

# Robust design algorithms for silicon photonics

by

Allan Elvio Sadun

Submitted to the Department of Electrical Engineering and Computer  
Science

in partial fulfillment of the requirements for the degree of

Master of Engineering in Electrical Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2019

© Massachusetts Institute of Technology 2019. All rights reserved.

Author .....  
Department of Electrical Engineering and Computer Science  
May 23rd, 2019

Certified by .....  
Dr. Luca Daniel  
Professor  
Thesis Supervisor

Accepted by .....  
Dr. Katrina LaCurts  
Chair, Master of Engineering Thesis Committee



# Robust design algorithms for silicon photonics

by

Allan Elvio Sadun

Submitted to the Department of Electrical Engineering and Computer Science  
on May 23rd, 2019, in partial fulfillment of the  
requirements for the degree of  
Master of Engineering in Electrical Engineering

## Abstract

Recent developments in inverse design have demonstrated the power of computer algorithms for designing new and more efficient silicon photonic devices. However, most photonic design remains in human hands, because of the need to design for manufacturability and mitigate the impact of process variations, where computational tools and models have lagged behind. In this work, I investigate the viability and limitations of modeling and efficiently predicting the effects of one type of process variation, line edge roughness, using the adjoint simulation method. I then develop a robust optimization framework for designing devices that mitigate the impact of line edge roughness, using a hybrid worst-case / average-case scenario construction procedure, and demonstrate its application on the design of a Y-branch splitter. The resulting design experiences 20% less variation in the output imbalance between its two ports compared to the nominal design it started with, in addition to a decrease in insertion loss. Although the optimization algorithm suffers from overfitting as currently implemented, it remains a successful proof-of-concept and its framework is broadly applicable.

Thesis Supervisor: Dr. Luca Daniel

Title: Professor



## Acknowledgments

This thesis would not have been possible without the assistance and support of many people. First and foremost, I am exceptionally indebted to my advisor Luca Daniel, who gave me financial support, connections, opportunities to travel, and incredible academic freedom, even as I spent months working in directions that ended up being unproductive. I enrolled in the MEng program at MIT in part to increase my technical expertise, but also to have the opportunity to struggle with formulating, scoping, and completing a research project, and to grow more independent, something which Luca pushed me to do far more than I expected.

This work was funded by AIM Photonics, as part of a Design for Manufacturability project. In that context, I worked closely with the members of Duane Boning's Statistical Metrology Group, who, in addition to being a great sounding board for ideas and teaching me about photolithography and photonic design tools, have keen eyes for careful data analysis and the kinds of errors that can occur in simulated experiments. I am particularly grateful to Zhengxing Zhang, who introduced me to the adjoint method and helped me debug my code as I implemented it, and Tsui-Wei Weng, who introduced me to uncertainty quantification and onboarded me into the photonics project. I would also like to acknowledge my officemates/classmates José Cruz Serralles, Nicolas Arango, and Lilly Chin, for their advice and friendship.

My time at MIT has been the most enriching of my life so far, thanks primarily to the many vibrant student communities and activities that are organized here. I could not have made it through this program without my sincere, supportive, and delightful housemates, who are all truly singular people. And I am beyond grateful for my family, who gave me life, and love, and brought me to this moment.

Lastly, the diverging colormap used for visualizing adjoint sensitivities was created by Kenneth Moreland, and referred to me by Lizhou Sha.



# Contents

<b>1</b>	<b>Introduction</b>	<b>13</b>
<b>2</b>	<b>Background</b>	<b>15</b>
2.1	Nanophotonic fabrication . . . . .	15
2.2	Photonic devices . . . . .	16
2.2.1	Ring resonators . . . . .	16
2.2.2	Directional couplers and Y-splitters . . . . .	18
2.2.3	Coupled-resonator optical filters . . . . .	19
2.3	The adjoint method . . . . .	20
2.3.1	General case . . . . .	20
2.3.2	Adjoint method for electromagnetism . . . . .	21
2.3.3	Adjoint method for S-parameters . . . . .	22
2.4	Optimization under uncertainty . . . . .	24
2.4.1	Model-based approaches . . . . .	25
2.4.2	Model-free approaches . . . . .	26
<b>3</b>	<b>Testing the adjoint method in the presence of roughness</b>	<b>27</b>
3.1	Notation and implementation details . . . . .	28
3.2	Effect of roughness on a smooth Y-splitter . . . . .	29
3.3	Effect of roughness on a smooth directional coupler . . . . .	33
3.4	Effect of roughness on a smooth ring resonator . . . . .	36
3.5	Effect of smooth design changes on a roughened Y-splitter . . . . .	38

<b>4</b>	<b>Robust shape optimization of Y-splitter against line edge roughness</b>	<b>41</b>
4.1	Algorithm . . . . .	41
4.2	Implementation on Y-splitter . . . . .	43
4.3	Analysis . . . . .	45
<b>5</b>	<b>Conclusions and Possibilities</b>	<b>49</b>



# List of Figures

2-1	Diagram of layers in a silicon-on-insulator wafer. . . . .	15
2-2	A ring resonator coupled to two bus waveguides in an add-drop configuration. . . . .	17
2-3	Top: a directional coupler. Bottom: a y-branch splitter. . . . .	19
3-1	The X (left) and Y (right) refractive indices reported by Lumerical’s conformal meshing algorithm on a material boundary. . . . .	29
3-2	Left: lower port S-parameters for a Y-splitter under 10 roughness families with correlation lengths of 20nm and 400nm, with constant-magnitude reference lines. Right: transmitted power through both ports for all 800 simulations, with averages for each correlation length of roughness, and zero-loss and zero-imbalance reference lines. . . . .	30
3-3	Heatmaps of the predicted sensitivity of various Y-splitter quantities to local outward wall motion, in $m^{-2}$ . . . . .	32
3-4	Normalized prediction errors, averaged across profiles, for Y-splitter total power loss and imbalance under 50nm-correlated roughness. . . . .	33
3-5	Electric field intensity in a directional coupler at 1550nm. . . . .	34
3-6	Transmitted power at 1500nm, 1550nm, and 1600nm through ports of a directional coupler for all 600 simulations, with averages for each correlation length of roughness, and a zero-loss reference line. . . . .	35
3-7	Normalized prediction errors, averaged across frequencies and profiles, for directional coupler transmission quantities under 50nm-correlated roughness. . . . .	37

