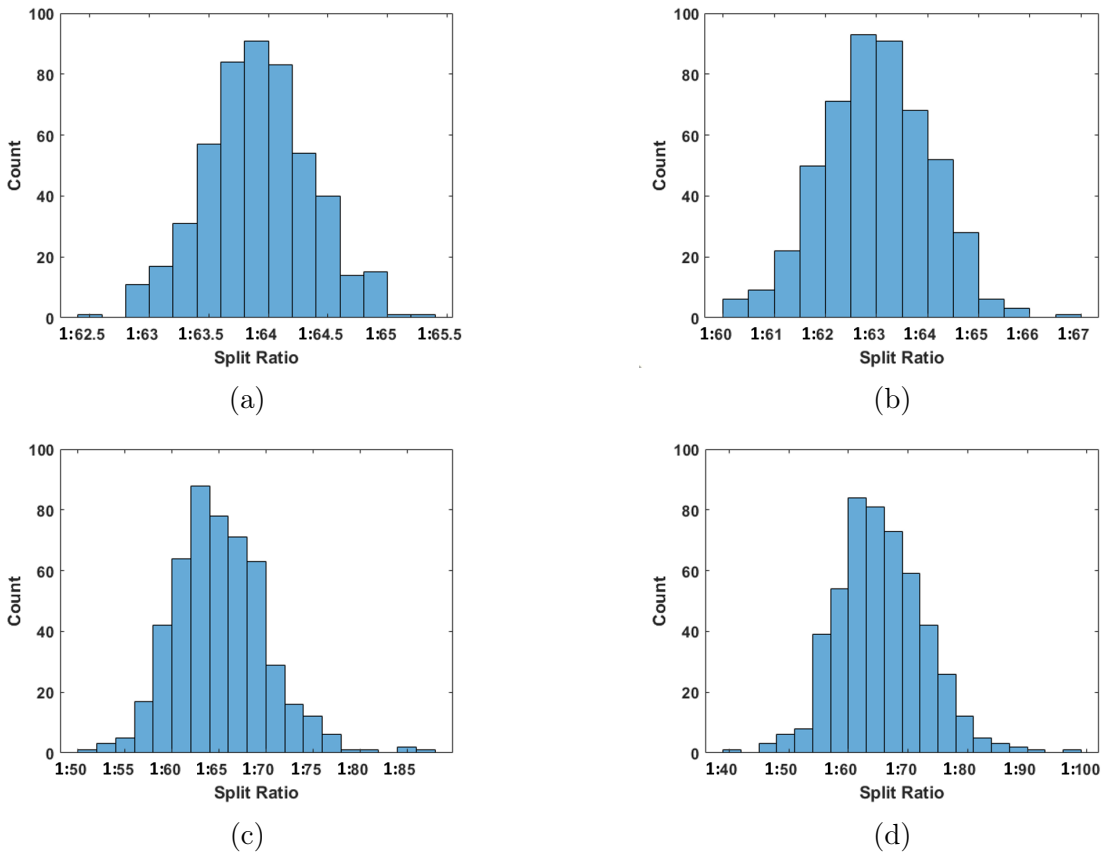
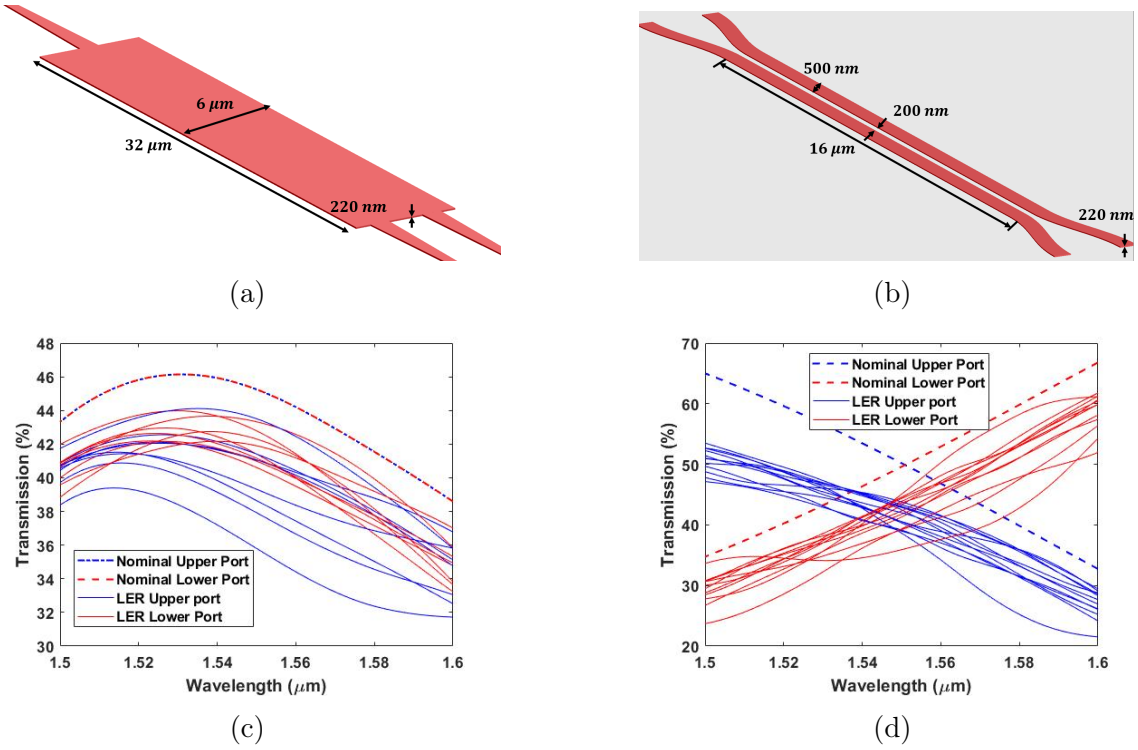


Figure 2-23: Split ratio calculated from connecting six generated Y-branches where this experiment is repeated 500 times for different A and L_c values. (a) $A = 1\ nm$ and $L_c = 25\ nm$, (b) $A = 3\ nm$ and $L_c = 20\ nm$, (c) $A = 6\ nm$ and $L_c = 60\ nm$, (d) $A = 12\ nm$ $L_c = 50\ nm$. Note differences in the horizontal scales.

mean (GPR based) and variance (polynomial based) models, and the output from the upper port for each is multiplied to obtain the final split ratio. This procedure is repeated 500 distinct times. Figure 2-23 shows the histogram for the 500 generated instantiations of six connected Y-branches. Ideally, the final output power would be 1/64 of the initial power (a 1:64 splitting ratio). As the A and/or L_c increase, the more instantiations there are that deviate from the ideal 1:64 ratio, indicating a larger shift from desired device behavior.

2.5 Application to Power Splitting Structure Comparison

The same methodology is applied to two different power splitting structures in addition to the Y-branch, including multimode interference (MMI) and 50/50 directional couplers, with 50/50 split ratio near 1550 nm, as shown in Fig 2-24. The sensitivity of each of these structures to LER is examined, where the imbalance and the excess loss experienced are calculated.



(a) (b) (c) (d)

Figure 2-24: (a) MMI geometry and, (b) directional coupler geometry, each to have 50/50 split at 1550 nm. (c) Transmission for MMI with and without LER. (d) Transmission for directional coupler with and without LER. Here LER with $A = 15$ nm and $L_c = 60$ nm is used.

Figure 2-25 shows the imbalance and excess loss (defined here as the extra loss the structure experience due to LER compared to the no LER case) distribution, experienced at wavelength of 1550 nm, for the three structures when LER of amplitude $A = 15$ nm and correlation length of $L_c = 60$ nm is applied to each, and 100 instantiations are simulated.

In the case of MMI, an imbalanced transmission occurs. We see that the MMI is the least sensitive to LER, as it encounters a 2% standard deviation imbalance and a maximum imbalance of 7%. Moreover, the excess loss it experiences has a mean of 8% and a standard deviation of 1.9%.

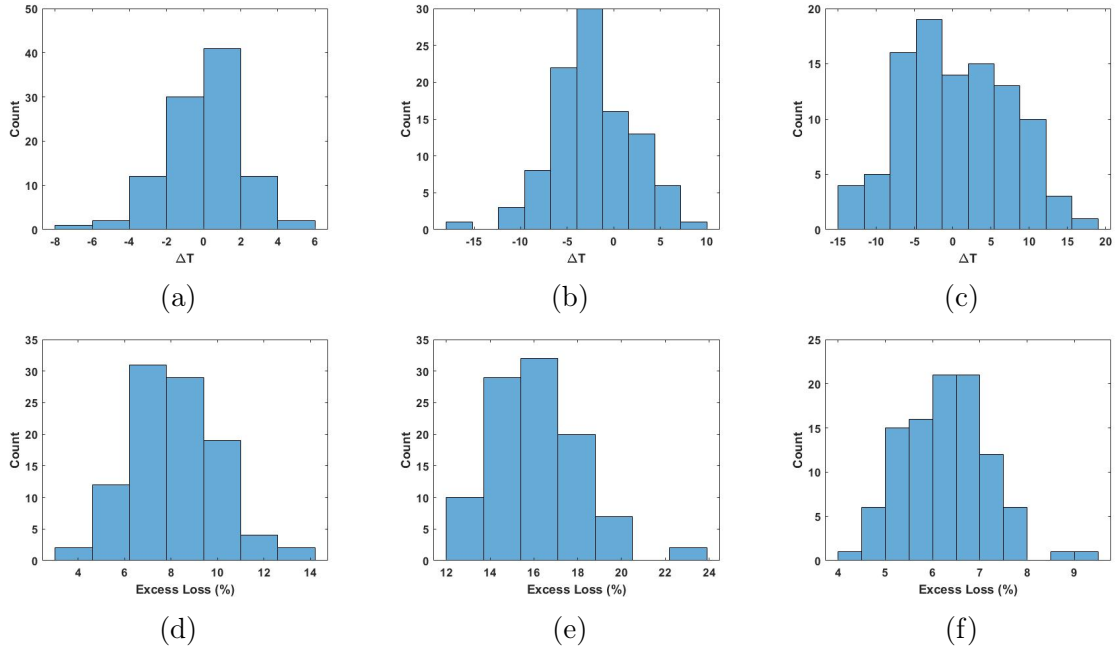


Figure 2-25: Imbalance in transmission at 1550 nm: (a) MMI, (b) directional coupler, (c) Y-branch. Excess loss at 1550 nm due to LER: (d) MMI, (e) directional coupler, (f) Y-branch.

We also observe that for the 50/50 directional coupler, the frequency of the 50/50 split point shifts, as shown in Fig. 2-26. This can be viewed as measure of variation or imbalance in the designed frequency of the 50/50 split. The standard deviation for the transmission imbalance at 1550 nm (due to the shift in the 50/50 split wavelength) is 3.5% with maximum imbalance of 15%. The excess loss for the directional coupler has a mean of 16% and a σ of 2%. Finally, the Y-branch imbalance standard deviation is 5% and has a maximum value of 17%, with an excess loss of 6% mean and 1% standard deviation, with this LER case.

Based upon this analysis we can see that there is a trade off between the footprint of the device and its sensitivity to LER. As the device becomes larger in size, it becomes less sensitive to the LER. Accordingly, choosing the device to use to achieve

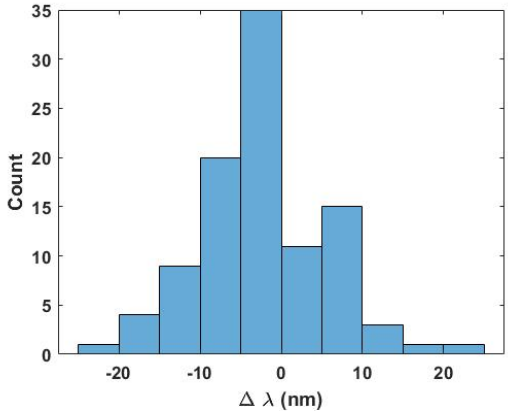


Figure 2-26: Histogram for the variation in the directional coupler wavelength at which the 50/50 split occurs, between the no LER nominal case (λ=1550 nm) and the LER cases.

splitting with a required ratio is dependent on the tolerance of the imbalance, as well as the area available for implementation.

The statistical compact modeling approach used here for random variations can be extended to other optical components, and combined with models for systematic geometric variations (as discussed in the next chapter) to serve as a basic building block for extended process design kits (PDKs) for silicon photonics. These PDKs would help photonic integrated circuit (PIC) designers predict the performance across process variations in the same way that is currently heavily employed in CMOS electronic IC design.

2.6 Summary

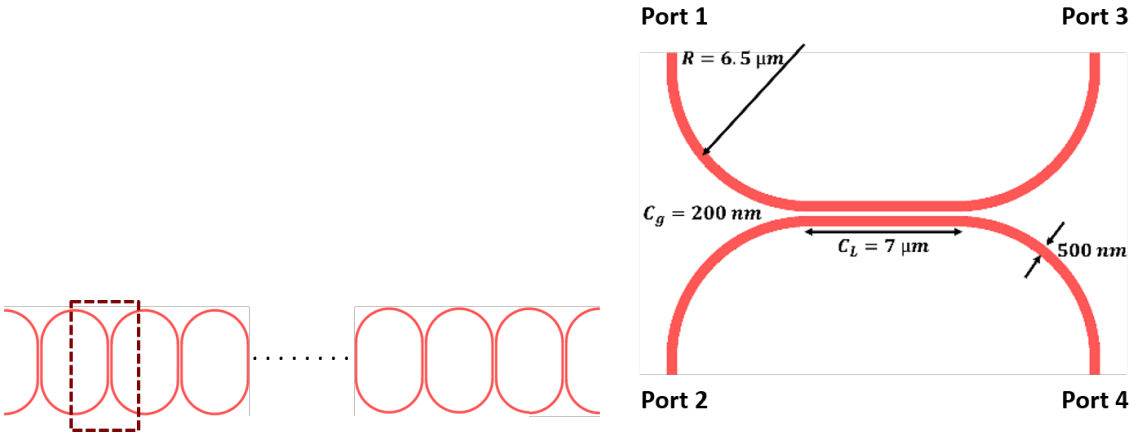
Line edge roughness with different combinations of amplitude and correlation length values is applied to many instantiations of a Y-branch through virtual fabrication. The optical transmission characteristics of the resulting structures are studied. Results show that LER causes unbalanced transmission between the upper and lower port, which differs substantially from the ideal (smooth) case where the power is split evenly between the two output ports. The amount of the imbalance is dependent on the statistical parameters of the LER (amplitude and correlation length). The largest

imbalance between the two output ports for a single run over all the wavelengths is 17% for large A and L_c at wavelength of 1550 nm across 50 ensemble simulations. However, the imbalance is relatively small ($\sim 1\%$) for small LER amplitude. In addition, LER increases the excess losses that the Y-branch experiences, as compared to the nominal case. These results motivate the desirability to develop statistical compact models for the LER effect on Y-branch. These models can then be used for efficient analysis of performance and yield in the face of these variations.

Chapter 3

CROW Compact Modeling

Significant computing and communication system performance gains can be achieved by transferring information using optical rather than electrical signals [37]. Accordingly, often on-chip optical buffers are needed that delay the optical signals. This raises the interest in coupled resonator optical waveguides (CROWs), where a number of ring waveguides are chained together, as illustrated in Fig. 3-1(a). The CROW serves as a slow wave structure as the signal needs to build up in all the rings, that can be used for buffering and storing data [37, 3, 38].



(a) (b)

Figure 3-1: (a) Schematic of a CROW where the dashed box represents the CROW constituting component. (b) Geometry of half ring used during CROW model development, with silicon waveguide nominal width W=500 nm, thickness T=220 nm, coupling length C_L=7 μm and coupling gap C_g=200 nm.

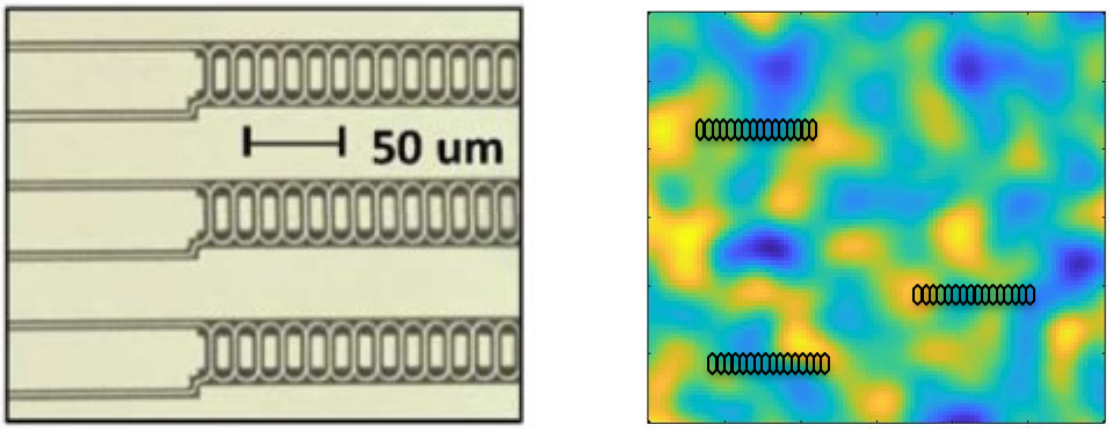
However, since CROWs are large structures extending hundreds of microns to millimeters in length, as in Fig. 3-2(a), depending on the number of constituent rings forming the CROW, the CROW can be challenged by the spatial variations within die or across the wafer, Fig. 3-2(b). These variations can cause a slight difference in the geometry of the rings than the intended nominal design values, which in turn can change the intended passband of the CROW, as seen in the simulation of Fig. 3-3. More importantly, these variations may cause the CROW to fail, if the spatial variations cause each group of rings to have different geometry from the neighboring rings. In such case, the resonances of the coupled rings lose their alignment, suppressing the transmitted power before arriving to the end, Fig. 3-4. Another challenge due to the large size of CROWs is the long simulation time, particularly for repetitive simulations needed to evaluate the impact of each design parameter.

In order to speed up simulations and optimization for CROWs in the presence of process variations, we develop a method to create variation-aware compact models that can be used to simulate and predict the CROW behavior against systematic and spatially correlated process variations. We demonstrate models for variations in CROW ring waveguide thickness and width as well as due to LER. Such proposed compact models enable feasible and fast simulations compared to FDTD or MODE [23] simulations, and enable variation impact and yield analysis.

In Section 3.1, the simulation setup is described with the variation-aware compact model development flow presented. Section 3.2 presents the variation-aware compact model based on S-parameter with example potential applications for the model shown. In Section 3.3, we develop S-parameter based compact models for LER and its application for yield and performance analysis is presented. Finally, Section 3.4 summarizes the chapter.

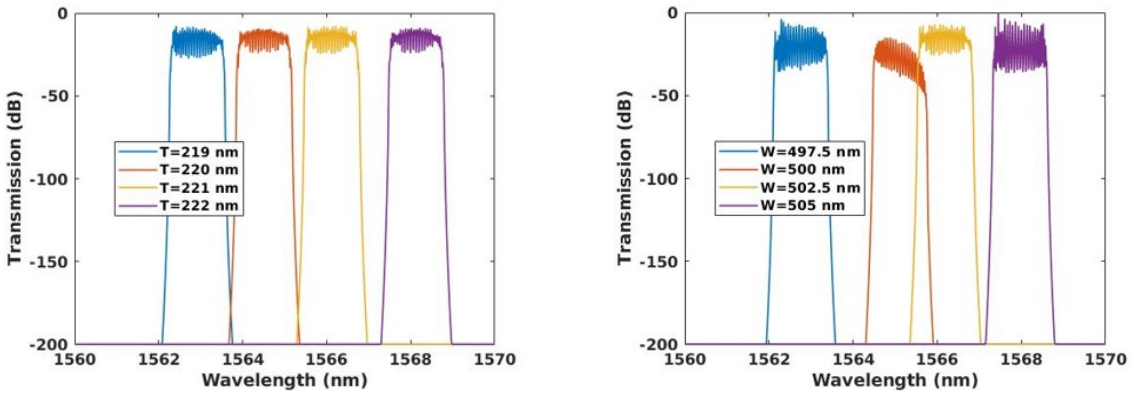
3.1 Simulation Setup

Generally, because a CROW consists of a large number of coupled rings, it is infeasible to simulate the whole structure directly using a device level simulator (FDTD) [22].



(a) (b)

Figure 3-2: (a) Long fabricated CROWs [3]. (b) Conceptual placement of CROW in a die subject to spatially varying geometry (e.g., silicon layer thickness) with each set of rings experiencing different geometry from nominal design due to the spatial variations.



(a) (b)

Figure 3-3: Change in the passband of a 28 ring CROW with all the rings experiencing the same geometric variation in. (a) Silicon layer thickness (T) of the rings away from nominal $T = 220\ nm$, and (b) width W of the rings away from nominal $W = 500\ nm$.

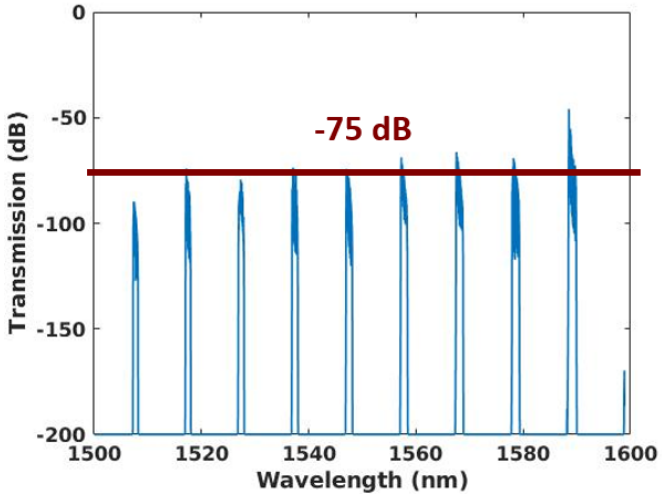


Figure 3-4: The response of CROW consisting of 28 ring with each ring experiencing a slightly different geometric variation. Overall resonance and response is suppressed below a usable level.

Instead, its simulation is based on generating the scattering parameters (S-parameters) for the constituent device component, the half ring as shown in Fig. 3-1(b), with a finite difference time domain (FDTD) simulator. The whole CROW is then simulated with a photonic circuit simulator (INTERCONNECT [39]) by connecting the half rings together. This simulation flow is shown in Fig. 3-5. The S-parameters generated in the FDTD simulator are specific to a certain half ring waveguide design (i.e., fixed thickness and width). So, varying the ring design requires re-generating the S-parameters; this is computationally expensive if a large number of half-ring variants must be considered for variation assessment. This highlights the need for a variation-aware compact model for device components, especially considering the fact that large numbers of these simulations are needed for Monte Carlo statistical simulations or during design optimization.

To develop such variation-aware compact models we start by determining a design of experiments (DOE) [40] for the silicon widths (W) and thicknesses (T) such that these values cover the design space of interest for the half ring. This design space is based on the controlled accuracy typically achieved for photonics 193 nm and 248 nm lithography process, with ± 20 nm for width, and ± 10 nm for thickness [41, 42, 43]. Then, the S-parameters for each of these DOE combinations are generated using

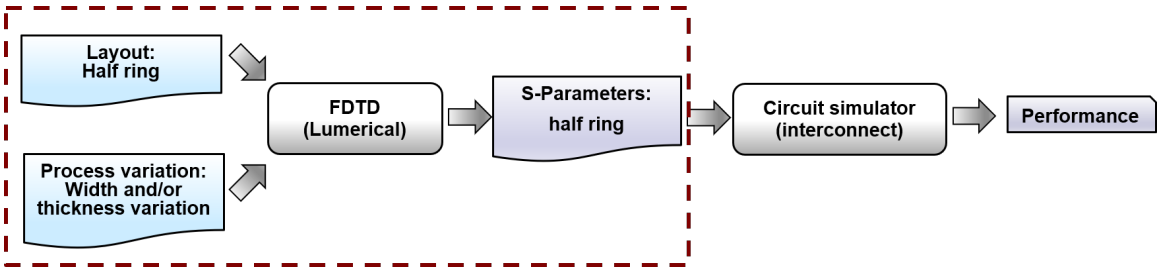


Figure 3-5: Simulation flow to evaluate the CROW performance for a single combination of width and thickness values.

Lumerical FDTD [22], and then used in MATLAB [44] to develop the variation-aware compact models for the S-parameters. Afterwards, these models are used to generate the S-parameters to be used in the circuit simulator to predict the response for new design and variation conditions. The new design flow with the variation-aware compact models is shown in Fig. 3-6.

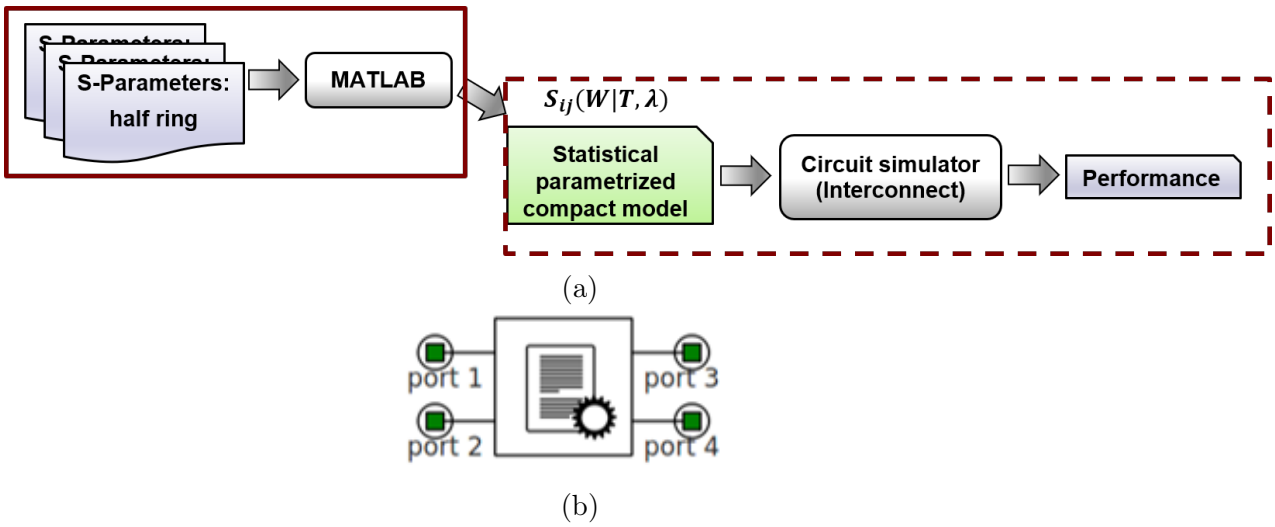


Figure 3-6: (a) Simulation flow (surrounded by dashed box) to obtain the performance for any width and thickness values using the variation-aware compact model. The steps surrounded by solid line box are done only once to generate the parametrized compact model. (b) The S-parameters generated from the parametrized compact model for the half ring can then be used in the circuit simulator.

3.2 Systematic Variations

In this section, we focus on the case when the rings experience systematic variations, i.e., variations in the silicon width and thickness. We consider two applications of the generated variation-aware compact models. The first case is when all rings experience the same variation. A second case then considers spatially varying perturbations.

3.2.1 Models

We first develop an S-parameter based parametrized variation-aware compact model for the CROW constituent device – a half ring with nominal design parameters – as shown in Fig 3-1(b). This model provides the S-parameters for the half ring under varying thickness or width, $S_{ij}(W|T, \lambda)$, with acceptable accuracy and much less computational cost than repeated FDTD simulations. We divide the data set generated based on our DOE (the S-parameters for different geometry) into a 70% training and 30% testing sets. Since the S-parameters can be fully defined by two terms at each wavelength, the magnitude and phase, the model is developed to predict both magnitude and phase of S-parameters using polynomial regression.

A third order polynomial in the design parameters, namely width (W), thickness (T), and operating wavelength (λ) issued to, model the S-parameter magnitude variations with width and thickness. The S-parameter magnitude model has an R^2 of 0.969 and 0.993 for width and thickness variation models, respectively. However, a fifth order polynomial is needed for the S-parameter phase variations, to achieve an R^2 of 0.99 for both width and thickness variations. Figure 3-7 shows the S_{13} comparison for the magnitude and phase using FDTD simulations directly, and using the developed compact models, when the width is varying. For the training set, the model has a difference from the direct simulation value for magnitude of 0.0009 and 0.05 rad for phase; however, for the test set these differences are 0.0007 and 0.09 rad for magnitude and phase, respectively.

