Behavioural Models, Parameter Extraction and Yield Prediction for Silicon Photonic Circuits

Gedragsmodellen, parameterextractie en voorspelling van opbrengst voor circuits in siliciumfotonica

Yufei Xing

Promotoren: prof. dr. ir. W. Bogaerts, prof. dr. ir. T. Dhaene
Proefschrift ingediend tot het behalen van de graad van
Doctor in de ingenieurswetenschappen: fotonica

UNIVERSITEIT
GENT

Vakgroep Informatietechnologie
Voorzitter: prof. dr. ir. B. Dhoedt
Faculteit Ingenieurswetenschappen en Architectuur
Academiejaar 2019 - 2020

ISBN 978-94-6355-311-7
NUR 950
Wettelijk depot: D/2019/10.500/119

Universiteit Gent
Faculteit Ingenieurswetenschappen
en Architectuur
Vakgroep Informatietechnologie

Promotoren:

Prof. dr. ir. Wim Bogaerts	Universiteit Gent
Prof. dr. ir. Tom Dhaene	Universiteit Gent

Examination Board:

Prof. dr. ir. Daniel de Zutter (Chair)	Universiteit Gent
Dr. ir. Pieter Dumon	Luceda Photonics
Dr. Xaveer J. M. Leijtens	Technische Universiteit Eindhoven, Nederland
Dr. ir. Domenico Spina	Universiteit Gent
Prof. dr. ir. Bart Kuyken	Universiteit Gent
Prof. dr. ir. Jeroen Beeckman	Universiteit Gent
Dr. ir. Umar Khan	Universiteit Gent
Dr. ir. Peter De Heyn	imec

Universiteit Gent
Faculteit Ingenieurswetenschappen en Architectuur

Vakgroep Informatietechnologie
Technologiepark-Zwijnaarde 126 iGent, B-9052 Gent, België

Proefschrift tot het bekomen van de graad van
Doctor in de ingenieurswetenschappen:
fotonica
Academiejaar 2019-2020

Acknowledgements

Five years is not short for a Ph.D., but I still feel so unwilling to finish it because these five years have been so memorable, warm, and enjoyable. I wish to thank all the people who kindly assisted me in achieving this milestone in my life.

I wish to express sincere gratefulness to my promoter, Professor Wim Bogaerts. You are such a great person that I always hope that I could be one-tenth as good as you are. Your knowledge, skillset, and philosophy are a fortune. Day in and day out, you inspire me with your passion and hard work. More than my academic mentor, you showed me how to manage a group with humor, how to face adversity with optimism and how to treat your students with empathy. Thank you for trusting me when the road is tough. Your trust gives me the confidence to face bigger challenges ahead.

I also want to thank my co-promoter, Professor Tom Dhaene. Thank you for offering kind help and sincere guidance whenever I needed it. The tools and skills of your group have greatly facilitated my research. I would like to recognize valuable assistance and input from Domenico. Without your knowledge, our collaboration on variability analysis would be impossible.

I think our design team has the funniest colleagues ever. I love how everyday Alfonso and Sarvagya exchanged intimate greetings. I also love to annoy Antonio every day who verbally refused but seemed to enjoy it somehow. Yinghao and I went through similar ups and downs in our Ph.D., I am so glad we both made it. Umar, your words and your story cheered me up when I was in 'trouble'. Lukas, thank you for translating the Dutch summary. Jiaxing, I am so proud to be your supervisor. Your attitude and hard work does not only lead to many good results but also inspire me to become a responsible person. Also, I would like to thank my Chinese teammates Ang, Mi, Hong, and Xiangfeng. Without your office talk, I would have graduated one year earlier :)

I would like to thank Michael for your patient help in automatic measurements for countless times. I would like to thank Martin and Joris from Luceda who listened to our feedback seriously and made IPKISS much easier to use. I would like to thank Ivo from the SUMO group for your amazing optimization toolbox and patient guidance. I would like to thank all my colleagues. Throughout the years, you all made Photonics Research Group such an amazing place to work.

My Chinese colleagues, you are the biggest reason why I don't want to leave.

To me, you are not only colleagues, but you are also my best friends. Thank all my senior brothers and sisters Zhechao, Yanlu, Yingtao, Weiqiang, Yunpeng, Chen, Bin, Ruijun, Jing, and Xin for welcoming me to the big family. Thank Qiangsheng, Yingchen for the good times together. Thank Haolan and Zheng for sharing growing pains with me. Thank youngster Chonghuai, Xiaomin, Xiaoning, Yuting, Xiangfeng, Yang, Zuyang, Chupao, Mi and Hong for making me feel young again. Your friendship is a treasure.

I would whole-heartedly thank my parents. Thank you for supporting me especially for buying me an Alfa Romeo. It brought me many speeding tickets but it also brought me my wife into my life from Zaventem.

Five years ago, I told myself finding the soulmate is the top priority of my Ph.D. Now, you are the top priority of my life. Love you, Mengyi!

November 2019, Gent

Yufei

Table of Contents

List of Figures

List of Tables

List of Acronyms

A

AWG	arrayed waveguide grating

C

CMA	covariance matrix adaptation
CMP	chemical mechanical polishing

D

DC	directional coupler

E

ES	evolution strategy

F

fab	semiconductor fabrication plant

G

GC	grating coupler

M

MZI	Mach-Zehnder interferometer

I

IC	integrated circuit
IDR	intra-die random
IDS	intra-die systematic
IWR	intra-wafer random
IWS	intra-wafer systematic

P

PCE	Polynomial Chaos Expansion
PCM	Process Control Monitoring
PD	pattern density
IC	photonics integrated circuit

R

RIE	reactive ion etching

S

SC	Stochastic Collocation
SD	standard deviation
SEM	Scanning electron microscopy

W

WDI	wafer-die interaction

Nederlandse samenvatting –Summary in Dutch–

Procesvariaties in gentegreerde fotonica-circuits

Silicon photonics is een aantrekkelijke technologie om kosten efficiënt fotonica op chip te integreren door zijn compatibiliteit met bestaande CMOS fabricage technologie. Co-integratie van elektronische en fotonische chips is eenvoudiger net door deze compatibiliteit. Het hoge index contrast tussen silicium en silicium dioxide zorgt voor sterke geleiding van het licht, waardoor zeer kleine componenten op grote schaal mogelijk zijn.

Het is echter zo dat dit hoge contrast en de kleine geometrische elementen ervoor zorgen dat deze fotonische circuits in silicium zeer gevoelig zijn aan variaties waar het fabricageproces aan onderworpen is. Bloodstellingsdosis, leeftijd van gebruikte resist, plasma dichtheid en de samenstelling van de slurry bij chemical mechanical polishing, zijn voorbeelden van parameters die bij de fabricage kunnen variëren die uiteindelijk leiden tot geometrische variaties op nanometer schaal zoals lijndikte, laagdikte en hoek van de zijwanden van waveguides. Deze variaties op componenten niveau vertalen zich op circuit niveau tot verschillende vertragingen tussen optische paden in een circuit, de optische signalen zijn uit balans. Dit zorgt voor een slechtere functionaliteit van het betreffende circuit, waarbij soms slechts een deel van een optisch circuit functioneel is. Hoe complexer de circuits, hoe groter de kans dat delen van het circuit niet werken.

Filters, zoals Arrayed Waveguide Gratings en ring resonatoren, zijn bijzonder gevoelig aan deze procesvariaties. Dit zorgt voor meer overspraak tussen kanalen, grotere verliezen, meer verbruik en een afwijking van de golflengte van een kanaal. Wanneer de impact van deze procesvariaties niet in acht genomen wordt zullen bepaalde componenten niet meer werken, zelfs met een variabel element die de fout tracht te compenseren. [1]. Variaties in het proces zijn dus een limiterende factor voor de opbrengst (yield), de fractie van werkende circuits gefabriceerd in dit proces. Zonder een goede analyse en praktische methode om de impact van proces variatie te voorspellen en te matigen, zal procesvariatie bij fabricage leiden tot een grotere kost en de mogelijkheden tot integratie beperken.

Uitdagingen voor het voorspellen van de opbrengst van Photonic Integrated Circuits fabricage

Om de opbrengst te voorspellen en te schatten hoe de degradatie van de prestaties van het geproduceerde circuit kan worden gecompenseerd, wordt het essentieel om voorspelling van de variatie in gedrag van het circuit in de gestandaardiseerde ontwerp workflow te integreren. Er zijn twee uitdagingen om een realistische opbrengstvoorspelling te maken voor fotonische gentegreerde schakelingen.

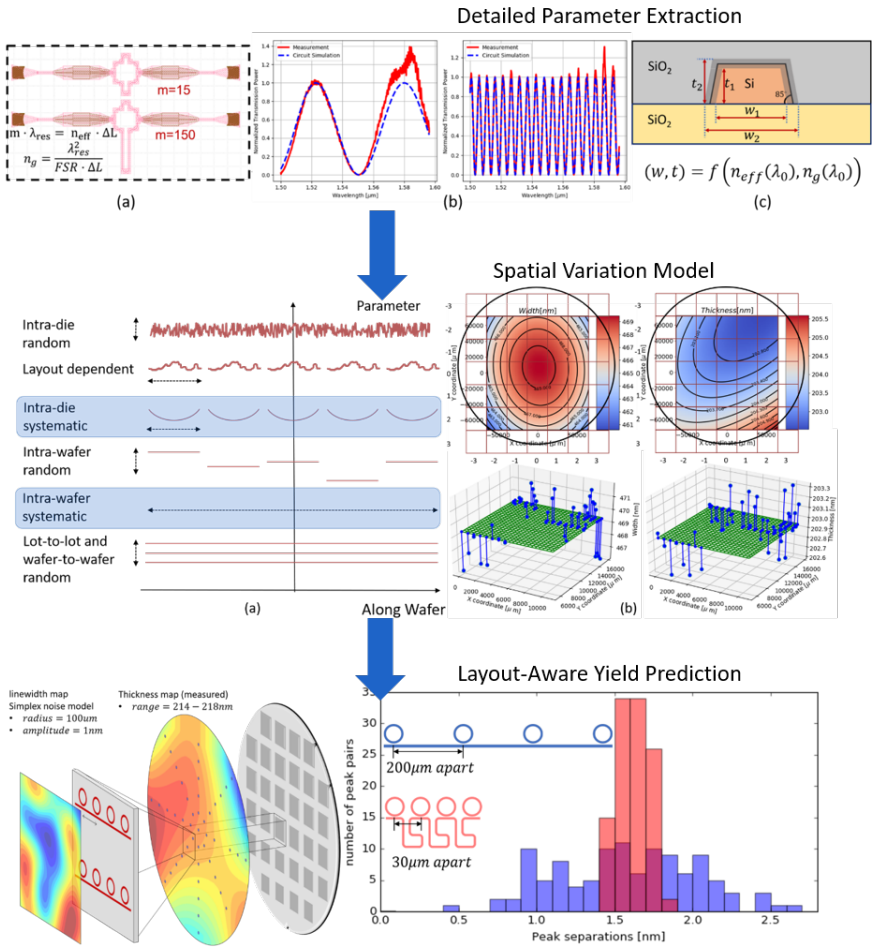
Ten eerste is het, anders dan micro-elektronica, minder intutief om "goedöf "slecht"te definiëren voor fotonische bouwstenen met behulp van één parameter. De effectieve index en koppelingscoëfficiënten hebben bijvoorbeeld beide invloed op de prestaties van de filter, maar ze hebben geen intrinsieke goede of slechte waarden. De voorspelling zou variaties in het gedrag van een circuit moeten berekenen die genduceerd worden door deze parametervariaties, aangezien uiteindelijk de mismatch tussen de respons van het ontworpen circuit en het gefabriceerde circuit de prestaties verslechtert.

Ten tweede zijn procesvariaties niet puur willekeurig. Omdat procesvariaties sterk locatieafhankelijk zijn, wordt de prestatie van het gefabriceerde circuit primair bepaald door de lay-out van het circuit en zijn positie op de wafer. Een realistische opbrengstvoorspelling vereist het kwantitatief vastleggen van de willekeurige en locatieafhankelijke variaties.

Bijdrage van dit werk aan het onderzoek

Om de opbrengst te voorspellen die rekening houdt met de lay-out, hebben we een techniek voorgesteld die variaties van gedragsparameters op laag niveau (effectieve index, koppelingscoëfficiënt) of gefabriceerde geometrische parameters (lijnbreedte, dikte) vertaald naar circuitprestaties op een hogere abstractie niveau. Door locatie-afhankelijke variaties toe te wijzen die de werkelijke variatie in fabricage nabootsen, kunnen we de circuitvariaties op hoog niveau inschatten met behulp van Monte-Carlo-simulaties. Deze techniek vereist drie stukjes informatie: een nauwkeurige verzameling van gegevens op de gefabriceerde wafer, een realistisch variatiemodel om systematische en willekeurige variaties op verschillende ruimtelijke niveaus te scheiden, en een virtuele wafer-kaart op basis van geëxtraheerde gegevens van gefabriceerde componenten om Monte-Carlo simulaties mogelijk te maken.

Het verzamelen van gegevens over gefabriceerde wafers biedt input voor variabiliteitsanalyse. We willen geometrische parameters van gefabriceerde golfgeleiders afleiden, omdat dit de meest fundamentele parameters van de lay-out zijn om de proceskwaliteit weer te geven. Het is echter erg duur om ze te verkrijgen met behulp van traditionele meettechnieken. We stelden een methode voor om silicium golfgeleider geometrie te extraheren met behulp van spectrale metingen van een paar MZI-circuits. We hebben een uitgebreide analyse uitgevoerd om de grenzen van de parameter te verkleinen, waardoor de nauwkeurigheid van de extractie



Figuur 1: Componenten voor realistische variabiliteitsanalyse en opbrengstvoorspelling.

aanzienlijk wordt verbeterd. Met behulp van de curve-fitting zijn we er ook in geslaagd om de extractiemethoden toleranter te maken voor ruis en variaties in de roosterkoppeling. We hebben een stapsgewijze workflow voorgesteld waarmee we lijnbreedte en dikte kunnen extraheren met sub-nanometer-nauwkeurigheid, waardoor het mogelijk is om gedetailleerde gefabriceerde wafer kaarten te extraheren voor verdere analyse. Het extraheren van de breedte en -dikte van een golfgeleider helpt bij het analyseren van de procesvariaties op een siliciumfotonica-chip. Vaak moeten we ook de kwaliteit van gefabriceerde directionele koppelaars kennen om de procesvariatie van koppelingen te analyseren en om de impact van variaties op de prestaties van de optische filters te schatten, waarvoor een compact gedragsmodel voor de directionele koppelaars vereist is. We hebben een compact model

van de directionele koppelaar, dat helpt bij het scheiden van componentontwerp van circuitontwerp. Het extraheren van koppelingscoëfficiënten is niet triviaal in de aanwezigheid van meetruis, uitlijningsfout en variatie van de roosterkoppeling. We hebben verschillende ontwerpen en methoden vergeleken en bewezen dat we, met behulp van curve-fitting en een goed rooster-koppelingsmodel, parameters van de directionele koppelaar met hoge nauwkeurigheid uit een MZI-circuit kunnen extraheren.

Om gelijktijdig meerdere parameters van on-chip golfgeleiders en directionele koppelaars uit optische metingen te extraheren, stelden we een gevouwen tweetraps MZI-circuit voor. Het compacte ontwerp is minder blootgesteld aan lokale variatie binnen het circuit, wat de nauwkeurigheid van de extractie aanzienlijk verbetert. Ook verkort het circuit de duur van optische metingen op waferschaal aanzienlijk, waardoor het bruikbaar is voor bewaking van procesbesturing. We hebben de heropstart van het wereldwijde optimalisatie-algoritme, CMA-ES, gebruikt om parameters te extraheren uit ingewikkelde spectra van het circuit. We hebben de workflow ook gellustreerd om gedetailleerde waferkaarten te plotten voor variabiliteitsanalyse van wafermetingen.

Om geëxtraheerde waferkaarten met informatie over de procesvariatie te analyseren, stelden we een additief hiërarchisch model voor dat variabiliteit op verschillende ruimtelijke niveaus kan omzetten in systematische en willekeurige variatie. Om willekeurige variatie in onze analyse te verminderen hebben we de fysieke oorsprong van deze procesvariaties besproken en een workflow samengesteld om de variabiliteit van intra-wafer te scheiden van het intra-die-niveau en een wafer-die interactie term gentroduceerd. We hebben het model toegepast om de metingen te analyseren op een 200 mm wafer gefabriceerd in imec's silicium fotonica-platform op basis van 193 nm lithografie. Het resultaat toont aan dat de systematische variatie de primaire bron van variatie is voor zowel lijnbreedte als dikte. Op die-niveau vonden we dat het systematische breedte-patroon nauw verwant is met de lokale patroondichtheid. Onze bevindingen helpen om de procesvariatie te identificeren en nieuwe ontwerpregels te creëren om de impact van de niet-uniformiteit te verminderen. Als de systematische variatie die we hebben waargenomen in het proces verder kan worden verfijnd en gecompenseerd, kunnen we de fotonische wafers voorzien met een aanzienlijk betere fabricage-uniformiteit in de toekomst.

In de volgende stap hebben we het CapheVE-raamwerk ontwikkeld waarmee we het ruimtelijke variabiliteitsmodel kunnen importeren en virtuele fabricagewafers kunnen genereren. Het raamwerk combineert het circuitmodel, parametergevoeligheid, circuit lay-out en procesvariabiliteitsmodel. Circuitparameters worden gewijzigd op basis van de gevoeligheid, de lay-out en de locatie op de wafer. Met CapheVE kunnen we de realisaties van het circuit over de wafer verspreiden en een Monte Carlo-simulatie uitvoeren om de respons van alle circuits te genereren.

Vervolgens kunnen we de gegenereerde antwoorden gebruiken voor lay-out bewuste variabiliteitsanalyse en opbrengstvoorspelling. We hebben interessante experimenten gedaan met het CapheVE- raamwerk. We genereerden virtuele wafers die een goede overeenkomst vertonen in statistische eigenschappen en ruimte-

lijke correlatie van de procesvariaties met de gefabriceerde wafer. We hebben het raamwerk ook gebruikt om opbrengstvoorspellingen te doen voor golflengte-demultiplexer en om de workflow voor parameter-extractie te valideren met behulp van de compacte tweetraps MZI.

We hebben ook stochastische analysemethoden geprobeerd om de kost van Monte Carlo-methoden voor opbrengstvoorspelling te verlagen. In vergelijking met de Monte Carlo-methode kan de stochastische analysemethode de simulatiekost aanzienlijk verlagen. We hebben de stochastische collocatiemethode toegepast om de variabiliteit van directionele koppelaars te analyseren. De methode heeft de kost van simulatietijd aanzienlijk verlaagd.

Conclusie en toekomst

In dit werk hebben we een compleet raamwerk voorgesteld om realistische opbrengstvoorspellingen te doen voor gentegreerde fotonica-schakelingen. Onze methoden om gefabriceerde geometrieën of golfgeleider en koppelingsgedragsparameters te extraheren helpen om gedetailleerde fabricage waferkaarten te verkrijgen met hoge nauwkeurigheid van spectrale metingen. Het variabiliteitsmodel dat we hebben opgebouwd, splitst procesvariaties op in systematische en willekeurige bijdragen in verschillende ruimtelijke niveaus. Dit model helpt de procesvariatie te identificeren en nieuwe ontwerpregels te creëren om de impact van de niet-uniformiteit te verminderen. We hebben het model ook gentegreerd in het layout bewuste opbrengstvoorspelling raamwerk CapheVE om virtuele proceskaarten te genereren die de kenmerken van geëxtraheerde waferkaarten vastleggen. Met dit raamwerk kunnen we circuitprestaties berekenen op basis van hun locaties op de wafer, waarmee we circuitopbrengsten kunnen afleiden met behulp van Monte Carlo-simulaties.

We hebben meer meetwaarden van de wafer nodig om het variabiliteitsmodel verder te valideren. Dit zou ook de waarde van wafer-to-wafer en lot-to-lot variatie in het model kunnen karakteriseren. We hebben de CMZI-structuur opgenomen in een fabricage proces waarbij we parameters over meerdere wafers en volledige dies zouden moeten kunnen extraheren. Met deze informatie kunnen we de modellering van de variatie in het niveau verbeteren.

Hoewel de relevantie van dit onderzoek voor opbrengstvoorspelling vrij duidelijk is, kan het alleen een praktische realiteit worden als het daadwerkelijk wordt gentroduceerd in de ontwerpstroom die wordt gebruikt door een aanzienlijk deel van de PIC-gemeenschap. Om dit te bewerkstelligen, moeten we de technieken voldoende robuust maken en integreren in de tools die door echte ontwerpers worden gebruikt. We werkten samen met Luceda Photonics (in het kader van het VLAIO-project MEPIC) om ervoor te zorgen dat onze technieken werkelijke ontwerpproblemen konden oplossen. Als resultaat van dit project zullen enkele van de technieken (bijvoorbeeld het CapheVE-framework) worden opgenomen in toekomstige releases van Luceda's IPKISS-framework.

Referenties

[1] Wim Bogaerts, Yufei Xing, and Umar Khan. *Layout-Aware Variability Analysis, Yield Prediction, and Optimization in Photonic Integrated Circuits*. IEEE Journal of Selected Topics in Quantum Electronics, 25(5):1–13, 9 2019.

English Summary

Process Variations in Photonics Integrated Circuits

Silicon photonics is a very attractive solution for low-cost, high-volume photonic integration for its compatibility with existing CMOS manufacturing technology. Also, tight integration with electronics is more feasible with silicon technology than with other photonics integrated circuits technologies. Its high contrast in the refractive index between silicon and silicon dioxide allows for strong light confinement, which facilitates small device footprint and large-scale integration.

However, the high material contrast and small feature size also make silicon photonics circuits very sensitive to process variations. Process variation such as exposure dose, resist age, plasma density, and chemical mechanical polishing slurry composition can lead to nanometer-scale variations in component geometries such as linewidth, layer thickness, sidewall angles. Variations in components geometries changes the optical properties of a device such as the effective index and the group index of a waveguide, the coupling ratio of a coupler and the center wavelength of interferometric structures. These variations at device level propagate and accumulate at circuit level such that optical delay has a random component and path imbalance is induced in the circuit, which deteriorates the circuit performance, making only a fraction of the fabricated circuits perform as intended. This fraction further shrinks with increasing circuit size and complexity. In particular, wavelength filters such as ring resonators, lattice filters, array waveguide gratings, and so on suffer significantly from process fabrication leading to increasing channel cross-talk, insertion loss, power consumption and deviation of channel wavelength. Neglecting the impact of process variations could lead to failure of the entire circuits, even with active tuning to compensate the error [1]. Process variation is a limiting factor to fabrication yield, i.e., the fraction of functional fabricated circuits. Without proper analysis and a practical method to predict and mitigate its impact, process variation will increase the cost of large-volume production and limit the scale of the integration.

Challenges in Yield Prediction for Photonics Integrated Circuits

To predict yield and estimate how to compensate for the degradation of fabricated circuit performance, it becomes essential to integrate prediction of the circuit performance variation in the standardized design workflow. There are two challenges to make a realistic yield prediction for photonic integrated circuits.

First, unlike microelectronics, it is less intuitive to define "good" or "bad" for photonics building blocks using one parameter. For example, the effective index and coupling coefficients both affect the performance of the wavelength filter, but they have no intrinsic good or bad values. The prediction should calculate the circuit performance induced by these parameter variations since ultimately it is the mismatch between the response of designed circuit and fabricated circuit that degrade the performance.

Second, process variations are not purely random. As process variations are highly location-dependent, the performance of the fabricated circuit is primarily determined by the layout of the circuit and its position on the wafer. A realistic yield prediction requires to capture the random and the location-dependent variations quantitatively.

Contribution of This Work

To make the layout-aware yield prediction, we proposed a technique that maps variations of low-level behavioral parameters (effective index, coupling coefficient) or fabricated geometry parameters (linewidth, thickness) to high-level circuit performance variations. By assigning location-aware variations that mimic the actual fabrication variation, we can derive the high-level circuit variations using Monte-Carlo simulations. This technique requires three pieces of information: an accurate collection of data on the fabricated wafer, a realistic variation model to separate systematic and random variations on different spatial levels, and a virtual fabrication wafer map based on extracted fabricated data to enable Monte-Carlo simulations.

Collecting data on fabricated wafers offers input for variability analysis. We want to derive geometry parameters of fabricated waveguide because they are the most fundamental layout parameters to reflect the process quality. However, it is very costly to obtain them using traditional measurement techniques. We proposed a method to extract silicon waveguide geometry using spectral measurements of a pair of MZI circuits. We carried out a comprehensive analysis to reduce the parameter bounds, which improves extraction accuracy significantly. Using the curve fitting method, we also managed to make the extraction methods more tolerant of measurement noise and grating coupler variations. We proposed a step-wise workflow that allows us to extract linewidth and thickness with sub-nanometer accuracy, which makes it possible to extract detailed fabrication wafer maps for further analysis.

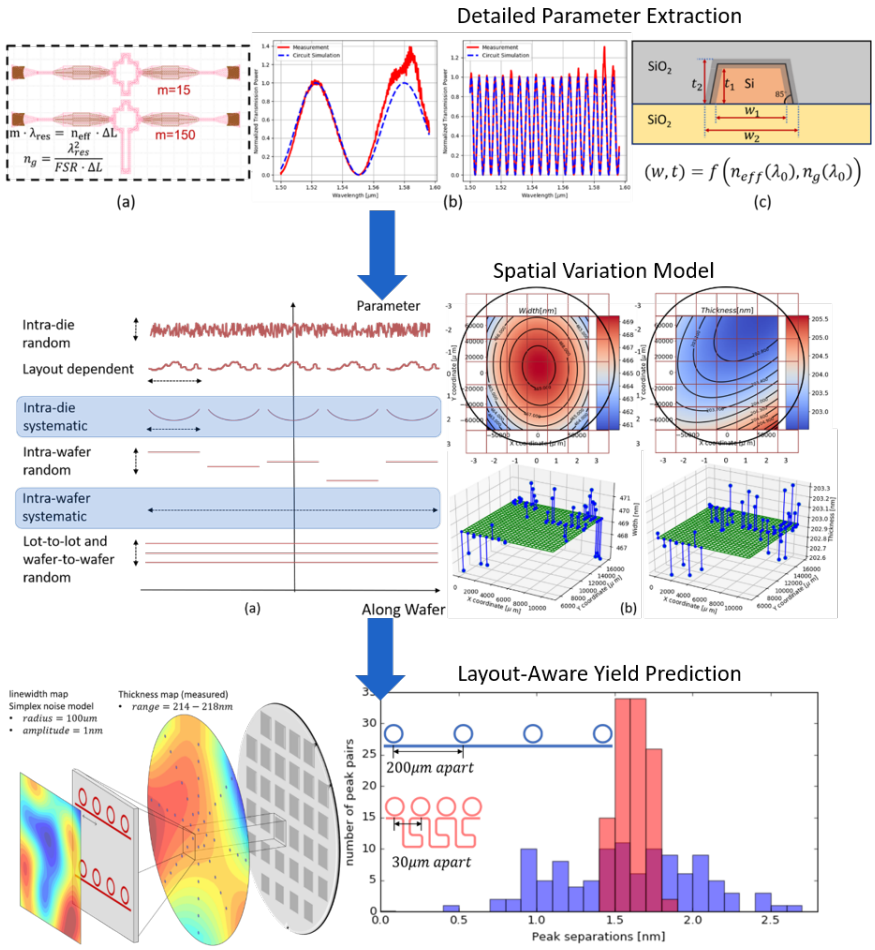


Figure 2: Components for realistic variability analysis and yield prediction.

Extracting waveguide width and thickness helps to analyze the process variations on a silicon photonics chip. Often, we also need to know the fabricated quality of directional couplers to analyze the process variation of couplers and to estimate the impact of variations on the performance of the optical filters, which requires a compact behavioral model for the directional coupler. We constructed and validated a compact model of the directional coupler that helps to separate component design from circuit design. Extracting coupling coefficients is not trivial in the presence of measurement noise, alignment error, and grating coupler variation. We compared different designs and methods and proved that using curve fitting and good grating coupler model, we can extract parameters of the directional coupler with high accuracy from an MZI circuit.

To simultaneously extract multiple parameters of on-chip waveguides and directional couplers from optical measurements, we proposed a folded two-stage MZI circuit. The compact design suffers less from local variation within the circuit, which significantly improves the accuracy of extraction. Also, the circuit greatly reduces the duration of wafer-scale optical measurements, making it useful for process control monitoring. We used the restart CMA-ES global optimization algorithm to extract parameters for complicated spectra of the circuit. We also illustrated the workflow to plot detailed wafer maps for variability analysis from wafer measurements.

To analyze extracted wafer maps that contain rich information about the process variation, we proposed an additive hierarchical model that can decompose variability on various spatial levels into systematic and random variation. We discussed the physical origins of these process variations and constructed a workflow to separate variability on intra-wafer, intra-die level, and introduced wafer-die interaction term to reduce random variation in our analysis. We applied the model to analyze the measurements on a 200 mm wafer fabricated in IMEC's silicon photonics platform based on 193 nm lithography. The result shows that the systematic variation is the primary source of variation for both linewidth and thickness. At the die level, we found that systematic width pattern is closely related to the local pattern density. Our findings help to identify the process variation and create new design rules to alleviate the impact of the non-uniformity. If the systematic variation we observed can be further refined and compensated in the process, we can foresee the photonics wafers with significantly better fabrication uniformity in the future.

In the next step, we developed the CapheVE framework that allows us to import the spatial variability model and generate virtual fabrication wafers. The framework combines the circuit model, parameter sensitivity, circuit layout, and process variability model. Circuit parameters are altered according to its sensitivity and layout and location on the wafer. Using CapheVE, we can place the instances of the circuit over the wafer, and run Monte Carlo simulation to generate the response for all circuits. We can then use the generated responses for layout-aware variability analysis and yield prediction. We did interesting experiments with the CapheVE framework. We generated virtual wafers that exhibit a good match in statistical properties and spatial correlation of the process variations with the fabricated wafer. We also used the framework to make yield prediction for wavelength de-multiplexer and to validate the parameter extraction workflow using the compact two-stage MZI.

We also tried stochastic analysis methods to reduce the cost of Monte Carlo methods for yield prediction. Compared to the Monte Carlo method, the stochastic analysis method could significantly reduce the simulation cost. We applied the stochastic collocation method to analyze the variability of DC. The method significantly reduced the cost of simulation time.

Conclusions and Perspectives

In this work, we proposed a complete framework to make realistic yield prediction for photonics integrated circuits. Our methods to extract fabricated geometries or waveguide and coupler behavioral parameters help to obtain detailed fabrication wafer maps with high accuracy from spectral measurements. The variability model we constructed decomposes process variations into systematic and random contributions on different spatial levels. This model helps to identify the process variation and create new design rules to alleviate the impact of the non-uniformity. We also integrated the model into the layout-aware yield prediction framework CapheVE to generate virtual process maps that captures the features of extracted wafer maps. This framework allows us to calculate circuits performances according to their locations on the wafer, with which we can derive circuit yield using Monte Carlo simulations.

We need more wafer measurements to validate the variability model further. Also, this could characterize the value of wafer-to-wafer and lot-to-lot variation in the model. We have included the CMZI structure on a fabrication run where we should be able to extract parameters over multiple wafers and full dies. With these information we might improve the modeling of the die-level variation.

While the relevance of this research on yield prediction is quite clear, it can only become a practical reality if it is actually introduced into the design flow that is used by a significant fraction of the PIC community. To bring this about, we should make the techniques sufficiently robust and integrate them into the tools that are used by actual designers. We worked together with Luceda Photonics (in the framework of the VLAIO project MEPIC), to make sure that our techniques could solve actual design problems. As a result of this project, some of the techniques (e.g., the CapheVE framework) will be incorporated in future releases of Luceda's IPKISS framework.

References

[1] Wim Bogaerts, Yufei Xing, and Umar Khan. *Layout-Aware Variability Analysis, Yield Prediction, and Optimization in Photonic Integrated Circuits*. IEEE Journal of Selected Topics in Quantum Electronics, 25(5):1–13, 9 2019.

1 Introduction

The major goal of this work is to answer a question: When I design a photonic circuit, what is the chance that it will work when it is fabricated? This 'chance' translates in what is called 'yield'. To answer the question, we need to make realistic yield prediction. In this work, we explored and researched on a few topics including compact behavioral models for photonic circuits, methods to make accurate and robust wafer-scale parameter extraction, spatial variability model of photonics, and layout-aware yield prediction framework. In this thesis, we will discuss the research we did on those topics and propose and verify a comprehensive workflow to make layout-aware yield prediction at the design stage.

1.1 Integrated Photonics

As has been largely experimented by microelectronics for a few decades, integration of circuits and functions on the same chip leads to significant reductions in assembly and test cost. Besides, the miniaturization process shrinks the footprint and lowers power consumption. The same benefits can apply to photonics.

Photonics is a technology to generate, modulate, propagate and detect light. Integrated photonics miniaturizes bulky photonic devices to process light on the surface of a chip. A photonic integrated circuit consists of functional building blocks connected by waveguides, which makes it a 'circuit'. These building blocks perform functions such as light generation, electro-optical signal modulation, photo detection, light distribution and wavelength filtering implemented as on-chip in-

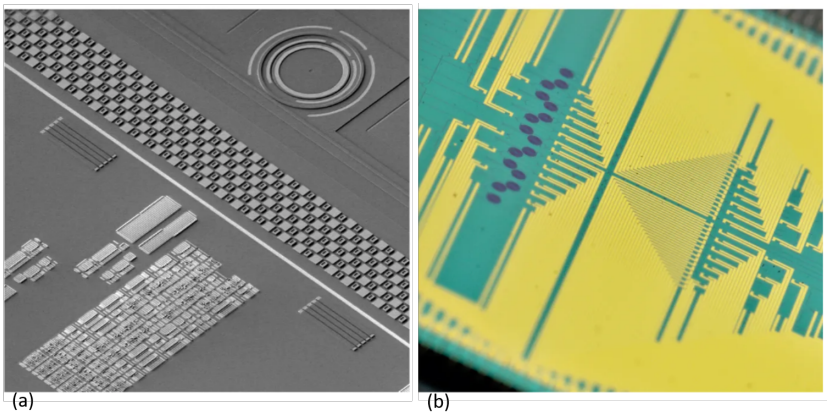


Figure 1.1: (a) Electronicphotonic silicon chip. Figure reproduced from ref. [2]. (b) Quantum silicon chip. Figure reproduced from ref. [3].

terferometers.

In integrated photonics, there co-exist several technologies. The most common of these are Silica-on-silicon, Lithium Niobate, III-V semiconductor materials and Silicon-on-insulator. Silica-on-Silicon uses doped silica as a core where undoped silica is used as a cladding. Silica-on-Silicon waveguide has a low index contrast (2% or lower). It requires a large bend radius to guide light which leads to large device footprint and low integration density. Lithium Niobate has suitable for electro-optical modulation and non-linear optics. But it has a complex process which is very expensive for large scale integration. III-V based technologies such as Indium Phosphide technology allows us to build lasers, optical amplifiers, and photodetectors on alloys III-V since it is a direct bandgap material able to generate and amplify light. It is also possible to integrate electronic elements with technology. Monolithically integrated InP multi-wavelength transceivers on the market can work up to 100Gbit/s [1] already.

Meanwhile, in the past decades silicon photonics has been growing at a breakneck pace precisely because it is basically compatible with the well-developed Silicon process infrastructure that supported very large scale integration on-chip and which led to the advancement of the highly integrated CMOS devices. This gives silicon photonics a significant leverage in reducing the cost, especially when the market needs high volumes of products. Also, tight integration with electronics is also more feasible with silicon technology than with other PIC technologies. [2]. In the foreseeable future, silicon photonics might still need hybrid integration with III-V to add lasers and amplifiers. Nevertheless, all the other photonics building blocks can be realized, such as waveguides, power splitters, de-multiplexers, modulators, detectors (Germanium monolithically integrated), etc.

Silicon photonics also benefits from the high index contrast of silicon, enabling strong light confinement and miniaturized circuits. The miniaturization is not yet the level in comparison with electronics. Still, it is significant. Photonic systems-on-chip have already integrated $>$ 1000 functional elements onto a few square millimeters circuit [4]. This paves the road for applications such as large switch matrices in data center [5], LIDAR [6], quantum computing [3], etc.

1.2 Process Variation in Silicon Photonics

Silicon Photonics is a promising solution for low-cost large-volume production of photonic circuits because of its compatibility with existing CMOS manufacturing technology. Also, it enables large-scale integration by its high contrast in refractive index (between silicon and silicon dioxide), which allows strong light confinement, and thus small footprint. However, the high material contrast and small feature size also make silicon photonic circuits very sensitive to nanometer scale variations in component geometries, which can be induced by process variations. Process variability induces changes in behavior of photonic circuits at different levels (Fig. 1.2). Process variation such as exposure dose, resist age, plasma density, and CMP slurry composition lead to variations in device geometry such as linewidth, layer thickness, sidewall angles, and doping profile variation. Process variation affects optical properties of a device such as effective index and group index of a waveguide, the coupling of a coupler and center wavelength of interferometric structures. The performance variations at device level propagate and accumulate at circuit level so that optical delay has a random component and path imbalance is induced in the circuit, which deteriorates the circuit performance, making only a fraction of the fabricated circuits perform as intended. This fraction further shrinks with the increasing circuit size and complexity. In particular, wavelength filters such as ring resonators, lattice filters, array waveguide gratings and so on suffer significantly from process fabrication leading to increasing channel cross-talk, insertion loss, power consumption and deviation of channel wavelength. Neglecting the impact of process variations could lead to failure even with tuning to compensate the error [7]. Process variation is a limiting factor to fabrication yield, i.e., the fraction of functional fabricated circuits. Without a good analysis and a practical method to predict and mitigate its impact, process variation will increase the cost of large-volume production and limit the scale of the integration.

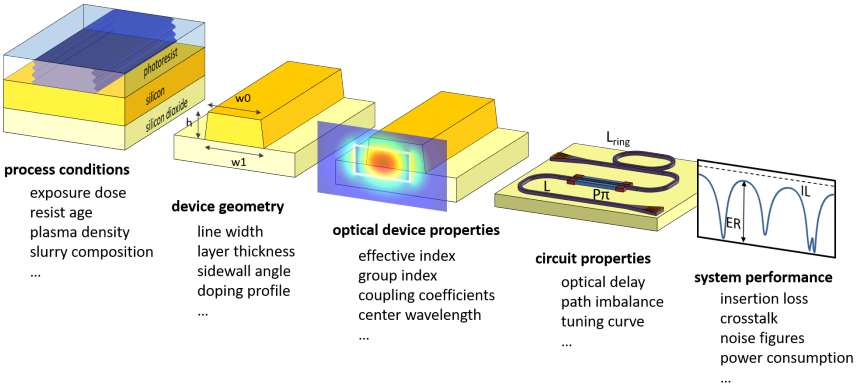


Figure 1.2: Describing variability presents at different levels.

1.3 Yield Prediction of Silicon Photonics

Process variations increase the cost of the delivered product and limit the capacity of a circuit. Predicting the circuit performance variation becomes essential in standardized design workflow to predict yield and estimate how to compensate for degradation of the fabricated chip.

In integrated electronics, the traditional way to assess the performance of circuits under the effects of variability is corner analysis, which calculates best and worst cases under a fabrication variation. The definition of best or worst is often clear (e.g. slow and fast transistors). Better corresponds to lower resistance, faster switching times, etc. However, it is less intuitive to define "good" or "bad" for photonics building blocks using one parameter. For example, the effective index and coupling coefficients both affect the performance of the wavelength filter, but they have no intrinsic good or bad values. It is the deviation from the designed value or the mismatch between values of components in the circuit that degrade the performance. To predict the yield of photonics circuits, we can map variations of low-level behavior parameters (effective index, coupling coefficient) or fabricated geometry parameters (linewidth, thickness) to high-level circuit performance variations. By assigning location-aware variations that mimic the actual fabrication variation, we can derive the high-level circuit variations using Monte-Carlo simulations. This technique requires three pieces of information: an accurate collection of data on the fabricated wafer, a realistic variation model to separate systematic and random variations on different spatial levels, and a virtual fabrication wafer map based on extracted fabricated data to enable Monte-Carlo simulations. This thesis will discuss our progress on these aspects to achieve realistic yield prediction.

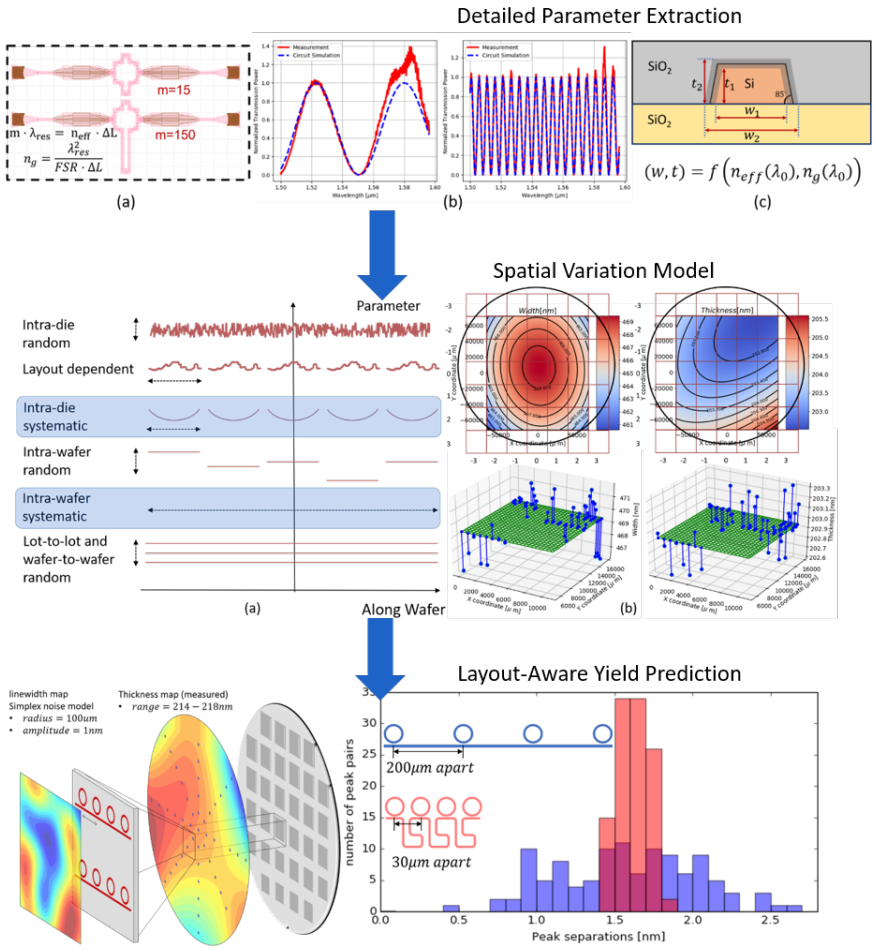


Figure 1.3: Components for realistic variability analysis and yield prediction.

1.4 Outline of the Thesis

1.4.1 Compact Behavioral Model and Parameter Extraction

Extracting compact model parameters of fabricated circuits is essential to get input data for performance evaluation [8] and variability analysis [9]. Extracting fabricated geometry is essential to map fabricated variations to circuit performance variations.

In Chapter 2, we will introduce a method to extract silicon waveguide geometry with sub-nanometer accuracy using spectral measurements. We will discuss how to obtain effective index and group index with high accuracy from a pair of Mach-

Zehnder Interferometers (*MZI*) and how to map them to waveguide width and thickness.

In Chapter 3, we will describe how we built and validated a compact behavioral model of the directional coupler. We will also compare a few circuit designs and methods to extract coupler model parameters. We will discuss how to accurately extract coupler model parameters in the presence of measurement noise and grating coupler process variations.

In Chapter 4, we will present a compact circuit to extract multiple parameters of waveguides and directional couplers simultaneously. The design suffers less from a local variation, which significantly improves the accuracy of extraction. The circuit also greatly reduces the duration of wafer-scale optical measurements, making it very useful for process control monitoring. We will illustrate the global optimization algorithm and the workflow to extract parameters from wafer measurements and plot detailed wafer maps for variability analysis.

1.4.2 Spatial Variability Model

To analyze the statistics of the parameters extracted from fabricated wafer maps and regenerate the statistics in the yield prediction, we require a variability model. In Chapter 5, we proposed a hierarchical model to separate the layout-dependent systematic process variation and random process variation on different spatial levels. Using the model, we decomposed variations of the measured wafer maps and found out quantitatively how each of them contributes to the total process variation. We also observed that die-level systematic linewidth variation is correlated with local pattern density.

1.4.3 Yield Prediction

Being able to estimate the yield at the design stage is crucial to improve the design for better process tolerance. Also, the prediction helps to estimate and reduce the cost of production. In Chapter 6, we will propose our solutions to two major issues to predict the yield for silicon photonics. In the first part, we will discuss methods based on the stochastic analysis to reduce simulation cost for yield prediction. In the second part, we will illustrate the CapheVE framework to make layout-aware yield prediction. We will describe how to generate virtual wafer maps from real fabrication statistics and how to include link process variability to performance variation of circuits in the framework. We will show the example where we used CapheVE to verify the extraction workflow proposed in Chapter 4 and the spatial variability model described in Chapter 5.

1.5 Funding and Collaborations

This work was carried out in the framework of the Flemish Research Foundation (FWO-Vlaanderen) project and Flemish Agency for Innovation and Entrepreneurship (VLAIO) with the MEPIC project. It involved the joint effort from the Photonic Research Group and the SUMO (SUrrogate MOdeling) Lab. It also involved a collaboration with Luceda Photonics.

1.6 Publications

1.6.1 Publications in International Journals

1. Y. Xing, M. Wang, A. Ruocco, J. Geessels, U. Khan, W. Bogaerts, "A Compact Silicon Photonics Circuit to Extract Multiple Parameters for Process Control Monitoring", OSA Continuum, (submitted).

2. M. Wang, A. Ribeiro, Y. Xing, W. Bogaerts, "Tolerant, Broadband Tunable 2×2 Coupler Circuit", Optics Express, (submitted).

3. B. Ouyang, Y. Xing, W. Bogaerts, J. Caro, "Silicon Ring Resonators with a Free Spectral Range Tolerant against Fabrication Variations", Optics Express, (submitted).

4. W. Bogaerts, Y. Xing, U. Khan, "Layout-Aware Variability Analysis, Yield Prediction and Optimization in Silicon Photonic Circuits", IEEE Journal on Selected Topics in Quantum Electronics, 25(5), pp. paper 6100413 (2019).

5. U. Khan, Y. Xing, Y. Ye, W. Bogaerts, "Photonic integrated circuits design in a foundry+fabless ecosystem", Journal of Selected Topics in Quantum Electronics, (2019).

6. Y. Xing, J. Dong, S. Dwivedi, U. Khan, W. Bogaerts, "Accurate Extraction of Fabricated Geometry Using Optical Measurement", Photonics Research, 6 (11), pp. 1008-1020 (2018).

7. Y. Ye, D. Spina, Y. Xing, W. Bogaerts, T. Dhaene, Numerical Modeling of a Linear Photonic System for Accurate and Efficient Time-Domain Simulations, Photonics Research, (2018)

8. Y. Xing, D. Spina, A. Li, T. Dhaene, W. Bogaerts, Stochastic Collocation for Device-level Variability Analysis in Integrated Photonics, Photonics Research, (2016)

9. Y. Xing, T. Ako, J. George, D. Korn, H. Yu, P. Verheyen, M. Pantouvaki, G. Lepage, P. Absil, A. Ruocco, C. Koos, J. Leuthold, K. Neyts, J. Beeckman,

W. Bogaerts, Digitally Controlled Phase Shifter using an SOI Slot Waveguide with Liquid Crystal Infiltration, Photonics Technology Letters, (2015)

1.6.2 Publications in International Conferences

1. U. Khan, Y. Xing, W. Bogaerts, "Parameter extraction, variability analysis and yield prediction of the photonic integrated circuits.", Advanced Photonics Congress, United States, (2019).

2. M. Wang, A. Ribeiro, Y. Xing, W. Bogaerts, "Tolerant, Broadband Tunable 2x2 Coupler Circuit", IEEE Photonics Society Summer Topicals, United States, (2019).

3. W. Bogaerts, Y. Xing, Y. Ye, U. Khan, J. Dong, J. Geessels, M. Fiers, D. Spina, T. Dhaene, "Predicting Yield of Photonic Circuits With Wafer-Scale Fabrication Variability", 2019 IEEE MTT-S International Conference on Numerical Electromagnetic and Multiphysics Modeling and Optimization., United States, (2019).

4. Y. Xing, M. Wang, A. Ruocco, J. Geessels, U. Khan, W. Bogaerts, "Extracting Multiple Parameters from a Compact Circuit for Performance Evaluation", European Conference on Integrated Optics (ECIO 2019), Belgium, pp. W.Po1.9 (2019).

5. U. Khan, Y. Xing, A. Ribeiro, W. Bogaerts, "Extracting Coupling Coefficients of Directional Couplers", European Conference on Integrated Optics (ECIO 2019), Belgium, pp. W.Po.1.8 (2019).

6. U. Khan, Y. Xing, W. Bogaerts, "Designing large-scale photonic integrated circuits", OPTICS 2019, Italy, (2019).

7. U. Khan, Y. Xing, W. Bogaerts, "Designing Large-Scale Photonic Integrated Circuits", OPTICS 2019, Italy, (Invited).

8. U. Khan, Y. Xing, W. Bogaerts, "Effect of Fabrication Imperfections on the Performance of Silicon-on-Insulator Arrayed Waveguide Gratings", Proceedings of the 23rd Annual Symposium of the IEEE Photonics Benelux Chapter, Belgium, (2018).

9. Y. Xing, J. Dong, U. Khan, Y. Ye, D. Spina, T. Dhaene, W. Bogaerts, "From Parameter Extraction, Variability Models to Yield Prediction", Latin America Optics & Photonics Conference, Peru, pp. paper W3E.1 (3 pages) (invited 2018).

10. Y. Ye, D. Spina, Y. Xing, W. Bogaerts, T. Dhaene, “Fast and Accurate Time-Domain Simulation of Passive Photonic Systems”, IEEE International Conference on Electromagnetics in Advanced Applications, Colombia, pp. 396-399 (2018).

11. Y. Xing, J. Dong, U. Khan, W. Bogaerts, “Hierarchical Model for Spatial Variations of Integrated Photonics”, IEEE International Conference on Group IV Photonics, Mexico, pp. 91-92 (WD4) (2018).

12. W. Bogaerts, U. Khan, Y. Xing, “Layout-Aware Yield Prediction of Photonic Circuits”, IEEE International Conference on Group IV Photonics, Mexico, pp. p93-93 (WD5) (2018).

13. Y. Xing, M. Khan, Y. Ye, W. Bogaerts, Extracting Fabricated Geometry on Die-Level, 2017 IEEE Photonics Scociety Benelux Annual Symposium, Netherlands, p.148-151 (2017)

14. Y. Xing, M. Khan, A. Ribeiro, W. Bogaerts, Behavior Model for Directional Coupler, 2017 IEEE Photonics Scociety Benelux Annual Symposium, Netherlands, (2017)

15. A. Li, Y. Xing, W. Bogaerts, An Integrated Tunable Reflector, European Conference on Integrated Optics (ECIO, Netherlands, p.paper W3.1 (2017)

16. Y. Xing, A. Li, R. Van Laer, R. Baets, W. Bogaerts, Backscattering Induced Transmission Noise and Length Dependent Attenuation in Silicon Waveguides, 2016 IEEE Photonics Society Benelux Annual Symposium, Belgium, (2016)

17. A. Li, Y. Xing, R. Van Laer, R. Baets, W. Bogaerts, Extreme Spectral Transmission Fluctuations in Silicon Nanowires Induced by Backscattering, IEEE International Conference on Group IV Photonics 2016, China, p.paper FB4 (2 pages) (2016)

18. Y. Xing, A. Li, R. Van Laer, R. Baets, W. Bogaerts, Backscatter Model for Nanoscale Silicon Waveguides , 24th International Workshop on Optical Wave & Waveguide Theory and Numerical Modelling (OWTNM 2016), Poland, p.paper OWTNM/O-18 (2016)

19. Y. Xing, D. Spina, A. Li, T. Dhaene, W. Bogaerts, Variability analysis of device-level photonics using stochastic collocation , SPIE Photonics Europe 2016, Belgium, (2016)

20. A. Li, Y. Xing, W. Bogaerts, A Novel Sensing Scheme Based on Resonance Splitting of Silicon Microrings, IEEE Photonics Society Benelux, Belgium, (2015).

21. A. Li, T. Van Vaerenbergh, P. De Heyn, Y. Xing, P. Bienstman, W. Bogaerts, Experimentally demonstrate the origin for asymmetric resonance splitting and contributions from couplers to backscattering in SOI microrings, Integrated Photonics Research, Silicon and Nano Photonics (IPR 2015), United States, p.IM2B.6 (2015)

22. Y. Xing, T. Ako, J. George, D. Korn, H. Yu, P. Verheyen, M. Pantouvaki, G. Lepage, P. Absil, C. Koos, J. Leuthold, J. Beeckman, W. Bogaerts, Direct Digital Control of an Efficient Silicon+Liquid Crystal Phase Shifter, 11th International Conference on Group IV Photonics (GFP), France, p.43-44 (2014)

References

[1] Fred Kish, Vikrant Lal, Peter Evans, Scott W. Corzine, Mehrdad Ziari, Tim Butrie, Mike Reffle, Huan-Shang Tsai, Andrew Dentai, Jacco Pleumeekers, Mark Missey, Matt Fisher, Sanjeev Murthy, Randal Salvatore, Parmijit Samra, Scott Demars, Naksup Kim, Adam James, Amir Hosseini, Pavel Studenkov, Matthias Lauermann, Ryan Going, Mingzhi Lu, Jiaming Zhang, Jie Tang, Jeff Bostak, Thomas Vallaitis, Matthias Kuntz, Don Pavinski, Andrew Karanicolas, Babak Behnia, Darrell Engel, Omer Khayam, Nikhil Modi, Mohammad Reza Chitgarha, Pierre Mertz, Wilson Ko, Robert Maher, John Osenbach, Jeffrey T. Rahn, Han Sun, Kuang-Tsan Wu, Matthew Mitchell, and David Welch. *System-on-Chip Photonic Integrated Circuits*. IEEE Journal of Selected Topics in Quantum Electronics, 24(1):1–20, 1 2018.

[2] Amir H. Atabaki, Sajjad Moazeni, Fabio Pavanello, Hayk Gevorgyan, Jelena Notaros, Luca Alloatti, Mark T. Wade, Chen Sun, Seth A. Kruger, Huaiyu Meng, Kenaish Al Qubaisi, Imbert Wang, Bohan Zhang, Anatol Khilo, Christopher V. Baiocco, Milo A. Popović, Vladimir M. Stojanović, and Rajeev J. Ram. *Integrating photonics with silicon nanoelectronics for the next generation of systems on a chip*. Nature, 556(7701):349–353, 4 2018.

[3] Jianwei Wang, Stefano Paesani, Yunhong Ding, Raffaele Santagati, Paul Skrzypczyk, Alexia Salavrakos, Jordi Tura, Remigiusz Augusiak, Laura Mančinska, Davide Bacco, Damien Bonneau, Joshua W. Silverstone, Qihuang Gong, Antonio Acín, Karsten Rottwitt, Leif K. Oxenløwe, Jeremy L. O'Brien, Anthony Laing, and Mark G. Thompson. *Multidimensional quantum entanglement with large-scale integrated optics*. Science, 360(6386):285–291, 4 2018.

[4] Francesco Testa, Claudio J. Oton, Christophe Kopp, Jong Moo Lee, Rubn Ortuño, Reinhard Enne, Stefano Tondini, Guido Chiaretti, Alberto Bianchi, Paolo Pintus, Min Su Kim, Daivid Fowler, Jos ngel Ayúcar, Michael Hofbauer, Mattia Mancinelli, Maryse Fournier, Giovan Battista Preve, Nikola Zecevic, Costanza L. Manganelli, Claudio Castellan, Gabriel Parès, Olivier Lemonnier, Fabrizio Gambini, Pierre Labeye, Marco Romagnoli, Lorenzo Pavesi, Horst Zimmermann, Fabrizio Di Pasquale, and Stefano Stracca. *Design and Implementation of an Integrated Reconfigurable Silicon Photonics Switch Matrix in IRIS Project*. IEEE Journal on Selected Topics in Quantum Electronics, 22(6), 11 2016.

[5] Russell A. Budd, Laurent Schares, Benjamin G. Lee, Fuad E. Doany, Christian Baks, Daniel M. Kuchta, Clint L. Schow, and Frank Libsch. *Semiconductor optical amplifier (SOA) packaging for scalable and gain-integrated sil-*

icon photonic switching platforms. In Proceedings - Electronic Components and Technology Conference, volume 2015-July, pages 1280–1286. Institute of Electrical and Electronics Engineers Inc., 7 2015.

[6] Christopher V. Poulton, Ami Yaacobi, David B. Cole, Matthew J. Byrd, Manan Raval, Diedrik Vermeulen, and Michael R. Watts. *Coherent solid-state LIDAR with silicon photonic optical phased arrays*. Optics Letters, 42(20):4091, 10 2017.

[7] Wim Bogaerts, Umar Khan, and Yufei Xing. *Layout-Aware Yield Prediction of Photonic Circuits*. In IEEE International Conference on Group IV Photonics, pages 1–2, Cancun, Mexico, 2018. Ieee.