3-8	Left: the ring resonator spectra produced by a single 50nm-correlation family of roughness profiles with amplitudes from -20nm to 20nm. Right: All 600 extracted resonant frequencies, with the average for each correlation length highlighted. . . . .	38
3-9	Trajectories of smooth design changes on a roughened Y-splitter, contrasted with the trajectories of the effects of smooth perturbations on a smooth Y-splitter. . . . .	40
3-10	Prediction error from linear and quadratic models on the effect of smooth design changes on the total loss from a roughened Y-splitter.	40
4-1	Snapshots of the designs produced by the algorithm as it progressed, for both its nominal-seeded run (top) and its triangle-seeded run (bottom).	44
4-2	Performance of the starting, intermediate, and final Y-splitter designs from the nominal-seeded optimization, under 100 Monte Carlo roughenings. . . . .	45
4-3	Penalty function values for the considered roughness scenarios (including “no roughness”) over the course of the optimization, with lines connecting results from the same scenario. . . . .	47
4-4	Penalty function values for some roughness scenarios over the course of the nominal-seeded optimization, with dashed lines illustrating how the penalty function was forecasted to evolve. . . . .	47

# List of Tables

3.1	Linear regime crossover points for various transmission quantities under various correlation lengths of roughness. N/A means the linear and quadratic prediction errors showed no sign of approaching crossing over.	31
4.1	Performance metrics of the starting, intermediate, and final Y-splitter designs under 100 Monte Carlo roughenings. . . . .	45



# Chapter 1

## Introduction

Silicon nanophotonics is a rapidly emerging new platform for a variety of applications, including sensing, quantum computing, supercomputing, and telecommunications [6]. These applications are well-suited to nanophotonic platforms because of their need for low energy consumption and/or high bandwidth. Nanophotonic platforms, in turn, have seen success being developed in silicon because of the decades of investment and research that the CMOS electronics industry has put into creating highly precise silicon microfabrication capabilities [2]. However, because nanophotonic devices are fundamentally very different kinds of devices than CMOS transistors (i.e. they are analog waveguides, rather than digital switches), integrated photonic circuits are often very sensitive to manufacturing process variations, and designing for manufacturability is increasingly important as greater reliability is demanded of silicon photonic devices.

As the field of silicon nanophotonics matures, there is also an increased need for computational tools to assist in the design and analysis of devices and circuits. Particularly promising has been the development of *inverse design* techniques [24][12], which, rather than using human design principles, create often non-intuitive devices from scratch using optimization. These inverse design algorithms are able to quickly optimize over an entire device's geometry by using the adjoint method, a powerful trick for computing the shape derivatives of electromagnetic quantities. However, when it comes to incorporating robustness, even inverse design approaches fall back

on heuristics, such as smoothness constraints or broadband performance [28], as there has not yet been a demonstration of inverse designing directly for robustness. As the authors of [19] put it:

Several key improvements would enable the widespread use of inverse-design methods in practical applications. First and foremost is improving the robustness of design methods to handle process variations in photolithography, which would enable high-throughput fabrication.

This thesis demonstrates a shape optimization algorithm that directly incorporates a multi-scenario approach to robustness, specifically robustness to line edge roughness. Chapter 2 gives background on photolithography process variations, a few types of passive photonic devices that could be optimized, the adjoint method in an electromagnetic context, and different frameworks for design optimization under uncertainty. Chapter 3 analyzes the effects of line edge roughness on a few different passive photonic devices, and determines the limitations of the adjoint method. Chapter 4 details the optimization algorithm used and evaluates its success, and Chapter 5 concludes. All experiments in this work were conducted in a commercial electromagnetic simulator, using a theoretical model of line edge roughness.

# Chapter 2

## Background

### 2.1 Nanophotonic fabrication

A common platform for photonics is silicon-on-insulator, in which silicon waveguides are surrounded by silicon dioxide to produce a four-layer wafer (see Figure 2-1. A typical dimension set (as in [6]) would be  $725\mu\text{m}$  of thickness for the substrate,  $2\mu\text{m}$  for the buried oxide, and  $220\text{nm}$  for the silicon device layer, in a wafer  $200\text{mm}$  wide. When manufacturing a silicon-on-insulator device, first the silicon device layer is deposited in its entirety. Next it is photolithographically etched to form the desired device shape, before the cladding oxide layer is deposited on top of it. There are three dominant types of variation coming out of this process:

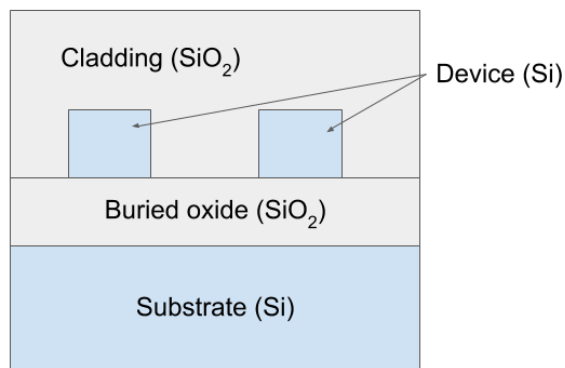


Figure 2-1: Diagram of layers in a silicon-on-insulator wafer.

- **Thickness variations.** The silicon deposition process is not perfect, and can lead to fluctuations in device thickness on the order of  $\pm 5$  nm [6][26].
- **Width variations.** If the photolithography etch process is under- or overdone, the resulting structures can end up too narrow or wide, also on the order of nanometers. This can come about due to miscalibration, but also due to local effects such as gas consumption, which depend on local pattern densities and other design factors [1]. Both thickness and width variations have correlation lengths on the order of millimeters, meaning they are mostly constant within a device [13].
- **Line edge roughness.** Due to the quantized nature of light and the sensitivity of the exposure step in photolithography, the etching process does not produce flat walls, but rather a roughness correlated on the length scale of tens of nanometers [16]. Roughness causes scattering losses that can be predicted using the Payne-Lacey model [11][3], which treats the perturbed edges of a waveguide as radiating dipole sources. Under the Payne-Lacey model, the greatest scattering losses come from fluctuations at spatial frequencies near  $\beta$ , the mode wavenumber in the waveguide. In addition to scattering loss, however, roughness can lead to systematic performance changes, if the random wall fluctuations happen to align in a sensitive part of the device.

## 2.2 Photonic devices

### 2.2.1 Ring resonators

At its most basic, a ring resonator is an optical waveguide that loops back on itself. At wavelengths which are an integer fraction of the optical path length of the loop, the ring resonates, allowing for highly wavelength-selective filters and switches, such as the add-drop filter in Figure 2-2. Off-resonance, the coupling between the ring and the waveguides is weak, but a signal at the right frequency will pass through.