The S-parameters generated from the variation-aware compact model can then be used in the photonic circuit simulator to simulate a whole CROW consisting of

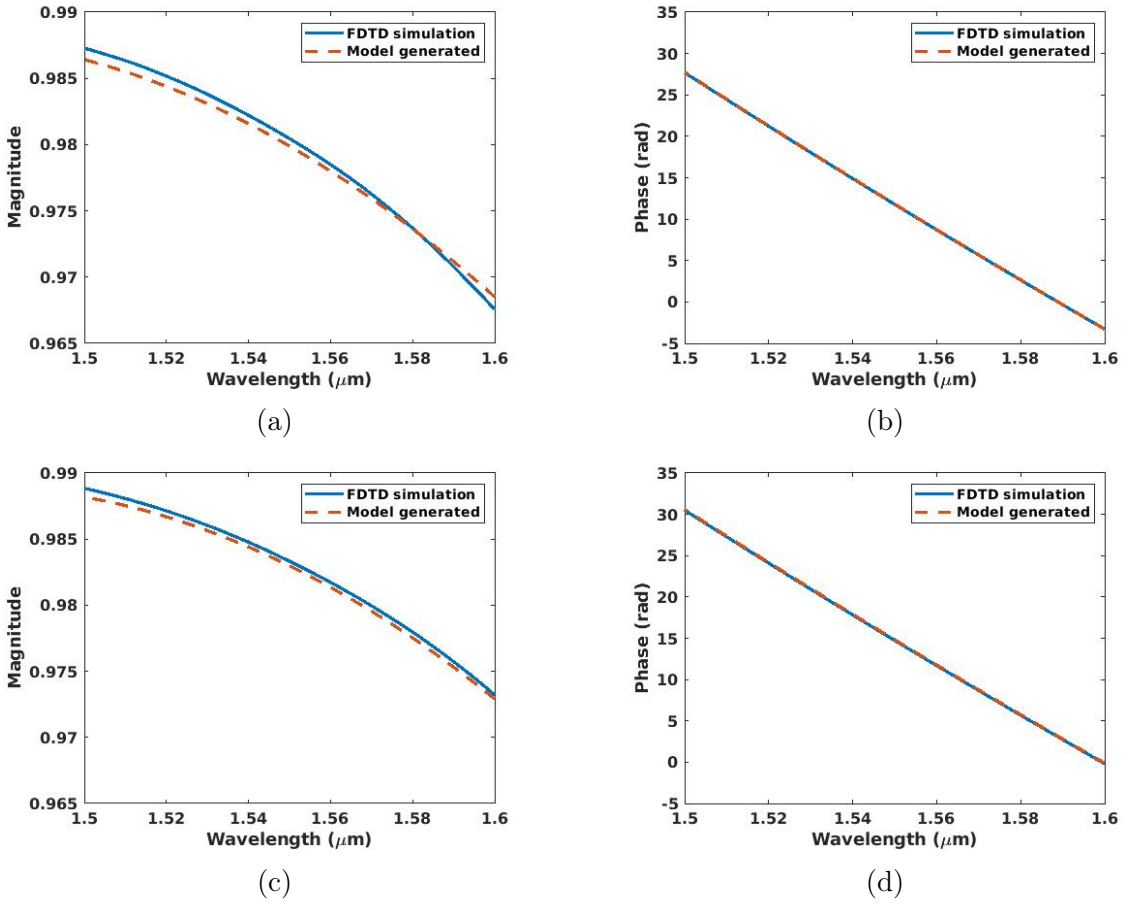


Figure 3-7: S_{13} for the half ring using both FDTD simulations and the developed model. (a) Magnitude in case of training with $W = 491\ nm$ and $T = 220\ nm$. (b) Phase in case of training with $W = 491\ nm$ and $T = 220\ nm$. (c) Magnitude in case of testing with $W = 504\ nm$ and $T = 220\ nm$. (d) Phase in case of testing with $W = 504\ nm$ and $T = 220\ nm$.

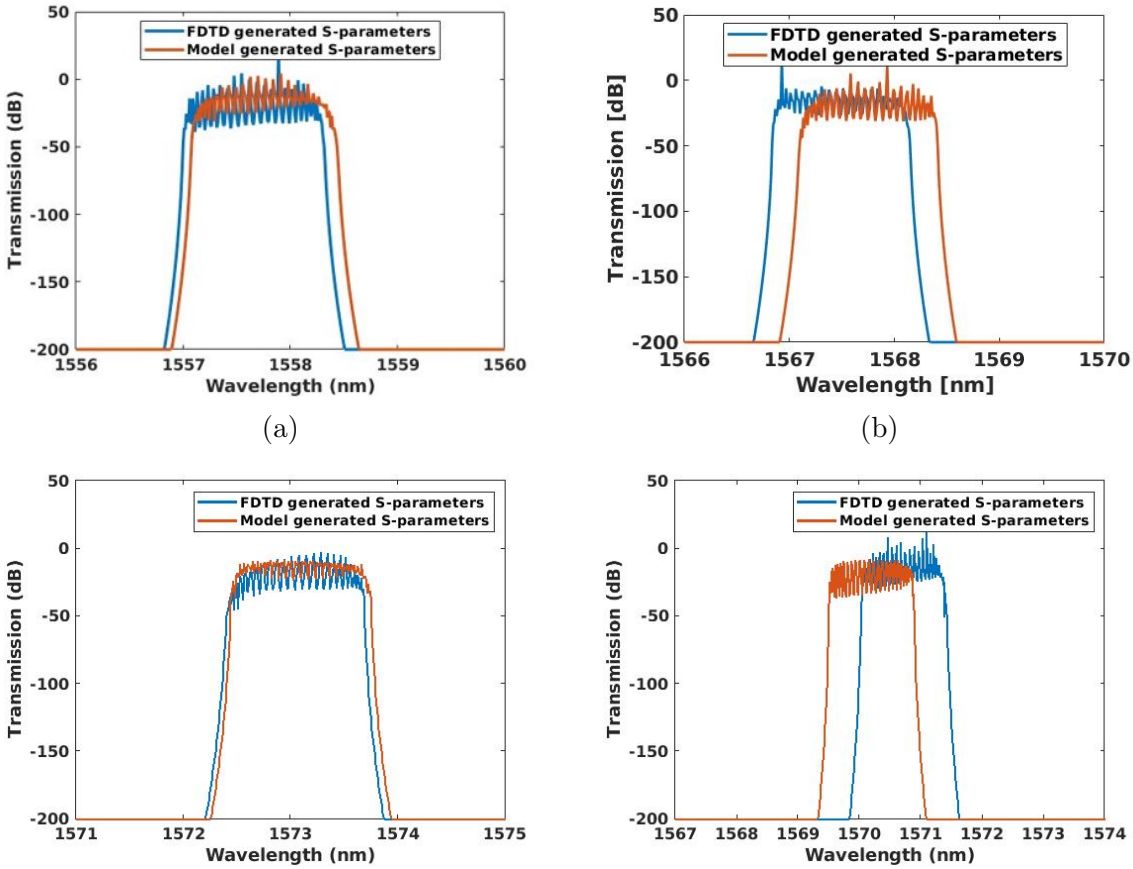


Figure 3-8: Behavior of CROW consisting of 28 rings simulated using the S-parameters generated from the FDTD simulations and generated by the compact model. (a) Training data with $W = 491\ nm$ and $T = 220\ nm$, (b) testing data with $W = 504\ nm$ and $T = 220\ nm$, (c) training data with $W = 500\ nm$ and $T = 225\ nm$, and (d) testing data with $W = 500\ nm$ and $T = 218\ nm$.

any number of rings, by connecting these compact model-based components together as indicated in Fig. 3-6(b). Figure 3-8 shows a comparison between the simulated performance of a 28 ring CROW using S-parameters generated directly from FDTD simulation, and S-parameters generated using the developed variation-aware compact model. As seen in Fig. 3-8, both result in similar passband transmissions (our most important parameter) and bandwidth for both the training and testing data. There is an offset in passband center wavelength of value $\Delta\lambda = 0.07\ nm$ for training data with $W = 491\ nm$ and $T = 220\ nm$ while $\Delta\lambda = 0.05\ nm$ for training data with $W = 500\ nm$ and $T = 225\ nm$. However, for the testing data, $\Delta\lambda = 0.24\ nm$ for the test data with $W = 504\ nm$ and $T = 220\ nm$ is observed, and $\Delta\lambda$ is about $0.5\ nm$ for the testing data with $W = 500\ nm$ and $T = 218\ nm$. The observed offset in the response of the CROW between the case when S-parameters are generated directly from FDTD simulation, and S-parameters are generated using the developed variation-aware compact model are random depending on different W and T values. These offsets can be attributed to the difference in the magnitude and phase response for S_{ij}, as shown in Fig. 3-7. Moreover, such offsets are modest, and worth the speedups made possible compared to the case when we use lower mesh (grid density) in FDTD simulations for speeding simulations, as shown in Fig. 3-9, where the yellow response corresponds to a lower mesh accuracy (Mesh accuracy 6) to avoid the highest accuracy (Mesh accuracy 8) which is computationally expensive (shown in blue). From this result, we see that the fitting errors introduced by the variation-aware compact model is smaller than errors that would typically arise from using a less accurate FDTD simulation. The compact model generates the S-parameters in a few seconds, while each FDTD-based S-parameter model requires $\sim$90 minutes to generate on similar hardware.

3.2.2 Model Applications

This S-parameter based parameterized compact models can be used to facilitate and speed up design optimization, and to run Monte Carlo simulations for variations analysis and yield prediction. Having a model that generate S-parameters much

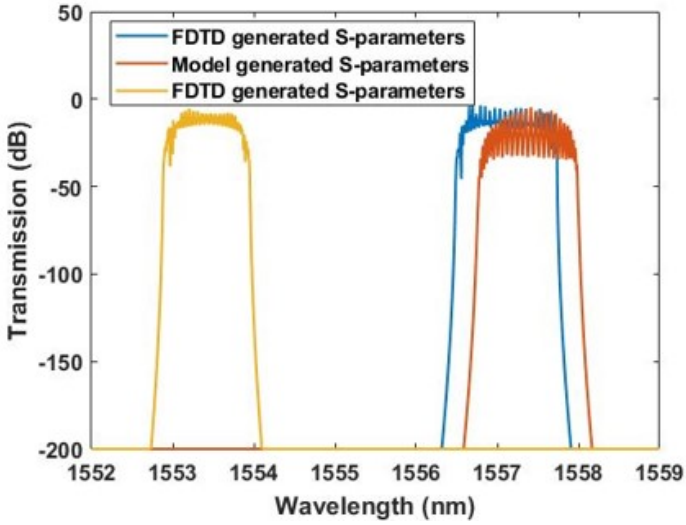
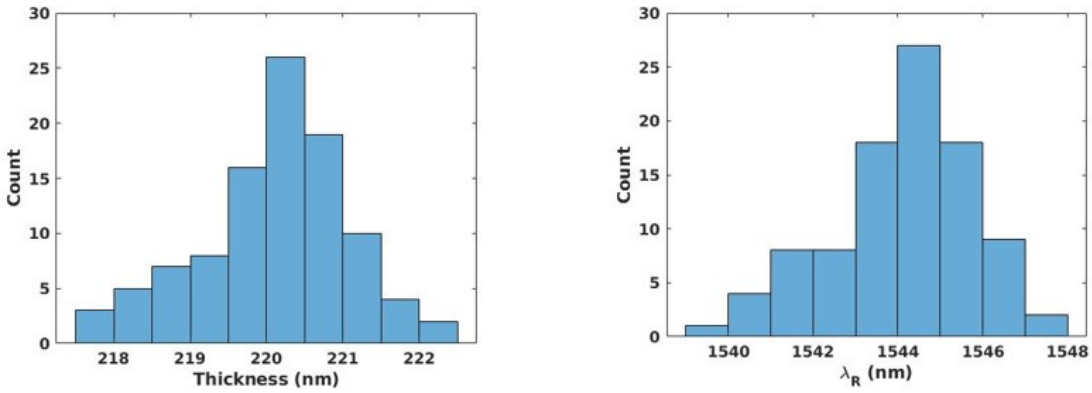


Figure 3-9: Behavior of CROW consisting of 28 rings simulated using the S-parameters generated from the FDTD simulations with different simulation accuracy and generated by the compact model with $W = 504\ nm$ and $T = 220\ nm$. The yellow response corresponds to a low simulation accuracy (Mesh accuracy 6).

faster than the direct simulations enables us to run many simulations in a very short time, making Monte Carlo simulation feasible using these models.

In our first application scenario, we use Monte Carlo analysis to examine when all of the rings in the CROW experience the same variation. The compact model is used to construct the S-parameters for the circuit level simulation, given random sample values for width and thickness. The compact model generates S-parameters for 100 half rings, each with a randomly sampled thickness as shown in the histogram in Fig. 3-10(a), assuming normally distributed thickness, $T = 220\ nm$ nominal, with a standard deviation 1 nm. Then, each S-parameter is used to build a 28 ring CROW, all with the same value for T. Accordingly, we have 100 different CROWs, each with a randomly sampled thickness. We then perform Lumerical INTERCONNECT simulation for the corresponding CROW. Figure 3-10(b) shows the resulting variation in resonance wavelength passband (λ_R) across these 100 different CROWs. A mean λ_R of 1544.2 nm and standard deviation of 1.4 nm is observed.

Another interesting application of these models is predicting the yield when the rings within the CROW experience different but spatially correlated variations in



(a) (b)

Figure 3-10: (a) Distribution of 100 ring thickness with variation around the nominal value $T = 220\ nm$ and a standard deviation of 1 nm. (b) Resulting resonance wavelength variation for 28 ring CROWs with these silicon thickness variations.

waveguide width or thickness that depend on the location within the die, defined as a function of the spatial correlation length (L_T) and amplitude (σ) as in Fig. 3-11.

As mentioned earlier, when the geometry of the rings within a CROW become slightly different, this makes the neighbouring rings lose their resonance alignment, and can suppress the resonance such that the response of the CROW becomes below usable levels as shown in Fig. 3-4. Accordingly, predicting the yield (the number of working CROWs) given the specific foundry width and thickness variation control (in terms of L_T and σ) becomes crucial. Variation impact and yield gives insight into the potential tuning power needed to have an acceptable yield, or to specify what foundry control level must be, i.e., the spatial variation correlation lengths, in order to have acceptable yield.

For yield prediction, we run 100 different instantiations of spatially correlated width and thickness process maps. Figure 3-11 shows an example of width and thickness spatial maps (along the length of the CROW), with geometric values for each ring. According to each, we generate S-parameters for each ring based on its local thickness and width values within the process spatial map. We consider working CROWs to be those where the pass band transmission is more than -20 dB. Yield simulations in response to width and thickness characteristic variations are shown Fig. 3-12, as a function of spatial correlation length (L_T) and amplitude (σ). From Fig. 3-12, we

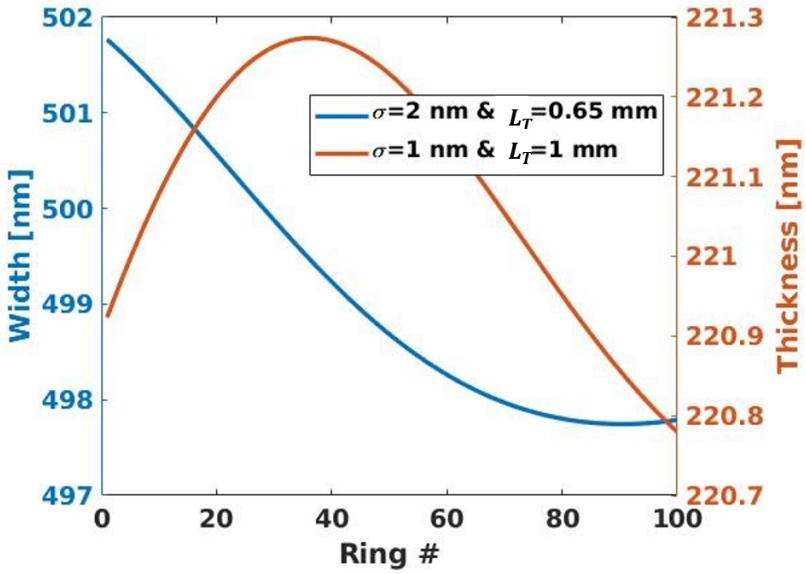


Figure 3-11: Spatially correlated variation map for a single instantiation of a CROW consisting of 100 rings.

can see that as the CROW length increases (more rings form the CROW), it becomes more sensitive to correlation length and amplitude of the spatial variation, because longer variations are sensed by the CROW. Also, we can see that the CROW is more sensitive to the spatial variations in thickness, since for the same amplitude or correlation length, the thickness variation (which is less than the width variation) causes a higher variation percentage in the yield.

3.3 Random Variations

In this section, we consider an important random process variation, specifically LER, and its impact on CROWs. We examine the case when LER is applied to the half ring, and evaluate how this can affect the response of the CROW and its passband.

3.3.1 Results

Following the procedure described in Chapter 2, we generate 30 different instantiations for the half ring geometry with LER applied to it using virtual fabrication with SEMulator3D [29] and, repeated for seven different combinations of A and L_c values.

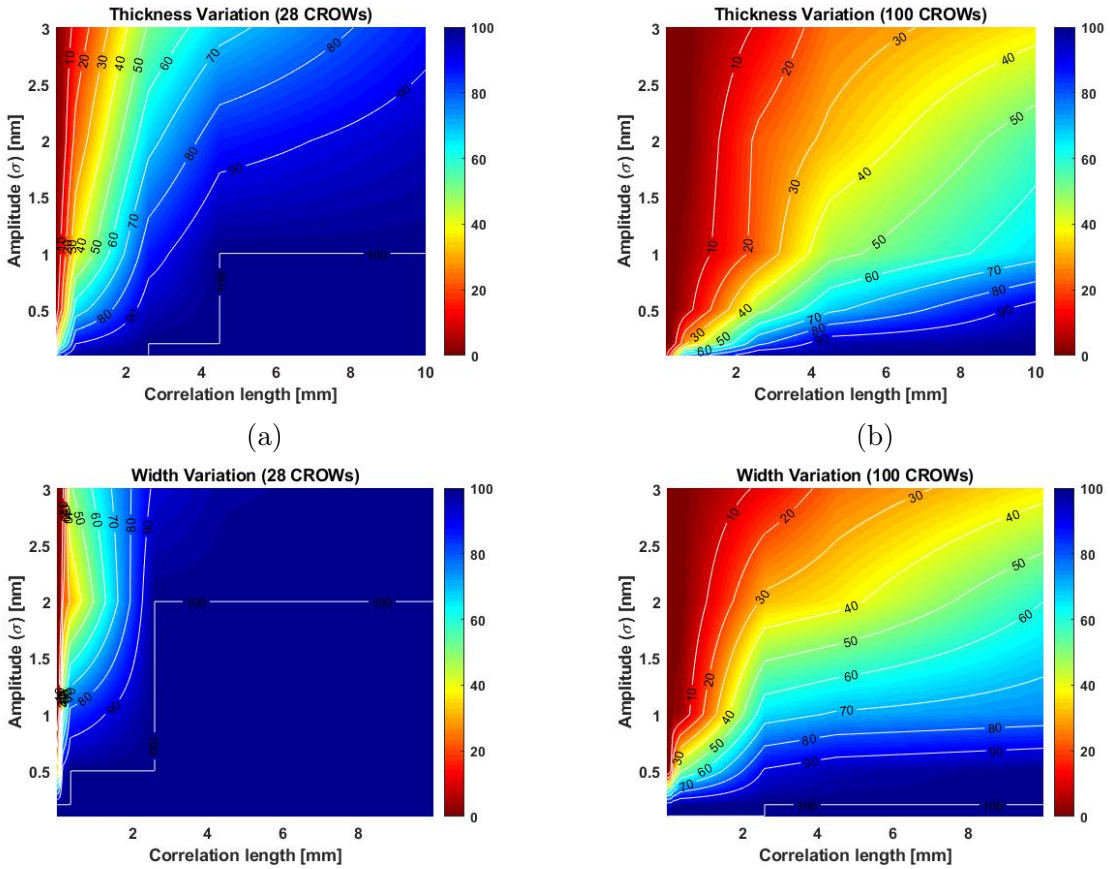


Figure 3-12: Yield (%) as a function of spatial variation correlation length (L_T) and amplitude (σ). (a) CROW composed of 28 rings experiencing spatial variations in silicon thickness, (b) CROW composed of 100 rings experiencing spatial variations in silicon thickness, (c) CROW composed of 28 rings experiencing spatial variations in silicon width, and (d) CROW composed of 100 rings experiencing spatial variations in silicon width. Yield is strongly impacted for longer 100 ring CROWs (right).

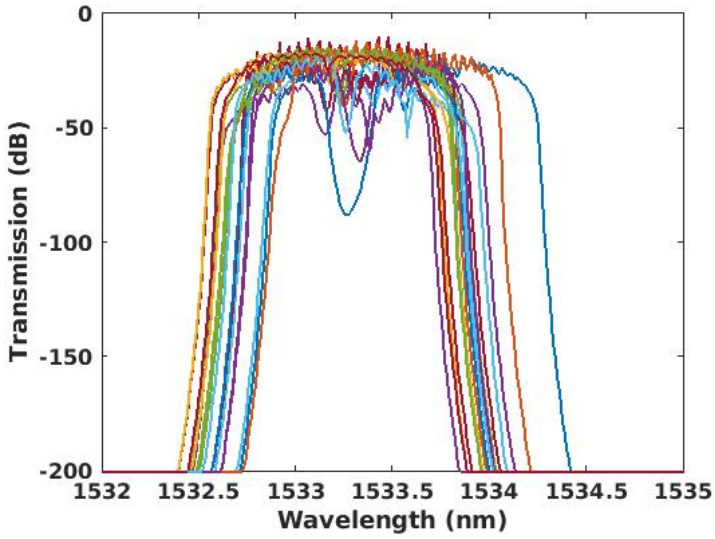
When LER is applied to the CROW rings, there are two potential types of responses, as shown in Fig. 3-13. The first case is shown in Fig. 3-13(a); for small LER A and L_c, we see that LER, which is a random process variation, causes randomness everywhere. Randomness in the resonance position is observed, where for each different instantiation, we see that the center of the passband is different. Also, the shape of the response is different, such that there are dips in the transmission with levels that vary from one instantiation to another. Moreover, the passband width is different for each instantiation. The second type of response happens for large values of LER amplitude and correlation length, where the LER causes the resonance to be largely suppressed, as in Fig. 3-13(b).

We focus on the small LER case, and seek to build variation-aware compact models that capture each of the observed effects. A first step in modeling is pre-processing of the data. We start by shifting all of the thirty instantiations for each A and L_c combination, such that all of them have the same starting passband resonance wavelength, as shown in Fig. 3-14. When doing this shift, we note that the passband width varies from one instantiation to another, and this shift is progressive with the wavelength as seen in Fig. 3-15. Accordingly, the shift is not constant for all resonance modes, and a mode-wise shift is applied.

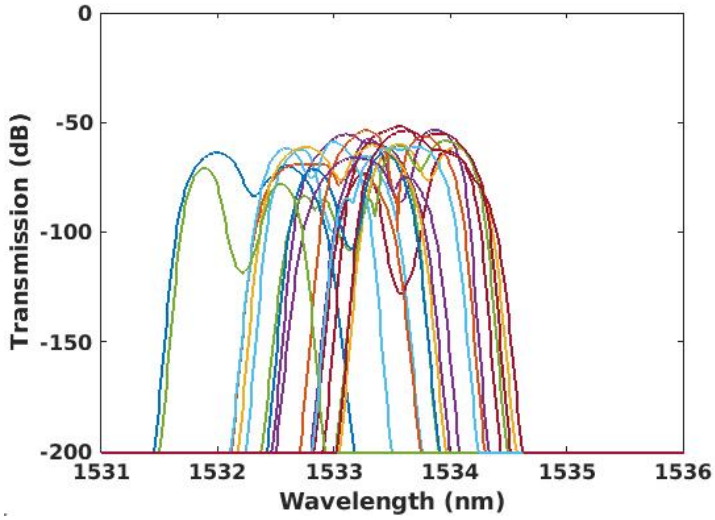
3.3.2 Models

Our variation-aware compact models are S-parameter based models; thus, we build models for both the S-parameter magnitude and phase, as a function of the process variation (LER amplitude and correlation length) and the operating wavelength (λ), such that the models are $S_{ij}(A, L_c, \lambda)$ where S_{ij} is the S-parameter for input port i and output port j, where both i and j can take values 1 through 4.

The real challenge is how to include all types of randomness in the response in the models as discussed in Section 3.3.1. A key step in modeling is choosing a single instantiation, which we call the "base instantiation," to be a representative of the behavior for each one of the LER amplitude and correlation length combinations. We then split our DOE simulation data into training and test sets with a split ratio of



(a)



(b)

Figure 3-13: Response of a 28 ring CROW when LER is applied. Different instantiations are displayed for (a) $A = 4\ nm$ and $L_c = 30\ nm$, and (b) $A = 7\ nm$ and $L_c = 120\ nm$.

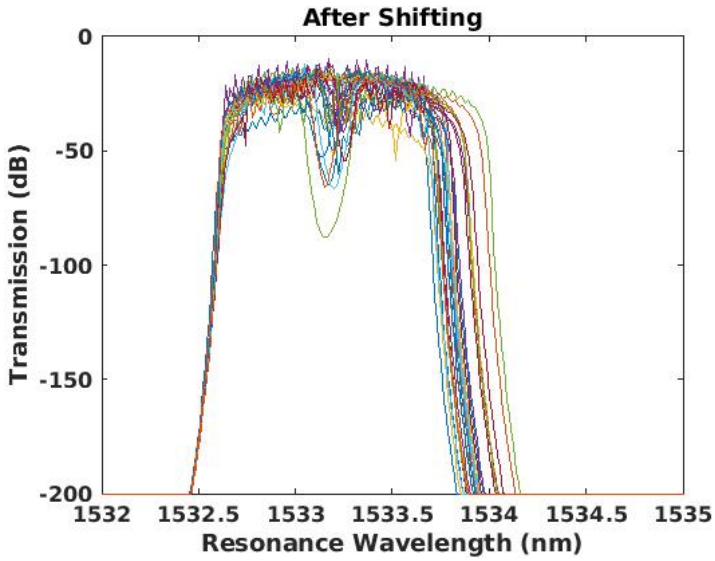


Figure 3-14: Thirty instantiations for LER with $A = 4\ nm$ and $L_c = 30\ nm$, after shift to start at the same wavelength.

70/30, respectively.

We build polynomial regression models for each of the magnitude and phase for each S_{ij}. The R^2 for all of the S-parameter models for phase and magnitude is at least 0.96. The models developed for the S-parameters are then used to generate a new full CROW instantiation, and compare it to simulation based S-parameters, as shown in Fig. 3-16.

As seen from Fig. 3-16, the model fails to capture the location of the dip. It is also difficult to generate new instantiations that mimic the range of variations that we see in Fig. 3-13. Accordingly, some post-processing for the models is required to be done to consider all the aforementioned types of randomness in the observed behavior in full simulations.

The first step for post-processing is to account for the randomness in the shift in resonance location for different instantiations with respect to the base instantiation (that we used in building the model). This shift is a mode-wise shift, meaning that the shift between instantiation i and instantiation j is not constant at every mode rather, it is mode dependent and random. Accordingly, this shift in the resonance can be decomposed into two parts: the random shift between instantiations which, we de-

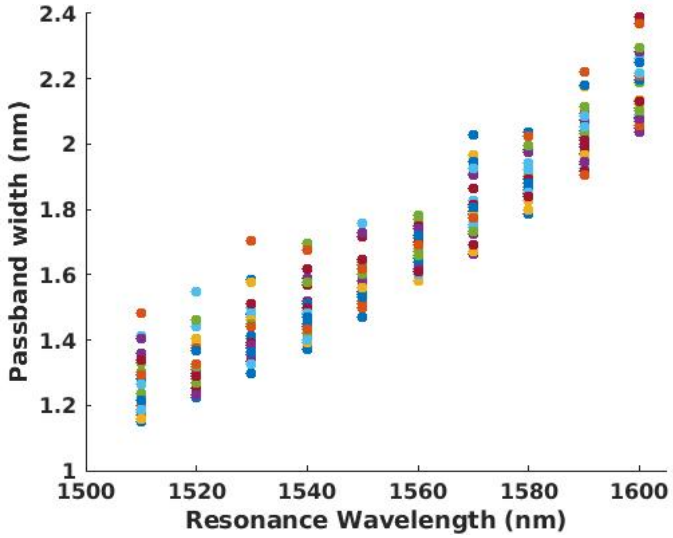


Figure 3-15: Passband width for the thirty different instantiations across the wavelength of interest (1500-1600 nm), showing the different resonance modes, for LER $A = 4\ nm$ and $L_c = 30\ nm$.

note as $N_{shift}(\mu_s, \sigma_s)$, and randomness within the instantiation $N_{shift/int}(\mu_{si}, \sigma_{si})$, i.e. within modes. The means and standard deviation are calculated from the ensemble of thirty instantiations. Figure 3-17(a) shows the wavelength shift between the base instantiation and the rest across the different modes from the ensemble simulations. Figure 3-17(b) shows the wavelength shift generated using the calculated means (μ_s and μ_{si}) and standard deviations (σ_s and σ_{si}).

When this extra post-processing step is added to the models of the S-parameters and the models used to generate LER instantiations, the result we obtain is shown in Fig. 3-18. Although the above post-processing step helps the model capture the change in the passband location due to LER, we can see that the model still fails to capture the change in the response shape (dip locations, levels and change in the passband width). Accordingly, we introduce another post-processing step to account for this. In this second post-processing step, we model the difference in the S-parameters between the base instantiation and the rest of the ensemble instantiations. This helps us capture the difference in the response (which is embedded in the difference between the S-parameters) in addition to the previously captured shift. This difference is random across the wavelength and has a mean ($\mu_d(\lambda)$) and a standard deviation($\sigma_d(\lambda)$),

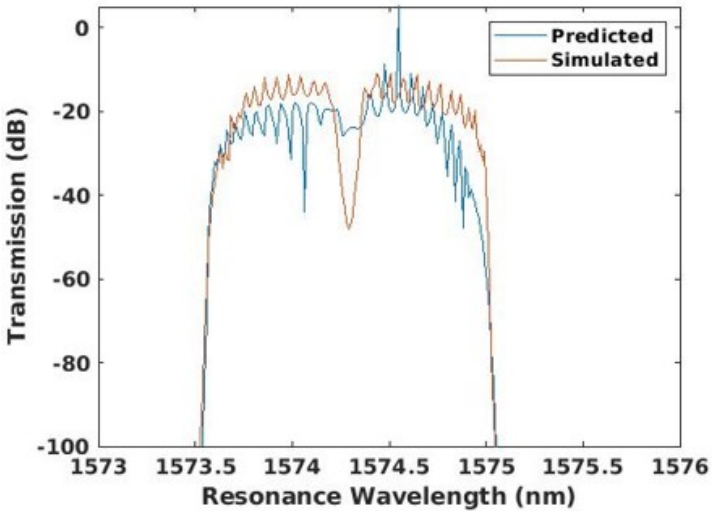


Figure 3-16: Comparison of the response behavior for a 28 ring CROW experiencing LER ($A = 4\ nm$ and $L_c = 30\ nm$) when S-parameters are generated using either virtual fabrications and FDTD simulations (simulated) or variation-aware compact models (predicted).