[8] Zeqin Lu, Jaspreet Jhoja, Jackson Klein, Xu Wang, Amy Liu, Jonas Flueckiger, James Pond, and Lukas Chrostowski. *Performance prediction for silicon photonics integrated circuits with layout-dependent correlated manufacturing variability*. Optics Express, 25(9):9712, 2017.

[9] Yufei Xing, Domenico Spina, Ang Li, Tom Dhaene, and Wim Bogaerts. *Stochastic collocation for device-level variability analysis in integrated photonics*. Photonics Research, 4(2):93, 2016.

2

Behavioral and Geometry Parameter Extraction of A Waveguide

Process variations have a significant impact on the performance of integrated photonics. To understand the behavior of process variations, we need methods to measure the variations accurately. Fabricated linewidth and thickness of waveguides are the most fundamental layout parameters that reflect the process quality. However, it is very costly to obtain them using current measurement techniques, which makes it difficult to obtain wafer maps to present process variations.

In this chapter, we will introduce an method to extract silicon waveguide geometry with sub-nanometer accuracy using spectral measurements. The effective index and group index of silicon on insulator (SOI) waveguides are extracted from the spectral measurements of a pair of Mach-Zehnder Interferometers (*MZI*). We built an accurate model mapping the SOI waveguide geometry to its effective index and group index to obtained accurate values of linewidth and thickness. We will analyze in detail how to set up the bounds for the effective index and group index to get the extraction with improved accuracy. We will also discuss principles to design the MZI extraction circuits. We applied the method on a die fabricated by IMEC multiproject wafer services and will present the result at the end of the chapter.

2.1 Process Monitoring for Silicon Photonic Waveguides

The sub-micron Silicon-on-Insulator (SOI) platform for silicon photonics offers tight confinement of light and compact integration of photonic devices. However, the high material index contrast also makes devices very sensitive to the geometry variation [1]. The variation introduced in fabrication often significantly deteriorates the device performance. Especially for spectral filters, geometry variation causes a shift in the spectrum and needs good compensation [2, 3]. Therefore, an accurate evaluation of the fabricated geometry helps to estimate how to compensate the performance error and make a sensible design.

2.1.1 Challenges in Metrology Measurement

Extracting the fabricated linewidth and layer thickness is essential in getting the input data for analyses such as performance evaluation, [4, 5], revising compact behavioral models [6], variability analysis in Chapter 5 and yield prediction in Chapter 6. However, metrology measurement of a fabricated photonic chip using a *scanning electron microscope* (SEM) is both expensive and destructive. Conventionally, the semiconductor fab only takes a few destructive cross-section pictures at given wafer locations on different dies, and this only during process development, not in production. The accuracy of such SEM-based measurements is usually limited to a few nm, which is good for a qualitative confirmation but not accurate enough for exact modeling. Ultra-precise methods such as *atomic force microscope* (AFM) are extremely time-consuming [7]. In production, non-destructive methods such as top-down SEM, ellipsometry, and scatterometry, are used to measure geometries such as line widths and layer thickness. The problem with such measurements is that they are collected in dedicated measurement sites, which are often not representative for the actual waveguide devices. The current methods are not capable of extracting accurate wafer maps of the actual device geometry on nanometer-scale and its variability.

An alternative approach is to use optical transmission measurements on the actual devices to extract the variation of geometry parameters. Investigating the spectral response of devices such as micro-disks or long Bragg gratings offers a more efficient way of characterizing manufacturing variations. However, most demonstrated methods either use dual-polarization measurements or request complex spectral reflection measurements from both ends of the device [4, 8]. Optical properties such as the effective index and group index can be extracted from interfering structures such as *Mach-Zehnder interferometers* (MZI) [3] and ring resonators [5, 9]. Recent research shows the possibility to correlate waveguide geometry with these behavioral parameters such as resonance wavelength and group

index by mapping them with a linear model [5]. Lu et al. measured spectral responses of ring resonators over the wafer and from this derived a geometry wafer map, demonstrating the potential of the optical method for wafer-scale geometry extraction. In this study, they used a ring resonator which consists of both straight and bent waveguides. Since straight and bend waveguides have different effective and group indices, the geometrical cross-section of a straight waveguide cannot be extracted accurately from a ring, without making assumptions on the straight and bend geometry that cannot be verified.

Even though geometry extraction through transmission measurements shows real practical potential, it is still challenging to characterize the manufacturing quality accurately. First of all, the geometry model that links geometry with the behaviour parameters should on one hand be very accurate, but on the other hand have as few parameters as possible. This is hard to achieve with combinations of different waveguide types (straight and bend), each with their own optical properties. Secondly, there is always noise and reflection-induced fringes in the spectral measurements, error in measurement alignment and variability in *grating couplers* (GC). It is essential to develop a method which is tolerant of the above-mentioned factors. Thirdly, geometry extraction requires a method to simultaneously extract both effective and group indices from the same device. Most importantly, it is not trivial to get the correct effective index from the spectrum of an interfering device [10]. We need a quantitative discussion to choose the right effective index from many possible solutions.

2.1.2 The Workflow of the Process Monitoring using Optical Measurement

In this chapter, we will address the above-mentioned challenges in a systematic way. As shown in Fig. 2.1, we first perform optical measurements on MZIs. We build a circuit model of the device and match the simulated transmission curve with measurements to get behavioral parameters of a waveguide such as n_{eff} and n_g. Next, we build an accurate model to map n_{eff} and n_g to waveguide width and thickness. Then, using the model, we obtain waveguide geometry parameters from extracted behavior parameters. We automated the optical measurement and repeated the geometry extraction on devices with the same design over the wafer. From the extraction, we perform a variability analysis and derive fabrication variation with high accuracy. In Section 2.2, we propose an improved geometry model to offer high modeling accuracy. In Section 2.3, we show that a curve fitting method is less sensitive to measurement noise and helps in removing the grating coupler envelope, and in Section 2.4, we show how to design two MZIs to extract the accurate and unique n_{eff} and n_g given the waveguide geometry variation range. Section 2.5 discusses the procedure of the waveguide geometry extraction from a

couple of MZIs. Finally, in Section 2.6 we apply the method to extract linewidths and thicknesses of SOI waveguides on a die fabricated by the IMEC MPW service.

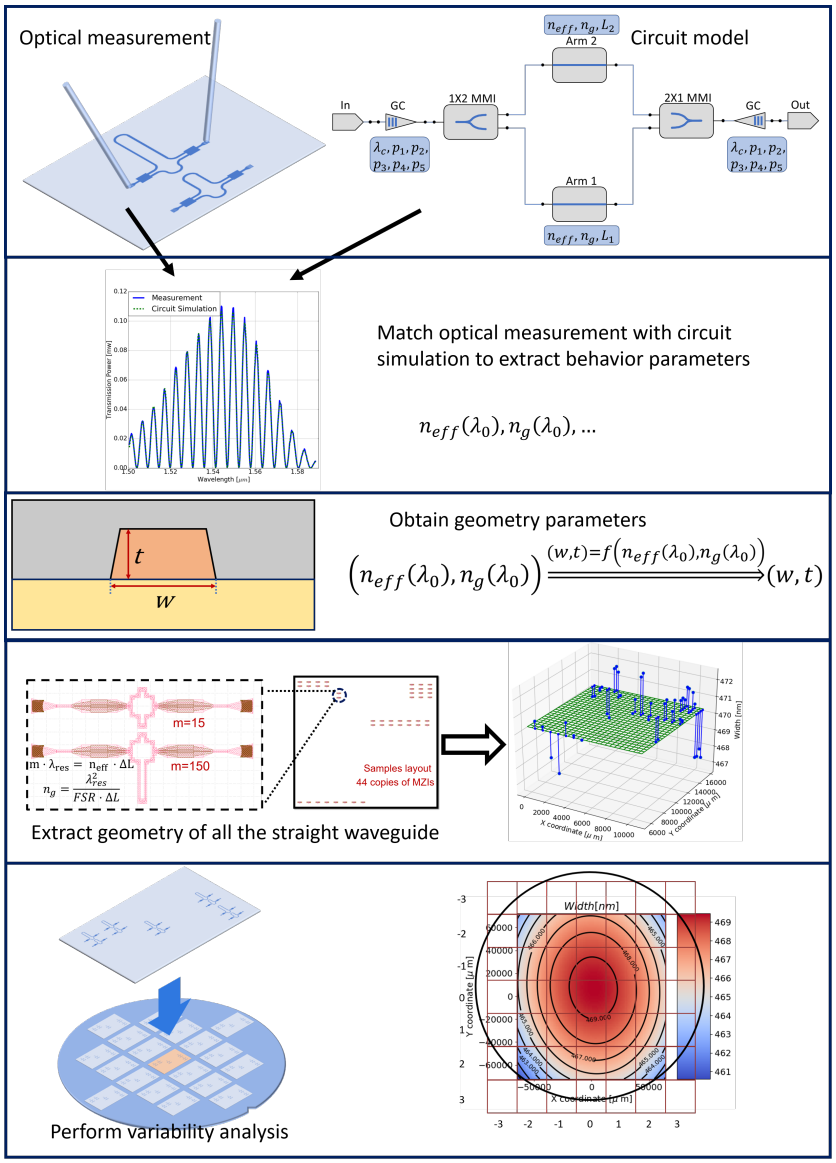


Figure 2.1: The workflow of extracting behavioral parameters and fabricated geometry using optical measurements.

2.2 An Accurate Geometry Model of Waveguide

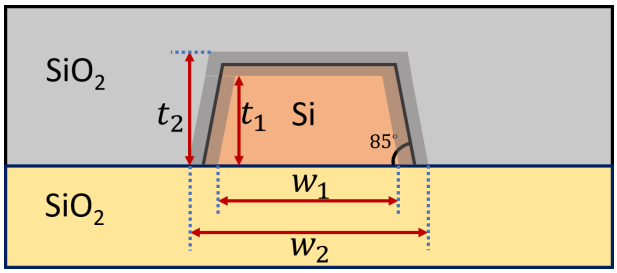


Figure 2.2: Cross-section schematic of an oxide-clad SOI strip waveguide with a 85° sidewall angle.

We do not measure linewidth w and thickness t directly, so we infer them from the results of optical transmission measurements. In particular, we first extract the effective index n_{eff} and group index n_g of the waveguide. To calculate the fabricated waveguide linewidth w and thickness t from the effective index n_{eff} and group index n_g, we require a geometry model that links n_{eff} and n_g to w and t. To build the model, we simulated oxide-clad Si waveguide cross-section (Fig. 2.2) with the COMSOL Multiphysics *Finite Element Method* (FEM) solver. According to the IMEC technology handbook of the run, a fabricated strip waveguide has a sidewall angle of 85°, so we simulated the waveguide with a 85° sidewall angle. The waveguide width is the bottom width of the trapezoidal. We swept width from 440 nm to 500 nm and thickness from 195 nm to 235 nm, and calculated n_{eff} and n_g at 1550 nm wavelength. The linewidth-thickness grid in Fig. 2.3 (a) can be mapped one-on-one to the simulated $n_{eff} - n_g$ grid (black solid line with circles) in Fig. 2.3 (b). Then, we fitted simulated n_{eff} and n_g using polynomials of w and t. We build first-, second- and third-order polynomial models (dashed lines Fig. 2.3 (b)). Both n_{eff} and n_g vary quite linearly with w and t. Nonetheless, the first-order fitted linear model shows a clear deviation (maximum 0.32 per cent error in n_{eff} and 0.40 per cent in n_g) from simulated n_{eff} and n_g. The third-order model matches the simulation very accurately(, as will be quantified in the discussion below). Then, to get w and t from spectral measurements, we wrote w and t , as a polynomial of n_{eff} and n_g at 1550 nm as:

$$w = p_0 + \Sigma_{j=1,i=1}^{n} p_{i,j} n_{eff}^i n_g^j \tag{2.1}$$

$$t = q_0 + \Sigma_{j=1,i=1}^{n} q_{i,j} n_{eff}^i n_g^j \tag{2.2}$$

where p_0 and q_0 are constant terms, $p_{i,j}$, and $q_{i,j}$ are coefficients in polynomials, and n is the order of the polynomial.

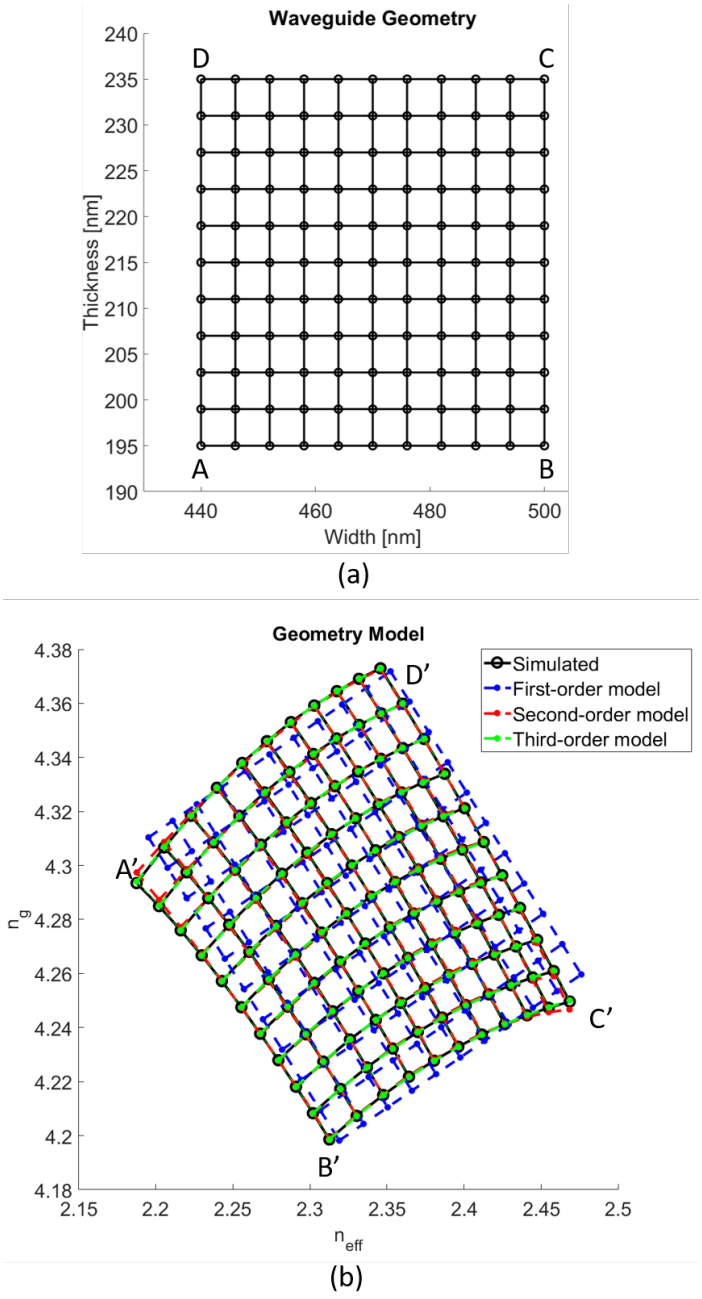


Figure 2.3: (a) width and thickness grid of strip waveguides; (b) effective and group index of strip waveguides on the geometry grid using the COMSOL FEM simulation, and the first-, the second- or the third-order polynomial mapping model.

Using a simulated model can introduce a *simulation error*, coming from a difference between the actual waveguide geometry (both dimensions, shape and material properties) and the trapezoidal geometry model we used in the mode solver. We have considered the sidewall angle in our model, but still we don't know the actual geometry, so it is very hard to compensate for this error. This means that some parameters will be only relative.

In addition, the *mapping error* is the difference between the simulated trapezoidal waveguide geometry and extracted trapezoidal waveguide geometry using the geometry model. To calculate the mapping error, we first use simulated n_{eff} and n_g for simulated geometry. Then, we calculate the waveguide geometry using a polynomial model with the simulated n_{eff} and n_g. The mapping error reduces significantly with the order of the polynomial.(Table. 2.1) The first-order model has a maximum error of several nanometers which is comparable to the reported intra-wafer manufacturing variations in width of 0.78 nm to 2.65 nm, and in thickness of 0.83 nm to 4.16 nm. [11–16] Obviously, the modelling error is too large to extract intra-die variations (variations between same devices on one die). For a good estimation of the fabricated geometry variation, especially to study variability on intra-wafer level and intra-die level, a lower mapping error is required. The third-order polynomial model has maximum error of 0.05 nm for both linewidth and thickness (Fig. 2.4), which is one magnitude smaller than the fabrication variation. A low modelling error makes geometry extraction more accurate and credible. Polynomial with a higher-order does not significantly improve the mapping accuracy so that we used the third-order polynomial as our model.

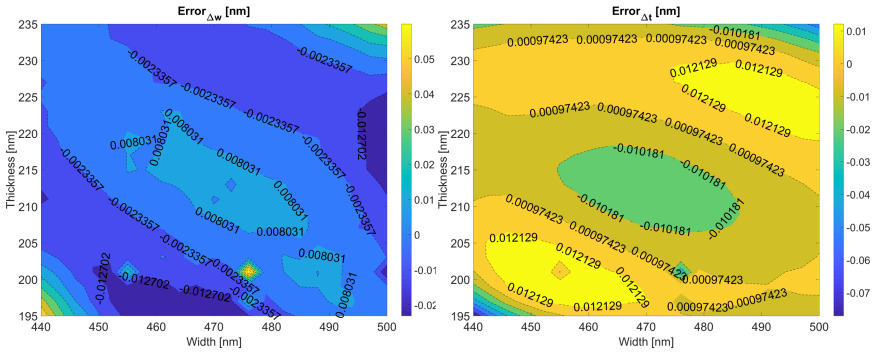


Figure 2.4: Error contour plot of the proposed third-order polynomial model where w ranges from 440 to 500 nm and thickness ranges from 195 to 235 nm. Left: width extraction error; Right: thickness extraction error.

	Error Δw [nm]	Error Δt [nm]
1st order	5.10	4.57
2nd order	0.14	0.18
3rd order	0.06	0.08

Table 2.1: Error of polynomial models (order = 1,2,3).

2.3 Extracting Effective Index and Group Index from an MZI

Before we can determine the waveguide geometry from n_{eff} and n_g, we need to experimentally measure those quantities of a device and isolate the values for a straight waveguide. This effectively means that we need a device where a transmission measurement can give us an accurate extraction of both n_{eff} and n_g for a straight waveguide only, and which is in the presence of measurement noise and variation in the coupling structures.

2.3.1 Using an MZI or a Ring Resonator?

In [5], ring resonators were used to extract n_{eff} and n_g. A ring resonator has sharp resonance peaks. Spectral measurement of a ring requires very fine resolution. Also, if you get asymmetric peak splitting, you would get a wrong peak position [17]. Moreover, a ring uses either a bent waveguide or it combines bends with straight sections, and the round trip path also includes the coupling sections which will also have different optical properties. As such we cannot isolate n_{eff} and n_g of the straight or bent waveguide. The alternative is to use an MZI with two arms that are identical but for the fact that one arm has a longer straight part than another [3]. Ideally, the n_{eff} and n_g of the two arms are identical so that the spectral response of the MZI is only dependent on the path length difference between two arms. We can use the MZI to measure varying waveguide geometry under process variation at different locations. In practice, the path length difference in a single MZI is also induced by a difference in the n_{eff} in the two arms because of local process variability. Also, there can be a difference in the n_{eff} that the bends will contribute. Those differences also lead to the extraction error in n_{eff} and n_g. However, since the distance between waveguide within an MZI this compact is within 100 μm, we can safely assume that the error will be much smaller than the device-to-device variation.

An interfering structure such as a ring or an MZI would have a constructive

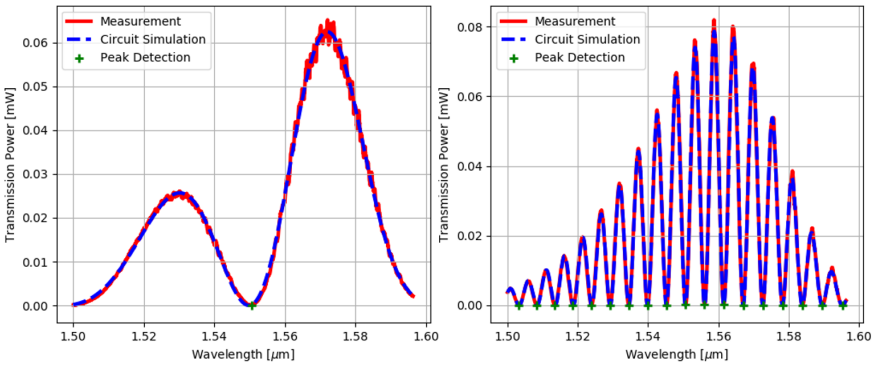


Figure 2.5: This figure shows the measured transmission spectrum (red solid) and fitted spectrum (blue dot) using the circuit model including the polynomial GC model. Also, valleys of the spectrum (green cross) are found by the peak detection method. Left: the low-order MZI. Right: the high-order MZI.

interference when interference order m is an integer and

$$m = \frac{n_{eff,0}\Delta L}{\lambda_{res}} \tag{2.3}$$

$$FSR = \frac{\lambda_{res}^2}{n_g \Delta L} \tag{2.4}$$

where $n_{eff,0}$ is the prior estimate at the resonance wavelength λ_{res} and ΔL is the physical path length difference.

2.3.2 Extraction Method: Resonance Detection vs. Curving Fitting

If we know the resonance order m, we could get n_{eff} by locating λ_{res} in the output spectrum. It is natural to apply the resonance detection method [5, 9] to locate resonances in the spectrum to get both n_{eff} and n_g. Extraction of parameters from resonance is accurate when the device has sharp resonance peaks, which is the case for a ring resonator. However, an MZI has a sinusoidal-like spectral transmission. Its curve is quite flat near a maximum or minimum. Especially when measurement noise is involved, it is hard to locate its peaks (constructive interference) or valleys (destructive interference) by peak detection method (Fig. 2.5 green cross indicates detected valleys). Using only maximum detection method leads to a significant error (Table. 2.2)in effective index and group index extraction making it not suitable for geometry extraction.

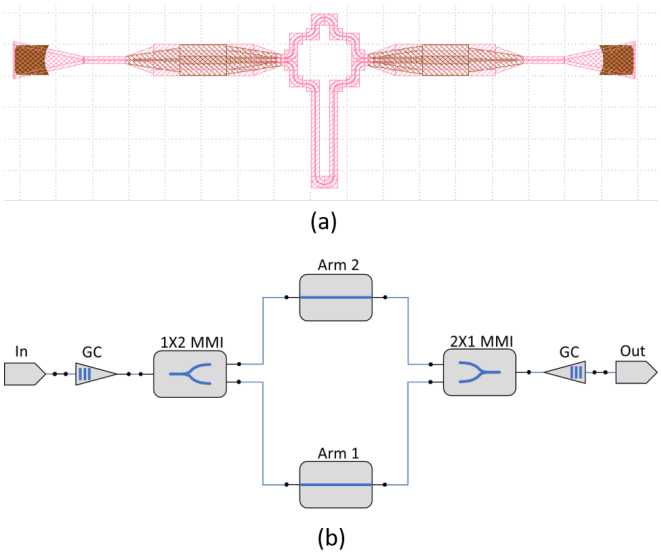


Figure 2.6: Top: layout of the MZI under test. Bottom: circuit schematic of the MZI.

To improve the extraction accuracy, we used a curve fitting technique. It extracts parameters by minimizing the difference between a circuit simulation and the measurement data. While maximum/minimum extraction only uses information at the peaks and ignores information on the rest of the spectrum, curve fitting method utilizes the information from the entire measured spectrum, which is more tolerant to the measurement noise and gives more reliable extraction.

We simulate the MZI circuit in Caphe, [18] a circuit simulator that calculates the scattering matrix of the circuit from defined scattering matrix of each component. We built a Caphe circuit model of the MZI the same way as our *device under test* (DUT), which has two types of components: waveguide and *multimode interference (MMI)* coupler.

	n_{eff}	n_g	width [nm]	thickness [nm]
Curve fitting using a GC model	2.319	4.291	466.0	211.8
Peak detection	2.318	4.302	462.0	213.8
Difference between two methods	0.001	0.009	4.0	2.0

Table 2.2: Comparison between the peak detection method and the curve fitting method. We applied both methods to extract parameters from the spectrum of a high-order MZI (Fig. 2.5 right)

2.3.2.1 Removing the Effect of the Grating Couplers

For automated measurement, light is vertically coupled to the DUT using a pair of *grating couplers* (GC). We need to remove the envelope of the GC before fitting the spectrum with the circuit model. We can remove the GC in two ways. One way we measure a pair of reference GCs, preferably close to the DUT. By subtracting the measured reference GC from the DUT spectrum in log-scale, we can normalize the transmission spectrum of the DUT. It is a very common practice in optical transmission measurements. However, this method is error-prone in the presence of fabrication variation in GCs, because the reference GC can be subtly different from the GCs connected to DUT. Even more, the input and output GC of the DUT can be different from each other. As shown in the Fig. 2.7, the MZI transmission after subtracting the reference is not as we would expect: the linear-scale spectrum should be a sine-like curve with maxima of the same amplitude. Also, the normalization cannot be correct, because you would have transmission larger than 1. The significant mismatch between measured and fitted spectrum introduces a large error in the extracted parameters.

Empirically, the transmission spectrum passing through input GC to output GC on the logarithmic scale (expressed in dB) can be well fitted by a fourth-order polynomial. So we use a fourth-order polynomial to represent the transmission spectrum of the combined input and output GC on the logarithmic scale. We include two GCs in the circuit model together with the MZI. As shown in the Fig. 2.5, the fitting is considerably enhanced where the simulation matches measurement nicely. From fitting, we can get circuit parameters such as effective index, group index and coefficients of the polynomial describing the GC.

2.3.2.2 Fitting Accuracy vs. MZI Order

In the transmission spectrum of the MZI, the positions of the peaks and valleys give information about the effective index n_{eff}. The periodicity of the transmission spectrum is determined by the group index n_g.

An MZI with low order m has only a few peaks/valleys in the measurement band, and therefore it will have a low accuracy on the extraction of n_g. On the other hand, a high-order MZI can give a high accuracy of n_g extraction. As explained by Dwivedi et al. in [3] the combination of a low-order MZI and a high-order MZI can give a good accuracy on both n_{eff} and n_g.

When noise is mingled in the measured spectrum, it will induce an additional uncertainty in the curve fitting. The uncertainty is proportional to the square root of the residual of the fitting. [19] Therefore, the spectral fringes and noise would increase the uncertainty of the extracted parameters, thus lowering the extraction accuracy. The uncertainties on the parameters are the standard deviations of the parameters as the fitting process takes place. If the goodness of fit depends strongly

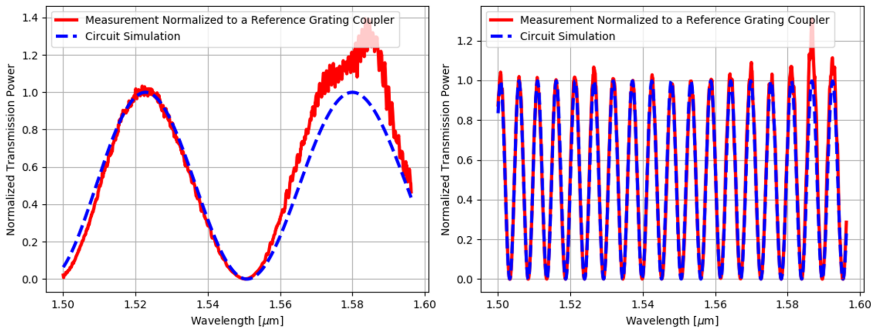


Figure 2.7: We removed the GC envelope using a reference GC near the DUT. Fabrication variation caused the measured spectrum after GC removal far from ideal (as shown by spectrum simulated by the circuit model) as ideally the peaks in the spectrum should have the same amplitude. Also, you have transmission larger than 1, so the normalization cannot be correct. After the GC removal, we fitted the measured spectrum with the circuit model (Fig. 2.6), not including GC. Red solid: measured transmission spectrum after removing the GC envelope using a reference GC. Blue dot: fitted spectrum using the circuit model. Left: the low-order MZI. Right: the high-order MZI.

on a particular fit parameter, the uncertainty will be low. Then, it is a question of how the fitting accuracy or uncertainty is related to the order of the MZI. In this work, we used the non-linear least-squares method to fit the transmission curves with a waveguide compact model.

Order	15	50	100	150
n_{eff}	2.336264	2.340747	2.339561	2.339729
n_g	4.28	4.290649	4.288268	4.288448
w [nm]	472.6998	469.3341	470.0631	470.0189
t [nm]	212.7182	215.6051	214.9168	214.9829
$Uncertainty_{n_{eff}}$	0.014705	2.18E-05	1.07E-05	7.39E-06
$Uncertainty_{n_g}$	0.718928	0.001366	0.000663	0.00046
$Uncertainty_w$ [nm]	270.3557	0.511691	0.248367	0.172499
$Uncertainty_t$ [nm]	137.6516	0.270922	0.130336	0.090586

Table 2.3: Fitting uncertainty vs. interference order

We built a circuit model of a waveguide with 470 nm width and 215 nm thickness and simulated the transmission spectrum. Then, we add a $\pm$ 0.2 dBm to the spectrum to emulate the typical spectral fringes led by reflection and measurement noise. Finally, we get n_{eff} and n_g using the curve fitting. The fitting uncertainty we presented is the estimate of 1.96 times standard deviations of each of the pa-

rameters, which provides confidence limits of approximately 95%. As shown in Table. 2.3, we did not get accurate n_g from an MZI with order = 15. This leads to a huge uncertainty in extracted width and thickness. The fitting uncertainty decreases with the interference order. Therefore, an increasing interference order of the MZI improves fitting accuracy.

2.4 Design Principle of the MZI for Geometry Extraction

When fitting the transmission curve of the MZI, the fitting algorithm implicitly assumes that the order m of the MZI is sufficiently accurate, i.e., that the peak near the center wavelength of 1550 nm corresponds with the designed order m. However, in the presence of fabrication variation, this is not necessarily the case, and as the designed order of the MZI increases, the uncertainty on the measured order increases. Therefore, the design parameters should be chosen such that the low-order MZI can be used to pin the order m of the device unambiguously [3], and make a good estimate of n_{eff}. The order of the high-order MZI should be chosen such that a maximum of information can be extracted, based on the estimate of n_{eff} obtained from the low-order MZI. We discuss the design process for these devices.

2.4.1 The Interference Order under a Given Effective Index Variation

Eq. 2.4 shows that if we know the resonance order m, we can calculate n_{eff} from the peak locating λ_{res} in the output spectrum. However, if fabrication variations can shift the spectrum more than half a *free spectral range* (FSR), we can no longer be certain of the order m. Therefore, we should design the MZI with sufficiently low order m such that the order at the center wavelength wavelength is always within $m \pm 0.5$, which means

$$\begin{aligned} \frac{n_{eff,0}\Delta L}{\lambda} &= m \\ \frac{(n_{eff,0} - \Delta n_{eff}/2)\Delta L}{\lambda} &> m - 0.5 \\ \frac{(n_{eff,0} + \Delta n_{eff}/2)\Delta L}{\lambda} &< m + 0.5 \end{aligned}$$

Given the variation in n_{eff} is Δn_{eff}, we can decide the order m that fulfills the condition and would give us sufficient confidence:

$$m < \frac{n_{eff,0}}{\Delta n_{eff}} \tag{2.5}$$

This is equivalent to a constraint on the length difference ΔL between the two arms of the MZI:

$$\Delta L < \frac{\lambda}{\Delta n_{eff}} \tag{2.6}$$

2.4.2 The Bounds for Effective Index and Group Index from the Geometry Variation

Within the measurement interval, the spectrum of an MZI looks quasi identical if we shift the interference order m by an integer number. Without a proper confidence interval on n_{eff}, there would be multiple solutions of n_{eff} to fit the spectrum. As n_{eff} and n_g can be mapped to linewidth w and thickness t, we can derive the bound of (n_{eff}, n_g) from the confident interval of the geometry parameters (w, t) which are supplied by the fab. As presented in section 2.2, n_g and n_{eff} can be accurately mapped on w and t by a third-order polynomial model. For simplicity of analysis in the derivation below, we use the linear geometry model where

$$n_{eff} = n_{eff_0} + \frac{\partial n_{eff}}{\partial w}(w - w_0) + \frac{\partial n_{eff}}{\partial t}(t - t_0)$$
$$n_g = n_{g_0} + \frac{\partial n_g}{\partial w}(w - w_0) + \frac{\partial n_g}{\partial t}(t - t_0)$$

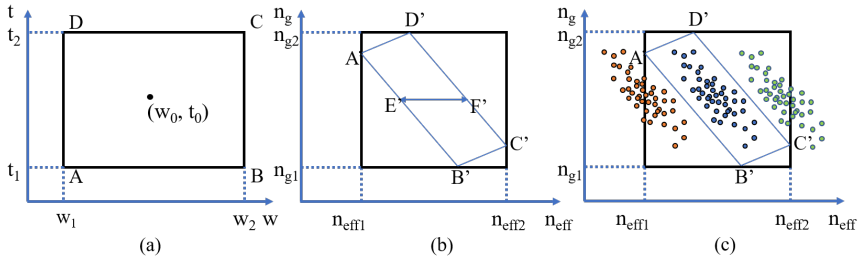


Figure 2.8: Bounds of the extraction. (a) The bound of width and thickness (b) Rectangle bound: reference [3], parallelogram: reduced bounds by linear transformation of geometry bounds (c) Rectangle bounds can not separate three groups of solutions (red, blue, green circles). The parallelogram cleanly isolates the correct solutions (blue circles).

When the bounds for linewidth and thickness form the rectangle $ABCD$ in (w, t) space (Fig. 2.8 a), the parameter range in (n_{eff}, n_g) space lies in-between

$n_{eff} \in [n_{eff_1}, n_{eff_2}]$ and $n_g \in [n_{g_1}, n_{g_2}]$:

$$n_{eff_1} = n_{eff_0} + \frac{\partial n_{eff}}{\partial w}(w_1 - w_0) + \frac{\partial n_{eff}}{\partial t}(t_1 - t_0) \tag{2.7}$$

$$n_{eff_2} = n_{eff_0} + \frac{\partial n_{eff}}{\partial w}(w_2 - w_0) + \frac{\partial n_{eff}}{\partial t}(t_2 - t_0) \tag{2.8}$$

$$n_{g_1} = n_{g_0} + \frac{\partial n_g}{\partial w}(w_2 - w_0) + \frac{\partial n_g}{\partial t}(t_1 - t_0) \tag{2.9}$$

$$n_{g_2} = n_{g_0} + \frac{\partial n_g}{\partial w}(w_1 - w_0) + \frac{\partial n_g}{\partial t}(t_2 - t_0) \tag{2.10}$$

where w_0 and t_0 are the nominal values for w and t variations. The range is a rectangle in (w, t) space whose centre is (w_0, t_0). Assume $\Delta w = w_2 - w_1$ and $\Delta t = t_2 - t_1$. As (w, t) and (n_{eff}, n_g) follow a near linear mapping, the vertices of the bound rectangle in the (w, t) space A,B,C,D can be mapped to the (n_{eff}, n_g) space as A',B',C',D'. Intuitively, the bound rectangle $ABCD$ in (w, t) space is linearly transformed into a parallelogram $A'B'C'D'$ in (n_{eff}, n_g) space. The tilted boundary $A'B'C'D'$ is within the original rectangular boundary but much smaller. For the fundamental TE mode of a SOI oxide-clad waveguide, $\frac{\partial n_g}{\partial w}$ is negative while $\frac{\partial n_{eff}}{\partial w}$, $\frac{\partial n_{eff}}{\partial t}$ and $\frac{\partial n_g}{\partial t}$ are positive. The parallelogram $A'B'C'D'$ then will be tilted as in Fig. 2.8 (b).

The bounds of the confidence interval for n_{eff} are fundamental to get a correct extraction. Extraction of the group index n_g does not pose that much of a problem, as the confidence interval is much larger, and there are no multiple solutions. Depending on whether we know n_g of the same waveguide, we can estimate the range for n_{eff} in two ways.

2.4.2.1 Low-order MZI: Estimating the Effective Index without information on the Group Index

Sometimes we cannot obtain accurate information of n_g, such as when we are using a low-order MZI. Then, as in [3], we can calculate the uncertainty Δn_{eff} from geometry variations Δw and Δt as

$$\Delta n_{eff,rectangle} = n_{eff_2} - n_{eff_1} = \frac{\partial n_{eff}}{\partial w}\Delta w + \frac{\partial n_{eff}}{\partial t}\Delta t \tag{2.11}$$

which is essentially corresponds to the width of the rectangle $ABCD$, or the horizontal distance $A'C'$.

2.4.2.2 High-order MZI: Estimating Effective Index with information on Group Index

The maximal range of the n_{eff} for a given n_g is $E'F'$ (Fig. 2.8 (b)), which is the maximal distance between two edges of the parallelogram at the given n_g.

The distance is dependent on the shape of the parallelogram. When A is higher than C ($n_(g, A) > n_(g, C)$), using Eq. 2.7 we can derive $\frac{\partial n_g}{\partial t}\Delta t < -\frac{\partial n_g}{\partial w}\Delta w$. Then $E'F'$ is the horizontal distance between line $A'B'$ and $C'D'$, and the range of n_{eff} is determined by the range of linewidth Δw. When A is lower than C ($n_(g, A) < n_(g, C)$), using Eq. 2.7 we can derive $\frac{\partial n_g}{\partial t}\Delta t > -\frac{\partial n_g}{\partial w}\Delta w$. Then $E'F'$ is the horizontal distance between line $A'C'$ and $B'D'$, and the range of n_{eff} is determined by the range of linewidth Δt.

When $\frac{\partial n_g}{\partial t}\Delta t < -\frac{\partial n_g}{\partial w}\Delta w$,

$$\Delta n_{eff,parallelogram} = \left(-\frac{\frac{\partial n_{eff}}{\partial w}\frac{\partial n_g}{\partial t}}{\frac{\partial n_g}{\partial w}} + \frac{\partial n_{eff}}{\partial t} \right) \Delta t \tag{2.12}$$

When $\frac{\partial n_g}{\partial t}\Delta t > -\frac{\partial n_g}{\partial w}\Delta w$,

$$\Delta n_{eff,parallelogram} = \left(\frac{\partial n_{eff}}{\partial w} - \frac{\frac{\partial n_g}{\partial w}\frac{\partial n_{eff}}{\partial t}}{\frac{\partial n_g}{\partial t}} \right) \Delta w \tag{2.13}$$

For the same geometry variation, an estimate of n_g reduces the uncertainty on Δn_{eff}. Fig. 2.8 (c) shows that we can separate three groups of solutions with the bound $A'B'C'D$ which are all located in the previous rectangle bound. As we will show in Chapter 5, thickness varies slowly across the wafer. The device-to-device thickness variation is much smaller than the device-to-device width variation. In this case, $\frac{\partial n_g}{\partial t}\Delta t < -\frac{\partial n_g}{\partial w}\Delta w$ is true for local variations. From Equation 2.11 we can calculate the n_{eff} boundary without knowledge on the n_g of the waveguide. Using 2.12, we can calculate the boundary knowing n_g. The ratio of Δn_{eff} between two situations is:

$$\frac{\Delta n_{eff,rectangle}}{\Delta n_{eff,parallelogram}} = a\frac{\Delta w}{\Delta t} + b$$

where

$$a = \frac{\frac{\partial n_{eff}}{\partial w}}{-\frac{\partial n_{eff}}{\partial w}\frac{\partial n_g}{\partial t}/\frac{\partial n_g}{\partial w} + \frac{\partial n_{eff}}{\partial t}}$$

$$b = \frac{\frac{\partial n_{eff}}{\partial t}}{-\frac{\partial n_{eff}}{\partial w}\frac{\partial n_g}{\partial t}/\frac{\partial n_g}{\partial w} + \frac{\partial n_{eff}}{\partial t}}$$

Both a and b in the equations are positive, so that the ratio is increasing with $\frac{\Delta w}{\Delta t}$. Intuitively, the smaller is AD compared to AB in Fig. 2.8 (a), the shorter EF is. A similar conclusion can be made when $\frac{\partial n_g}{\partial t}\Delta t < -\frac{\partial n_g}{\partial w}\Delta w$, where the ratio is increasing with $\frac{\Delta t}{\Delta w}$. So if the Δt is much smaller than the Δw, we can reduce the boundary of n_{eff} from knowing n_g. This is very helpful because with a small Δn_{eff} (refer Equation 2.5), we can design a high-order MZI to extract the local variation that improves the accuracy of extraction.

2.5 Geometry Extraction on a Fabricated Silicon Wafer

2.5.1 Extracting Effective Index and Group Index from Two MZIs

The total process variation (intra-die, die-to-die, and wafer-to-wafer) on an isolated waveguide on SOI platform is large. The variation can be several tens of nanometers for both linewidth and thickness. As discussed in the previous section, to capture the large variation using an MZI, we should choose a sufficiently low order m. However, as discussed in Section 2.3.2.2, this low-order MZI suffers from a low accuracy on n_g extraction. On the other hand, high-order MZI can offer good accuracy of n_g extraction. So a combination of the two devices can give us both essential optical parameters. So we can extract n_{eff} and n_g using a low order MZI and a high order MZI:

1. Extract a good estimate of n_{eff} from the low-order MZI
2. Extract an accurate n_g from the high-order MZI

Even though the devices are close together, they do not have the same n_{eff} and n_g because of local variations. To accurately map the waveguide geometry, we need to extract both n_{eff} and n_g from the same waveguide. In the following discussion, we will present how to obtain accurate n_{eff} and n_g both of the high order MZI in three step below.

1. Extract a good estimate of n_{eff} from the low-order MZI. Also, extract an accurate n_g from the high-order MZI.
2. Obtain the n_{eff} die map by an interpolation of the estimated n_{eff} from the low-order MZI. The map offers the average n_{eff} at each location where we can remove the local variation.
3. Use interpolated $n_{eff,\mu}$ at each location as a reference, and extract accurate n_{eff} from the high-order MZI using its n_g to limit the n_{eff} boundary.

We will discuss why we set up this step-by-step extraction procedure in the following section.

2.5.2 Extracting Inter-die and Intra-die Variability in Three Steps

In a wafer-scale fabrication process, we can identify different levels of process variations. As we will explain in Chapter 5, for each die, all the variations that originate at levels such as lot-to-lot, wafer-to-wafer, and intra-wafer (die-to-die) variations have the same impact on every device in the die. Since these variations has the same impact on every device in the die, we categorize these higher-level spatial variations together as the global variation. Meanwhile, we get also the intra-die (device-to-device) variation that affects devices differently on the same die. We can further decompose the intra-die variation into location-dependent variation and local variation. The location-dependent variation is the variation depending on the location of the device on the die. It can be caused by the continuous variation of thickness, photoresist spinning effects or plasma distributions, and other equipment non-uniformity that affecting the fabricated geometry varied spatially. On the other hand, the local variation we define here induces local disparities between devices placed close together (less than a few hundred microns apart). It includes intrinsic variability such as thickness fluctuations and width variations caused by pattern density non-uniformity, which in Chapter 5 are categorized into intra-die variation. The sum of the three variations V gives us the process variation of a device.

$$V_{total} = V_{inter-die} + V_{location-dependent} + V_{local} \tag{2.14}$$

The total process variation is considerably larger than the local (intra-die) variation. The total linewidth and thickness variation of an isolated waveguide on SOI platform can amount to tens of nanometers, [20] while intra-die variation is typically only a few nanometers. [5, 13, 20, 21]

With our two MZIs, we address variations on the different levels in three steps. The first step, we extract n_{eff} from a low-order MZI. Since the n_g extracted from the low-order MZI is very inaccurate, we estimate the range of n_{eff} without the information of n_g by substituting geometry variation by the total variation in Eq. 2.11:

$$\Delta n_{eff,total} = \frac{\partial n_{eff}}{\partial w}\Delta w_{total} + \frac{\partial n_{eff}}{\partial t}\Delta t_{total}$$

Given the range, we derive fairly accurate value of n_{eff} from the low-order MZI. In a second step, we obtain the n_{eff} map over the die by interpolation. The map offers the average n_{eff} of waveguides placed at each location where we can remove the local variation, and the inter-die variation and location-dependent variation

together determines the average value. In the third step, we use interpolated $n_{eff,\mu}$ at each location as a reference. Now, rather than the total variation we can only deal with the much smaller local variation. Since we can accurately extract n_g on the high-order MZI, the range for n_{eff} under the local variation is estimated by substituting geometry variation using the local variation in Eq. 2.12 and 2.13.

$$\Delta n_{eff,local} = \left(-\frac{\frac{\partial n_{eff}}{\partial w}\frac{\partial n_g}{\partial t}}{\frac{\partial n_g}{\partial w}} + \frac{\partial n_{eff}}{\partial t} \right) \Delta t_{local}$$

Because our analysis decreases the bound of extraction, we can use a much higher-order MZI to get n_{eff} and n_g simultaneously and accurately.

2.5.3 The Specification of the Two MZIs

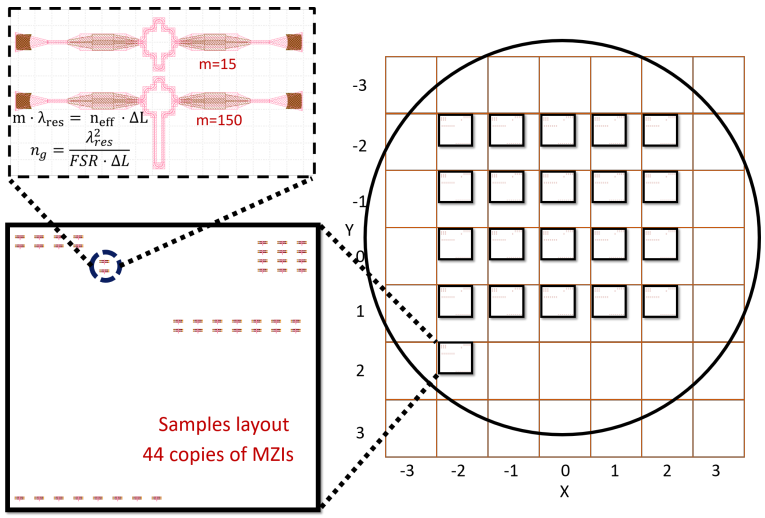


Figure 2.9: Left top: low-order and high-order MZI we used for geometry extraction. Left bottom: locations of two devices on a die. Right: locations of dies on the wafer. Red grid indicates dies on the wafer. The black circle is the boundary of the wafer.

The MZIs each consist of two waveguide arms and two 50-50 MMI coupler (Fig. 2.9 left). Our devices are fabricated by the IMEC *multi project wafer* (MPW) service. We design the waveguide with a linewidth of 450 nm. According to the technology handbook, the fabricated waveguide has a sidewall angle of 85°. Also, the 450 nm waveguide on the mask is measured to have 470 nm mean value and $\pm$ 20 nm variations in fabrication. The nominal thickness is 215 nm and the variation is $\pm$ 10 nm. The pre-estimated $n_{eff,0}$ of 470 nm $\times$ 215 nm waveguide is 2.340.

Using Eq. 2.11 we calculated the variation of $n_{eff,total}$ is

$$\begin{aligned}\Delta n_{eff,total} &= \frac{\partial n_{eff}}{\partial w}\Delta w_{total} + \frac{\partial n_{eff}}{\partial t}\Delta t_{total} \\ &= 0.002055 \times 40nm + 0.003916 \times 20nm \\ &= 0.16.\end{aligned}$$

From the third-order model we calculate that the $\Delta n_{eff,total}$ is also 0.16. The arm length difference of the low-order MZI is $\Delta L < \frac{1.55\mu m}{0.16} = 9.7\mu m$ and the order of $m_l < \frac{n_{eff,0}}{\Delta n_{eff,total}} \approx 15.00$.

Only a few references report typical fabricated geometry maps of silicon waveguides on die-level. Thickness maps in SOI depend largely on the qualities of the source wafer. Linewidth maps depend much more on the actual fabrication process and will be very different for devices fabricated with deep UV lithography or e-beam lithography, and vary between fabs. From the research of Lu et al. [5], the thickness varies slowly over the die, and the maximum difference between neighboring thickness is 0.5 nm. So we assume as a worst case that the thickness is slowly changing locally with only $\Delta t_{local} = 2.00$ nm. We also assume that $\Delta w_{local} = 15.00$ nm, which is significant. Using Eq. 2.12, we calculated the variation of $n_{eff,intra-die}$. Notice that we have no pre-estimate of the nominal value of w and t locally so that they can be any value within the specified fabrication window. We calculated $-\dfrac{\dfrac{\partial n_{eff}}{\partial w}\dfrac{\partial n_g}{\partial t}}{\dfrac{\partial n_g}{\partial w}} + \dfrac{\partial n_{eff}}{\partial t}$ for $w \in [450, 490]$ nm and $t \in [205, 225]$ nm. Its value is in between 0.0045 and 0.0074. Therefore,the max local n_{eff} variation we can surely extract is

$$\Delta n_{eff,intra-die} = \left(-\frac{\dfrac{\partial n_{eff}}{\partial w}\dfrac{\partial n_g}{\partial t}}{\dfrac{\partial n_g}{\partial w}} + \frac{\partial n_{eff}}{\partial t} \right) \Delta t_{intra-die} = 0.0074 \times 2 = 0.0148,$$

which is the same calculated using the third-order model. The arm length difference of the high-order MZI $\Delta L_h < \dfrac{\lambda}{\Delta n_{eff,local}} = 104.7\mu m$ where the order is around $m \approx 158$. The confidence limit estimated with the information on n_g on the order is 2.93 $\times$ that calculated by the original method without the information on n_g. Based on the estimation, we design the low-order MZI to have an order around $m = 15$ at 1550 nm and order of the high-order MZI is $m = 150$.

We assumed that we cannot obtain accurate n_g on the low-order MZI so that the designed low-order MZI is only tolerant to $\pm$ 20 width variation and $\pm$ 10 thickness variation. One improvement can be made by estimating the lower-order n_g from the accurate high-order n_g. We can derive a n_g die map by fitting all the high-order n_g on the die, and use that as a reference to estimate the range of

the low-order n_g. The n_g estimation gives information on n_g that should easily increase the range of n_{eff} a few times when we use the low-order MZI. Therefore, the two-MZI pair can be tolerant to process variation much larger than $\pm$ 20 width variation and $\pm$ 10 thickness variation, which makes the design very robust. This can be useful when we do not have a priori knowledge about the technology. The design is sufficiently tolerant that it can be extended for 'blind' first iterations.

2.6 Measurements and Results

2.6.1 Laser Calibration and Stability Test

In parameter extraction experiments using optical measurements, the value of the parameter should be only determined by the circuit layout and device geometry. However, an erroneous wavelength shift in the spectral measurement would lead to a serious error in the extraction of a parameter such as the effective index and group index. Problems such as laser drift and instability may rise such a measurement error.

The long-term usage of a tunable laser is likely to cause wavelength drift when it is lack of a good wavelength calibration. The drift is usually a linear transformation of sweeping wavelength that might be led by the aging of the mechanics. As a result, we would observe as wavelength shifting and broadening. Such an issue is not that noticeable if absolute value peak wavelength or free spectral range is not the focus. However, it will cause significant errors in our measurements. Therefore, laser calibration before and after the wafer-level or die-level measurement is very important.

Meanwhile, we also need to make sure that the laser is stable throughout the measurement. Month-long wafer-scale measurement might suffer from the bad thermal control or unstable mechanics of the wavelength tuning. In the end, it might lead to a variation of the swept wavelength during the lengthy measurement. Therefore, thre should also be a stability test in the measurement workflow. We describe the procedure of the laser calibration and the stability test is in Appendix A.

2.6.2 Extraction of Die Maps

We automated the optical measurements on 21 dies on the same wafer. The optical measurement was conducted in our clean room with the room temperature controlled at 20 degree Celsius.

On each die, we distributed 44 copies of the MZI pair (Fig. 2.9 left bottom) and repeated the fitting for all MZI blocks. With the estimated process variations on different levels, we set up the bound for n_{eff} and follow the three-step procedure to extract n_{eff} and n_g of high-order MZIs (Fig. 2.5 right). Each point in the scatter

plot represents extracted n_{eff} and n_g of one waveguide on the die. All the points are gathered in one group as confined by the bound. The average fitting uncertainty for n_{eff} is 1.1×10^{-5} and the average fitting uncertainty for n_g is 1.0×10^{-3}. These fitting uncertainties propagate to fitting uncertainties of 0.33 nm in width and 0.18 nm in thickness. Adding the mapping error of the geometry model in section 2.2, the total extraction uncertainty and error for width w and thickness t are:

$$Error_w = 0.31 + 0.06 = 0.37nm$$
$$Error_t = 0.18 + 0.08 = 0.26nm$$

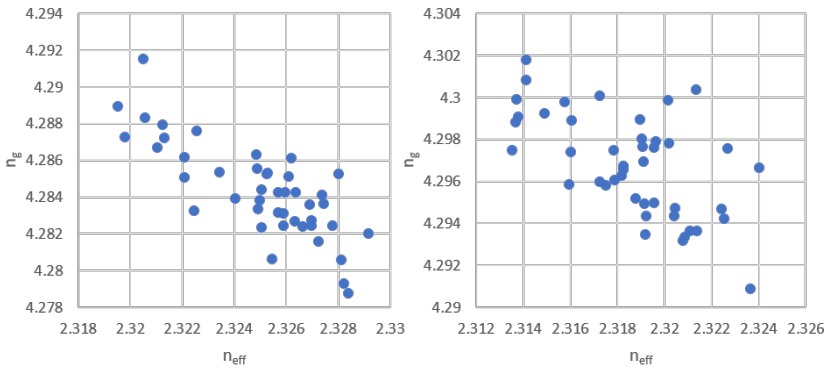


Figure 2.10: Extracted n_{eff} and n_g of the high-order MZI. Left: die (X=0, Y=0); Right: die (X=-2, Y=2).

In reality, a fabricated waveguide has varied width and thickness along its length. Therefore, the width and thickness we extracted are only averaged values over the waveguide. Even more, n_{eff} and n_g we extracted reflects the difference of the optical lengths of two waveguides, which are given by the integral of the effective index and the physical length. So, even we can use our technique to extract width and thickness with sub-nanometer accuracy, these values only reflect the averaged values of waveguides in the MZI circuit.

Using the geometry model, we mapped n_{eff} and n_g to width w and thickness t of the high-order MZI arms. Extracted linewidth on the die (X=0, Y=0) in the wafer center (Fig. 2.9 right) ranges from 468.8 nm to 471.9 nm and thickness ranges from 211.4 nm to 212.3 nm. The standard deviations are 1.26 nm and 0.30 nm respectively. Extracted linewidth on the die (X=-2, Y=2) near the boundary of the wafer ranges from 461.4 nm to 466.8 nm and thickness ranges from 212.1 nm to 214.0 nm. The standard deviations are 1.26 nm and 0.30 nm respectively. For both dies, we observed a very weak correlation (correlation coefficient = -0.2856) between the linewidth and the thickness. We fitted width and thickness to

its location on the die with a linear model, and the green grid is the fitted map (Fig. 2.11) that indicates the location-dependent variation. We did observe a systematic trend for width w on this die, but the trend is quite flat on each die with a systematic variation of less than 1 nm. Meanwhile, the width w shows an obvious local variation with a maximum of 3 nm. For thickness t, both dies exhibits location-dependency which might be the result of slow varying systematic variation over the wafer. Local variation has a maximum of 0.5 nm, well below the local variation range we set in the extraction. The maximum location-dependent difference in thickness on the die (X=0, Y=0) is 0.4 nm while on the die (X=-2, Y=2) is 1.5 nm.

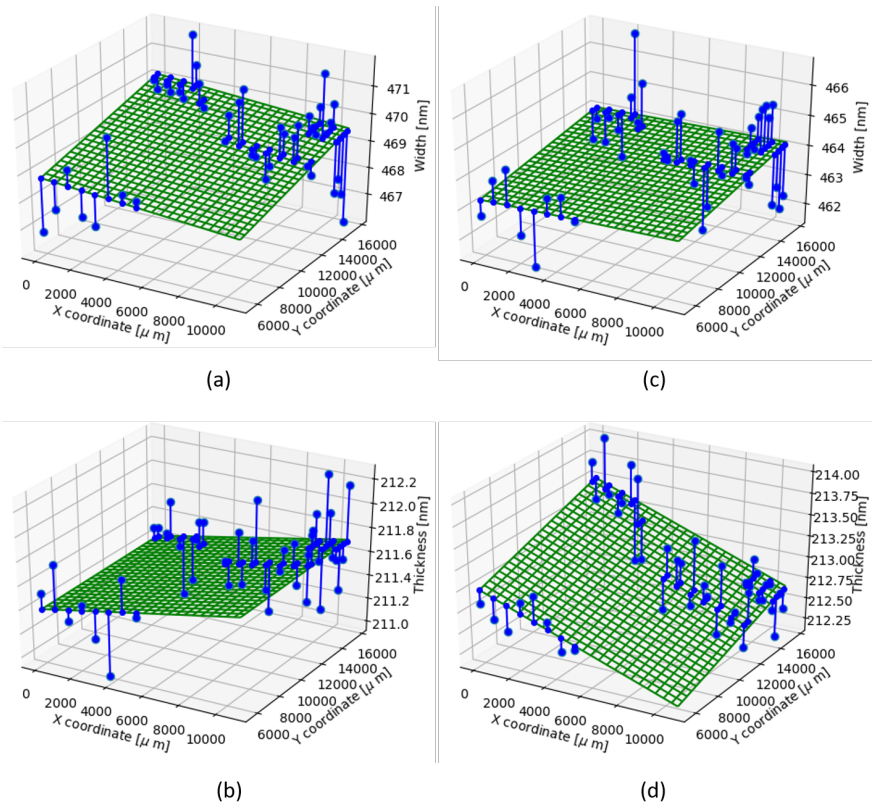


Figure 2.11: X and Y coordinates give the locations of the MZIs on two dies. Blue solid dot: extracted value. Green grid: fitted map of extracted values using a linear polynomial. (a) extracted width map of die (X=0, Y=0) (in the center of the wafer). (b) extracted thickness map of die (X=0, Y=0). (c) extracted width map of die (X=-2, Y=2) (near the edge of the wafer). (d) extracted thickness map of die (X=-2, Y=2).