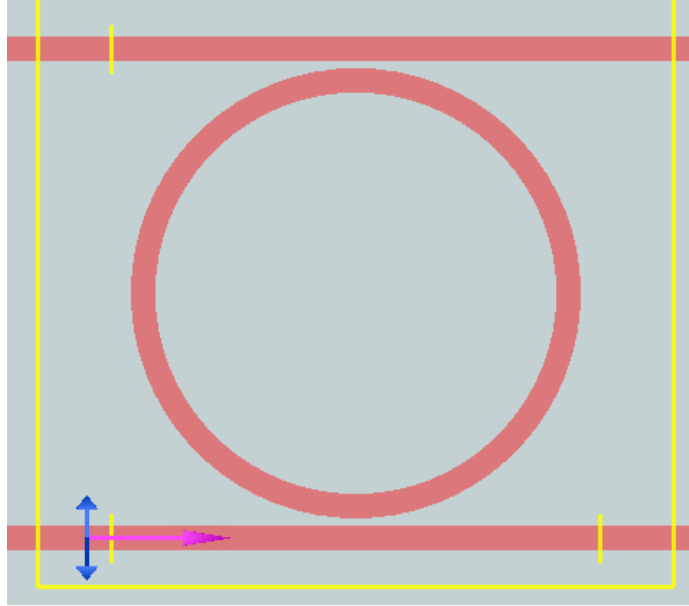


Figure 2-2: A ring resonator coupled to two bus waveguides in an add-drop configuration.

Ring resonators are described by their free spectral range (FSR), which is the spacing between consecutive resonant wavelengths, and their quality factor ( $Q$ ), which measures both the ability of the resonator to hold energy without losses as well as the spectral width of the resonances. There is a tradeoff between these two values, as in order to operate a ring resonator as a bandpass filter, the FSR must be larger than the spectral band of the system's operation. Increasing the FSR requires decreasing the size of the ring, which increases bending losses and degrades the quality factor  $Q$ .

Although sidewall roughness affects ring resonators by introducing loss and decreasing  $Q$ , ring resonators are more strongly affected by width and thickness variations, which alter the group index and shift the resonant wavelength, something that most applications are strongly sensitive to. It is therefore common to manufacture resonators alongside micro-heaters, which can (if controlled) use the thermo-optic effect to adjust the effective index post-fabrication.

If a resonator is made of doped silicon and equipped with an electrical contact, then its resonance can be directly shifted more quickly than using thermal tuning. This forms the basis for most active photonic switches. However, the electrical contacts can disrupt the mode confinement of the ring resonator, particularly if the ring is

small. One approach for dealing with this is to slowly vary the waveguide width going around the ring, so as to adiabatically shift the optical mode to one which is less sensitive to disruptions. These adiabatic resonators have been shown to not only be less disrupted by the presence of an electrical contact, but also less sensitive to width and thickness variations.

### 2.2.2 Directional couplers and Y-splitters

A directional coupler is a compact four-port device, typically designed as two parallel waveguides (as shown in Figure 2-3). As light propagates down one of the waveguides, it slowly couples into the other waveguide, leading to a splitting or combining effect. (This can be designed to achieve whatever coupling ratio is desired, whether 50/50 or 99/1.) Mathematically, the coupling can be precisely understood by comparing the effective indices of the symmetric and antisymmetric “supermodes” in the two-waveguide system. But generally, the coupling decreases exponentially with gap width and increases sinusoidally with coupling length. Directional couplers are distinguished from adiabatic couplers, which are much longer to avoid higher-order mode excitation, and multimode interference couplers, which join the two waveguides into a single cavity but are also larger [29].

Because effective mode indices are wavelength-dependent, the coupling ratio of a directional coupler is wavelength-dependent as well. There have been a few different design strategies to try and cancel out the wavelength dependence and create a broadband directional coupler, including superimposing a sub-wavelength grating on the coupling region [29], adding an asymmetrical “phase control” region where one waveguide is wider than the other [14], and adding an intermediately-etched slab surrounding the waveguide [9]. Interestingly, when a computer design algorithm attempted this task, it produced a design reminiscent of sub-wavelength gratings, but through oscillations in the waveguide thickness [25].

In situations such as interferometry where a 50/50 power split is desired, the symmetry of a single-input Y-branch design (also in Figure 2-3) is more reliable than that of using one side of a directional coupler. Although a Y-branch splitter may

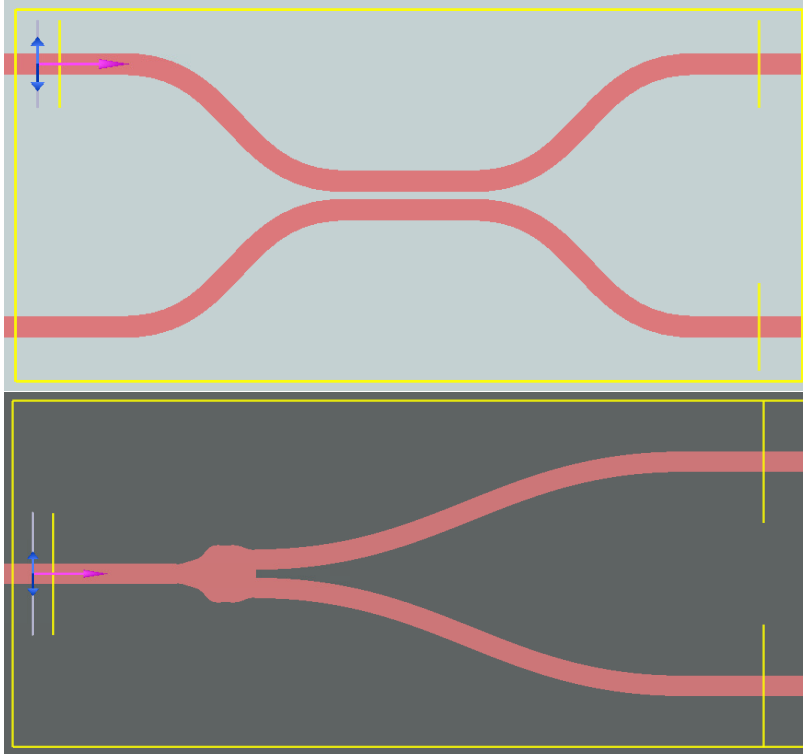


Figure 2-3: Top: a directional coupler. Bottom: a y-branch splitter.

still have insertion loss or backreflections, uniform width or thickness variations will not lead to a difference in output between the two arms. However, because line edge roughness is a random process, it can still imbalance a Y-splitter, as can a vertical gradient in width or thickness variations.

### 2.2.3 Coupled-resonator optical filters

As discussed previously, a single ring resonator between two bus waveguides acts as a bandpass filter. Its bandwidth is controlled by the quality of the resonator and the coupling coefficients; however, its roll-off is uncontrolled, and the filter contrast goes down as the coupling strength goes up. A sharper (higher-order) filter can be achieved by putting multiple ring resonators in series, and its characteristics can be tuned by adjusting the coupling coefficients between the rings [17]. The downsides of using a coupled-resonator optical filter, however, are that more tuners and footprint are required to deal with each resonator, and that there is increased variability due

to the number of coupling coefficients which might drift due to width or thickness variations [5]. Due to their relative simplicity, coupled-resonator filters are a useful test case for robust optimization techniques at the system level [30][27].