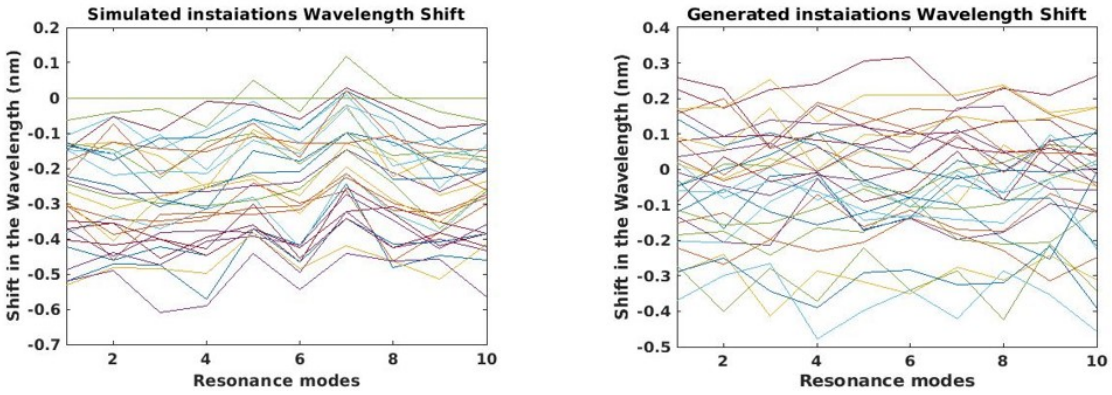
which we denote as $N_{diff}(\mu_d(\lambda), \sigma_d(\lambda))$. Figure 3-19(a) shows the difference between the base instantiation S-parameters and others calculated for the simulated S_{13} magnitude. Figure 3-19(b) shows this difference for S_{13} magnitude generated using the random distribution $N_{diff}(\mu_d(\lambda), \sigma_d(\lambda))$, where $\mu_d(\lambda)$ and $\sigma_d(\lambda)$ are calculated using the ensemble of thirty instantiations.

Now to generate an instantiation, we do the two post-processing steps and accordingly the model for any of the S-parameters becomes:

$$S^*_{ij}|_{inst}(A, L_c, \lambda) = S_{ij}|_{pred}(A, L_c, \lambda) + N_{shift}(\mu_s, \sigma_s) + N_{shift/int}(\mu_{si}, \sigma_{si}) + N_{diff}(\mu_d(\lambda), \sigma_d(\lambda)) \tag{3.1}$$

where $S_{ij}|_{pred}(A, L_c, \lambda)$ is the model developed using the base instantiations without any processing, the terms $N_{shift}(\mu_s, \sigma_s)$and $N_{shift/i}(\mu_{si}, \sigma_{si})$ account for the shift in the passband location, and finally the term $N_{diff}(\mu_d(\lambda), \sigma_d(\lambda))$ accounts for the across wavelength difference in the response (S-parameter magnitude and phase) between the base instantiation and the other instantiations.

Using the model in Eq. 3.1 to generate different instantiations for S-parameters of



(a) (b)

Figure 3-17: Shift in the resonance location compared to the base instantiation for LER with $A = 4\ nm$ and $L_c = 30\ nm$ as, (a) calculated from the simulated instantiations we have, and (b) generated using the calculated means and standard deviations for between-instantiation shift and across instantiations shift.

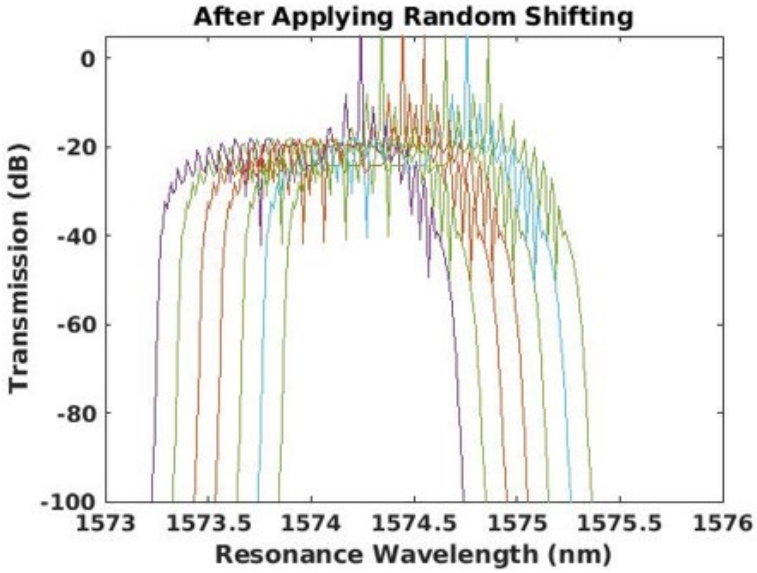
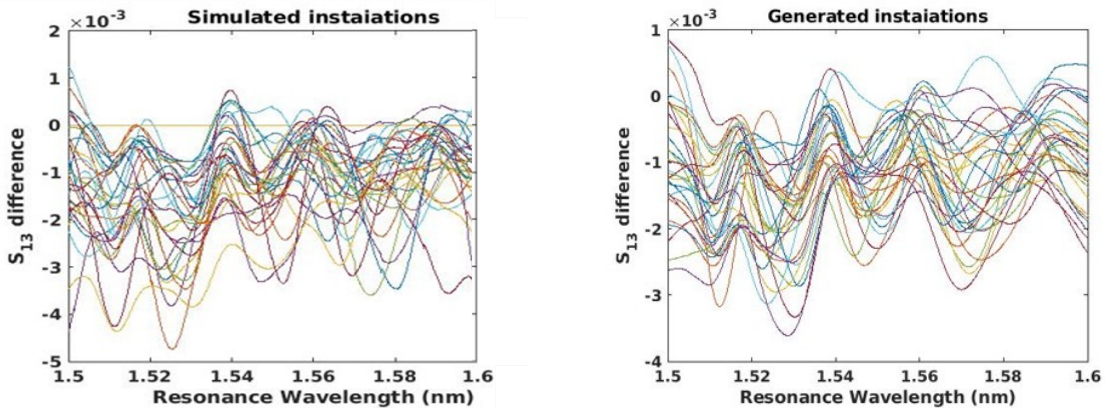


Figure 3-18: Response for a 28 ring CROW with generated S-parameter instantiations using the developed model, and after applying the shifting post-processing step, with LER of $A = 4\ nm$ and $L_c = 30\ nm$.



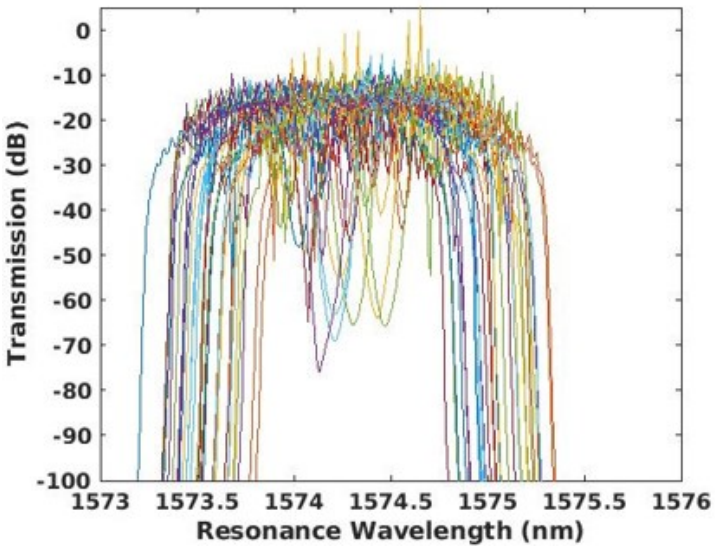
(a) (b)

Figure 3-19: Difference in the magnitude of S_{13} as compared to the base instantiation for LER with $A = 4\ nm$ and $L_c = 30\ nm$. (a) Calculated from the simulated instantiations we have, (b) generated using the calculated means and standard deviations for the difference using the ensemble.

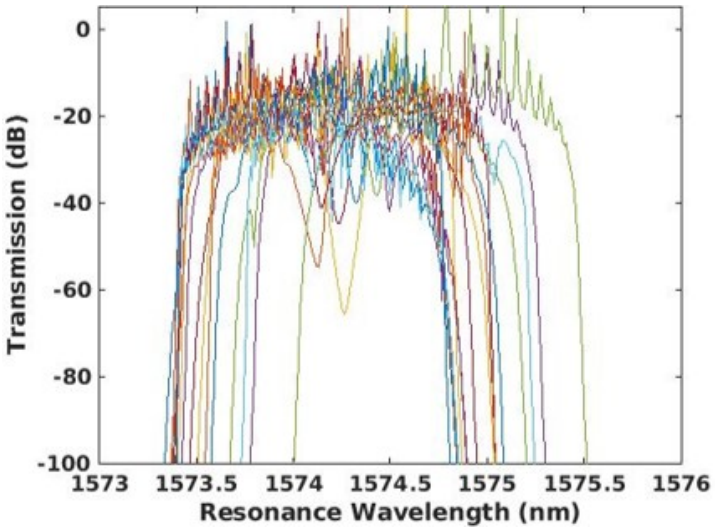
the half ring with LER applied to its side walls, we get the response shown in Fig. 3-20(b). When comparing the response between the different instantiations when the S-parameters are directly simulated in FDTD to the case when the S-parameters are generated based on the model defined by Eq. 3.1, we see that the models capture to a good extent representative response randomness in passband width, passband location as well as the dip locations and levels.

3.3.3 Model Application

The models developed for the half ring S-parameters in the presence of LER can be used to predict the yield under specific foundry LER parameters (the amplitude and correlation length). We use the models developed to generate half ring S-parameters for 100 instantiations for different amplitudes and correlation length combinations. These S-parameters are then used to build a 28 ring CROW (by connecting the half rings together) and calculate the yield, i.e., the fraction of working CROWs. We consider any CROW that has a transmission level less than -20 dB or has a dip in the transmission that goes lower than -50 dB to be a defective CROW. Figure 3-21 shows the calculated yield. We can see that LER has a serious effect on the

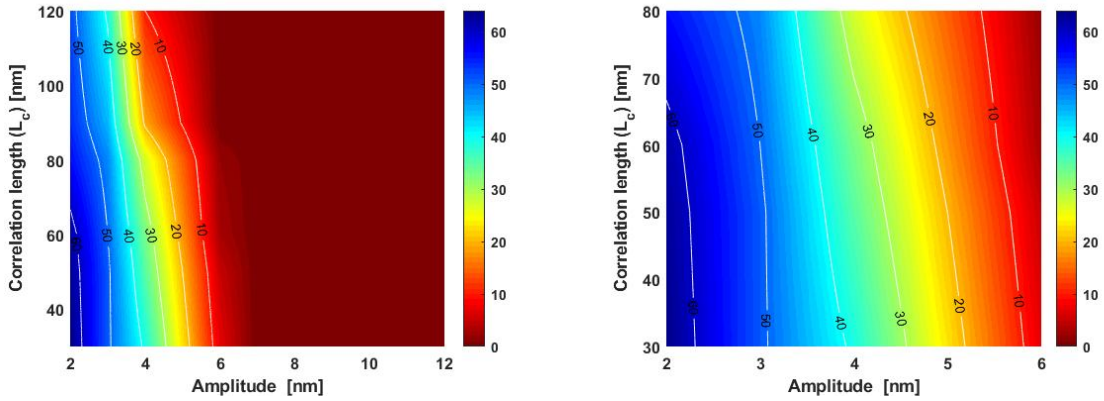


(a)



(b)

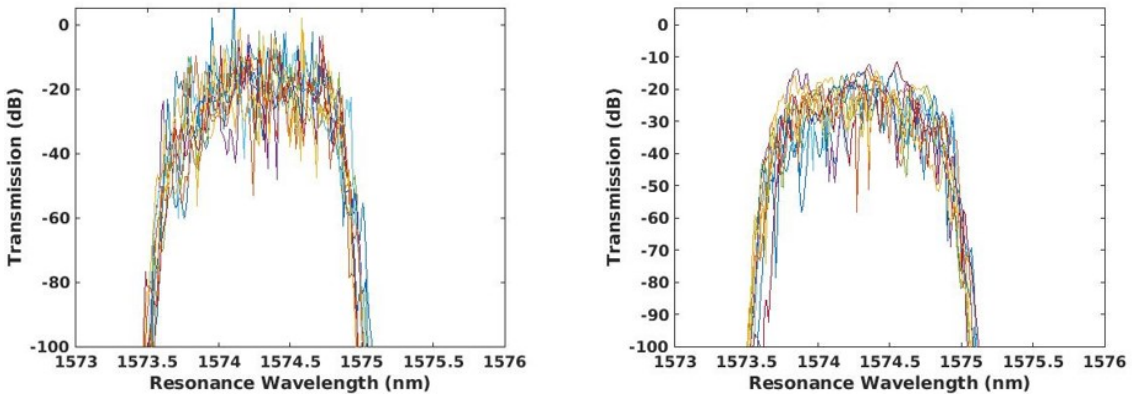
Figure 3-20: Response for a 28 ring CROW with LER of $A = 4\ nm$ and $L_c = 30\ nm$ at different instantiations when: (a) S-parameters are generated directly from the FDTD simulations, and (b) S-parameters are generated based on the model in Eq. 3.1.



(a) (b)

Figure 3-21: CROW yield (%) calculated when LER is applied to the half ring (CROW constitute parameters), (a) for different values of amplitude and correlation length, and (b) zoomed view on the working range of amplitude and correlation length.

CROW performance, and that the amplitude rather than the correlation length is more impactful, where, for $A > 6\ nm$ the yield is almost always zero. This is an expected behavior, because substantial LER will cause the geometries of the rings to be different from one ring to another, as it changes the geometry of the half ring. In particular, the coupling gap deviates from the intended design value ($C_g = 200\ nm$), and can be different for each ring.



(a) (b)

Figure 3-22: Response of a 28 ring CROW with each ring experiencing a different instantiation for LER of $A = 4\ nm$ and $L_c = 30\ nm$ when S-parameters are generated using: (a) compact model, and (b) FDTD simulations.

We further use the S-parameters generated from the model to simulate 28 ring

CROWs, with each ring experiencing a different LER instantiation, as shown in Fig. 3-22(a). Comparing it with the case of building CROWs using simulation generated S-parameters, Fig 3-22(b), we can see that both agree well.

3.4 Summary

We present S-parameter based compact models for a constituent half ring coupler geometry. These models can be used to simulate larger circuits (full CROWs) and predict their performance in the presence of process variations, either geometric or random (LER). In addition to the prediction capability of these models, they also enable fast approximate simulations, where a simulation can be done directly in the circuit simulator in a few seconds, compared to hours needed using traditional simulations using device and circuit level simulators. The compact model demonstrated can serve as another building block for variation-aware process design kits (PDKs) for photonics. Such models enable silicon photonic designers to run Monte Carlo simulations to predict performance and yield as shown in this chapter, of their silicon photonic devices and circuits given the variation characteristics of a specific foundry.

Chapter 4

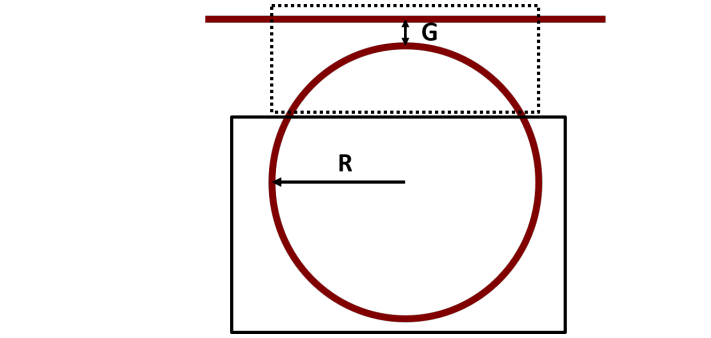
Large Silicon Nitride Ring Resonator Modeling

Silicon nitride based ring resonators are crucial for many applications such as optical clocks [45], frequency combs [46, 47], and metrology and sensing [48]. Accordingly, when using silicon nitride (Si_3N_4) rather than silicon (Si), a lower index difference ($\Delta n \sim 0.5$) results; this low Δn offers better tolerance for fabrication variations.

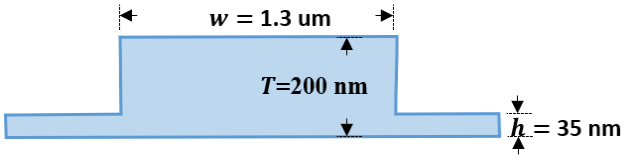
Moreover, the ring resonators are usually challenging devices in full physical simulation, because the ring simulation usually takes a long time for the signal to fully propagate in the device. Accordingly, FDTD ring simulations are time consuming, and rings are usually simulated by partitioning the ring into smaller parts. Each part takes less time; however, when the rings have large radii ($R \geq 10\ \mu m$), even the coupling region section simulation still takes substantial time. So, designers in many cases resort to fabricating and measuring performance rather than simulating many structures, examining many different designs in order to choose the ring parameters that meet required specifications.

In this chapter, we consider an all pass Si_3N_4 rib ring as in Fig. 4-1(a-b). We seek to speed up the simulation and optimization by proposing a methodology to develop compact models that will help the designer predict the performance for different design parameters of the silicon nitride ring. In addition, once we have fabricated rings, we can use measurements to tune the model to the particular fabrication process. We can

also use these measurements to extract geometric variations that the rings experience during fabrication.



(a)



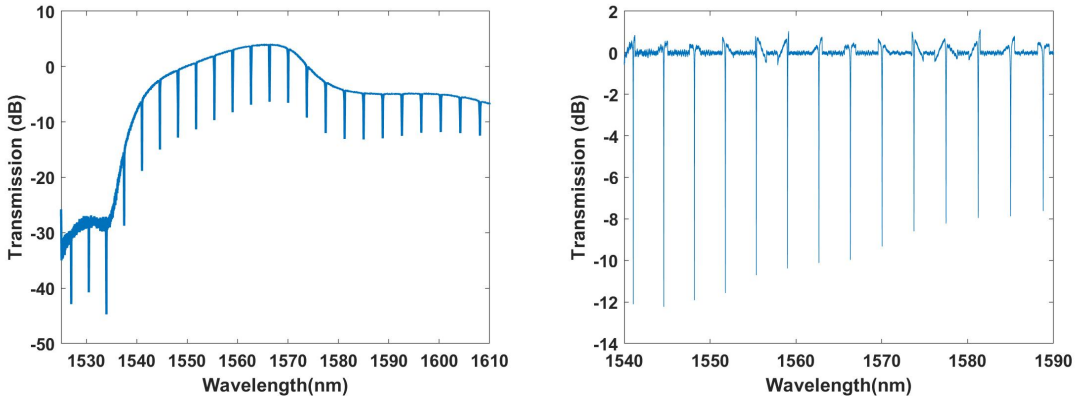
(b)

(c)

Figure 4-1: (a) The silicon nitride all pass ring used, (b) geometry of the rib silicon nitride ring fabricated, (c) layout of the different fabricated rings.

4.1 Fabrication and Measurements[1]

Different all pass rib silicon nitride rings as shown in Fig. 4-1(a-b) are designed, with radii (R) ranging from 20 μm to 160 μm and gaps (G) ranging from 0.2 μm to 1.6 μm. We have a total of 48 unique rings fabricated on a 6 inch wafer as shown in Fig. 4-1(c). The insertion loss (IL) for each of these rings is measured for wavelength ranging from 1525 to 1610 nm with a large number of points (2^{16} points) sufficient to locate the resonance locations (to within 1.297×10^{-3} nm). Figure 4-2(a) shows the insertion loss as directly measured. A first step is to detrend the IL, i.e., remove the effect of the grating coupler used for measurement, resulting in the ring transmission as shown in Fig. 4-2(b).



(a) (b)

Figure 4-2: (a) The insertion loss measured for $R = 60$ μm and $G = 1$ μm. (b) The measured insertion loss for the ring after detrending.

After detrending all 48 rings, some of the rings are found to be non-functional. Specifically, all the rings having $R = 20$ μm and $R = 40$ μm with $G > 1$ μm are non-functional; hence, only 39 rings are working, and used in subsequent analysis below.

[1]These rings are designed, fabricated, and measured by Carlos A Rios Ocampo at MIT. The analysis in this chapter is in collaboration with Carlos A Rios Ocampo and Zhengxing Zhang.

4.1.1 Features Extraction

For each of the working rings, we extract features from the measured data, including free spectral range (FSR), effective index (n_{eff}), group index (n_g), resonance wavelength for each mode (λ_{res}), coupling coefficient (t), and loss (α).

We start by calculating the FSR and the resonance locations for each resonance mode. To calculate the effective index, we need to consider the resonance mode, with n_{eff} calculated based on the relation

$$n_{eff} = \frac{M \cdot \lambda_{res}}{L} \tag{4.1}$$

where M is the resonance mode, and L is the ring circumference such that $L = 2\pi R$.

To find M, each ring is simulated in MODE [23] and the simulated n_{eff} near a resonance of 1550 nm is identified. For the given ring geometry and identified n_{eff}, the corresponding M is calculated. This resonance mode M is then used to calculate the n_{eff} for each of the extracted resonances for that ring. The calculated n_{eff} values are shown in Fig. 4-3, where each cross represents an n_{eff} for a specific resonance wavelength for each of the rings. Finally, to choose the n_{eff} values that are closest to the simulation, i.e., to choose the resonance wavelength, we compare the extracted values to the simulated values (solid line drawn in Fig. 4-3). Accordingly, the data points in black stars are the points that we choose to be our effective refractive indices for our rings, with the corresponding wavelength as our chosen λ_{res}.

Now, to calculate the group index (n_g) we use the relation:

$$n_g = \frac{\lambda_{res}^2}{FSR \cdot L} \tag{4.2}$$

where λ_{res} is the resonance wavelength. The calculated n_g is different for each of the rings, with an example for one ring geometry shown in Fig. 4-4.

The final feature that we care to extract is the coupling coefficient (t). However, calculating t is challenging as it is an indirect calculation. To do so, we need to find the extinction ratio (ε) as well as the finess (f), based on [49]. This is because the extinction ratio, as shown in Eq. 4.3, can be calculated from the measurement data

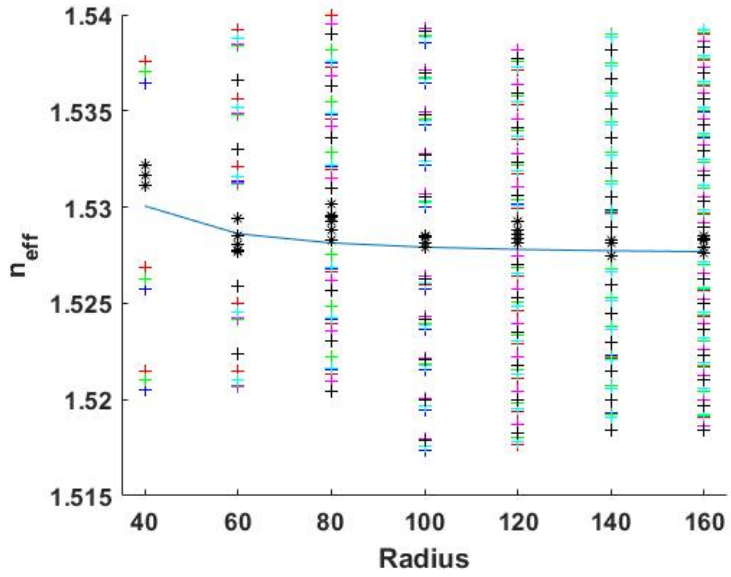


Figure 4-3: The effective refractive indices n_{eff} calculated from the measurement data (+ symbols), compared to the MODE simulated values (line).

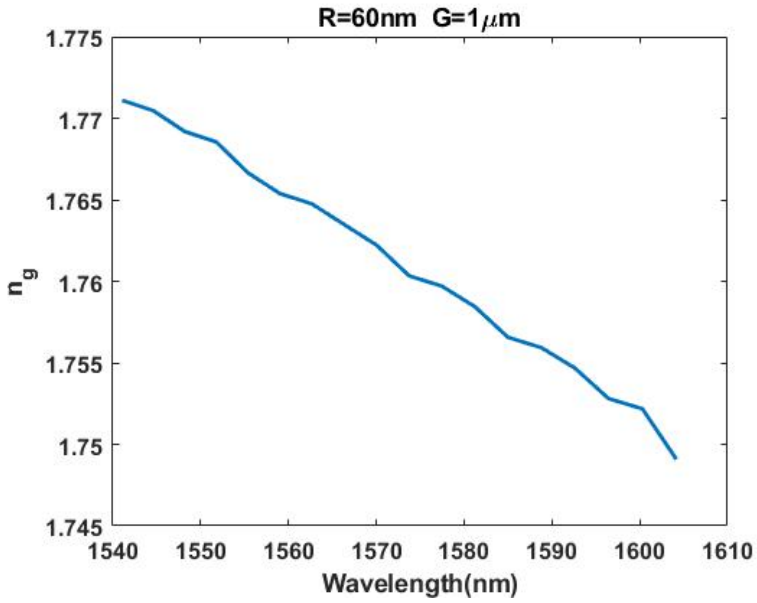


Figure 4-4: Group index calculated from measurement data for a silicon nitride ring having $R = 60$ μm and $G = 1$ μm.

by dividing the maximum and minimum transmissions, but it is also a function of the coupling coefficient (t) as well as the loss (α). Similarly, the finess can be calculated from the measurements by dividing the FSR and the FWHM as in Eq. 4.4, and it can also be expressed in terms of t and α, as in Eq. 4.5:

$$\varepsilon = \frac{T_{max}}{T_{min}} = [\frac{\alpha + t}{\alpha - t}\frac{1 - \alpha t}{1 + \alpha t}]^2 \tag{4.3}$$

$$f = \frac{FSR}{FWHM} \tag{4.4}$$

$$\cos(\frac{\pi}{f}) = \frac{2\alpha t}{1 + \alpha^2 t^2} \tag{4.5}$$

where T_{max} and T_{min} are the maximum and minimum transmission, respectively, and FWHM is the full width half maximum value for the transmission.

Hence, calculating ε and f from the measurements and from equations 4.3 and 4.5 based on solving these two equations, will result in two roots: one for the loss and the other for the coupling [49]. In order to find the coupling coefficient, i.e., decide which of the roots is t, we need to use the fact that, for rings, if the gap is kept constant then the coupling coefficient is supposed to be the same for different radii. Also, if the radius is kept constant, the loss should be constant for different gaps. Accordingly, for each radius, we start by identifying the loss to be the responses that are similar, as in Fig. 4-5, and for each gap, we identify the responses that are similar to be the coupling coefficient, as shown in Fig. 4-6. In this way, the coupling coefficient for each ring is extracted.

4.1.2 Geometric Extraction

Fabricating the structure enables us to infer the geometric spatial variations in silicon nitride width (W), thickness (T) and height (h). To find the spatial maps for these variations, we need to relate the change in the parameter features that we extracted to the change in the geometry [42]. For our geometric variations, we have three values;

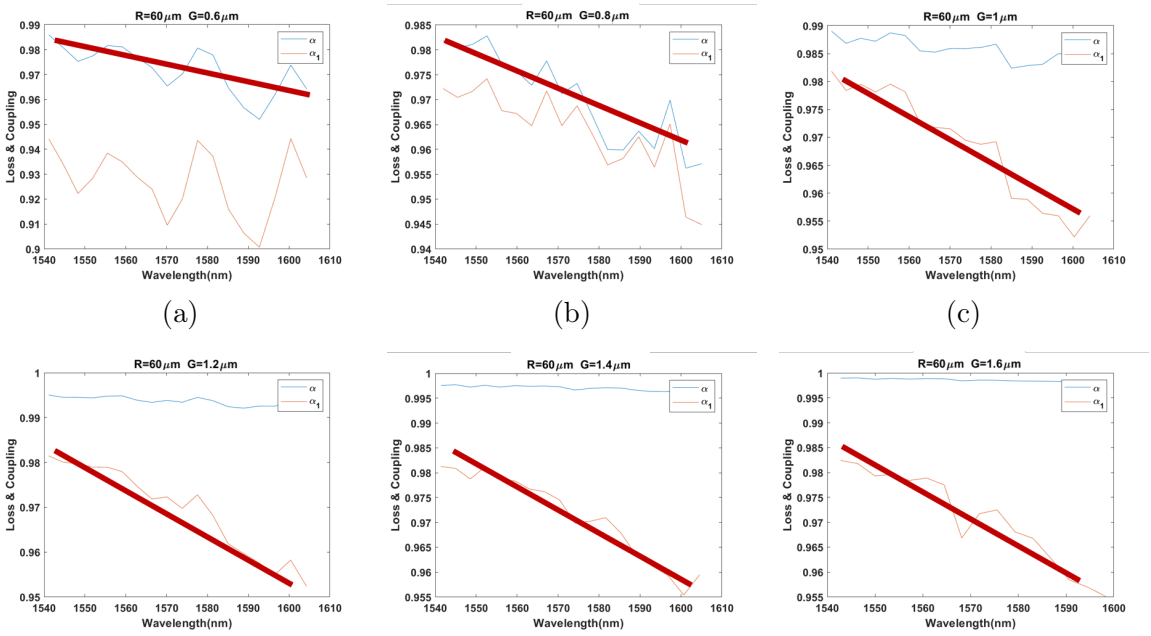


Figure 4-5: Inferring the loss from solving Eq. 4.3 and Eq. 4.5 by using the fact that, for the same radius ($R = 60$ μm in these figures), the loss is the same for different gaps. (a) $G = 0.6$ μm, (b) $G = 0.8$ μm, (c) $G = 1$ μm, (d) $G = 1.2$ μm, (e) $G = 1.4$ μm, (f) $G = 1.6$ μm. The red line highlights the loss α extracted in each case.