	Width w [nm]		Thickness t [nm]	
Die number	(X=0, Y=0)	(X=-2, Y=2)	(X=0, Y=0)	(X=-2, Y=2)
Mean, μ	469.34	463.88	211.58	212.86
Standard deviation, σ	1.26	1.18	0.30	0.42
Extraction Error	0.37		0.26	

Table 2.4: Statistical results for the manufacturing variations of a 200-mm wafer fabricated through a 193-nm DUV lithography process

2.6.3 Extraction of a Simple Wafer Map

We also extracted width w and thickness t wafer map of one pair of MZIs that shares the same location on every die. The dots represent the locations of the measured devices. The extracted wafer map (Fig. 2.12) shows an explicit location dependence of fabricated geometry. We fit a parameter wafer map of w, t using a second-order bivariate polynomial. The slow-varying trend of linewidth matches the dome-like radial symmetric pattern of the wafer-level systematic variation. The width systematic variation ranges from 459 to 465 nm while the random part has a maximum 2 nm contribution. The thickness also present a strong location dependence. Its systematic variation ranges from 211.5 to 214.5 nm. Its random variation has a maximum of less than 1 nm.

2.7 Conclusion

In this chapter, we showed how to extract waveguide geometry from optical transmission measurement with a sub-nanometer accuracy. We replaced the linear mapping model between (w, t) and (n_{eff}, n_g) with an accurate third-order geometry model to obtain accurate waveguide geometry from its effective index and group index. The curve fitting method is less sensitive to measurement noise and helps in removing grating coupler envelope. We discussed how to set parameter bounds under a given process variation, which helps to choose the correct set of extracted parameter values from multiple solutions. With the information of the group index of a waveguide, we can reduce the parameter bounds for the effective index, allowing us to use a higher-order MZI to improve fitting accuracy. We proposed a procedure to separate different levels of process variation so that our method can deal with a total variation of several tens of nanometers and still obtain accurate linewidth and thickness extraction. We applied the method to measurement data from two dies and presented the linewidth and thickness map on die-level. We also applied the method to extract one pair of MZIs in 21 dies and presented a simple wafer map of fabricated geometry.

One work we need to do in the future is to valid extracted geometries with FIB measurements. One initial validation is in [3] where he also used MZIs to

extract waveguide effective and group index. He measured waveguide with 470, 604 and 805 nm width and 211 nm thickness with XSEM measurements. He has found very good matching between measured and simulated effective and group index over the C-band, which proves that we can get geometry parameters with confidence from spectral measurements of MZI circuits.

Extracting waveguide width and thickness helps to analyze the process variations on a silicon photonics chip. Often, we also need to know the fabricated quality of directional couplers to analyze the process variation of couplers and to estimate the impact of variations on the performance of the optical filters. Next chapter, we will discuss how to build the behavioral model of the directional coupler and how to extract coupler parameters accurately from optical measurements.

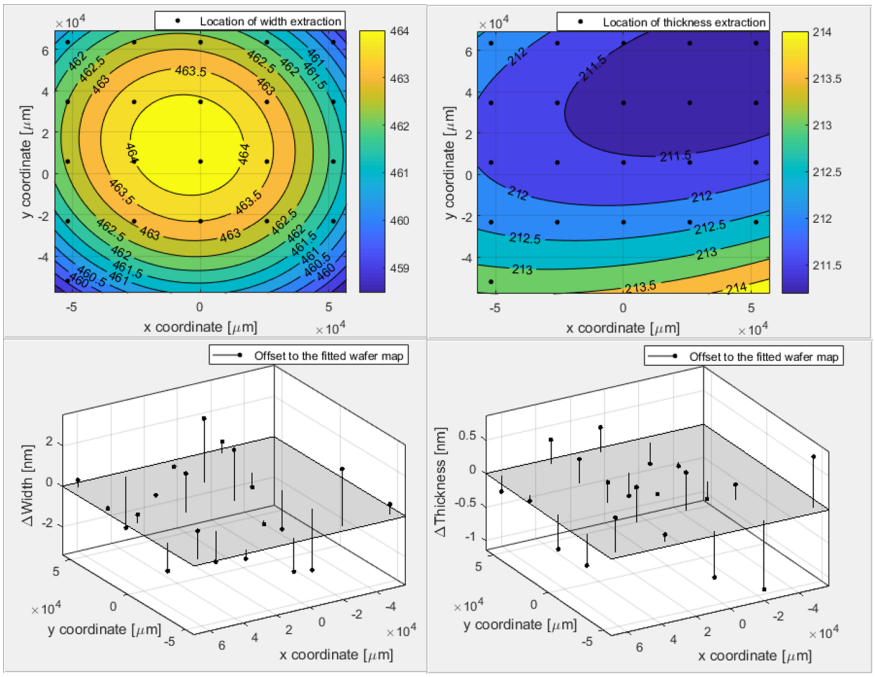


Figure 2.12: We extracted the linewidth and thickness on the same device over 21 dies on the wafer. Left top: systematic linewidth variation; Left bottom: random linewidth variation; Right top: systematic thickness variation; Right bottom: random thickness variation

References

[1] Shibnath Pathak, Dries Van Thourhout, Wim Bogaerts, R $Δ$$θ$ $Δ$$λ$, and R $\Delta\theta$ $\Delta\lambda$. *Design trade-offs for silicon-on-insulator-based AWGs for (de)multiplexer applications*. Optics Letters, 38(16):2961, 8 2013.

[2] Antonio Ribeiro, Sarvagya Dwivedi, and Wim Bogaerts. *A Thermally Tunable but Athermal Silicon MZI Filter*. In 18th European Conference in Integrated Optics 2016 (ECIO), pages 0–25, Warsaw, Poland, 2016.

[3] Sarvagya Dwivedi, Herbert D'heer, Wim Bogaerts, Herbert D 'heer, Wim Bogaerts, Herbert D'heer, and Wim Bogaerts. *Maximizing fabrication and thermal tolerances of all-silicon fir wavelength filters*. IEEE Photonics Technology Letters, 27(8):871–874, 4 2015.

[4] William a Zortman, Douglas C Trotter, and Michael R Watts. *Silicon photonics manufacturing*. Optics Express, 18(23):23598, 2010.

[5] Zeqin Lu, Jaspreet Jhoja, Jackson Klein, Xu Wang, Amy Liu, Jonas Flueckiger, James Pond, and Lukas Chrostowski. *Performance prediction for silicon photonics integrated circuits with layout-dependent correlated manufacturing variability*. Optics Express, 25(9):9712, 2017.

[6] Wim Bogaerts, Martin Fiers, and Pieter Dumon. *Design Challenges in Silicon Photonics*. IEEE Journal of Selected Topics in Quantum Electronics, 20(4):1–8, 2014.

[7] Nicolas Ayotte, Alexandre D. Simard, and Sophie Larochelle. *Long integrated bragg gratings for SoI wafer metrology*. IEEE Photonics Technology Letters, 27(7):755–758, 2015.

[8] Xu Wang, Wei Shi, Han Yun, Samantha Grist, Nicolas A. F. Jaeger, and Lukas Chrostowski. *Narrow-band waveguide Bragg gratings on SOI wafers with CMOS-compatible fabrication process*. Optics Express, 20(14):15547, 2012.

[9] Tsuyoshi Horikawa, Daisuke Shimura, Hiroyuki Takahashi, Jun Ushida, Yohei Sobu, Akemi Shiina, Masatoshi Tokushima, Seok-Hwan Jeong, Keizo Kinoshita, and Tohru Mogami. *Extraction of SOI Thickness Deviation based on Resonant Wavelength Analysis for Silicon Photonics Devices*. 2017 IEEE SOI-3D-Subthreshold Microelectronics Technology Unified Conference (S3S), (1):3–5, 10 2017.

[10] Wim Bogaerts and Lukas Chrostowski. *Silicon Photonics Circuit Design: Methods, Tools and Challenges*. Laser & Photonics Reviews, 12(4):1700237, 4 2018.

[11] Shankar Kumar S.K. Selvaraja, Wim Bogaerts, Pieter Dumon, Dries Van Thourhout, and Roel Baets. *Subnanometer Linewidth Uniformity in Silicon Nanophotonic Waveguide Devices Using CMOS Fabrication Technology*. IEEE Journal of Selected Topics in Quantum Electronics, 16(1):316–324, 1 2010.

[12] Shankar Kumar Selvaraja, Erik Rosseel, Luis Fernandez, Martin Tabat, Wim Bogaerts, John Hautala, Philippe Absil, Shankar Kumar Selvaraja, Erik Rosseel, Luis Fernandez, Martin Tabat, Wim Bogaerts, John Hautala, Philippe Absil, Shankar Kumar Selvaraja, Erik Rosseel, Luis Fernandez, Martin Tabat, Wim Bogaerts, John Hautala, Philippe Absil, Shankar Kumar Selvaraja, Erik Rosseel, Luis Fernandez, Martin Tabat, Wim Bogaerts, John Hautala, and Philippe Absil. *SOI thickness uniformity improvement using corrective etching for silicon nano-photonic device*. In IEEE International Conference on Group IV Photonics GFP, pages 71–73. IEEE, 9 2011.

[13] Shankar Kumar Selvaraja. *Wafer-Scale Fabrication Technology for Silicon Photonic Integrated Circuits*. PhD thesis, University of Gent, 2011.

[14] Raymond G. Beausoleil, Andrei Faraon, David Fattal, Marco Fiorentino, Zhen Peng, and Charles Santori. *Devices and architectures for large-scale integrated silicon photonics circuits*. In Louay A. Eldada and El-Hang Lee, editors, Optoelectronic Integrated Circuits XIII, Proc. of SPIE, volume 7942, page 794204. International Society for Optics and Photonics, 2 2011.

[15] Xi Chen, Moustafa Mohamed, Zheng Li, Li Shang, and Alan R. Mickelson. *Process variation in silicon photonic devices*. Applied Optics, 52(31):7638, 11 2013.

[16] Shankar K. Selvaraja, Gustaf Winroth, Sabrina Locorotondo, Gayle Murdoch, Alexey Milenin, Christie Delvaux, Patrick Ong, Shibnath Pathak, Weiqiang Xie, Gunther Sterckx, Guy Lepage, Dries Van Thourhout, Wim Bogaerts, Joris Van Campenhout, and Philippe Absil. *193Nm Immersion Lithography for High-Performance Silicon Photonic Circuits*. In 27th Optical Microlithography Conference as part of the SPIE Advanced Lithography Symposium, volume 9052, page 90520F, 2014.

[17] Ang Li, Thomas Van Vaerenbergh, Peter De Heyn, Peter Bienstman, and Wim Bogaerts. *Backscattering in silicon microring resonators: a quantitative analysis*. Laser & Photonics Reviews, 10(3):420–431, 5 2016.

[18] Martin Fiers, Thomas Van Vaerenbergh, Joni Dambre, and Peter Bienstman. *CAPHE: Time-domain and Frequency-domain Modeling of Nonlinear Optical Components*. Advanced Photonics Congress, page IM2B.3, 2012.

[19] *scipy.optimize.curve_fit SciPy v1.3.2 Reference Guide*.

[20] Shankar Kumar Selvaraja, Erik Rosseel, Luis Fernandez, Martin Tabat, Wim Bogaerts, John Hautala, and Philippe Absil. *SOI thickness uniformity improvement using wafer-scale corrective etching for silicon nano-photonic device*. In The 2011 Annual Symposium of the IEEE Photonics Benelux Chapter, page 289292, 2011.

[21] Dan-Xia Xu, Jens H. Schmid, Graham T. Reed, Goran Z. Mashanovich, David J. Thomson, Milos Nedeljkovic, Xia Chen, Dries Van Thourhout, Shahram Keyvaninia, and Shankar K. Selvaraja. *Silicon Photonic Integration PlatformHave We Found the Sweet Spot?* IEEE Journal of Selected Topics in Quantum Electronics, 20(4):189–205, 7 2014.

3

Compact Model and Parameter Extraction of A Directional Coupler

A high-level integration of photonics requires a separation of components design and circuit design. A behavioral model describes the characteristics of a component accurately with a minimal set of parameters. It simplifies the circuit design workflow and facilitates parameter extraction and performance evaluation of components and circuits.

In this chapter, we will describe the behavioral model of a directional coupler, one of the essential building blocks in passive photonic integrated circuits. We validated the model with FDTD simulations and on-chip measurements. Extracting coupling coefficients is not trivial in the presence of measurement noise, alignment error, and grating coupler variation. We compared different designs and methods to extract behavioral parameters of the directional coupler. Using a curve fitting method and good grating coupler model, we can extract parameters of the directional coupler with high accuracy from a Mach-Zehnder interferometer circuit.

3.1 Behavioral Model of the Directional Coupler

3.1.1 Directional Coupler

A *directional coupler* (DC) is one of the fundamental building blocks in photonics circuits. It is a four-port 2×2 reciprocal passive optical component used for splitting and combining optical power. A DC can be easily designed to have an

arbitrary coupling ratio. It is interesting for a ring resonator where a low coupling ratio is needed to achieve a high resonator quality factor. It is also interesting for a feed-forward filter design such as a lattice filter where several coupling ratios are synthesized to achieve an optimal filter shape. The arbitrary coupling ratio is otherwise quite difficult to design using other types of couplers. Compared to a *multimode interference* (MMI) coupler, a DC on *Silicon-On-Insulator* (SOI) is also significantly more compact. A 50-50 DC on SOI is normally around 10 μm while a 2×2 MMI can stretch more than 100 μm. A DC is also waveguide-based and has less back-reflection than MMI and Y-junction like splitters. An additional benefit of directional couplers compared to the MMI is that it operates based on two modes (even and odd supermodes). We can engineer it to have zero loss. If you can get a lossless coupling from the waveguide modes to the even or odd modes, there is no further intrinsic loss mechanism in the device for any coupling ratio. These merits make DC frequently used in ring resonators, *Mach-Zehnder interferometers* (MZI), and optical filter circuits.

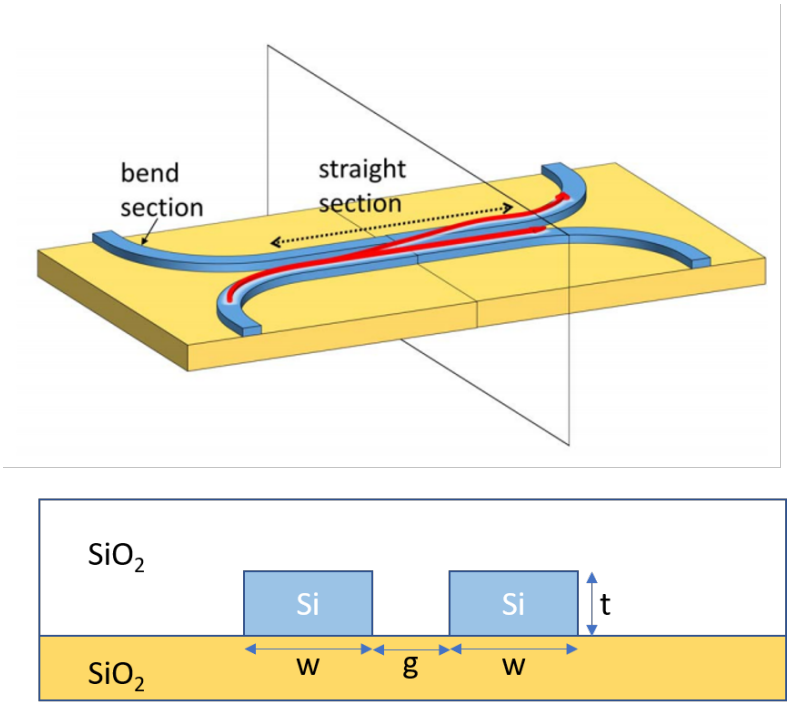


Figure 3.1: The upper plot shows the perspective view of a symmetric DC. Red arrows present the flow of light. Part of the light is coupled from bottom waveguide to the above one. The cross-section is amplified in the lower plot.

A DC consists of two waveguides with an identical cross-section (Figure 3.1).

They are brought so closed to each other that light starts coupling evanescently. The evanescent coupling is determined by the cross-section geometrical parameters. Waveguide with a large width or thickness confines light more, and less light will be coupled to another waveguide. A narrow gap between waveguides increases an overlap between the evanescent field and the adjacent waveguide, which couples light between waveguide faster. The power is gradually transferred from one waveguide to another along the propagation so that the power coupling is coupler length dependent. Meanwhile, apart from the coupling in the straight section, there is also a small couping in the bends that adds to the total coupling.

3.1.2 The Behavioral Model of the Directional Coupler

To describe the characteristics of DC, we need to build a behavioral model for it. Photonics would benefit from a higher level of integration that means not only an increasing complexity but also functionality. Design for high-level integration requires a good separation of component design and circuit design. The behavioral model is meant to simplify the design workflow for a circuit designer. Accurate design for photonic building blocks requires computationally expensive, full-vectorial electromagnetic simulations (e.g., FDTD simulations). To design complex circuits, we need efficient behavioral component models, as FDTD simulations for complex circuits are impossible. If we can use a few parameters to capture the behavior of every component in the circuit sufficiently accurate, then a designer does not have to repeat time-consuming electromagnetic simulations of a component. Instead, one can focus on circuit design and still rely on the overall simulations to be accurate.

The behavioral model is also useful for performance evaluation and process control monitoring that we will discuss in Chapter 4. For a component like a DC, that is sensitive to the component geometry changes, a process variation can lead to a significant performance variation. The variation in each DC accumulates in a circuit, which deteriorates the circuit performance. For example, it can lower the *extinction ratio* (ER) and increase the crosstalk in optical filters. Therefore, in many cases, we want to monitor the fabricated quality of DC to reflect the performance of DC-based optical filters. Also, the statistics of DC behavioral parameters helps to estimate the yield of DC-based circuits and the compensation required to tune circuits to get a desired optical response. To monitor such parameter statistics, we need to build a behavioral model of DC.

A behavioral model should use a minimal set of parameters that captures the behavior of the component. More parameters are not only unnecessary but can also include over-fitting of the behavior, which makes it likely to extract incorrect parameters from a measurement. According to the Coupled Mode Theory (CMT) [1], a DC with two identical parallel waveguides has an odd super mode and an

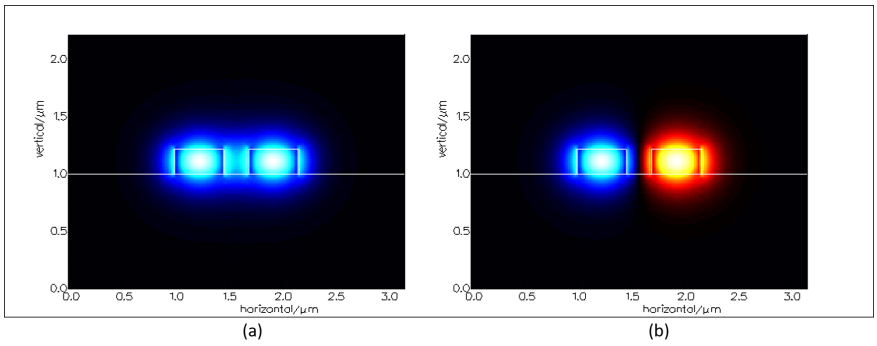


Figure 3.2: Mode profile of the TE electric field of (a) even and (b) odd supermode.

even super mode (Figure 3.2).

The coupling between them brings power K_{cross} in one waveguide to another as:

$$K_{cross} = A \cdot sin^2(\kappa' L), \tag{3.1}$$

where L is the length of the coupler, κ' is the field coupling coefficient is the amplitude coupling per unit length that determines the strength of the coupling in the straight coupling section. The field coupling coefficient depends on the difference between the effective indices of the odd (n_{odd}) and even (n_{even}) modes as:

$$\kappa' = \frac{\pi}{\lambda}(n_{odd} - n_{even}) \tag{3.2}$$

As in Figure 3.3, a DC consists of not only a straight section but a couple of bend sections as well. Both contribute to the overall coupling of the component. So, the model should consider the power coupling in these bends sections, too. Additional contributions from these bends result in the expression of the power coupling K_{cross} that can be expressed as:

$$K_{cross} = A \cdot sin^2(\kappa' L + \kappa_0) \tag{3.3}$$

where κ' is the field coupling coefficient of the straight DC section, and κ_0 is the lumped power coupling contributed by two bends. The coupling in a DC is dispersive. The light of longer wavelength is less confined in the waveguide and has a faster coupling in the DC. We expand such wavelength-dependency in κ' and κ_0 into a polynomial series as:

$$\kappa'(\lambda) = \kappa'|_{\lambda=\lambda_0} + \frac{d\kappa'}{d\lambda}(\lambda - \lambda_0) + \frac{1}{2}\frac{d^2\kappa'}{d\lambda^2}(\lambda - \lambda_0)^2 \tag{3.4}$$

$$\kappa_0(\lambda) = \kappa_0|_{\lambda=\lambda_0} + \frac{d\kappa_0}{d\lambda}(\lambda - \lambda_0) + \frac{1}{2}\frac{d^2\kappa_0}{d\lambda^2}(\lambda - \lambda_0)^2 \tag{3.5}$$

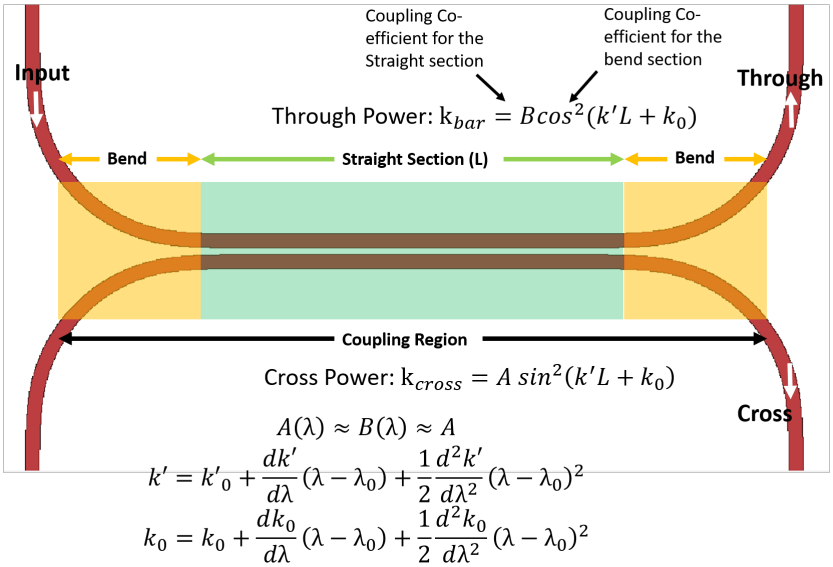


Figure 3.3: Layout of a DC consists of parallel straight waveguides with coupler length of L and two bends. The behavioral model separate contribution of straight section from two bends. Also, the dispersion is considered in the coupling coefficients.

For this work, we use second-order polynomials and neglect higher orders because we found the higher-order terms to be very small both in simulation and experiment. However, the model can easily be extended to more dispersive devices. Substituting it into the equation above, the power at the 'cross' port becomes:

$$K_{cross}(\lambda) = A \cdot sin^2(\kappa'(\lambda) \cdot L + \kappa_0(\lambda)) \tag{3.6}$$

The power at the 'through' port becomes:

$$K_{through}(\lambda) = B \cdot cos^2(\kappa'(\lambda) \cdot L + \kappa_0(\lambda)) \tag{3.7}$$

A and B can also be length-dependent since loss becomes larger with longer length. We ignore this aspect here for the moment because the length of the DC that we discussed in the chapter is very short. We just assume that $A = B = 1$.

3.1.3 Verification of the Behavioral Model

In this section, we will verify the behavioral model with FDTD simulations and on-chip measurements.

3.1.3.1 Verification from FDTD simulations

First, we verified the model by comparing it to the three dimensional (3D) FDTD simulations. We fixed the cross-section of the DC (450 nm $\times$ 220 nm oxide-clad silicon waveguides with a 250 nm gap) and used the same bend radius for all DCs. We swept the coupler length in the Lumerical FDTD simulation and obtained the scattering matrix of the device, from which we have the cross-coupling power transmission. The cross-section and the bend radii were same, so all the devices had the same field coupling $\kappa'(\lambda)$ and bend coupling $\kappa_0(\lambda)$ coefficients. According to Eq. 3.3, the cross-coupling power K_{cross} follows a sinusoidal relation with the coupler length (L). We obtained the bend coupling coefficient κ_0 from the zero coupler length DC. The fitting is more accurate if we choose at least one length value residing on the right side of the maximum power coupling. Then, we have samples to cover more than one fourth of a cycle of the sinusoidal curve. As shown in Figure 3.4 (a), $\kappa'(\lambda)$ and $\kappa_0(\lambda)$ are derived by fitting the K_{cross} vs L graph at 1550 nm, where we observed a very good fitting. We plotted the graph for each simulated wavelength and observed very good fitting for all wavelengths. The $\kappa'(\lambda)$ obtained at 1550 nm is 0.042.

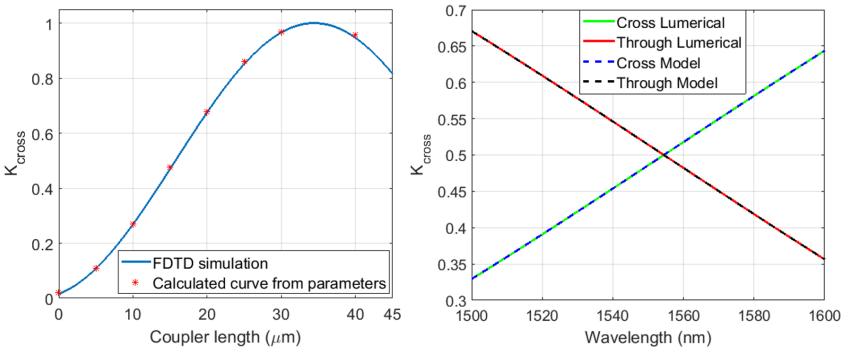


Figure 3.4: (a) The red stars in the graph show the power couplings calculated using the FDTD simulations. The blue plot shows the fitted power couplings using our behavioral model. Such graphs are generated for each wavelength and fitted to extract the values of field coupling $\kappa'(\lambda)$) and bend coupling $\kappa_0(\lambda)$. The plot was calculated for a wavelength of 1550 nm. (b) A comparison showing a perfect match between the DC spectra generated using the commercially available FDTD solver from Lumerical and our dispersive behavioral model.

To confirm the extracted values of $\kappa'(\lambda)$, the cross-section was simulated using the Fimmwave *Film Mode Matching* (FMM) solver. The FMM solver is desirable to simulate rectangular structures because it has no intrinsic grid and discretization, which reduces the discretization error. It is also very fast to calculate the even and odd mode of a longitudinally-invariant section. But it cannot easily capture

sloped sidewalls and bends, so we used it only to simulate modes of rectangular waveguide-based structures. We calculated the difference between the even and odd mode effective indices and obtained that $\kappa' = 0.0427$ at 1550 nm using $\kappa' = \frac{\pi}{\lambda}(n_{odd} - n_{even})$. The value is close to the extracted value of 0.042 extracted from the FDTD simulations.

We derived $\kappa'(\lambda)$ and $\kappa_0(\lambda)$ from FDTD simulations over a wide wavelength range from 1500 to 1600 nm. We obtained κ_0 at one wavelength from the simulation that the coupler length is zero. We used power transmission at one wavelength from eight couplers with different coupler length to plot the K_{cross} vs. L graph and derive κ'. We repeated the process for all wavelength to get $\kappa'(\lambda)$ and $\kappa_0(\lambda)$. Using Eq. 3.6 and 3.7, we found that spectral responses calculated by our model match pretty well to the responses calculated by the FDTD solver from Lumerical (Figure 3.4 (b)). The good matching validated that separating length-dependent straight coupler and lumped bends contributions are accurate to model the coupler. Also, using the derivative up to the second-order is sufficient to describe the dispersion of the coupling.

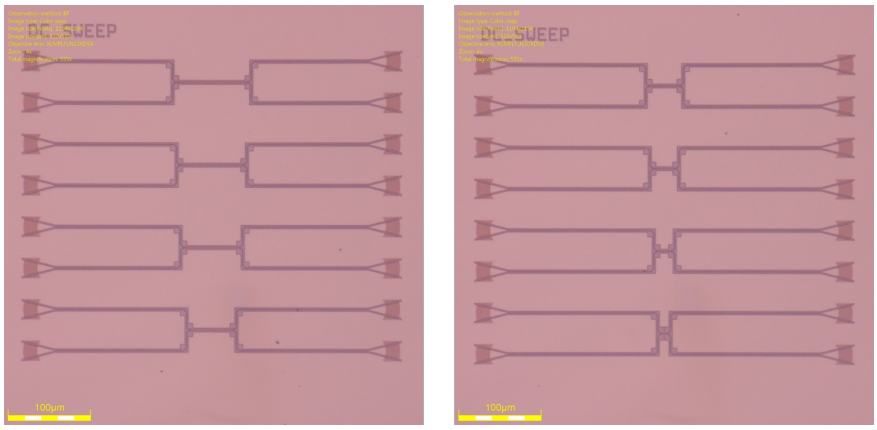


Figure 3.5: The microscopic picture of the 8 DCs of lengths 0.15 μm, 10 μm, 20 μm, 30 μm, 40 μm, 60 μm, 70 μm and 80 μm fabricated using ebeam lithography. DC circuits were fabricated at RMIT Melbourne.

3.1.3.2 Verification from Measurements

In the second part, we checked the model with reality and extracted its parameters from measurements. To validate the model from fabricated devices, 8 DCs (Figure 3.5) of lengths 0.15 μm, 10 μm, 20 μm, 30 μm, 40 μm, 60 μm, 70 μm and 80 μm were fabricated using ebeam lithography through the Australian Silicon Photonics prototyping service at RMIT Melbourne. All DCs had the same designed cross-section ($450nm \times 220nm$) and the same bend sections. Fabricated devices were

air-clad from the top.

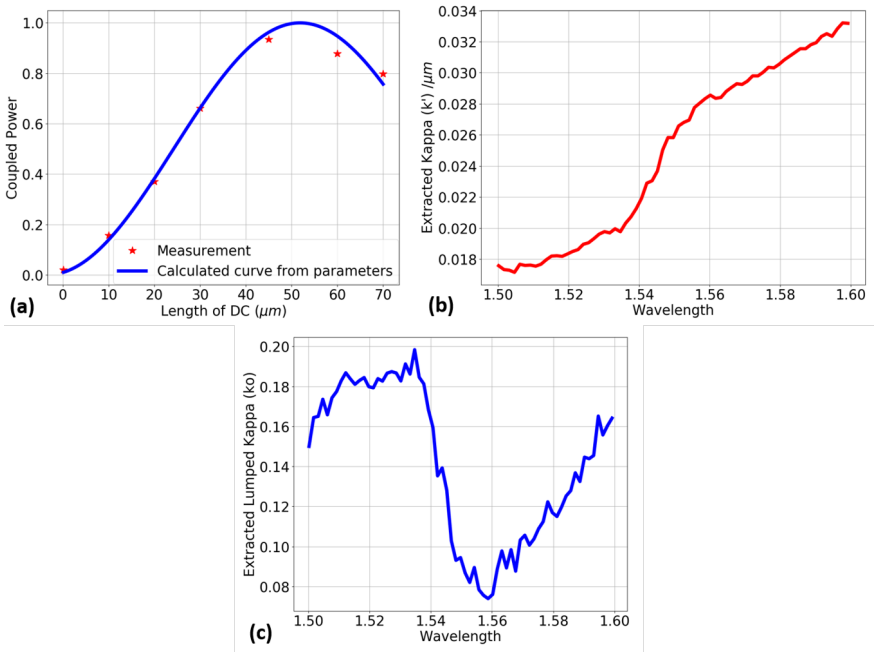


Figure 3.6: (a) The plot shows the measured and the fitted power coupling for different of coupler length L. (b) The extracted field coupling coefficient κ'. (c) The plot shows the extracted lumped coupling coefficient κ_0 of the bend.

We validated the model the same way as we validated the model with FDTD simulations. We used the power coupling at a fixed wavelength for multiple DC lengths and fitted the K_{cross} vs. L graph. The parameters are extracted from the fitting. As in Figure 3.6 (a), we repeated the process for all wavelengths to get the coupling coefficients as shown in Figure 3.6 (b) and (c). The coupling coefficients increase with increasing wavelength as we expected. The extracted values of field coupling are generated for each wavelength and fitted to extract the values of field coupling $\kappa'(\lambda)$) and bend coupling $\kappa_0(\lambda)$. However, they are not varying smoothly with the wavelength because the fitting is performed independently for each wavelength. In this procedure, no effort was made to remove the oscillations in the transmission spectrum from reflection from measurement noise, variations in the *grating couplers* (GC), and interference fringes caused by back-reflections. Still, the extracted $\kappa'(\lambda)$) and $\kappa_0(\lambda)$ from measurement are comparable in magnitude with the results we got from FDTD simulations. However, we need to improve the method to extract accurate and to smooth dispersive coupling parameters to get a more convincing validation, which we will discuss in the following section.

3.2 Methods to Extract Directional Coupler Parameters

After validating the model of DC, we now need a reliable method to extract model parameters from optical measurements for further analysis, which is not trivial. As we discussed in the last section, measurement noise can significantly deteriorate the extraction accuracy of DC parameters. Also, the process variation of grating couplers adds significant error in a normalized 2×2 DC transmission. In the following, we will discuss what a better extraction method and the optimal circuit design to extract DC parameters is.

3.2.1 Extraction from 'Naked' DCs

3.2.1.1 Curve Fitting Method

The first way to extract parameters from 'Naked' DCs is introduced in the last section. Using the power coupling K_{cross} at a fixed wavelength for multiple DC lengths L, we can extract the parameters by the fitting of the K_{cross} vs. L graph. Dispersive coupling is derived by repeating the process for all wavelengths. However, this method is very sensitive to any effect that can introduce a wavelength-dependent device-to-device variation in the transmitted power: grating coupler variation, interferometric oscillations due to parasitic cavities created by back-reflection and detector noise in the spectrum. Besides, to fit the K_{cross} vs. L graph, we need to sweep a sufficient number of coupler length to get a good fit. In our case, we used eight couplers. As explained, at least one coupler length value should reside on the right side of the maximum power coupling to cover more than half a cycle of the sinusoidal curve. A large number of devices requires not only more measurements, but also more space on-chip that leads to an increasing process variation between DCs and increasing modeling error.

Fitting the power coupling for every individual wavelength is not a good strategy, because the wavelengths are treated independently, while we know the coefficients vary only slowly with wavelength. Therefore, an alternative method is to extract parameters by fitting the measured spectrum with the behavioral model. The curve fitting method finds the parameters of the DC model by matching the measured spectrum with the simulated spectrum. The curve fitting method addresses the major drawbacks of the first method. Firstly, there is only one set of solution of DC parameters set that offers the optimal matching between measurement and simulation. Secondly, the dispersive coupling is more tolerant of measurement noise. Rather than measuring the ER at a few wavelengths, the curve fitting uses the information on the entire measured spectrum, and it is much more tolerant to measurement noises. Thirdly, fitting the spectrum with a model smooths the error in the dispersive coupling coefficients. Since the coupling is smooth over wave-

length as also validated by the FDTD simulations, when we fit the entire spectrum with the model, the fitting would average the spectral noise.

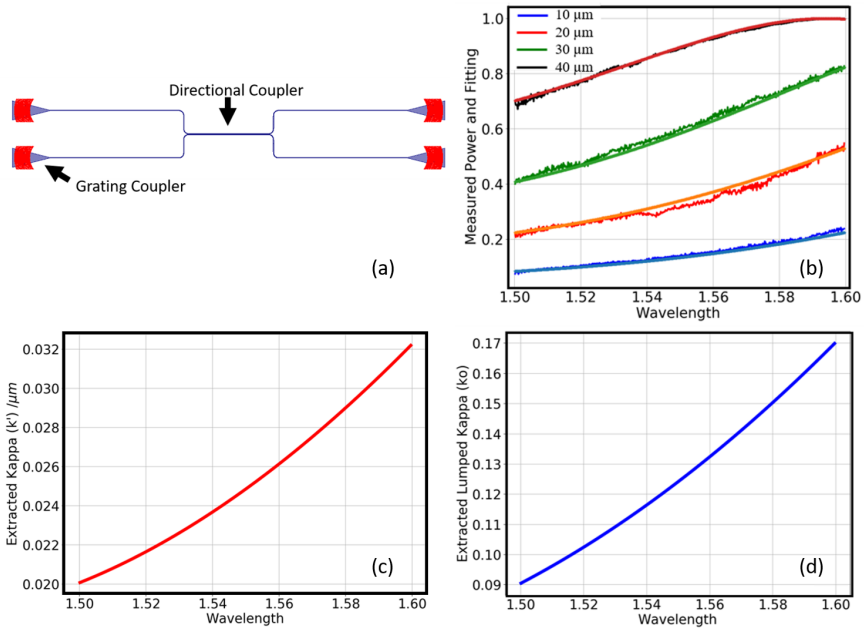


Figure 3.7: (a) Layout and port positions of the DC circuit. (b) Measured and the fitted spectra from the coupled port of the DCs with coupling lengths of 10, 20, 30 and 40 μm. (c) Extracted field coupling coefficient using a 'naked' DC. (d) Extracted lumped coupling coefficient.

To separate $\kappa'(\lambda)$ from $\kappa_0(\lambda)$, the curve fitting method fits the transmission spectrum of a few DCs simultaneously. In principle, we could extract all the parameters just from two DCs of any coupler length (the additional devices make the method more robust), while the first method required at least eight devices to get a decent fit. So, it is more practical and cost-saving. From the same measurements to validate the DC model, we chose DCs of lengths 10 μm, 20 μm, 30 μm and 45 μm for the fitting where the choice of the coupler length is arbitrary. We are not taking into account device-to-device variability, so $\kappa'(\lambda)$) and $\kappa_0(\lambda)$ are assumed identical for these four lengths. It is a much simpler procedure since we can extract all six parameters in a single operation. Measured spectra of four devices were fitted simultaneously to the circuit model shown in Figure 3.7 (c) using the *non-linear least-squares minimization and curve-fitting* (lmfit) tool for Python. It can be noticed that the extracted $\kappa'(\lambda)$ (Figure 3.7 (c)) and $\kappa_0(\lambda)$ (Figure 3.7 (d)) using both methods are comparable with an exception that extracted parameters

vary much more smoothly over the wavelengths using the second approach. The comparable results are a good validation of the behavioral model.

3.2.1.2 Extraction Error Led by the Grating Couplers Variation

In the measurements of the naked DC circuits, we measure the spectrum with the DC connected to GCs on two ends. The transmission includes the GC envelope. We normally would assume that all GCs are identical, so that GC transmission envelopes from 'in' to 'through' and 'in' to 'cross' (Figure 3.7 (a)) are the same. Then, we could remove the GC envelope by normalizing two transmissions from 'through' and 'cross' port:

$$P_{through,norm} = \frac{P_{through,withGC}}{P_{through,withGC} + P_{cross,withGC}}$$
$$P_{cross,norm} = \frac{P_{cross,withGC}}{P_{through,withGC} + P_{cross,withGC}}$$

However, the transmission spectra of the GC are not identical due to process variation and alignment variations during the measurements. In [2], transmission spectra are shown for the optical coupling between a grating coupler and a fiber for 64 grating couplers on a single SOI wafer. The standard deviation of maximum wavelength λ_{max} on a wafer can be 1.9 nm while the maximum coupling C_{max} has a standard deviation of 0.07 dB. This variation leads to the normalization error of the DC transmission spectra.

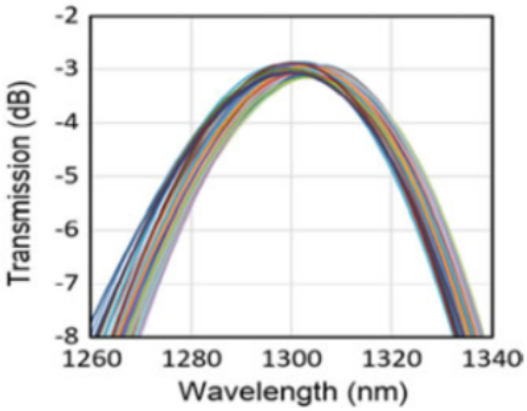


Figure 3.8: The figure is taken from [2]. It shows transmission spectra for the optical coupling between a grating coupler and a fiber for 64 grating couplers on a single SOI wafer.

We can estimate the extraction error led by this GC variation using the circuit simulator Caphe [3]. We design the circuit where the 'naked' DC is connected with two GCs. Since the size of the DC is small, we eliminated and ignored any loss in the directional coupler (which can be wavelength-dependent). We design the same circuit as the one we measured in Fig. 3.7 (a), which consists of three DC circuits with coupler length of 6.65, 12.91 and 19.17 μm. Each DC has the same set of coupling coefficients calculated for a DC cross-section of 450 nm $\times$ 220 nm with a gap of 250 nm. The corresponding coupling coefficients are displayed in Table 3.2.

As mentioned in Section 2.3 of Chapter 2, the transmission spectra from fiber to GC can be very well fitted by a fourth-order polynomial as:

$$P_{GC,logarithmic} = \sum_{i=1}^{4} p_i \cdot (\lambda - \lambda_{max})^i + C_{max} \tag{3.8}$$

Where λ_{max} is the wavelength of maximal coupling and C_{max} is the value of the maximal coupling on logarithmic scale. We assumed that the process variation between two GCs in a circuit is smaller than the variation on a single wafer shown in [2]. In each DC circuit, we assume the difference between maximum wavelengths λ_{max} between two GCs transmission spectra $\Delta\lambda_{max} = \lambda_{max,in-through} - \lambda_{max,in-cross} = 1$ nm. The maximum coupling difference ΔC_{max} is led by the process variation or an alignment error. We assumed $\Delta C_{max} = C_{max,through} - C_{max,cross} = 0.3$ dB between two GCs in the circuits, which is roughly 4σ of the variation reported in [2]. Since the shapes of the GC envelopes are almost identical, we can assume the coefficients p_i in the polynomial are the same for all GCs.

We also add the background amplitude noise in the spectrum to mimic the spectral fringes due to reflection and back-coupling in the circuit, and measurement noise. We set the noise level in the simulated spectrum similar to what we usually measured from the spectral measurement of the naked DC circuit. The noise follows a normal distribution with a standard deviation of 0.2 dB. Then, we could generate spectra that emulate the real measurements where measurement noise, alignment errors in the circuit simulator, and process variation of GCs are taken into account.

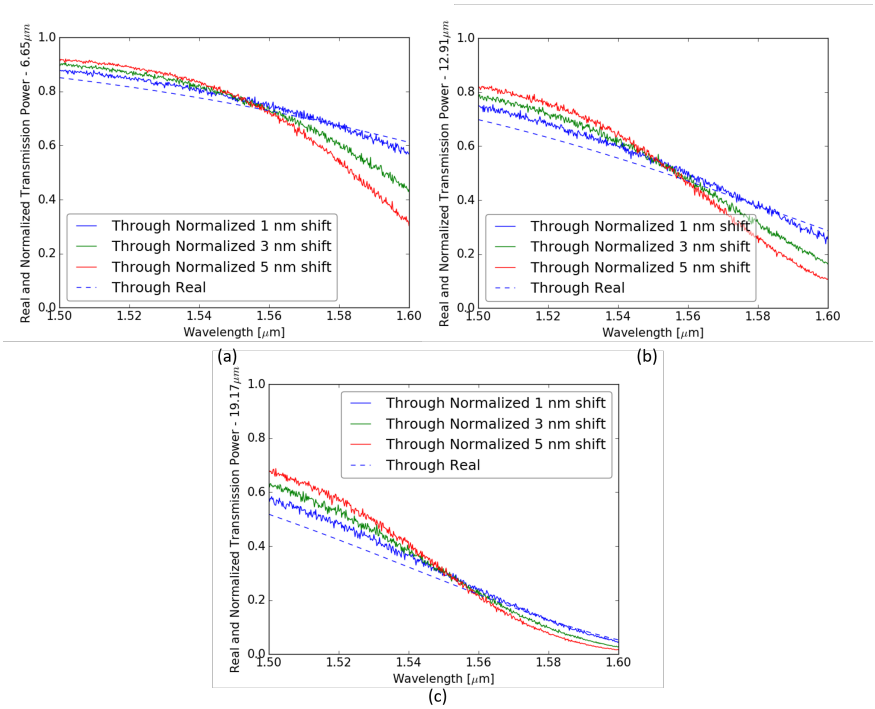


Figure 3.9: Red solid: simulated measurements with variations that mimic the variations you get from a measurement. Blue dash: normalized power transmission at 'through' port. There is a clear mismatch between normalized spectra and actual spectra. (a) L=6.65 μm. (b) L=12.91 μm. (c) L=19.17 μm. The mismatch increases noticeably with the difference of maximum wavelength between two GCs transmission spectra $\Delta\lambda_{max}$. Blue, green and red curves are normalized DC transmissions with GC λ_{max} differing by 1,3 and 5 nm.

From the generated simulated measurements, we normalized the power at cross port K_{cross} using Eq. 3.8. Figure 3.9 shows when the variations of GCs present themselves in the circuit, there is a clear mismatch between actual K_{cross} that generated from circuit simulator without adding GC (blue dash) and the normalized $K_{cross,norm}$ (solid). The mismatch increases with the assumed difference of maximum wavelength between two GCs transmission spectra $\Delta\lambda_{max}$. For $\Delta\lambda_{max} = 1nm$, as shown in Figure 3.10, a good fitting is obtained between normalized spectra and the simulated spectra from fitting. Nonetheless, as shown in Table 3.2, the GC variation leads to a 3.45% relative error in the extracted coupling κ'. So, we need to find a way to reduce the impact of the GC variation and extract DC parameters with high accuracy.

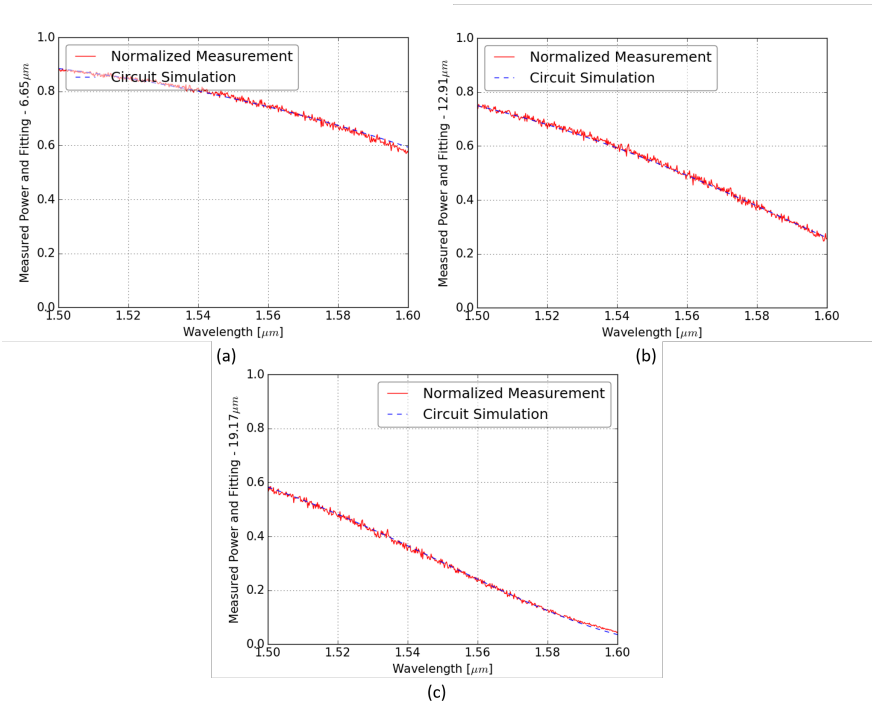


Figure 3.10: Red solid: normalized power transmission at 'through' port. Blue dash: Caphe circuit simulation. (a) L=6.65 μm. (b) L=12.91 μm. (c) L=19.17 μm.

3.2.1.3 Grating Coupler Model: Does It (Always) Work to Remove the GC Envelope?

One way to remove the impact of GC variation is to also include a model for GC in the circuit model. If we can fit the GC parameters with confidence, we know the shape of each GC. Then, we can remove the GC envelope.

To do that, we apply the fourth-order polynomial model of the GC and fit each DC circuit that consists of two GCs and a DC. For the moment, we assume the maximum coupling wavelength λ_{max} and the maximum coupling C_{max} will vary. The shape of the GC transmission envelope will not change, which means that the polynomial coefficients of the GC model are assumed same for both GC envelopes in the circuits. Figure 3.11 shows the result of fitting of the power at 'through' port. Very good fitting is observed for all three DC circuits.

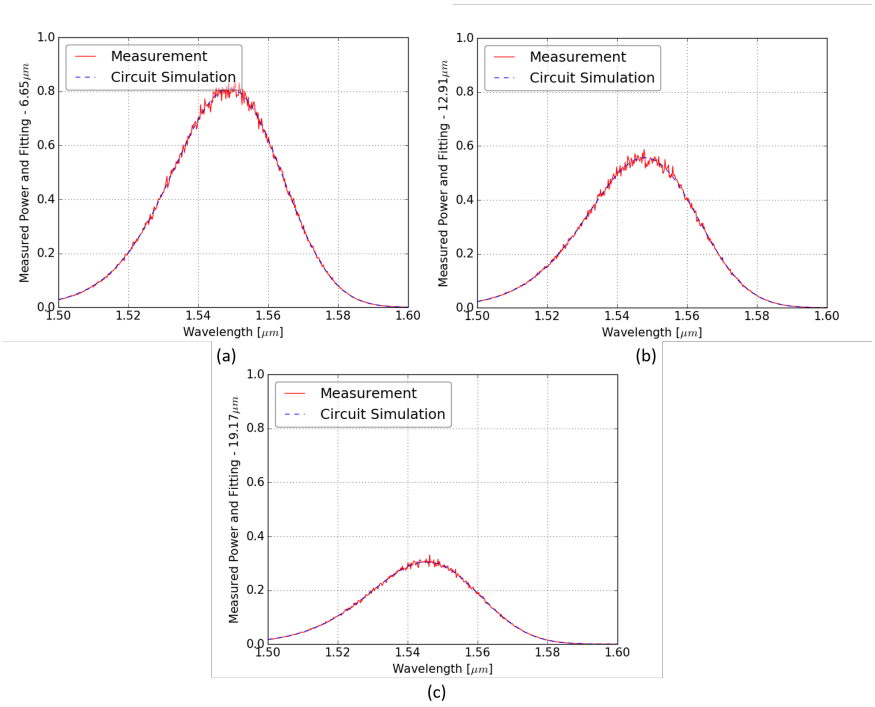


Figure 3.11: Red solid: Simulated measurement generate with GC variation and measurement noise. Blue dash: Caphe circuit simulation that incorporates polynomial GC model. (a) L=6.65 μm. (b) L=12.91 μm. (c) L=19.17 μm.

	$\Delta\lambda_{max}$ [nm]	ΔC_{max}
Values Used to Generate Simulated Measurements	1.0	0.30
DC (L=6.65 μm)	1.1	0.61
DC (L=12.91 μm)	0.1	-0.16
DC (L=19.17 μm)	1.4	0.80

Table 3.1: GC parameters we assumed to generate the simulated measurements and the extracted parameters.

However, as shown in Table 3.1, $\Delta\lambda_{max}$ and ΔC_{max} we extracted are far from the ones we used to mimic the GC variation. So, we did not extract the true GC envelopes from the circuit. The good curve matching and significant mismatch in parameters extracted means it is hard to separate the GC envelope from a 'naked' DC circuit. As the spectrum of DC transmission is gradually changed over wavelength, we can also fit its transmission on log scale by a series of polynomials. Consequently, different combinations of DC parameters and GC parameters might

lead to a similar circuit response. Therefore, it is not possible with the current techniques to remove the GC envelope from such as circuits and get accurate DC parameters. Again, it has been evidenced by the extracted DC parameters in Table 3.2, where there is a big difference between the extracted and the actual values of the coefficients.

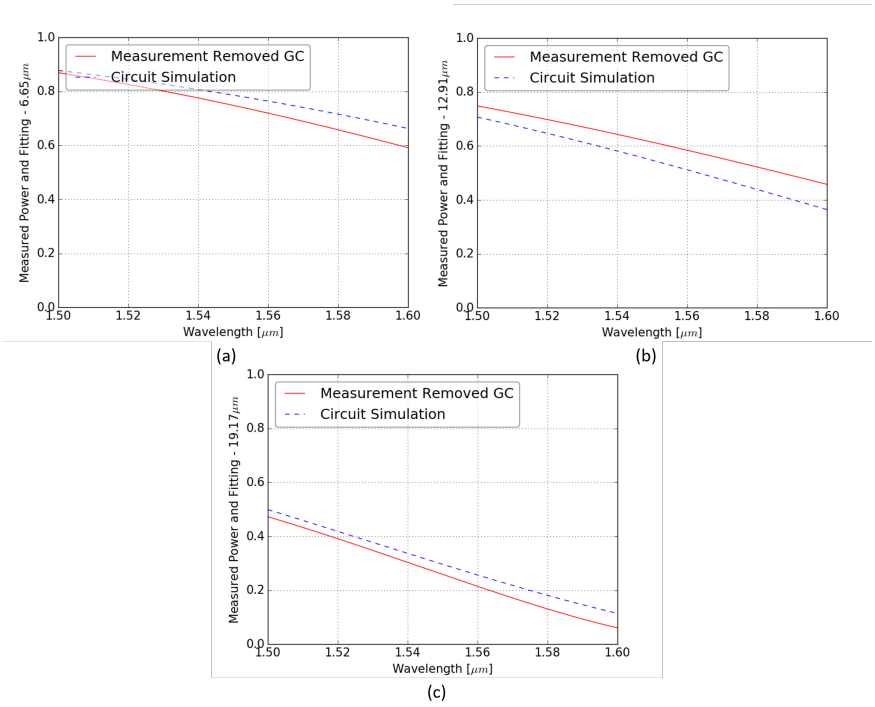


Figure 3.12: Red solid: simulated Measurement generate with GC variation and measurement noise. Blue dash: Caphe circuit simulation that incorporates polynomial GC model. (a) L=6.65 μm. (b) L=12.91 μm. (c) L=19.17 μm.

	Simulated measurements	Remove GC with the model		Remove GC by normalization	
		Values	Relative Error [%]	Values	Relative Error [%]
κ'	0.0406	0.0411	1.23	0.0392	3.45
$\frac{d\kappa'}{d\lambda}$	0.2382	0.1419	40.43	0.2673	12.22
$\frac{d^2\kappa'}{d\lambda^2}$	1.0047	0.1890	81.19	1.9896	98.03
κ_0	0.2548	0.2071	18.72	0.2360	7.38
$\frac{d\kappa_0}{d\lambda}$	1.2107	1.6782	38.61	1.6782	38.61

Table 3.2: Parameters extracted using a GC model and using normalized transmission from the 'naked' DC circuits.

3.2.2 MZI with a 1×2 MMI and a DC: Extraction from the Extinction Ratio

The similarity in the GC and DC transmissions make it hard to remove the impact of GC from a 'naked' DC circuit. Therefore, our idea is to design a circuit whose transmission spectrum differs significantly from GC transmission.

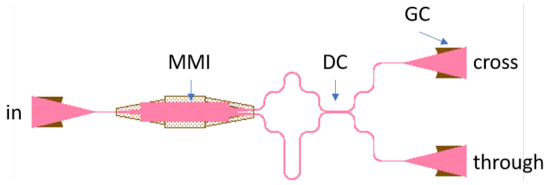


Figure 3.13: Layout of the MZI circuit with two identical DCs.