## 2.3 The adjoint method

Let  $A(\mathbf{p})\mathbf{x} = \mathbf{b}$  be a linear system parameterized by a  $P$ -dimensional vector of parameters  $\mathbf{p}$ , and let  $\mathbf{f}(\mathbf{x})$  be an  $M$ -valued performance function (possibly nonlinear) of the system solution  $\mathbf{x}$ . It is often useful to know the first-order sensitivity  $\frac{d\mathbf{f}}{d\mathbf{p}}$  of the system performance to system parameters. To estimate this naively using finite difference methods, however, requires  $O(P)$  system solves, which can be prohibitively expensive if  $A$  and  $P$  are both large. The adjoint state method is a numerical technique which allows this quantity to be computed directly, using only  $O(M)$  system solves of the adjoint equation,  $A(\mathbf{p})^T \mathbf{c} = \mathbf{d}$ .

### 2.3.1 General case

The general principle of the adjoint method has been explained in a number of different ways, however for this derivation we follow the approach of [25].

$$\begin{aligned}
 A\mathbf{x} = \mathbf{b} &\implies \frac{dA}{d\mathbf{p}}\mathbf{x} + A\frac{d\mathbf{x}}{d\mathbf{p}} = 0 \implies \frac{d\mathbf{x}}{d\mathbf{p}} = -A^{-1}\frac{dA}{d\mathbf{p}}\mathbf{x} \\
 \frac{d\mathbf{f}(\mathbf{x})}{d\mathbf{p}} &= \frac{d\mathbf{f}}{d\mathbf{x}}\frac{d\mathbf{x}}{d\mathbf{p}} = -\frac{d\mathbf{f}}{d\mathbf{x}}A^{-1}\frac{dA}{d\mathbf{p}}\mathbf{x}
 \end{aligned} \tag{2.1}$$

Computing the right-hand-side of the above expression requires the application of  $A^{-1}$ , i.e. a linear system solve. If  $A^{-1}$  is applied on the right, the system solve will take  $O(P)$  time, as each dimension of  $\mathbf{p}$  yields a different column vector  $\frac{dA}{dp_i}\mathbf{x}$ . However, in the adjoint method,  $A^{-1}$  is applied on the left, through solving:

$$A^T \mathbf{c} = \frac{d\mathbf{f}}{d\mathbf{x}}$$

$$\implies \frac{d\mathbf{f}(\mathbf{x})}{d\mathbf{p}} = -\mathbf{c}^T \frac{dA}{d\mathbf{p}} \mathbf{x} \quad (2.2)$$

This will take  $O(M)$  time, as each dimension of  $\mathbf{f}$  yields a different row vector  $\frac{df_i}{d\mathbf{x}}$ . It is also important to note that a forward solve is required to obtain the base value of  $\mathbf{x}$ , and that the value of  $\mathbf{c}$  does depend on the definition of the output quantities  $\mathbf{f}$ . Once  $\mathbf{c}$  and  $\mathbf{x}$  are computed, however, they do not need to be re-computed, so no matter what the dimensionality, no additional system solves will be required to compute the linear sensitivity of  $\mathbf{f}$  to arbitrary small fluctuations in  $\mathbf{p}$ . If  $\mathbf{p}$  describes the geometry of a device's shape, then the shape gradient can be very high-dimensional, and so the adjoint method saves considerable computational effort.

### 2.3.2 Adjoint method for electromagnetism

In a linear material with no free charges, Maxwell's equations can be written as

$$\begin{bmatrix} -\frac{d}{dt}(\epsilon \cdot) & \nabla \times \\ -\nabla \times & -\frac{d}{dt}(\mu \cdot) \end{bmatrix} \begin{bmatrix} \mathbf{E} \\ \mathbf{H} \end{bmatrix} = \vec{\xi}, \quad (2.3)$$

where  $\vec{\xi}$  is a vector of electric and magnetic current sources. This is in the form  $A\mathbf{x} = \mathbf{b}$  from above, and in particular the adjoint  $A^\dagger$  is simply the time-reversal of  $A$  - suggesting that given an electromagnetic simulator capable of solving Maxwell's equations, the same simulator can be used to solve the adjoint equation and get the sensitivity of a field functional to material geometry.

A detailed derivation using the reciprocity of Green's functions can be found in [18], showing that this is indeed the case. Suppose we have a single-valued merit functional of the  $\mathbf{E}$  and  $\mathbf{H}$  fields over a 3D region  $\Omega$ , defined as

$$F[\mathbf{E}, \mathbf{H}] = \int_{\Omega} f(\mathbf{E}(\mathbf{r}), \mathbf{H}(\mathbf{r}), \mathbf{r}) d^3\mathbf{r}.$$

Suppose further that those fields are a result of fixed sources interacting with a material with real permittivity  $\epsilon(\mathbf{r})$  and permeability  $\mu = \mu_0$ . Then if  $f$  is a holomorphic function of  $\mathbf{E}$  and  $\mathbf{H}$  (i.e. the Wirtinger derivatives  $\frac{\partial f}{\partial \mathbf{E}}$  and  $\frac{\partial f}{\partial \mathbf{H}}$  of the conjugate

fields are both zero), the variation  $\delta F$  caused by an infinitesimal change  $\delta\epsilon(\mathbf{r})$  in permittivity can be calculated<sup>1</sup> as

$$\delta F = \int_{\Omega} \mathbf{E}^F(\mathbf{r}) \cdot \mathbf{E}^A(\mathbf{r}) \delta\epsilon(\mathbf{r}) d^3\mathbf{r}, \quad (2.4)$$

where  $\mathbf{E}^F$  is the electric field produced by the unperturbed forward simulation, and  $\mathbf{E}^A$  is the electric field produced by the ‘‘adjoint simulation’’. The adjoint simulation is a normal electromagnetic system solve, but with a polarization dipole source  $\mathbf{P} = \frac{\partial f}{\partial \mathbf{E}}$  and a magnetization dipole source  $\mathbf{M} = -\frac{1}{\mu_0} \frac{\partial f}{\partial \mathbf{H}}$  instead of whatever sources produced the forward simulation.

If the material fluctuation is caused by an infinitesimally shifting boundary wall  $\partial\Omega$  between two permittivities  $\epsilon_1$  and  $\epsilon_2$ , rather than an infinitesimal change in the permittivity, then the above formula is ill-defined, as  $\mathbf{E}$  (specifically,  $\mathbf{E}_{\perp}$ ) is discontinuous on a boundary. The variation is instead given by

$$\delta F = \int_{\partial\Omega} \delta y_n \left[ (\epsilon_2 - \epsilon_1) (\mathbf{E}_{\parallel}^F \cdot \mathbf{E}_{\parallel}^A) + \left( \frac{1}{\epsilon_1} - \frac{1}{\epsilon_2} \right) (\mathbf{D}_{\perp}^F \cdot \mathbf{D}_{\perp}^A) \right] dA, \quad (2.5)$$

where  $\delta y_n(\mathbf{r})$  is the displacement of the boundary wall in the normal direction at position  $\mathbf{r}$ .