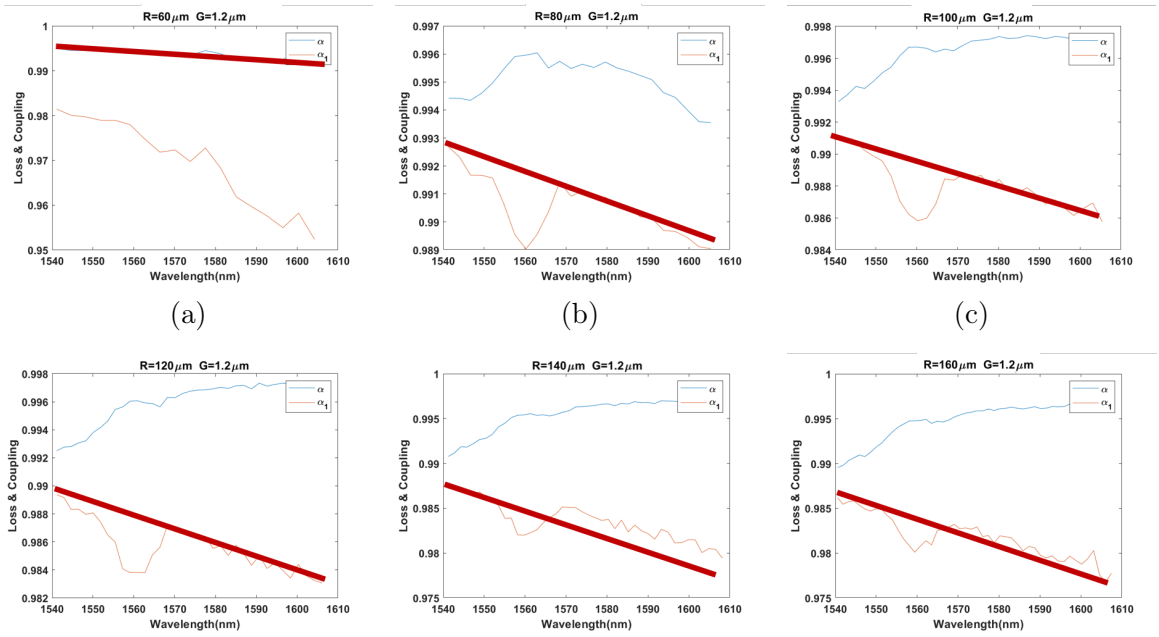


Figure 4-6: Inferring the coupling from solving Eq. 4.3 and Eq. 4.5 by using the fact that, for the same gap ($G = 1.2$ μm in these figures), the coupling coefficient is the same for different radii. (a) $R = 60$ μm, (b) $R = 80$ μm, (c) $R = 100$ μm, (d) $R = 120$ μm, (e) $R = 140$ μm, (f) $R = 160$ μm. The red line highlights coupling coefficient t extracted in each case.

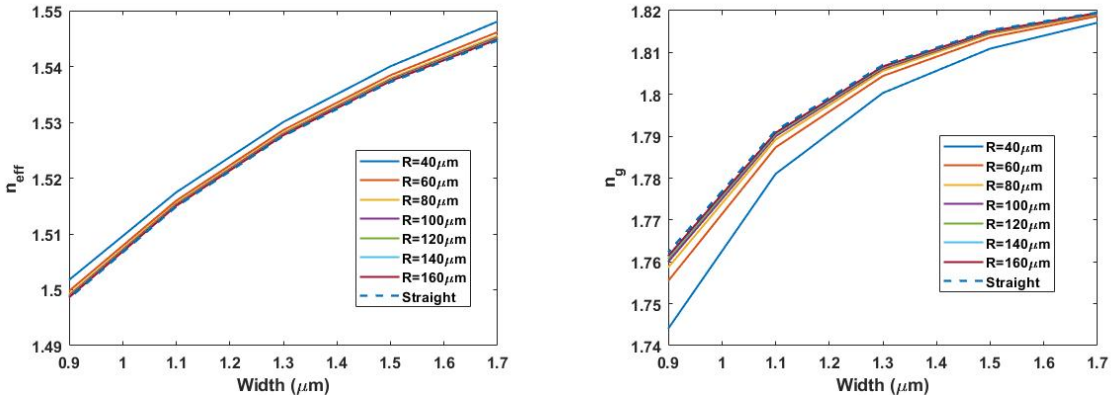
so we need three feature parameters. We relate geometry variations to extracted measurement feature as:

$$\begin{bmatrix} \Delta W \\ \Delta T \\ \Delta h \end{bmatrix} = \begin{bmatrix} \frac{\delta n_g}{\delta W} & \frac{\delta n_g}{\delta T} & \frac{\delta n_g}{\delta h} \\ \frac{\delta \lambda_{res}}{\delta W} & \frac{\delta \lambda_{res}}{\delta T} & \frac{\delta \lambda_{res}}{\delta h} \\ \frac{\delta t}{\delta W} & \frac{\delta t}{\delta T} & \frac{\delta t}{\delta h} \end{bmatrix}^{-1} \begin{bmatrix} \Delta n_g \\ \Delta \lambda_{res} \\ \Delta t \end{bmatrix} \tag{4.6}$$

where Δn_g, $\Delta \lambda_{res}$, and Δt are the differences between the simulation-generated and fabrication-calculated values for group index, resonance wavelength, and coupling coefficient, respectively. Here ΔW, ΔT and Δh are the variation in the ring width, thickness, and height, respectively. In addition $\frac{\delta n_g}{\delta W}$ and $\frac{\delta n_{eff}}{\delta W}$ are the sensitivity of the group and effective indices to the width variations. Similarly, $\frac{\delta n_g}{\delta T}$, $\frac{\delta n_{eff}}{\delta T}$, $\frac{\delta n_g}{\delta h}$ and $\frac{\delta n_{eff}}{\delta h}$ are the sensitivities of group and effective indices to the thickness and height variations. Finally, $\frac{\delta t}{\delta W}$, $\frac{\delta t}{\delta T}$, and $\frac{\delta t}{\delta h}$ are the sensitivities of the coupling coefficient due to variations in width, thickness, and height, respectively.

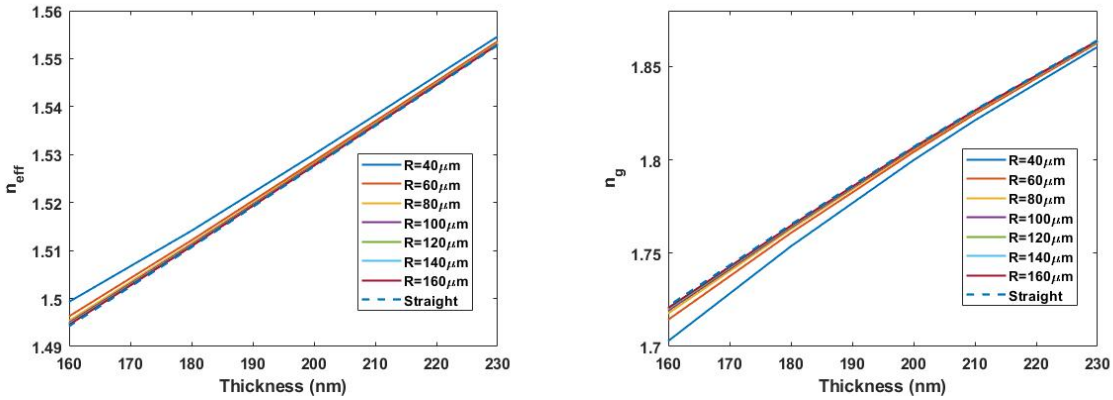
To calculate the sensitivity matrix, i.e., the sensitivities of the features due to variations in the geometry, we use MODE [23] simulations, where we vary the width, thickness or height and calculate the n_{eff} and n_g values at 1550 nm. Figures 4-7 shows the sensitivity for width variation, where this value is different for each radius. Similarly, Figs. 4-8 and 4-9 show the sensitivity at 1550 nm for thickness and height variations, respectively. For coupling sensitivity calculations, FDTD [22] simulations are used to find the coupling for each of the rings, as the coupling is different for each ring and gap value. Figure 4-10 shows the coupling sensitivity for selected rings as simulated with FDTD.

Now, after all of the sensitivity matrix entries are calculated using simulations for each of the silicon nitride rings, we need to find the difference between the designed (simulation) values and the fabrication-extracted values for the effective index, group index and coupling, i.e., find Δn_{eff}, Δn_g, and Δt. Using FDTD simulations for coupling calculation and MODE simulations for effective and group index calculation we find the differences (Δn_{eff}, Δn_g). Figure 4-11 shows the simulated and fabrication extracted values for selected rings.



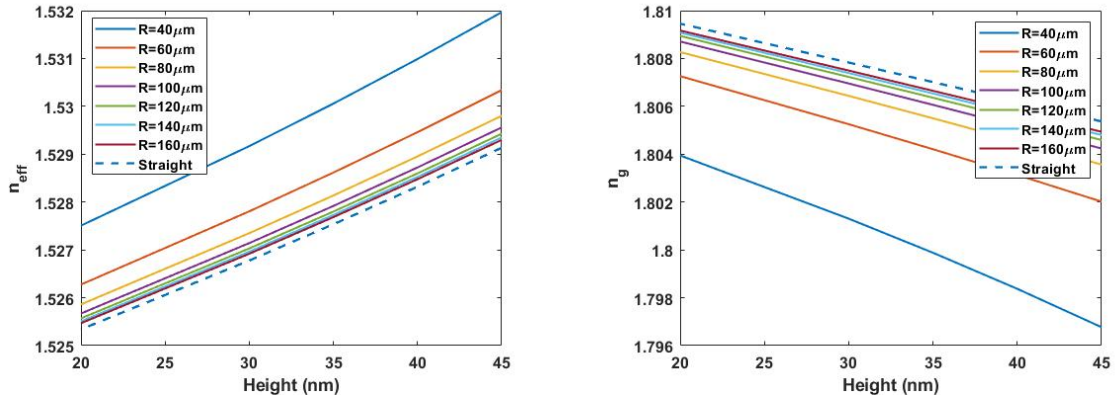
(a) (b)

Figure 4-7: The sensitivity due to width variations at 1550 nm in (a) effective index, and (b) group index.



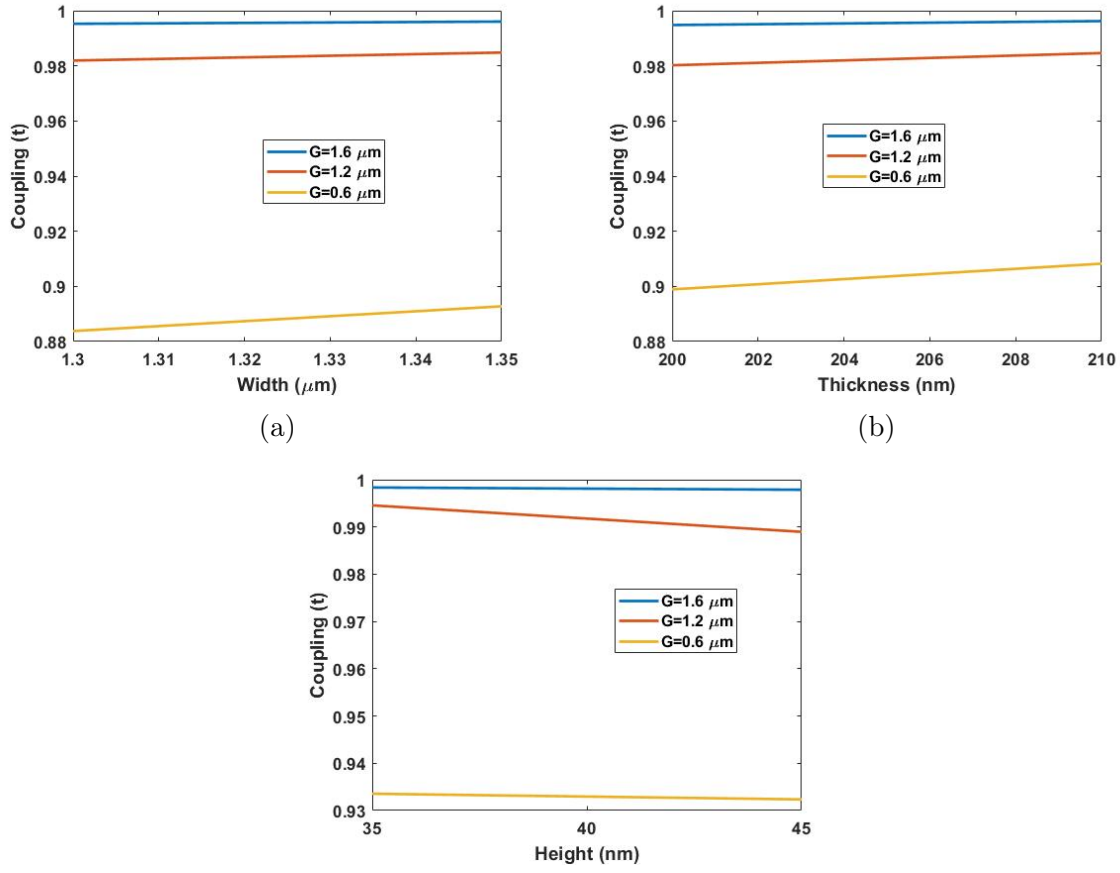
(a) (b)

Figure 4-8: The sensitivity due to thickness variations at 1550 nm in (a) effective index, and (b) group index.



(a) (b)

Figure 4-9: The sensitivity due to height variations at 1550 nm in (a) effective index, and (b) group index.



(a) (b)

(c)

Figure 4-10: The coupling coefficient variation at 1550 nm due to variations in silicon nitride ring (a) width for $R = 60$ μm, (b) thickness for $R = 100$ μm, and (c) height for $R = 120$ μm.

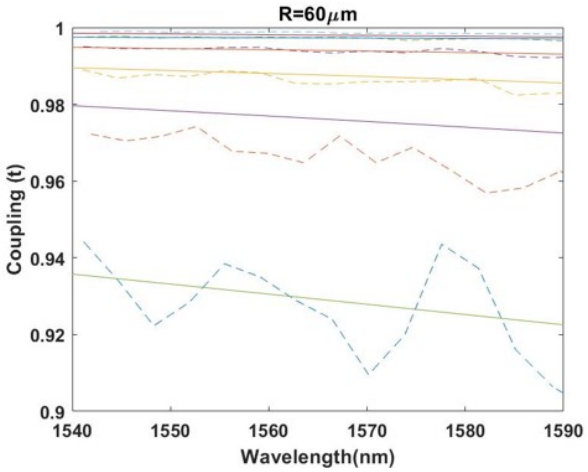
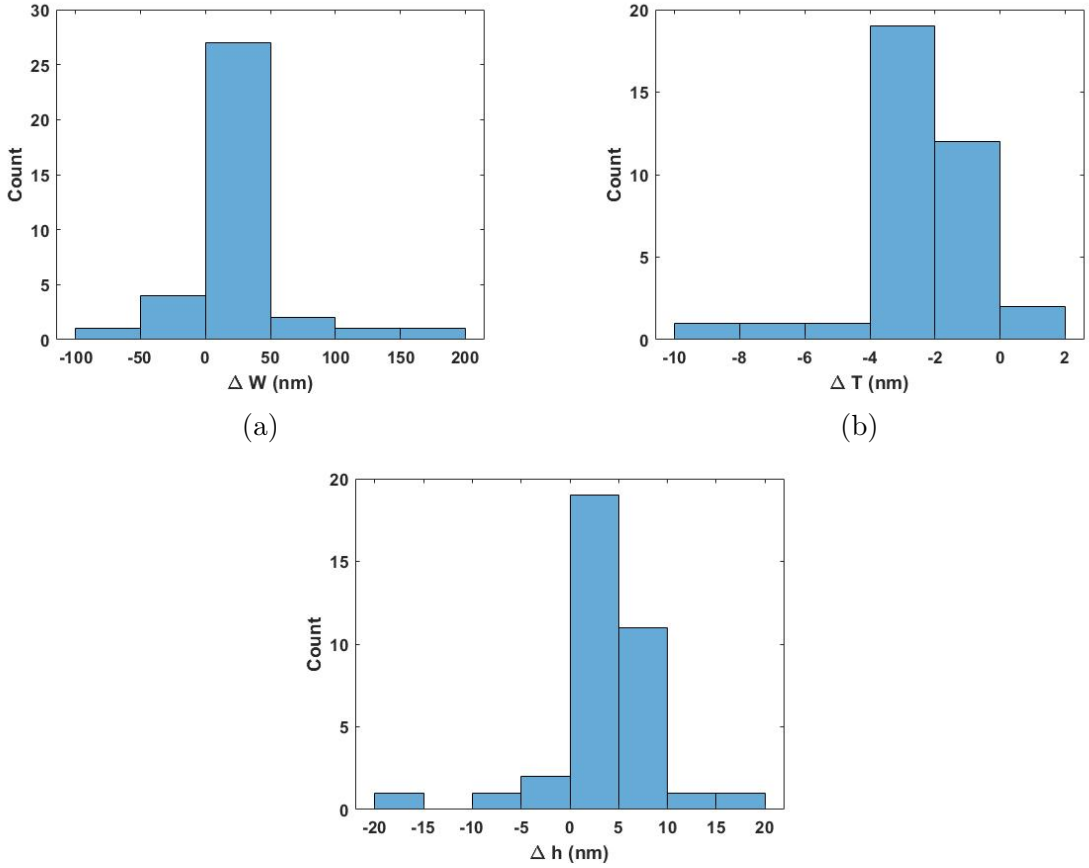


Figure 4-11: The coupling coefficient for silicon nitride rings with $R = 60$ μm and different gaps. The solid lines are the simulation values and the dashed lines are the fabrication-extracted values.

Using the calculated sensitivity matrix and the difference between fabrication and simulations for the features, Eq. 4.6 can be used to find the variations in geometry, i.e., in width, thickness, and height, due to fabrication. Figure 4-12 shows the histogram of the extracted variations in width, thickness, and height. For the width variation (ΔW), it has a standard deviation of 42 nm; the standard deviation for ΔT is 1.76 nm; and finally, for the height variation (Δh), the standard deviation is 5.53 nm. Figure 4-13 shows the spatial variation for the extracted geometries overlaid on the wafer layout for the multiple rings.

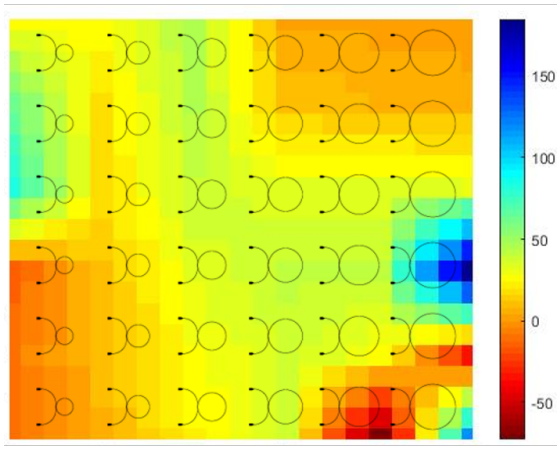
4.2 Models

Since the simulation of a whole ring with large radius (like our silicon nitride rings having $R \geq 20$ μm) using FDTD can be difficult and slow, it is common to divide it into sub-components and simulate each separately. Specifically the half ring is simulated to obtain the coupling coefficient or S-parameters using FDTD [22] as shown surrounded by the dotted box in Fig. 4-1(a), and the rest of the remaining passive bent sections, shown surrounded by the solid box in Fig. 4-1(a), are characterized using MODE [23] simulations. However, in this case of silicon nitride rings that have

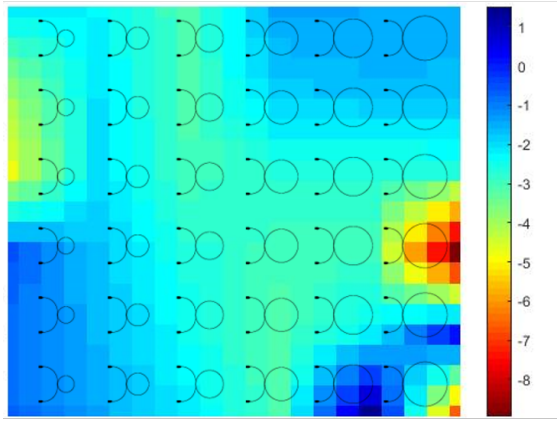


(a) (b) (c)

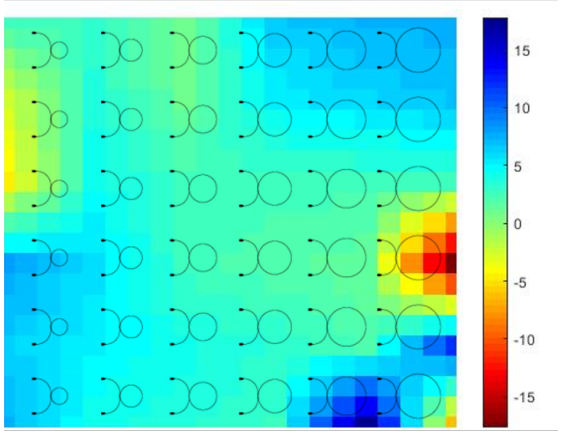
Figure 4-12: The histograms for the extracted silicon nitride geometric variations using Eq. 4.6 for (a) width, (b) thickness, and (c) height.



(a)



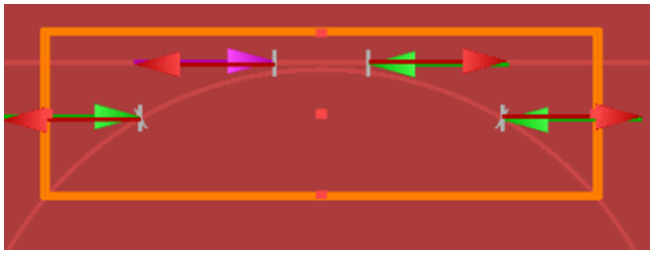
(b)



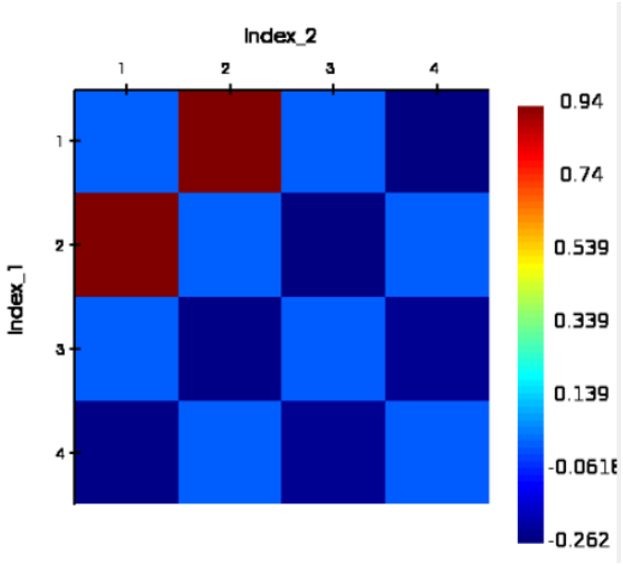
(c)

Figure 4-13: The spatial variation map of the extracted silicon nitride geometric variations in Fig. 4-12, for (a) width, (b) thickness, and (c) height.

radii that can reach up to 160 μm, even simulating the half ring requires a long time in FDTD; this can reach more than 12 hours for large radii. So, to obtain the coupling coefficient and S-parameters for a given ring, we choose to simulate a smaller portion of the ring. In particular, we simulate 1/6 of the ring as shown in Fig. 4-14(a); the resulting S-parameters from the simulation are shown in Fig. 4-14(b).



(a)



(b)

Figure 4-14: (a) Silicon nitride ring simulation setup in FDTD, where the ports are placed at angles 30^o and -30^o with respect to the center of the ring. (b) The S-parameters at 1550 nm for ring with $R = 60$ μm and $G = 0.6$ μm.

As mentioned earlier, the long simulation times for generating S-parameters motivates the development of compact models for these silicon nitride rings, i.e., for parameterized models able to generate $S_{ij}(R, G, \lambda)$ where i and j are the input and output ports. To develop the compact models, we start by choosing a DOE consisting of 13 different radius and gap combinations for simulations. Using these, the corresponding 13 S-parameter spectra are generated using FDTD simulation. We further

divide these 13 S-parameter datasets into training and test sets, with ten used for training and the rest for testing. For the different S-parameters, we make use of symmetry as observed in Fig. 4-14, such that:

$$S_{11} = S_{22},\ S_{12} = S_{21}$$
$$S_{13} = S_{24},\ S_{14} = S_{23}$$
$$S_{31} = S_{42},\ S_{32} = S_{41}$$
$$S_{33} = S_{44},\ S_{34} = S_{43}$$

A third-order polynomial regression (in R, G, and λ) is used for the S-parameter magnitude and phase models. The coefficient of determination (R^2) for all of the models is at least 0.95; i.e., the model is capable of explaining 95% of the data spread in training data. Figure 4-15 shows the S_{12} magnitude and phase response for a training and a test set. For the phase models, the difference between the model and the simulation is very small, $\sim$0.1 rad; and for the magnitude, this difference for the training set is 0.005, and for the test set it is 0.008.

To simulate the whole ring, in addition to the S-parameters corresponding to the coupling region characteristic making up 1/6 of the ring behavior, the rest of the ring consisting of the bent waveguide section in Fig. 4-14(a) can be characterized in MODE. Using MODE simulation results for the bent section along with the S-parameters, either generated from direct simulation or generated by model, we simulate the whole ring in the circuit simulator INTERCONNECT [39] using the circuit shown in Fig. 4-16(a). We obtain the responses in Fig. 4-16(b). It is seen that the behavior of the silicon nitride ring using the model-generated and the simulation-generated S-parameters are in good agreement.

We further compare the behavior of the rings as fabricated and measured with circuit simulations using the S-parameters generated from compact model; the results are displayed in Fig. 4-17. We see that there is a slight difference in the FSR between the fabricated (3.62) and the simulated (3.56) ring. This could be due to the geometric variations that the fabricated rings experience, as shown previously in Section 4.1.2.

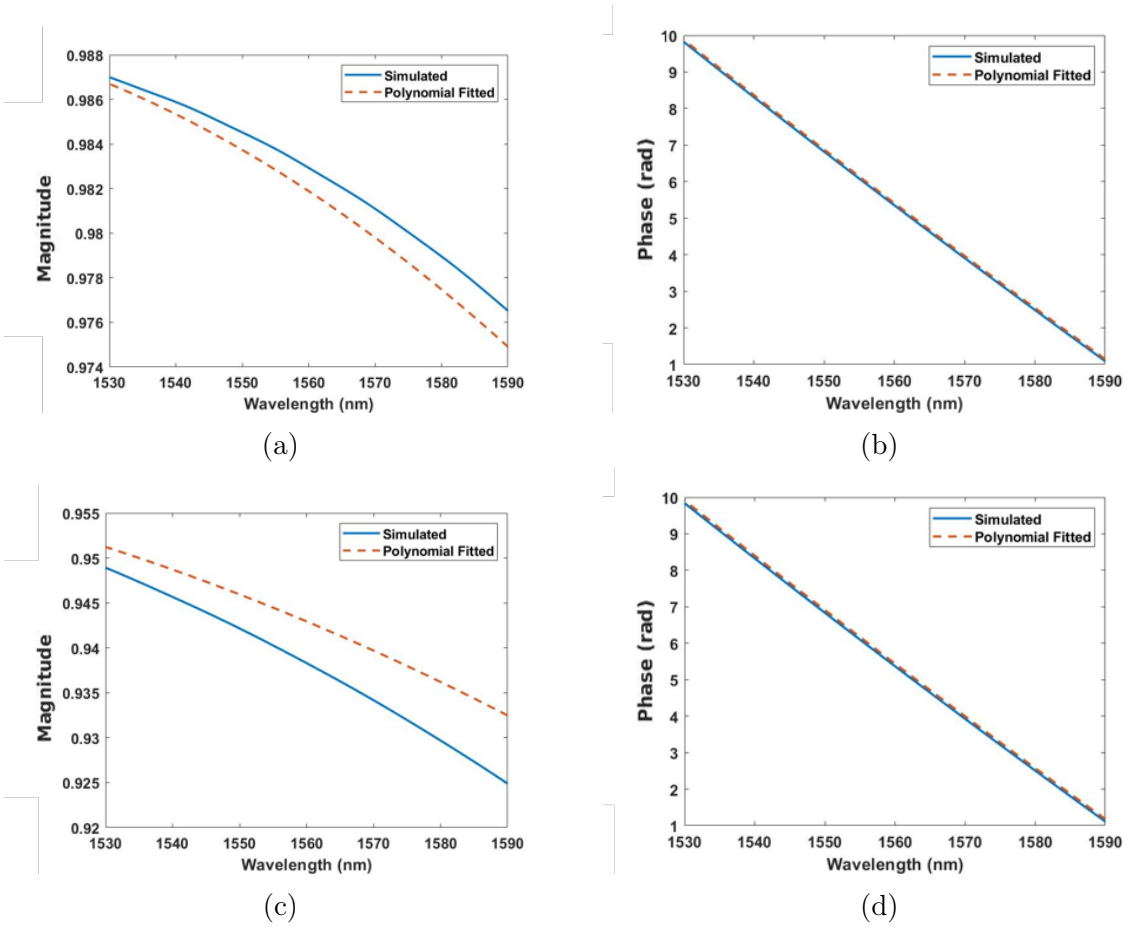
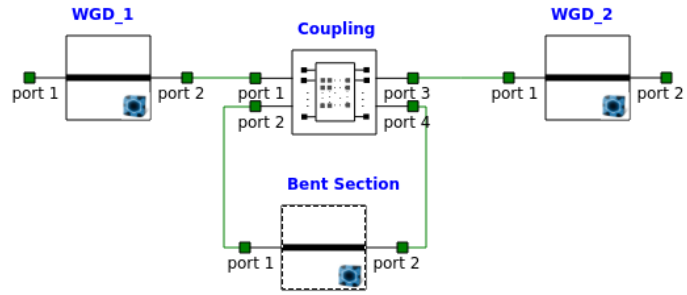
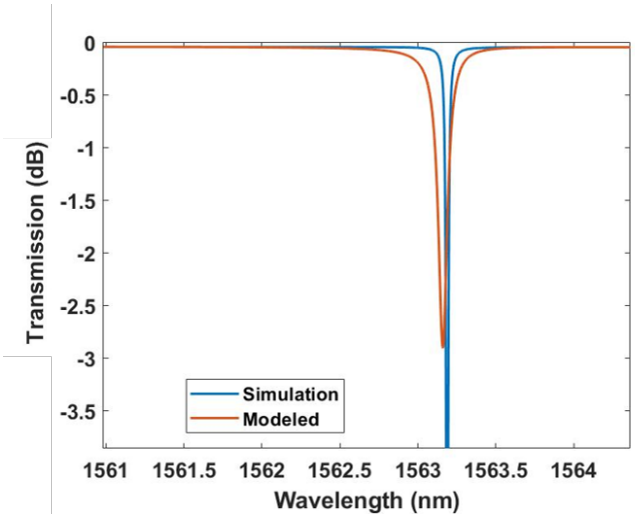


Figure 4-15: Comparison of S_{12} polynomial regression modeling and direct simulation behavior for silicon nitride rings. (a) Magnitude for training set with $R = 80$ μm and $G = 0.8$ μm, and (b) phase for training set with $R = 80$ μm and $G = 0.8$ μm. (c) Magnitude for test set with $R = 60$ μm and $G = 0.6$ μm, and (d) phase for test set with $R = 60$ μm and $G = 0.6$ μm.