The first circuit we look at is the DC-based MZI. [4] [1] As in Figure 3.13, the imbalanced MZI consists of a DC splitter and an MMI combiner with 50% combining ratio at 1550 nm. When the phase difference in two arms is the odd times of π, a constructive interference occurs in the MZI output resulting in a maximum; when the phase difference is the even times of π, a destructive interference occurs. When both splitting and combining ratios are 50%, a complete constructive or destructive interference occurs. We need to mention that the phase difference between through and cross port is π. When the splitting or combining ratios are different, we can extract the imbalance from the wavelength dependent ratio between adjacent minima and maxima of the output spectrum. The output field E_{out} is calculated using the matrix formalism:

$$[T_{E_{out}}] = [T_{combiner}] \cdot [T_{interferometer}] \cdot [T_{splitter}] \cdot [T_{E_{in}}] \tag{3.9}$$

$$E_{out} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} e^{-j\phi_1} & 0 \\ 0 & e^{-j\phi_2} \end{bmatrix} \begin{bmatrix} \sqrt{1-\kappa^2} & j\kappa \\ j\kappa & \sqrt{1-\kappa^2} \end{bmatrix} \begin{bmatrix} 0 \\ E_{in} \end{bmatrix} \tag{3.10}$$

In [4], a detailed derivation is given on how to determine the coupling coefficient κ from the extinction ratio between the upper and lower envelope of the normalized intensity (The method is illustrated in Figure 3.14). The DCs are assumed to be lossless. From the wavelength dependent ER, we can determine the wavelength dependent cross power coupling K_{cross}:

$$K_{cross} = \frac{1}{2} \pm \frac{1}{2}\sqrt{1 - \frac{ER-1}{ER+1}} \tag{3.11}$$

[1] This method is also described in: http://photonics.intec.ugent.be/publications/publications.asp?ID=3105

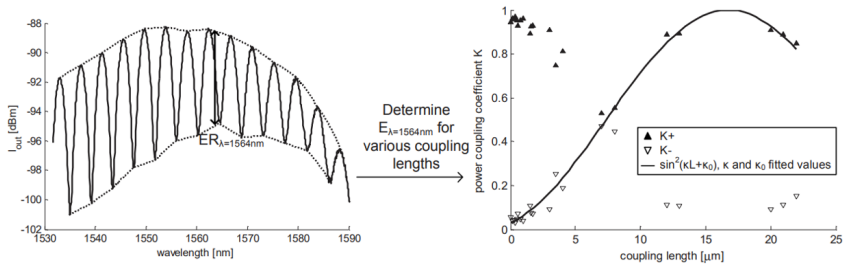


Figure 3.14: The figure is taken from [4]. The figure shows the method to obtain the wavelength dependent coupling coefficient κ' as well as the lumped coupling coefficient κ_0 of a DC from the ER in the output intensity of an MZI.

The design of the circuit makes it possible to extract coupling coefficients from the ER without a transmission normalization. We need to fit or interpolate the upper envelope and the lower envelope from the measured 'cross' transmission with the influence of the GC. K_{cross} can be calculated from the ER, which is the distance between two envelopes. Such approach separates DC related ER features in the spectrum from the envelope of the GC, which also proves that we can separate DC and GC transmission with proper circuit design.

There are a few major drawbacks of this method. First, for each coupler length L, there are two solutions for the cross power coupling K_{cross}. We can only guess the right group of the solution by choosing the samples that are close to the curve $sin^2(\kappa' \cdot L + \kappa_0)$. The guess could be not objective. We can substitute the estimated κ' and κ_0 in the sinusoidal curve. Still, it is hard to pick up the right solution when the K_{cross} is near 50%. Second, an accurate ER depends on the accurate measurement of maxima and minima in the spectrum because the positions of maxima and minima determine the accuracy of the upper and lower envelope. However, the noise and limited resolution of the spectrum all lead to errors in determining maxima and minima locations. This error is particularly large when the coupling ratio is around 50%. Thirdly, the method calculates the power coupling at a few wavelengths and interpolates the coupling over the spectrum. Again, the noise in the spectrum leads to significant errors in coupling coefficients, which can be observed in Figure 3.15.

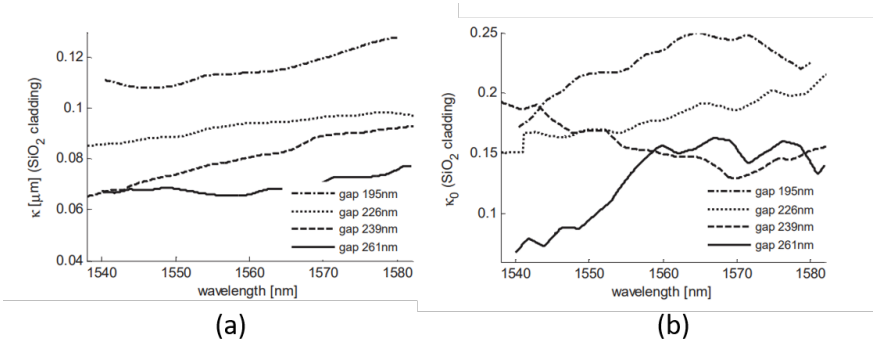


Figure 3.15: The figure is taken from [4]. Measured wavelength dependent coupling coefficient κ' (a) and lumped coupling coefficient κ_0 (b) of a DC versus wavelength. The coefficients are not smooth over wavelength because they are obtained from interpolated upper and lower transmission envelopes.

3.2.3 From MZI with two Identical DCs: Spectrum Curve Fitting with a GC Model

Even though the approach to extract coupling from MZI is not desirable from a few reasons, the circuit itself has the DC and GC contribute differently to the features of the transmission spectra. This section, we will illustrate our experiment to remove GC envelope and improve extraction accuracy using a GC model.

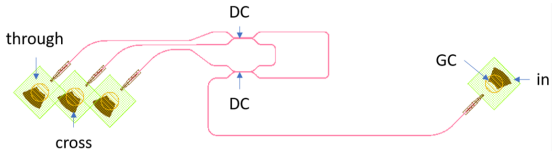


Figure 3.16: Layout of the MZI circuit with two identical DCs.

We design an MZI with two identical DCs (Figure 3.16). Similarly, the experiment consists of three MZIs with DC coupler lengths L of 6.65, 12.91, and 19.17 μm. As assumed in 3.2.1.2, we add measurement noise of 0.3 dB and include GC variations by assuming the difference of maximum wavelength λ_{max} between two GCs transmission spectra is 1.0 nm. As shown in Figure 3.17, we fit the simulated measurement at the 'through' port with a circuit model of the MZI with the polynomial GC model. Excellent matching between simulated measurements (red solid) and circuit simulations (blue dash) presents in the figure. Also, the extracted the $\Delta\lambda_{max}$ [nm] and ΔC_{max} match well with the parameters we used to introduce GC variation. It proves that we can remove the GC envelope using the curve fitting for this design.

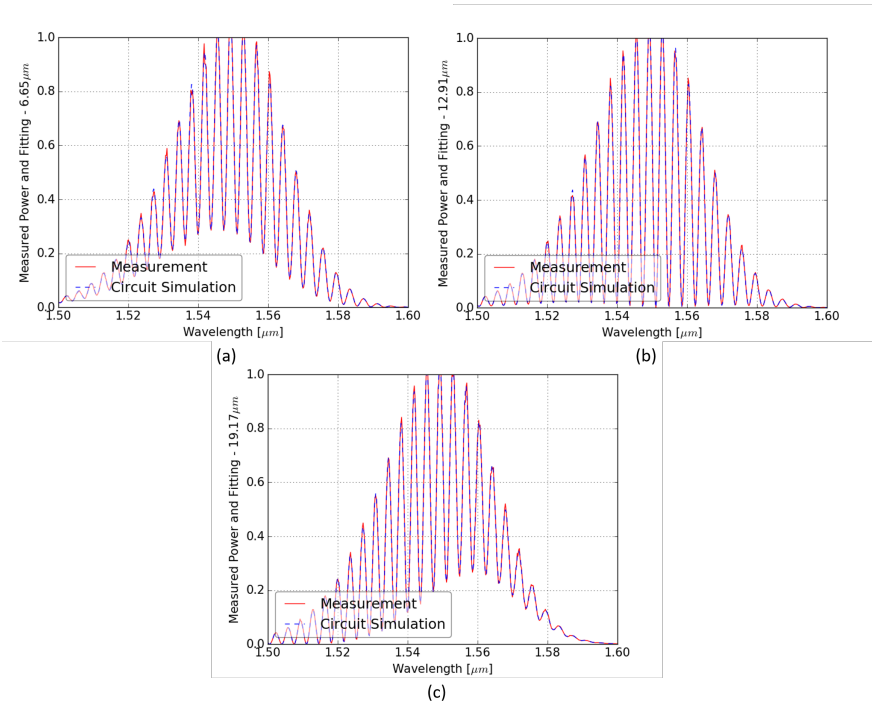


Figure 3.17: Fitting of MZI spectra with the GC model. Red solid: simulated measurement generated with GC variation and measurement noise. Blue dash: Caphe circuit simulation that incorporates polynomial GC model. (a) L=6.65 μm. (b) L=12.91 μm. (c) L=19.17 μm.

After successfully removing the GC envelope, we fitted three 'through' transmission spectra simultaneously with Caphe circuit simulations. Figure 3.18 presents the result of the fitting. Simulations and measurements match excellently. The spectral response of the MZI is sinusoidal that is significantly differed from the GC envelope. So, it is possible to separate such an impact.

	$\Delta\lambda_{max}$ [nm]	ΔC_{max}
Values Used to Generate Simulated Measurements	1.0	0.30
DC (L=6.65 μm)	1.0	0.30
DC (L=12.91 μm)	1.0	0.29
DC (L=19.17 μm)	1.0	0.31

Table 3.3: GC parameters we assumed to generate the simulated measurements and the extracted parameters.

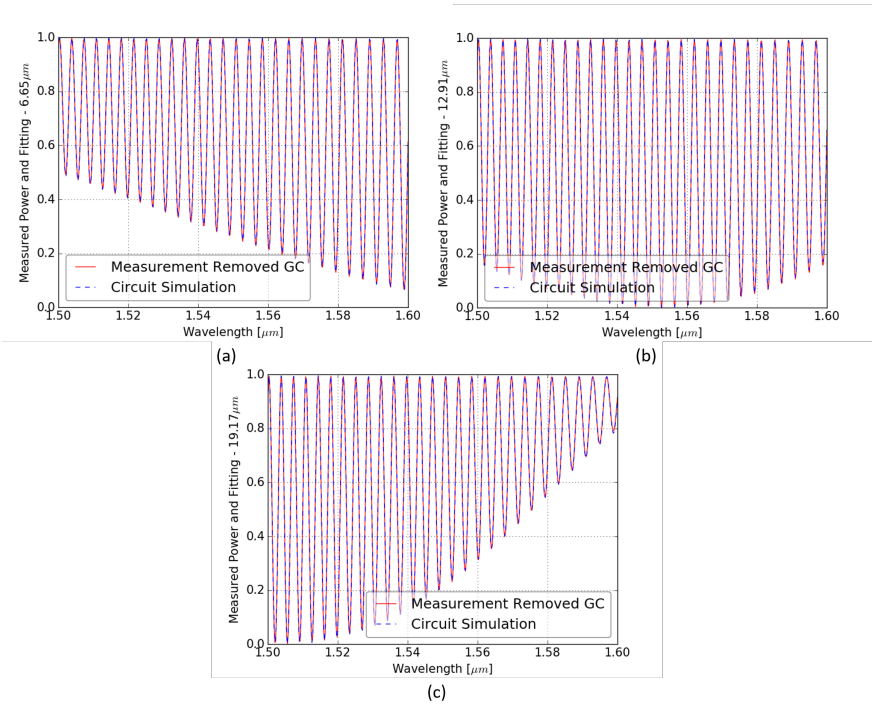


Figure 3.18: Fitting of MZI spectra after removing GC envelope with the GC model. Red solid: simulated Measurement generate with GC variation and measurement noise. Blue dash: Caphe circuit simulation that incorporates polynomial GC model. (a) L=6.65 μm. (b) L=12.91 μm. (c) L=19.17 μm.

This circuit design and curve-fitting give a high accuracy of the DC coupling coefficients (Table 3.4) in the presence of measurement noise and the GC variation. If we use the normalization method, the extracted parameters from the MZI circuit have significantly larger extraction errors, as shown in Table 3.4. Also, compared to all the methods mentioned above, this method ensures the highest accuracy. The result confirms that to extract the DC parameters with high accuracy, we need both the curve-fitting method and good circuit design to deal with measurement noise and the GC variation.

	simulated measurements	Remove GC with the model		Remove GC by normalization	
		Values	Relative Error [%]	Values	Relative Error [%]
κ'	0.0406	0.0404	0.49	0.0432	6.40
$\frac{d\kappa'}{d\lambda}$	0.2382	0.2463	3.40	0.2096	12.01
$\frac{d\kappa'^2}{d^2\lambda}$	1.0047	0.8681	13.60	2.4744	146.28
κ_0	0.2548	0.2495	2.08	0.2117	16.92
$\frac{d\kappa_0}{d\lambda}$	1.2107	1.0612	12.35	1.6782	38.61

Table 3.4: Parameters extracted using a GC model and using normalized transmission from the MZI circuits that each has two identical DCs.

3.3 Conclusion

In the chapter, we discussed how we built the behavioral model of DC. We validated the dispersive model from both FDTD simulations and on-chip measurements. To extract the coupling coefficients accurately in the presence of measurement noise and a GC variation, we should combine a sensible circuit design and the use of curve fitting method. We found out that we can use the curve fitting method to extract parameters from the MZI circuit that comprises two identical DCs. The circuit allows us to remove the GC variation in the spectra and improves the extraction accuracy significantly. We provided a detailed discussion and presented results on the coupling coefficients extraction.

The method discussed in the chapter requires at least three circuits with different coupler lengths to extract one set of the coupler parameters. If we remove the GC envelope using normalization, we need two optical measurements for each circuit. Therefore, at least six measurements are needed to extract one set of values. Also, the distance between extraction circuits indicates a local variation that hampers the assumption that all DCs in the extraction circuits suffers the same process variation. It will introduce an error in the extraction. In the next chapter, we will propose a circuit that allows us to address these problems.

References

[1] Wei-Ping Huang. *Coupled-mode theory for optical waveguides: an overview*. Journal of the Optical Society of America A, 11(3):963, 3 1994.

[2] Tsuyoshi Horikawa, Daisuke Shimura, Hideaki Okayama, Seok Hwan Jeong, Hiroyuki Takahashi, Jun Ushida, Yohei Sobu, Akemi Shiina, Masatoshi Tokushima, Keizo Kinoshita, and Tohru Mogami. *A 300-mm Silicon Photonics Platform for Large-Scale Device Integration*. IEEE Journal of Selected Topics in Quantum Electronics, 2018.

[3] Martin Fiers, Thomas Van Vaerenbergh, Ken Caluwaerts, Dries Vande Ginste, Benjamin Schrauwen, Joni Dambre, and Peter Bienstman. *Time-domain and frequency-domain modeling of nonlinear optical components at the circuit-level using a node-based approach*. Journal of the Optical Society of America B, 29(5):896, 5 2012.

[4] Katrien De Vos. *Label-free Silicon Photonics Biosensor Platform with Microring Resonators*. PhD thesis, Ghent University, 2010.

4

Process Control Monitoring Using a Compact Silicon Photonics Circuit to Extract Multiple Parameters

In Chapter 2 and 3, we presented a few designs and techniques to extract waveguide and directional coupler parameters. In this chapter, we present a compact circuit to simultaneously extract multiple parameters of on-chip waveguides and directional couplers from optical measurements. The compact design suffers less from local variation with in the circuit which greatly improves the accuracy of extraction. Also, the circuit greatly reduces the duration of wafer-scale optical measurements, making it useful for process control monitoring and detailed wafer-level variability analysis. The parameter extraction from the complicated spectrum requires a global optimization to replace the standard curve fitting method. We will discuss the design requirements and illustrate and how to setup the extraction using the Restart-CMA-ES global optimization algorithm. Since the circuit is designed to extract detailed wafer maps that will be illustrated in Chapter 5, we will also discuss the workflow to extract wafer maps in the last part of the chapter.

4.1 Background and Challenges in Process Control Monitoring

4.1.1 Progress in Process Monitoring for Photonics Chips

To monitor the performance of fabricated circuits and extract process variations, it is essential to do *process control monitoring* (PCM). PCM monitors the essential properties and variations of the fabrication process. PCM extracts technology specific parameters across the wafer (and between wafers and lots) that offer the input data for device-level and circuit-level variability analysis that we will describe in Chapter 5. The devices or circuits for parameter extraction should be compact so they can be placed at various locations to construct a granular map of the process variation on the fabricated chips as input for location-dependent variability analysis.

As we explained in Chapter 2, for variability analysis, we preferably measure parameters on the final fabricated circuits and at a large number of sites to obtain the variability contributions at different length scales. Optical transmission measurements provide a desirable alternative to measure fabricated geometry. We introduced in Chapter 2 an *Mach-Zehnder interferometers* (MZIs) to extract the average effective and group indices along the path of a delay line. Because silicon waveguides are extremely sensitive to geometry variations, the effective and group indices can be mapped uniquely onto waveguide linewidth and thickness, which allow us to derive small variations in the fabricated waveguide geometry with a sub-nanometer accuracy as described in Chapter 2.

In addition to waveguide parameters, the parameters of a *directional coupler* (DC) are also essential in the performance of optical filters. In Chapter 3, we described different methods to extract obtain coupling coefficients of a DC. A typical optical filter measurement captures the power transmission of at least two DCs (preferably 3 or more) with different coupling length to separate the length-dependent coupling and the contribution of the bends. To eliminate the effect of the *grating couplers* (GCs), we measured the two outputs of the directional coupler and normalized the transmission to the total power. In total, 3 DCs and 6 optical measurements are required for the extraction, and then we have to assume the properties of these three DCs are identical. Any variation in linewidth, thickness, and gap among the DCs will introduce extraction errors. Therefore, it is desirable to bring three DCs as closely together on the chip to reduce the extraction error caused by the local variations.

4.1.2 Challenges in Current Monitoring Circuit Designs

There are two major challenges in current designs described in Chapter 2 and Chapter 3 of monitoring circuits. First, the footprint of the design needs further

reduction. A compact monitoring circuit reduces local variation within the circuit that improves extraction accuracy. In the two MZIs design for n_{eff} and n_g extraction, we assumed two MZIs are on the same location on the die and they only suffer random device-to-device variation between them. In the design to extract DC parameters, we assumed that DCs in three MZIs experience the same fabrication variation so that they have identical cross-section and bend sections. Such assumptions are less convincing when the monitoring circuit is not compact, so different location in the circuit suffers significantly different variation. Process variation has a significant deterministic contribution that can be determined by its location on a wafer. We will describe in Chapter 5 that the fabrication variation is correlated between nearby sites. For example, on the intra-die level, the linewidth is correlated in a several-hundred-micron range, because the local pattern density affects it on the scale of 100 μm (Chapter. 5 Section 5.6). To extract the detailed location dependence of the deterministic variation and to observe such a short-range spatial correlation, we need to design the monitoring circuit as compact as possible. Besides, we want the circuit to be so compact that we can easily squeeze it in various locations on-chip and place many replicates of it very close to the region of interest.

The second challenge is the significant time cost of the optical measurement for process monitoring, especially when extracting parameters from multiple circuits. The time cost of the optical measurement becomes substantial when we are collecting a detailed wafer map. If each die requires hundreds of monitoring circuits, a wafer would easily contain thousands of such circuits. The automated measurement on all circuits on-wafer would require several months. In particular, when you use separate circuits described in previous chapters, for example, to extract waveguide parameters and DC parameters, the number of measurements can be multiplied a few times further. If we could reduce the number of optical measurements for each monitoring circuit, we will significantly reduce the measurement time.

4.2 Design of the Monitoring Circuit

4.2.1 Footprint of the Two-Stage MZI Monitoring Circuit

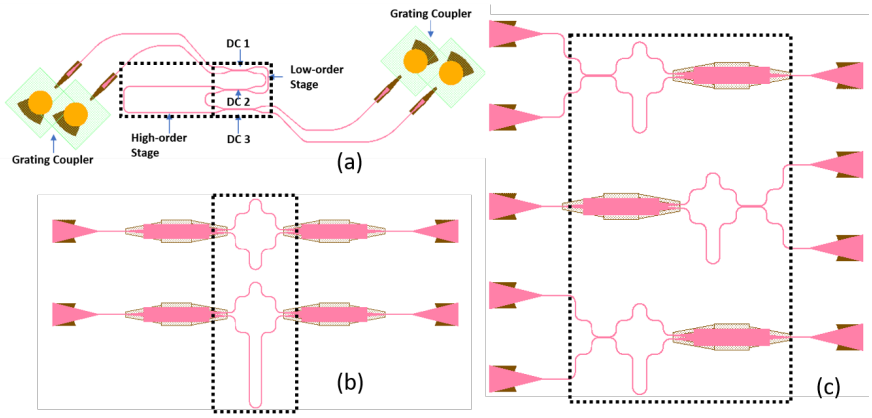


Figure 4.1: (a) The layout of the folded two-stage MZI. We used low-reflection GCs [1] to reduce measurement noise due to parasitic back-reflections. The large dotted frame indicates the region to extract process variation of the waveguide. The small dotted frame on the right indicates the region to extract process variation of the DC. (b) The layout of the two MZIs to extract waveguide parameters used in Chapter 2. The design has a footprint of 350 μm × 180 μm. The region to extract process variation of the waveguide covers an area of 55 μm × 180 μm. (c) The layout of the three MZIs to extract DC parameter used in Chapter 3. The design has a footprint of 285 μm × 280 μm. The region to extract process variation of the DC covers an area of 150 μm × 220 μm.

To reduce the footprint of test structures and the number of optical measurements for performance evaluation, we propose a two-stage MZI design, shown in Figure 4.1(a), with which we can simultaneously extract effective and group indices of a waveguide and parameters of the used DCs. The design wraps the low-order and high-order MZI into one circuit with two inputs and two outputs. Including low-reflection GCs [1], the circuit has a footprint of 400 μm × 100 μm which can be further reduced by quick adjustment of the layout. The region to extract process variation of the waveguide includes arms of both the low-order stage and the high-order stage. Grating couplers and entrance waveguides are not included because they do not determine the extracte waveguide and DC parameters. It covers an area of 120 μm × 40 μm. It is 1.6× smaller in terms of footprint and 2.1× smaller in terms of the region to extract variation compared to the two-MZI design in Figure 4.1(b) used in Chapter 2 for waveguide parameter extraction. For DC parameter extraction, the region of interest is the small rectangle that covers three DCs with an area of 45 μm × 40 μm. Our monitoring circuit is 2.0× smaller in terms of footprint and 18.3× smaller in terms of the region to extract variation

compared to the three-MZI design in Figure 4.1(c) used in Chapter 3. The design is organized to be very compact which reduces the local variation between waveguides and DCs, which in turn improves the extraction accuracy. Moreover, the circuit requires only 2 optical measurements instead of 8 to extract all waveguide and DC parameters, which significantly reduces the cost of the automated optical measurements.

4.2.2 Design Principle

We design the two-stage MZI using the same rules for the low and high order as described in Chapter 2. We used a combination of a low and a high order MZI stage to extract the effective and group indices of straight waveguides. The fabrication variation can shift the spectrum of an MZI by more than one *free spectral range* (FSR), making it difficult to identify the correct discrete interference order, resulting into multiple solutions for the effective index. So, we designed the order of one MZI stage sufficiently low, such that its spectrum will not shift more than one FSR under the expected process variation. This low-order stage offers a local reference for the effective index. The second, high-order stage has many more interference orders within the measurement range, offering more spectral features for accurate extraction of both effective and group indices. The order of the high-order stage is designed such that we can still estimate effective index reliably based on the local reference effective index extracted from the low-order stage. The low-order stage provides a local reference point for the effective index, and the high-order stage enables accurate extraction.

We based our designs on the specifications in imec's technology handbook for the iSiPP50G silicon photonics platform. For the waveguides, the standard deviation in linewidth is specified as 5.3 nm over the wafer, while the thickness has a standard deviation of 0.7 nm. For a safe design based on a 6σ spread, we targeted waveguides of 470$\pm$15 nm line width and 210$\pm$5 nm thickness. We assume the waveguide is rectangular with 90° sidewall, where the specified sidewall angle is above 85°. As we are mostly interested in relative variations on the wafer and between wafers, this deviation from the vertical is not a big issue. As explained in Chapter 2, the n_g extraction from the low-stage is inaccurate. Without information of n_g on the low-order stage, we estimate the tolerance of its n_{eff} by Equation 2.11 in Chapter 2:

$$\begin{aligned}\Delta n_{eff,total} &= \frac{\partial n_{eff}}{\partial w}\Delta w_{total} + \frac{\partial n_{eff}}{\partial t}\Delta t_{total} \\ &= 0.0019\ nm^{-1} \times 30\ nm + 0.0040\ nm^{-1} \times 10\ nm = 0.097.\end{aligned} \tag{4.1}$$

Then $L_{low\ order} < \frac{\lambda}{\Delta n_{eff,total}} = 16.0\mu m$. We estimate the local variation from the maximum difference between an extracted parameter with an interpolated wafer

map. From Equation 2.12 in Chapter 2, we know that when the width variation is significantly larger than the thickness variation that $\Delta w > -\frac{\partial w}{\partial n_g} \cdot \frac{\partial n_g}{\partial t}\Delta t$, the range of n_{eff} is determined largely by Δt. From Chapter 2, we also know that thickness varies smoothly over the wafer, with local variations of less than $\pm$ 0.6 nm on the two dies we measured. So here we assume the maximum local variation (within the MZI circuit) is below ± 0.8 nm. In the high-order MZI, the extraction of n_g is much more accurate, as we cover more interference orders in the measurement range. For $w \in [455, 485]$ nm and $t \in [205, 215]$ nm, we can now, knowing the accurate local n_g, estimate the range of the high-order n_{eff} by Equation 2.12 in Chapter 2:

$$\begin{aligned}\Delta n_{eff,local} &= \left(-\frac{\frac{\partial n_{eff}}{\partial w}\frac{\partial n_g}{\partial t}}{\frac{\partial n_g}{\partial w}} + \frac{\partial n_{eff}}{\partial t} \right) \Delta t_{local} \\ &= 0.0064\ nm^{-1} \times 0.8\ nm \times 2 = 0.0102. \end{aligned} \tag{4.2}$$

Then $L_{high\ order} < \frac{\lambda}{\Delta n_{eff,local}} = 152.0 \mu m$. From the analysis, we choose the arm length difference of the low-order stage as 15 μm and the high-order as 150 μm.

To extract DC parameters, we put three DCs connecting the two MZI stages, and the coupler lengths correspond to a nominal 25%, 50%, 75% cross coupling power at 1550 nm. The gap between the waveguides in the DC is 250 nm, and the corresponding coupler length in three DCs are 6.65 μm, 12.91 μm, 19.17 μm. To further reduce the footprint of the device, we also folded the MZI as shown in Figure 4.1(a) so that we reduced the distance between the pairs of arms and the three DCs. This should reduce local variation and improve extraction accuracy.

4.3 Extracting Multiple Parameters Using the Restart-CMA-ES method

4.3.1 Circuit Model

We extract the parameters of the folded MZI circuits by matching a simulated spectrum with the measured spectrum. This requires a behavioral model for the circuit and its constituent components. As in Figure 4.2, for a waveguide arm of the MZI, we use two compact model parameters, namely effective index n_{eff} and group index n_g at $\lambda_0 = 1550\ nm$. The effective index n_{eff} at a given wavelength is then:

$$n_{eff}(\lambda) = n_{eff} - (\lambda - \lambda_0) \cdot \frac{n_g - n_{eff}}{\lambda_0} \tag{4.3}$$

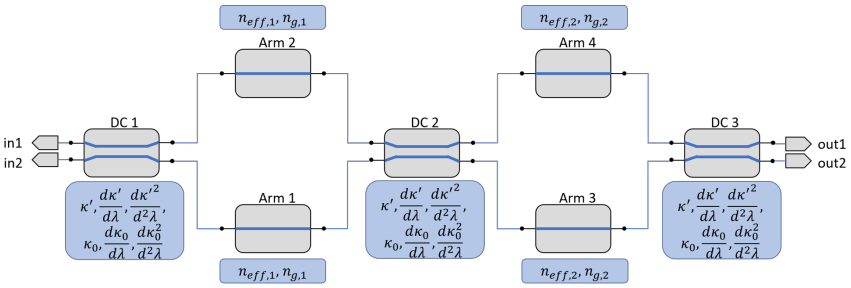


Figure 4.2: The circuit model of the device. Two MZI stages have different n_{eff} and n_g led by the local fabrication variation.

As we have explained in Chapter 3, a DC has coupling contributions from two parts: the straight coupling section and its two bends. When we neglect insertion loss, the power at the coupled port is:

$$K_{coupled}(\lambda) = sin^2(\kappa'(\lambda)L_{coupler} + \kappa_0(\lambda)) \tag{4.4}$$

As explained in Chapter 3, the DC model has six parameters, namely the length-specific coupling coefficient of the straight coupling part κ' and its first and second-order derivative $\frac{\partial\kappa'}{\partial\lambda}$ and $\frac{\partial^2\kappa'}{\partial\lambda^2}$, and the lumped power coupling of the bend κ_0 and its first and second-order derivative $\frac{\partial\kappa_0}{\partial\lambda}$ and $\frac{\partial^2\kappa_0}{\partial\lambda^2}$.

We implemented the compact model of the two-stage MZI in the IPKISS circuit simulator CAPHE of Luceda Photonics [2]. We then try to match the simulated spectrum to the measured optical spectrum by adjusting the model parameters. To remove the effect of GCs in the spectrum, we measured both the spectra from port *in1* to *out1* and *in1* to *out2* and normalized the transmission spectra to the sum of the two spectra. The solid red curve in Figure 4.8 shows a typical normalized measured spectrum from port *in1* to port *out1*.

4.3.2 The Problem to Use the Standard Curve Fitting Method

Standard curve fitting methods (e.g., from the scientific python package 'scipy') are capable of extracting parameters from a single MZI response as described in Chapter 2. However, it becomes difficult to use these curve fitting to extract parameters from the two-stage MZI. As shown in Figure 4.3, the spectrum of the device is more complicated. Extracting the circuit parameters is an optimization to minimize the target function that is the difference between the simulated and measured spectrum. To find the right and unique parameter set, we are not interested in a local minimum of the target function. As displayed in Figure 4.3, the classical curve-fitting methods often fail to handle the non-convex parameter landscape and will obtain a local optimum instead of the global optimum.

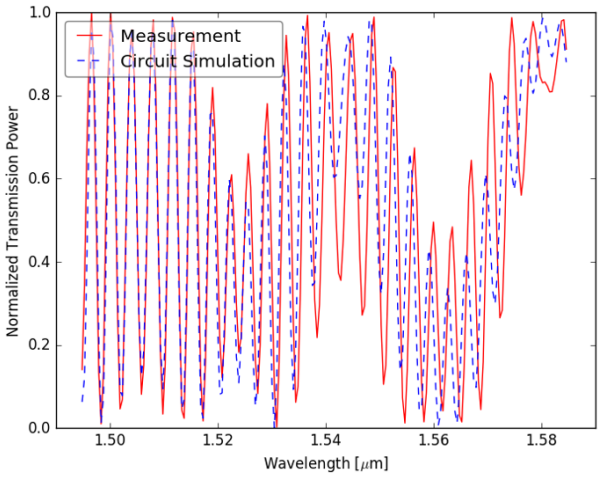


Figure 4.3: A bad match between simulated and measured spectra is obtained by the standard curve fitting method. Red: measured spectrum. Blue: simulated spectrum by CAPHE.

4.3.3 Parameter Extraction Using the Restart CMA-ES Algorithm

4.3.3.1 CMA-ES Algorithm

Instead, we can use smart global optimization algorithms that adaptively choose the samples to drastically reduce the number of simulations for the non-convex parameter landscape optimization. Covariance Matrix Adaptation Evolutionary Strategy (CMA-ES) is an optimization method that adaptively chooses its searching path and searching range (Figure 4.3). The algorithm chooses samples of the population of a new generation based on the samples offering the best optimization of the previous generation [3]. The CMA-ES greatly reduces the sample number in the extraction and is especially powerful to extract multiple parameters simultaneously. Also, unlike other optimization technique, it has only a few parameters to set up, which is easy and intuitive to use. The technique has been implemented in many programming languages, and you can access the open source code from [4]. The technique is very efficient when we apply it to extract a spectrum with complex features, but it does not always guarantee to find the global optimum.

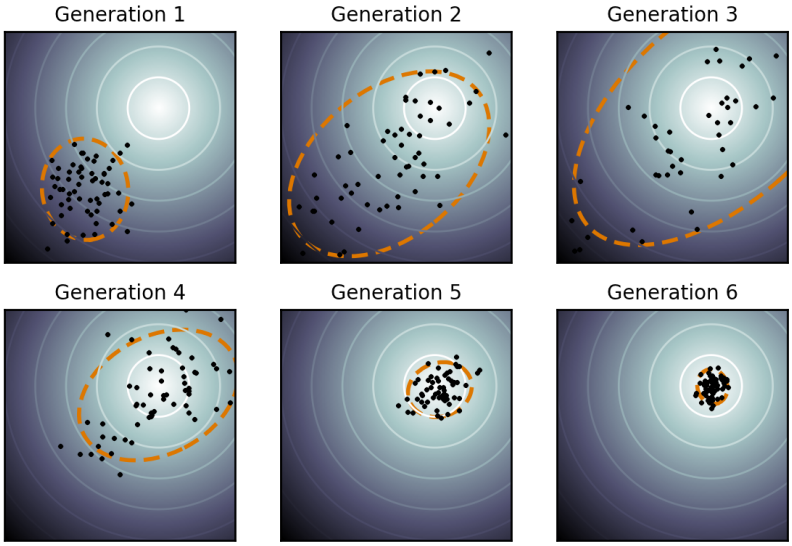


Figure 4.4: Illustration of an actual optimization run with CMA on a simple two-dimensional problem. The spherical optimization landscape is depicted with solid lines of equal f-values. The population (dots) is much larger than necessary but clearly shows how the distribution of the population (dotted line) changes during the optimization. On this simple problem, the population concentrates over the global optimum within a few generations. The figure is from Wikipedia. [5]

4.3.3.2 Parameter Extraction using the Restart-CMA-ES Algorithm

A variation, the Restart-CMA-ES method, is a global optimization method which is suitable for our purpose because it guarantees to locate the global optimum. In particular, we adopted the technique described in [6]. The Restart-CMA-ES method will restart the optimization after one search if it only finds a local optimum. We decide that the optimization reaches the global optimum when the objective function is below a predefined value that is small enough to distinguish a global optimum and a local optimum. How to choose the predefined value will be discussed in Section 4.4.2. After each restart, we could also increase the population size, so the search characteristic becomes more global after each restart. The loop stops when the error between the simulation and measurement is below the defined threshold, which indicates the global optimum is obtained. We validated the algorithm with simulated samples with ± 0.2 dBm noise to emulate the typical measurement noise. It works robustly for given waveguide variation (w$\in$[465,485] nm, t$\in$[205,215] nm) and DC gap $\in$[100,400] nm.

4.4 How to Set Up the CMA-ES Algorithm

The Restart CMA-ES global optimization technique is very powerful and robust. In theory, it guarantees to locate the global optimum with high accuracy if we build the correct circuit model, set a right parameter boundary and use a sufficient amount of evaluations. In practice, when we are extracting a wafer map, the process control monitoring requires to repeat the extraction to obtain parameters from thousands of circuits accurately. Noise in the spectral measurements also means the mismatch between simulation and measurement is not zero. We want to obtain accurate parameter, and meanwhile consider the cost of the strategy which is down to the choice of algorithm parameters.

The CMA-ES algorithm does not require a tedious parameter tuning for its application. [5]. In fact, the choice of internal parameters in the algorithm is not left to the users. Instead, it is part of the algorithm design. On the user side, we can choose the population size that determines the convergence rate. Initial step-size is also an option left to users. Termination criteria such as function tolerance, maximum evaluations, minimum change in the parameters can be chosen to change the cost of the algorithm. We will discuss how to choose these parameters below.

4.4.1 Initial Guess and Step Size

The CMA-ES implementation chooses the search direction and step size adaptively. The step size here is the step size of the first generation. It is an initial standard deviation of the variable. We have normalized all the variables according to the variable range such that the same standard deviations can be reasonably applied to all variables. The variable range is set between 0 and 2, which means the variable v with a range from $v - \Delta v$ to $v + \Delta v$ should be linearly mapped to 0 to 2.

When we choose a small step size, the algorithm can quickly converge to a local optimum. This is very useful when we have a good initial guess. For example, when we have a solution from a previous optimization and we want to improve the fitting error of the solution, we can use a refined step size and search in the neighborhood with more evaluations. This helps to quickly converge to the solution.

For most of the cases, we did not know where the solution is in our search range determined by the process variation. We are rather looking for the 'global' optimum. According to the tutorial, [5], the initial step size should be about one-fourth of the search domain width, where the optimum of the strategy is to be expected. In terms of the initial guess, We could only assume it within the searching range. One way is to assume the initial guess right in the middle of the parameter space. However, we do not want to fix the position of the initial guess. If the right solution is near the boundary of the searching range, the initial guess in the middle makes it quite unlikely to find it. Instead, we assume a random initial guess to

add robustness to the algorithm. At each restart, we choose it randomly following an uniform distribution. Choosing the initial guess randomly within the range makes the algorithm more tolerant and more likely to find the optimum after a few restarts.

4.4.2 Global or Local Optimum?

The standard to judge if an obtained optimum is global or local is to look at the mismatch between the simulation and the measurement. 'f-value' in the CMA-ES algorithm is the mismatch between the measured spectrum and the simulated spectrum. In the ideal case, there should be no mismatch between these two. In practice, measurement noise and mismatch between the circuit model and the circuit response introduce mismatches. Measurement noise is around 0.2 dBm for most of our measurement, which is around 5% relative error on each wavelength. There is also mismatch between the circuit model we assumed and the real circuit behavior. For instance, the assumption that three DCs share the same set of coupling parameters is based on the assumption that there is no local variation in the circuit. The deviation from the assumption introduces a mismatch. Also, in the procedure to remove the GC envelope, we assumed an identical GC response for all GCs in the circuit. The mismatch in the GC performance introduces error, especially in the region of low grating coupling efficiency. The small coupling leads to small optical transmission in the spectrum range, thus reducing the signal-noise-ratio and increasing the mismatch. We can remove the part of the spectrum with a low grating coupling efficiency to alleviate the issue.

The mismatch between the spectrum simulated at a local optimum and the measured one is much larger than the mismatch caused by noise and modelling errors. For example, a local optimum can be a spectrum that is one interference order off from the right order, and it leads to significant mismatch that is at least $0.10 \times$ number of wavelengths in the spectrum. For instance, the local optimum we found using the curve fitting in Figure 4.5 (b) has a mismatch of $0.15 \times$ number of wavelengths. In practice, we tested with thousands of spectra measured on the wafer, the mismatch driven by noise, or GC is below $0.02 \times$ time of wavelengths of the power spectrum. When it is over that value, we assume that the optimization did not find the global optimum. We use this standard to decide if the algorithm needs a restart and determine if we get a global optimum after a few restarts.

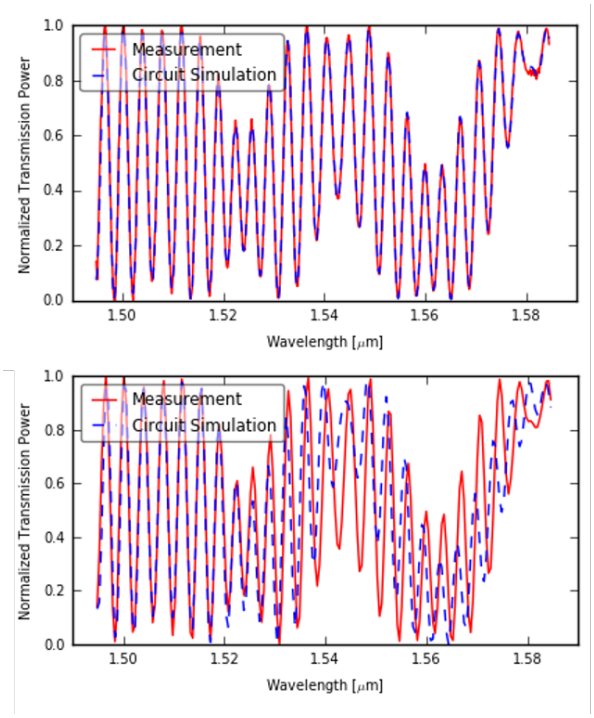


Figure 4.5: Two types of optimum. Red solid curve is the measured spectrum with 200 wavelengths. Blue dash is the simulated spectrum. (a) A very good match that corresponds to the global optimum. The value of the target function 'f-value' of the global optimum is 2.65, which is below 0.02 $\times$ time of wavelengths. (b) A bad match that corresponds to a Local optimum obtained by CMA-ES. The local optimum is has the smallest 'f-value' among of all observed local optimums, whose 'f-value' is 30.73, which is 0.15 $\times$ time of wavelengths of the power spectrum.

4.4.3 Population Size

Population size is the number of evaluations sampled in the parameter space in each generation. It determines how global the search is. Restart-CMA-ES is a global optimization method. However, 'global' is a word that requires attention. There are two ways to operate the Restart-CMA-ES: increasing the population size after each restart which makes the search more global, or using the same size after each restart. As shown in Figure 4.4, the radius of the search circle reduces generation by generation to approach the optimum. If you have a large population, there are more samples in the searching circle which captures the more terrain in the circle. More samples likely to find the global optimum

The 'global' approach increases the population size after restart so that the samples in the population cover more details of the searching area. The method

is more robust and reliable to find the optimum. However, most of the time, the choice has a slow convergence to the right solution. [3] Sometimes, it fails even to find a local optimum within 20,000 evaluations. So if we did not have a good initial guess, the 'global' approach requires too many evaluations to find the global optimum. On the other hand, the restart scheme with the same size of the search span is more 'local', it converges faster and is more efficient. However, it has a larger tendency to find one local optimum rather than a global one.

The default population size of each generation is given as: $4 + floor(3 \times ln(N))$. In our case, there are 10 parameters in a circuit simulation, so the default population size is 10. To see how convergence is related to the population size, we tried population size of 5, 10, 20, 40, 100 and did the optimization without restart. For each population size, we run 10 times of optimization. As in Table. 4.1, the number of evaluation increases with the population size almost linearly when the population size is above 20. Below 20, the number of evaluations does not change significantly. On the other hand, we also test the the successful rate to obtain global optimum without restart. The rate to obtain a global optimum increases with the population size. In our test, the rate did not significantly improve the population size above 20, which might because we tried only 10 times of optimization for each population size. Still, it is reasonable to choose population size as 20 which fast converges and at the same time ensures a good coverage of searching space.

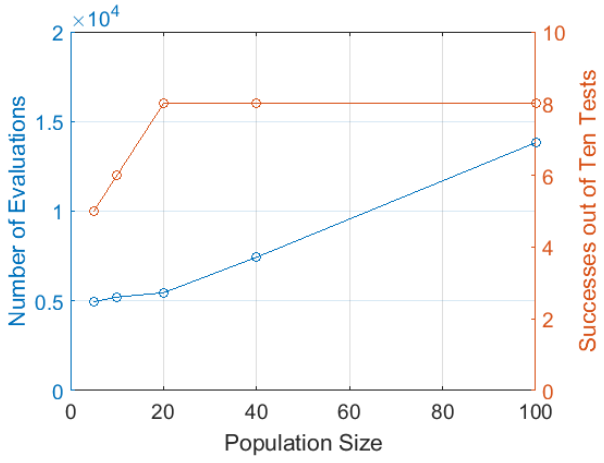


Figure 4.6: Population size vs. number of evaluations to find an optimum. The optimization stops when the tolerance in the function value is below 10^{-7}.

Population Size	5	10	20	40	100
Number of Evaluations	4946	5218	5452	7430	13820
Successful Tests	5	6	8	8	8

Table 4.1: The number of evaluations required to find an optimum and the rate to find a global optimum under different population size.

We set the 'local' method with a population size of 20. For a majority of our measurements, the 'local' method without increasing population size is capable of finding the solution with a maximum number of evaluations in each restart set at 20,000. We allow the algorithm to restart a maximum of 3 times so it would not run forever if there is an error in measurement or there is a fault in the optimization set up. If it did fail to find a solution in some spectrum, we switched to the 'global' approach to process a difficult spectrum. We set the 'global' method with the initial population size of 20. After each restart, the population size is 50% larger than the previous generation. The global method is also limited to a maximum of 3 restarts. As shown in Figure4.6, an optimum can be found within 20,000 evaluations even with the largest population size we used in the 'global' approach. So we limit the evaluations to 20,000. If we did not find the global optimum after three restarts, we would just rerun the global algorithm. Most of the time it works within two reruns.

By using the 'local' and 'global' method to target different spectra, we can combine the merits of both approaches.

4.4.4 Termination Criteria

Termination criteria tell the program to stop when it meets one or more preset criteria. There are a few options we used to stop the optimization.

- 'TolFun' : It is the tolerance in the function value. Stop if the range of the best objective function values of the last $10 + [30 * \mathrm{n}/\lambda]$ generations and all function values of the recent generation is below TolFun. n is the number of variables in the searching space. λ is the step size. The smaller TolFun means more evaluations that increases optimization cost and reduces the fitting error. The optimal TolFun depends on the problem.

  We did a convergence test to find out the optimal TolFun for our problem. We set the TolFun as 10^{-12}, which is a very small number so that the optimization can run sufficiently long to get the solution with low fitting errors. As in Figure 4.7 (a), the optimization falls into the vicinity of the solution after around 2,500 evaluations, when the TolFun is around 10^{-3}. Followed evaluations continue to reduce the fitting error as shown in Figure 4.7 (b). From Figure4.7 (a) and (b), we observed that after the algorithm start to

search in the close vicinity of the global optimum and will not jump to another adjacent local optimum, it takes around 2,000 iterations to lower TolFun by 2 magnitudes and the fitting error by 1 magnitude. We can calculate the fitting error for width and thickness from the fitting errors of $n_{eff,2}$ and $n_{g,2}$. The TolFun approaches 10^{-6} after 4,000 evaluations. The fitting errors of both width and thickness drop below 0.1 nm, which is sufficiently small for variability analysis. To be on the safer side, we find TolFun as 10^{-7} ensuring a solution with sufficiently low fitting errors on one hand, and still being cost-effective on the other hand.

- 'maxfevals': It is the maximum evaluations of one restart. More evaluations do not only increase the odds to find the optimum, but also helps to reduce the fitting error (which is proportional to the radius of searching) with more time approaching the optimum. However, if we set the evaluation maximum necessarily large, it is a waste when the algorithm cannot find an optimum. It is better to restart the searching for such cases. On the other hand, we should allow the algorithm to find an optimum with sufficiently small fitting error under most situations, which reduces the reprocessing of unsuccessful samples in the following steps. We set the number to be 20,000. As shown in Figure 4.7 (c), most of the time, the algorithm can find an optimum in around 8000 evaluations with sufficient parameter accuracy. We use 20,000 as the maximum to make sure in the majority of the cases, the optimum is found and the fitting error is reduced to an acceptable level.

As shown in Figure 4.8, we obtained excellent matching between simulated and measured spectra using the restart-CMA-ES with increasing population after each restart. To extract ten parameters with high accuracy, usually, the optimization requires less than 20,000 evaluations. The behavior parameters have been extracted with good accuracy (Table 4.2). Since the circuit is very simple, we can test in the future a simple and very fast transfer matrix method without having to run a full-fledged circuit simulator. If we apply a simple evaluation of an analytical formula, the extraction should be greatly accelerated.

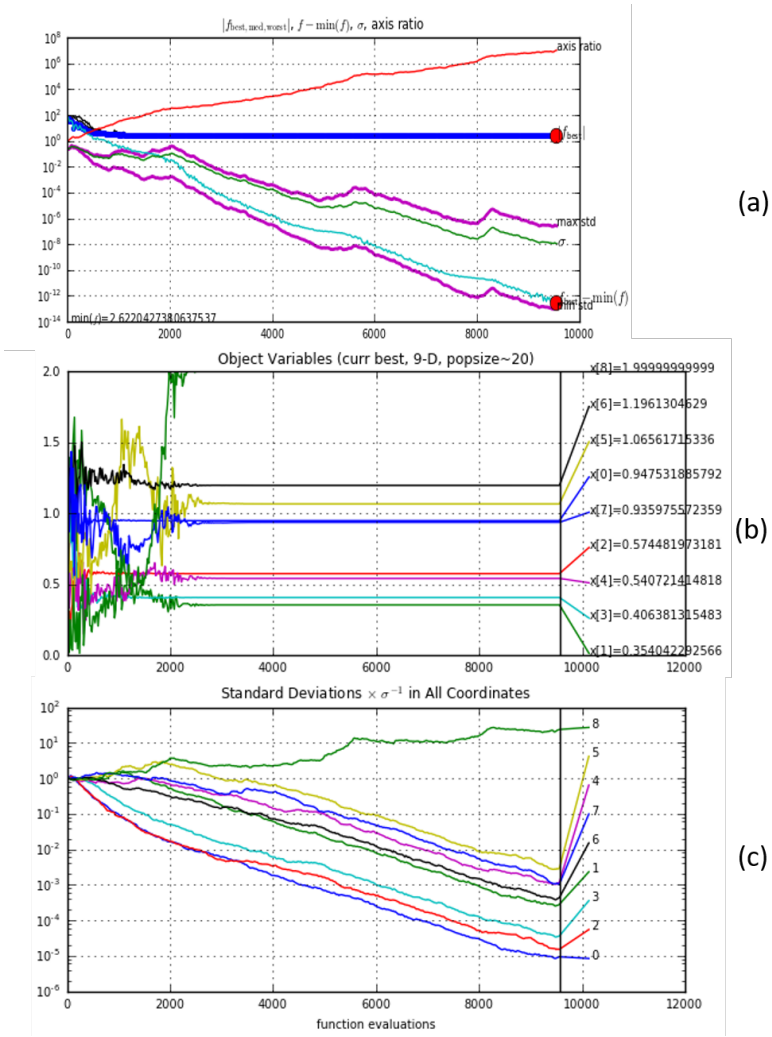


Figure 4.7: (a) The blue curve is the minimal target function value after each evaluation. The light blue curve is the 'TolFun' (f_{best}-f(min)) which is the difference between target function and minimum target function found. We stopped the optimization when the 'TolFun' value reaches 10^{-12}. (b) x(1) to x(8) are $n_{eff,1}$, $n_{g,1}$, $n_{eff,2}$, $n_{g,2}$, $\frac{d\kappa'}{d\lambda}$, $\frac{d\kappa'^2}{d^2\lambda}$, κ_0, $\frac{d\kappa_0}{d\lambda}$, $\frac{d\kappa_0^2}{d^2\lambda}$ normalized to their variation range. All normalized parameters range from -1 to 1. In this case, the optimization finds a good solution after around 2500 evaluations. Further evaluations can reduce the fitting errors of the parameters. (c) Curve 1 to 8 indicates standard deviation of the normalized parameters of the circuit. The fitting errors of variables are defined as the twice of the normalized standard deviation of variables. At 3000 evaluations, the converted width and thickness fitting errors are both below 0.1 nm.

	Obtained Value	Fitting Error		Obtained Value	Fitting Error
$n_{eff,1}$	2.356	$1.456 \cdot 10^{-6}$	$\frac{d\kappa'}{d\lambda}$	$2.149 \cdot 10^{-1}$	$9.147 \cdot 10^{-5}$
$n_{g,1}$	4.228	$1.322 \cdot 10^{-4}$	$\frac{d\kappa'^2}{d^2\lambda}$	1.990	4.060
$n_{eff,2}$	2.356	$2.284 \cdot 10^{-7}$	κ_0	$2.315 \cdot 10^{-1}$	$7.852 \cdot 10^{-5}$
$n_{g,2}$	4.220	$2.105 \cdot 10^{-5}$	$\frac{d\kappa_0}{d\lambda}$	1.438	$1.266 \cdot 10^{-2}$
κ'	$4.173 \cdot 10^{-2}$	$5.863 \cdot 10^{-6}$	$\frac{d\kappa_0^2}{d^2\lambda}$	$8.110 \cdot 10^{-1}$	$6.325 \cdot 10^{-2}$

Table 4.2: Obtained parameter values and fitting errors using the Restart CMA-ES method.

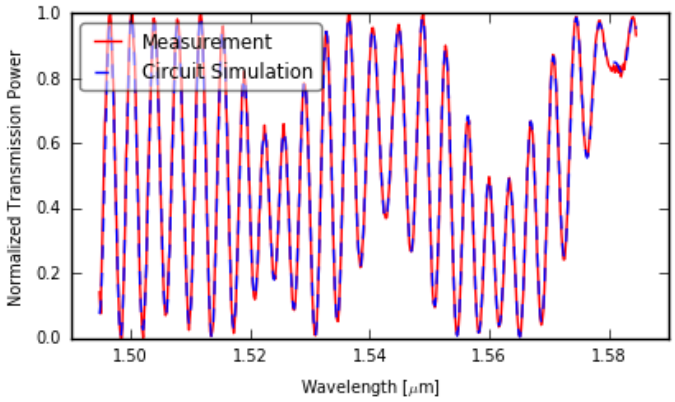


Figure 4.8: A good match between simulated and measured spectra is achieved by the restart-CMA-ES method. Red: measured spectrum. Blue: simulated spectrum by CAPHE.

Then, we mapped the width and thickness of the high-order stage arm from $n_{eff,2}$ and $n_{g,2}$ (Table 4.3). As explained in Chapter 2, the extraction of geometry parameters includes several errors, from the model, the simulations, the mapping, and the fitting procedure. The modeling error is the mismatch between the compact circuit model and the actual fabricated circuit behavior; for example, assuming identical parameters κ' and κ_0 for the three DCs while the fabricated DCs have some disparity. The simulation error is the difference between the actual waveguide geometry (the shape, dimension and material properties) and the rectangular geometry we used in the mode solver. This error is hard to compensate, but it is relative and will not affect the trends in the extracted parameters. The mapping error is the difference between the simulated waveguide geometry and extracted waveguide geometry. The mapping error of width and thickness are 0.06 nm and 0.08 nm respectively when we apply a third-order polynomial fitted model. The fitting error is estimated by twice the standard deviation of each parameter obtained by the fitting, which provides confidence limits of approximately 95%. Extracted

Parameter	Extracted Value	Fitting Error	Mapping Error	Total Error
Width	474.68 nm	0.01 nm	0.06 nm	0.07 nm
Thickness	208.35 nm	0.01 nm	0.08 nm	0.09 nm

Table 4.3: Extracted waveguide width and thickness of the high-order stage arm.

width and thickness each have a 0.01 nm fitting error.

4.5 Results on a Silicon Photonics Die

We automated the optical measurements on 117 copies of the two-stage MZI on the same die (Figure 4.9). We measured test circuits in our clean room with the temperature controlled at 20 degree Celsius using a calibrated laser. We first extracted all ten parameters for each circuit. Then, we interpolated the $n_{eff,1}$ extracted from the lower-stage to get a reference wafer map of the n_{eff}. After that, we used the n_{eff} wafer map as a reference at each location for the high-order stage and revised the derived high-order $n_{eff,2}$ values to bracket them in the boundary defined by the local variation. Figure 4.10 presents the extracted effective and group indices of the high-order stage. We then used the geometry model to map $n_{eff,2}$ and $n_{g,2}$ to width w and thickness t of the high-order MZI arms. The extracted linewidth on die (X=0, Y=0) in the wafer center ranges from 468.9 nm to 479.5 nm (Figure 4.11. a) and thickness ranges from 207.6 nm to 209.6 nm (Figure 4.11. b). The standard deviations are 1.9 nm and 0.5 nm respectively.

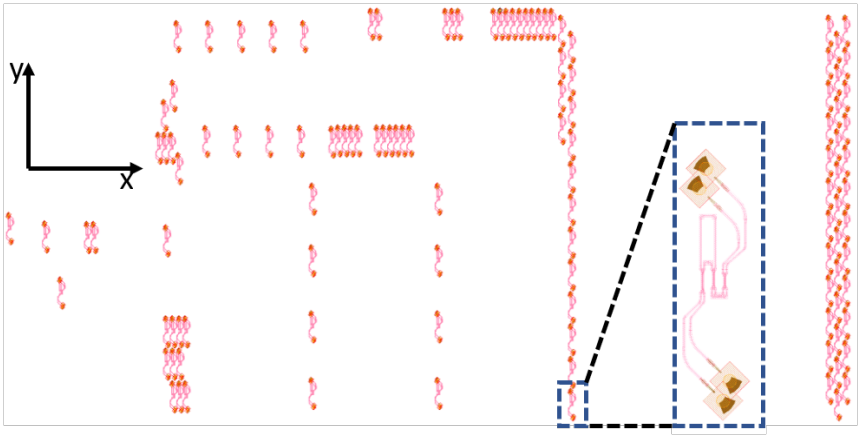


Figure 4.9: Locations of the folded two-stage MZIs on a die.

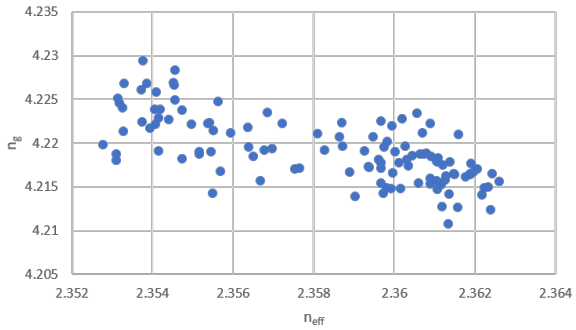


Figure 4.10: Extracted $n_{eff,2}$ and $n_{g,2}$ of die (X=0, Y=0) (in the center of the wafer).

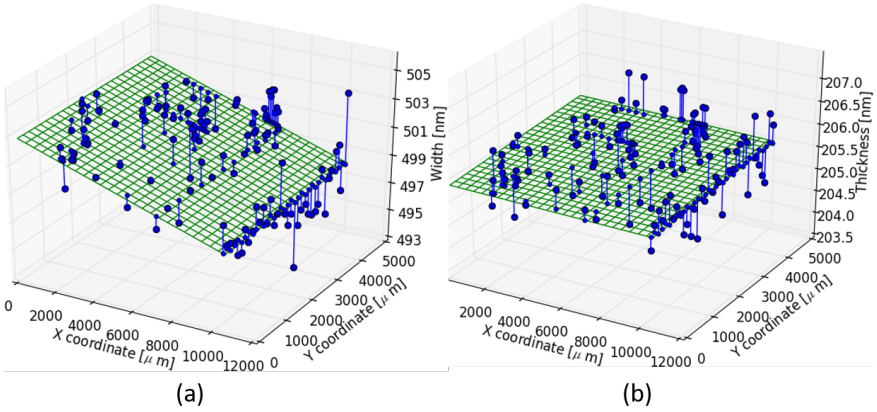


Figure 4.11: (a) Extracted width map and (b) thickness map of the die. x and y indicate the locations of the MZIs on the die. Blue dots: extracted value. Green grid: fitted map of extracted values using a linear function.