### 2.3.3 Adjoint method for S-parameters

For two guided modes  $(\mathbf{E}_a, \mathbf{H}_a)$  and  $(\mathbf{E}_b, \mathbf{H}_b)$  on a cross-sectional surface  $\chi$ , an inner product can be defined as

$$\langle \mathbf{E}_a, \mathbf{H}_a | \mathbf{E}_b, \mathbf{H}_b \rangle = \frac{1}{4} \int_{\chi} \hat{n} \cdot (\mathbf{E}_a^* \times \mathbf{H}_b + \mathbf{E}_b \times \mathbf{H}_a^*) dA \quad (2.6)$$

---

<sup>1</sup>If  $f$  is not holomorphic but is always real, the variation  $\delta F$  is twice the real part of what is written. If neither condition applies, things are more complicated; [18] only derives the real case.

such that at a given frequency  $\omega$ , distinct modes are orthogonal. This inner product is used to define the mode expansion coefficient of a field  $(\mathbf{E}, \mathbf{H})$  in a mode  $(\mathbf{E}_m, \mathbf{H}_m)$ :

$$a = \frac{\langle \mathbf{E}_m, \mathbf{H}_m | \mathbf{E}, \mathbf{H} \rangle}{\langle \mathbf{E}_m, \mathbf{H}_m | \mathbf{E}_m, \mathbf{H}_m \rangle} = \frac{1}{4N_m} \int_{\chi} \hat{n} \cdot (\mathbf{E}_m^* \times \mathbf{H} + \mathbf{E} \times \mathbf{H}_m^*) dA \quad (2.7)$$

( $N_m = \langle \mathbf{E}_m, \mathbf{H}_m | \mathbf{E}_m, \mathbf{H}_m \rangle$  is the modal power. In simulations, it is easier to normalize after the fact than to ensure that all modes have unit norm.) This coefficient is an integral of a quantity holomorphic in  $(\mathbf{E}, \mathbf{H})$ , and so the adjoint method can be used to calculate its variation. The necessary adjoint sources are a surface polarization  $\mathbf{P} = \frac{\partial f}{\partial \mathbf{E}} = -\frac{\delta(\chi)}{4N_m} (\hat{n} \times \mathbf{H}_m^*)$  and a surface magnetization  $\mathbf{M} = -\frac{1}{\mu_0} \frac{\partial f}{\partial \mathbf{H}} = -\frac{\delta(\chi)}{4N_m \mu_0} (\hat{n} \times \mathbf{E}_m^*)$ .

However, by converting these sources to electric and magnetic current sources<sup>2</sup>, we can see that these adjoint sources are exactly the equivalent source conditions[21] for a backwards-propagating mode source  $\frac{j\omega}{4N_m} (\mathbf{E}_m^*, -\mathbf{H}_m^*)$ . Since an S-parameter between an input and an output mode can be defined from a forward simulation as the ratio of power-normalized mode expansion coefficients

$$S_{io} = \frac{a_o \sqrt{N_o}}{a_i \sqrt{N_i}}, \quad (2.8)$$

the variation in  $S_{io}$  can therefore be calculated using the following procedure:

1. Simulate a forward propagation of the input mode. Record  $a_i$ ,  $a_o$ ,  $N_i$ , and the electric field  $\mathbf{E}^F$ .
2. Simulate a backward propagation of the output mode. Record  $b_o$  (the backwards mode amplitude at the output port),  $N_o$ , and the electric field  $\mathbf{E}^A$ .
3. Use  $\mathbf{E}^F$ ,  $\mathbf{E}^A$ , the expected material perturbation, and equation (2.4) or (2.5) as appropriate to calculate  $\delta F$ .
4. Normalize:

$$\delta S_{io} = \frac{j\omega}{4a_i b_o \sqrt{N_i N_o}} \delta F \quad (2.9)$$

---

<sup>2</sup> $\mathbf{J} = d\mathbf{P}/dt = j\omega\mathbf{P}$ ,  $\mathbf{K} = d\mathbf{M}/dt = j\omega\mathbf{M}$

Note that equation (2.9) is symmetric with respect to interchanging the forward and adjoint simulations - this makes sense, as  $S_{io} = S_{oi}$ . Note also that the perturbation information was only used in step (3), after the simulations were run - essentially, the adjoint method calculates the entire shape gradient at once. This is what makes it so useful for design optimization.

## 2.4 Optimization under uncertainty

In engineering, there is often a need to optimize some kind of parameterized design  $\mathbf{p}$  in a way which performs well under a variety of uncontrolled scenarios parameterized by  $\nu$ . If performance is described by a function  $F(\mathbf{p}, \nu)$ , then two approaches for formulating an optimization problem are *worst-case optimization*, i.e. seeking  $\min_{\mathbf{p}} \max_{\nu \in \mathcal{V}} F(\mathbf{p}, \nu)$  for some prescribed set of acceptable scenarios  $\mathcal{V}$ , or *average-case optimization*, i.e. seeking  $\min_{\mathbf{p}} \int_{\mathcal{V}} F(\mathbf{p}, \nu) \rho(\nu) d\nu$  for some prescribed probability distribution  $\rho(\nu)$ . Worst-case optimization is generally, but not always, referred to as *robust* optimization, and average-case *stochastic* optimization.

It is often the case that compared to designs which perform well in the best-case, designs which perform well in the worst-case are also designs that perform well in the average-case, and vice versa. This is because robustness is often a matter of simply not having any sharply sensitive components, and correlates with other traits such as broadband performance [12]. Consequently, a third approach for resilient optimization is to address resilience indirectly, and do a deterministic optimization with a regularization penalty for the kinds of design features that tend to be sensitive to variation (e.g. resonances, tight bends, or small gaps). Given that  $\rho(\nu)$  and  $\mathcal{V}$  are often arbitrary or estimated anyway, which approach is chosen is often a matter of computational convenience.

If  $F$  is convex or closed-form, there exist many well-defined algorithms for solving the worst-case optimization problem. Similarly, if  $F$  is cheap to calculate, or can only be calculated with partial control over the arguments (i.e. if real-world experiments are being conducted), different algorithms are available. This thesis is focused on the



type of problem where  $F$  can be simulated for any  $\mathbf{p}$  and  $\nu$ , and where an algorithm can re-evaluate after each simulation because the limiting factor is the simulation time in  $F$ . (To use the classification in [10], where the noise is exogenous, constructive, and generative.)

### 2.4.1 Model-based approaches

When the dimensionality of  $\mathbf{p}$  and  $\nu$  are small, a common strategy is to sample a variety of points  $(\mathbf{p}, \nu)$  in the design space, construct an approximate response surface model for  $F$  which is easier to compute, and optimize over the resulting surrogate model to find a new design point for  $\mathbf{p}$ . This is similar to the task of *uncertainty quantification* - of computing properties (i.e. moments) of the probability distribution of  $F(\nu)$  given  $\rho(\nu)$  and only a few samples of  $F$  - and so the modeling and sampling tools used are largely the same.