Also, the extinction ratio of the simulated ring is less than that of the fabricated ring, since the fabricated ring has more losses.



(a)



(b)

Figure 4-16: (a) Circuit used in INTERCONNECT to simulate the silicon nitride rings. (b) Using simulated and model-generated S-parameters to simulate a ring with $R = 60$ μm and $G = 0.6$ μm.

4.3 Summary

We demonstrate a methodology to provide S-parameter based predictive models as a function of the silicon nitride ring coupling gap (G), radius (R), as well as the operating wavelength (λ). These models can be used to generate the S-parameters for a given ring design that can be used in circuit simulation to analyze or optimize ring performance, for such parameters as free spectral range (FSR), resonance

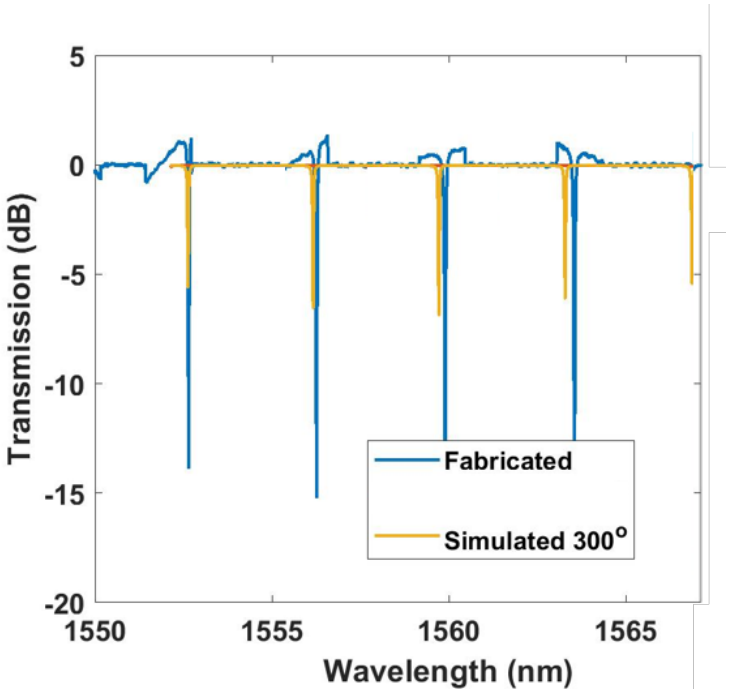


Figure 4-17: Behavior for a ring with $R = 60$ μm and $G = 0.6$ μm, from fabrication and using the model-generated S-parameters in circuit simulation.

locations (λ_{res}), and the group index (n_g). We validate and test the model using fabricated rings, where we compare the performance extracted from fabrication with the model's predicted performance, for various design parameters. This compact modeling approach can help designers narrow, if not pre-define exactly, the parameters that can be used to achieve the desired performance specification, and thus speed up design simulation and optimization. Moreover, as we have fabricated devices, we use measurement data to predict the within-wafer silicon nitride width, height and thickness variations for the foundry in which the devices are fabricated.

Chapter 5

Variation Extraction and Decomposition from Measurements

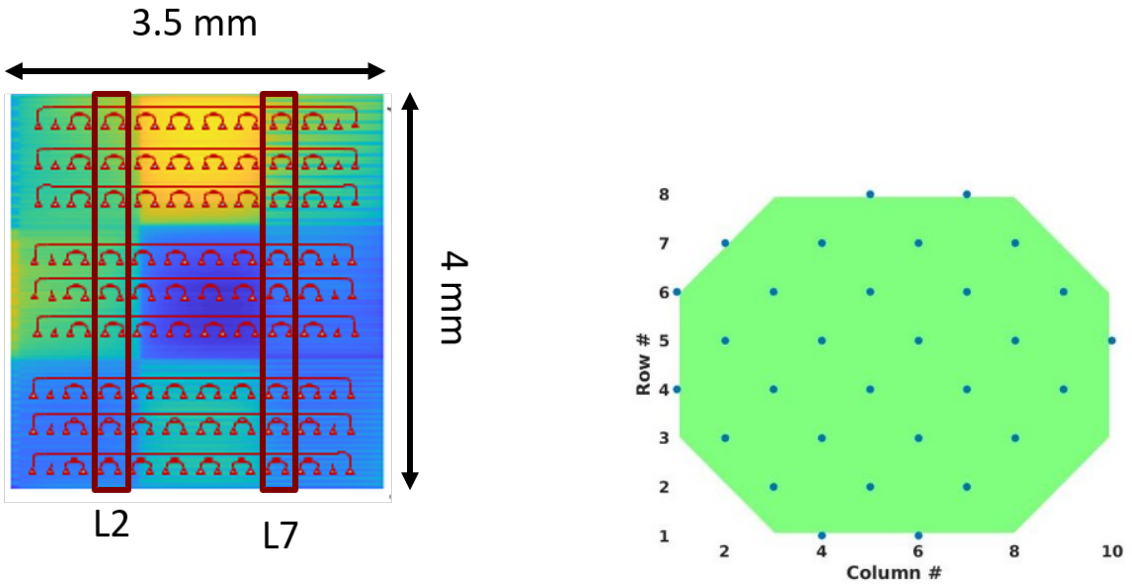
The DFM methods presented in previous chapters enable prediction of photonic behavior in the presence of process variations. However, process variations can depend on the foundry, process steps, location within the wafer, or surrounding layout context within a die (e.g., layout pattern density) [50]. Accordingly, a complete DFM flow (as in Fig. 1-4) requires methods to extract and decompose [51] variation dependencies and models from measurements.

In this chapter, we present a case study demonstrating methods to extract variation statistics, including the use of variation test chips. Section 5.1 presents the designed test structure and test chip, and summarizes measurements from a fabricated test wafer. The extraction of geometric variations from these measurements is described in Section 5.2. This includes decomposition into die-to-die variation, silicon layout pattern density dependent variation, and residual variation.

5.1 Test Structure and Measurements

We use a slab ring resonator with a radius of 5 μm, width of 500 nm, and thickness of 220 nm, as the test structure in our test chip. The overall chip is 3.5 $mm \times 4$ mm in size, and is contained within a larger multi-project wafer (MWP) reticle. The slab ring

resonator is replicated within the test chip and placed in different pattern densities, Fig. 5-1(a). We introduce nine different regions with specified layout pattern density of etched silicon, ranging from about 5% to 55%, as created by different designed dummy fill pattern densities. The test chip is then replicated across the wafer, with every other chip measured, Fig. 5-1(b). With these replicas, we obtain both the statistics of that ring resonator behavior, and are able to extract the underlying spatial variations.



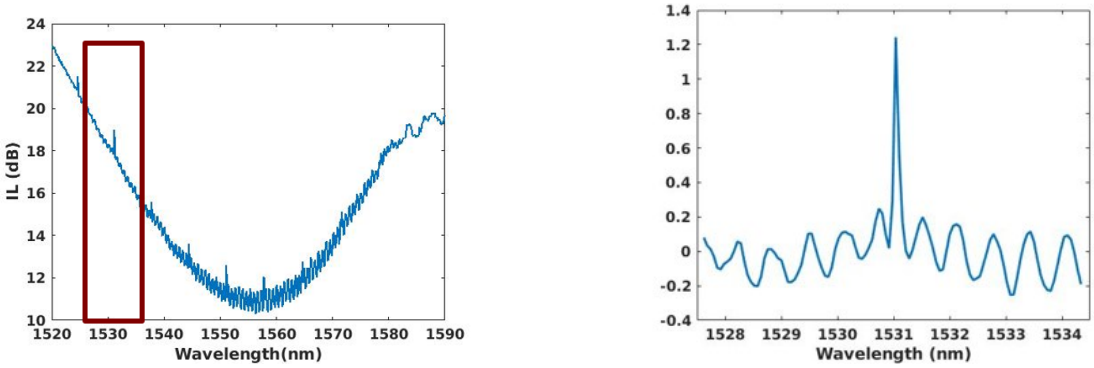
(a) (b)

Figure 5-1: (a) Slab ring resonators, in red box, replicated within the test chip with different pattern densities across the test chip. (b) Every other chip on the 300 nm wafer is measured, with locations as shown.

The insertion loss for each of these ring replicas is measured using grating couplers, as shown in Fig. 5-2(a). After detrending, an example (normalized) transmission is shown in Fig. 5-2(b). From these, we extract the group index as well as the resonance wavelength (λ_{res}). This is done by extracting the locations of resonance and the free spectral range (FSR). Then the group index is found such that:

$$n_g = \frac{\lambda_{res}^2}{FSR \cdot L} \tag{5.1}$$

where L is the circumference of the ring, $L = 2\pi R$ with designed ring radius R.



(a) (b)

Figure 5-2: (a) Insertion loss measured for one of the fabricated rings including grating coupler effects. (b) The insertion loss after de-trending and removing the effect of the measurement setup.

Both of the extracted quantities (n_g and λ_{res}) vary with the width and thickness [41], and are related by the first order sensitivity (sensitivity matrix) with respect to width and thickness as in Eq. 5.2. The sensitivity matrix can be calculated using FDTD simulation for our designed structure. The width results are shown in Fig. 5-3, giving $\frac{\delta\lambda_{res}}{\delta W} = 0.5859\ (nm/nm)$ and $\frac{\delta n_g}{\delta W} = -0.00165\ (/nm)$. For thickness, the results are shown in Fig. 5-4, and give $\frac{\Delta\lambda_{res}}{\delta T} = 1.3633\ (nm/nm)$ and $\frac{\delta n_g}{\Delta T} = 0.00109\ (/nm)$.

$$\begin{bmatrix} \Delta W \\ \Delta T \end{bmatrix} = \begin{bmatrix} \frac{\delta n_g}{\delta W} & \frac{\delta n_g}{\delta T} \\ \frac{\delta \lambda_{res}}{\delta W} & \frac{\delta \lambda_{res}}{\delta T} \end{bmatrix}^{-1} \begin{bmatrix} \Delta n_g \\ \Delta \lambda_{res} \end{bmatrix} \tag{5.2}$$

5.2 Geometry Extraction

We can use Eq. 5.2 to infer width and thickness variations for each test chip fabricated in a specific foundry. In our experiment, our extractions are done for the resonance mode near wavelength of 1531 nm. One thing to note is that the silicon layer thickness is assumed to be constant within each individual test chip, due correlation lengths of 4 nm to 5 nm within wafer [41, 52, 53]; accordingly, thickness variations are considered to be a die-to-die variation (with 30 values only). Using this information with Eq. 5.2, we extract the values for ΔW and ΔT geometric variations.

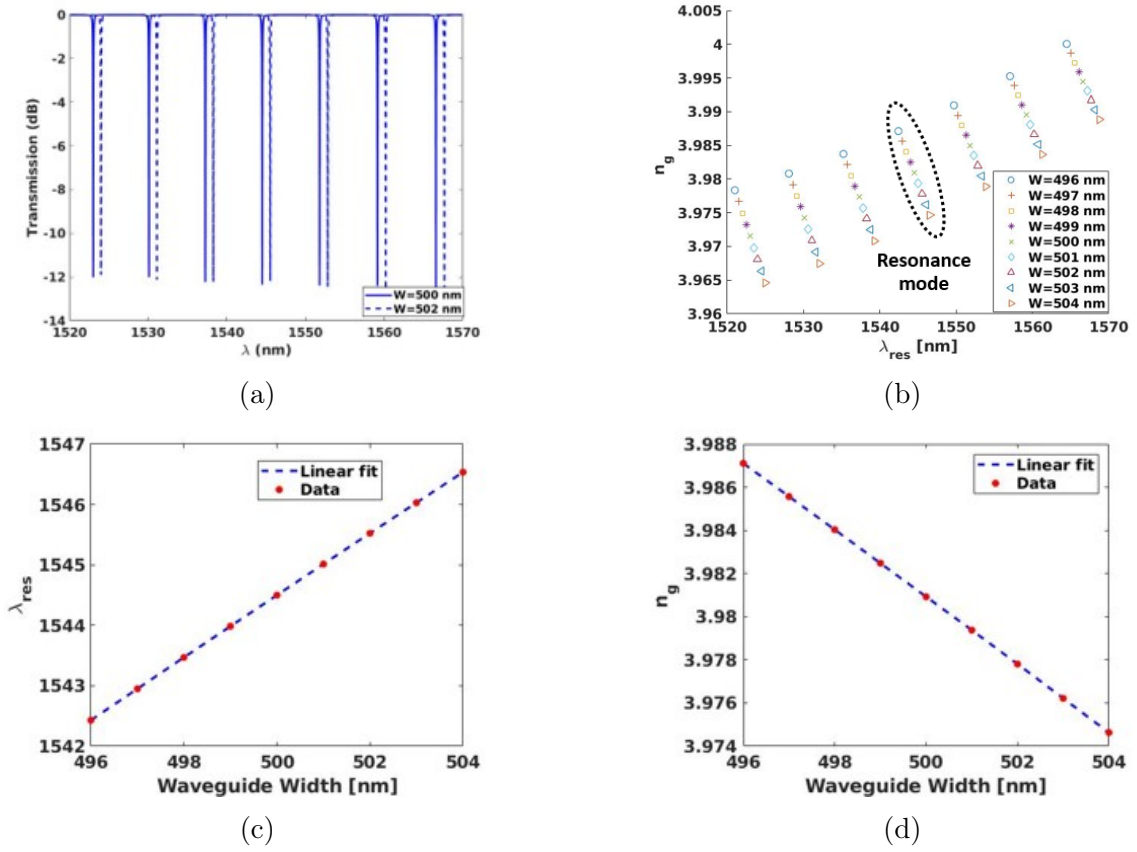


Figure 5-3: The sensitivity due to width variations. (a) Change in λ_{res}, (b) change in group index (n_g), (c) the first order sensitivity in λ_{res} due to width ($\frac{\delta\lambda_{res}}{\delta W}$) , (d) the first order sensitivity in n_g due to width ($\frac{\delta n_g}{\delta W}$).

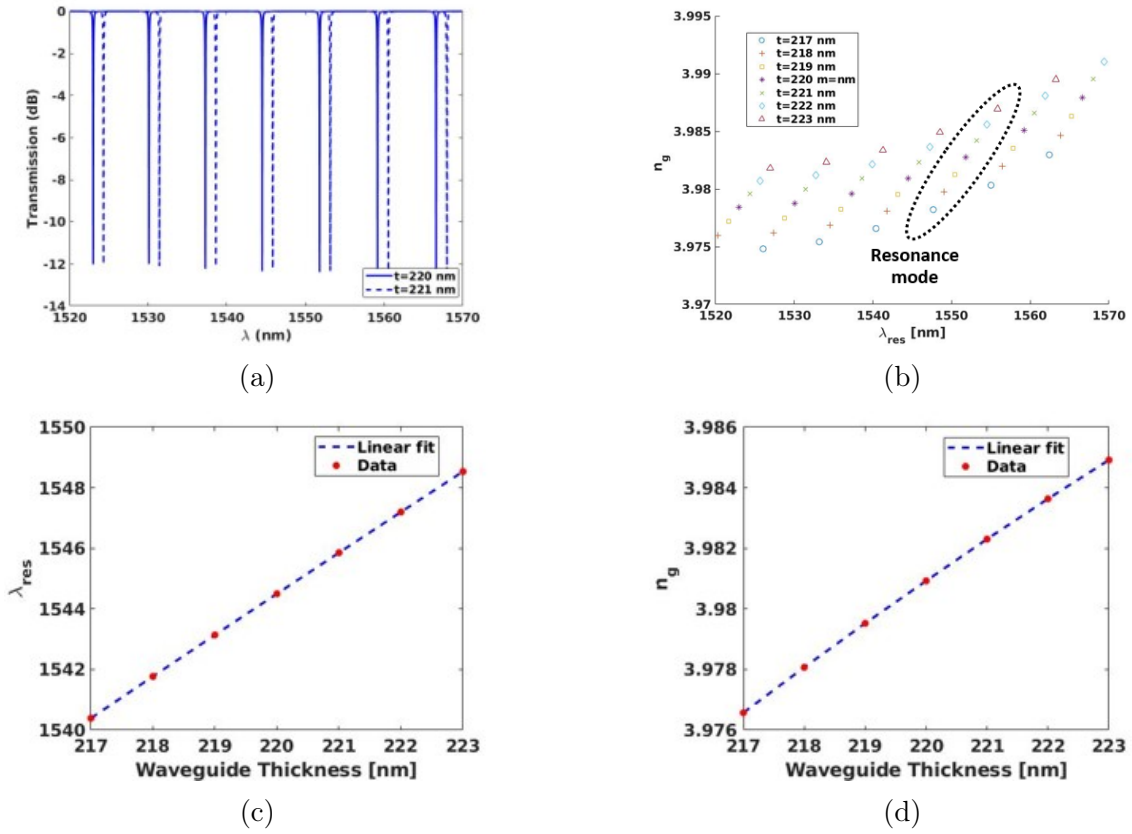
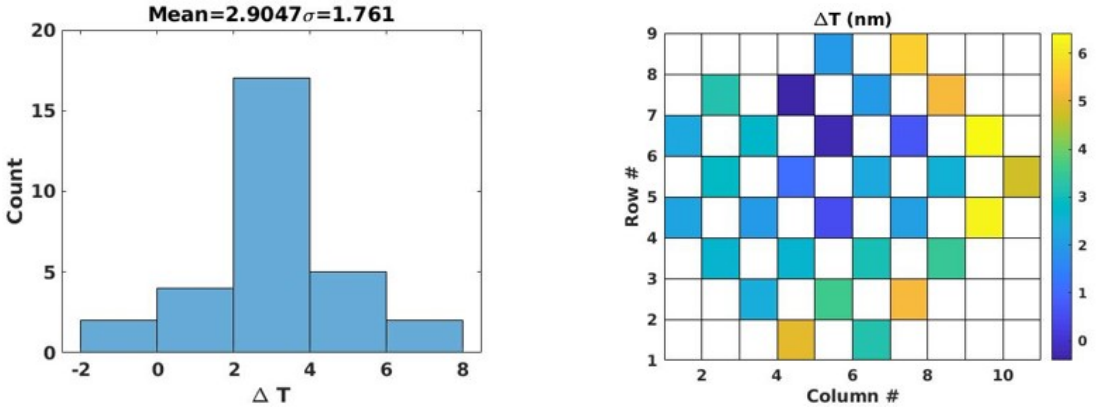


Figure 5-4: The sensitivity due to thickness variations. (a) Change in λ_{res}, (b) change in group index (n_g), (c) the first order sensitivity in λ_{res} due to thickness ($\frac{\delta\lambda_{res}}{\delta T}$) , (d) the first order sensitivity in n_g due to thickness ($\frac{\delta n_g}{\delta T}$).



(a) (b)

Figure 5-5: (a) Die-to-die variation statistics for the thickness variation ΔT across the wafer. (b) Spatial variation map for the thickness variation across the wafer.

For thickness, the histogram of the extracted values is shown in Fig. 5-5(a), where ΔT is found to have a mean of 2.9047 nm and variance of 1.76 nm. The spatial map of the ΔT variation across the wafer is shown in Fig. 5-5(b), which indicates a center-low thickness variation pattern at the wafer scale.

For the width variations ΔW, the variations are assumed to depend or vary based on location within the test chip as well as across the wafer. Thus, we get an overall larger distribution in device-to-device variation statistics. Figure 5-6 shows the histogram for the extracted width variations, where ΔW has a mean of -6.72 nm and a standard deviation of 4.4789 nm.

However, with width we are also interested in a systematic width variation dependency due to the plasma etch process. Etch variations are known to depend on the local or effective pattern density of structures nearby to or around any given structure, resulting in within die variation [54, 55].

Accordingly, it is important to quantify and extract the effect of the effective layout pattern density on the etch width. To calculate the effective pattern density, we need to find the length scale or correlation length (σ) over which we average layout pattern density to get the effective pattern density. For our data, a correlation length of 500 μm is found to give the least mean square error for a second order fit of ΔW and ρ_{eff}, Fig. 5-7. These results are similar to those reported elsewhere for waveguide

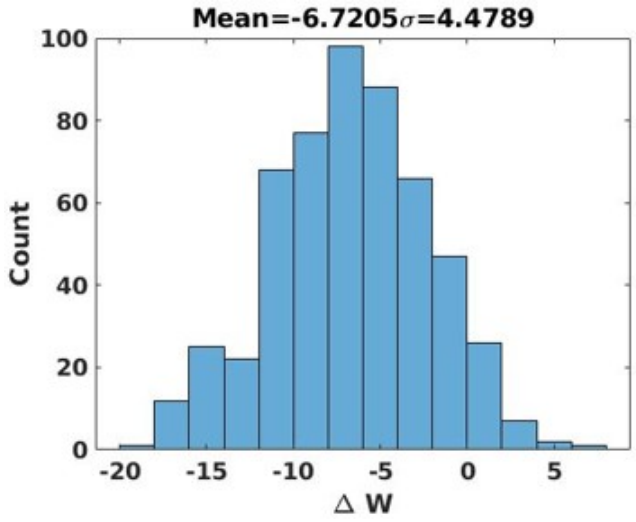


Figure 5-6: Histogram for the extracted width variation, ΔW, including both within-die and die-to-die variations.

width variations in a 193 nm lithography process [55].

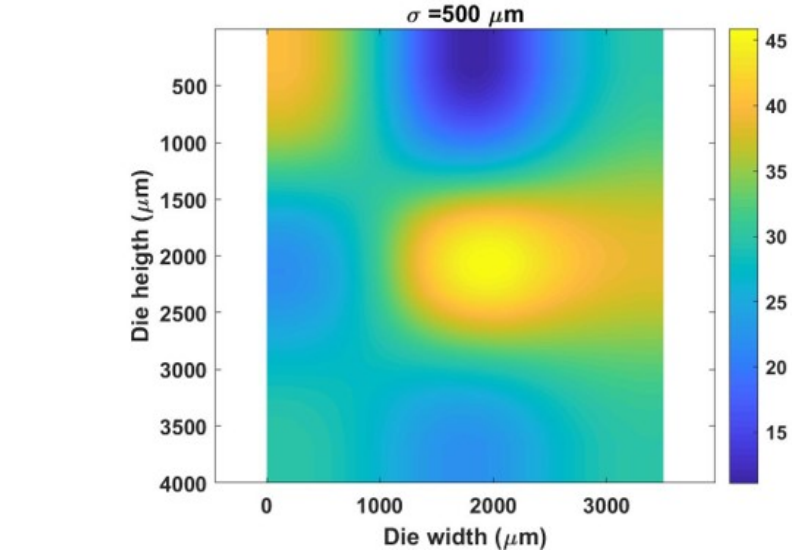
The width variations can be further decomposed into multiple components. First, die-to-die is the variation in mean width within a test chip, as shown in Fig. 5-8(a). Second is width variation due to effective layout pattern density, as in Fig. 5-8(b). This can be very useful in evaluating dummy fill strategies or to make layout modifications to ensure that width variations, are kept within accepted range. Finally, there is some residual or random width variations as shown in Fig. 5-8(c).

Accordingly, we can describe width variation as:

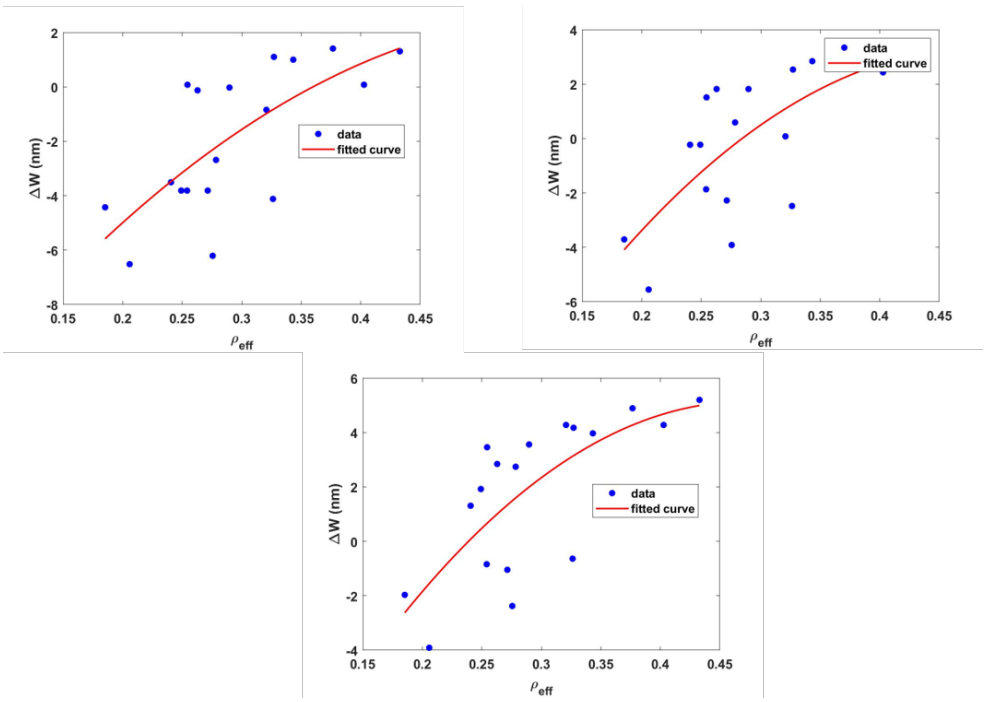
$$W_{i,x,y} = W_{nom} + \Delta W_{d_i} + f(\rho_{eff}, x, y) + \Delta W_{res} \tag{5.3}$$

where $W_{i,x,y}$ is the width of device in die i at position x and y within the die. Here, W_{nom} is the nominal width (it is 500 nm in our test structure). Variation ΔW_{d_i} is the die offset, or in other word, the variation in the device width due to its presence in die i. This is the mean width variation within the die from extraction. The function $f(\rho_{eff}, x, y)$ is the shift in the width of the device at position x and y due to the effective pattern density. Finally, ΔW_{res} is random/residual variation not included or explained by the die mean offset or the effective pattern density.

Such methods that include design of test structures (e.g., ring resonators) and



(a)



(b)

Figure 5-7: (a) Effective layout pattern density for the test chip shown in Fig. 5-1, based on extracted spatial averaging length of 500 μm. (b) The fitting of ρ_{eff} and ΔW for different dies.

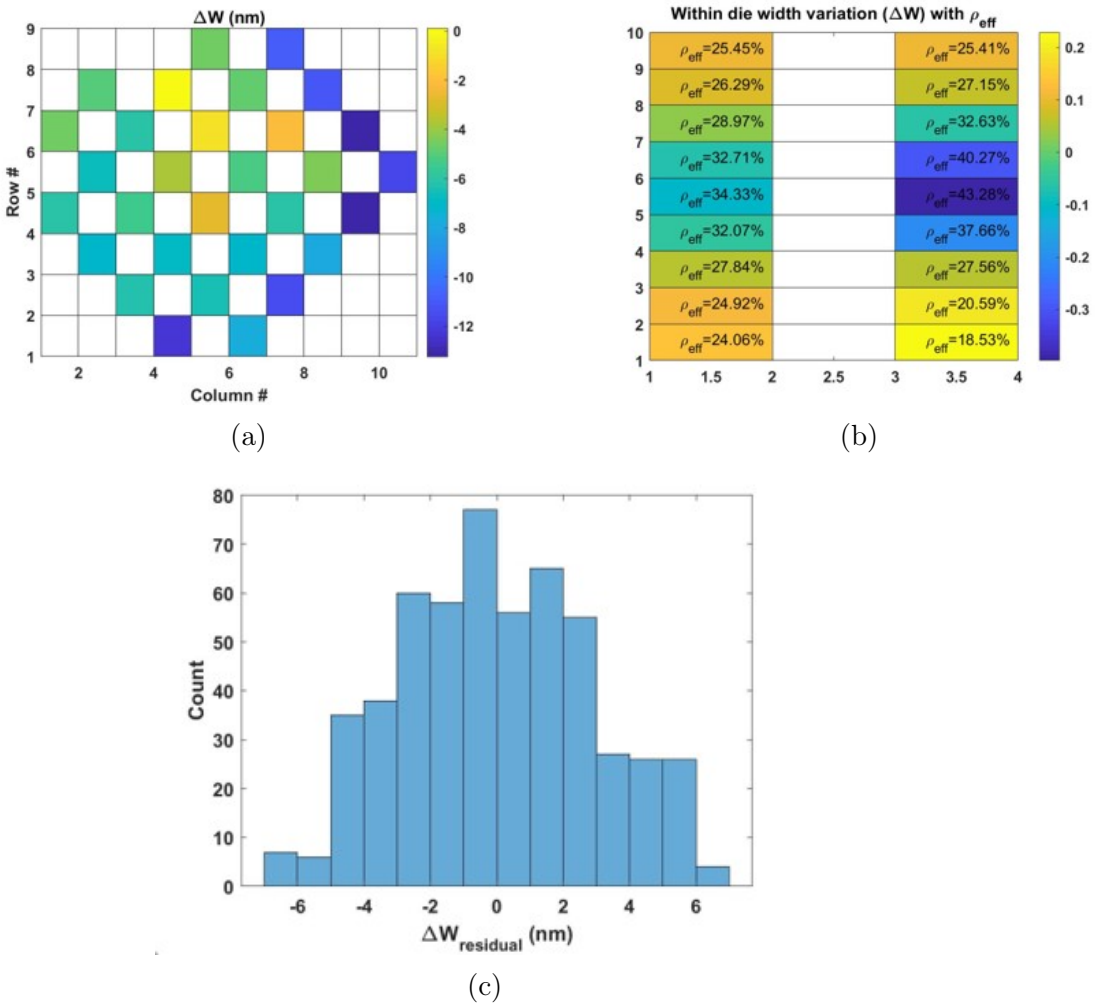


Figure 5-8: (a) Spatial variation map for the mean width within each die across the wafer. (b) Width variation ΔW as a function of effective layout pattern density. (c) Residual width variation.

variation decomposition using test structure measurements provide statistics and dependencies of the underlying process variations. The variation statistics and models can then be used in conjunction with variation-aware compact models to analyze variation yield impact and to provide photonic design rule guidelines for acceptable yield. This information will help designers when placing their components in a photonic integrated circuit in order to maximize the yield and minimize the effect of process variations.