4.5.1 Our Workflow to Extract Wafer Map

Our ultimate purpose is to use the PCM circuit practically to obtain fabricated wafer maps as we will show in Chapter 5. We have discussed how to extract multiple parameters from a single circuit. To extract the parameters on a complete wafer is challenging for a couple of reasons. First, the extraction from a complicated spectrum requires many evaluations, which can be time-consuming. If we cannot reduce the cost of circuit simulation, extraction from thousands of circuits will be impractical. Second, the number of PCM circuits on the wafer is enormous. Even with very efficient circuit simulator, it would still take a month to generate a wafer map. Third, global optimization could be time-consuming if we apply to

the same strategy for all types of spectrum. In Appendix D, we discussed in detail to how to extract parameters on a fabricated wafer. From these discussion, we propose our workflow to extract fabricated wafer maps in this section.

Fig 4.12 shows the workflow to obtain the process wafer map. We parallel the extraction of the die map to speed up the extraction. We organized the result from every die to plot a process wafer map. We applied the workflow and obtained the wafer maps of a 8-inch wafer. The extracted results will be presented and discussed in the following Chapter 5.

1. Generate normalized spectrum from the measured spectrum to remove the grating coupler envelope.

2. Use the 'local' Restart CMA-ES to extract all the behavioral parameters. Set the maximum evaluation of each restart to 20,000. You get reference n_{eff} from the lower arm and accurate n_g on the high-order stage. Label the missing measurements and samples with a large mismatch between simulation and measurements.

3. Use the more 'global' Restart CMA-ES optimization to process the spectrum with a large mismatch. Set the maximum of the evaluation sufficiently large that it guarantees to find the global optimum.

4. Fit the n_{eff} of the low-order stage with a plane, which removes the local variation. Use the plane as a reference and correct the n_{eff} of the high-order stage. Make sure the corrected n_{eff} is within half order distance from the reference plane.

5. After you have one process sample on each die, plot a simple wafer map to see if there is a strange deviation from the expected wafer map.

6. Re-extract parameter for the circuit with the fitting error above our set limit (e.g., width 1 nm and thickness 0.5 nm). Use the parameter previously extracted as the initial guess. Also, limit the range of the parameter sufficiently small to improve efficiency.

7. Map n_{eff} and n_g to obtain waveguide geometry, plot the linewidth and thickness die map for visual inspection

8. Label the circuit with more than 1 nm offset from the fitted thickness plane die map. Correct the corresponding n_{eff} to push the thickness of the sample back to the reasonable thickness range.

9. Parallel the extraction procedure for multiple dies at the same time.

10. Remeasure circuits that are missing. Check if the circuit with a large mismatch in fitting has a saturated spectrum. Check if a large fitting error occurs on certain circuits on many dies.

11. Process remeasured circuits.

12. Organized data from all the process dies and plot a wafer map.

4.6 Conclusion

In conclusion, we discussed how to design a compact folded two-stage MZI that can be used to extract fabrication parameters. We applied the Restart-CMA-ES global optimization algorithm to extract multiple waveguides and DC parameters from only two optical measurements of the circuit. We illustrated how to setup the algorithm in practice to obtain the global optimum efficiently. We then mapped the fabricated geometry parameters from the extracted effective and group indices. The compact device is especially useful for process monitoring and extracting detailed wafer maps for performance evaluation and variability analysis.

The wafer-scale optical measurement and extraction of process wafer maps are very time-consuming and prone to errors. To efficiently use the PCM circuit for process monitoring on wafer scale and obtain fabricated wafer maps, we discussed the benefits of applying the step-wise workflow and parallelization to process the data. We described our workflow to extract fabricated wafer maps. In the following Chapter 5, we will present the obtained accurate wafer maps which we obtained using the extraction workflow.

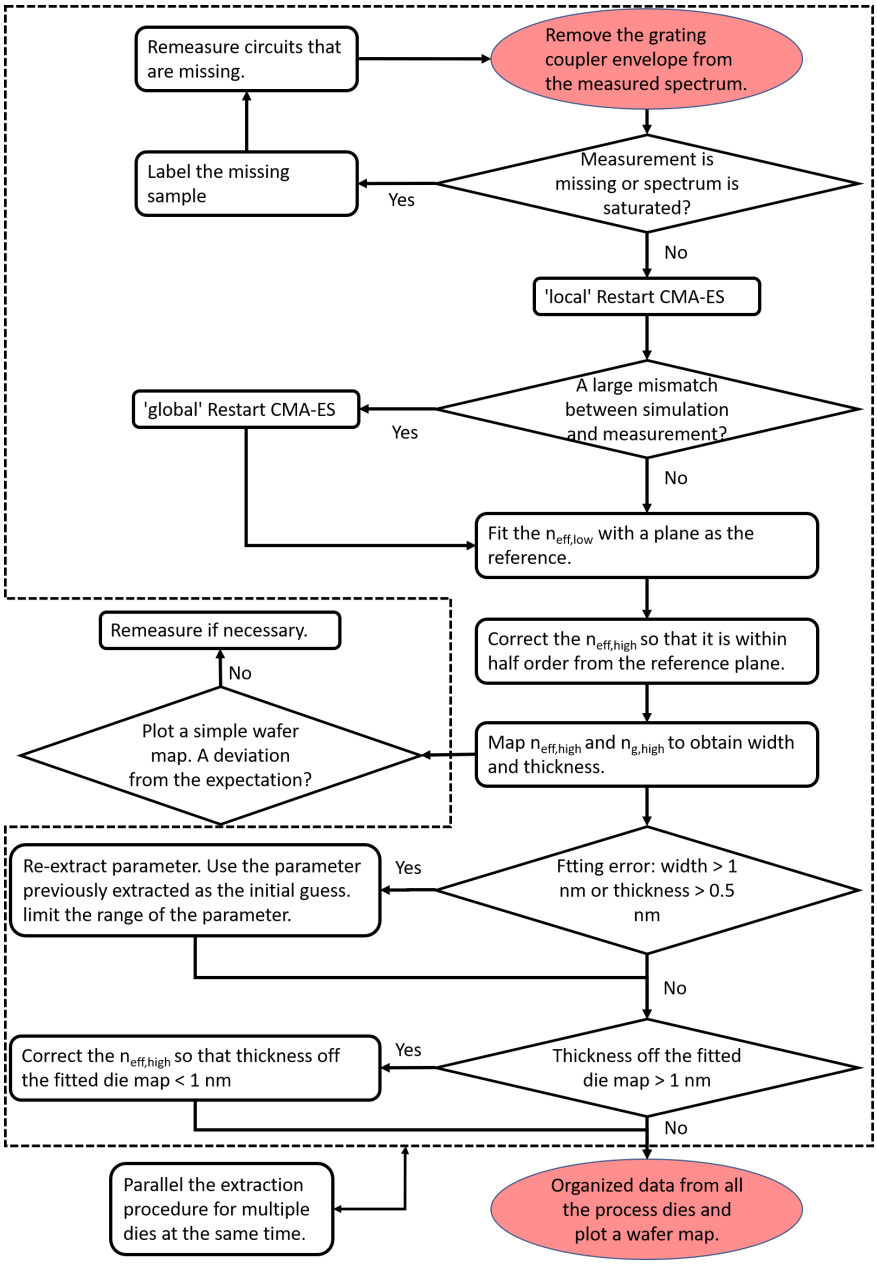


Figure 4.12: Workflow to extract process wafer map.

References

[1] Yanlu Li, Diedrik Vermeulen, Yannick De Koninck, Gunay Yurtsever, Gnther Roelkens, and Roel Baets. *Compact grating couplers on silicon-on-insulator with reduced backreflection*. Optics letters, 37(21):4356–4358, 2012.

[2] Martin Fiers, Thomas Van Vaerenbergh, Ken Caluwaerts, Dries Vande Ginste, Benjamin Schrauwen, Joni Dambre, and Peter Bienstman. *Time-domain and frequency-domain modeling of nonlinear optical components at the circuit-level using a node-based approach*. Journal of the Optical Society of America B, 29(5):896, 5 2012.

[3] Nikolaus Hansen. *The CMA evolution strategy: A tutorial*. Technical report, 4 2016.

[4] Nikolaus Hansen. *CMA-ES Source Code*.

[5] *CMA-ES*.

[6] Anne Auger and Nikolaus Hansen. *A restart CMA evolution strategy with increasing population size*. In Evolutionary Computation, 2005. The 2005 IEEE Congress on, volume 2, pages 1769–1776. IEEE, 2005.

5

Process Variation and Spatial Variability of Silicon Photonics

Silicon Photonics is very sensitive to process variation. Using the parameter extraction method described in 4, we present in this chapter an extracted granular map of process variation on a Silicon Photonics wafer. We proposed a hierarchical model to separate the layout-dependent and location-dependent systematic process variation and random process variation on different spatial levels. Using the model, we decomposed variations and found out quantitatively how each of them contributes to the total process variation. We also observed that die-level systematic linewidth variation are correlated with local pattern density. Based on our findings, we could generate a virtual fabrication wafer map based on real fabrication statistics for a realistic yield prediction that we will discuss in the next chapter. Our observation does not only reflect the quality of the fabrication process and indicate directions of improvement, but also gives a few guidelines for designers to design consciously to reduce the impact of the process variation.

5.1 Background of the Research

5.1.1 Variability in Silicon Photonics

As in Fig. 5.1, process variability presents in photonic circuits at different levels. Process variation such as exposure dose, resist age, plasma density, and *chemical mechanical polishing* (CMP) slurry composition lead to non-uniformity in device

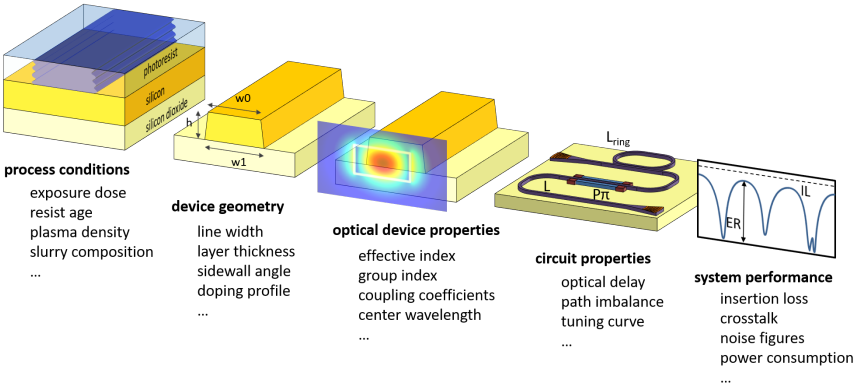


Figure 5.1: Describing variability presents at different levels. Extracted from [3]

geometry such as linewidth, layer thickness, sidewall angles, and doping profile variation. The process non-uniformity introduces variability in optical properties of a device such as effective index and group index of a waveguide or coupling coefficients of a coupler and center wavelength of a filter. The performance variation at device-level would propagate and accumulate at circuit-level such that optical delay has a random component and path imbalance presents in the circuit, which deteriorates the circuit performance, making only a fraction of circuits perform as intended. This fraction further shrinks with the increasing circuit size. In particular, wavelength filters such as ring resonators, lattice filters, *array waveguide gratings* (AWG) and so on suffer significantly from process non-uniformity leading to increasing channel cross-talk, insertion loss, power consumption and deviation of channel wavelength. On the Silicon-on-Insulator (SOI) platform, 1 nm change in linewidth can lead to 1 nm spectrum shift, while 1 nm change in thickness even leads to around 2 nm spectrum shift [1].Table. 5.1 lists within-wafer fabrication variations from various platforms. 6σ of both width and thickness variation within a wafer are usually over 15 nm. Neglecting the impact of process variation could lead to circuit failure even with tuning to compensate the error [2]. Process variation becomes a limiting factor to the high fabrication yield, i.e., the fraction of functional fabricated circuits. Without a good analysis and a practical method to predict, then mitigate its impact, process variation will increase the cost of large-volume production and limit the scale of the integration.

Process	Within-wafer variations		
	σ_{Δ_w}	σ_{Δ_t}	Fabricated Length
imec 193-nm dry lithography [4, 5]	2.59 nm	2 nm	
imec wafer-scale corrective etching [6]		0.83 nm	
imec wafer-scale corrective etching [7]		3.64 nm	
imec 193-nm immersion lithography [8]	2.53 nm		
200 mm wafer, 193-nm dry lithography [9]	0.78 nm		
300 mm wafer, 193-nm immersion lithography [9]	2.65 nm		
BAE 248-nm lithography [10]			4.17 ± 0.42 mm
IME 248-nm lithography [11]			5 mm
imec 193-nm lithography [12]		2.4 nm	
248-nm lithography [13]	2 nm		
248-nm DUV lithography process [14]	3.86 nm	1.32 nm	

Table 5.1: Literature results for within-wafer fabrication variations.

5.1.2 The Need of the Fabrication Wafer Map for Layout-Aware Variability Analysis

Variability analysis is the study to relate fabrication non-uniformity to the fabricated device and circuit performance variability. Since there is a lack of detailed wafer maps and statistics of fabrication variations, initial variability studies often assumed that fabricated geometry parameters on a photonic chip such as linewidth and thickness [15, 16] or device behavior parameters such as coupling coefficients [17] are totally random and location-independent. Also, since back then there is no way to extract detailed wafer map to analyze the statistics of fabrication and what kind of distributions the random parameters are, we often assume that the random parameters follow the normal distribution.

In practice, non-uniformity on a wafer is strongly location-dependent and layout-dependent. For example, a wafer is processed with chemical mechanical polishing so that there is a gradual change in the thickness profile across the wafer. In this case, the thickness variation is determined by its location on the wafer or in a die. Also, the conditions of the lithographic process are correlated for structures nearby. As we will show in Section 5.6, the width variation is correlated with the pattern density around the design, which is the average density of the layout nearby. So the actual layout on the chip influences the process variation and also contributes to a systematic variation repeated on each die. These layout-dependent and location-dependent variation have similar impact on the adjacent structures and circuits on-chip. These structures will suffer correlated process variation and behave similarly. In summary, process variation consists of both location-dependent and layout-dependent systematic contributions and random contributions that are independent of location.

A realistic variability analysis and yield prediction should incorporate both the spatially correlated deterministic variation and the random variation [3]. Recently,

studies shift to layout-aware yield prediction that considers the spatial correlation of process variability. Like studies described in [14, 18], in next chapter, we can generate a virtual wafer map to emulate the real process variations. Circuit parameters are revised by its location so that its performance is also influenced by position. Using the Monte-Carlo simulation that locates circuits on various wafer positions, we could calculate the performance variability and estimate the yield of the circuits.

This missing link here is how to generate a realistic virtual wafer map. To generate it, we need to do two things: First, obtain an accurate and granular wafer map of the process variation. Second, build a realistic model that decomposes variation into sensible categories to makes sense of the large and noisy data set of the fabrication map.

5.1.3 Method We used to Obtain the Fabrication Wafer Maps

Location-dependent variability analysis requires an accurate and granular fabrication wafer map to offer rich details for analysis. Linewidth and thickness are the most interesting in variability analysis because they are fundamental parameters that reflect the fabrication quality and determine the behavioral variability of fabricated passive circuits. Getting a linewidth or thickness wafer map is not trivial.

As described in Chapter 2 and 4, we are able to use automated optical measurement to extract geometry parameters on wafer-scale. We can obtain behavioral parameters such as effective index and group index from the measured spectral transmission of interfering structures. Then, we can map behavioral parameters to linewidth and thickness. The wafer maps we analyzed in this chapter are derived by the compact circuits and the workflow illustrated in Chapter 4. With the compact design, we should be able to distribute the compact monitoring circuits over the wafer to extract multiple process parameters and generate detailed wafer maps with a sub-nanometer accuracy.

We present a granular wafer map of process linewidth and thickness variation. To make sense of the data, we will present in the following section a hierarchical model to separate process variation into systematic and random contributions on different spatial levels. We will also describe a procedure to separate spatial variations, and the result qualitatively and quantitatively shows interesting features of each contribution. We observed a moderate correlation between intra-die systematic linewidth variation and local pattern density.

5.2 Classification of Process Variation

5.2.1 Environmental, Temporal, and Spatial Variations

Variation is the difference in a parameter between its designed value and fabricated value. Process variability can be environmental, temporal, or spatial. Environmental variability is the change in the operating conditions such as temperature variation within chip led by varying thermal profile, variation in surrounding temperature, power supply voltage fluctuation and even cosmic radiation variation. Temporal variability is the parameter change over time. It includes nano-second effect such as the SOI history effect and self-heating effect, to year-long effect such as dielectric material deterioration. Spatial variability is the non-uniformity that depends on circuit location on the wafer or the distance between two circuits. Environmental and temporal variability is often addressed by reliability models while the spatial variability requires a statistical model that is the focus in this research.

5.2.2 Systematic Variation and Random Variation

Process variation consists of systematic variation and random variation. Systematic variation, or deterministic variation, denotes a repeatable deterministic pattern that can be determined by a circuit's location. On the wafer level, the slow-varying profile of process variation over the wafer such as chamber effects and the slow-varying plasma density distribution that contributes to deterministic across-wafer patterns. It can also be repeated pattern on every die induced by a non-ideality in the photomask and optical proximity effects. The systematic variation can also be due to non-ideality of the lithographic system such as defocus and misalignment; and pattern density-dependent effects such as CMP.

Random variation or stochastic variation refers to stochastic residuals after removing the systematic part from the raw data. It is unpredictable and random in nature, and it cannot be determined by its location. Non-uniformity such as macro-scale random fluctuation in the fabrication process such as exposure dose from die-to-die, to micro-scale difference such as random linewidth led by sidewall roughness [19] and photoresist granularity [20]; and atomic-scale oxide-thickness variation [21].

Systematic and random variation impact the device and circuit performance in different ways. Systematic variation is determined by the device or circuit locations on a wafer. It contributes to the spatially related difference on the wafer. It directly adds to a spatial correlation among devices. On the other hand, random variation is independent of circuit position and from the modeling point of view, it is usually treated as independent fluctuation. In practice, the classification of systematic or random variation is not absolute. Often, the location of the design or

spatially related process features (such as local pattern density) is not available to the designers. So, it is impossible to estimate the variation with certainty. In such cases, systematic variations are treated as random. In addition, as we proposed to separate variation on different spatial levels, the systematic variation is supposed to be fitted perfectly by the proposed analytic model, the random variation is the residual after fitting the total variation with the model. However, the mismatch between the model and the real systematic variation can lead to an overestimated random variation. For example, we could assume that on the wafer level, the systematic width variation is radial symmetrical that can be described by a bi-variate polynomial. If the real variation cannot be fully described by the polynomial, when we decompose the variation, the mismatch would partly contribute to the random variation at the die level.

5.3 Hierarchical Spatial Variability Model

5.3.1 Physical Origins of Different Variations

The spatial variability is the non-uniformity that depends on circuit location on the wafer or the distance between two circuits. As shown in Fig. 5.2, a process-related device parameter has its variations from sources at different spatial levels. Lot-to-lot fabrication suffers a variation from tool drift, resist aging, or a change of wafer supplier. Wafer-to-wafer variation is mainly led by tool priming, the difference in the layer thickness of wafers and non-uniformity in the chamber environment. Wafer-level non-uniformity originates from layer thickness, photoresist spinning effects, and plasma distributions. Fluctuation in exposure dose and imaging focus add to the random die-to-die variation. On the die level, low-frequency change in layer thickness, local pattern density, and error in the photomask lead to a systematic pattern that repeats on each die. Also, intrinsic randomness in layer thickness and waveguide sidewalls raise the device-to-device random variation. Besides, the wafer-die interaction accounts for deformation around wafer-edge or other effects we will discuss in this chapter.

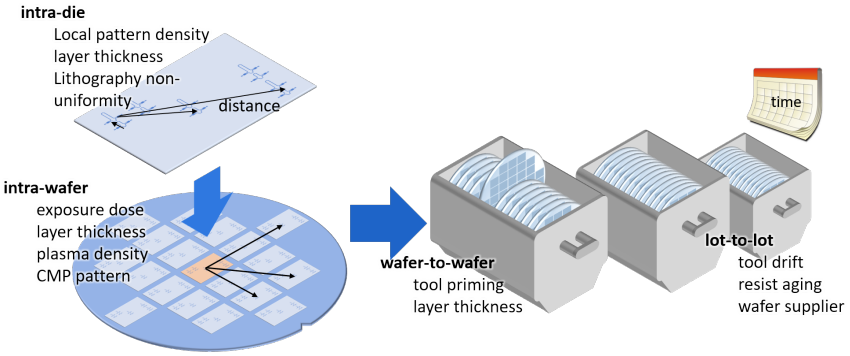


Figure 5.2: Illustration of sources of process variability on different spatial levels.

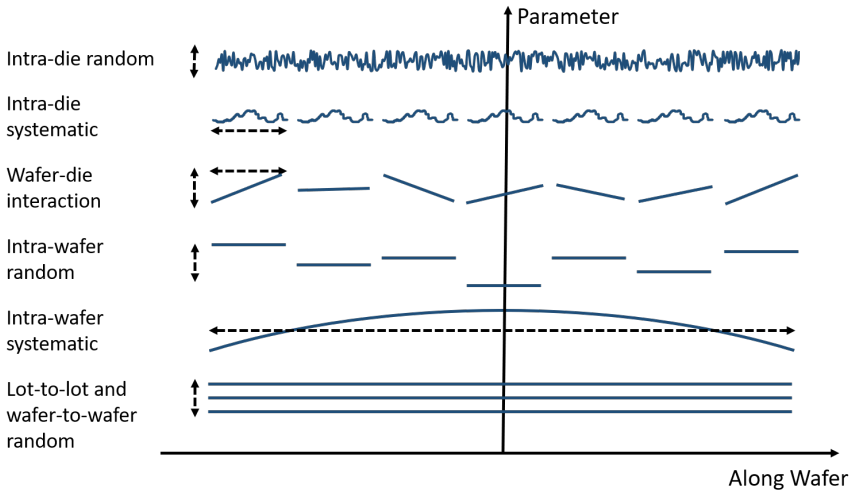


Figure 5.3: Illustration of decomposing spatial variability of device parameter at different levels. The hierarchical model is based on [22].

Since physical origins of process variation work on different spatial scales quite independently, process variation allows for a spatial decomposition. Based on observation of the extracted wafer map, we revised spatial models described in [22] and [23], and proposed an additive hierarchical spatial variability model of fabricated integrated photonics as shown in Fig. 5.3. The model decomposes total spatial variation into lot-to-lot, wafer-to-wafer, die-to-die and device-to-device variations with systematic and random components [22]. We also include a wafer-die interaction in our variability model that helps to reduce the random contribution

in the variation. It captures a die-level variation that is highly correlated within a die, but the correlation varies randomly with the die location on the wafer. The decomposition helps to identify the origin of process variation to improve the fabrication and facilitates the generation of faithful virtual fabrication maps for yield prediction.

5.3.2 Lot-to-Lot and Wafer-to-Wafer Variation

In the state-of-the-art semiconductor fabs today, manufacturing involves various batch processes that apply to multiple wafers together to increase the one-time wafer throughput. For examples, the *chemical vapor deposition* (CVD) heats several wafers in the furnace with reactive gas to deposit a thin layer on the surface of wafers. [24] The batches are usually referred to as lots. One lot conventionally contains 25 wafers, and wafers in a lot are processed with the same process condition. Between lots, there would be a change in the process condition which leads to a lot-to-lot variation. Within one lot, tool priming and non-uniformity in the chamber environment bring the difference between wafers in one batch, which causes within-lot variation. In processes such as lithographic imaging and *reactive ion etching* (RIE), each wafer is processed individually. Naturally, there is a wafer-to-wafer variation. In theory, determining the systematic signature of lot-to-lot and wafer-to-wafer variations requires time-series models [24] and long-term monitoring of lots of wafers. Due to the complexity of the process involved in the processing, it is challenging to separate systematic and random variation. In practice, without losing generality of the further discussion, we assume variability above wafer-level is random and follows the normal distribution.

$$V_{LTL} + V_{WTW} = \mathcal{N}(0, \sigma^2_{LTL,WTW}) \tag{5.1}$$

5.3.3 Intra-wafer systematic variation

Across the wafer, there are many steps that lead to the process variation of circuits. Interestingly, these non-uniformities follow a radial pattern by the nature of the equipment. Many processes followed a 'center-fed' or 'edge-fed' style which has different boundary conditions near the edge. For example, *post-exposure bake* (PEB) in the lithography smooths the standing wave induced roughness on the sidewall. PEB temperature is higher in the middle of the wafer and decreases outwards [25]. Similarly, chamber wall conditions also cause etching rate non-uniformity. The second cause for the radial pattern in variation is the rotation of the wafer. Wafers are rotated during some processes to reduce non-uniformity [24]. Errors in the position and rotation of the wafer stage during exposure increase from the wafer center outwards. Similar patterns are also in variations in wafer stage vibration and the distortion of the wafer with respect to the exposure pattern.

We also observed in several wafers that width variation shows radial pattern over the wafer. It means that the etching rate in the etching suffers non-uniformity. One explanation can be that in the center of the wafer, patterns that consume more plasma than near the edge, because outside the wafer there is no pattern to etch. The decreasing plasma concentration in the center leads to slower etching rate, therefore broader waveguide linewidth. An alternative reason for radial plasma distribution is the field profile of the magnetic RF fields that heat the plasma: the heating is not perfectly uniform. These variations across the wafer produce the systematic intra-wafer variation that varies slowly across the wafer. The variation exhibits a symmetric radial pattern, which often has a dome-like or bull-eye shape. We can describe the systematic variation by a bi-variate polynomial as:

$$V_{IWS} = \sum_{i+j\leq 2 \& i,j\geq 0}^{i,j} p_{ij} \cdot x_w^i \cdot y_w^j \tag{5.2}$$

5.3.4 Intra-Wafer Random Variation

On the wafer level, there is also random variation led by the exposure fluctuation, depth of focus change in the lithographic imaging from die to die. We define it as the averaged randomness over a die, so within a die, structures endure the same amount of *Intra-Wafer Random* (IWR) variation. Since when a step moves to the next location, lithography process patterns a die layout, the linewidth of the pattern suffers more significant IWR variation than parameters like thickness. We model IWR as a normal distributed random variable:

$$V_{IWR} = \mathcal{N}(0, \sigma_{IWR}^2) \tag{5.3}$$

5.3.5 Wafer-Die Interaction Variation

Wafer-die interaction captures differences in intra-die variation as one move from one die to another across a wafer. [23] For example, lithographic imaging does not only fluctuate in the exposure dose but also varies on the incident angle on the wafer. The slightest deviation from a vertical incidence on a die would lead to an exposure dose profile that is location-dependent. As a result, on top of the IWR variation that is the same within each die, there is also a variation that is determined by its location in the die. However, unlike IDS variation, the actual dependency is a result of the process parameters such as the angle of incidence in the lithographic imaging which might vary randomly from die to die.

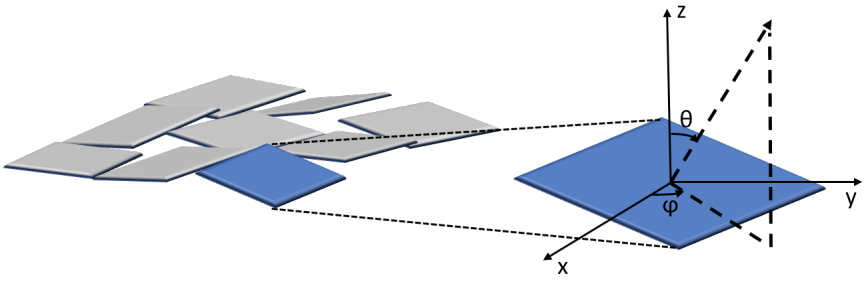


Figure 5.4: Illustration of the wafer-die interaction. After removing the systematic variation on the wafer and the die level, the residuals on each die can be fitted by a plane, which is the WDI variation. The normal of each plane pointing upwards and has a little deviation from the zenith direction. The normal can be described by inclination θ and azimuth ϕ.

In principle, WDI variation can also be systematic where the die variation may be attenuated or accentuated depending on its location on the wafer. For example, the edge of the wafer is often less uniform and less controlled. Consequently, the die variation on edge might be significantly worse than near the center. In our study, we did not observe in our data the systematic variation in WDI. We found that the random WDI on each die can be adequately represented by a plane as in Fig. 5.4. The plane as be expressed as:

$$V_{WDI} = f(x_d, y_d) = a * x_d + b * y_d \tag{5.4}$$

$$\theta = arctan(\tfrac{b}{a}) \tag{5.5}$$

$$\phi = arccos(\tfrac{-1}{\sqrt{a^2+b^2+(-1)^2}}) \tag{5.6}$$

where the a and b are coefficients in the plane equation, phi and θ are the inclination and the azimuth of the normal of the WDI plane.

5.3.6 Intra-Die Systematic Variation

Variability from die-to-die and wafer-to-wafer were the dominant sources of variability in the IC industry. [26] As an increasing level of the integration accommodates more components, the field size has to grow gradually. Increasing die size reduces the controllability of the fabrication process on the die level and increases the intra-die variability that becomes more significant.

A significant component of the intra-die variation is systematic. It originates from the layout and pattern-dependent errors on the die level, such as an error on the photomask. Also, the CMP rate is related to the pattern density which varies the oxide thickness on wafer [27]. Fabs add dummy samples to patternless space to mitigate the non-uniformity in pattern density to reduce thickness variation. We

also observed the significant correlation between the pattern density and the *intra-die systematic* (IDS) width variation. We will discuss such correlation in 5.6. In theory, if we can trace back the IDS variation to all of its physical origins and study how they correlated, we can derive the IDS as a function of the parameters $f_i(x_{die}, y_{die})$ that describes the physical origins. In practice, due to the limited access and knowledge on the process, we can explain part of the correlation with causes such as the pattern density. We could treat the part cannot be explained as a correlated random variation with a correlation length. Then, we imitate the variation by a coherent noise map with a right correlation length l. The IDS variation can be expressed as:

$$V_{IDS}(x_{die}, y_{die}) = \sum^{i} p_i \cdot f_i(x_{die}, y_{die}) + CoherentNoiseMap(l) \quad (5.7)$$

where (x_{die}, y_{die}) are coordinates of the location on the die; p_i maps the impact of the parameter $f_i(x_{die}, y_{die})$, such as pattern density over the die, to the IDS variation $V_{IDS}(x_{die}, y_{die})$.

5.3.7 Intra-Die Random Variation

Intra-die random (IDR) variation is the device-to-device disparity or local mismatch. IDR variation includes intrinsic variability like atomic oxide thickness variation, *line-edge roughness* (LER) due to photoresist granularity. The IDR variation dominates at the sub-micron level while the size of a photonic device is beyond several tens of micron. The distance between devices is far beyond the correlation length of the IDR variation. We can regard then each device suffers IDR variation independent of their location. We can model IDR variation as a normal distribution independent to the other variation components:

$$V_{IDR} = \mathcal{N}(0, \sigma^2_{IDR}) \quad (5.8)$$

5.4 A Workflow to Separate Spatial Variations on Different Levels

After we have the hierarchical model, the next step is to separate levels of variation from the measured or extracted data on a fabricated wafer. There are a few methods in literature to separate variations of CMOS integrated circuits. [28] uses filtering, spline, and regression-based approaches to separate wafer-level effect. Then, they separated the die-level effect by the spatial Fourier transform method. This method requires a good choice of parameters in the separation procedure, which is neither easy nor intuitive and sometimes subjective. Also, the extracted systematic variation is an interpolated map which is not analytic nor can be related to process

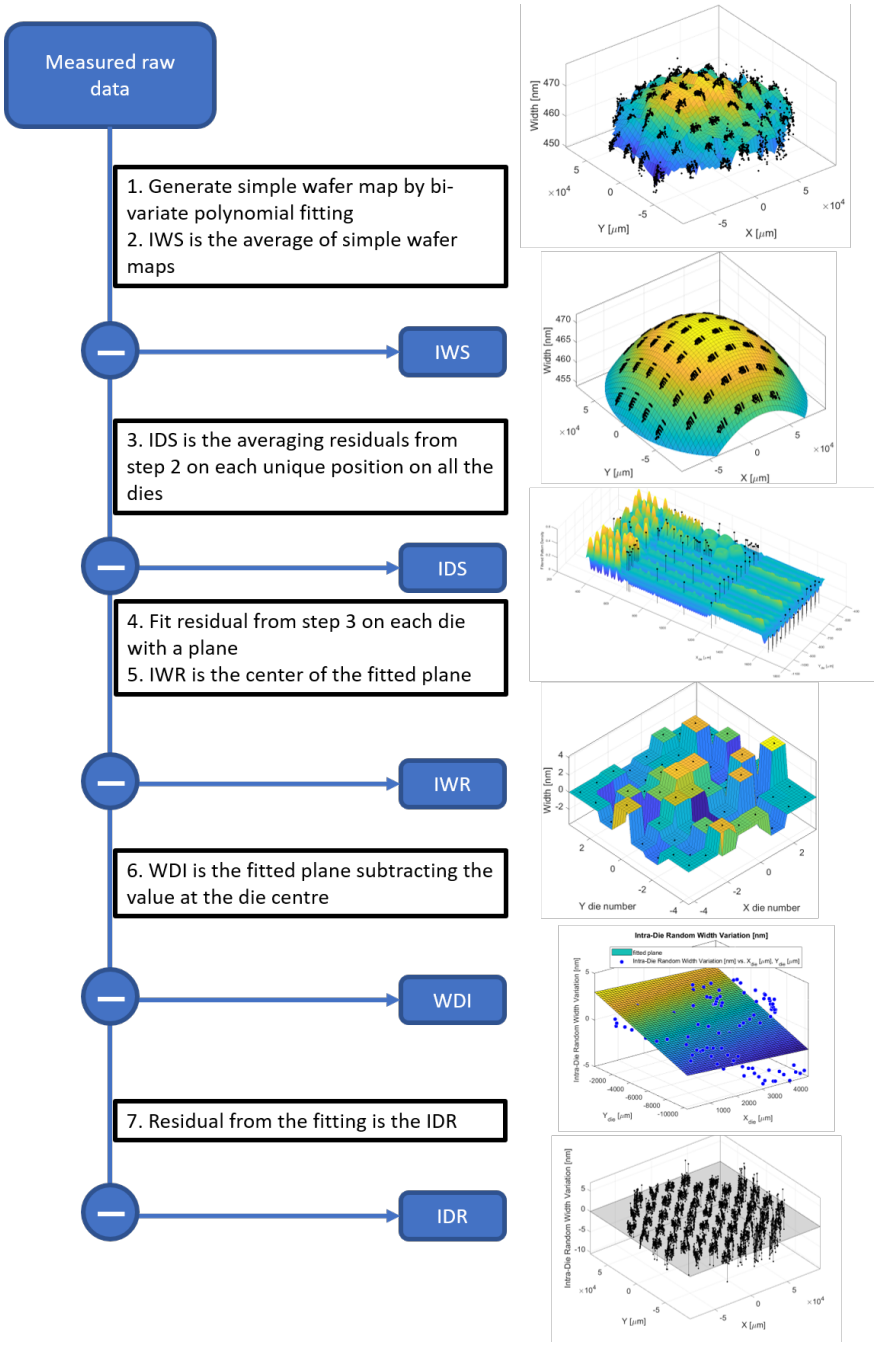


Figure 5.5: Proposed workflow to decompose variation on different spatial levels.

parameters immediately, making it difficult to apply the extracted model for yield prediction. [29] found in CMOS fabrication, systematic variations on intra-wafer and intra-die level can both be described by parabolic function. From our observation on the measured photonics wafer, we cannot simply describe the IDS variation as a parabolic function. The IDS variation is largely dependent on the pattern density that is not parabolic. Also, we observed that WDI variation introduces a within-die variation that cannot be described by IWR variation.

Based on the hierarchical spatial variability model and observation of our measured data on the fabricated map, we will introduce below a workflow to separate systematic and random variations.

5.4.1 Intra-Wafer Systematic Variation

First, we separate the intra-wafer systematic variation. IWS variation is the slow varying radial symmetric non-uniformity across the wafer. There are two ways to derive the IWS variation: First, fit all the data on the wafer with the IWS model. Second, fit a simple wafer map for each unique sample on a die, and then get the IWS wafer map by averaging simple wafer maps. The first should work even with the presence of lower-level variations in the raw data. The low-frequency profile should filter out the high-frequency variation on low levels and will not induce a significant error in estimation. However, the second method gets the IWS map by averaging simple wafer maps is preferred. In case we have more samples over the wafer for a certain die location, we would not overweight the location to obtain the IWS map.

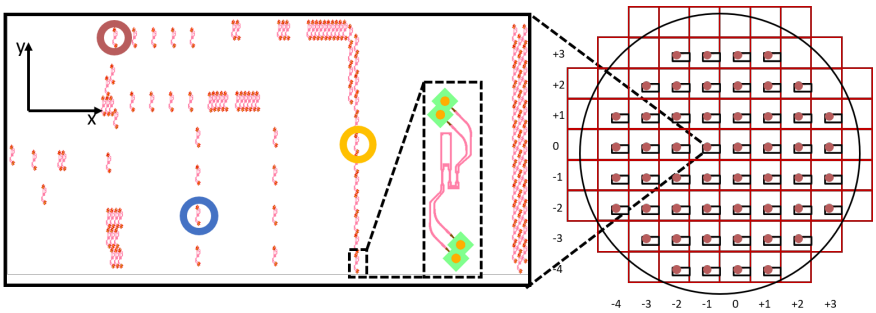


Figure 5.6: The illustration of devices on the die. Red, blue and yellow circles represent three unique samples on the die. Red circles on the wafer map represent locations of the sample on the wafer.

Using the second method as shown in Fig. 5.7, we take devices at an unique location (Fig. 5.6 Left) in a die over all of dies on the wafer (Fig. 5.6 Right), and fit a simple parameter wafer map using a second-order bi-variate polynomial in (5.2),

which averages out the intra-wafer and intra-die random variation. Then, average all the simple wafer maps to derive the IWS variation which is also a bi-variate polynomial.

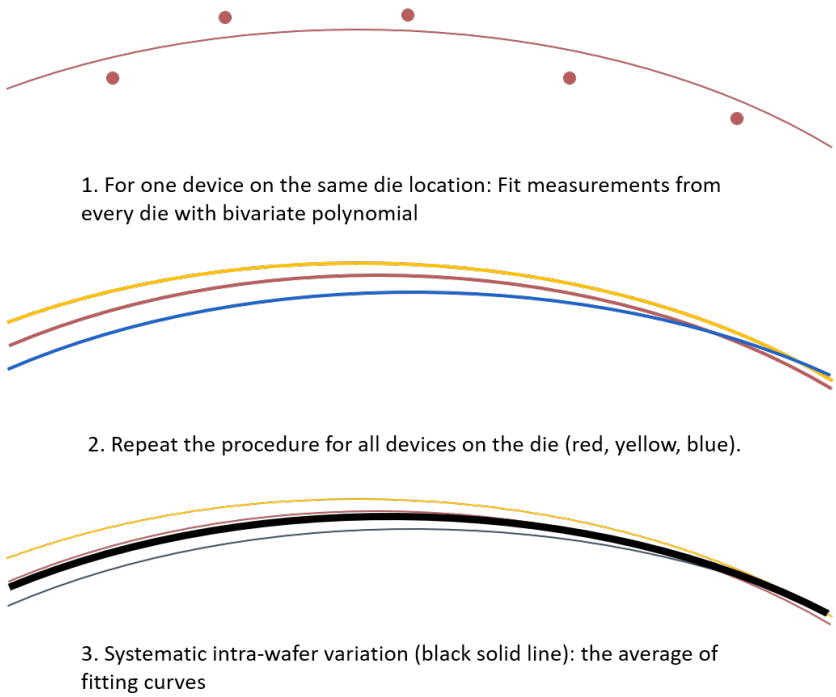


Figure 5.7: The workflow to decompose IWS variation.

5.4.2 Intra-Die Systematic Variation

The second step we separate the IDS variation. We removed the IWS from the raw data. In the residual, there are contributions from IWR and IDR which both follow the normal distribution with a zero mean. As defined in Section 5.3.5, WDI on each die is a plane with a small random deviation from horizontal. By definition, WDI variation also has a zero mean. Therefore, IDR, IWR, and WDI can all be averaged out over all the dies on the wafer. As shown in Fig. 5.8, at each unique die position, we average the data over all the dies. The randomness in the residual from the last step is averaged out, and we could obtain the IDS that is the repeated contribution on each die. Fig. 5.8 shows that the average offset between the raw data and the IWS map gives the IDS variation.

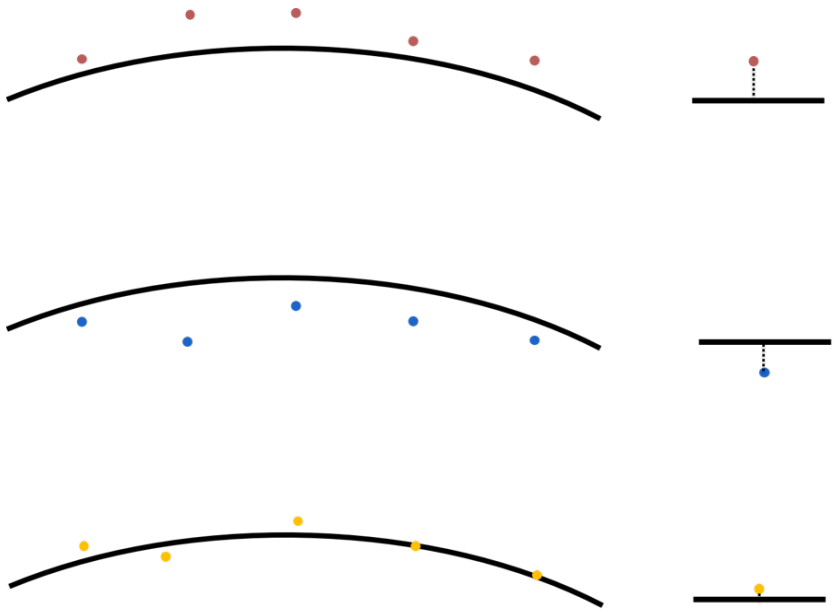


Figure 5.8: Left: Black solid curves indicate IWS variation. Points with the same color are parameters measured on the identical locations on different dies. Right: The average offset between measured parameter and IWS variation is the IDS variation.

5.4.3 Intra-Wafer Random Variation, Intra-Die Random Variation, and Wafer-Die Interaction

We remove the IDS from the data. The residual from the first two steps is the sum of IWR, IDR, and WDI.

IWR is the die-to-die randomness and has identical values within a die. Device-to-device random IDR variation can be averaged over a die. WDI is a plane with a zero mean as defined, so it has centrosymmetry around the die center. WDI can also be averaged out over a die. So we can average all the residuals on a die to derive IWR variation of the die. However, since we only have limited samples on an uneven sampling grid, the IWR variation derived by averaging the residue is biased by the sampling.

A better way is to derive IWR variation is to fit the residual with a plane. As shown in Fig. 5.9, the center of the fitted plane is the IWR variation, which is the unbiased average of the residual. The fitted plane subtracting the IWR leaves the WDI variation with a zero mean. The remaining residual is the offset from the plane of the fitting, which is the IDR variation.

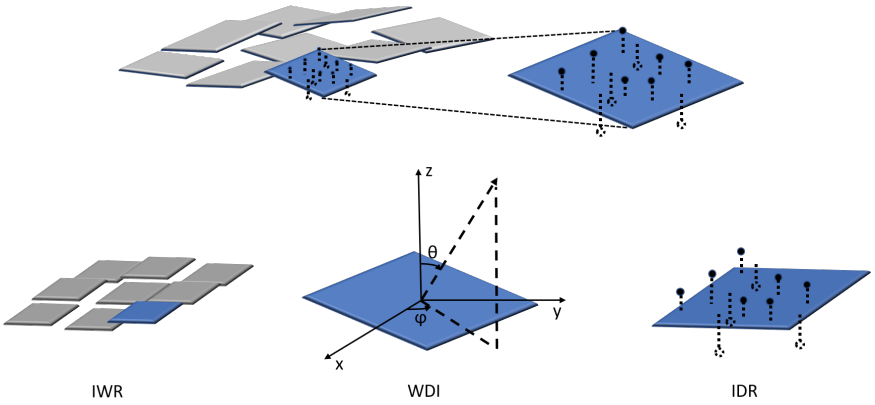


Figure 5.9: The workflow to decompose IWR, WDI, and IDR variation using the fitted plane. The center of the fitted plane derives the IWR of the die. The fitted plane subtracting the IWR is the WDI of the die. WDI has zero value at the die center. IDR is the residual of the fitting.

5.4.4 Evaluation of Decomposed Data

The purpose of variation decomposition is to understand how a process condition such as polish, pattern density leads to the wafer and the die topography variation qualitatively and quantitatively. Visual inspection on topography across the wafer gives an insight into the uniformity of the polishing and effect of the slow-varying plasma concentration. The die-level pattern could also be inspected visually that offers an indication of the layout and pattern density related non-uniformity. We modeled the wafer level systematic variation as a low-frequency bi-variate polynomial while modeling the die-level systematic variation as a function of the layout factors for detailed understanding. The focus of the wafer-die interaction should be on minimizing the wafer- and die-level variations which also helps to identify the remaining systematic variation. The residual is random noise, and we should ensure the layout-correlation is low by correlation examination, to gain confidence that the systematic sources of variation have indeed been taken into account. In many pieces of research on CMOS variability, they neglected the wafer-die interaction since it has a negligible contribution to overall variation. But in our observation, we need to pick it up from the residual since we found a clear and significant pattern in it. So, it is necessary to do a correlation analysis on the "random" residual. We can add further terms in our model if a new pattern can be well observed and understood in the residual.

5.5 Results and Analysis on a Silicon Photonics Wafer using 193 nm Lithography

In this section, we applied the model and use the workflow to analyze the detailed wafer maps we extracted from a 200 mm silicon photonics wafer. We will mainly focus on two fundamental process parameters: linewidth and thickness. The analysis confirms the validity of the hierarchical model and the proposed workflow. The result helps us to understand the process variation of different spatial scales quantitatively.

5.5.1 Specifications of the Monitoring Circuit

To get process parameters such as linewidth and thickness, and behavioral parameters of the DC, we used the folded two-stage MZI described in Chapter 4 as our monitoring circuit. Our design was fabricated by imec's deep UV lithography on a 200 mm Silicon-on-Insulator wafer. The wafer consists of 52 complete dies. We scattered 117 duplicates of the PCM circuit (Fig. 5.11) on a $5mm \times 10mm$ block in each die with a die size of $21.84mm \times 21.84mm$. We distributed some circuits densely (80 μm horizontally and 400 μm vertically) in a mesh that consists of several horizontal and vertical arrays to identify the correlation length of the process variation. We also include some sparsely distributed circuits to capture the long-range systematic variation.

We measured the wafer in our cleanroom using an automatic optical measurement setup (Fig. 5.10). Before and after the wafer measurement, we calibrated the tunable laser by CO_2 and NH_3 gas cells. We also did a stability test to make sure the laser did not drift during the several-week-long measurement. More details about calibration and stability test of the tunable laser can be found in A.

After setting up an alignment reference, the machine would move and align input and output fibers to the top of grating couplers of a PCM according to coordinates of two GCs extracted from the .gds file of the design. We measured all the PCM circuits on the wafer in the controlled environment of 25 degree Celsius. Optical transmission of the circuits is measured from 1500 nm to 1600 nm. As explained in 4.3, we need at least 500 wavelengths over the spectrum to extract circuit parameters from curve fitting. So, the spectral resolution can be as low as 200 pm. On the other hand, since we want to reduce the measurement time of the wafer, we want to lower measurement resolution which means less time spent to sweep laser. The lowest sweeping step of the laser can be correctly set is 20 pm. In the case, each measurement contains 5000 wavelengths in the spectrum and takes around 2 minutes, which is limited by the sweeping of laser wavelength and the alignment to the coupler. The whole wafer measurement consists of 12168 measurements on 6084 samples. It takes a minimum of 17 days to complete the

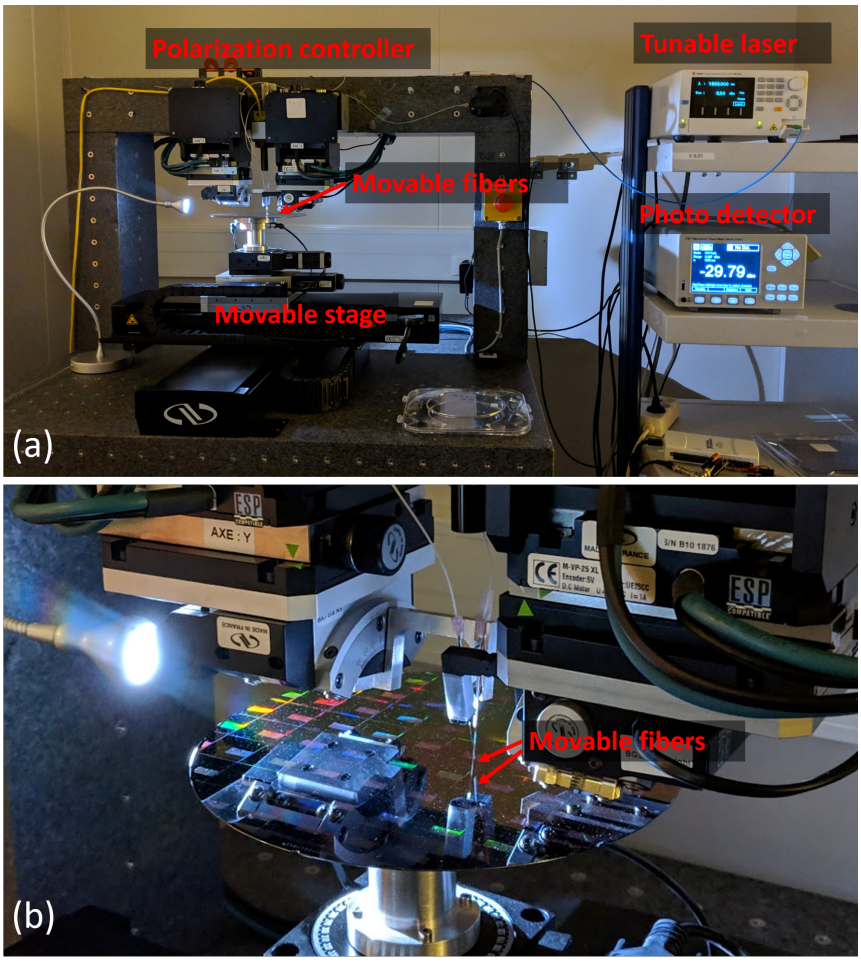


Figure 5.10: (a) Automatic setup in the cleanroom. (b) A close-up of the wafer under wafer-scale measurement.

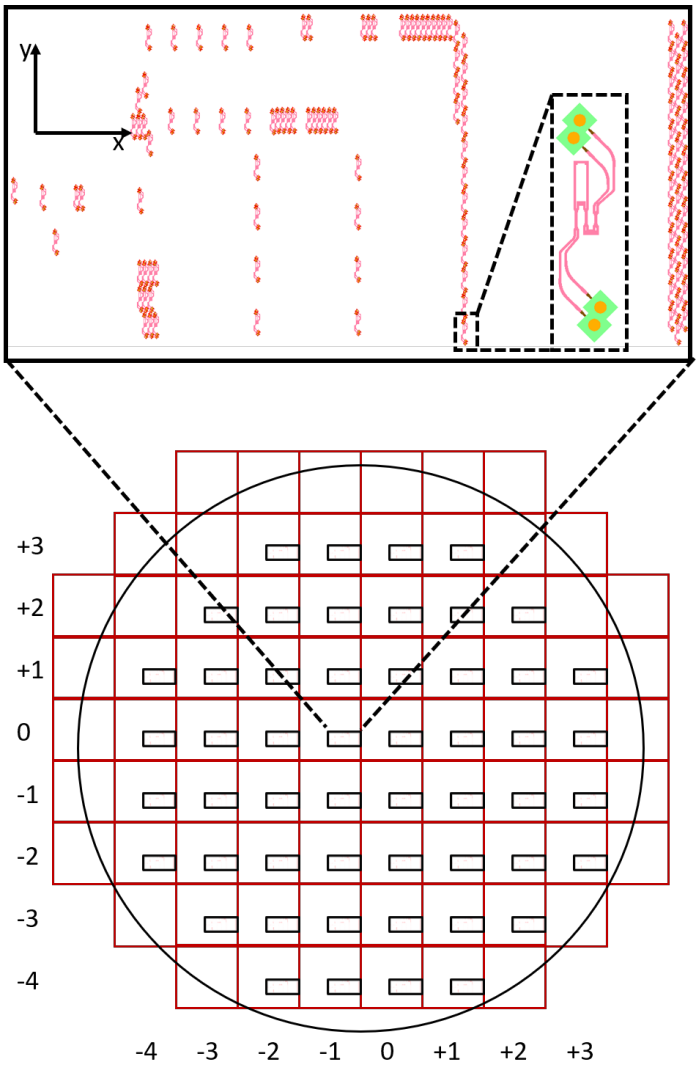


Figure 5.11: Location of the monitoring circuits on the die and the wafer. Inset is the layout of the cascaded MZI monitoring circuit.

wafer measurement. The time of the wafer-scale measurement can be greatly reduced with improvement on the measurement hardware. A tunable laser with fast wavelength sweeping and alignment instruments such as hexapods piezo-actuators with fast control loops can decrease the one spectral measurements from minutes to seconds.

5.5.2 Results of Decomposed Variations

5.5.2.1 Raw Data

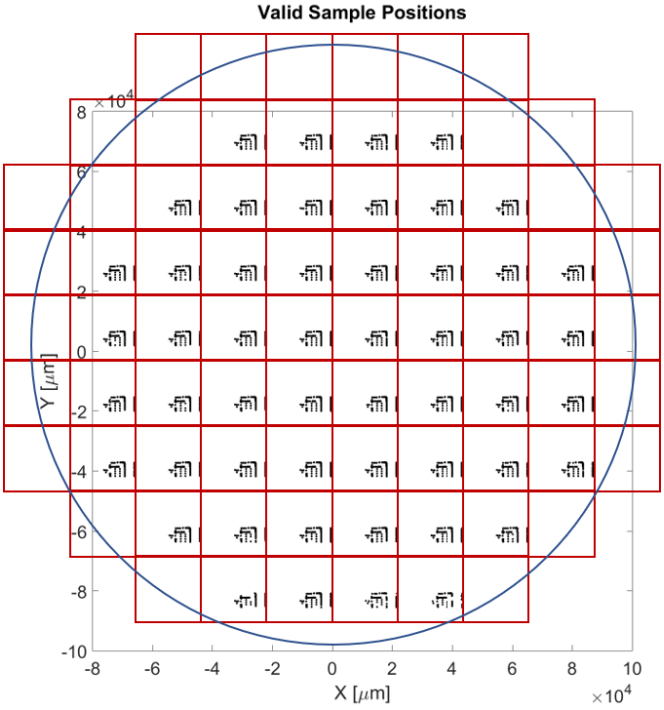


Figure 5.12: On-wafer positions of valid samples those have maximum width fitting error of 1.00 nm and thickness fitting error of 0.50 nm.

We processed the optical measurements with the technique described in 4. Among all 6084 samples on the wafer, we got 5841 valid samples. Some samples are missing one or both optical measurements, which may occur due to unsuccessful alignment in the automated process. Some optical measurements have a bad estimation of the dynamic range which results in saturation in the spectrum. There are also measurements fitted with relatively large fitting errors due to very poor alignment. After removing those 'invalid' samples, we got 5841 valid samples

(Fig. 5.12). We set the standard of valid samples with the maximal fitting errors for width and thickness are 1.00 nm (Fig. 5.13 (a)) and 0.50 nm (Fig. 5.13 (b)) respectively. Among the valid samples, the average fitting error is 0.15 nm for width and 0.08 nm for thickness.

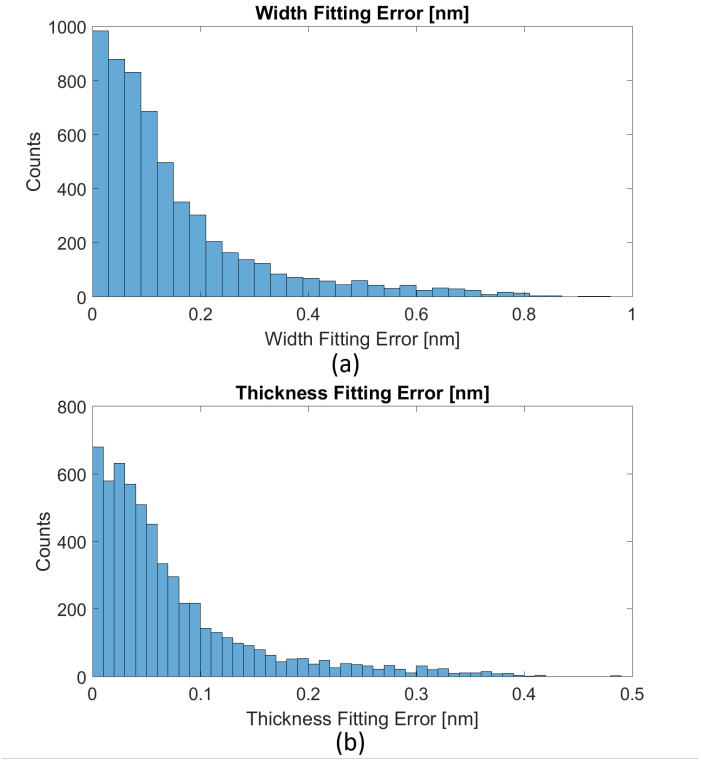


Figure 5.13: Histogram of the fitting error on the extracted linewidth and thickness. (a) The maximum fitting error of linewidth is 1.00 nm. (b) The maximum fitting error of thickness is 1.00 nm.