Two of the most popular model types used are kriging (also known as Gaussian process emulation) models and polynomial chaos expansions. In kriging, the function  $F$  is assumed to be an instance of a random Gaussian process with some variance and covariance structure, and the data points are assumed to be perfect observations which can be used to condition the distributions of the unsampled points (and to fit higher-level parameters such as the correlation lengths). In polynomial chaos expansions, the function  $F$  is assumed to be a sum of relatively low-order polynomial basis terms, and the data points are used to fit the basis coefficients (and in sparse polynomial chaos expansions, to infer which of the basis elements are to be used at all). Advantages of kriging include that kriging models are perfect interpolations, and that every prediction comes with an uncertainty attached - however, one key advantage of polynomial chaos is that polynomial models are very cheap to optimize over and compute moments of. Neither method has been shown to be intrinsically superior, even in head-to-head comparisons [23], and both have been used in design optimization under uncertainty [4][30][27].

Adaptively picking the most efficient sample points is an active field of research. In polynomial chaos expansions, the basis elements form a numerical matrix over

the selected data points, and so alphabetic optimality criteria such as D-optimality or S-optimality can be used to encourage the fitting process to be an orthonormal projection [8]. In kriging approaches, the uncertainties can be used to create an expected improvement criterion, which weights the amount of information provided by sampling a given point with the anticipated usefulness of that information to the final optimization or moment calculation [27]. Another important consideration in both cases is longer-term planning - greedily selecting the highest-payoff point at each iteration can be less efficient than selecting a group of points all at once, if k-means (as in [7]) or some other method is used to ensure the group is spread out.

### 2.4.2 Model-free approaches

When the dimensionality of a space is large, the number of parameters required to define a surrogate model can skyrocket. Without large numbers of data points to avoid overfitting, models have to become very simple, and consequently lose predictive power. This manifests differently in  $\mathbf{p}$ -space and  $\nu$ -space; in  $\mathbf{p}$ -space, if a gradient is available (i.e. through the adjoint method), optimization can proceed via gradient descent provided there is some kind of aggregation in  $\nu$ -space.

This aggregation can come in a few different forms, most of them defined in the context of stochastic optimization problems [10]. In sample average approximation (SAA), a fixed set of  $\nu$ 's is determined a priori, and any function evaluations or gradients are averaged over all  $\nu$ 's in the set before being fed to a deterministic optimization algorithm. This has the advantage of stability, but does not converge. In stochastic gradient descent (SGD) (popular in machine learning, where different values of  $\nu$  correspond to different training data), the set of  $\nu$ 's varies with each iteration, and the expectation is that over many iterations the gradients will average out and converge to the right answer. In both cases, picking points in  $\nu$ -space without a model can proceed either randomly, e.g. with Latin hypercube sampling, or pseudo-randomly, e.g. with Sobol sequences.

# Chapter 3

## Testing the adjoint method in the presence of roughness

The adjoint method is powerful and has been validated many times in the literature, but it has limitations. Chief among them is the fact that it is a first-order local technique, and so using it to approximate the response surface to a design only works on a scale limited by the magnitude of higher-order effects. Additionally, it requires accurate values of the electromagnetic fields to be calculated exactly at material boundaries, which is exactly where the fields are the most volatile and simulators are most likely to stumble. In the case of line-edge roughness, both of these issues are exacerbated; simulations and interpolations are more difficult because the boundaries are jagged, and scattering losses are likely to be dominated by second-order effects, as what matters is more often the amount of roughness than whether a particular wiggle goes out or in. In optimization, this is particularly pernicious, as “optimized” designs are particularly likely to have no first-order dependence on roughness, even as their second-order sensitivities are worse than that of other designs. Therefore, in each application that relies on the adjoint method, and especially in this one, it is important to first determine the extent of its applicability in that particular context and implementation.

This chapter describes experiments carried out to try to determine those extents, guided around the following questions:

1. How accurately does the adjoint method measure the first-order dependence of transmission quantities?
2. How large can a design change or perturbation be before its effect on performance is no longer first-order? How does this vary by type of perturbation?
3. How do the above limitations vary by choice of quantity or device?

### 3.1 Notation and implementation details

I carried out electromagnetic simulations using Lumerical’s MODE Solutions software, using the varFDTD [15] 2.5D simulator to reduce computation time. varFDTD is a time-domain simulation technique, which calculates frequency-domain field information and S-parameters from the result of a single broadband pulse.

Device geometry in MODE Solutions is defined using polygons of points  $\{\mathbf{r}_i\}$  in 2D space. I parameterized perturbations to device geometry using a vector of real numbers  $\mathbf{p} = \{p_i\}$ , where  $p_i$  is the displacement of point  $i$  from its base position  $\mathbf{r}_i$  in the direction normal to the material boundary<sup>1</sup>. Along the z-axis, all devices were 220nm thick Si, cladded in SiO<sub>2</sub> (using the Palik materials set), with input and output waveguides 500nm wide.

All device input and output ports were associated with a mode source and a mode monitor. To avoid near-field artifacts, the mode sources were spatially separated from the mode monitors (the phase effect of this propagation distance is taken into account through  $a_i$  and  $b_o$ ). To avoid interpolation phase artifacts in calculating S-parameters, all mode monitors were placed at FDTD mesh boundaries.

However, interpolation artifacts are much harder to avoid when estimating the field at the material boundary. varFDTD monitors only report fields and refractive indices (themselves the products of a proprietary conformal meshing algorithm, see Figure 3-1) on a static rectangular grid, which does not line up with the boundary. Rather than try to reverse-engineer the meshing algorithm, I calculated D-fields in each mesh cell

---

<sup>1</sup>i.e. perpendicular to  $\mathbf{r}_{i+1} - \mathbf{r}_{i-1}$

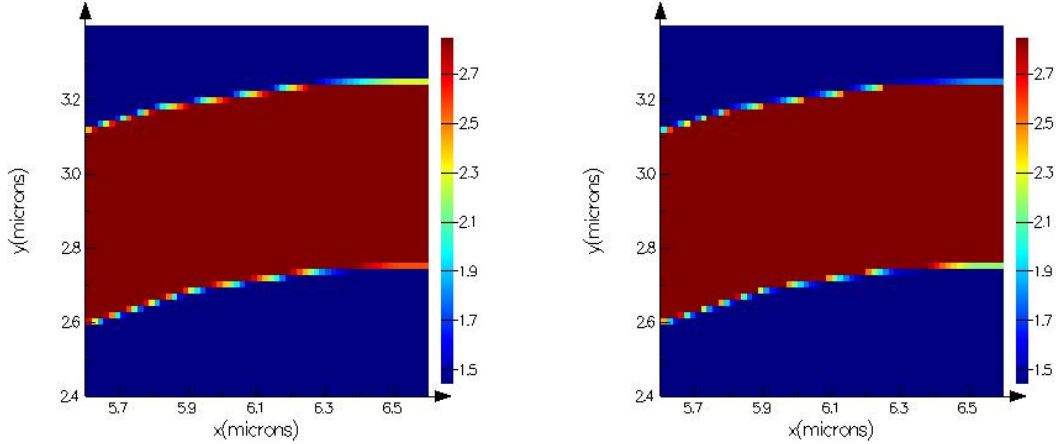


Figure 3-1: The X (left) and Y (right) refractive indices reported by Lumerical’s conformal meshing algorithm on a material boundary.