5.3 Summary

We design and measure test chips based on a slab ring resonator test structure to extract key process variations. We explore layout dependencies by applying variable pattern densities within the die. The die is repeated across the wafer, enabling die-to-die variation extraction. The aim of this method is to extract both the effect of pattern density on the behavior of the test structure, as well as extract the spatial correlation relation within die and across the wafer. Based on test structure measurement and geometric variation extraction, we then separate variations due to pattern density effect, location within wafer, and random residuals not captured in the pattern density effect or wafer location effect. This information will help designers to analyze and maximize yield, in conjunction with variation-aware compact models.

Chapter 6

Active Devices: Modulators

In the previous chapters, we have seen that process variations, both systematic and random, can cause substantial shift and degradation in photonic performance and consequently negatively affect yield. Accordingly, it becomes clear that for a photonic integrated circuit (PIC), the availability of means to correct for these process variations becomes crucial. For this purpose, modulation rises as a well-known approach to correct for both operating condition and fabrication variations [56, 57].

More generally, optical modulation is an essential functionality in optical interconnect solutions. Optical modulators are key building-blocks in silicon photonics, as they are used to modulate or vary the propagating light beam properties. Being an important photonic component, we are also interested in understanding and modeling how modulators are impacted by process variation.

In this chapter, we study active photonic components, specifically elements used to implement modulators. This enables us both to understand variation impact in these devices, and to examine their use to correct or compensate for inherent process variations in PICs. In Section 6.1, we review dependencies of the index of refraction on carrier concentration, particularly PN junctions. We describe our simulation setup in Section 6.2, and model sensitivities to different process variations. In Chapter 7, we use these modules for larger PIC variation analysis and variation compensation.

6.1 Plasma Dispersion Effect

Optical modulation in photonic circuits is implemented by devices that cause direct changes in optical intensity via absorption, or cause changes in the refractive index of the material (and hence the phase of a propagating wave) that can be converted to an intensity change via an interferometer or a resonator (e.g., a ring resonator and Mach-Zehnder interferometer) [58]. Applying an electric field is an attractive way to achieve modulation, since modulation can be achieved with small current, with low power as well as fast response.

The plasma dispersion effect [59] defines the relation between the refractive index, both real and imaginary parts, and the carrier concentrations modified either by removing or injecting carriers. The change in the refractive index real part (Δn) and imaginary part, absorption ($\Delta\alpha$), according to the plasma dispersion effect is updated in [60] and defined as:

$$\Delta n\ (1550\ nm) = -5.4 \times 10^{-22}\Delta N^{1.011} - 1.53 \times 10^{-18}\Delta P^{0.838} \tag{6.1}$$

$$\Delta\alpha\ (1550\ nm) = 8.88 \times 10^{-22}\Delta N^{1.167} + 5.84 \times 10^{-20}\Delta P^{1.109}\ [cm^{-1}] \tag{6.2}$$

where ΔN and ΔP are the changes in carrier densities for electrons and holes (cm^{-3}), respectively. The above relations for Δn and $\Delta\alpha$ variations can be extended to include wavelength dependence as well [5], as follows:

$$\Delta n(\lambda) = -3.6 \times 10^{-10}\lambda^2\Delta N - 3.51 \times 10^{-6}\lambda^2\Delta P^{0.8} \tag{6.3}$$

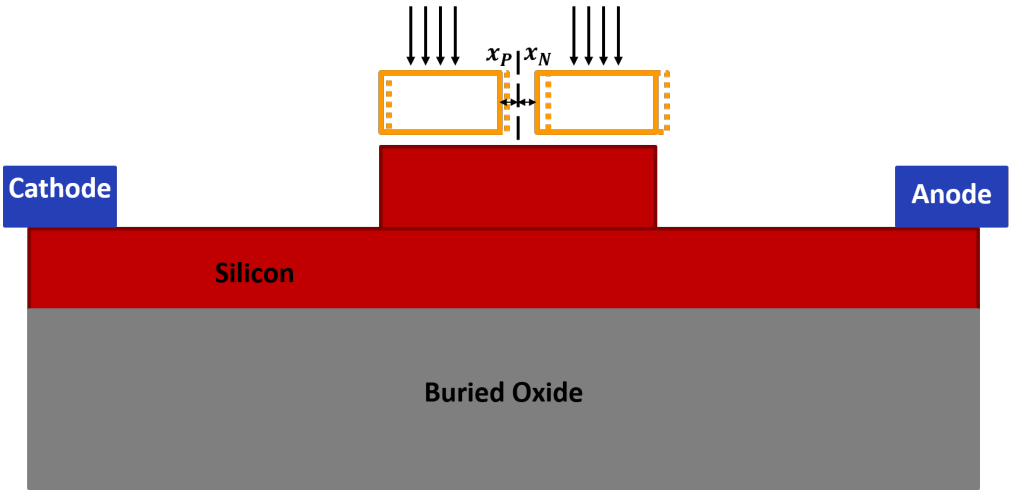
$$\Delta\alpha(\lambda) = 3.52 \times 10^{-6}\lambda^2\Delta N + 2.4 \times 10^{-6}\lambda^2\Delta P\ [cm^{-1}] \tag{6.4}$$

where λ is the wavelength in meters.

6.1.1 PN Junction

Accordingly, to change the refractive index, we need to modify the charge density. Manipulation of the charge density can be achieved using the depletion mechanism

as in a PN junction, as shown in Fig. 6-1. In our application, the PN junction is a rib waveguide that is doped in such a way that enables modulation in silicon photonics [5] using the plasma dispersion effect for controllable electro-optical phase shift. The doping is done with a mask-defined doping boundary. When the PN junction is reverse biased, a depletion region with no free carriers is formed; consequently, the effective index increases and the absorption decreases based on Eqs. 6.3 and 6.4. This depletion region formed between the N and P doping regions increases as the applied reverse voltage increases as shown in Fig. 6-5. We examine integrating this PN semiconductor junction into devices (e.g., ring resonator) to achieve efficient modulation, relying on the plasma-dispersion effect.



(a)

Figure 6-1: A cross section of the PN junction schematic, where x_n and x_p are the N and P doping mask offsets.

6.2 Structure Modulated

Ring resonators are often modulated and used for high speed data transfer [61, 62]. In our study, we examine a ring resonator modulated with a PN junction. The geometry of the silicon ring modulator used is adopted from [63], having a radius of 10 μm, width of 450 nm, thickness of 220 nm, and gap of 400 nm.

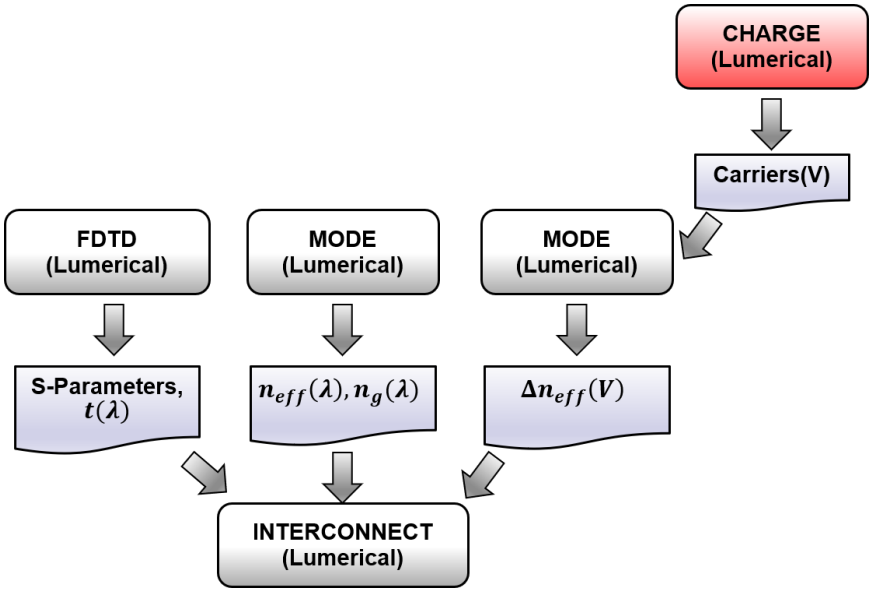


Figure 6-2: Simulation flow for a ring modulator, including both active and passive photonic device components.

6.2.1 Simulation setup

Usually for simulating an optical ring modulator, the device is divided into some number of sub-components, where the designer uses a simulation flow, as shown in Fig. 6-2, to characterize each of these sub-components. We start with an FDTD [22] simulation to obtain the coupling coefficient (t) and S-parameters as a function of frequency for the waveguide coupler section (surrounded by a solid box in Fig. 6-3). We use MODE [23] simulations to characterize the passive (un-modulated) straight and bent waveguides (surrounded by dashed box in Fig. 6-3), where effective index, group index, and dispersion as a function of frequency of interest are calculated. For the active, modulated, waveguide section (surrounded by dotted box in Fig. 6-3), CHARGE [24] simulator is used to obtain the spatial distribution of charge carriers as a function of bias voltage applied. These results are then used in MODE to calculate the perturbation in effective index as a function of bias voltage. Finally, these simulation results for the different sub-components are combined in the photonic circuit simulator INTERCONNECT [39] to simulate the whole structure of the ring modulator.

Applying different reverse bias voltages to the PN region changes the depletion

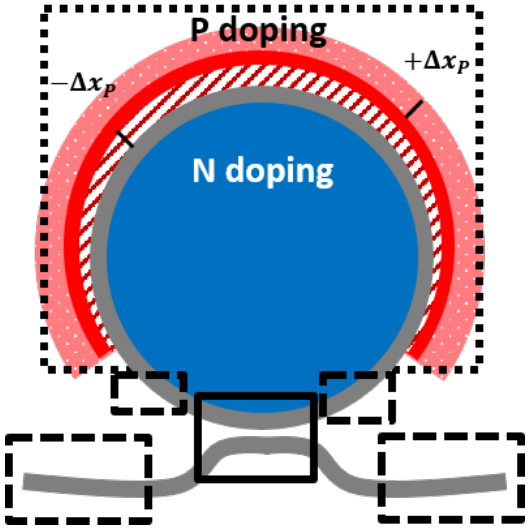


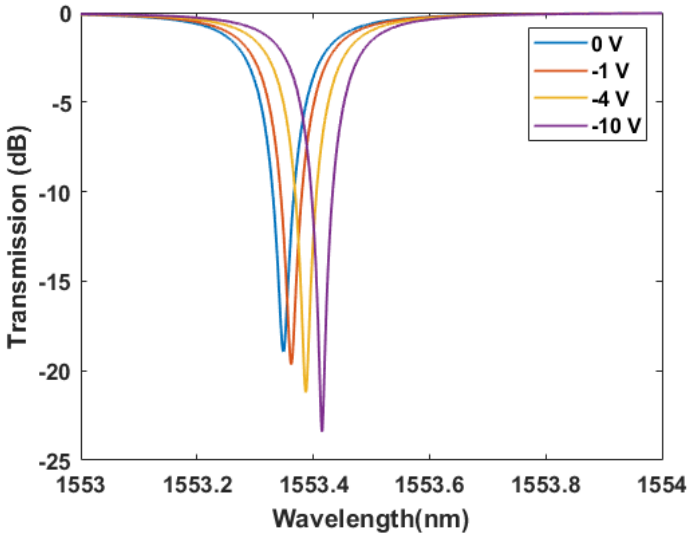
Figure 6-3: Ring modulator used in our analysis. Solid box surrounds the coupler waveguide; dashed boxes surround the passive straight and bent waveguides, and finally the dotted box surrounds the active waveguide.The dotted pattern in the P doping shows the over exposure case while the dashed pattern represents the under exposure case.

region width, as seen in Fig. 6-5 as more free carriers move towards the applied voltage leaving behind charged ions and thus the carrier concentration changes, that changes the effective refractive index, which in turn causes a shift in the resonance locations as seen in Fig. 6-4.

We examine the impact of variations in the active modulation section (CHARGE simulation). Specifically, we are interested in the effect of the doping concentration variation and mask exposure variation on the response of the modulated ring. To analyze this, we use CHARGE simulation to understand variation impact on the active PN region.

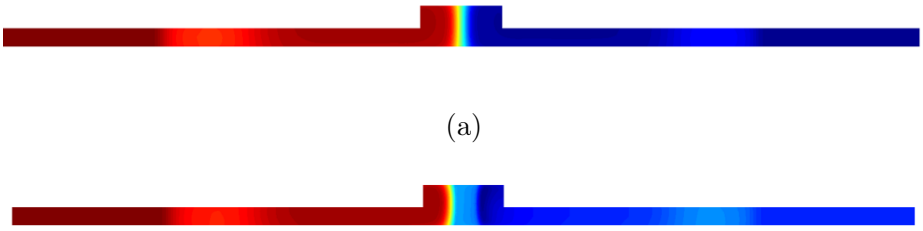
6.2.2 Mask Exposure Variation

For the PN junction, the doping area is defined by a mask that specifies the area within which dopants are allowed to be incorporated into the silicon. When the mask exposure step experiences variation, this causes a variation in the PN junction geometry; i.e., doping locations are shrunk or expanded around the ring from where the intended locations are, as shown in Fig. 6-3. This mask exposure variation can



(a)

Figure 6-4: The effect of varying the reverse voltage applied to the PN junction on resonance wavelength of the ring resonator.



(a)

(b)

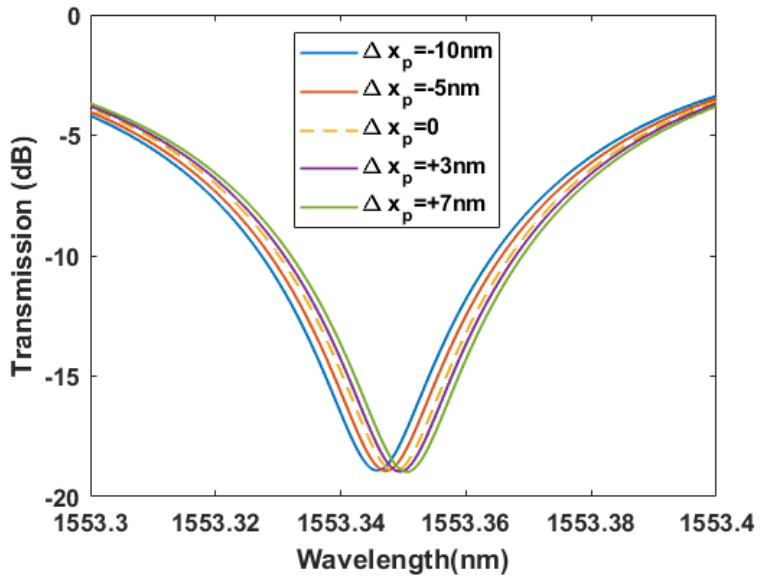
Figure 6-5: Charge distribution in the PN junction with (a) 0 V applied, and (b) -4 V reverse bias applied.

affect device performance. To model the effect of this exposure variation on the modulator directly would require many simulations, particularly if we are to consider many different mask exposure variations.

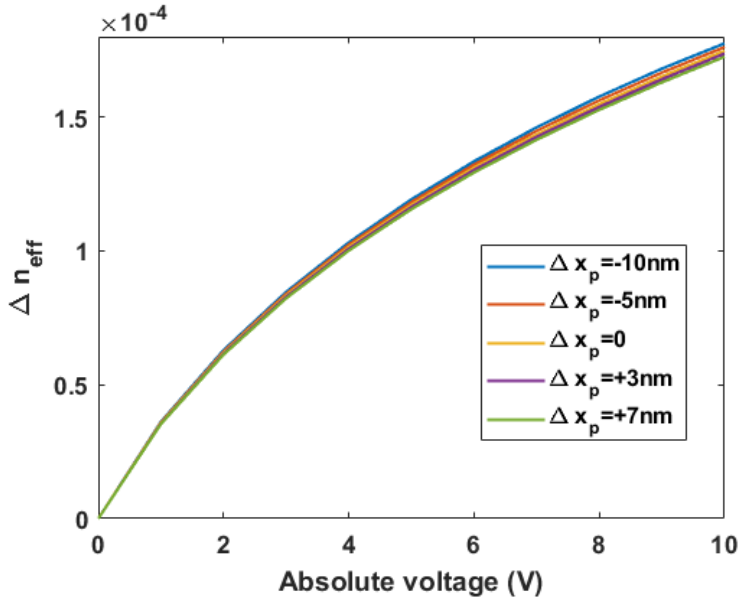
As with the passive components in earlier chapters, here we seek to develop simplified compact models that can predict the effect of mask exposure variation on the different performance parameters. We start by specifying a DOE for the mask position relative to the nominal case. The nominal N-doping mask edge is located 50 nm in the positive direction away from the waveguide center ($x_n = 50\ nm$), and the P-doping mask edge is located 50 nm in the negative direction ($x_p = 50\ nm$), as shown in Fig. 6-1.

For our DOE, we consider maximum variations up to 20% ($\pm 10\ nm$), such that either masks can be expanded or shrunk 10 nm in either direction, i.e., towards or away from the waveguide center. Using the combined simulation setup described in Section 6.2.1, we perform simulations at nominal, $\pm 1\ nm$ to $\pm 10\ nm$ with a step of 1 nm for mask edge variations. The change in the ring resonance frequency due to mask exposure variation is shown Fig. 6-6(a). These can be attributed to the change in the effective index as a function of voltage, shown in Fig. 6-6(b).

We now build models that capture the effect of the mask exposure variation on the various performance parameters of the modulator. We fit regression models for each of our key performance parameters f_i, including extinction ratio, FSR, capacitance, and resonance wavelength as a function of the N and P mask exposure variation and the bias voltage, $f_i(\Delta x_N, \Delta x_P, V)$. Our DOE simulation points are divided into training and test sets (70/30 split ratio). A third-order polynomial regression is used to develop the response surface models. The R^2 values for all of these responses is at least $\sim$0.97. Figure 6-7 shows selected response surface models, illustrating fits for extinction ratio and quality factor as a function of mask exposure variation on the performance.

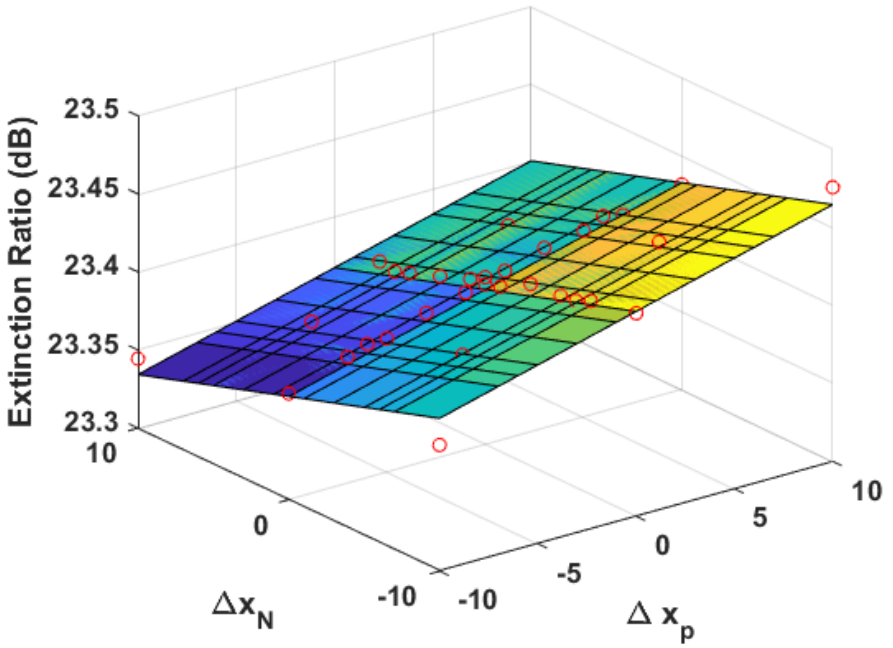


(a)

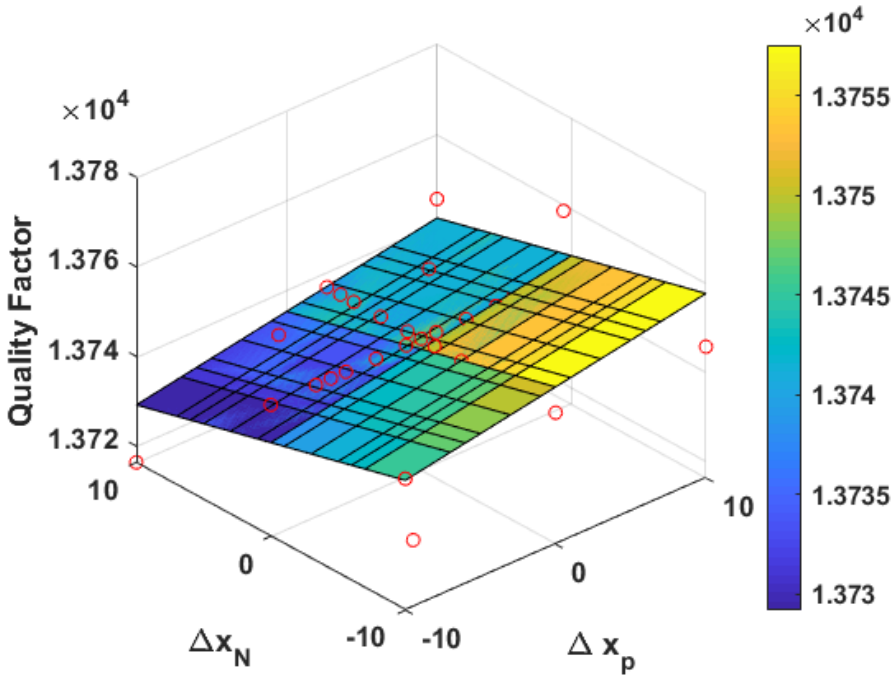


(b)

Figure 6-6: The effect of a PN junction P doping side mask exposure variation on, (a) the resonance wavelength (b) the effective refractive index variation (Δn_{eff}) as a function of applied voltage, for different ΔP.



(a)



(b)

Figure 6-7: Response surface models developed at -10 V for mask location variation impact, (a) the extinction ratio, and (b) the ring quality factor. The red dots represent the DOE data points used for training and testing.

6.2.3 Doping Concentration

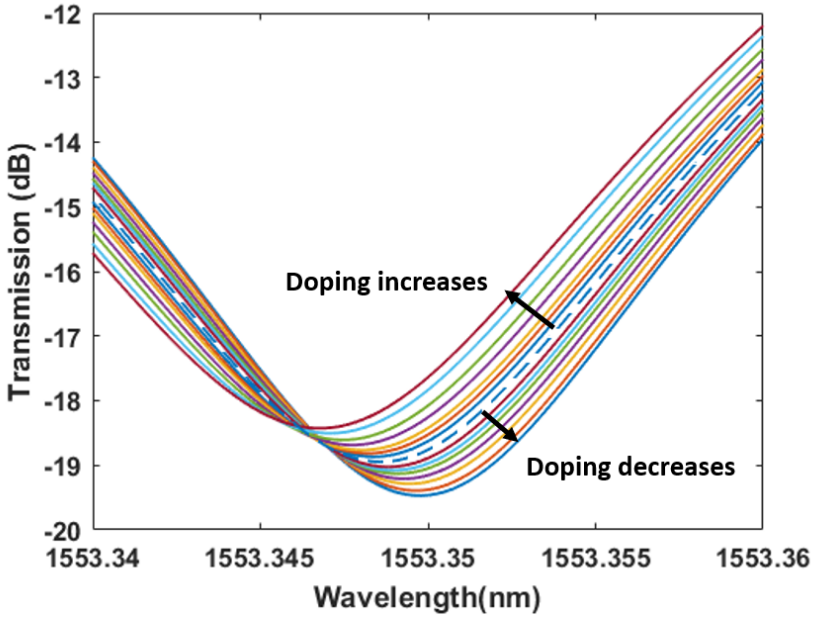
The doping concentration can vary from one location to another within the die or wafer, due to dose, energy, angle, or other ion implant dependencies. Accordingly, understanding the effect of doping concentration variation is also of interest.

We start by considering up to 20% variation in the doping concentration in either N or P doping nominal ($2.5\times10^{17}\ cm^{-3}$) values. The resulting variation in the resonance location is shown in Fig. 6-8. This is caused by the variation in the effective index, Fig. 6-9. From Figs. 6-8 and 6-9, we can see that the variations in the P-doping are more impactful than the N-doping variation, since the spread of the variation in Δn_{eff} is larger with the P doping variation as reflected by a larger shift in the resonance location. This is expected based on Eq. 6.3.

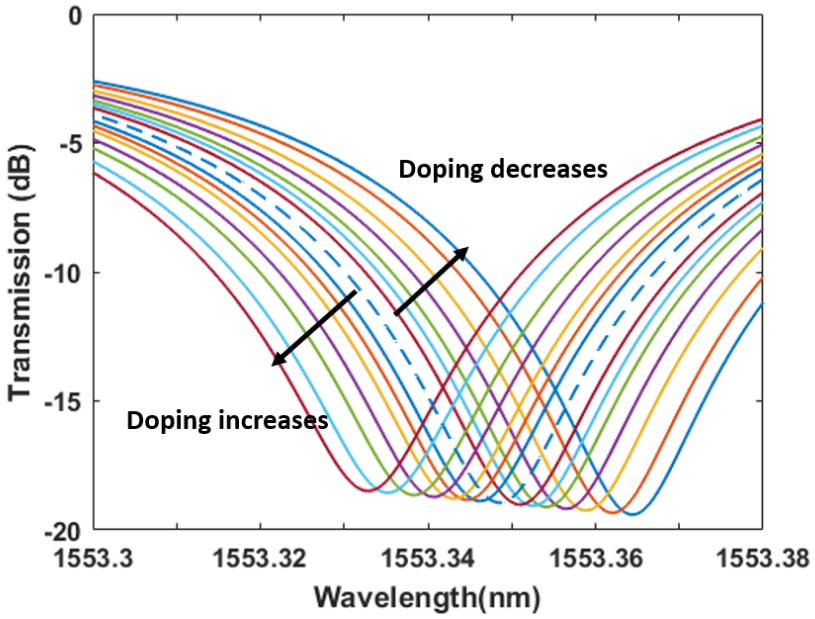
Similar to the case of mask exposure variation, we build response surface models that capture the effect of the doping concentration variation on the various performance parameters. In this case, functions f_i for each performance parameter are developed to depend on the N and P doping concentration and the bias voltage with the form $f_i(\Delta N', \Delta P', V)$, where $\Delta N' = \frac{N}{N_{nominal}}$ and $\Delta P' = \frac{P}{P_{nominal}}$. Our DOE simulation data points are again divided into training and test sets, with 70/30 split ratio. Polynomial regression, third-order in the features, is used to fit the response surface models. The R^2 values for all of these responses is at least $\sim$0.98. Figures 6-11 and 6-10 show example response surface models that capture the effect of doping concentration variation on these performance parameters.

6.3 Summary

The effect of representative process variations within the active region of a silicon ring modulator on the modulation is examined. Specifically, the effect of mask exposure and doping concentration variation is analyzed. The effect of the doping variation is greater than the effect of mask exposure variation. The mask exposure variation causes a variation in the resonance wavelength of almost 0.005 nm when there is an over exposure 10 nm in the mask location. However, for the case of doping variation,

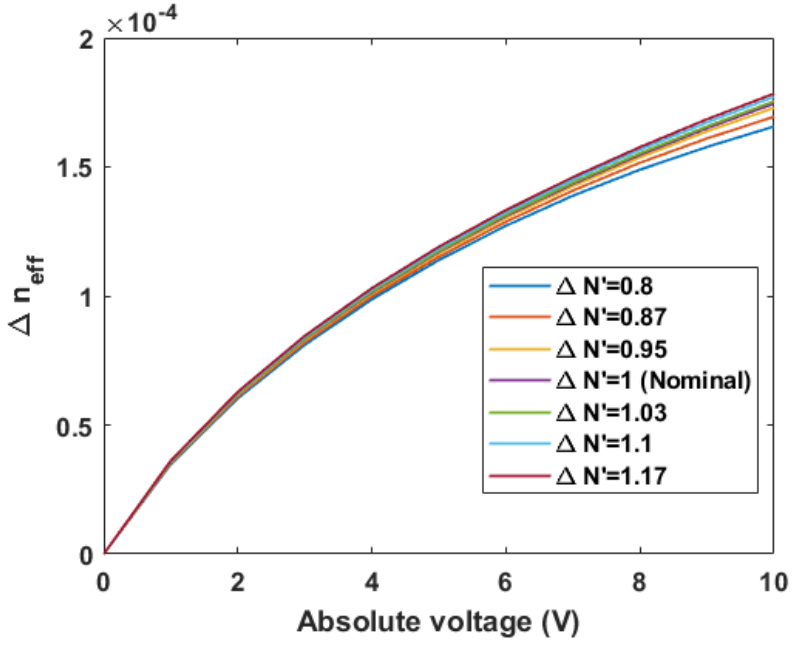


(a)

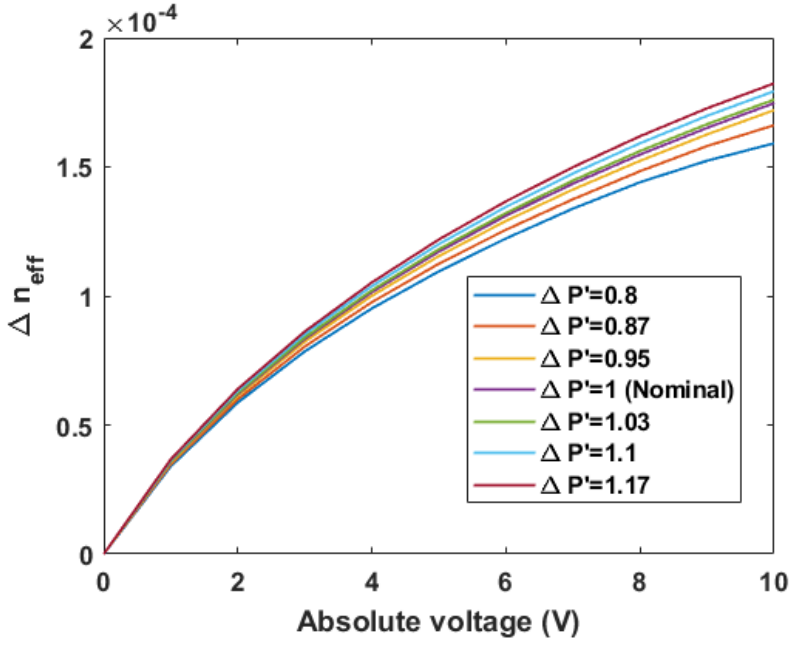


(b)

Figure 6-8: The effect of doping variation at 0 V on the resonance wavelength when (a) N-doping changes from -20% to +20%, and (b) P-doping changes from -20% to +20%. The dashed line is the nominal doping case.