Derived interpolated wafer map in Fig. 5.15 is a granular wafer map obtained for processed silicon photonics wafer. We observed that the linewidth tends to be wide in the center, and it narrows down towards the perimeter of the wafer. The trend of width variation follows a dome-like shape. The average width is 464.7 nm, where the target value claimed in the technology handbook of the run is 470.0 nm. The measured width has a standard deviation (SD) of 4.6 nm. The maximum value on the wafer is 476.0 nm, while the minimum is 450.8 nm. The thickness of the wafer varies like a slope with its left-bottom leaning towards right-top. On the edge of the wafer, the change in thickness reveals some abruptness may due to

the imperfection of the polishing near the wafer perimeter. The average thickness is 210.3 nm where the target value is 215.0 nm. The measured width has an SD of 0.8 nm. The maximum value on the wafer is 214.3 nm, while the minimum is 208.4 nm.

Fig. 5.14 displays the histogram of both parameters. Obviously, both distribution are not simply normal distributions.

	Width	Thickness
Mean [nm]	464.7	210.3
Standard Deviation [nm]	4.6	0.8
Max [nm]	476.0	214.3
Min [nm]	450.8	208.4
Max-Min [nm]	25.2	5.9

Table 5.2: Statistics of measured width and thickness.

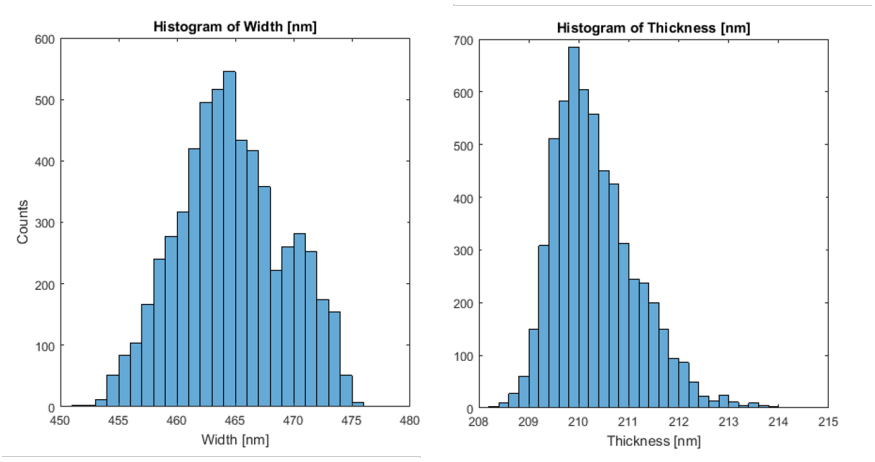


Figure 5.14: Histogram of measured width and thickness over the wafer.

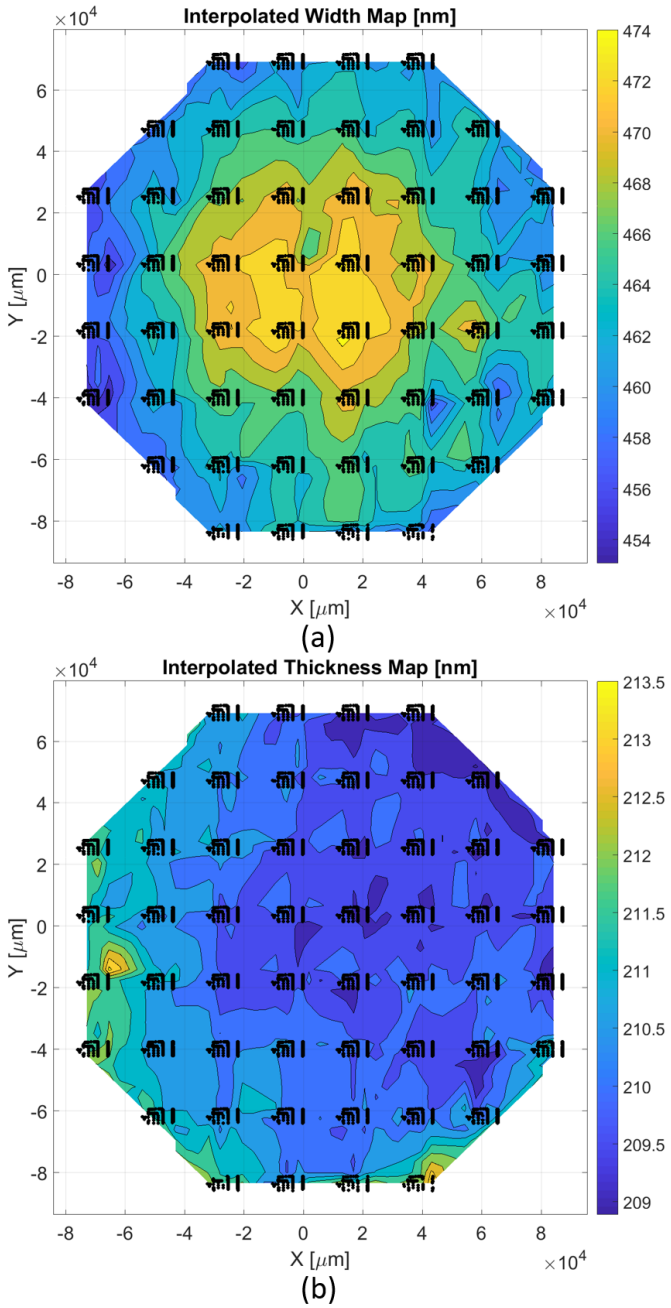


Figure 5.15: Interpolated wafer map of (a) linewidth and (b) thickness. Black dot indicates the site of samples.

5.5.2.2 Intra-Wafer Systematic Variation

Since we did not have multiple wafers to separate the variation above wafer level, here we did not separate the design value and variation above the wafer from the raw data. We put all these contributions on the wafer level. On the intra-wafer level, exposure dose, layer thickness, plasma density or CMP pattern has a low-frequency non-uniformity and lead to a gradual change in width and thickness profile. We fitted the raw extraction data of each unique device on every die with a second-order bi-variate polynomial, which gives a simple wafer map. Then, we averaged 117 simple wafer maps to get the IWS wafer map. As shown in Fig. 5.16 (a), IWS width variation shows a quite symmetric radial pattern with its center near to the center of the wafer. The Equation 5.2 can describe the dome-like shape. Coefficients of the polynomial are displayed in Table. 5.3

The very radial systematic width IWS contour might be a result of the plasma concentration profile over the wafer. Linewidth is narrow under a high local plasma concentration thus fast etching rate. In the wafer center, patterns fill up the chip and consume plasma, lowering the plasma concentration. So, in the center of the wafer, the etching rate is slow, and the linewidth is wide. Towards to rim, the plasma concentration is high because there are no patterns to etch outside the wafer. So, the plasma concentration should have a reverse dome-like profile, which in turn lead to a dome-like width variation profile. The IWS variation contributes to a significant of width variation. The maximum difference led by IWS variation among measured samples is 16.4 nm, consisting of 65% of the total variation. If we use the polynomial model to calculate the IWS width on all 52 complete dies, the maximum difference is 19.4 nm.

IWS variation of thickness (Fig. 5.16 (b)) is like a slanted plane leaning from the left bottom towards the right top. The polishing process might be the reason that shapes the trend. Also, the small variation in the cut of Si layer in the SOI production might cause a variation in layer thickness. The IWS variation counts for a maximum of 2.6 nm difference among measured samples, which is 44% of total thickness variation. The model calculates that the maximum difference over 52 dies is 3.2 nm.

	Width	Thickness
p_{20}	-1.90·10^{-0}	1.68·10^{-10}
p_{11}	-9.65·10^{-11}	-1.05·10^{-10}
p_{10}	4.36·10^{-5}	-1.30·10^{-5}
p_{02}	-1.57·10^{-9}	1.39·10^{-10}
p_{01}	-2.06·10^{-5}	-4.25·10^{-6}
p_{00}	471.0858	209.7866

Table 5.3: Coefficients of the bi-variate polynomial to describe the IWS width and thickness variation.

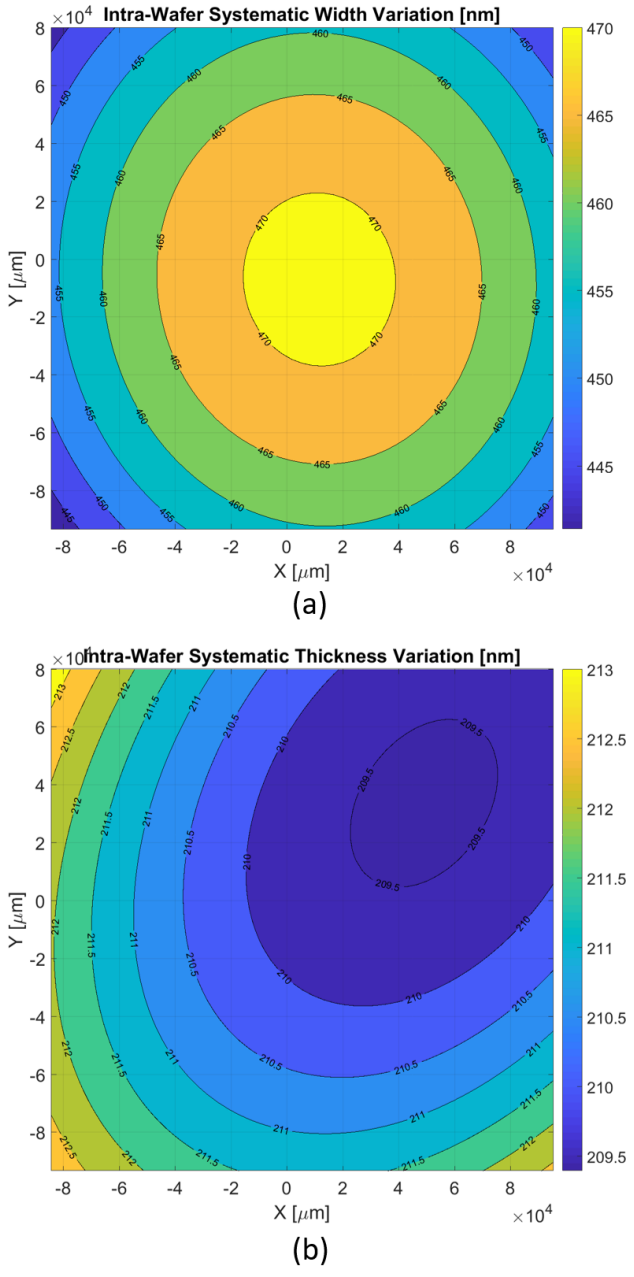


Figure 5.16: IWS variation of (a) width and (b) thickness.

	IWS Width [nm]	IWS Thickness [nm]
Max (samples)	471.3835	211.9885
Min (samples)	454.9698	209.4005
Max (52 dies)	471.4	212.6
Min (52 dies)	452.0	209.4
Max-Min (samples)	16.4137	2.5880

Table 5.4: IWS width and thickness variation. We listed both maximum and minimum IWS variation among samples on the wafer. We also applied the polynomial to calculate the IWS variation on all possible positions on the 52 dies.

The IWS model we built here is from the analysis of the data we measured on a single wafer by the EP5814 imec run. The results we obtained from two wafers from EP4695 illustrated in Chapter 2 Figure 2.12 and [30] also show very similar patterns. The radial pattern of the IWS width variation and the slanted plane of the IWS thickness variation on each wafer we measured are good initial validation of the IWS model.

To further validate the IWS model, we have also put the CMZI circuit on the wafers of a dedicated run, where we could design in the full field of the mask. We distributed the monitoring circuits over the entire die to obtain more insights on the die-level variation we will discuss next. From this run we will also obtain more wafers to analyze the IWS variation. Since IWS is a systematic variation that repeats on every wafer, from multiple wafers we can average the coefficients of the polynomial fitting of each wafer to derive IWS model coefficients. The variation on each coefficient explains the wafer-to-wafer variation that is not studied in this research.

5.5.2.3 Intra-Die Systematic Variation

On intra-die level, systematic variation reflects the repeated fabrication variation introduced by the photomask error or pattern density caused etching speed non-uniformity. As introduced in Section 5.4, after removing the IWS variation from raw extraction data, we extracted the IDS by averaging the residuals of the same sample on all dies. From Fig. 5.17, we observed that the IDS width variation are very correlated locally. The maximum IDS width variation is 1.52 nm, and the minimum is -2.52 nm. The repeated variation on die-level counts for a maximum of 4.04 nm difference, which is 16% of total width variation. The IDS width is significantly larger near array waveguide gratings and spirals, where the pattern density is large (i.e. large fraction is etched). The variation tends to be negative near the east boundary that is also the border of the die (where there is a strip between two dies with no tiling placed) where the pattern density is low. We have observed a moderate to a strong correlation between pattern density and IDS width variation. We will describe it in detail in Section. 5.6.

IDS thickness variation has a maximum of 0.40 nm and a minimum of -0.51 nm. The maximum difference is 0.91 nm which is 15% of total thickness variation. We did not observe a correlation between IDS width and thickness variation. Also, we did not find an association between pattern density and IDS thickness variation.

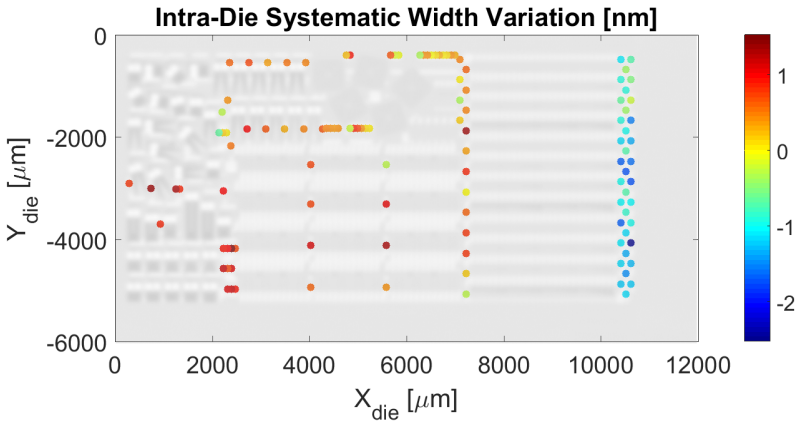


Figure 5.17: Scatter plot of the intra-die systematic width variation. The plot is overlay with the layout in gray scale. The color of the marker indicates the value of the IDS width variation.

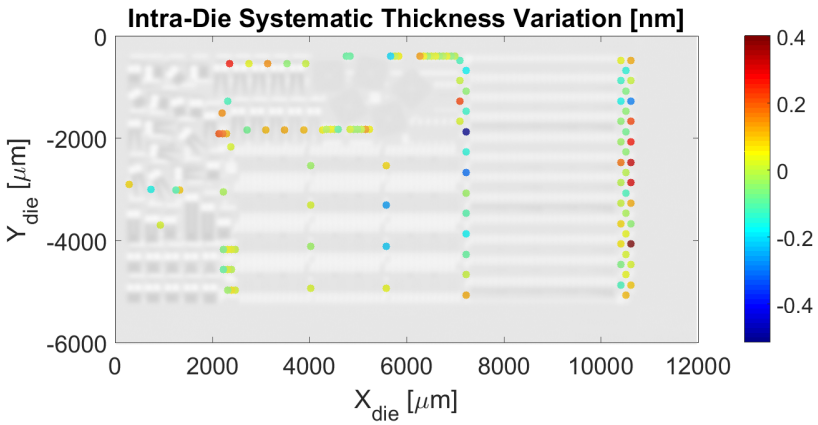


Figure 5.18: Scatter plot of the intra-die systematic thickness variation. The plot is overlay with the layout in gray scale. The color of the marker indicates the value of the IWS thickness variation.

	IDS Width [nm]	IDS Thickness [nm]
Max	1.52	0.4
Min	2.52	-0.51
Max-Min	4.04	0.91

Table 5.5: Intra-die systematic variation.

5.5.2.4 Intra-Wafer Random Variation

Intra-wafer random variation is the die-to-die random variation. The IWR width variation might be led by a focus or exposure fluctuation in the imaging of lithography process from die to die. As in Fig. 5.19 (a), IWR width variation shows no spatial correlation. It has a maximum of 3.83 nm and a minimum of -3.40 nm. The difference in the variation can be as large as 7.23 nm that is equal to 28.7% of the total width variation.

IWR thickness variation shows an interesting deviation from the systematic flat plane on the wafer level. We observed that prominent variations tend to appear near the edge of the wafer. We did not know the exact process of the fabrication. This offset might be caused by the faster polishing rate near the border of the wafer and imperfection in the process to correct the thickness using the corrective etching [7]. Except for dies near the edge, other dies on the wafer have minimal IWR variation with similar value compared with the fitting error (0.08 nm).

One way to improve the modeling of the IWR thickness variation is to divide the variation into two location-dependent parts: a contribution from dies on the edge of the wafer and a contribution from dies in the middlecenter. They can be assumed as normal distributions with different standard deviation. The die on the edge is likely to have a large deviation. Now, we have only one wafer map to analyze, so that we could not verify our findings on the edge. Also, the number of dies on the edge from one wafer is insufficient to obtain the statistical properties like the standard deviation of the randomness on the edge. As mentioned, we can improve the IWR variation model with more wafers to arrive.

	IWR Width	IWR Thickness
Mean [nm]	0.04	0.00
SD [nm]	1.68	0.33
Max [nm]	3.83	0.84
Min [nm]	-3.40	-1.04
Max-Min [nm]	7.23	1.88

Table 5.6: Statistics of IWR width and thickness variation.

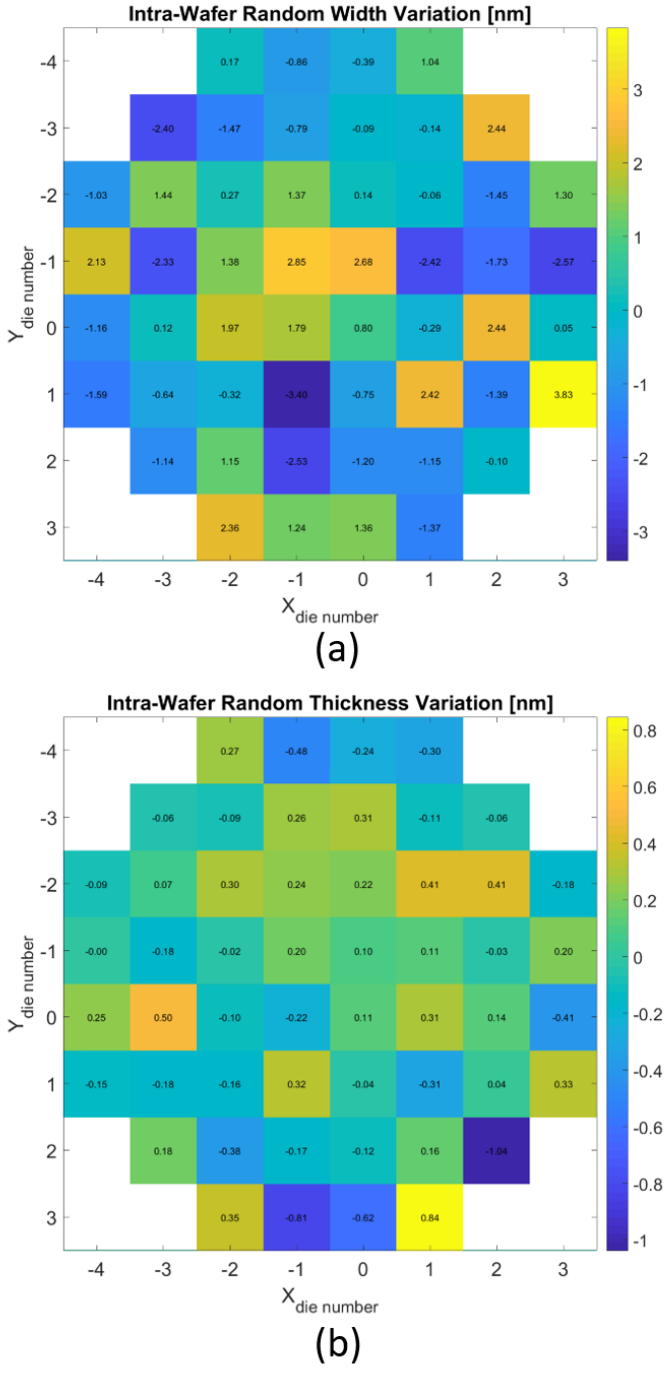


Figure 5.19: IWR variation of (a) width and (b) thickness.

5.5.2.5 Wafer-Die Interaction Variation

After we removed the systematic width variation at wafer and die level, we could observe that within each die the residual is not totally random. Instead, it shows a strong spatial correlation and the trend obviously can be fitted by a plane. The R squared value of the fitting has an average of 0.92. It means a significant contribution (92%) in the residual can be explained by the plane, which we deem as WDI width variation. If we neglect the plane-like component, we would significantly overestimate the random variation on die level.

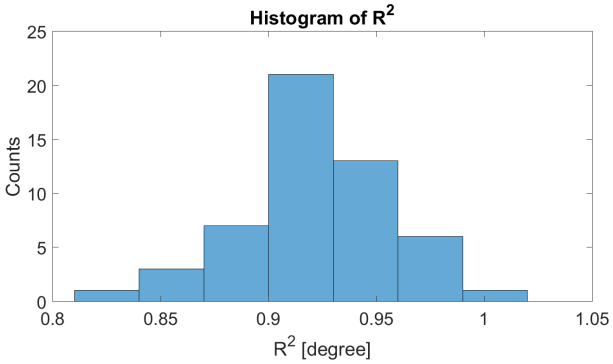


Figure 5.20: Histogram of the R squared value of fitting the width residual with a plane. The R squared value has an average of 0.92. It means a significant contribution (92%) in the residual can be attributed to the WDI variation.

Fig. 5.21 shows the residual after removing systematic variation of width on four randomly chosen dies. The plane that fits the die residual has a very small deviation from being a horizontal plane. We observed that the direction of the plane normal does not correlate with its location. As explained in Section 5.3.5, we describe the plane by inclination θ and azimuth ϕ. Fig. 5.4 shows that θ is a random variation with mean value of 179.9999844 with a standard variation of 1.216e-5. ϕ behaves like a uniform distribution with range between $-\pi$ and π. No association between location of the die and its angles is observed.

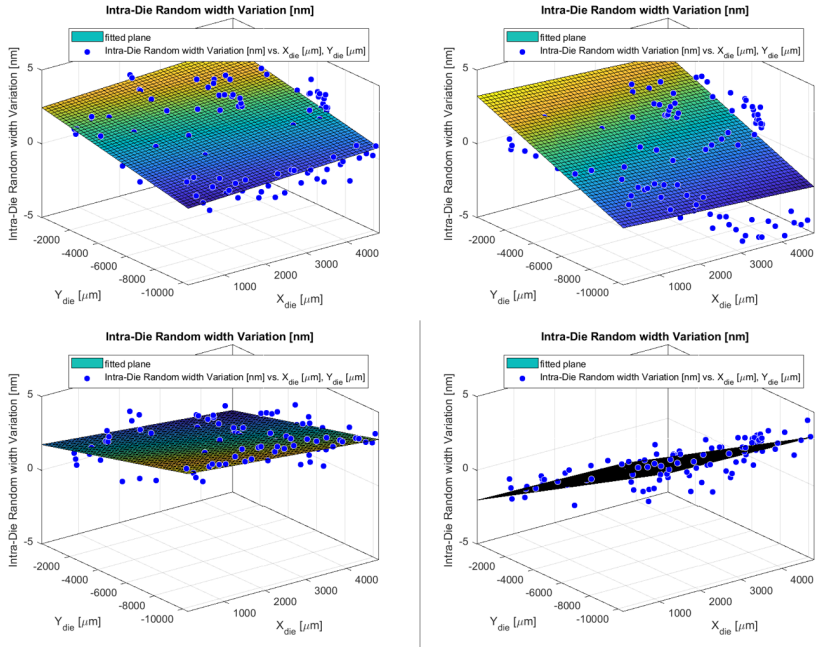


Figure 5.21: The figure shows the width residual on four randomly chosen dies on the wafer. Blue dots are residuals after removing IWS and IDS variation from the raw data. Fitted the plane is the WDI of the width variation plus the IWR offset. The distance between the fitted plane and the blue dot is the IDR variation. WDI contributes largely to the residual, and it explains the correlated spatial variation in the residual.

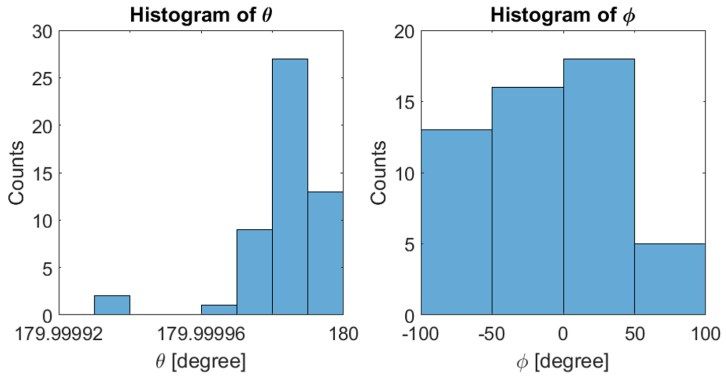


Figure 5.22: Histogram of (a) inclination and (b) azimuth of the WDI width variation.

Unlike residual in the width variation, thickness shows very little plane-like contribution on the die level. Figure 5.24 shows that we fitted the residual with a plane. The R squared value of the fitting has an average of 0.09 (Fig. 5.23), so that the WDI variation contributes little in the residual who is mostly (91%) random. The contribution of the WDI might be the low-frequency non-uniformity on the die level due to polishing that is not captured by the IWS.

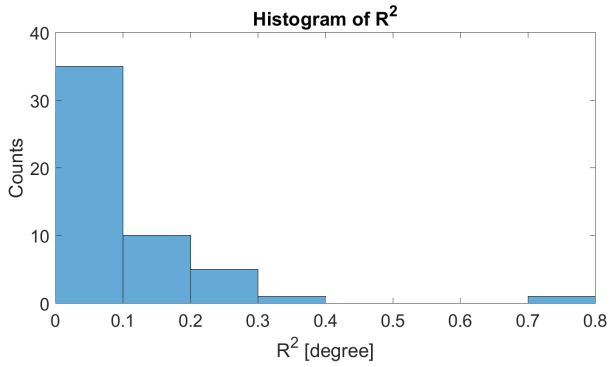


Figure 5.23: Histogram of the R squared value of fitting the thickness residual with a plane. The R squared value has an average of 0.09. It means a little contribution (9%) in the residual can be attributed to the WDI variation.

WDI variation is systematic within a die, but random from die to die. As by definition, the WDI is zero at the die center. We estimate the contribution of the WDI variation by its maximum on the die, which can be calculated by the diagonal of the die and the inclination angle of the WDI.

$$V_{WDI,max} = |0.5 \times diagonal \times sin(\theta)| \tag{5.9}$$

Obviously, the WDI variation increases with the size of a die. The die we measured has a size of 21840 μm $\times$ 21840 μm. The diagonal has a length of 39866 μm. We use the average inclination θ to calculate the maximum WDI variation on a die. The maximum WDI width variation estimated is 4.2 nm while maximum WDI thickness variation is 0.8 nm.

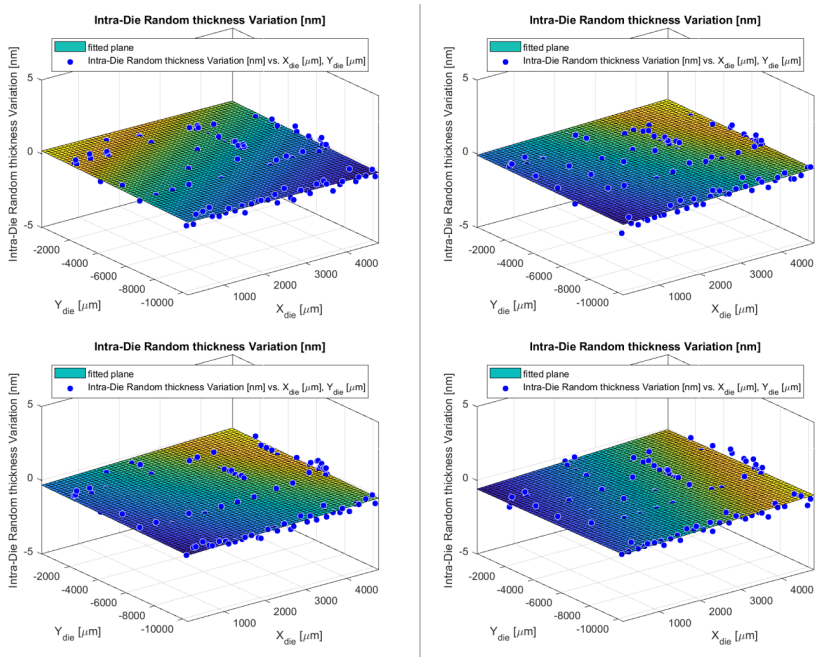


Figure 5.24: The figure shows the thickness residual on four randomly chosen dies on the wafer. Blue dots are residuals after removing IWS and IDS variation from the raw data. Fitted the plane is the WDI of the width variation plus the IWR offset. The distance between the fitted plane and the blue dot is the IDR variation. WDI contributes little to the residual, and the residual are mostly location-independent randomness.

	Width	Thickness
WDI_{max} [nm]	4.2	0.8
μ_θ	179.9999844	179.99999687

Table 5.7: Maximum WDI on a die estimated by the average inclination angle.

5.5.2.6 Intra-Die Random Variation

IDR variation is the random variation between circuit or device. It is the residual after removing all the systematic variations and random variations on larger spatial scales. We did correlation test of the residual which is IDR variation. For both IDR width and thickness variation, we did not observe spatial correlation. The IDR width variation has standard deviation of 1.10 nm. The IDR thickness variation has a standard deviation of 0.34 nm.

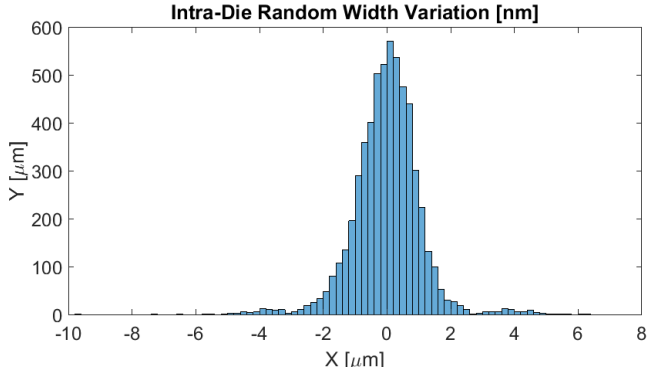


Figure 5.25: Histogram of the IDR width variation on the single wafer.

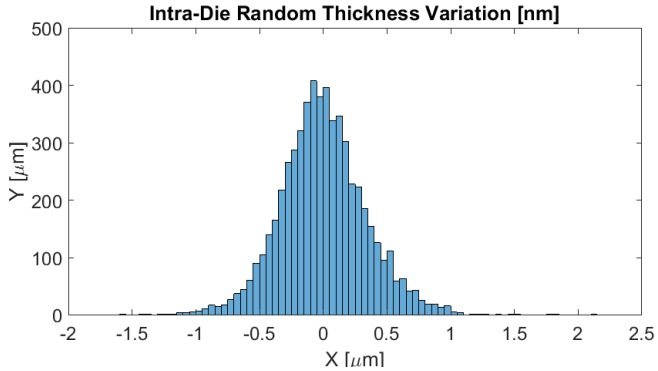


Figure 5.26: Histogram of the IDR thickness variation on the single wafer.

	IDR Width	IDR Thickness
Mean [nm]	0.00	0.00
SD [nm]	1.10	0.34
Max [nm]	6.24	2.11
Min [nm]	-9.61	-1.56
Max-Min [nm]	15.85	3.68

Table 5.8: Statistics of IDR width and thickness variation.

5.5.3 Discussion

The analysis above helps us to understand the location dependency of the process variation. Obviously, process variation is not simply random and cannot be simply described as a normal distribution. Instead, the location-dependent systematic

variation contributes significantly to the performance non-uniformity. It is hard to directly compare the value of systematic variation to the statistic moments of random variation. Nevertheless, we can calculate the contribution of the systematic variation in the total variation, and estimate random contribution that takes up the remaining part.

IWS and IDS works on different scale, and a die is much smaller compared to a wafer. We can think that when IWS is the maximum, we can also find all IDS values in the roughly same wafer location. So, IWS and IDS can approximately reach maximum and minimum at the same location. Then, we can represent the contribution percentage of the systematic variation by using sum of systematic variation divided by the total variation measured.

	Variation		Percentage [%]			
	Width	Thickness	Width		Thickness	
IWS [nm]	16.41	2.59	65.1	81.1	43.9	59.3
IDS [nm]	4.04	0.91	16.0		15.4	
IWR [nm]	1.68	0.33	5.2	19.9	10.6	40.7
IDR [nm]	1.10	0.34	3.4		10.9	
WDI_{block} [nm]	3.4	0.6	10.5		19.2	
Max Variation [nm]	25.2	5.9				

Table 5.9: A comparison of different spatial levels of variation.

For width variation, the sum of systematic variation is 20.45 nm which is equivalent to 81.1% of the 25.2 nm total variation measured on the wafer. It clearly shows that the systematic contribution is the main source of process variation. On the other hand, the random variation contributes to 19.9% in the variation measured. Still, in the random contribution, the biggest part is the WDI variation, which is location-dependent within a die but the dependency is random from die to die. For thickness, the sum of systematic variation is 3.50 nm which is equivalent to 59.3% of the 5.9 nm total variation measured on the wafer. Also, the systematic contribution is the main source of the thickness process variation. On the other hand, the random variation contributes to 40.7% in the variation measured. Random variation in the thickness is far more non-negligible than that in width. The big contribution from the systematic variation further proves the importance of the location-dependent variability analysis. This also indicates that if we can compensate the pattern of the systematic variation, we can significantly limit the process varaition in the future.

In terms of spatial level, most variation of both width and thickness is on the wafer level. To a circuits on chip, wafer level variation behaves like common-mode variation that impact different parts of the circuits similarly. Die-level variation impact the circuit components differently. For example, a circuit like MZI are

sensitive to differential-mode phase imbalance that changes the phase in two phase arms differently rather than the common-mode variation that contributes to the two phase arms equally. The die-level variation should be carefully considered and compensated in the circuits that are MZI-based.

As we mentioned in the introduction of the chapter, 1 nm change in linewidth can lead to 1 nm spectrum shift, while 1 nm change in thickness even leads to around 2 nm spectrum shift. The total width variation is four times that of the thickness, so that width variation leads to more than 25 nm spectrum shift and it is the main source of the performance degradation on our wafer.

5.6 Correlation between Pattern Density and Intra-Die Systematic Width Variation

Intra-die systematic variation counts for a large portion of the total fabrication variation. Since it works on die level, it determines the device and circuit disparity in a die. The variation increases difference between components in the same circuit and a good understanding of its process origin is very helpful to reduce such variation.

One of the possible cause of the IDS variation is the varying profile of the *pattern density* (PD). An area of high PD consumes plasma fast, which leads to a low plasma concentration and a decreasing etching rate. Also, high PD area creates more reaction products (wastes) which slow the etching rates and create deposited residues. Given the low etching rate, the sidewall of the structure is less eroded so that a waveguide will be patterned with a wider linewidth in the dense-patterned locality on the chip. Varying PD leads to a change in the plasma concentration and etching rate, which ultimately results in die-level linewidth variation. Since the pattern is duplicated on each die, the pattern is systematic on die level.

5.6.1 How to Generate Pattern Density from the Layout Mask

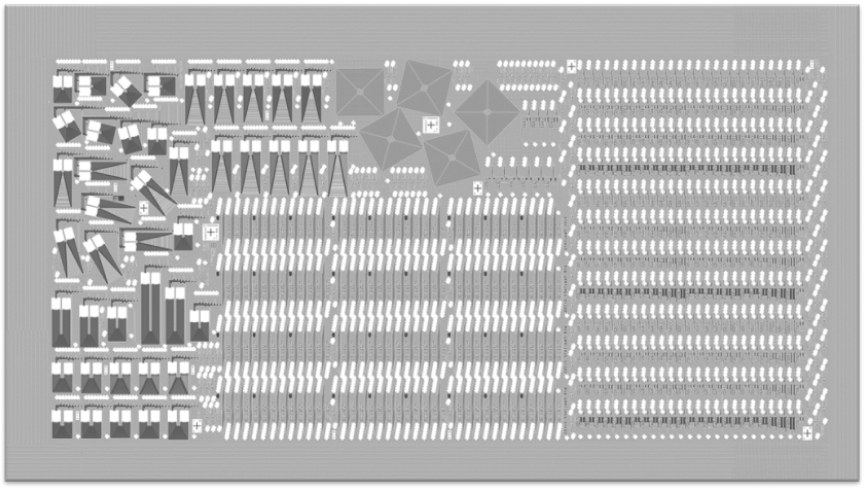


Figure 5.27: Layout of the our block. There are high-density patterns like AWG and spirals on the left and top left. On the right edge, the pattern is less dense.

When comparing the pattern on the design (Figure 5.27) and IDS width (Figure 5.17) variation, a clear correlation between PD and IDS width variation can be visually observed. To study quantitatively what the strength of correlation is and in which range the pattern impacts fabricated width, we correlated the PD map with IDS width map. PD by definition is the portion of area with design in a window over the windowed area on-chip. In a passive photonic chip, we can approximately think that the pattern is defined on waveguide layer. So, the PD on a photonics chip is:

$$PD = \frac{Area\ of\ Waveguide\ Layer\ in\ the\ Window}{Window\ Area} \tag{5.10}$$

The equation shows that the value of PD is dependent on how we choose the window area. The choice is quite subjective. First, it is decided by the window size. When we choose a large window, the value of PD at each location is averaged over a large area. So, the PD would have a low-frequency profile. In addition to size, the shape of window also decides how the PD is calculated. When a semiconductor fab tries to determine if the pattern is too dense locally in the *design rule checking* (DRC) procedure, it would normally use a moving rectangle window to calculate the PD.

Our purpose is to find out how pattern or the density of pattern affects the fabrication variation that differs width from devices next to each other. Obviously,

the fabricated linewidth is mostly influenced by the pattern close to it. The pattern long distance away should have little impact on the fabrication locally. The choice of window should reflect the distance-dependent impact. If the window size is over-estimated, the impact of distant pattern will be over-calculated. If the window size is under-estimated, we will under-estimate the impact from a distance. So, the PD calculated with the chosen window should be most correlated to the IDS width variation. The window with proper size should tell that within which range the pattern affects the IDS variation. In other words, it should tell outside which window, pattern density will have little influence on our fabricated design. To study in which range PD affects fabrication, it is more realistic and reasonable to use a round window so that it imitates the impact of pattern on the etching rate, which relies on its distance to the point of interest. To be more precise, a pattern close to the point of interest should have a stronger impact, and the impact should reduce gradually with the distance. So we assume that the impact of pattern on local PD can be filtered by a 2-dimensional Gaussian filter so that the impact decreases gradually over distance:

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{5.11}$$

where x is the distance from the origin in the horizontal axis, y is the distance from the origin in the vertical axis, and σ is the standard deviation of the Gaussian distribution. The Gaussian filter works as a round window with entrance weighting decreasing as the pattern moves away from the window center. σ intuitively is the window size and presents the radial length of the round window in which pattern that has an strong impact on points of interest.

To generate the PD map using a Gaussian filter, we generate the PD map from a high-resolution bitmap sampling of the gds file of the die layout. The resolution should be high enough to distinguish the pattern on the waveguide layer and the tiling so that the pattern density calculation is accurate. First, we deleted all the layers except for waveguide layer that represents pattern in the gds file. We have complete pattern information from our block covering one eighth of area of the die. However, little information on the pattern outside our block on the die is given. We assumed that outside the block, the area is covered by tiling with same density as the tiling on our block. Then, we exported the gds to a very high resolution grey scale bmp file. After that, we processed the exported image using the Gaussian filter with various standard deviation (Fig. 5.28). The image presents the pattern density map under a given window size σ.

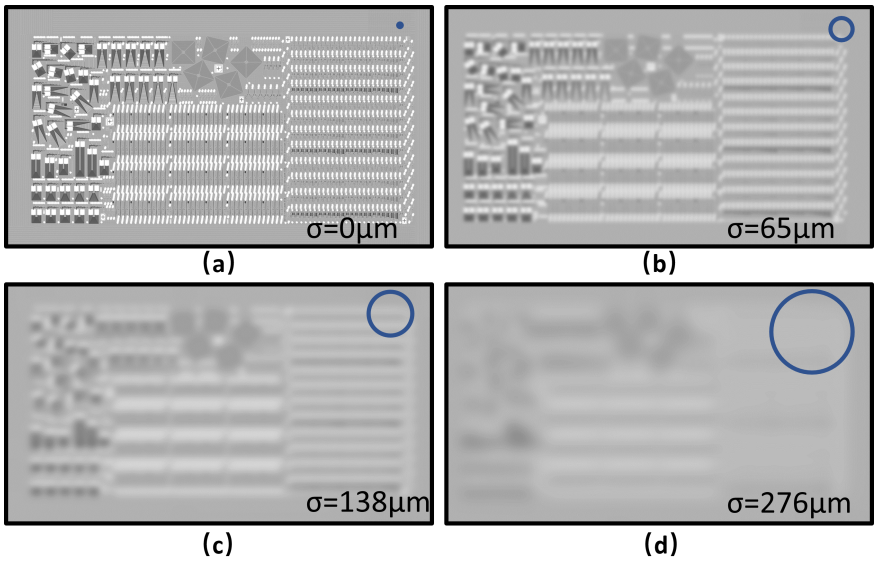


Figure 5.28: We filtered the pattern image with 2d Gaussian filter with different sigma. (a),(b),(c),(d) are examples of PD calculated using sigma of 0, 69, 138, 276 μm respectively. The blue circle is the window that has a radius of 3 σ, which indicates the assumed region that pattern density has impact on the IDS width variation.

5.6.2 Window Size and Correlation Strength

After we obtained the filtered pattern image, we calculated the correlation between two arrays: the IDS width variation of the samples on the die and corresponding PD determined by σ. Fig. 5.29 exhibits an example of PD contour with IDS width variation of the pattern in the block with $\sigma = 69\mu m$. We swept σ from 0 to 918.5 μm to find out the optimum window size σ. We observed that the correlation quickly increases with σ increasing from 0. When $\sigma = 69\mu m$, IDS width variation become most correlated with the pattern density. As the σ is further increased, the two variables become less correlated. $\sigma = 69\mu m$ is the optimal window size that reflects the impact of pattern density on IDS width variation. When $\sigma = 69\mu m$, PD is moderately positively correlated with local pattern density with a correlation coefficient of 0.57. Three arrays of samples on the east edge might be have inaccurate PD calculated since we have little information of the pattern outside the east border. If we remove those samples from our correlation analysis, the correlation further increases to 0.62, which means that PD and IDS width variation has a strong positive correlation.

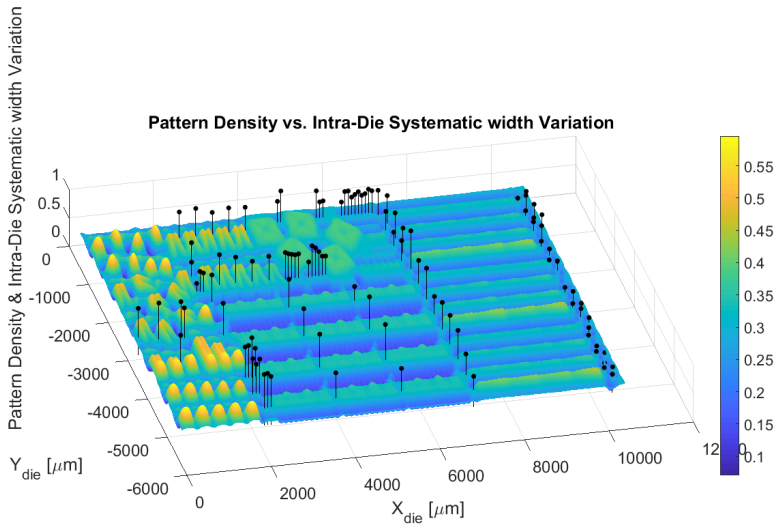


Figure 5.29: Contour plot of filtered PD vs. IDS width variation. 3D contour presents the PD image processed by the Gaussian filter with radius of 69 μm. The stem plot with black head shows the IDS width variation.

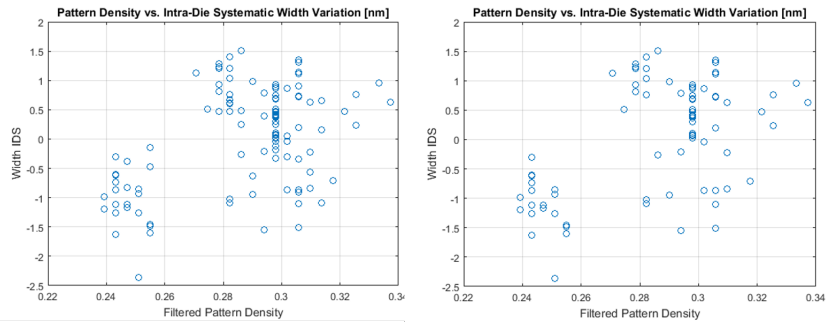


Figure 5.30: Pattern density vs. IDS width variation when the filter radius is 69 μm. Left: all 117 samples. Right 87 samples where east border samples ar excluded.

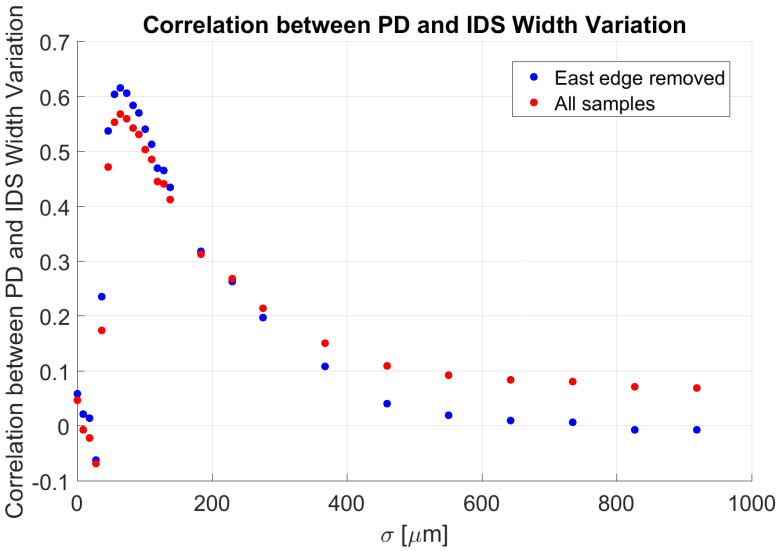


Figure 5.31: Correlation between IDS width variation and the PD calculated using the Gaussian filter. The IDS width variation and the PD are chosen at the same location on the die.

From Equation 5.11, we can conclude pattern density at one point only affects the etching rate within a circle whose radius is 3σ. The patterns outside the 3σ radius circle have little impact on the etching rate thus the IDS width variation. In our case, IDS width variation is immune to any pattern outside the $3\sigma \approx 200\mu m$ circle. This has interesting implication that if we put a circuit about $200\mu m$ away from a highly dense pattern, it would not suffer much from the pattern. So when we evaluate the pattern density in the design, we should use a Gaussian circular window instead of using the large rectangle window, a Gaussian filter with proper size reflects more how pattern affects fabricated with. Even simpler, we can just force the design to be ˜200 μm away from dense patterns.

5.7 Conclusion

In this chapter we discussed about the hierarchical spatial variability model and applied the model to analyze measured data on a fabricated 200 mm Silicon photonic wafer. We first discussed the variability in integrated Photonics and introduced the necessity to obtain fabrication wafer map for layout-aware variability analysis. Then, we described the recent progress to obtain the fabrication wafer map with optical measurements.

We focus our study to separate variability on various spatial levels into systematic and random variation. We discussed the physical origins of process variation.

Based on that, we proposed a model to separate variability on intra-wafer, intra-die level and introduced wafer-die interaction term to reduce random variation in our analysis. We also illustrate our workflow to separate variation components.

We applied the model and the workflow to process the measurements on a 200 mm wafer fabricated by the imec 193 nm lithography. The result shows that the intra-wafer systematic variation is the major source of variation for both linewidth and thickness. We observed the width variation has a systematic dome-like profile across the wafer. Thickness non-uniformity across the wafer looks like a slanted plane with a few mismatches around wafer edge. On die level, we found repeated systematic width pattern are closely related to the local pattern density. Our analysis showed that the intra-die systematic width variation are affected by the pattern within a distance of ˜200 μm to the site. Our findings helps to identify the process variation and create new design rules to alleviate the impact of the non-uniformity. If the systematic variation we observed can be further identified and compensated in the process, we can foresee the photonics wafer with less fabrication uniformity in the future.

References

[1] Sarvagya Dwivedi, Herbert D'heer, Wim Bogaerts, Herbert D 'heer, Wim Bogaerts, Herbert D'heer, and Wim Bogaerts. *Maximizing fabrication and thermal tolerances of all-silicon fir wavelength filters*. IEEE Photonics Technology Letters, 27(8):871–874, 4 2015.

[2] Wim Bogaerts, Umar Khan, and Yufei Xing. *Layout-Aware Yield Prediction of Photonic Circuits*. In IEEE International Conference on Group IV Photonics, pages 1–2, Cancun, Mexico, 2018. Ieee.

[3] Wim Bogaerts and Lukas Chrostowski. *Silicon Photonics Circuit Design: Methods, Tools and Challenges*. Laser & Photonics Reviews, 12(4):1700237, 4 2018.

[4] Shankar Kumar S.K. Selvaraja, Wim Bogaerts, Pieter Dumon, Dries Van Thourhout, and Roel Baets. *Subnanometer Linewidth Uniformity in Silicon Nanophotonic Waveguide Devices Using CMOS Fabrication Technology*. IEEE Journal of Selected Topics in Quantum Electronics, 16(1):316–324, 1 2010.

[5] Shankar Kumar Selvaraja. *Wafer-Scale Fabrication Technology for Silicon Photonic Integrated Circuits*. PhD thesis, University of Gent, 2011.

[6] Shankar Kumar Selvaraja, Erik Rosseel, Luis Fernandez, Martin Tabat, Wim Bogaerts, John Hautala, and Philippe Absil. *SOI thickness uniformity improvement using wafer-scale corrective etching for silicon nano-photonic device*. In The 2011 Annual Symposium of the IEEE Photonics Benelux Chapter, page 289292, 2011.

[7] Shankar Kumar Selvaraja, Erik Rosseel, Luis Fernandez, Martin Tabat, Wim Bogaerts, John Hautala, Philippe Absil, Shankar Kumar Selvaraja, Erik Rosseel, Luis Fernandez, Martin Tabat, Wim Bogaerts, John Hautala, Philippe Absil, Shankar Kumar Selvaraja, Erik Rosseel, Luis Fernandez, Martin Tabat, Wim Bogaerts, John Hautala, Philippe Absil, Shankar Kumar Selvaraja, Erik Rosseel, Luis Fernandez, Martin Tabat, Wim Bogaerts, John Hautala, and Philippe Absil. *SOI thickness uniformity improvement using corrective etching for silicon nano-photonic device*. In IEEE International Conference on Group IV Photonics GFP, pages 71–73. IEEE, 9 2011.

[8] Dan-Xia Xu, Jens H. Schmid, Graham T. Reed, Goran Z. Mashanovich, David J. Thomson, Milos Nedeljkovic, Xia Chen, Dries Van Thourhout, Shahram Keyvaninia, and Shankar K. Selvaraja. *Silicon Photonic Integration PlatformHave We Found the Sweet Spot?* IEEE Journal of Selected Topics in Quantum Electronics, 20(4):189–205, 7 2014.

[9] Shankar K. Selvaraja, Gustaf Winroth, Sabrina Locorotondo, Gayle Murdoch, Alexey Milenin, Christie Delvaux, Patrick Ong, Shibnath Pathak, Weiqiang Xie, Gunther Sterckx, Guy Lepage, Dries Van Thourhout, Wim Bogaerts, Joris Van Campenhout, and Philippe Absil. *193Nm Immersion Lithography for High-Performance Silicon Photonic Circuits*. In 27th Optical Microlithography Conference as part of the SPIE Advanced Lithography Symposium, volume 9052, page 90520F, 2014.

[10] Yisu Yang, Yangjin Ma, Hang Guan, Yang Liu, Steven Danziger, Stewart Ocheltree, Keren Bergman, Tom Baehr-Jones, and Michael Hochberg. *Phase coherence length in silicon photonic platform*. Optics Express, 23(13):16890, 6 2015.

[11] Lukas Chrostowski, X. Xu Wang, Jonas Flueckiger, Y. Wu, Y. Wang, and S. Talebi Fard. *Impact of fabrication non-uniformity on chip-scale silicon photonic integrated circuits*. In Conference on Optical Fiber Communication, Technical Digest Series, pages Th2A–37, Washington, D.C., 3 2014. OSA, OSA.

[12] Xu Wang, Wei Shi, Han Yun, Samantha Grist, Nicolas A. F. Jaeger, and Lukas Chrostowski. *Narrow-band waveguide Bragg gratings on SOI wafers with CMOS-compatible fabrication process*. Optics Express, 20(14):15547, 2012.

[13] Raymond G. Beausoleil, Andrei Faraon, David Fattal, Marco Fiorentino, Zhen Peng, and Charles Santori. *Devices and architectures for large-scale integrated silicon photonics circuits*. In Louay A. Eldada and El-Hang Lee, editors, Optoelectronic Integrated Circuits XIII, Proc. of SPIE, volume 7942, page 794204. International Society for Optics and Photonics, 2 2011.

[14] Zeqin Lu, Jaspreet Jhoja, Jackson Klein, Xu Wang, Amy Liu, Jonas Flueckiger, James Pond, and Lukas Chrostowski. *Performance prediction for silicon photonics integrated circuits with layout-dependent correlated manufacturing variability*. Optics Express, 25(9):9712, 2017.

[15] Yufei Xing, Domenico Spina, Ang Li, Tom Dhaene, and Wim Bogaerts. *Stochastic collocation for device-level variability analysis in integrated photonics*. Photonics Research, 4(2):93, 2016.

[16] Tsui-Wei Weng, Zheng Zhang, Zhan Su, Youssef Marzouk, Andrea Melloni, and Luca Daniel. *Uncertainty quantification of silicon photonic devices with correlated and non-Gaussian random parameters*. Optics Express, 23(4):4242, 2015.

[17] Tsui Wei Weng, Daniele Melati, Andrea Ivano Melloni, Luca Daniel, and others. *Stochastic simulation and robust design optimization of integrated photonic filters*. Nanophotonics, 6(1):299–308, 2017.

[18] Wim Bogaerts, Yufei Xing, and Umar Khan. *Layout-Aware Variability Analysis, Yield Prediction, and Optimization in Photonic Integrated Circuits*. IEEE Journal of Selected Topics in Quantum Electronics, 25(5):1–13, 9 2019.

[19] Wim Bogaerts, Roel Baets, Senior Student Member, Pieter Dumon, Senior Student Member, Vincent Wiaux, Stephan Beckx, Dirk Taillaert, Senior Student Member, Bert Luyssaert, Senior Student Member, Joris Van Campenhout, and Senior Student Member. *Nanophotonic Waveguides in Silicon-on-Insulator Fabricated With CMOS Technology*. Journal of Lightwave Technology, 23(1):401–412, 1 2005.

[20] J.A. Croon, G. Storms, S. Winkelmeier, I. Pollentier, M. Ercken, S. Decoutere, W. Sansen, and H.E. Maes. *Line edge roughness: characterization, modeling and impact on device behavior*. In Digest. International Electron Devices Meeting,, pages 307–310. IEEE.

[21] B. Cheng, S. Roy, G. Roy, F. Adamu-Lema, and A. Asenov. *Impact of intrinsic parameter fluctuations in decanano MOSFETs on yield and functionality of SRAM cells*. Solid-State Electronics, 49(5):740–746, 5 2005.

[22] Kun Qian, Borivoje Nikolić, and Costas J. Spanos. *Hierarchical modeling of spatial variability with a 45nm example*. In Vivek K Singh and Michael L Rieger, editors, Design for Manufacturability through Design-Process Integration III, volume 7275, page 727505, 3 2009.

[23] Brian E. B.E. Stine, Duane S. D.S. Boning, and James E. J.E. Chung. *Analysis and decomposition of spatial variation in integrated circuit processes and devices*. IEEE Transactions on Semiconductor Manufacturing, 10(1):24–41, 1997.

[24] Gary S May and Costas J Spanos. *FUNDAMENTALS OF SEMICONDUCTOR MANUFACTURING AND PROCESS CONTROL*. Wiley-Interscience, 2006.