by component-wise multiplying the reported permittivities  $\epsilon_x$  and  $\epsilon_y$  by the reported E-field components, then use the material boundary direction to calculate  $\mathbf{E}_{\parallel}$  and  $\mathbf{D}_{\perp}$ , then interpolated the rotated fields (which should each be less discontinuous than the component-wise values) to get a value at each specific boundary vertex. I rejected an alternate approach which involved fitting a locally linear field pattern in the 3-mesh-cell vicinity of the boundary instead of interpolating between the single-cell nearest mesh points, as fitting produced smoother but less accurate results.

### 3.2 Effect of roughness on a smooth Y-splitter

A series of 800 simulations were run on roughened versions of a Y-branch splitter, following the design in [31]. For each of the four correlation lengths [20nm, 50nm, 100nm, 400nm], ten roughness scenarios were generated using a Gaussian correlation function and an amplitude of 20nm. These scenarios were then scaled by a series of 20 values between -1 and 1, in order to create a tightly sampled trajectory for finite difference analysis. Roughness was applied to the inner and outer edges of the splitter, starting 1  $\mu\text{m}$  before the cavity begins, and ending 1.5  $\mu\text{m}$  after the arms straighten out. In addition, forward and adjoint simulations were run on the base design, allowing adjoint predictions of the effects of each scenario on the Y-splitter.

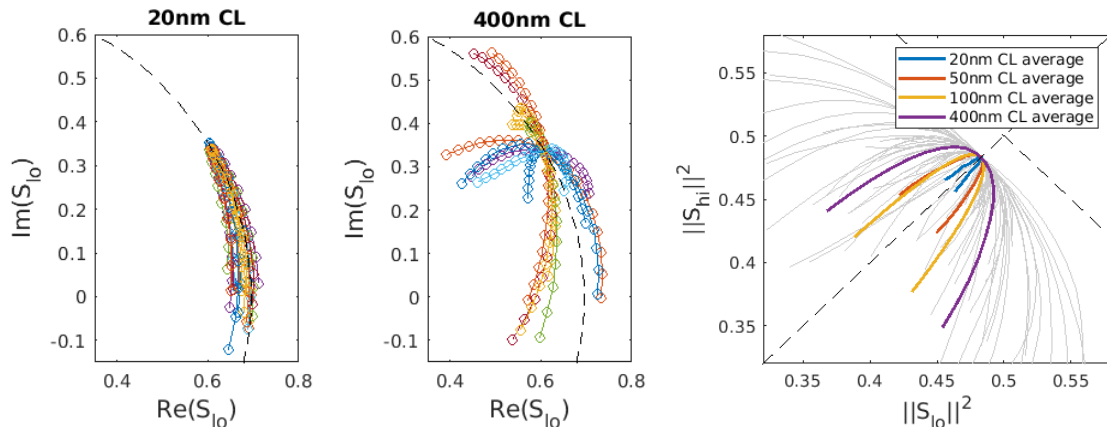


Figure 3-2: Left: lower port S-parameters for a Y-splitter under 10 roughness families with correlation lengths of 20nm and 400nm, with constant-magnitude reference lines. Right: transmitted power through both ports for all 800 simulations, with averages for each correlation length of roughness, and zero-loss and zero-imbalance CL reference lines.

All simulations were done at a free-space wavelength of 1550nm, using the lowest-order TE modes ( $2\pi/\beta \approx 622\text{nm}$ ) for both input and analysis. The geometry of the Y-splitter was discretized using points at most 20nm apart (closer, in the bends of the cavity), and the simulation mesh cells were 20nm by 20nm.

Results of the simulations are shown in Figure 3-2, illustrating two basic facts - first, that the effects of fine-grained roughness on phase are much stronger than on magnitude, and second, that in both magnitude and phase, for amplitudes up to 20nm, the effects of any kind of roughness are smooth and at worst second-order. A third observation is that while the linear effect of a fluctuation can point in any direction, the quadratic effect seems to always point in the direction of decreased phase and magnitude. This suggests that the quadratic effect is dominated by scattering loss gradually sapping the energy of the optical beam, with associated phase loss, rather than any nonlinear dependencies in sensitive sections of the device. Lastly, although roughness with a correlation length of 400nm is not physically realistic, it elicited stronger linear and quadratic reactions (at least in magnitude) than finer perturbations, which is as expected as 400nm is closer to the modal wavelength.

Qualitatively, the predictions of the adjoint method can be seen in Figure 3-3. The strong phase sensitivity can easily be seen, as can the way that asymmetric

	20nm CL	50nm CL	100nm CL	400nm CL
$S_{lo}$	3nm	8nm	10.5nm	>20nm
$\ S_{lo}\ ^2$	>20nm	14.5nm	14nm	>20nm
$\ S_{lo}\ ^2 + \ S_{hi}\ ^2$	3nm	2.5nm	1.5nm	2nm
$\ S_{lo}\ ^2 - \ S_{hi}\ ^2$	N/A	N/A	N/A	N/A

Table 3.1: Linear regime crossover points for various transmission quantities under various correlation lengths of roughness. N/A means the linear and quadratic prediction errors showed no sign of approaching crossing over.

fluctuations just before the resonant cavity can strongly direct optical energy towards one port or another. Notably, however, there is nowhere with a strong linear effect on overall loss, as loss is already close to a minimum in the base design, so it increases in both directions of any given perturbation.

A more quantitative way of assessing the success of the adjoint method is through prediction error. If the total power transmitted through both ports is plotted across a single family of scenarios, then the resulting curve can be fit linearly or quadratically, either using a best-fit linear slope or the adjoint-predicted linear slope, and each model will have its own mean square error, plotted in Figure 3-4. Since the adjoint prediction errors are nearly identical to the best-fit linear prediction errors, it is evident that the adjoint method is correctly calculating the linear slopes. The crossover points between the best-fit linear model and the best-fit quadratic model confirm and explain the results in [32]; for fluctuations beyond 5nm in amplitude, while the linear regime remains effective at predicting power imbalance, no adjoint method centered around a smooth design can predict overall insertion loss.

The full set of crossover points are tabulated in Table 3.1. For all correlation lengths, imbalance was overwhelmingly linear and loss was overwhelmingly quadratic. Interestingly, however, when predicting the complex S-parameters (i.e. predicting their phases), the fine-grained fluctuations caused generally quadratic phase fluctuations, whereas the coarse fluctuations were more likely to have an aggregate linear effect.