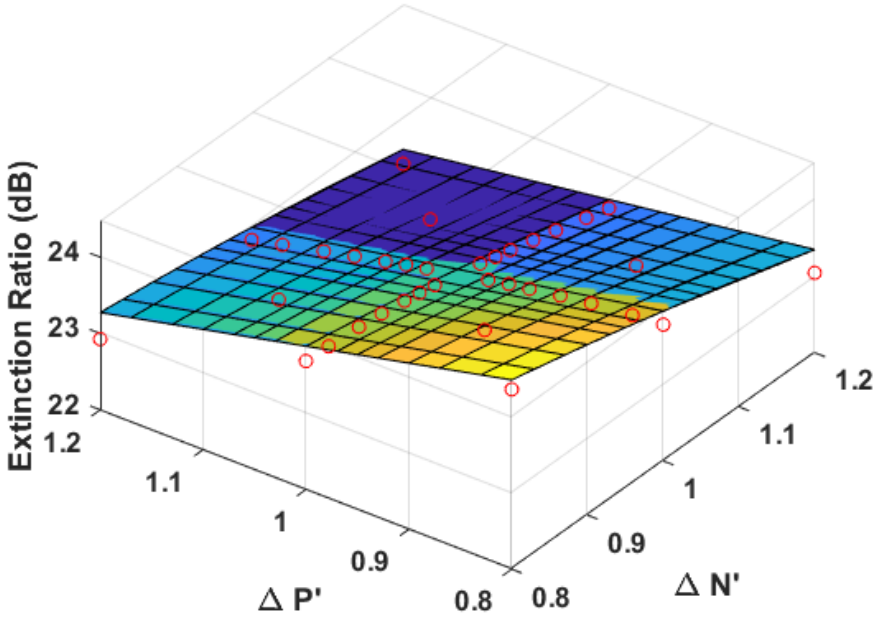


(a)

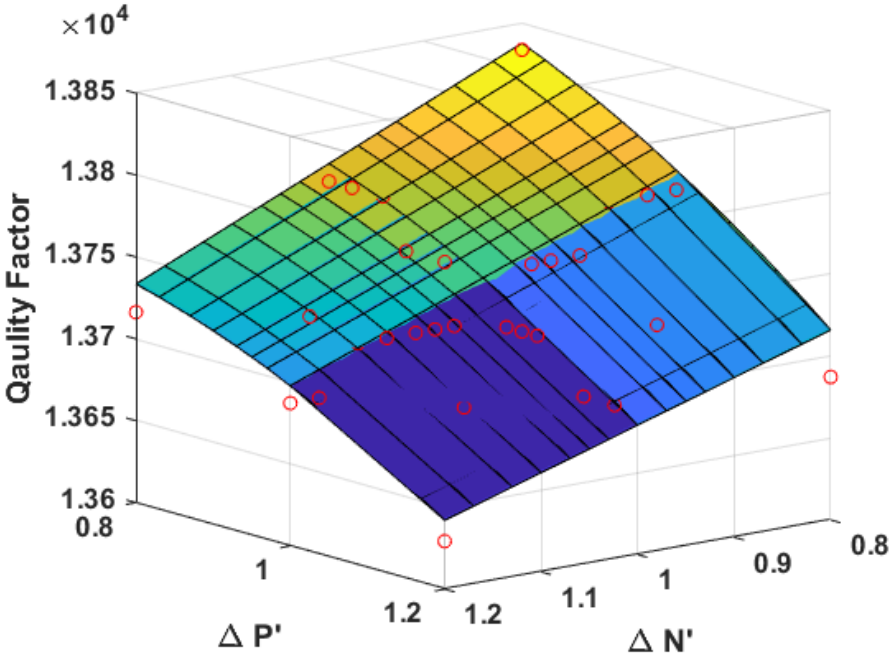


(b)

Figure 6-9: The effect of doping variation at 0 V on the PN junction effective index (Δn_{eff}) on PN junction when (a) N-doping is changing, and (b) P-doping is changing.

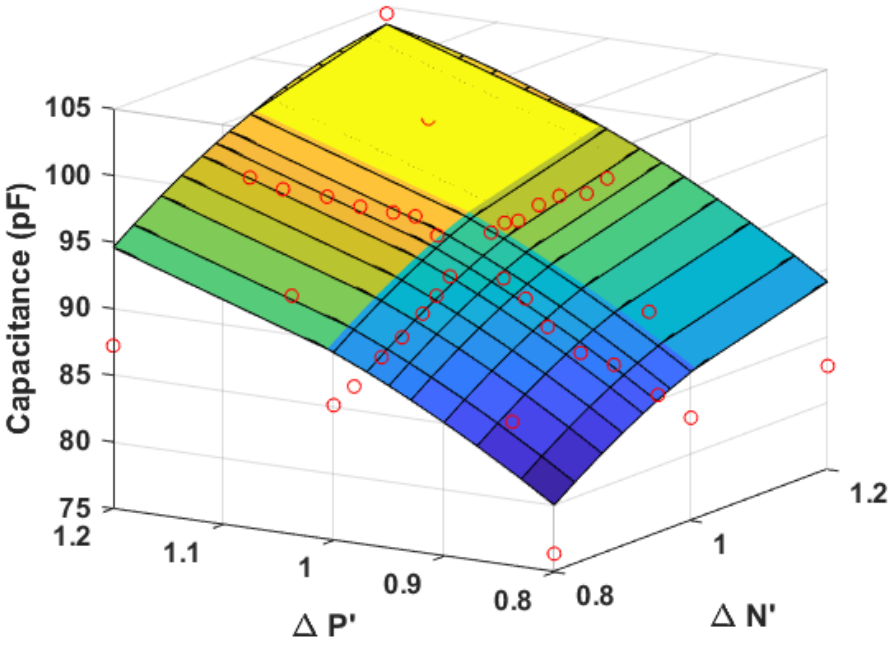


(a)



(b)

Figure 6-10: Response surface models at -10 V for the doping concentration variation impact, (a) the extinction ratio, and (b) the ring quality factor. The red dots represent the data points used for training and testing.



(a)

Figure 6-11: Response surface model for the doping concentration variation impact on PN junction capacitance at -10 V.

the resonance wavelength shift is only 0.02 nm with a 10% change in the doping concentration. These results show that the variation in the active (modulation) section in the silicon ring resonator, i.e., due to mask exposure and doping concentration, have much smaller impact than do geometric variation and LER discussed in previous chapters. However, the compact models relating voltage to modulation will find additional application in Chapter 7, to variation compensation in PICs.

Chapter 7

Photonic Integrated Circuit Simulation with Variation Impact Analysis

In this chapter, we use the models and analysis methods developed and presented in the previous chapters, to simulate a larger photonic integrated circuit (PIC) system. Importantly, these models and methods enable us to analyze the impact of process variations on PIC performance. We compare the time and computational requirements needed for the simulation using the traditional simulation methods and the use of our developed models and analysis.

Section 7.1 overviews the PICs considered here, and maps the components to compact models developed in previous chapters. In Sections 7.2 and 7.3, we use circuit simulation to analyze PIC behavior and obtain eye diagram metrics for checking the interference of signal, and calculate the delay an N CROW achieves. In Section 7.4, we present thermal modulation for compensating for process variations, and the power budget required for such compensation. Section 7.5 shows the effects of environmental variations and compares it to process variations. Finally, Section 7.6 summarizes the chapter.

7.1 PIC Systems

We analyze three PICs systems examples, shown in Fig 7-1, using the previously developed compact models in previous chapters. Specifically, we use the compact model for the half ring presented in Chapter 3 for variation analysis, and the modulator models presented in Chapter 6 for power and compensation analysis. A transceiver is an important building block in telecommunications, used to transmit and receive data over a data link. A key photonic building block of an optical transceiver is the ring, either as a modulator at the transmitting end or as a filter at the receiving end, as shown in Fig. 7-1(a). In addition to transceivers, we use the compact model to build a CROW, as shown in Fig. 7-1(b). Thus, we first investigate the behavior of the PIC using the half ring compact models we developed earlier, when it is fed by a modulating signal generated by a pseudo-random bit sequence with bit-rate of 1 $GBits/s$, shown in Fig. 7-2, and having a power level of 0 dBm (1 mW) to obtain the delay the signal experiences and the response to a modulated input. Moreover, we use the variation-aware compact models developed for the half ring to thermally modulate ring, as shown in Fig. 7-1(c). This PIC setup will enable calculating the power budget required to compensate for various process variations.

7.2 Eye Diagram

An eye diagram [64] represents the signal when repetitively sampled with respect to time. It is a valuable means to qualitatively analyze and evaluate the behavior of the system. The eye diagram calculated for the a single ring (transmitter ring modulator) is shown in Fig. 7-3 which shows that the output drops about 35% which is related to the transmission level of -2.3 dB. We also feed the modulated signal into a 28 ring CROW, to consider how this long structure would affect the signal. The output signal at the drop port of the 28 ring CROW is shown in Fig. 7-4(a), and the eye diagram of the signal received at the drop port at the passband frequency is shown in Fig. 7-4(b). We see that the output signal has a level of ~ 0.08 mW which is an

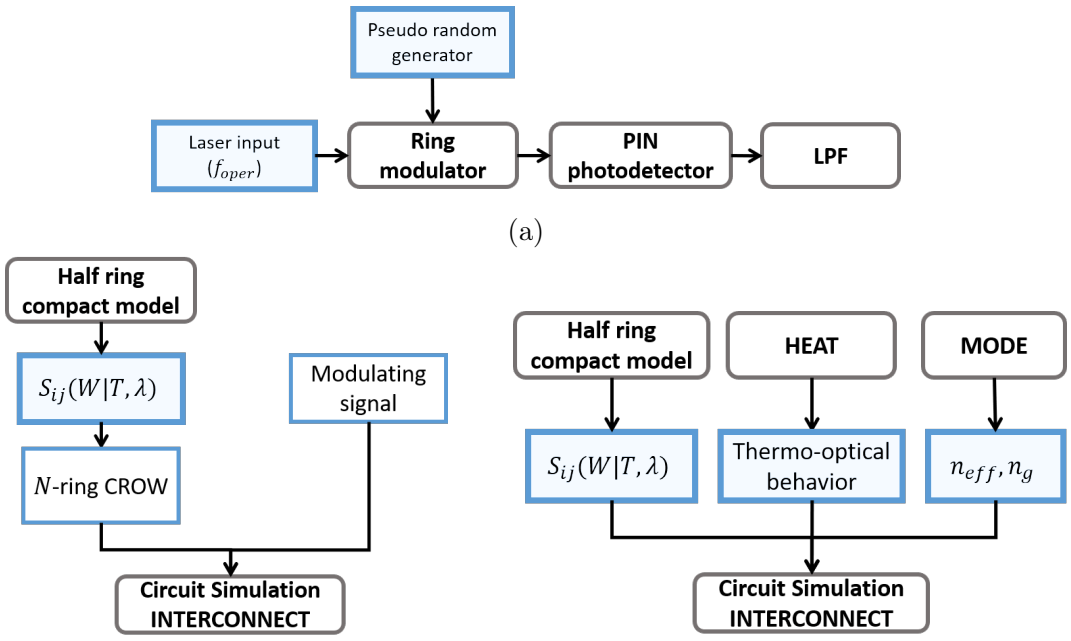


Figure 7-1: (a) Block diagram for simulating transceiver circuit. (b) Simulation flow for PIC consisting of N-rings having a modulated input signal using the developed compact models foe the half ring in Chapter 3. (b) Simulation flow for thermal modulation of a ring using the developed compact models in Chapters 3 and 6.

expected value based on the response of the 28 ring CROW shown in Fig. 3-3, which has a passband with $-11\ dB$. The eye diagram shows that the output signal will be experiencing a bit error rate (BER) of 0, in this ideal case when there is no variation or circuit and environmental noise.

7.3 Delay

The CROWs [3] are typically used as delay units or buffers, so measuring or being able to predict the delay that an N-ring CROW can provide becomes important, i.e., to determine the number of rings (N) needed to achieve a required delay. Accordingly, we use the compact models developed for the half ring to analyze an N-ring CROW. The circuit simulation setup is shown in Fig. 7-5. We feed the circuit with an optical impulse at $\lambda = 1554.17\ nm$ (the CROW passband center) at time t_0, and measure the time t_1 when the signal arrives at the drop port (port 2 in Fig. 7-5). The time

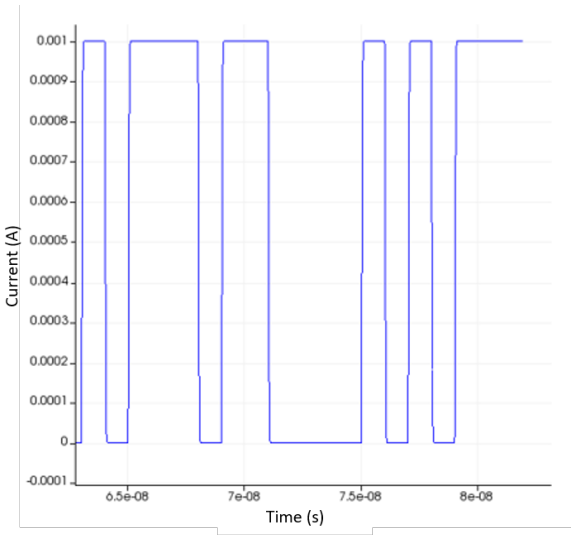


Figure 7-2: Level of the input sequence with bitrate of 1 $GBits/s$ and power of 1 mW, used to drive the ring modulator.

$t_1 - t_0$ is the delay this N-ring CROW would provide.

A 28 ring CROW and a 100 ring CROW are examined. For the 28 ring CROW, the delay is found to be 60 ps, while for the 100 ring CROW it is 172 ps. These results are close to those in [37] which reports 220 ps for fabricated CROWs. This discrepancy between the simulated and literature value is attributed to several factors, among which is the slight difference in the CROW geometry.

It is to be noted that, in all of the above simulations, with either 1, 28 or 100 rings, the simulation time to obtain eye diagram or find the delay is the INTERCONNECT circuit simulation time; this ranges from seconds to a few minutes depending on the circuit size. However, without the use of our S-parameter compact models, one would have to simulate the half ring to get its S-parameters. In that case, the simulation time would be dominated by the FDTD simulation step to generate the ring S-parameters, which takes hours. By doing these simulations up front using DOE and constructing the parameterized compact models, rapid PIC circuit evaluation, exploration, and optimization becomes possible.

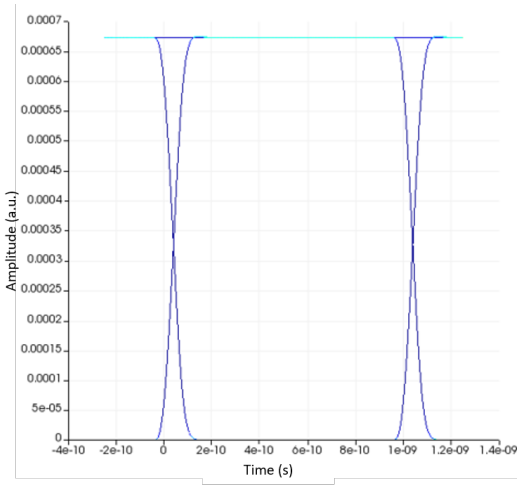


Figure 7-3: Eye diagram for a single ring, calculated using the compact models developed in Chapter 3 for the half ring.

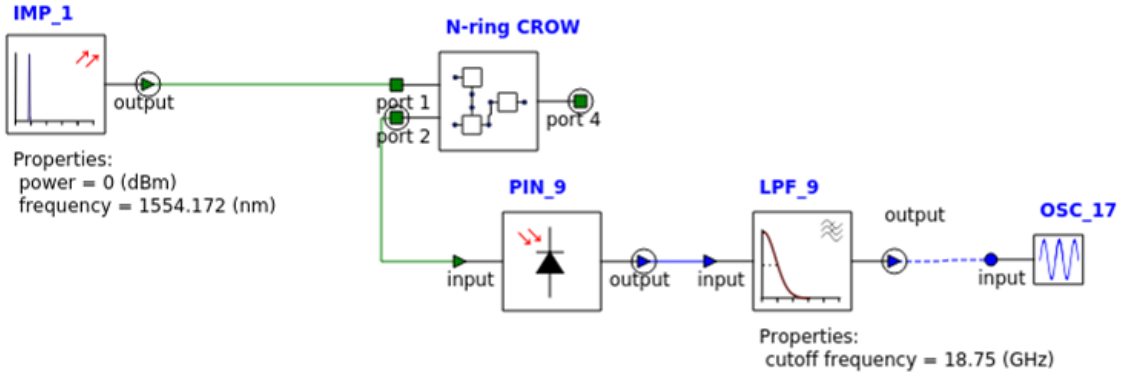
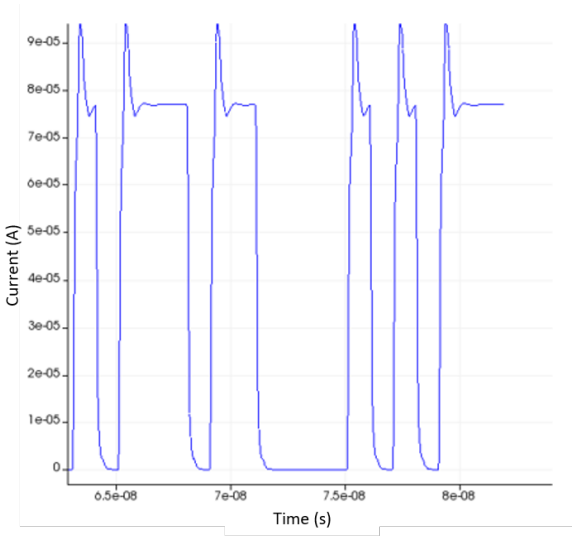


Figure 7-5: The circuit simulation setup used for measuring the delay of an N ring CROW.

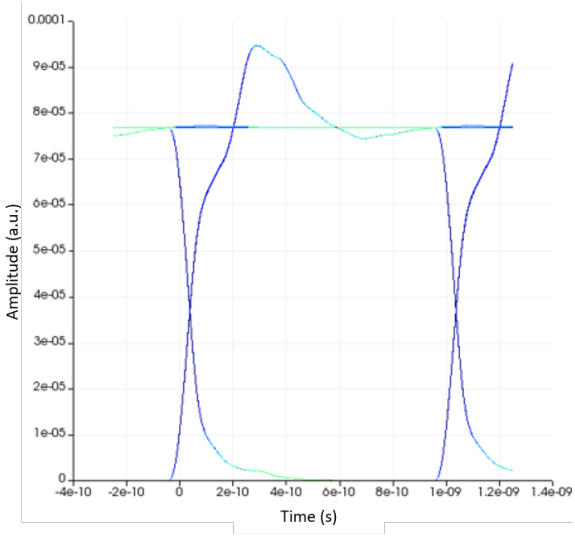
7.4 Thermal Modulation and Compensation

In Chapter 6, we analyzed the electro-optical modulation of a ring resonator, and considered its variation sensitivity. In terms of basic operation, electro-optical modulation of the ring causes a small shift or change in the resonance frequency, and advantageously has a fast response time. Accordingly, electro-optical modulation is usually used for data transfer.

For coarse tuning of the resonance frequency, i.e., to correct for process variations, thermal modulation [65] is usually the best option. This is due to the the large



(a)



(b)

Figure 7-4: (a) The output signal from the drop port after passing through a low pass filter (LPF) of the 28 ring CROW. (b) Eye diagram for a 28 ring CROW, calculated using the compact models developed in Chapter 3 for the half ring.

thermo-optic coefficient for silicon ($\frac{dn}{dT}$ $[K^{-1}]$), which enables substantial change in the effective index n_{eff} with modest changes in temperature, T [66]. However, thermal modulation comes with a slow response compared to the requirements in modern communications and interconnects, and so is only used for signal modulation in very low bitrate scenarios.

7.4.1 Simulation Setup

To implement and investigate the effect of thermal modulation, a heater is placed above the device, such that when a voltage is applied to the heater/wire, current flows and the power $P = V \cdot I$ resistively heats the wire. Thus the underlying photonic component experiences a higher temperature, which in turn changes the refractive index.

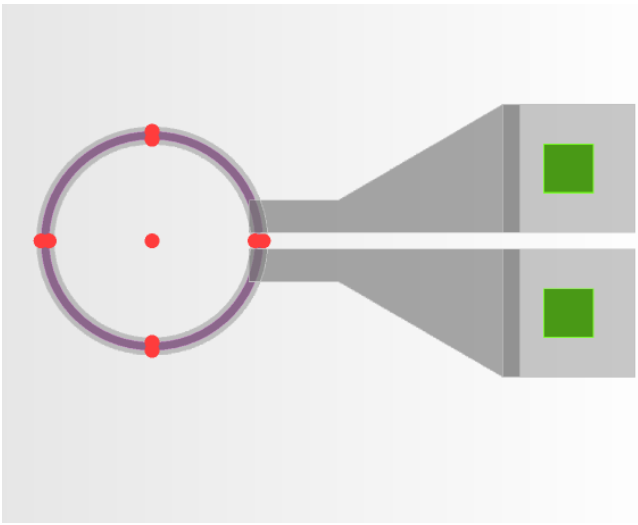
For our analysis, we again use the half ring models developed in Chapter 3, but now augmented to account for thermal impact. Specifically, we need to correct for the group delay τ_g, such that the S-parameter based models previously developed for the half ring can be used. However, for thermal modulation we heat the ring curvature sections, and thus simulate those sections using HEAT [24]. So, the light path for the ring sections that are described by the S-parameters, i.e., the sections that are not modulated, changes from $2\pi R + 2L_c$ to $2L_c$ where R is the ring radius and L_c is the coupling length. Accordingly, the group delay also changes, since the group delay is defined as:

$$\tau_g = \frac{L \cdot n_g}{c} \tag{7.1}$$

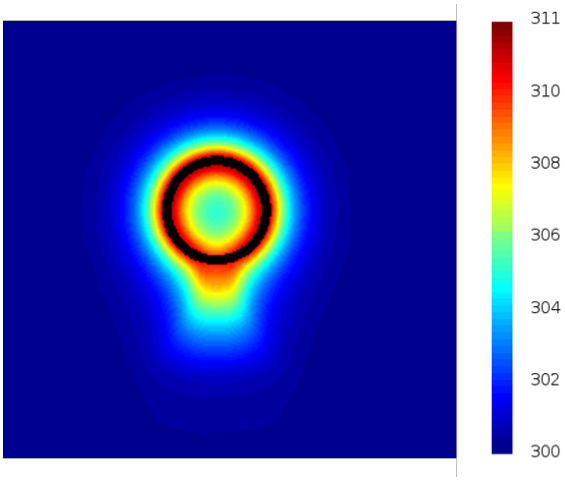
where L is the length of the ring or the path the light experiences, n_g is the group index, and c is the speed of light.

We place a heater above a ring with radius 6.5 μm, the same size as the half ring analyzed in Chapter 3, as shown in Fig. 7-6(a). We apply voltage to the heater, so that temperature increase affects the underlying ring, Fig. 7-6(b).

Using the thermo-optic coefficient of silicon, $\frac{dn}{dT} = 1.86 \times 10^{-4} K^{-1}$, along with the



(a)



(b)

Figure 7-6: (a) Layout used to simulate the thermally modulated ring in HEAT. (b) The heat distribution in the ring due to applying a voltage of 0.4 V.

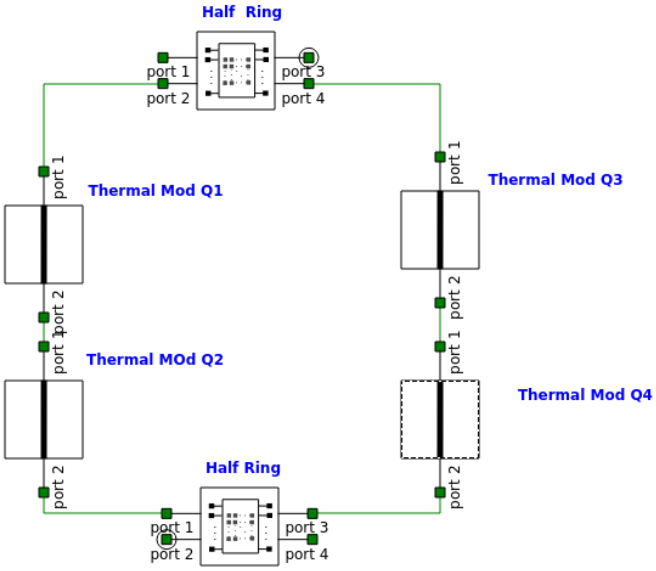
silicon effective index, group index, the heat change across the curved section of the ring, and the S-parameters generated from the original ring models after modification in the circuit simulator, Fig. 7-7(a), we obtain the modulated behavior. The response is shown in Fig. 7-7(b), where we see the change in the resonance frequency resulting from changing the applied voltage. From these results, we see that for thermal modulation, a large shift in the resonance can be achieved. Thus, if we need to correct for substantial resonance shifts due to geometric variations, such as those in Fig. 3-3 where shifts of more than 1 nm are seen, thermal modulation can be used.

7.4.2 Power for Modulation

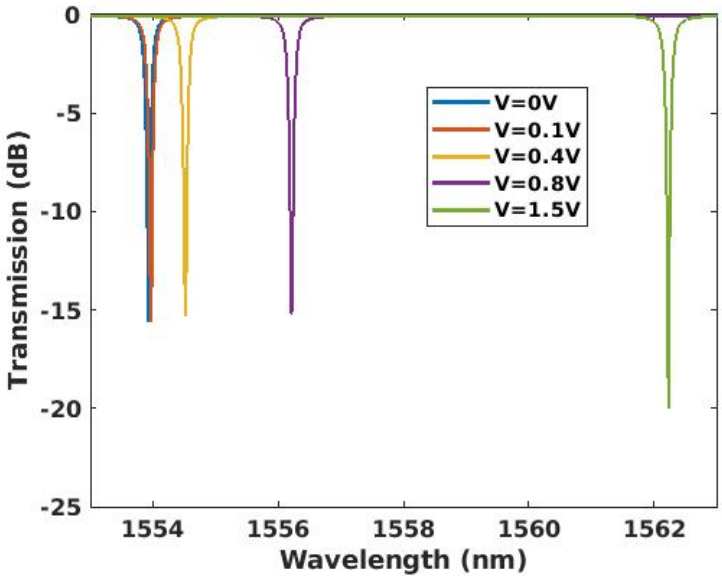
As discussed above, thermal modulation can be used for coarse frequency tuning and correcting for resonance shifts due to variations. However, unlike electro-optical modulation, thermal modulation requires a relatively substantial flow of current through resistive heater element. Thus, it is important to calculate and predict the power consumed for correction of such process variations.

The current that flows through the heater when a voltage is applied increases with increasing voltage, i.e., the power needed for larger correction increases. Figure 7-8(a) shows the circuit simulation results for a ring that is thermally modulated. We observe resonance location change by varying the power that is consumed. The change in the resonance wavelength ($\Delta\lambda$) with respect to power is shown in Fig. 7-8(b); this sensitivity is $\frac{\Delta\lambda}{\Delta P} = 0.1\ nm/0.172\ mW$. Thus, to achieve a 0.1 nm shift in the resonance wavelength, 0.172 mW of power is needed.

Now, to find the power needed to correct for a geometric variation that affects resonance, we need to analyze how these variations influence the resonance location. To do this, we use the S-parameters based compact to consider rings experiencing different geometric variations. Figure 7-9 shows the variation in the resonance location due to variation in the ring silicon thickness or width. In the case of thickness variations, the variations are large, such that 1 nm variation in thickness causes almost a 2 nm variation in the resonance location. The change in the resonance wavelength ($\Delta\lambda$) with respect to the variation in the geometry (ΔT and ΔW) is shown in Fig. 7-10.