[25] David A. Steele, Anthony Coniglio, Cherry Tang, Bhanwar Singh, Steve Nip, and Costas J. Spanos. *Characterizing post-exposure bake processing for transient- and steady-state conditions in the context of critical dimension control*. volume 4689, page 517. International Society for Optics and Photonics, 7 2002.

[26] C Chiang and J Kawa. *Design for manufacturability and yield for nano-scale CMOS*. Springer, 2007.

[27] Dennis Okumu Ouma. *Modeling of Chemical Mechanical Polishing for Dielectric Planarization by*. 1998.

[28] Duane S. Boning and James E. Chung. *Statistical metrology: understanding spatial variation in semiconductor manufacturing</title>*. volume 2874, pages 16–26. International Society for Optics and Photonics, 9 1996.

[29] Kun Qian. *Variability Modeling and Statistical Parameter Extraction for CMOS Devices*. PhD thesis, 2015.

[30] Yufei Xing, Jiaxing Dong, Umar Khan, and Wim Bogaerts. *Hierarchical Model for Spatial Variations of Integrated Photonics*. pages 1–2, Cancun, Mexico, 8 2018.

6

Realistic Yield Prediction of Silicon Photonics

Silicon photonics is sensitive to process variation. Being able to estimate the yield at the design stage is crucial to improve the design for better process tolerance. Also, the prediction helps to estimate and reduce the cost of chip development and production. There are two significant issues in the yield prediction for silicon photonics. First, the simulation cost of Monte-Carlo experiment is very high to obtain the statistics on a fabricated wafer for the prediction. Second, a yield prediction can only be realistic when it is layout-aware. In this chapter, we will first discuss methods based on the stochastic analysis to reduce simulation cost for yield prediction. In the second part, we will illustrate the CapheVE framework to make layout-aware yield prediction.

6.1 Challenges in Yield Prediction

Silicon photonics is the technology to integrate a large number of optical functions on a chip using CMOS technology. It has the potential to reach low-cost, high-volume manufacturing. However, it is very sensitive to process variations that limit the complexity of circuits and volume of functioning products.

As we introduced in the previous chapters, variability due to fabrication processes affects the performance of photonic circuits and limits the complexity of the circuits that can be implemented. Being able to estimate the yield at the design

stage is crucial to improve the design for better process tolerance. Also, designers would like to obtain a sensible estimation on yield to calculate the cost and return of the production at the design stage. Variability analysis and yield prediction can also help to estimate the power budget to compensate for the process variation using techniques such as thermal tuning. Therefore, a proper variability analysis on the performance of the photonic components and circuits has become crucial.

The *Monte Carlo* (MC) method [1] is considered the standard approach for variability analysis, thanks to its accuracy and ease of implementation. Unfortunately, the MC analysis has a slow convergence rate, so it requires a large number of data points (simulations or measurements). Therefore, MC has a very high computational cost, considering that accurate simulations of photonic devices can be time and resource-intensive. In Section 6.2, we will introduce how to use stochastic analysis algorithms to reduce the simulation cost of yield prediction.

Another problem in yield prediction is how to include layout-awareness in the prediction workflow. As process variations are highly location-dependent, the performance of the fabricated circuit is primarily determined by the layout of the circuit and its position on the wafer. Just assuming the process variations to be purely random will only give a qualitative estimation of the statistics. A realistic yield prediction requires to capture the random and the location-dependent variations quantitatively. In Section 6.3, we will illustrate how to make layout-aware yield estimation using the CapheVE framework and our wafer-scale variability model described in Chapter 5.

6.2 Stochastic Analysis Algorithms to Reduce the Simulation Cost of Yield Prediction

The standard approach of variability analysis is based on the MC method. It injects random values into the many circuit parameters to extract performance metrics (e.g., crosstalk, insertion loss, bandwidth). MC repeats this experiment by sampling with the probability distributions assigned to the input parameters. After the experiment is repeated a sufficiently large number of times, the MC experiment will generate a probability distribution of the output for analysis. The Monte-Carlo method [1] is easy to implement and has good accuracy. But the method has a slow convergence, i.e., it requires tens of thousands of experiments to provide accurate statistics of the outcome. In photonics, an experiment in the MC method is either a measurement or a simulation, which can be quite expensive. Therefore, repeating the experiment thousands of times is impractically expensive for variability analysis.

To lower the cost of variability analysis, we can replace expensive simulations or measurements with cheap models. These models use analytical functions to de-

scribe relations between device or circuit performances and design parameters. In this section, we will introduce two methods to build a cheap stochastic model. We will give a brief introduction to the *generalized polynomial chaos* (gPC) method. We will also introduce the *stochastic collocation* (SC) method with an example to do variability analysis on the *directional coupler* (DC) with parameter uncertainties.

6.2.1 Polynomial Chaos Expansion Methods

To build the cheap stochastic model, one of the modeling technique in variability analysis is the gPC method. The gPC expansion has been applied in several domains as an efficient alternative to the classic MC method [2–4]. It is a non-sample based approach that expresses a stochastic process as a series of orthogonal basis functions [5]. Recently, it has been proposed for the variability analysis of photonic devices [6, 7] and circuit tolerance optimization [8], which saves computation cost considerably.

The gPC-based modeling approach aims at expressing a stochastic process as a series of orthogonal basis functions with suitable coefficients and gives an analytic representation of the variability of the system on the random variables under consideration [9]. Let $u(\vec{x}, \epsilon)$ be a quantity of interest of a device or circuit under process variations, for example, the bandwidth of a filter, the coupling coefficient of a DC, or the effective index of a waveguide. The vector $\vec{x}$ is a set of constant design variables, such as the width or thickness of a waveguide, the gap of a DC, or the nominal radius of a ring resonator. ϵ is a random vector describing process variations, such as the variations in the waveguide width, thickness, the gap of a DC, and the radius of a ring. The idea of gPC is to approximate $u(\vec{x,}\epsilon)$ by a set of basis functions as:

$$u(\vec{x}, \epsilon) = \sum_{i=0}^{\infty} a_i(\vec{x})\phi_i(\vec{\epsilon}) \tag{6.1}$$

where $\phi_i(\vec{\epsilon})$ is a multivariate polynomial and a_i is the corresponding coefficient. i is the non-negative integer multi-index. A perfect approximation is achieved with an infinitely large polynomial order. However, for practical applications, the equation above must be truncated to a limited number of basis functions $M + 1$ via suitable truncation schemes, leading to:

$$u(\vec{x}, \epsilon) \approx \sum_{i=0}^{M} a_i(\vec{x})\phi_i(\vec{\epsilon}) \tag{6.2}$$

We can prove that $M + 1 = \frac{(d+p)!}{p!d!}$, where d is the number of random variables, and p is the order of the polynomials. The coefficients a_i can be obtained

either intrusively by stochastic Galerkin method or sample-based methods such as stochastic collocation method and regression method. To obtain the coefficients, we need to evaluate at least $M + 1$ simulations or experiments. The optimal convergence rate of a gPC model can be achieved when we choose a basis functions $\phi_i(\vec{\epsilon})$ corresponding to the distribution of the random variables ϵ [5]. For some well-known distributions, such as uniform, Gaussian, Beta, and Gamma distribution, the corresponding bases are also known, and they are Legendre, Hermite, Jacobi, and Laguerre polynomials, respectively. By choosing the proper bases, we can use a lower order of the polynomials p where a smaller number of evaluations is required to calculate the polynomial coefficients.

Once we obtain the polynomials-based gPC model, we can perform variability analysis using the cheap stochastic model efficiently and accurately. *Cumulative distribution function* (CDF) and PDF can be efficiently obtained via the (inexpensive) sampling of equation (6.2). One advantage of the gPC method is that intead of running expensive simulations, stochastic moments like the mean μ and variance σ^2 of Y can be analytically calculated based on the PC expansion coefficients as:

$$\mu(u(\epsilon)) = a_0 \sigma^2(u(\epsilon)) = \sum_{i=0}^{M} a_i^2. \tag{6.3}$$

Moreover, the PC model can also be used for sensitivity analysis. Sensitivity can be determined easily from the PC expansion coefficients [10, 11].

6.2.2 Stochastic Collocation Methods

The *stochastic collocation* (SC) method is an efficient alternative to characterize photonic devices under the effect of uncertainty. The fundamental principle of the SC approach is to approximate the unknown stochastic solution by interpolation functions in the stochastic space. By repeatedly solving (sampling) the deterministic problem at a pre-determined set of nodes in the stochastic space, we can construct the interpolation. This approach offers similar high accuracy and efficiency as the stochastic gPC method, but at the same time, it is easier to implement, like sampling-based methods (e.g., MC approach).

6.2.2.1 General Knowledge about Stochastic Collocation Methods

SC methods are based on interpolation schemes to compute stochastic quantities. Collocation points are a pre-determined set of nodes in the stochastic space [12]. We can construct the interpolation by repeatedly solving the deterministic problem

at collocation points. Indeed, a stochastic process $\mathbf{Y}(\boldsymbol{\xi})$ can be expressed as

$$\mathbf{Y}(\boldsymbol{\xi}) = \sum_{i=1}^{Q} \mathbf{Y}(\boldsymbol{\xi}_i) L_i(\boldsymbol{\xi}) \tag{6.4}$$

where $\boldsymbol{\xi}$ denote the N stochastic parameters and $L_i(\boldsymbol{\xi})$ represents the interpolation basis functions.

For a photonic device, the process $\mathbf{Y}$ could correspond to the functional parameters such as the waveguide propagating constants and the coupling coefficients in coupling devices. The stochastic variables ξ_i correspond to device properties affected in a stochastic way by fabrication and operational conditions (e.g., waveguide linewidth or temperature).

In (6.4) different types of interpolation schemes can be adopted (e.g. piecewise linear [12, 13], Lagrange [5, 9] or multivariate simplicial methods [14]). However, the key issue for this approach is the selection of the support nodes, such that using the minimal number of nodes one achieves a good approximation.

For example, if the Lagrange interpolation scheme is chosen, the element L_i in (6.4) for a one-dimensional interpolation can be expressed as

$$L_i(\xi) = \prod_{i=1,\, i\neq j}^{Q} \frac{\xi - \xi_i}{\xi_j - \xi_i} \tag{6.5}$$

where L_i is equal to 1 for $\xi = \xi_j$ and is equal to 0 for $\xi = \xi_i$. Next, for interpolation in multiple dimensions, a tensor-product approach can be used and equation (6.4) becomes

$$\mathbf{Y}(\boldsymbol{\xi}) = \sum_{i_1=1}^{Q_{k_1}} \cdots \sum_{i_N=1}^{Q_{k_N}} \mathbf{Y}\left(\xi_{i_1}^{k_1}, \ldots, \xi_{i_N}^{k_N}\right) \left(L_{i_1}^{k_1} \otimes \cdots \otimes L_{i_N}^{k_N}\right) \tag{6.6}$$

where ξ_i^k is the $i-$th node in the $k-$th direction and the total number of nodes used in (6.6) is

$$Q = \prod_{n=1}^{N} Q_{k_n} \tag{6.7}$$

As it can be seen from (6.7), the number of nodes required by the full tensor product increases rapidly with the number of random parameters N. For example, if three random variables are considered and 10 collocation points are used for each parameter, a total of 1000 nodes are required by the full tensor product approach. Hence, the performance of the photonic device under study must be evaluated for 1000 different combinations of the random variables considered, leading to high computational time.

The required number of nodes can be significantly reduced by adopting sparse grids in the stochastic space, based on the Smolyak algorithm [12, 15–18]. By

choosing the collocation points correctly, Smolyak algorithm drastically reduces the total number of nodes used in the interpolation with respect to the full tensor product approach while preserving a high level of accuracy.

It is important to remark that the SC models are expanded using interpolation functions of *independent random variables* $\boldsymbol{\xi}$ [9]. In the general case of correlated random variables, decorrelation can be obtained via a variable transformation, such as the Nataf transformation [19] or the Karhunen-Loéve expansion [20].

The stochastic moments (mean, variance) can be computed utilizing analytical formulas and then very efficiently, once the analytical form of the interpolation functions $\{L_i(\boldsymbol{\xi})\}_{n=1}^{N}$ has been decided. For example, if the random variables $\boldsymbol{\xi}$ are defined in the sample space $\boldsymbol{\Omega}$, the mean of $\mathbf{Y}(\boldsymbol{\xi})$ is defined as

$$\mu(\mathbf{Y}(\boldsymbol{\xi})) = \int_{\Omega} \mathbf{Y}(\boldsymbol{\xi}) W(\boldsymbol{\xi}) d\boldsymbol{\xi} \tag{6.8}$$

where $W(\boldsymbol{\xi})$ is the joint *probability density function* (PDF) of the random variables $\boldsymbol{\xi}$. Using equation (6.4) in (6.8) leads to

$$\mu(\mathbf{Y}(\boldsymbol{\xi})) = \int_{\boldsymbol{\Omega}} \sum_{i=1}^{Q} \mathbf{Y}(\boldsymbol{\xi}_i) L_i(\boldsymbol{\xi}) W(\boldsymbol{\xi}) d\boldsymbol{\xi} \tag{6.9}$$

which depends only on the interpolation functions $L_i(\boldsymbol{\xi})$ and joint PDF $W(\boldsymbol{\xi})$. Note that, if the choice of the interpolation functions and probability measure does not allow an analytic computation of the stochastic moments like (6.9), an efficient numerical solution can be used (e.g., by MC analysis of the interpolation model or numerical integration). Finally, it is important to remark that it is not possible to define a priori the speed-up of a generic SC modeling technique compared to the MC method. Indeed, the number of nodes needed to compute an accurate SC model (which is directly related to the efficiency of SC methods, as described above) cannot be decided upfront, since it depends on the following factors:

- the impact of the chosen random variables $\boldsymbol{\xi}$ on the variations of the stochastic process considered $\boldsymbol{Y}$ (dynamic stochastic processes require a higher number of collocation points);
- the interpolation scheme L_i adopted (the more powerful the interpolation scheme, the fewer nodes are needed);
- the sampling strategy adopted (efficient sampling strategy limit the number of collocation points used);
- the number of random variables considered (the higher the number of variables, the more collocation points are needed).

However, it has been proven in the literature that, for a limited number of random variables (indicatively less than ten) SC methods are much more efficient with respect to the MC analysis, see [5, 9, 12]. For stochastic processes depending on a high number of random variables, the efficiency of SC methods is significantly reduced.

Two approaches can be used to increase the efficiency of an SC modeling technique. Using nested sampling schemes allows to adaptively choose the collocation points (additional details are provided in Section 6.2.2.6 and Appendix C). Adopting adaptive sparse grids [18] reduces the nodes requirement, which is especially useful when a high number of random variables is considered. For a more detailed reference on SC methods, we refer the reader to [5, 9, 12, 18].

6.2.2.2 Directional Coupler Example

We demonstrate the use of SC for integrated photonics through the analysis of a DC in a silicon photonics platform. As explained in Chapter 3, the power coupling $K_{cross}(L)$ in a DC can be expressed as:

$$K_{cross}(L) = sin^2(\kappa' L + \kappa_0) \tag{6.10}$$

The rate of coupling is defined as the field coupling coefficient κ', which is determined by the geometry of the coupler cross-section, such as the waveguide width, thickness, and gap between the waveguide cores.

We assume that, for simplicity, the two waveguides in a DC are identical. As a result, the straight section of the DC layout is defined by three parameters: the waveguide width w, thickness t, and gap g (Figure 6.1). Furthermore, we assume that, in the lithography process, the centers of the two waveguides are located at the designed position. It is a good assumption for optical lithography techniques but might be less accurate for e-beam written devices. With this assumption, the sum of the gap g and $2\times$ the half-waveguide width w is constant, as shown in Figure 6.1. Therefore, in our example, we can describe the full geometry of the directional coupler with only two parameters: w and t.

In this study, we will use the SC technique to find out how geometry variability influences the DC performance, namely the coupling coefficient κ'. Indeed, due to the fabrication variations, the fabricated linewidth w, thickness t, and gap g are different on the value is chosen during the design phase. To prove the robustness and modeling power of the proposed approach, we assume the width w and thickness t of the DC as correlated random variables, rather than independent, following the Gaussian distribution. It is not an unrealistic assumption: thickness variations could induce over-etching on the sidewalls.

It is good to note that the SC methods can deal with random variables with an arbitrary distribution. It is therefore not necessary that the t and w adhere to a

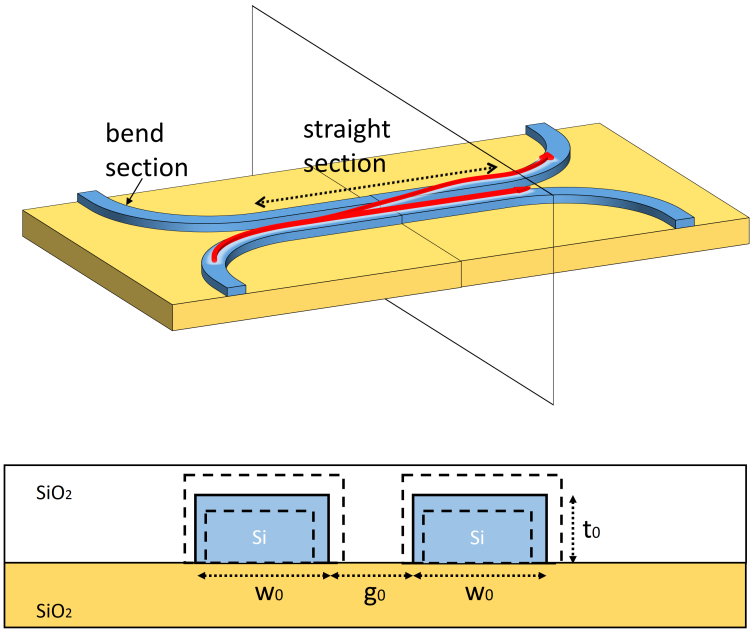


Figure 6.1: The upper plot shows the perspective view of a symmetric DC. Red arrows present the flow of light. Part of the light is coupled from bottom waveguide to the above one. The cross section is amplified in the lower plot. The mean width and thickness of the DC are w_0 and t_0, respectively. The width w and thickness t of the fabricated DC are indicated as dashed boxes. The refractive indexes are $n_{si} = 3.44$, $n_{SiO_2} = 1.45$.

Gaussian distribution.

6.2.2.3 Simulation Setup

According to the theory of supermodes, we can write the coupling coefficient κ' as

$$\kappa' = \frac{\pi}{\lambda}(n_{eff\_o} - n_{eff\_e}) \tag{6.11}$$

where n_{eff_o} and n_{eff_e} are the effective index of asymmetrical and symmetrical supermodes in DC. For our silicon photonics devices, we assume the wavelength to $\lambda = 1.55\ \mu m$.

Next, the nominal value of the width and thickness are $w_0 = 450$ nm and $t_0 = 220$ nm, respectively, while we fix the sum of width w and gap g at 650 nm.

To calculate κ of a given geometry, we define the DC structure accordingly and simulate n_{eff_o} and n_{eff_e} in the mode solver Fimmwave using its Film Matching Mode (FMM) solver. For later performance comparisons, all simulations are performed on a computer with an Intel Core i5 2500 quad-core CPU clocked at 3.3 GHz and 8GB of memory.

6.2.2.4 Problem Definition

As mentioned, we considered the coupling coefficient κ' of a directional coupler as a stochastic process depending on two correlated random variables with Gaussian PDFs: the width w and the thickness t. Hence, the joint PDF of the two random variables considered is defined as

$$W_{\boldsymbol{\eta}} = \frac{1}{2\pi det(\mathbf{C})^{\frac{1}{2}}} exp\left(-\frac{1}{2}\left(\boldsymbol{\eta} - \boldsymbol{\mu}\right)^T \mathbf{C}^{-1}\left(\boldsymbol{\eta} - \boldsymbol{\mu}\right)\right) \tag{6.12}$$

where $\boldsymbol{\eta} = [w\ t]^T$ is the vector of the correlated random variables considered, the vector $\boldsymbol{\mu} = [w_0\ t_0]^T$ contains the corresponding nominal values (mean values) w_0 and t_0, and the matrix $\mathbf{C}$ is the covariance matrix. The symbol $det(\cdot)$ represents the matrix determinant operator. The covariance matrix is defined as

$$\mathbf{C} = \begin{bmatrix} (w_0\sigma_w)^2 & \rho w_0\sigma_w t_0\sigma_t \\ \rho w_0\sigma_w t_0\sigma_t & (t_0\sigma_t)^2 \end{bmatrix}$$

where the symbols σ_w and σ_t are the normalized standard deviations of the w and t, while ρ is the correlation coefficient of these two random variables. The correlation coefficient $|\rho| < 1$ denotes the strength of correlation: the random variables considered are independent if $\rho = 0$ and strongly correlated if $|\rho| = 1$. Note that, by describing this example in terms of normalized standard deviations, we make further analysis independent of the actual nominal values of our 2 random variables.

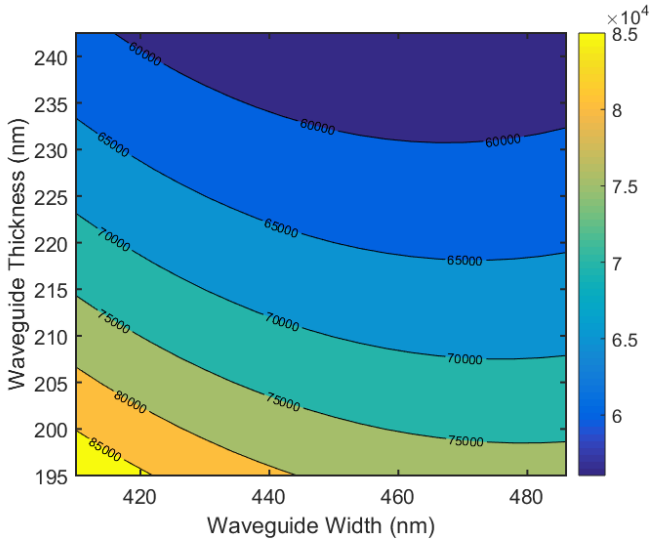


Figure 6.2: 2D contour plot of field coupling coefficient vs. waveguide width and thickness.

When we validate the robustness of the proposed method, σ_w and σ_t are chosen equal to 2% and the correlation coefficient $\rho = 0.9$, which is a challenging example to study since the coupling coefficient is quite dynamic with respect to the parameters considered, see Figure 6.2. The proposed method is discussed in details in the following and summarized Figure 6.3.

6.2.2.5 Variable Transformation

Now, SC methods in the form (6.4) deal with independent random variables. Hence, to fit the problem into the SC framework, first of all, it is necessary to express the coupling coefficient in two independent Gaussian random variables, starting from the correlated random variables $\boldsymbol{\eta}$, defined by (6.12). As mentioned, such decorrelation can be obtained via a variable transformation. Thanks to the Karhunen-Loéve expansion [20], it is possible to express the vector of correlated Gaussian random variables $\boldsymbol{\eta}$ in the vector of uncorrelated Gaussian random variables with zero mean and unit variance $\boldsymbol{\xi} = [\xi_1, \xi_2]^T$ as

$$\boldsymbol{\eta} = \boldsymbol{\mu} + \mathbf{V}\mathbf{E}^{\frac{1}{2}}\boldsymbol{\xi} \tag{6.13}$$

Where $\mathbf{E}$ and $\mathbf{V}$ are the diagonal matrix of the eigenvalues and the full matrix of the eigenvectors of the covariance matrix $\mathbf{C}$, respectively. Since uncorrelated Gaussian random variables are also independent, we have now expressed the coupling

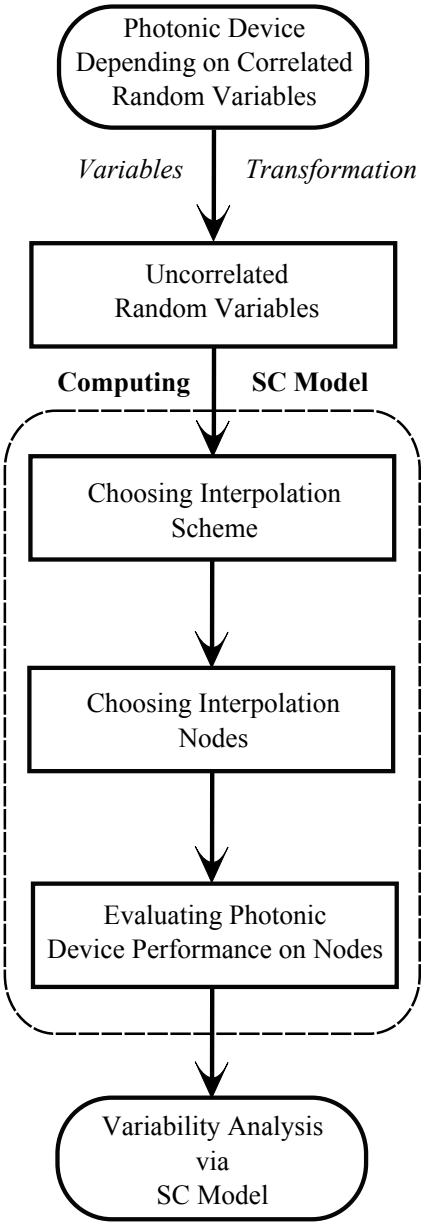


Figure 6.3: Flow chart of the proposed technique.

coefficient as a stochastic process which depends on the pair of independent Gaussian random variables (ξ_1, ξ_2). An accurate description of the Karhunen-Loéve expansion for Gaussian random variables is given in Appendix B.

6.2.2.6 SC Model Computation

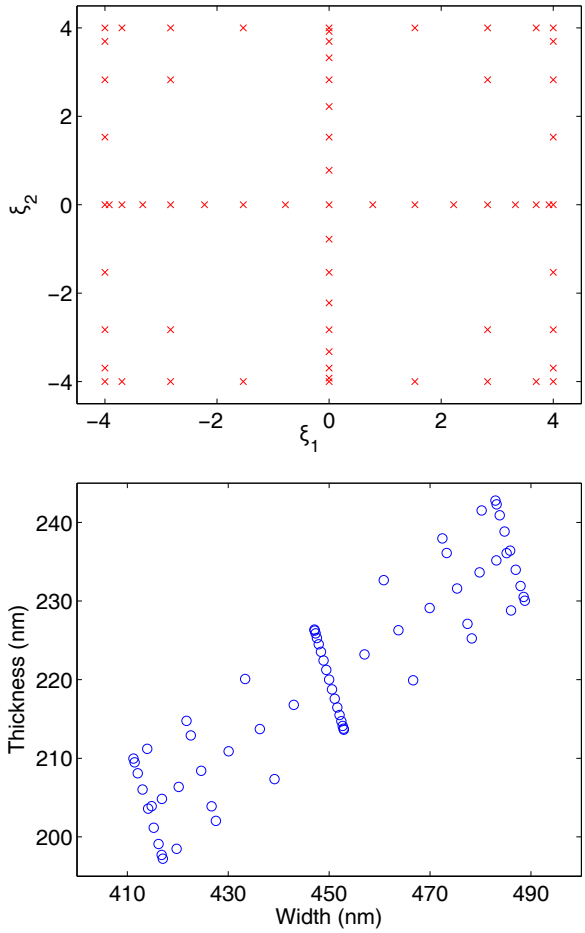


Figure 6.4: Sampling grid where samples are chosen by the Smolyak Algorithm. Top: the red exes $(\times)$ represent the interpolation nodes for the normalized independent random variables ξ_1 and ξ_2 used to build the SC model. Bottom: the blue circles $(\circ)$ are the corresponding values for the correlated random variables w and t used to compute the coupling coefficients in Fimmwave.

To compute an SC in the form (6.4), the first step is choosing the interpolation scheme: the Lagrange interpolation scheme is adopted in this example for its modeling power and ease of implementation. Next, a rule which guarantees a good quality of the approximation must be used to choose the collocation points for each random variable: ξ_1 and ξ_2. In this example, the Clenshaw Curtis rule is adopted [15]: the collocation points for each random variable are the extrema of the Chebyshev polynomials. Now, the total number of nodes could be obtained using the full tensor products of the nodes chosen for each random variable, but it would not be efficient. Instead, the nodes are chosen over a sparse grid based on the Smolyak algorithm. Indeed, the adoption of the Smolyak algorithm allows building our SC model by using only a subset of all the collocation points given by the full tensor product [15]. Furthermore, the collocation points chosen by the Smolyak algorithm based on the Clenshaw-Curtis rule are nested: if additional nodes are required to model the DC accurately, the nodes already computed are kept in the new sparse grid, reducing the number of evaluation of the DC coupling coefficient. See Appendix C for additional details on the Smolyak algorithm. As a result, only 65 collocation points (Figure 6.4) are required to build the desired SC model, and the values of the coupling coefficient at the interpolation nodes are computed using the Film Matching Mode (FMM) solver Fimmwave.

6.2.2.7 Directional Coupler Variability Analysis

Finally, the variability analysis for the coupling coefficient of the directional coupler under study is performed using an SC model depending on the pair of independent Gaussian random variables (ξ_1, ξ_2) and the results obtained are validated through comparison with an MC analysis based on the Fimmwave FMM solver On the directional coupler cross-section for a couple of correlated random variables (w, t). To compare the performance of the two methods, the same set of 10000 samples for the pair of correlated random variables (w, t) (see Figure 6.5). The corresponding values for the independent random variables (ξ_1, ξ_2) are used to estimate the device variability features.

The proposed method shows excellent accuracy compared with the classical MC analysis, as shown in Table 6.1, Figs. 6.6, 6.7. In particular, the mean and the standard deviation of the coupling coefficient obtained employing the two methods are reported in Table 6.1: the relative error in the estimation of the mean and the standard deviation is only 9.0×10^{-5} and 5.6×10^{-3}, respectively. Apart from stochastic moments, more complicated functions of the stochastic process under study can be estimated: the probability density and *cumulative distribution function* (CDF) of κ obtained utilizing the two methods considered are in excellent agreement, as shown in Figure 6.7.

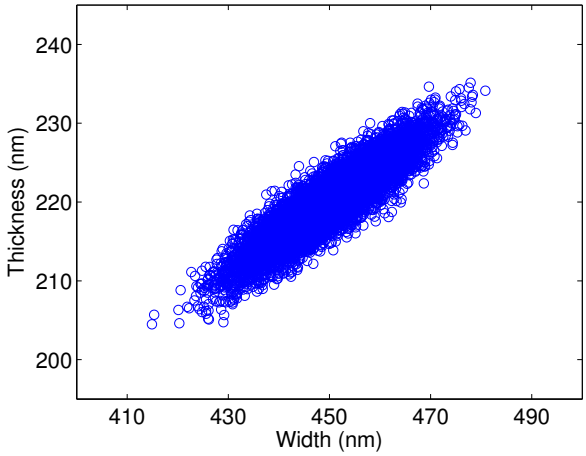


Figure 6.5: Sampling points used to perform the MC analysis through direct Fimmwave simulations for the correlated random variables (w, t). The corresponding values for the independent random variables (ξ_1, ξ_2) are used to evaluate the SC model computed.

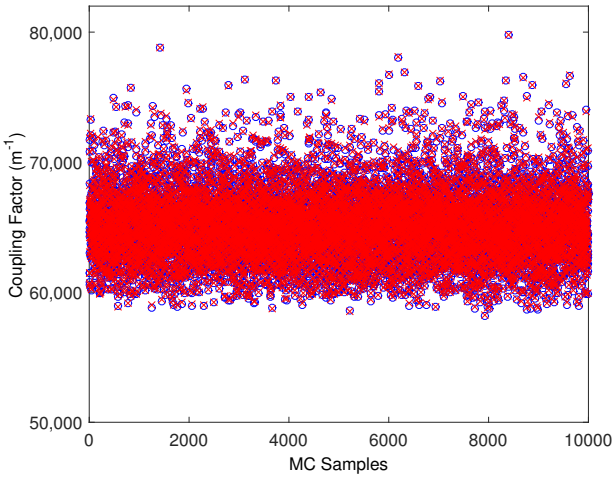


Figure 6.6: Blue circles $(\circ)$: coupling coefficient computed via the MC analysis for the 10000 (w, t) samples shown in Figure 6.5. Red $(\times)$-markers: corresponding values obtained by evaluating the SC model.

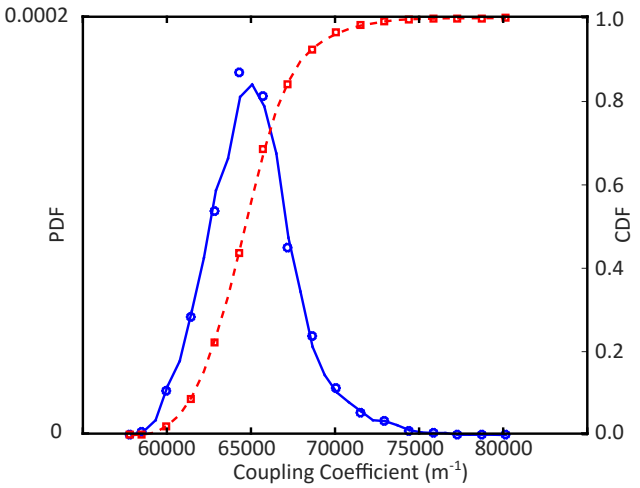


Figure 6.7: The *probability density function* PDF and CDF of the coupling coefficient for $\lambda = 1.55\mu m$. The blue solid and red dashed line are PDF and CDF obtained by means of the SC model, respectively, while the blue circles and red squares represent the same quantities computed by means of the MC analysis.

	Monte Carlo	Stochastic Collocation
Mean value	65160	65166
S.t.d value	2616.9	2631.4

Table 6.1: Performance summary of Stochastic Collocation and Monte Carlo simulation.

Furthermore, as presented in Table 6.2, the SC method has dramatically saved computational cost. Note that, the SC method took a two-step procedure to perform the same variability analysis. Initially, SC required 65 simulations to compute the coupling coefficient at the collocation points. Next, we used the SC model over 10000 samples of the independent random variables in the Monte-Carlo method. Hence, the total computational time of the SC method is 8 min and 59 s, which represents a speed-up of a factor $146\times$ with respect to the MC analysis performed in Fimmwave for the couple of correlated random variables (w, t), which required 21 h 53 min 14 s.

Variability analysis technique		Simulator	Number of points	Computation time
Monte Carlo		Fimmwave FMM solver	10000	21 h 53 min 14 s
Stochastic Collocation	Stochastic Modeling	Fimmwave FMM solver	65	8 min 32 s
	MC using stochastic model	SC stochastic model	10000	27 s
	Total time			8 min 59 s

Table 6.2: Computation time of Stochastic Collocation and Monte Carlo simulation.

6.3 Layout-Aware Yield Prediction Using CapheVE

6.3.1 Caphe Variability Extension (CapheVE) Framework

Assuming purely random variations, we could only estimate the yield of fabricated circuits qualitatively. The real process variations are layout-dependent. As we introduced in Chapter 5, process variations are correlated locally, so the circuit placed closed-by would behave similarly. Also, the systematic variation is the major source of contribution to the process variation. So, the process variation is largely location-dependent. Layout-aware variability analysis should include location-related information of the process variation to predict the circuit performance accurately.

A layout-aware yield prediction should allow us to import the model to describe the process variation, and revise the circuit performance according to its location on the wafer. Monte-Carlo simulation is the most straightforward implementation to do such analysis. We should be able to generate the virtual wafer maps that mimic the realistic process variations. Then, we can put the layout of duplicated designs over the virtual wafer where the circuit performance is revised concerning its location. By comparing all the location-dependent circuit responses over the wafer with the design target, we can predict the yield of the circuit.

Monte-Carlo simulations require a large sample set. Electromagnetic simulations are too expensive for such an analysis. Instead, we need compact behavioral models of all the components to calculate circuit response, which is way faster. Once the global variables such as the waveguide width and thickness change, the model parameters should also be revised to generate the layout-aware circuit response. Therefore, the prediction framework should include four essential elements [21]:

1. Compact models of all building blocks in the circuits.

2. The sensitivity of model parameters to the global variables such as waveguide width and thickness changes.

3. A variability model that describes how global variables changes over the wafer.

4. The circuit layout with locations and orientations of the individual building blocks.

In the Photonics Research Group, we have implemented such a simulation scheme on top of the IPKISS design framework by Luceda Photonics [21]. IPKISS is a parametric design tool for photonic integrated circuits that combines layout, connectivity, and circuit model into parametric cells. It also has a built-in photonic circuit simulator, *Caphe*, that supports both frequency-domain and time-domain simulations, with efficient circuit models that can be custom-written in Python [22].

The circuit design flow for a photonic circuit is depicted in the top part of Figure 6.8. Starting from a component library in a process design kit (PDK), a circuit is composed of parametric building blocks, and a mask layout is generated [23]. The resulting circuit is then simulated, and the design is iterated until it meets the specifications.

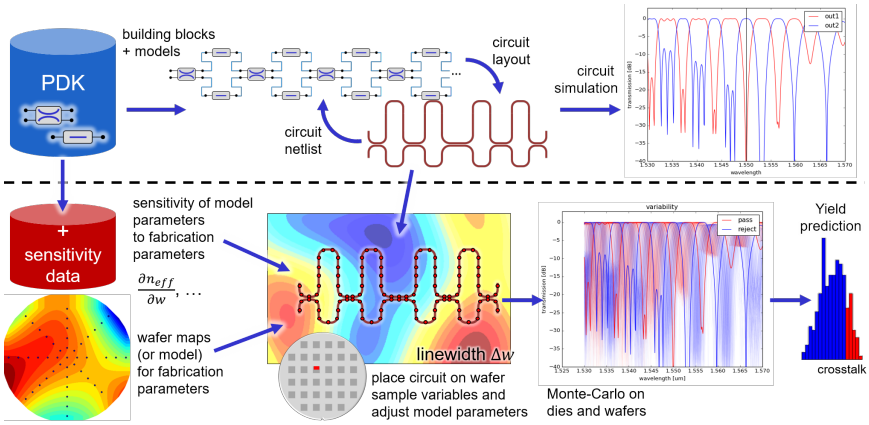


Figure 6.8: Adding layout-aware variability modeling to the photonic circuit design flow. Above the dotted line, the classical photonic circuit design flow is depicted [23], starting from PDK blocks and composing circuits first as a schematic and then as a layout. Only the nominal circuit response is simulated. We extend the PDK models with sensitivity data (either from measurement or simulation) and generate wafer maps of global variables (linewidth, thickness) as we described in Chapter 5. We can then perform MC simulations by placing the circuit on different wafer positions (and different virtual wafers). From the many circuit responses, we can then extract the yield, i.e., the fraction of circuits that meet the specifications set out in the system requirements.

When we extend this design flow with variability analysis and yield prediction, we did not impose any restrictions on how the fab has constructed its circuit models, and how designers generate their layout in the tool. The IPKISS framework, which is written in Python, can be easily extended with additional functionality [24]. Therefore, we created the necessary data structures and processes without perturbing the original framework and without requiring the fab to change their circuit models.

The *Caphe Variability Extensions* (CapheVE) framework combine these four elements with the existing circuit models and layout. First, it positions the circuit on the wafer, then evaluates the local linewidth and thickness for each sample point within the circuit, and for building blocks with multiple samples, these values are aggregated. Using the sensitivity matrix, CapheVE updates model parameters for each instance. Then, it launches a circuit simulation. This is repeated for multiple positions on the wafer. In this process, the original circuit design is not altered.

Based on the results, plotting and data analysis routines from scientific Python libraries can be used to evaluate the impact of the variability or predict the yield of the circuit after fabrication. Because the whole process is scriptable from Python, this simulation routine can easily be embedded in an optimization loop to optimize a circuit for yield, rather than for ultimate performance.

6.3.2 Cascaded MZI Example

In this section, we use the *cascaded Mach-Zehnder interferometer* (CMZI) as an example to illustrate the yield estimation workflow using CapheVE. Since we only have one wafer to build the hierarchical model, there is still more validation to be made on IWS and IWR variation with more wafers to measure in the future. So the example using the CapheVE based on the single wafer data is more an estimation than a prediction. There are two purposes of such an experiment. First, we want to validate the spatial variation model proposed in Chapter 5. To do that, we import the spatial variability model that mimics the process variation extracted on the EP5814 wafer (Chapter 5) and use the variability model to generate virtual wafers in CapheVE. If the virtual wafer shares similar statistics with the extracted wafer maps from optical measurements, the experiment will be a good validation of the spatial variation model. It also gives us the confidence to use the model for realistic yield estimation.

Second, we can validate the extraction workflow proposed in Chapter 4. After the validation of the variability model and the generated virtual wafer maps, we put the CMZI circuits over the virtual wafer. Then, we run MC simulations to generate the spectral responses at each location. The circuits locations are the same as the fabricated wafer. Then, we use the parameter extraction method described in Chapter 4 to extract width and thickness wafer maps from spectral transmission

generated by CapheVE. We will compare the extracted wafer maps with the virtual wafers we imported. By comparing the difference between the two sets of maps, we can obtain the accuracy of the extraction.

6.3.3 Generate Virtual Fabrication Map from the Hierarchical Spatial Variation Model

The CapheVE framework allows us to import additive wafer maps to include a comprehensive model of variations. As we introduced in Chapter 5, process variability can be decomposed into different spatial levels. From the wafer maps extracted from optical measurements, we modeled the systematic variations and random variations quantitatively. Now, we can use the obtained spatial variation model to reconstruct virtual wafers that mimic the statistics and spatial correlation in the variation.

As in Figure 6.9 and 6.10, we generate the wafer map of linewidth and thickness at each level separately. Since the variation model is additive, the final wafer map is an addition of variation maps at each level.

On the wafer-level, *intra-wafer systematic* (IWS) variation describes the slow-varying tendency over the wafer, which is one of the dominant variations for both linewidth and thickness. We generate intra-wafer systematic variation (Figure 6.9 (a) and 6.10 (a)) using the bi-variate polynomials. *Intra-wafer random* (IWR) variation explains the difference in die-to-die averages. Generation of intra-wafer random variation maps is also straightforward. Since no spatial correlation is observed in IWR, we assign each die a random value that follows a zero-mean Gaussian distribution. We set the standard deviation the same as what we extracted from the measured wafer.

Intra-die systematic (IDS) variation accounts for the short-distance correlation observed within a die. IDS width variation is correlated with pattern density. The correlation coefficient is 0.62, so that 62% of the variation can be explained as the deterministic impact of the local pattern density. As explained in 5, the pattern density map is generated by processing the pattern of the waveguide layer on the chip with a Gaussian filter. In particular, the window radius is 65 μm. The pattern density map is stored in a bitmap image file that can be read by Python as an array. We use the linear relation between the normalized pattern density and the IDS width to calculate the pattern density related contribution in IDS width variation map. In addition to the pattern density related contribution, there is still a remaining 38% contribution in the IDS width variation that we do not know the origin. This contribution is also repeated on each die, and we assume it is spatially correlated. (But it could also be mask errors that might occur randomly and are spatially uncorrelated.) We generate the contribution using the coherent noise map with a radius of 500 μm and make the pattern repeat on every die. If we have

further information on how to model this part of the IDS width variation in the future, we can replace the coherent noise map with an updated model.

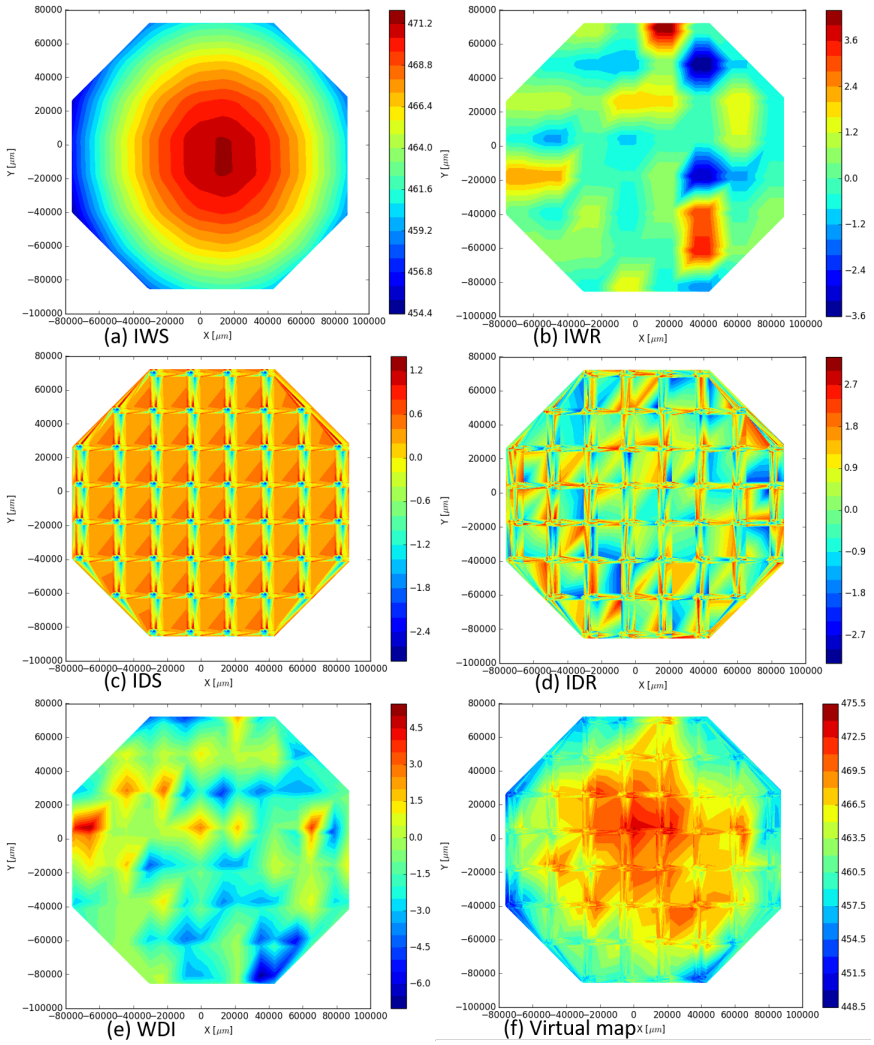


Figure 6.9: Generated width wafer map. (a) Intra-wafer systematic variation. (b) Intra-wafer random variation. (c) Intra-die systematic variation (d) Intra-die random variation (e) Wafer-die interaction variation (f) Virtual fabricated wafer map.

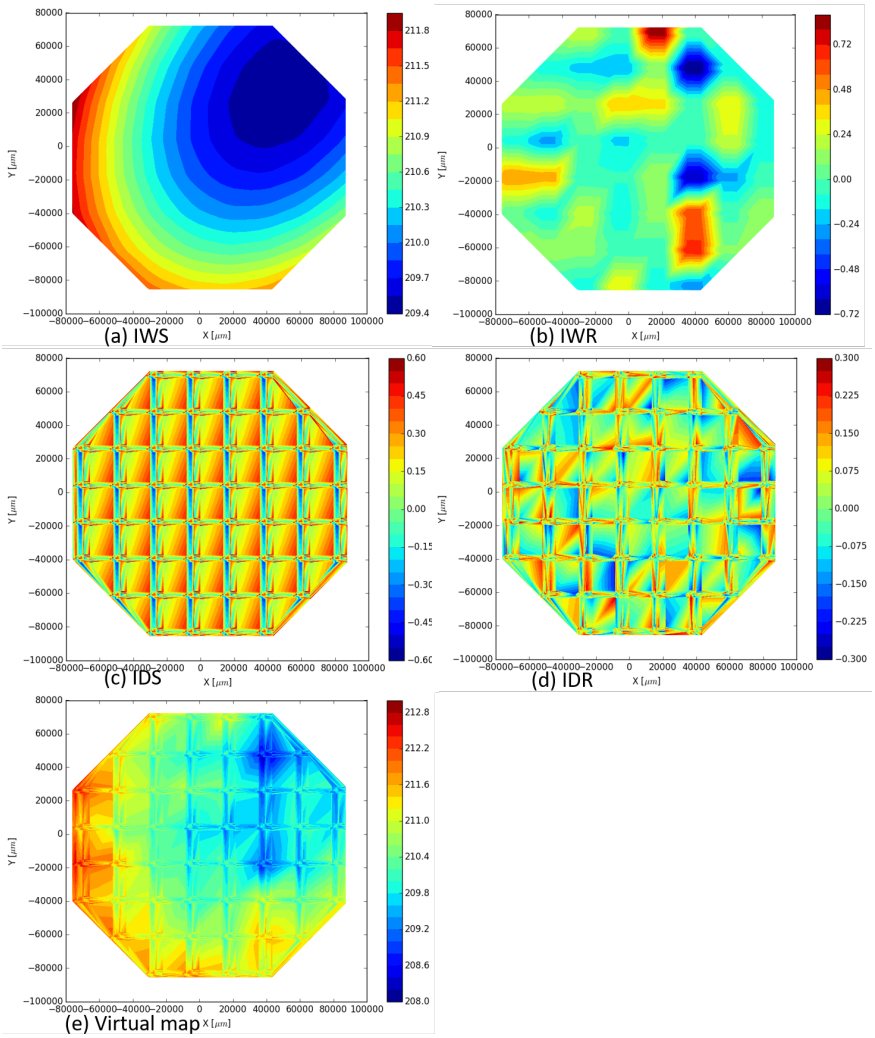


Figure 6.10: Generated thickness wafer map. (a) Intra-wafer systematic variation. (b) Intra-wafer random variation. (c) Intra-die systematic variation (d) Intra-die random variation (e) Virtual fabricated wafer map.

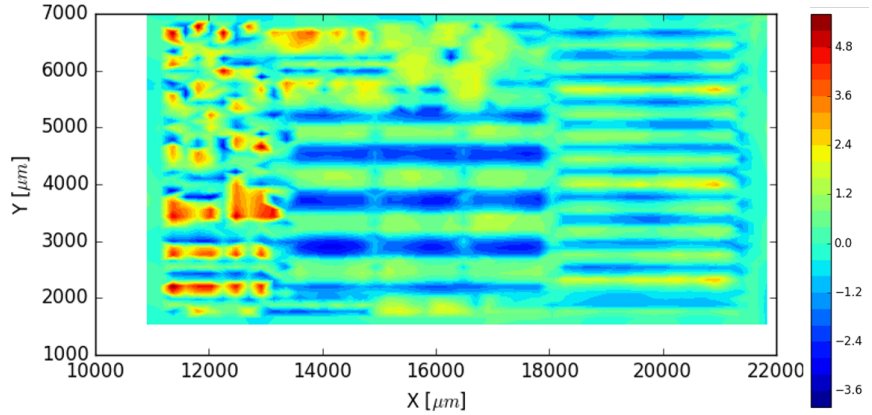


Figure 6.11: Generated IDS width map. The IDS map has both a pattern density correlated contribution and a non-correlated contribution. The figure shows IDS variation over an area of a block which is one-eighth of a die.

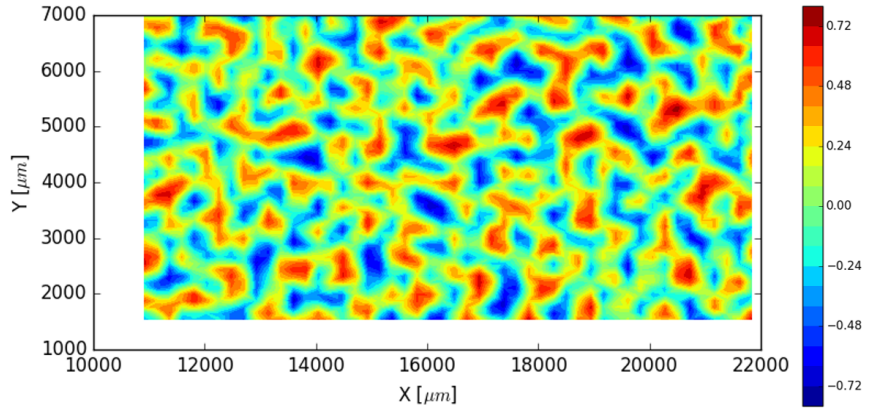


Figure 6.12: Generated IDS thickness map. The figure shows IDS variation over an area of a block which is one eighth of a die.

We also generate IDS thickness variation map 6.12. Since we did not know yet how IDS thickness variation is affected by the layout of circuit or process conditions, we need to generate the IDS thickness variation with an guess. We generated the IDS variation with the coherent noise map. We use the coherent noise map to make the variation more arbitrary and less intentional. The amplitude of the noise is set as the maximum IDS thickness variation we extracted on the measured wafer.

Intra-die random (IDR) variation denotes the device-to-device variations that are not location-dependent. However, within a device or a small circuit, both width

and thickness should still be spatially correlated. We know of no mechanisms where thickness abruptly changes with 1nm from one point to another. The width should also vary continuously along a waveguide. We saw the device-to-device random variation, which seems to be spatially uncorrelated because the correlation length of such randomness is smaller than the device-to-device distance. Therefore, we use the spatially correlated randomness generated by the coherent noise map. The correlation length is set to be 100 μm, which is smaller than the minimum circuit footprint we put on the wafer.

Wafer-die interaction (WDI) variation explains a significant part of location-dependent width variation within a die. WDI map on each die is a plane whose normal has a random deviation from the z-axis (Figure 5.22). As explained in 5.3.5, we describe the plane by inclination θ and azimuth ϕ. θ is defined as a normal distribution, while ϕ is a random value between 0 and 2π. We generate WDI width map from the two random variables. Thickness variation does not have a significant contribution from WDI, so we do not add WDI variation to thickness variation map.

Figure 6.9 (a-e) and 6.10 (a-d) present the virtual wafer map of linewidth and thickness at each level separately. Figure 6.9 (f) and 6.10 (e) present the final wafer maps.

Figure 6.13 shows the generated virtual width and thickness map over the block with the CMZI design. The design block is located in the center of the wafer. Clearly, we do see the impact of several variations here, especially evident in the width map. We can view in the width map a contribution from the pattern density correlation. Also, samples on the right side of the block tend to have a large width. The WDI variation on the die leads the tendency. The IDR variation as randomness that is seen as the blurry background in the map.

The virtual wafer does not have to be same as the fabricated wafer since every fabricated wafer is different. Nonetheless, the virtual wafer should mimic the real wafer in at least two ways:

- It should present the similar statistics of global variables over the wafer as the real wafer.

- It should present a similar spatial correlation of global variables as the real wafer.

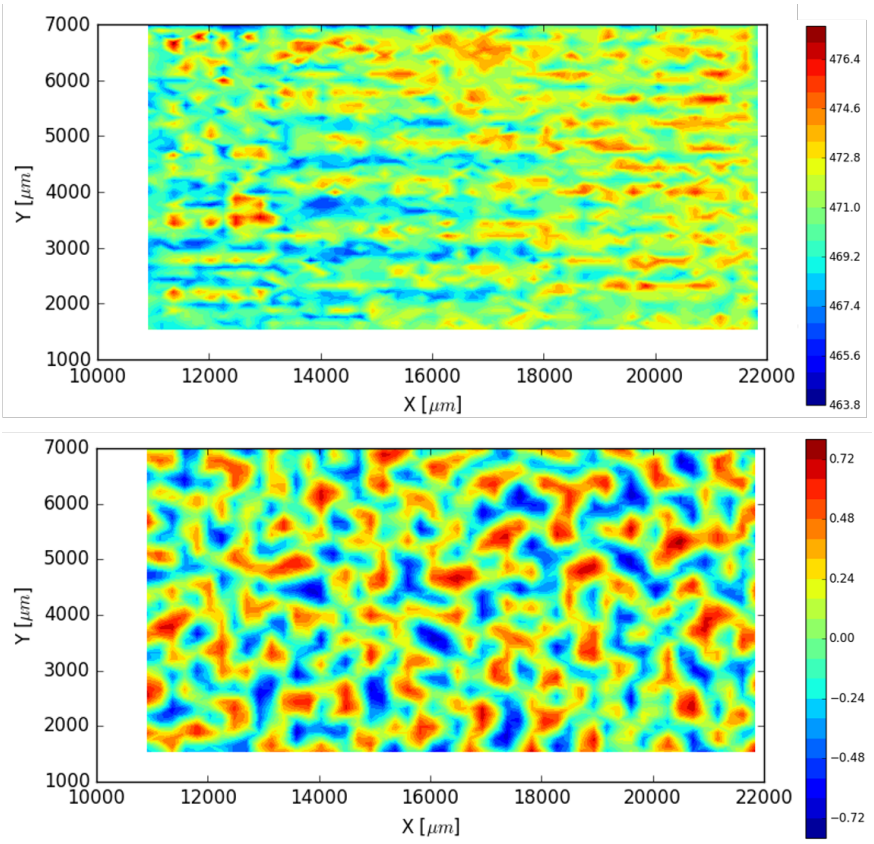


Figure 6.13: Generated virtual map on the block with the CMZI design.

The first criterion is evident. The second criterion ensures the yield estimation captures the correlation between neighboring locations on a chip. A realistic virtual map should have the feature so that it can be used to include the layout-dependency of variations and help to test process-tolerant circuits.

To compare the statistics of linewidth and thickness on the virtual wafer with the fabricated wafer, we sample them at the same locations where CMZIs are located on the fabricated wafer. The sampled locations are shown in Figure 6.14.

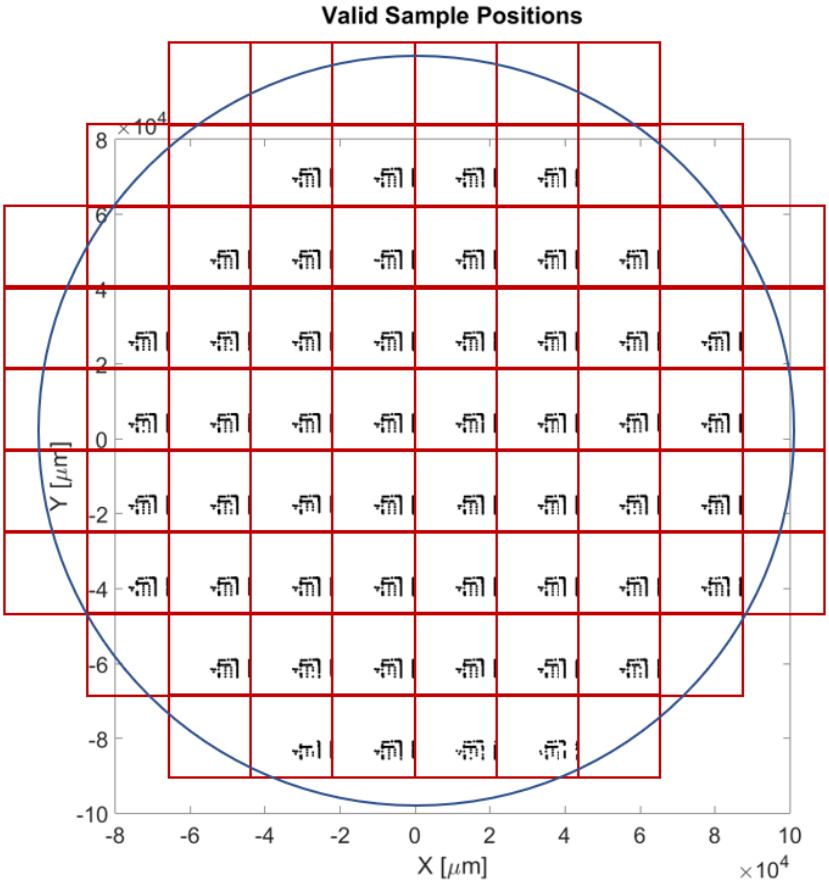


Figure 6.14: Locations of CMZI on the wafer. It share the same locations as on the fabricated wafer. There are 117 (circuits per die) $\times$ 52 (dies) circuits in total.

Figure 6.15 and 6.16 show the histogram of the variables on the fabricated wafer and two generated virtual wafers. As shown in Table 6.3, the mean and standard deviation values on each generate wafer differ slightly, but they matched fabricated wafer statistically. The match in statistical moments is a good validation of the variability model. The shape of the width histograms of the generated wafers match well with the shape of the fabricated wafer. For thickness, there is a larger population on the extreme side (thickness ¿ 213 nm) of the distribution on the fabricated wafer. As mentioned in 5, this is because there is a prominent thickness variation near the edge of the wafer. This deviation is not quantitatively captured by the hierarchical model yet, which causes the mismatch between histograms of fabricated thickness wafer and the generated wafers.

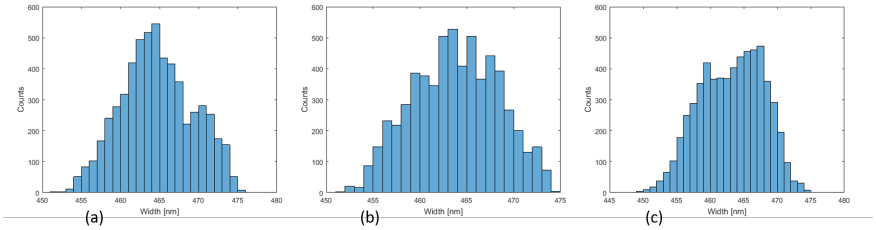


Figure 6.15: Histogram of width on the wafer map. (a) Extracted wafer maps. (b) Virtual wafer map a. (c) Virtual wafer map b.

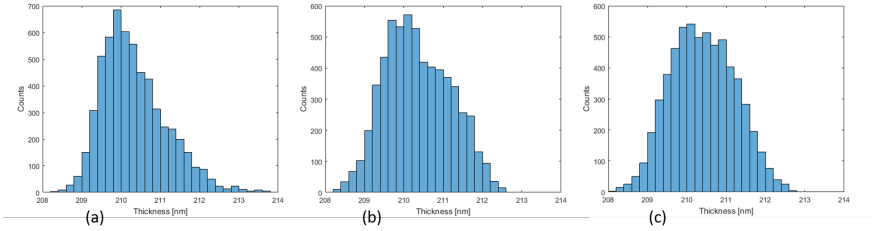


Figure 6.16: Histogram of thickness on the wafer map. (a) Extracted wafer maps. (b) Virtual wafer map a. (c) Virtual wafer map b.

Width	Fabricated Wafer	Virtual Wafer 1	Virtual Wafer 2
Mean [nm]	464.6792	465.5518	463.1819
Standard Deviation [nm]	4.5894	4.4429	4.7052
Thickness	Fabricated Wafer	Virtual Wafer 1	Virtual Wafer 2
Mean [nm]	210.3328	210.4073	210.3547
Standard Deviation [nm]	0.8249	0.8132	0.8321

Table 6.3: The mean and standard deviation of sampled width and thickness on a fabricated wafer and two virtual wafers.

To analyze the spatial correlation of width and thickness, we plot the mismatch in the variables like width and thickness related to the distance between circuits across a chip. Here, we focus on mismatch within a chip because the knowledge on spatial correlation is most interesting to estimate and minimize such impact in a circuit. Since the circuit size cannot go beyond the size of a die, we only plot the mismatches between circuits that are on the same chip. We separate the data into ten groups with an incremental distance of 1000 μm (0-1000 μm, 1000-2000 μm,..., 9000-10000 μm). We calculate the mean, 25, and 75 percentile of each category.

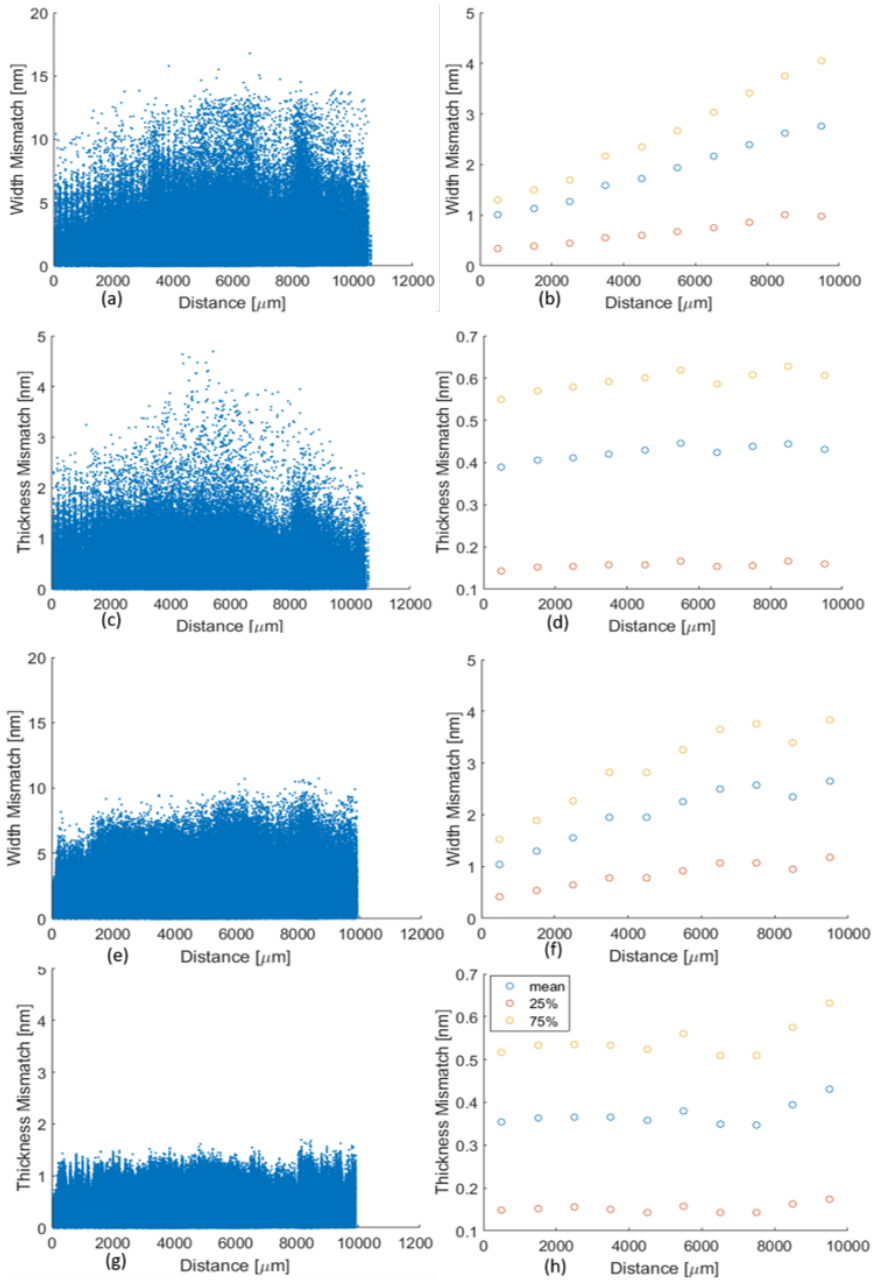


Figure 6.17: (a) Width mismatch vs. the distance between devices across the chip on the fabricated wafer. (b) The mean, 25 and 75 percentile of each category (0-1000, 1000-2000,..., 9000-10000 μm). The lower (orange) and upper (yellow) dots are at the 25%/75% quantiles. The middle circles (blue) indicate the mean. (c) Thickness mismatch on the fabricated wafer. (d) Mean, 25, and 75 percentile of thickness mismatch on the fabricated wafer. (e) Width Mismatch on the virtual wafer. (f) Mean, 25 and 75 percentile of width mismatch on the virtual wafer. (g) Thickness Mismatch on the virtual wafer. (h) Mean, 25, and 75 percentile of thickness mismatch on the virtual wafer.

Figure 6.17 compares the distance-related mismatches on the fabricated wafer (a-d) and on the virtual wafer (e-h). Scatter plots present the difference in width and thickness, versus distance between them. The width mismatch is gradually increasing with the distance between two circuits. It is more evident from the percentile plot in Figure 6.17 (b). Not only on average the mismatch increases steadily, but also the range of center 50 percent (25 to 75 percentile) of mismatches increases gradually against the circuit distance. Unlike the width mismatch, thickness mismatch is less distance related. It does not increase much against the distance between two circuits.

Scatter plots (Figure 6.17 (e) and (g)) of the virtual wafer show similar correlations as of the fabricated wafer. From the percentile map shown in Figure 6.17 (f) and (h), we also see good matches for both width and thickness. The matches prove the spatial variation model captures the spatial correlation statistically. We observed in the scatter plot of the fabricated wafers shows a few samples with larger mismatches in each distance. This should come from the modeling of the IDR variations. As shown in 5.25 and 5.26, the distributions of IDR behave like a normal distribution. But there is excessive population outside the 2σ range than there should be for a normal distribution. This mismatch explains why there are larger mismatches on a fabricated wafer. If we came up with distribution to describe the IDR variation better, we could further improve our variability model.

6.3.4 Sampling points for each components

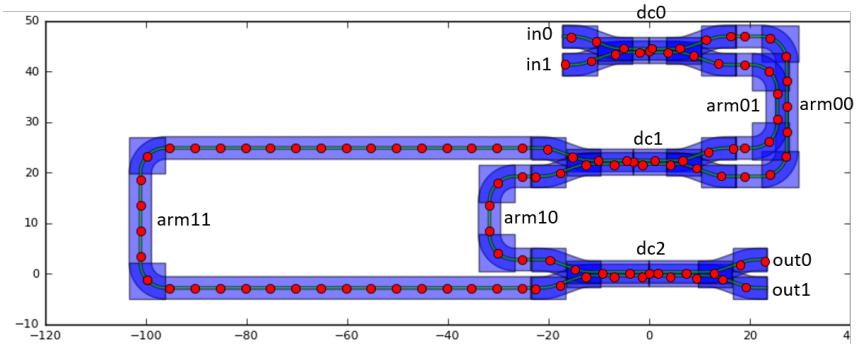


Figure 6.18: The figure shows the layout of the CMZI. To model the effect of variability, the width and thickness variations are sampled on multiple locations in the layout, both for the waveguides and for the directional couplers.

After the generation of the variability map, we should determine where in the layout the global variables such as linewidth and thickness are evaluated. This requires a data structure to determine a set of sampling points in each component.

For most (small) building blocks, a single sampling point is sufficient, but longer waveguides are automatically sampled with regular spacing (which is parametric).

As shown in Figure 6.18, both the waveguide and DC are sampled every 5 μm. Then, the sampled width and thickness are aggregated over the entire waveguide or DC. We use the aggregated values as global variable values that change the model parameters in the next step.

6.3.5 Sensitivity of Model Parameters

After we derived the global variables for each component, we need to calculate the change to the model parameters. In our case, we need to change n_{eff} and n_g of waveguide and coupling coefficients of DC with respect to width and thickness variations. This would require a sensitivity model to map process variations to model parameter change.

To describe the sensitivity, we annotated the existing circuit models with a variability matrix, describing how every circuit model parameter varies with changes in local linewidth and thickness. These annotations can be of the form of Taylor series:

$$C(p) = C_0 + \sum_{i=1}^{n} \frac{1}{n!} \frac{\partial^i C}{\partial X^i} \Delta X(p)^i \tag{6.14}$$

In practice, we will truncate the polynomial to a finite order n. The perturbation of a circuit model parameter C by a global parameter X at position p could be implemented as an n$^{\text{th}}$-order perturbation. For this, the sensitivity of C to X should be known. This can be characterized through measurements, or through simulations. The actual sensitivity data is not generated automatically; if this is not supplied by the fab, it is up to the designer to evaluate this by running simulations or experimentally characterize fabricated devices. By default, the sensitivity of component parameters is set to zero. The good news is: the sensitivity is usually fairly constant even if the geometry of the devices changes slightly.

In our case, calculate the n_{eff} and n_g using the Fimmwave *film-matching mode* (FMM) solver. We swept waveguide width and thickness and used a third-order derivative model and the derivative sensitivity model of n_{eff} and n_g is:

$$n_{eff} = n_{eff_0} + \sum_{i=1}^{3} \frac{\partial^i n_{eff}}{\partial w^i}(w - w_0)^i + \sum_{j=1}^{3} \frac{\partial^j n_{eff}}{\partial t^j}(t - t_0)^j \tag{6.15}$$

$$n_g = n_{g_0} + \sum_{i=1}^{3} \frac{\partial^i n_g}{\partial w^i}(w - w_0)^i + \sum_{j=1}^{3} \frac{\partial^j n_g}{\partial t^j}(t - t_0)^j \tag{6.16}$$

Similarly, we calculated the derivative model for coupling coefficients κ' and $\frac{\partial \kappa'}{\partial \lambda}$ using the Fimmwave FMM solver and the lumped coupling coefficients κ_0 using the Lumerical FDTD simulations. The sensitivity implementation is not limited to the third-order derivatives. It can also use higher-order polynomial expressions, or a custom Python function. The sensitivity data structure is added to the existing models.

6.3.6 Yield Estimation on Fabricated CMZI

To test and verify the extraction workflow we proposed in Chapter 4, we use CapheVE to emulate the fabrication and optical measurements of the CMZI. Table 6.4 shows an example of two CMZIs randomly chosen on the center die on the virtual wafer. The width and thickness of each component in the circuit are altered according to its position. From all the 6084 samples on the wafer, the difference in width in a circuit between low-order arms is always below 0.2 nm. The difference between low-order and high-order arms is below 5.0 nm, which is a larger value because of the larger distance between two stages. The difference in thickness in a circuit is smaller than 0.6 nm. The circuit parameters are revised by the calculated sensitivity (Table 6.5). The difference in width between the low-order stage and high-order stage is smaller than the CMZI design can tolerant. So, we should be able to extract a unique n_{eff} from the CMZI spectrum.

component:variable	CMZI 1 [nm]	CMZI 2 [nm]
arm00: linewidth	3.733	-2.087
arm00: thickness	-0.217	-0.162
arm01: linewidth	3.675	-2.098
arm01: thickness	-0.224	-0.161
arm10: linewidth	3.658	-2.665
arm10: thickness	-0.109	-0.177
arm11: linewidth	3.197	-1.814
arm11: thickness	-0.032	-0.123
dc0: linewidth	3.925	-2.141
dc0: thickness	-0.113	-0.126
dc1: linewidth	3.776	-2.705
dc1: thickness	-0.155	-0.183
dc2: linewidth	3.759	-2.746
dc2: thickness	-0.224	-0.230

Table 6.4: Linewidth and thickness variations in a CMZI circuit at two different sampling locations on the wafer map.

Notice that two arms in the high-order stage arm10 and arm11 have a width difference. From the data we obtain over the virtual wafer, this difference can be as large as 3 nm. When we extract the width from fitting the spectrum, we assume arm width in these two arms is identical. As explained, we extracted the averaged width of these two arms. Quantitatively, we can link the averaged $n_{eff,averaged}$ with n_{eff} on arm10 and arm11 as:

$$n_{eff,averaged}(L_{arm11} - L_{arm10}) = n_{eff,arm11}L_{arm11} - n_{eff,arm10}L_{arm10}$$

component: parameter	CMZI 1	CMZI 2
arm00: n_{eff}	2.360	2.349
arm00: n_g	4.224	4.235
arm01: n_{eff}	2.360	2.349
arm01: n_g	4.224	4.235
arm10: n_{eff}	2.360	2.348
arm10: n_g	4.224	4.236
arm11: n_{eff}	2.360	2.350
arm11: n_g	4.225	4.234
dc0: κ'	0.045	0.045
dc0: $\frac{d\kappa'}{d\lambda}$	0.211	0.216
dc1: κ'	0.045	0.045
dc1: $\frac{d\kappa'}{d\lambda}$	0.211	0.217
dc2: κ'	0.045	0.045
dc2: $\frac{d\kappa'}{d\lambda}$	0.212	0.217

Table 6.5: Changes in circuit parameters by process variations at the same wafer locations of Table 6.4.

The same can apply to the n_g.

$$n_{g,averaged}(L_{arm11} - L_{arm10}) = n_{g,arm11}L_{arm11} - n_{g,arm10}L_{arm10}$$

Assuming the linear relation between width and n_{eff} and n_g, we can approximate:

$$\begin{aligned} \Delta w_{averaged}(L_{arm11} - L_{arm10}) &= \Delta w_{arm11}L_{arm11} - \Delta w_{arm10}L_{arm10} \\ \Delta t_{averaged}(L_{arm11} - L_{arm10}) &= \Delta t_{arm11}L_{arm11} - \Delta t_{arm10}L_{arm10} \end{aligned}$$

Therefore, the extracted width variation $\Delta w_{extracted}$ using the CMA-ES workflow should be compared with the averaged width variation on the high-order stage $\Delta w_{averaged}$

Figure 6.19 shows the 117 transmission spectra from port “in0” to port “out0” generated by the Caphe circuit simulator on the central (X=0, y=0) die. The generated spectra can shift 2 nm from the nominal spectrum where no variation is included. Also, the shapes of the spectra are altered by the change in the coupling, which led to the extinction ratio change.

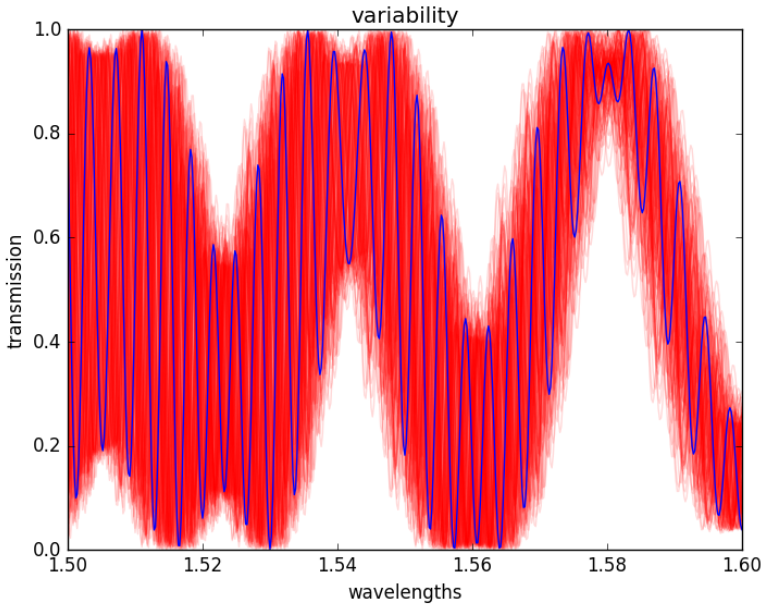


Figure 6.19: Red: 117 transmission spectra from port “in0” to port “out0” generated by the Caphe circuit simulator on the central (X=0, y=0) die. Blue: nominal spectrum.

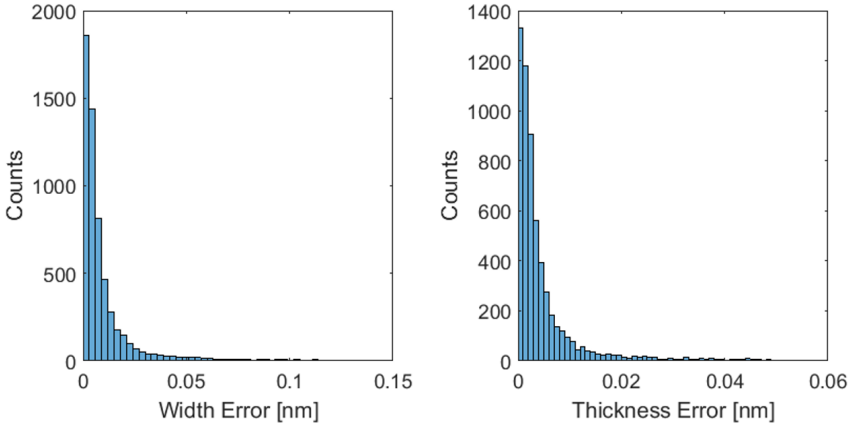


Figure 6.20: To model the effect of variability, the width and thickness variations are sampled on multiple locations in the layout, both for the waveguides and for the directional couplers.

After we obtained the spectra for all the CMZI circuits over the wafer, We used

the extraction workflow in Chapter 4 to extract width and thickness wafer map. Figure 6.20 shows the difference between extracted width variation $\Delta w_{extracted}$ and thickness variation $\Delta w_{extracted}$ compared to averaged width variation $\Delta w_{averaged}$ and the thickness variation $\Delta t_{averaged}$. The difference in extracted width is less than 0.15 nm for width and 0.05 nm for thickness. This is a very good validation of our workflow.

6.4 Conclusion and Future work

In this chapter, we introduced two stochastic analysis methods to do variability analysis. Compared to the MC methods, the stochastic analysis method could significantly reduce the simulation cost. We briefly introduced the gPC methods that have been used to do variability analysis and tolerance optimization for photonics. We also introduced SC methods in details with an example to analyze the variability of DC.

In the second part, we introduced how to analyze the variability of photonic circuits the CapheVE framework. The framework allows us to import the spatial variability model presented in Chapter 5 to make layout-aware yield estimation. The generated virtual wafers exhibit a good match in statistical properties and spatial correlation of the process variations with the fabricated wafer. We used CapheVE to simulate the CMZI spectral responses over a virtual wafer. Using the extraction workflow introduced in Chapter 4, we obtained the linewidth and thickness maps with very high accuracy. It is a good validation of the workflow.

Stochastic analysis methods such as gPC-based methods help to reduce the cost of MC simulations. However, the fact such method assuming process-related variables are purely random leaves a question mark how to integrate such methods with the layout-aware variability analysis such as our approach using CapheVE. Integration of stochastic analysis methods and layout variability analysis might be the next step of our research.

References

[1] George Fishman. *Monte Carlo: concepts, algorithms, and applications*. Springer Science & Business Media, 2013.

[2] Domenico Spina, Francesco Ferranti, Giulio Antonini, Tom Dhaene, and Luc Knockaert. *Efficient Variability Analysis of Electromagnetic Systems Via Polynomial Chaos and Model Order Reduction*. IEEE Transactions on Components, Packaging and Manufacturing Technology, 4(6):1038–1051, 6 2014.

[3] Peter R Johnston. *Defibrillation thresholds: a generalised polynomial chaos study*. In Computing in Cardiology 2014, pages 197–200. IEEE, 2014.

[4] Kwang-Ki K. Kim and Richard D. Braatz. *Generalized polynomial chaos expansion approaches to approximate stochastic receding horizon control with applications to probabilistic collision checking and avoidance*. In 2012 IEEE International Conference on Control Applications, pages 350–355. IEEE, 10 2012.

[5] Dongbin Xiu. *Fast numerical methods for stochastic computations: a review*. Communications in computational physics, 5(2-4):242–272, 2009.

[6] Dario Cassano, Francesco Morichetti, and Andrea Melloni. *Statistical Analysis of Photonic Integrated Circuits Via Polynomial-Chaos Expansion*. In Advanced Photonics 2013, page JT3A.8, Washington, D.C., 7 2013. OSA.

[7] Tsui-Wei Weng, Zheng Zhang, Zhan Su, Youssef Marzouk, Andrea Melloni, and Luca Daniel. *Uncertainty quantification of silicon photonic devices with correlated and non-Gaussian random parameters*. Optics Express, 23(4):4242, 2015.

[8] Tsui Wei Weng, Daniele Melati, Andrea Ivano Melloni, Luca Daniel, and others. *Stochastic simulation and robust design optimization of integrated photonic filters*. Nanophotonics, 6(1):299–308, 2017.

[9] Michael Eldred. *Recent Advances in Non-Intrusive Polynomial Chaos and Stochastic Collocation Methods for Uncertainty Analysis and Design*. 50th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, (May), 5 2009.

[10] Bruno Sudret. *Global sensitivity analysis using polynomial chaos expansions*. Reliability Engineering & System Safety, 93(7):964–979, 7 2008.

[11] Alessandra Petrocchi, Arun Kaintura, Gustavo Avolio, Domenico Spina, Tom Dhaene, Antonio Raffo, and Dominique M. M.-P. Schreurs. *Measurement Uncertainty Propagation in Transistor Model Parameters via Polynomial Chaos Expansion*. IEEE Microwave and Wireless Components Letters, 27(6):572–574, 6 2017.

[12] Nitin Agarwal and N. R. Aluru. *Weighted Smolyak algorithm for solution of stochastic differential equations on non-uniform probability measures*. International Journal for Numerical Methods in Engineering, 85(11):1365–1389, 3 2011.

[13] Nitin Agarwal and Narayana R. Aluru. *Stochastic Analysis of Electrostatic MEMS Subjected to Parameter Variations*. Journal of Microelectromechanical Systems, 18(6):1454–1468, 12 2009.

[14] Alan Weiser and Sergio E Zarantonello. *A note on piecewise linear and multilinear table interpolation in many dimensions*. Mathematics of Computation, 50(181):189–196, 1988.

[15] Volker Barthelmann, Erich Novak, and Klaus Ritter. *High dimensional polynomial interpolation on sparse grids*. Advances in Computational Mathematics, 12(4):273–288, 2000.

[16] Erich Novak and Klaus Ritter. *High dimensional integration of smooth functions over cubes*. Numerische Mathematik, 75(1):79–97, 1996.

[17] Erich Novak and Klaus Ritter. *Simple cubature formulas with high polynomial exactness*. Constructive approximation, 15(4):499–522, 1999.

[18] Baskar Ganapathysubramanian and Nicholas Zabaras. *Sparse grid collocation schemes for stochastic natural convection problems*. Journal of Computational Physics, 225(1):652–685, 2007.

[19] Armen Der Kiureghian and Pei-Ling Liu. *Structural reliability under incomplete probability information*. Journal of Engineering Mechanics, 112(1):85–104, 1986.

[20] Michel Loeve. *Elementary probability theory*. In Probability Theory I, pages 1–52. Springer, 1977.

[21] Wim Bogaerts, Yufei Xing, and Umar Khan. *Layout-Aware Variability Analysis, Yield Prediction, and Optimization in Photonic Integrated Circuits*. IEEE Journal of Selected Topics in Quantum Electronics, 25(5):1–13, 9 2019.

[22] Martin Fiers, Thomas Van Vaerenbergh, Joni Dambre, and Peter Bienstman. *CAPHE: Time-domain and Frequency-domain Modeling of Nonlinear Optical Components*. Advanced Photonics Congress, page IM2B.3, 2012.

[23] Wim Bogaerts, Umar Khan, and Yufei Xing. *Layout-Aware Yield Prediction of Photonic Circuits*. In IEEE International Conference on Group IV Photonics, pages 1–2, Cancun, Mexico, 2018. Ieee.

[24] Wim Bogaerts, Martin Fiers, Massimo Sivilotti, Pieter Dumon, Massimo Sivilotti, and Pieter Dumon. *The IPKISS photonic design framework*. In Optical Fiber Communication Conference, page W1E.1, Washington, D.C., 3 2016. OSA.

7 Conclusion and Future Work

7.1 Conclusion

Silicon photonics is an attractive platform for photonics integration. Its high material contrast and tight light confinement allow a high-level of miniaturization of circuits and functions on-chip, making it a very competitive integration solution. This work targets a crucial challenge for large-scale high-volume photonics integration: how to make a realistic yield prediction, so we can design circuits that can actually work?.

A realistic yield prediction mimics the effects of the physics and chemistry in the actual fabrication flow, and therefore should be aware of the layout and the location of the circuit on a chip. To make such a prediction, we identified three essential steps. First, extract circuit parameters and fabrication wafer maps accurately and robustly. Second, build a variability model to analyze and reconstruct the spatial and statistical features of the wafer maps. Third, integrate the variability model into the design workflow to make realistic yield prediction.

To obtain wafer maps of linewidth and thickness of fabricated wafers, we developed a method to extract waveguide geometry from spectral measurements of two configurations of Mach-Zehnder interferometers with sub-nanometer accuracy. Using a curve fitting method, we made the extraction of waveguide effective index and group index less sensitive to measurement noise. We also used curve fitting to remove the impact of grating coupler envelope in the extraction. We offered analysis to reduce the bounds of the effective index with knowledge of the group

index. The analysis allows us to use a higher-order MZI to improve fitting accuracy. We proposed a procedure to separate different levels of process variations so that our method can deal with a total variation of several tens of nanometers and still obtain accurate linewidth and thickness extraction. We also built an accurate third-order geometry model that we can greatly enhance the accuracy of mapped waveguide geometry from its effective index and group index.

Often, we also need to know the fabricated quality of couplers to estimate performance variation of the optical filters. To evaluate coupler characteristics, we introduced a compact behavioral model of the directional coupler and validated the dispersive model from both FDTD simulations and on-chip measurements. An accurate behavioral model also helps to replace the expensive electromagnetic simulations for components. Using the compact model, a designer can focus on circuit-level design and still has confidence in the simulated response of a circuit. We tested a few methods and circuit designs to extract the coupler model parameters. Our experiments and discussions revealed that a combination of the curve fitting method and an MZI-based circuit design helps us to extract parameters accurately, even in the presence of measurement noise and a grating coupler variation. We provided a detailed discussion and presented results on the coupling coefficients extraction.

To simultaneously extract waveguide and directional coupler performance using the mentioned methods, traditional measurements require at least five circuits, which take up large footprint, introduce considerable local variation induced extraction error, and longer measurement time. Also, it means such circuits can not capture the short-distance features of spatial variations on the chip. To solve the problem, we designed a compact folded two-stage MZI that can be used to extract fabrication parameters. The compact design suffers less from local variation within the circuit, which significantly improves the accuracy of extraction. Also, the circuit greatly reduces the duration of wafer-scale optical measurements, making it useful for process control monitoring and detailed wafer-level variability analysis. The transmission spectra of the circuit are complicated, which requires robust and efficient global optimization method to obtain circuit parameters. We applied the Restart-CMA-ES global optimization algorithm to extract multiple waveguides and DC parameters from only two optical measurements of the circuit. We illustrated how to set up the algorithm in practice to obtain the global optimum efficiently.

The wafer-scale optical measurement and extraction of process wafer maps are very time-consuming and prone to errors. To efficiently use the circuit for process monitoring on wafer-scale and obtain fabricated wafer maps, we applied a stepwise workflow and parallelization to process the data. Using the compact circuit and the workflow, we extracted detailed wafer maps for performance evaluation and variability analysis.

Wafer maps contain rich information about the process variation. We proposed an additive hierarchical model that can decompose variability on various spatial levels into systematic and random variation. We discussed the physical origins of these process variations. Based on that, we proposed a workflow to separate variability on intra-wafer, intra-die level, and introduced wafer-die interaction term to reduce random variation in our analysis. We applied the model and the workflow to process the measurements on a 200 mm wafer fabricated in IMEC's silicon photonics platform based on 193 nm lithography. The result shows that the intra-wafer systematic variation is the primary source of variation for both linewidth and thickness. We observed that the width variation has a systematic dome-like profile across the wafer. Thickness non-uniformity across the wafer looks like a slanted plane with a few mismatches around the wafer edge. At the die level, we found that systematic width pattern is closely related to the local pattern density. Our analysis showed that the intra-die systematic width variation is affected by the pattern within a distance of 195 μm to the site. Our findings help to identify the process variation and create new design rules to alleviate the impact of the non-uniformity. If the systematic variation we observed can be further refined and compensated in the process, we can foresee the photonics wafers with significantly better fabrication uniformity in the future.

To make a realistic yield prediction, we developed the CapheVE framework that combines circuit model, parameter sensitivity, circuit layout, and process variability model. The framework allows us to import the spatial variability model and generate virtual fabrication wafers. Circuit parameters are altered according to its sensitivity and layout and location on the wafer. Using CapheVE, we can place the instances of the circuit over the wafer, and run Monte Carlo simulation to generate the response for all circuits. We can then use the generated responses for layout-aware variability analysis and yield prediction. Using CapheVE, we generated virtual wafers that exhibit a good match in statistical properties and spatial correlation of the process variations with the fabricated wafer. We also used the framework to make yield prediction for wavelength de-multiplexer and to validate the parameter extraction workflow using the compact two-stage MZI.

To reduce the cost of Monte Carlo methods for yield prediction, we tried stochastic analysis methods to do variability analysis. Compared to the Monte Carlo method, the stochastic analysis method could significantly reduce the simulation cost. We applied the stochastic collocation method to analyze the variability of DC. The method reduced the cost of simulation time by 146 times.

In this work, we have also observed surprisingly strong fluctuation in the high-contrast waveguide transmission spectrum. We observe that fluctuations in the spectrum increase drastically with the waveguide length. The 7-cm-long waveguides produce fluctuations above 15 dB. We attributed such fluctuations to the backscattering induced by sidewall roughness. To understand the phenomenon,

we derived a circuit model that models waveguides as a series of very short cascaded lumped sections with a defined loss. To incorporate backscattering, we also introduced a reflection with a random phase change between sections. Our model explains and accurately captures fluctuations in long silicon waveguides. To model the waveguide with backscattering, we also proposed to explain the phenomenon by the wave transport theory in the random media. We found the light backscattering in the waveguide is governed by the Fokker-Planck Equation. The initial result generated by the equation shows a good match with our circuit model.

7.2 Future Work

In the scope of this research, we can still find improvements in the following directions:

1. The folded two-stage MZI is a good circuit for process monitoring control. We can still improve the extraction accuracy for directional coupler parameters by removing the impact of grating coupler variation in transmission. We need to test if the removal works for this circuit. We also need to test what is the cost of this method. Incorporating a grating coupler model in the extraction would add at least five more parameters which might increase the extraction time significantly.

2. We need more wafer measurements to validate the variability model further. Also, this could characterize the value of wafer-to-wafer variation in the model. Now, our extraction circuits are located only in one-eighth of a die as we participated in a multi-project wafer run. We have included the CMZI structure on a fabrication run where we should be able to extract parameters over multiple wafers and full dies. With these information we might improve the modeling of the die-level variation.

3. Now the simulation cost of the layout-aware yield prediction is still high because we used the Monte Carlo method that requires a considerable number of evaluations. Stochastic analysis methods such as gPC-based methods help to reduce the cost of MC simulations. However, the fact such method assumes that the process-related variables are purely random leaves a question mark how to integrate such methods with the layout-aware variability analysis such as our approach using CapheVE where we capture location-dependent correlations between elements in a circuit. Integration of stochastic analysis methods and layout variability analysis it a logical step to push this research to a next level.

4. While the relevance of this research (i.e. yield prediction) is quite clear, it can only become a practical reality if it is actually introduced into the design

flow that is used by a significant fraction of the PIC community. To bring this about, the techniques should become be made sufficiently robust and integrated in the tools that are used by actual designers. We worked together with Luceda Photonics (in the framework of the VLAIO project MEPIC) , to make sure that our techniques could solve actual design problems. As a result of this project, some of the techniques (e.g. the CapheVE framework) will be incorporated in future releases of Luceda's IPKISS framework.

A

Laser Calibration and Stability Test

In Chapter 2, we mention that laser calibration is required to get accurate and consistent behavioral parameters from optical measurements. In this appendix, we will describe in detail the procedure for the laser calibration.

A.1 Pre- and Post- Measurement Laser Calibration

In parameter extraction experiment using optical measurement, the value of the parameter should be only determined by the circuit layout and device geometry. However, an erroneous wavelength shift in the spectral measurement would lead to a serious error in the extraction of a parameter such as the effective index and group index. Problems such as laser drifting and instability may rise such a measurement error.

A tunable laser might suffer from wavelength drifting over the year-long usage, especially when it is lack of good maintenance. The drifting could be the aging of the mechanics to sweep the wavelength that results in a linear transformation of the sweeping wavelength. Problems such as wavelength shifting and broadening occur. Such an issue is not that noticeable if absolute value peak wavelength or free spectral range is not the focus.

To calibrate the tunable laser, we need to use a reference of the spectrum with absolute dip values. The absorption spectrum of the gas cell has dips at a few certain wavelengths when the temperature is fixed. We can measure the gas cell, and compare the measured dips which wavelengths are read from the laser with

literature to calibrate the tunable laser.

For example, when we calibrate our tunable laser with a tunable range from 1500 to 1630 nm, we want to find gas with absorption dips also spread out in that range. Ammonia (NH_3) gas has many absorption dips located from 1500 to 1550 nm. We have an ammonia gas cell with a cell length of 2.5 cm. We can find the transmittance spectrum of the gas cell from an online database such as HITRAN. [1] HITRAN is an acronym for high-resolution transmission molecular absorption database. HITRAN is a compilation of spectroscopic parameters that a variety of computer codes use to predict and simulate the transmission and emission of light in the atmosphere. We simulate the transmittance using HITRAN with a temperature of 20 degree Celsius as the controlled temperature in our cleanroom. The pressure of the gas cell is 740 torr. Fig. A.1 is the simulated transmittance spectrum.

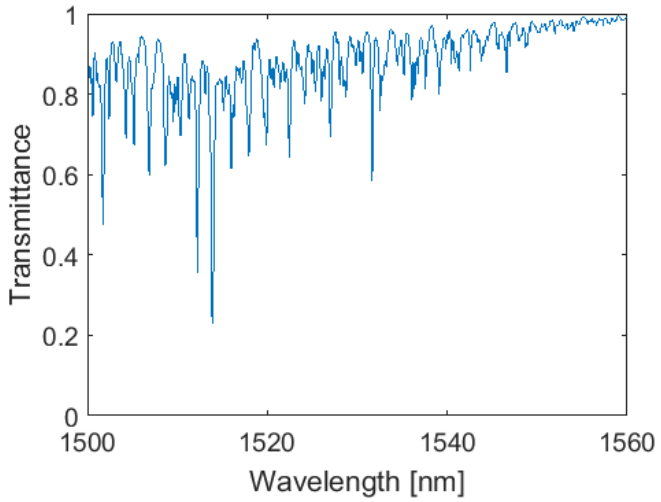


Figure A.1: Transmittance spectrum of the ammonia gas cell obtained from HITRAN database. A.1

As in Fig. A.2, in the next step, we measured the gas cell with the tunable laser we want to calibrate. We found out the most distinguished dips and matched measured transmission dips with the HITRAN database. We fitted a linear relation between actual wavelengths $\lambda_{database}$ of the dips referred from the database and the dip wavelengths λ_{laser} used by the tunable laser.

$$\lambda_{laser} = p_0 + p_1 \cdot \lambda_{database} \tag{A.1}$$

As in Fig. A.3, we observed a perfect linear relation between two sets of

wavelengths and the residual of the fitting is negligible. Using the linear relation, we can calibrate the laser and correct the wavelength of the measured spectrum.

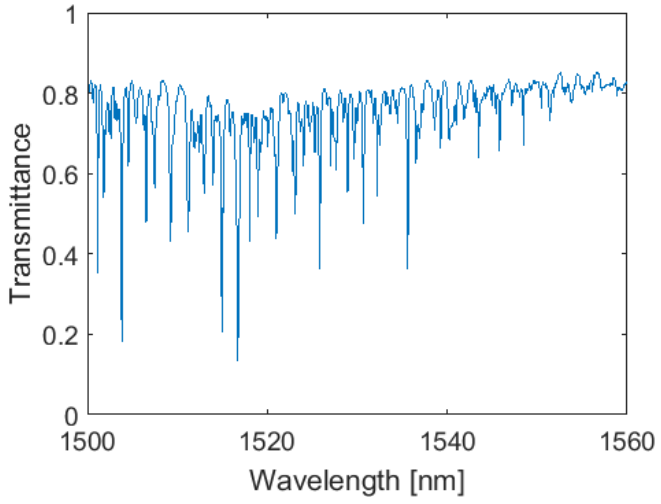


Figure A.2: Transmittance spectrum of the ammonia gas cell measured by a tunable laser that needs calibration.

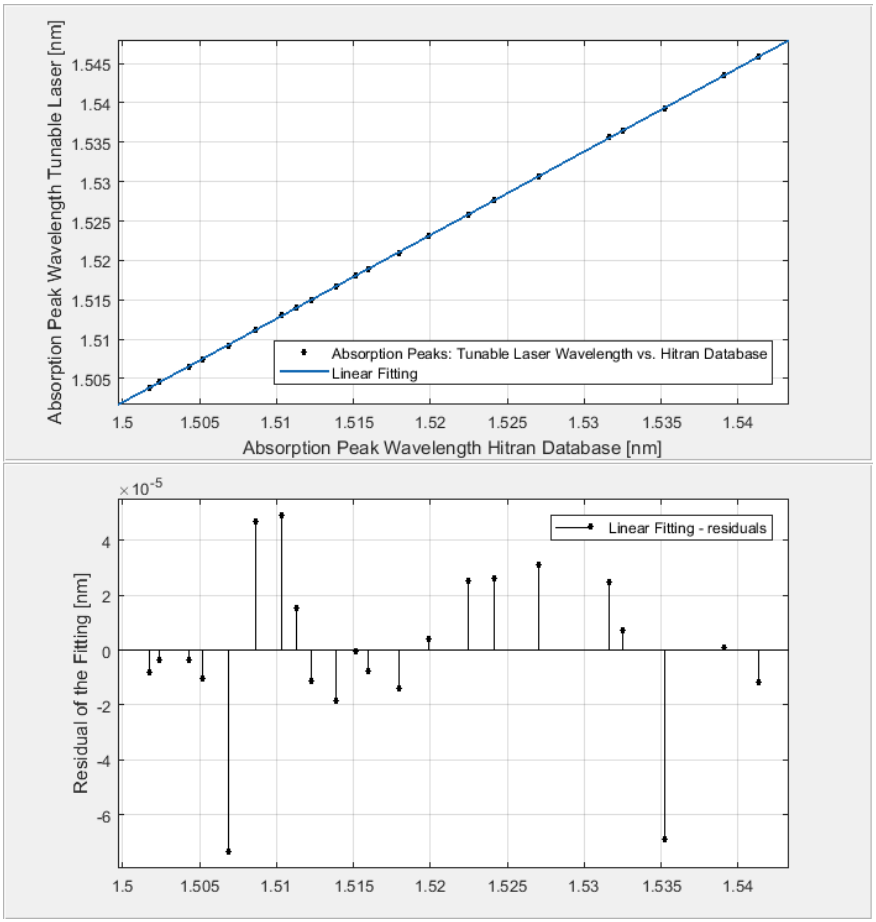


Figure A.3: Dips in the transmittance spectrum measured by the tunable laser follows a linear relation with the actual dips wavelengths simulated by the HITRAN database. Top: linear fitting. Bottom: the residual of the fitting.

A.2 Stability Test During the Measurement

After the calibration of the tunable laser, we still need to ensure the laser is stable throughout the measurement. Since wafer-scale measurement takes weeks to months to finish, bad thermal control or unstable mechanics of the wavelength tuning might lead to a variation of the swept wavelength during the lengthy measurement. So we need to incorporate stability test in the measurement workflow.

The most straightforward way is to measure the gas cell once a day during the whole measurement. If the spectra are identical, the laser is stable. Assembly of the gas cell might be impractical during the wafer-scale measurement. Alter-

natively, you can measure the spectrum of an interfering device such as a ring resonator periodically.

In one of our wafer measurement, we have 117 devices under test on each die. We measure one to two devices on each die over the wafer before the measurement. After the complete wafer measurement, we have two measurements on this one to two reference devices. We could compare the spectra. Since there is one to two measurement on each die, we can test the laser stability throughout the wafer measurement.

References

[1] *The HITRAN Database*.

B

Karhunen-Loéve expansion

B.1 Karhunen-Loéve expansion and Correlated Gaussian Random Variables

Let us assume that the correlation matrix $\mathbf{C}^{N\times N}$ for the random variables $\boldsymbol{\eta}$ under study is symmetric and positive-definite. Then, $\mathbf{C}$ and can be diagonalized as

$$\mathbf{C} = \mathbf{V}\mathbf{E}\mathbf{V}^T \tag{B.1}$$

Thanks to (B.1), equation (6.12) becomes

$$W_{\boldsymbol{\eta}} = \frac{1}{2\pi det(\mathbf{E})^{\frac{1}{2}}} exp\left(-\frac{1}{2}\left(\boldsymbol{\eta}-\boldsymbol{\mu}\right)^T \mathbf{V}\mathbf{E}^{-1}\mathbf{V}^T\left(\boldsymbol{\eta}-\boldsymbol{\mu}\right)\right) \tag{B.2}$$

Hence, the Karhunen-Loéve expansion is a simple change of variables for correlated Gaussian random variables following the non-degenerate multivariate normal distribution (6.12). Furthermore, it is possible to express the joint probability density function (B.2) with respect to a vector of independent Gaussian random variable $\boldsymbol{x}$, with zero mean and variance equal to $[\mathbf{E}_{ii}]_{i=1}^{N}$, as

$$W_{\boldsymbol{x}} = \frac{1}{2\pi det(\mathbf{E})^{\frac{1}{2}}} exp\left(-\frac{1}{2}\boldsymbol{x}^T\mathbf{E}^{-1}\boldsymbol{x}\right) \tag{B.3}$$

where

$$\boldsymbol{x} = \mathbf{V}^T\left(\boldsymbol{\eta}-\boldsymbol{\mu}\right) \tag{B.4}$$

Finally, the vector $\boldsymbol{x}$ can be written as

$$\boldsymbol{x} = \mathbf{E}^{\frac{1}{2}}\boldsymbol{\xi} \tag{B.5}$$

where $\boldsymbol{\xi}$ is a vector of normalized Gaussian random variables with zero mean and unitary variance. Equation (6.13) can be obtained by combining (B.4) and (B.5).

C

Smolyak algorithm

C.1 Smolyak algorithm

Let us express a stochastic process $\mathbf{Y}$ depending on one random variable ξ by means of the Lagrange interpolation scheme as [1]

$$\mathbf{U}(\xi) = \sum_{i=1}^{Q} \mathbf{Y}\left(\xi_i\right) L_i\left(\xi\right) \tag{C.1}$$

where L_i is given by equation (6.5). The Q nodes can be chosen from a node distribution which guarantees a good quality of the approximation (i.e. the extrema of the Chebyshev polynomials). Extending (C.1) to the case of multiple random variables can be performed via tensor product, as it has been shown in Section 6.2.2, and equation (C.1) becomes

$$\mathbf{Y}\left(\boldsymbol{\xi}\right) = \mathbf{U}^{k_1} \otimes \cdots \otimes \mathbf{U}^{k_N} = \sum_{i_1=1}^{Q_{k_1}} \cdots \sum_{i_N=1}^{Q_{k_N}} \mathbf{Y}\left(\xi_{i_1}^{k_1}, \ldots, \xi_{i_N}^{k_N}\right) \left(L_{i_1}^{k_1} \otimes \cdots \otimes L_{i_N}^{k_N}\right) \tag{C.2}$$

where $\mathbf{U}^{k_j}$ represents the interpolation scheme in the form (C.1) with respect to the random variable ξ_j and N is the number of random parameters considered. The total number of nodes required to compute (C.2) is the given by the product of the nodes used for each random parameter, as shown in (6.7). Clearly, the required number of nodes grows very quickly with respect to the number of parameters

considered. Indeed, if only two nodes are used for each random variable, the total number of points required for a full-tensor product interpolation is $Q = 2^N$.

The Smolyak algorithm allows to build multi-dimensional interpolation functions based on a minimal number of nodes by expressing the desired interpolation as a linear combination of tensor products. In particular, the property of the one-dimensional interpolation is conserved for higher dimensions. Indeed, the sparse interpolant $\boldsymbol{A}_{q,N}$ given by the Smolyak algorithm is

$$\boldsymbol{A}_{q,N}(\boldsymbol{\xi}) = \sum_{q-N+1 \leq |\boldsymbol{k}| \leq q} (-1)^{q-|\boldsymbol{k}|} \begin{pmatrix} N-1 \\ q-\boldsymbol{k} \end{pmatrix} (\mathbf{U}^{k_1} \otimes \cdots \otimes \mathbf{U}^{k_N}) \tag{C.3}$$

where $q - N$ is the order of interpolation, $A_{N-1,N} = 0$ and $\boldsymbol{k} = (k_1, \ldots, k_N)$ with $|\boldsymbol{k}| = k_1 + \cdots + k_N$. Hence, the interpolation function is built by adding a combination of one dimensional interpolant of order k_j with the constraint that the total sum $|\boldsymbol{k}|$ across all parameters is between $q - N + 1$ and q. Note that, k_j can be considered as the interpolation level along the j-th direction.

Let us denote Θ as the set of points utilized in the one-dimensional function interpolation. According to (C.3), the stochastic process $\mathbf{Y}$ must be computed at the nodes of the sparse grid $\mathbf{H}_{q,N}$ given by

$$\mathbf{H}_{q,N} = \bigcup_{q-N+1 \leq |\boldsymbol{k}| \leq q} \Theta_1^{k_1} \times \cdots \times \Theta_N^{k_N} \tag{C.4}$$

It is important to notice that, by choosing a suitable node distribution, such as Chebyshev or Gauss-Lobatto points, the sets of collocation points Θ^k are nested. Hence, the sparse grid of order q contains all the nodes computed for the sparse grid of order $q - 1$ and the stochastic process $\mathbf{Y}$ must be evaluated only on few new collocation points.

References

[1] Baskar Ganapathysubramanian and Nicholas Zabaras. *Sparse grid collocation schemes for stochastic natural convection problems*. Journal of Computational Physics, 225(1):652–685, 2007.

D

Problems and the Procedure to Extract and Correct Parameters on a Fabricated Wafer

D.1 Problems and the Procedure to Extract and Correct Parameters on a Fabricated Wafer

Our ultimate purpose is to use the PCM circuit practically to obtain fabricated wafer maps as we will show in Chapter 5. We have discussed how to extract multiple parameters one circuits. To extract the parameters on a complete wafer is challenging for a couple of reasons. First, the extraction from a complicated spectrum requires many evaluations, which can be time-consuming. If we cannot reduce the cost of circuit simulation, extraction from thousands of circuits will be impractical. Second, the number of PCM circuits on the wafer is enormous. Even with very efficient circuit simulator, it would still take a month to generate a wafer map. Third, global optimization could be time-consuming if we apply to the same strategy for all types of spectrum. In this section, we discuss the problems and solutions. We also propose our workflow to extract fabricated wafer maps.

D.1.1 Reducing the Cost of Circuit Simulation

First, the extraction from a complicated spectrum requires many evaluations. If each evaluation of circuit simulation is not fast enough, the extraction can be very

time-consuming. In our case, the extraction from on circuit using the CMA-ES optimization method takes around 8,000 evaluations to find a good solution. However, we set the program to restart once it reaches 20,000 evaluations without finding a good solution. Using the old Caphe circuit simulation, it takes 1 second for one circuit simulation with 200 wavelengths in the spectrum. It will cost more than several hours to process just one circuit. It is challenging even to try different parameter and boundary setting to get the CMA-ES working consistently. Now, with the help of people from Luceda and the release of the Caphe, we can run the same circuit simulation 30 times faster. Now, most of the time we can obtain a solution in around 6 minutes. It does not only help us to test the algorithm and make it running, but also very important if we want to use it for wafer map extraction.

D.1.2 Step-Wise Workflow

Second, the number of PCM circuits on the wafer is enormous. Our wafer consists of more than 6,000 circuits to be analyzed. In the best case when we can found a solution in around 6 minutes, it still costs more than 25 days to process the complete data set. Not to mention that many circuits need a few restarts to find the global solution which further increases the extraction cost. Meanwhile, for such a large amount of dataset, we can hardly avoid an error in measurements and extraction. In the automated measurement, one dies might be measured with a mistaken die number if we set the parameter or coordinate wrongly in the measurement. Also, if the tunable laser is not stable, there would be laser wavelength drifting that makes the entire measurement wasted. In addition, the spectrum for a particular die could be mostly saturated when we choose an improper wavelength to determine the photodetector dynamic range. Spotting that type of error at an early stage rather than noticing that after the monthly-long measurement and lengthy extraction workflow saves time and unnecessary repetitions.

It is reasonable to generate intermediate results for analysis and inspect possible errors occurred in the measurements and extraction. Using the step-wise workflow can offer not only intermediate result but also makes sure the result is correct before moving to the next. Another reason is that we can speed up the extraction process by using a different strategy to deal with diverse spectrum in a few steps rather than in one step. Here we discuss in detail under which circumstances, and we should follow the step-wise workflow to process the wafer data.

D.1.2.1 Global Optimization in Two Steps

The Restart CMA-ES global optimization technique is very powerful and robust. If we have the correct model, correct measurement with the noise controlled at a reasonable level, right parameter boundary and sufficient evaluations, we should always be able to find the correct extraction. However, 'global' is a word requires

attention. There are two ways to operate the Restart CMA-ES. To increase the population size after each restart which makes the search more global, or use the same size after each restart. They have different advantages. You would suppose that you can find the optimum quickly if you use the more global approach. You increase the population size after restart so that the samples in the population cover more searching area. However most of the time, we found that the choice actually makes it slower to converge to the right solution. Sometimes, it fails even to find a local optimization within 20,000 evaluation. So the 'global' approach covers larger search range and could always find the optimum, but it requires far more evaluations to find it. On the other hand, the restart scheme with the same size of the search span is more local but quite efficient. For a majority of our measurements, when we did not find a global optimum in one search with maximum evaluations of 20,000, we can restart the search with the same setting and find a right solution after a few restarts. But it did fail to find a solution in some spectrum. So, we used the 'local' approach to process all the data. If it fails to work, we use the 'global' method to process the difficult ones. In this way, we can combine the merits of both approaches.

D.1.2.2 Improving Fitting Accuracy in Two Steps

As shown in Figure 4.4, the radius of the search circle reduces generation by generation to approach the optimum. The size of the radius is the standard deviation error on the parameter. In other words, the radius is proportional to the fitting error. To improve the fitting accuracy, we could use lower function tolerance to make sure that the difference between two function evaluations is sufficiently small and in most cases, it means the searching radius or the fitting error is enough small. However, this increases the number of evaluations significantly. For example, if you decrease the function tolerance from 10^{-6} to 10^{-9}, it usually means 3000 more evaluations. In many cases, we can obtain a low fitting error with 10^{-6} tolerance. Also, even we set the lower function tolerance, the algorithm is very likely to terminate because it reaches 20,000 maximum evaluations.

In practice, there are quite some extractions find the global optimum (f-value ¡ 0.02 $\times$ time of wavelengths of the power spectrum) with a slightly big fitting error because the optimization terminates at 20,000 evaluations with a low f-value. Of course, we can also set the fitting error as an option to restart the searching. However, searching to reduce fitting error is quite different from searching for a global optimum. In this case, we are already very close to an optimum, and we just want more evaluations in the neighborhood around the obtained parameter to reduce the searching radius or fitting error.

What we do is after we batch-processed the data on the wafer, we pick up the sample that reaches a global optimum while having a fitting error larger than our standard. We can rerun a very local optimization around the obtained parameters

in the searching space, and the algorithm will quickly find out the new parameters (which has little difference compared to the previous parameters) with a low fitting error. In practice, we convert the fitting error of the behavioral parameters to width and thickness fitting errors. We limit the fitting error to 1 nm for width and 0.5 nm for thickness. The two-step approach limits the fitting error with reduced extraction cost.

D.1.3 Parallelization

There are different ways of parallelization. We chose to divide our data into dies and process several dies simultaneously. For example, we used two servers each with 16 cores and each core process data on one die. The parallelization can reduce the processing time from around one month to just one day.

We divided the data by die for a few reasons. First, certain steps in our workflow need die maps. For example, we need to a fitted die map of the n_{eff} of the low-order stage as a reference to correct the order of the high-order n_{eff}. Another instance is when we need a thickness die map to detect obvious outliers. If we group data by die, we do not have to wait for the whole wafer to be processed and could proceed to steps afterward to get the die map. Also, the optical measurement is executed die after die. We can start to process the finished dies while waiting for the measurements on other dies. Lastly, when we have processed the first samples on every die, we could use them to plot a simple wafer map that helps to identify if we have a wrong numbering on the dies in the automated measurements.