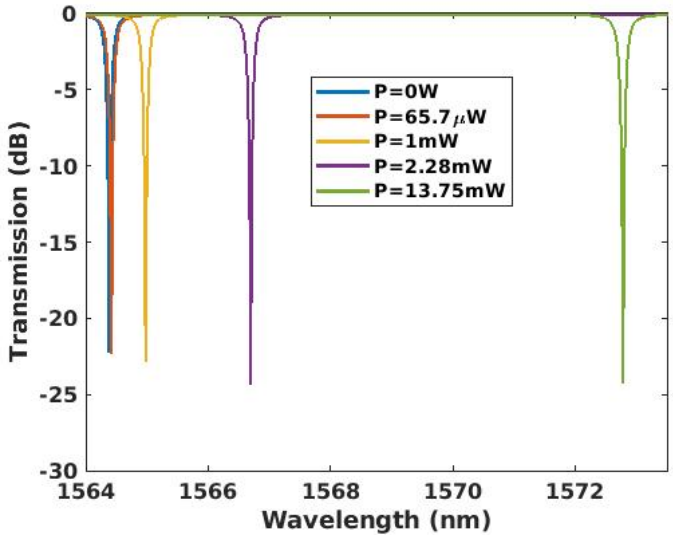


(a)

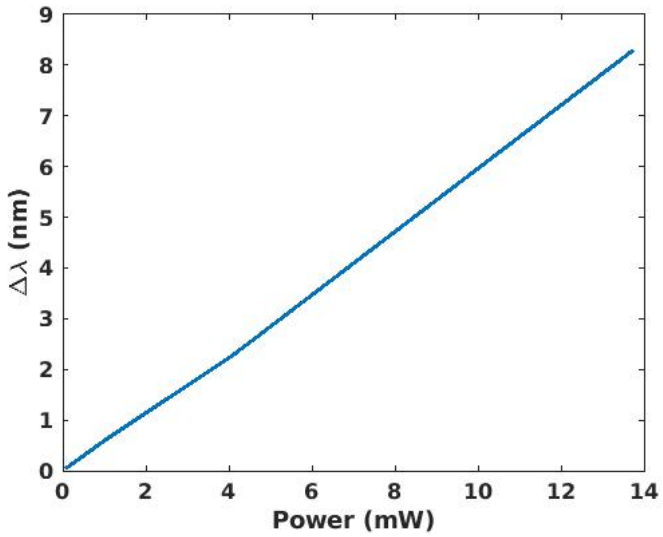


(b)

Figure 7-7: (a) Circuit simulation setup used to simulate the thermally modulated ring in INTERCONNECT. (b) Response of a ring thermally modulated with $W = 500\ nm$ and $T = 220\ nm$ by different applied voltages.

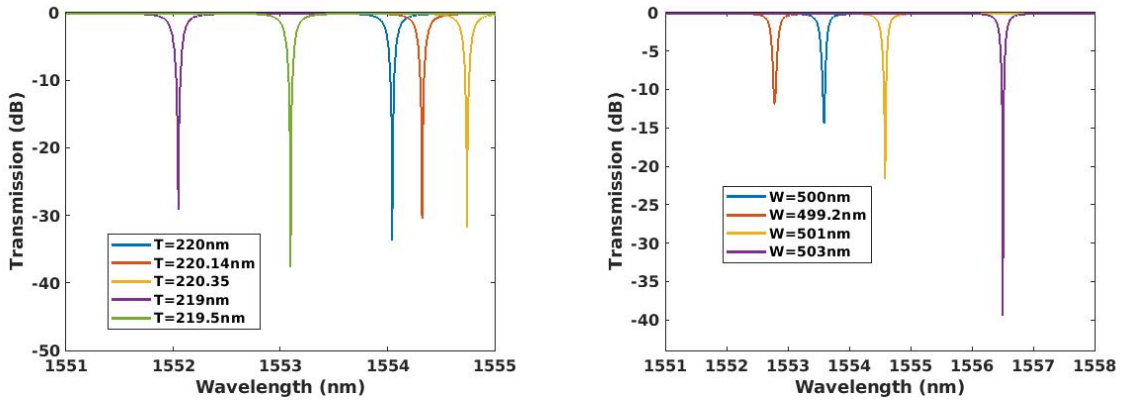


(a)



(b)

Figure 7-8: (a) The change in the resonance location for different applied thermal tuning power, for a ring with $T = 220\ nm$ and $W = 500\ nm$. (b) The resonance shift with respect to the power consumption in the heater.



(a) (b)

Figure 7-9: (a) Change in the resonance location for ring resonators. (a) $W = 500\ nm$ and varying thickness. (b) $T = 220\ nm$ and varying width.

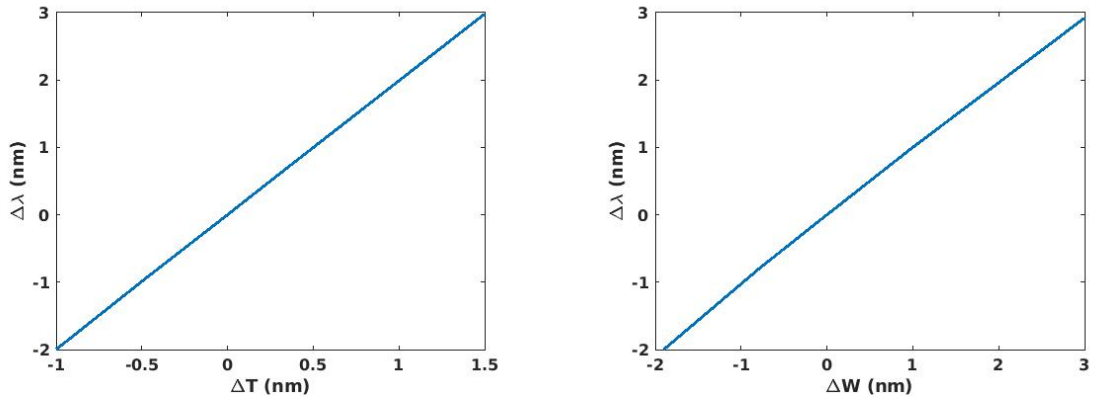
The sensitivity for thickness is calculated to be $\frac{\Delta\lambda}{\Delta T} = 2\ nm/1\ nm$. For the width variations, the sensitivity is $\frac{\Delta\lambda}{\Delta W} = 1\ nm/1\ nm$, i.e., a 1 nm variation in the width causes a 1 nm shift in the resonance wavelength.

Combining the two sensitivities, $\frac{\Delta\lambda}{\Delta T}$ and $\frac{\Delta\lambda}{\Delta P}$ we find that 3.44 mW ia required to correct for the shift caused by a 1 nm change in the thickness, i.e., $\frac{\Delta P}{\Delta T} = 3.44\,mW/1\ nm$. For the width, the power per width variation is $\frac{\Delta P}{\Delta W} = 1.72\ mW/1\ nm$.

When thinking of using thermal modulation to correct for LER effect on a ring, it turns out that this cannot be achieved efficiently because LER not only shift the resonance location but also distorts the S-parameters, as shown in Fig. 7-11 and consequently the response. For LER, when using thermal modulation we can shift the resonance location due to the variations in n_{eff} due to temperature change however, the distortion in the S-parameter is still existing to minimize or remove its effect, the solution will include extra processing steps while fabricating to smear these perturbations.

7.5 Environmental Variations and Compensation

Environmental variations are systematic variations that arise during operation of the circuit. For example, depending on how photonic and electronic components



(a) (b)

Figure 7-10: Variation in the resonance wavelength with respect to (a) variations in thickness relative to nominal thickness of 220 nm, and (b) variations in width relative to nominal thickness of 500 nm.

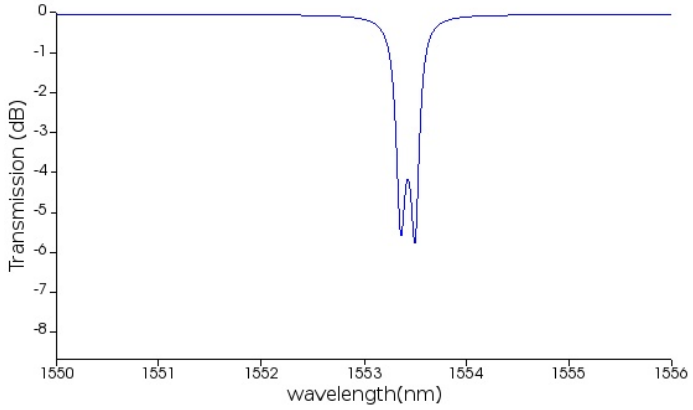


Figure 7-11: The response of a ring experiencing LER with $A = 4\,nm$ and $L_c = 30\,nm$.

are distributed within the chip, some components may dissipate power which can cause variation in local temperatures within the chip [11]. From the above results for thermal modulation, we can see that applied heater power of 1 mW causes a local change in the temperature of $\sim$10.8 ^{o}C. So, for example, if the chip experiences a temperature variation of 10 ^{o}C, using the above calculated sensitivities, this would affect the performance of an affected ring resonator, causing a $\sim$ 0.6 nm shift in the resonance wavelength. Similarly, a 30 ^{o}C perturbation would cause a shift of $\sim$ 1.75 nm in resonant wavelength. Comparing this to the shift in resonance due to process variations introduced, we see that geometric process variations can be more serious and impactful than this example environmental variation. For a 0.5% variation in the thickness (1 nm for a 220 nm silicon thickness), a shift of 2 nm in the resonance of our ring is observed, and 3.44 mW power is needed to correct for it.

7.6 Summary

In this chapter, we use our previously developed variation-aware compact models, with some corrections, to analyze photonic integrated circuits (PIC). These circuits are based primarily on rings. We consider modulating a ring for data transmission; this is typically done using electrical modulation due it its ability to support high frequency and fast communication. We also use the models to analyze CROWs, examining the resulting eye diagram, BER, and signal delay the signal experience. Finally, we consider thermal modulation to correct for the geometric variations, and analyze the power needed to correct for example process variations.

Chapter 8

Conclusions and Future Work

Silicon photonics is becoming an attractive technology platform to meet future demanding communications speed and power requirements, as well as to enable new photonic functionalities in sensing, quantum computation, and other applications. However, it lacks variation-aware compact models against process variations. Thus, producing high yield photonic integrated circuits has been challenging, due to the difficulty in design, analysis, and optimization of photonic components and circuits to account for process variations.

8.1 Contribution

In this thesis, we present methodologies to enable understanding, analyzing and modeling the effect of process variations on photonic components, both passive and active. Such methods support variation-aware photonic design for manufacturability (DFM).

For systematic process variations, S-parameter based variation-aware compact models are developed for geometric variations. In the case of CROWs, we perform simulations of the fundamental CROW sub-components (the half ring) and considering geometric variations that cover the values we expect due to fabrication variations. Based on these simulations, compact models are developed providing an input-output relationship between the geometric variations (width and thickness in this case) and the half ring S-parameter response. The models are then used to interpret the per-

formance, behavior and predict yield of full CROWs in the presence of geometric variations. Moreover, when using the compact models in circuit simulation, substantial speed-up in the simulation can be achieved, enabling Monte Carlo simulations and design exploration. We further extend the method used in developing CROW compact models, to develop models for large silicon nitride rings. Simulation results using these models are compared with fabrication data and found to be in acceptable agreement. Such models again provide results faster than full FDTD simulation.

In the case of line edge roughness (LER), which is a random process variation, statistical analysis can help predict and express the uncertainty in photonic device behavior due to the effect of LER. For the Y-branch, such statistical analysis is performed using ensemble virtual fabrication. We further develop behavioral models for the Y-branch transmission statistics in the presence of LER. These models help predict performance and yield, enabling Monte Carlo simulations in much less time than having to go through the ensemble simulations for each yield analysis. Variation-aware compact models are also developed for CROWs in the presence of LER, where these models take into account various types of randomness in CROW behavior introduced by the LER. These models are used to efficiently predict the yield and performance using circuit simulations.

We examine the effect of process variations on active devices, and show how the variation in the active region of a PN junction silicon ring, either doping concentration or mask exposure variation, can change slightly the behavior from the nominal design. Response surface models are developed for different performance metrics in the presence of these variations. Also, thermal modulation is examined, and the power needed to correct for the effect of geometry on performance is highlighted for representative resonant photonic devices.

Since all of these variations are foundry specific, methods for extraction and decomposition of spatial variations including wafer-level, pattern-dependent, and random variations, are presented. These extracted variations can then be used in conjunction with variation-aware compact models for ensemble study of variation impact. Finally, we use our developed compact models to simulate representative photonic in-

tegrated circuits, and calculate the power budget required to tune the circuits to compensate for process and environmental variations.

Together, these methods and models can be applied to different photonic components to help build extended process design kits (PDKs) that will enable silicon photonic designers to predict and optimize behavior, performance, and yield of complex silicon photonic devices and circuits in the face of unavoidable manufacturing variation, just as IC designers do today.

8.2 Future Work

The methods and models presented in the thesis are just the start for the extended PDKs; there is still much work to be done to help silicon photonics designers effectively and efficiently design for manufacturability like the case with IC designers.

Future work includes extending these proposed methods for modeling other components, and to examine other process variations, like etch lateral ratio. However, fabricating and measuring device responses and comparing to the models is a crucial step to further validate the methodologies we have presented.

Moreover, the compact models themselves could be extended. For rings, more parameters could be included, i.e., such that the S-parameter ring compact model would have the form $S_{ij}(R, G, L_c, \lambda, T, W)$ where R is the ring radius, G is the coupling gap, L_c is the coupling length, λ is the operating wavelength, T is the silicon thickness, and W is the silicon width. Such an extended model could facilitate joint design space and variation exploration.

Another interesting opportunity is developing semi-empirical models using a mixture of measurements and simulations. This could help generate better models, where measurements will give insight into effects that cannot be directly included and captured in current physical simulations.

An essential step for silicon photonics growth is design for robustness against these process variations, such that device designs are created that can overcome or compensate for the effect of variations.

Finally, we recommend that variation-aware compact models and design guide rules should be made available in extended PDKs, based on and provided by the specific foundry. These variation-aware models and rules are crucial to enable PIC designers to improve and enhance both performance and yield.

Bibliography

[1] J. Sun, E. Timurdogan, A. Yaacobi, E. S. Hosseini, and M. R. Watts, "Large-scale nanophotonic phased array," *Nature*, vol. 493, no. 7431, pp. 195–199, 2013.

[2] Y. Zhang, S. Yang, A. E.-J. Lim, G.-Q. Lo, C. Galland, T. Baehr-Jones, and M. Hochberg, "A compact and low loss Y-junction for submicron silicon waveguide," *Optics Express*, vol. 21, no. 1, pp. 1310–1316, 2013.

[3] M. L. Cooper, G. Gupta, W. M. Green, S. Assefa, F. Xia, Y. A. Vlasov, and S. Mookherjea, "235-ring coupled-resonator optical waveguides," in *Conference on Lasers and Electro-Optics*, p. CTuHH3, Optical Society of America, 2010.

[4] J. E. Bowers, T. Komljenovic, M. Davenport, J. Hulme, A. Y. Liu, C. T. Santis, A. Spott, S. Srinivasan, E. J. Stanton, and C. Zhang, "Recent advances in silicon photonic integrated circuits," in *Next-Generation Optical Communication: Components, Sub-Systems, and Systems V*, vol. 9774, p. 977402, International Society for Optics and Photonics, 2016.

[5] L. Chrostowski and M. Hochberg, *Silicon Photonics Design: From Devices to Systems*. Cambridge University Press, 2015.

[6] C. Sun, M. T. Wade, Y. Lee, J. S. Orcutt, L. Alloatti, M. S. Georgas, A. S. Waterman, J. M. Shainline, R. R. Avizienis, S. Lin, B. R. Moss, R. Kumar, F. Pavanello, A. H. Atabaki, A. J. Cook, Henry M.and Ou, Y.-H. Leu, Jonathan C.and Chen, K. Asanović, R. J. Ram, M. A. Popović, and V. M. Stojanović, "Single-chip microprocessor that communicates directly using light," *Nature*, vol. 528, no. 7583, p. 534, 2015.

[7] C. García-Meca, S. Lechago, A. Brimont, A. Griol, S. Mas, L. Sánchez, L. Bellieres, N. S. Losilla, and J. Martí, "On-chip wireless silicon photonics: from reconfigurable interconnects to lab-on-chip devices," *Light: Science & Applications*, vol. 6, no. 9, pp. e17053–e17053, 2017.

[8] K. Zinoviev, L. G. Carrascosa, J. Sánchez del Río, B. Sepulveda, C. Domínguez, and L. M. Lechuga, "Silicon photonic biosensors for lab-on-a-chip applications," *Advances in Optical Technologies*, 2008.

[9] C. V. Poulton, A. Yaacobi, D. B. Cole, M. J. Byrd, M. Raval, D. Vermeulen, and M. R. Watts, "Coherent solid-state LIDAR with silicon photonic optical phased arrays," *Optics Letters*, vol. 42, no. 20, pp. 4091–4094, 2017.

[10] P. Bhargava, T. Kim, C. V. Poulton, J. Notaros, A. Yaacobi, E. Timurdogan, C. Baiocco, N. Fahrenkopf, S. Kruger, T. Ngai, Y. Timalsina, M. R. Watts, and V. Stojanović, "Fully integrated coherent lidar in 3d-integrated silicon photonics/65nm CMOS," in *2019 Symposium on VLSI Circuits*, pp. C262–C263, IEEE, 2019.

[11] M. Orshansky, S. Nassif, and D. Boning, *Design for Manufacturability and Statistical Design: A Constructive Approach*. Springer Science & Business Media, 2007.

[12] W. A. Zortman, D. C. Trotter, and M. R. Watts, "Silicon photonics manufacturing," *Optics Express*, vol. 18, no. 23, p. 23598, 2010.

[13] T.-H. Yen and Y.-J. Hung, "Fabrication-tolerant CWDM (de)multiplexer based on cascaded Mach-Zehnder interferometers on silicon-on-insulator," *Journal of Lightwave Technology*, vol. 39, no. 1, pp. 146–153, 2021.

[14] D. Melati, A. Melloni, and F. Morichetti, "Real photonic waveguides: guiding light through imperfections," *Advances in Optics and Photonics*, vol. 6, no. 2, pp. 156–224, 2014.

[15] R. Wu, C.-H. Chen, T.-C. Huang, R. Beausoleil, and K.-T. Cheng, "Spatial pattern analysis of process variations in silicon microring modulators," in *IEEE Optical Interconnects Conf.*, pp. 116–117, 2016.

[16] S. I. El-Henawy, Z. Zhang, R. Miller, and D. S. Boning, "Photonic device sensitivity analysis methods: towards process variation-aware silicon photonics design," in *Optical Modeling and System Alignment*, vol. 11103, p. 111030P, International Society for Optics and Photonics, 2019.

[17] T.-W. Weng, Z. Zhang, Z. Su, Y. Marzouk, A. Melloni, and L. Daniel, "Uncertainty quantification of silicon photonic devices with correlated and non-gaussian random parameters," *Optics Express*, vol. 23, no. 4, pp. 4242–4254, 2015.

[18] J. M. F. Cabanillas, D. M. Kita, A. Khilo, F. Sedgwick, J. Fini, M. A. Popovic, and M. T. Wade, "Demonstration and fabrication tolerance study of a low-loss, ultra-broadband rapid adiabatic 3-dB coupler in a next-generation 45 nm monolithic electronic-photonic platform," in *Optical Fiber Communication Conference (OFC)*, p. Tu5B.4, OSA, 2021.

[19] Z. Zhang, S. I. El-Henawy, A. Sadun, R. Miller, L. Daniel, J. K. White, and D. S. Boning, "Adjoint-based sensitivity analysis for silicon photonic variations," in *2019 IEEE MTT-S International Conference on Numerical Electromagnetic and Multiphysics Modeling and Optimization (NEMO)*, pp. 1–4, IEEE, 2019.

[20] C. A. Mack, "Generating random rough edges, surfaces, and volumes," *Applied Optics*, vol. 52, no. 7, pp. 1472–1480, 2013.

[21] A. Asenov, S. Kaya, and A. R. Brown, "Intrinsic parameter fluctuations in decananometer mosfets introduced by gate line edge roughness," *IEEE Transactions on Electron Devices*, vol. 50, no. 5, pp. 1254–1260, 2003.

[22] https://www.lumerical.com/products/fdtd-solutions.

[23] https://www.lumerical.com/products/mode.

[24] https://www.lumerical.com/products/charge.

[25] F. Payne and J. Lacey, "A theoretical analysis of scattering loss from planar optical waveguides," *Optical and Quantum Electronics*, vol. 26, no. 10, pp. 977–986, 1994.

[26] D. K. Sparacin, S. J. Spector, and L. C. Kimerling, "Silicon waveguide sidewall smoothing by wet chemical oxidation," *Journal of Lightwave Technology*, vol. 23, no. 8, p. 2455, 2005.

[27] N. Marchack, M. Khater, J. Orcutt, J. Chang, S. Holmes, T. Barwicz, S. Kamlapurkar, W. Green, and S. Engelmann, "Reducing line edge roughness in Si and SiN through plasma etch chemistry optimization for photonic waveguide applications," in *Advanced Etch Technology for Nanopatterning VI*, vol. 10149, p. 101490F, International Society for Optics and Photonics, 2017.

[28] S. I. El-Henawy, R. Miller, and D. S. Boning, "Effects of a random process variation on the transfer characteristics of a fundamental photonic integrated circuit component," in *Optical Modeling and Performance Predictions X*, vol. 10743, p. 107430O, International Society for Optics and Photonics, 2018.

[29] https://www.coventor.com/products/semulator3d/.

[30] A. D. Simard, N. Ayotte, Y. Painchaud, S. Bedard, and S. LaRochelle, "Impact of sidewall roughness on integrated bragg gratings," *Journal of Lightwave Technology*, vol. 29, no. 24, pp. 3693–3704, 2011.

[31] E. Jaberansary, T. M. B. Masaud, M. Milosevic, M. Nedeljkovic, G. Z. Mashanovich, and H. M. Chong, "Scattering loss estimation using 2-d Fourier analysis and modeling of sidewall roughness on optical waveguides," *IEEE Photonics Journal*, vol. 5, no. 3, pp. 6601010–6601010, 2013.

[32] K. K. Lee, D. R. Lim, H.-C. Luan, A. Agarwal, J. Foresi, and L. C. Kimerling, "Effect of size and roughness on light transmission in a Si/SiO2 waveguide: Experiments and model," *Applied Physics Letters*, vol. 77, no. 11, pp. 1617–1619, 2000.

[33] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning*, vol. 112. Springer, 2013.

[34] M. Frean and P. Boyle, "Using gaussian processes to optimize expensive functions," in *Australasian Joint Conference on Artificial Intelligence*, pp. 258–267, Springer, 2008.

[35] C. K. Williams and C. E. Rasmussen, "Gaussian processes for regression," in *Advances in Neural Information Processing Systems*, pp. 514–520, 1996.

[36] C. E. Rasmussen, "Gaussian processes in machine learning," in *Summer School on Machine Learning*, pp. 63–71, Springer, 2003.

[37] F. Xia, L. Sekaric, and Y. Vlasov, "Ultracompact optical buffers on a silicon chip," *Nature Photonics*, vol. 1, no. 1, pp. 65–71, 2007.

[38] W. Bogaerts, P. De Heyn, T. Van Vaerenbergh, K. De Vos, S. Kumar Selvaraja, T. Claes, P. Dumon, P. Bienstman, D. Van Thourhout, and R. Baets, "Silicon microring resonators," *Laser & Photonics Reviews*, vol. 6, no. 1, pp. 47–73, 2012.

[39] https://www.lumerical.com/products/interconnect.

[40] A. I. Khuri and S. Mukhopadhyay, "Response surface methodology," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 2, no. 2, pp. 128–149, 2010.

[41] Z. Lu, J. Jhoja, J. Klein, X. Wang, A. Liu, J. Flueckiger, J. Pond, and L. Chrostowski, "Performance prediction for silicon photonics integrated circuits with layout-dependent correlated manufacturing variability," *Optics Express*, vol. 25, no. 9, pp. 9712–9733, 2017.

[42] Y. Xing, J. Dong, S. Dwivedi, U. Khan, and W. Bogaerts, "Accurate extraction of fabricated geometry using optical measurement," *Photonics Research*, vol. 6, no. 11, pp. 1008–1020, 2018.

[43] S. Selvaraja, L. Fernandez, M. Vanslembrouck, J.-L. Everaert, P. Dumon, J. Van Campenhout, W. Bogaerts, and P. Absil, "Si photonic device uniformity improvement using wafer-scale location specific processing," in *IEEE Photonics Conference 2012*, pp. 725–726, IEEE, 2012.

[44] https://www.mathworks.com/products/matlab.html.

[45] S. B. Papp, K. Beha, P. Del'Haye, F. Quinlan, H. Lee, K. J. Vahala, and S. A. Diddams, "Microresonator frequency comb optical clock," *Optica*, vol. 1, no. 1, pp. 10–14, 2014.

[46] T. J. Kippenberg, R. Holzwarth, and S. A. Diddams, "Microresonator-based optical frequency combs," *Science*, vol. 332, no. 6029, pp. 555–559, 2011.

[47] L. Razzari, D. Duchesne, M. Ferrera, R. Morandotti, S. Chu, B. Little, and D. Moss, "Cmos-compatible integrated optical hyper-parametric oscillator," *Nature Photonics*, vol. 4, no. 1, pp. 41–45, 2010.

[48] T. Udem, R. Holzwarth, and T. W. Hänsch, "Optical frequency metrology," *Nature*, vol. 416, no. 6877, pp. 233–237, 2002.

[49] W. McKinnon, D.-X. Xu, C. Storey, E. Post, A. Densmore, A. Delâge, P. Waldron, J. Schmid, and S. Janz, "Extracting coupling and loss coefficients from a ring resonator," *Optics Express*, vol. 17, no. 21, pp. 18971–18982, 2009.

[50] D. Boning, K. Balakrishnan, H. Cai, N. Drego, A. Farahanchi, K. M. Gettings, D. Lim, A. Somani, H. Taylor, D. Truque, and X. Xie, "Variation," *IEEE Trans. Semicond. Manuf.*, vol. 21, no. 1, pp. 63–71, 2008.

[51] B. Stine, D. Boning, and J. Chung, "Analysis and decomposition of spatial variation in integrated circuit process and devices," *IEEE Trans. Semicond. Manuf.*, vol. 10, no. 1, pp. 24–41, 1997.

[52] Y. Yang, Y. Ma, H. Guan, Y. Liu, S. Danziger, S. Ocheltree, K. Bergman, T. Baehr-Jones, and M. Hochberg, "Phase coherence length in silicon photonic platform," *Optics Express*, vol. 23, no. 13, pp. 16890–16902, 2015.

[53] L. Chrostowski, X. Wang, J. Flueckiger, Y. Wu, Y. Wang, and S. T. Fard, "Impact of fabrication non-uniformity on chip-scale silicon photonic integrated circuits," in *Optical Fiber Communication Conference*, pp. Th2A–37, Optical Society of America, 2014.

[54] K. O. Abrokwah, P. R. Chidambaram, and D. S. Boning, "Pattern based prediction for plasma etch," *IEEE Trans. Semicond. Manuf.*, vol. 20, no. 2, pp. 77–86, 2007.

[55] Y. Xing, J. Dong, U. Khan, and W. Bogaerts, "Correlation between pattern density and linewidth variation in silicon photonics waveguides," *Optics Express*, vol. 28, pp. 7961–7968, Mar 2020.

[56] G. T. Reed and C. J. Png, "Silicon optical modulators," *Materials Today*, vol. 8, no. 1, pp. 40–50, 2005.

[57] G. T. Reed, G. Mashanovich, F. Y. Gardes, and D. Thomson, "Silicon optical modulators," *Nature Photonics*, vol. 4, no. 8, pp. 518–526, 2010.

[58] X. Xiao, H. Xu, X. Li, Y. Hu, K. Xiong, Z. Li, T. Chu, Y. Yu, and J. Yu, "25 Gbit/s silicon microring modulator based on misalignment-tolerant interleaved PN junctions," *Optics Express*, vol. 20, no. 3, pp. 2507–2515, 2012.

[59] R. A. Soref and B. R. Bennett, "Kramers-Kronig analysis of electro-optical switching in silicon," in *Integrated Optical Circuit Engineering IV*, vol. 704, pp. 32–37, International Society for Optics and Photonics, 1987.

[60] M. Nedeljkovic, R. Soref, and G. Z. Mashanovich, "Free-carrier electrorefraction and electroabsorption modulation predictions for silicon over the 1–14-μm infrared wavelength range," *IEEE Photonics Journal*, vol. 3, no. 6, pp. 1171–1180, 2011.

[61] G. Li, A. V. Krishnamoorthy, I. Shubin, J. Yao, Y. Luo, H. Thacker, X. Zheng, K. Raj, and J. E. Cunningham, "Ring resonator modulators in silicon for interchip photonic links," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 19, no. 6, pp. 95–113, 2013.

[62] W. Bogaerts, P. De Heyn, T. Van Vaerenbergh, K. De Vos, S. Kumar Selvaraja, T. Claes, P. Dumon, P. Bienstman, D. Van Thourhout, and R. Baets, "Silicon microring resonators," *Laser & Photonics Reviews*, vol. 6, no. 1, pp. 47–73, 2012.

[63] T. Baehr-Jones, R. Ding, Y. Liu, A. Ayazi, T. Pinguet, N. C. Harris, M. Streshinsky, P. Lee, Y. Zhang, A. E.-J. Lim, T.-Y. Liow, S. H.-G. Teo, G.-Q. Lo, and H. Michael, "Ultralow drive voltage silicon traveling-wave modulator," *Optics Express*, vol. 20, no. 11, pp. 12014–12020, 2012.

[64] G. Breed, "Analyzing signals using the eye diagram," *High Frequency Electronics*, vol. 4, no. 11, pp. 50–53, 2005.

[65] S. A. Clark, B. Culshaw, E. J. Dawnay, and I. E. Day, "Thermo-optic phase modulators in SIMOX material," in *Integrated Optics Devices IV*, vol. 3936, pp. 16–24, International Society for Optics and Photonics, 2000.

[66] J. Komma, C. Schwarz, G. Hofmann, D. Heinert, and R. Nawrodt, "Thermo-optic coefficient of silicon at 1550 nm and cryogenic temperatures," *Applied Physics Letters*, vol. 101, no. 4, p. 041905, 2012.