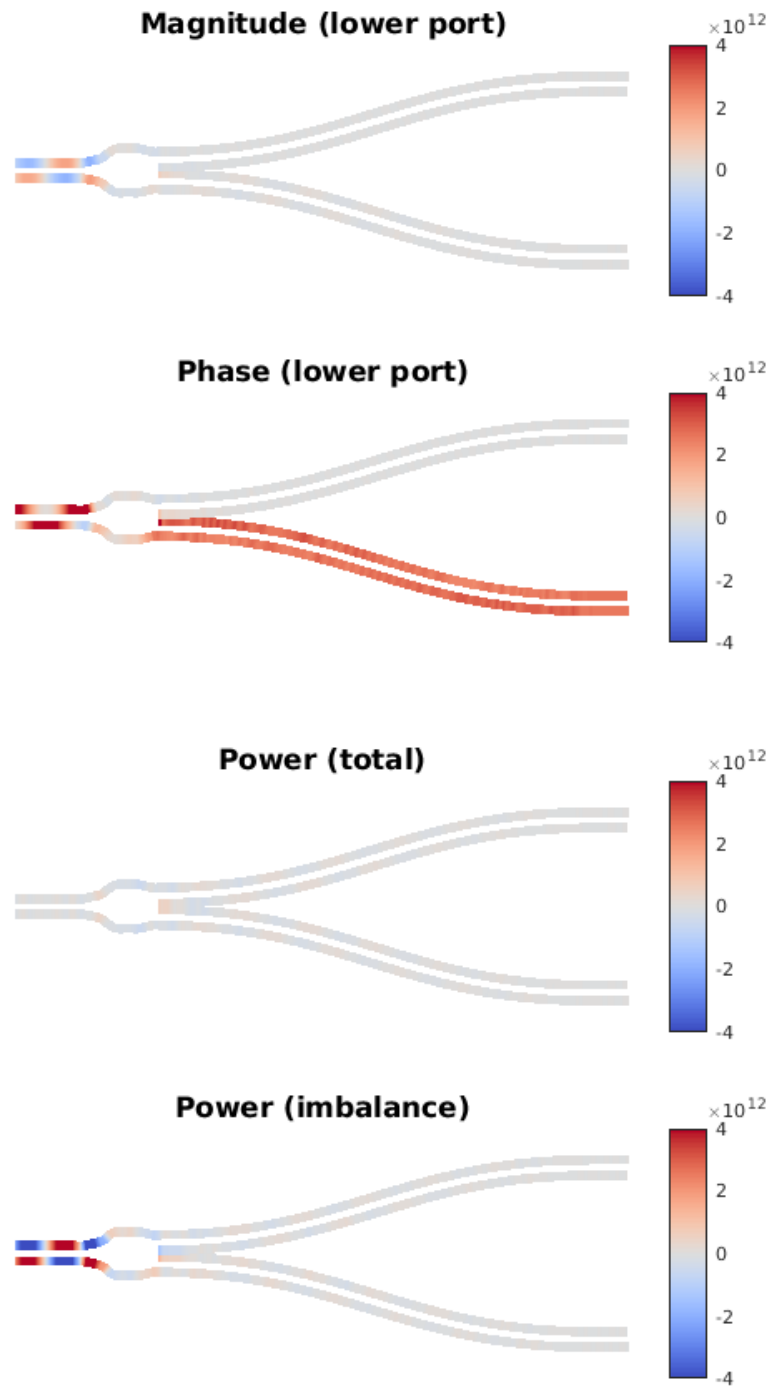


Figure 3-3: Heatmaps of the predicted sensitivity of various Y-splitter quantities to local outward wall motion, in  $\text{m}^{-2}$ .



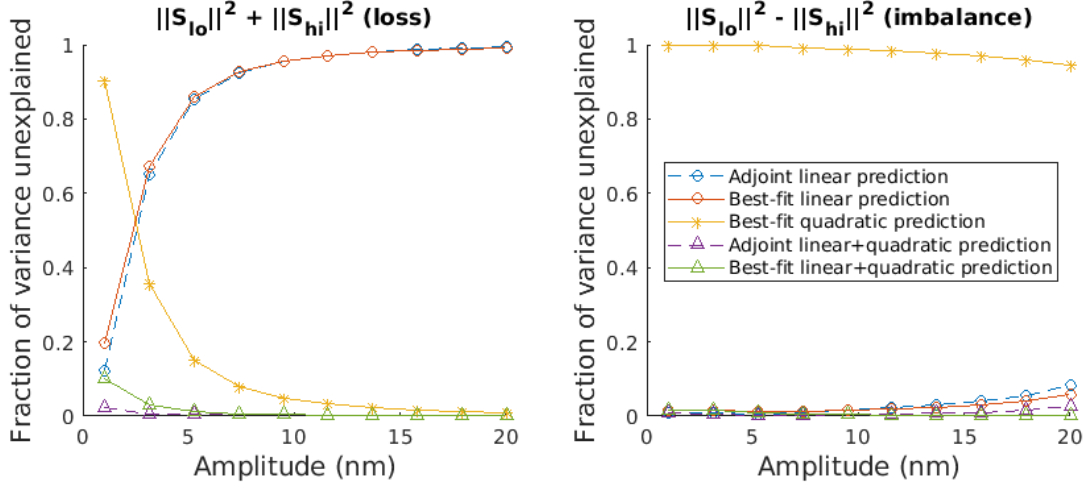


Figure 3-4: Normalized prediction errors, averaged across profiles, for Y-splitter total power loss and imbalance under 50nm-correlated roughness.

### 3.3 Effect of roughness on a smooth directional coupler

A series of 600 simulations were run on a directional coupler design, with a coupling length of  $3 \mu\text{m}$  and a coupling gap of 150nm, as shown in figure 3-5. The experimental conditions were identical to the Y-splitter experiments in the previous section (ten families of twenty scenarios each, ranging from -20nm to 20nm in amplitude, per correlation length) but with only three correlation lengths (20nm, 50nm, 100nm) and with data collected at nine frequency points, running from 1500nm to 1600nm. The unroughened coupling ratio is not constant with frequency, but varies approximately linearly between 15/82 at 1500nm and 24/73 at 1600nm. Roughness was applied to all four edges - both inside and outside, both top and bottom waveguide - but only in the horizontal coupling region in the center. The geometry of the coupler was discretized using points 30nm apart, and the simulation mesh cells were 20nm by 20nm, except in the coupling gap, where they were 10nm wide in the  $y$  direction (perpendicular to signal flow).

Compared to the Y-branch splitter, the behavior of the directional coupler was much less sensitive. At any given frequency, the transmitted power at either port never varied by more than 0.069, compared to 0.189 for the Y-splitter at the same

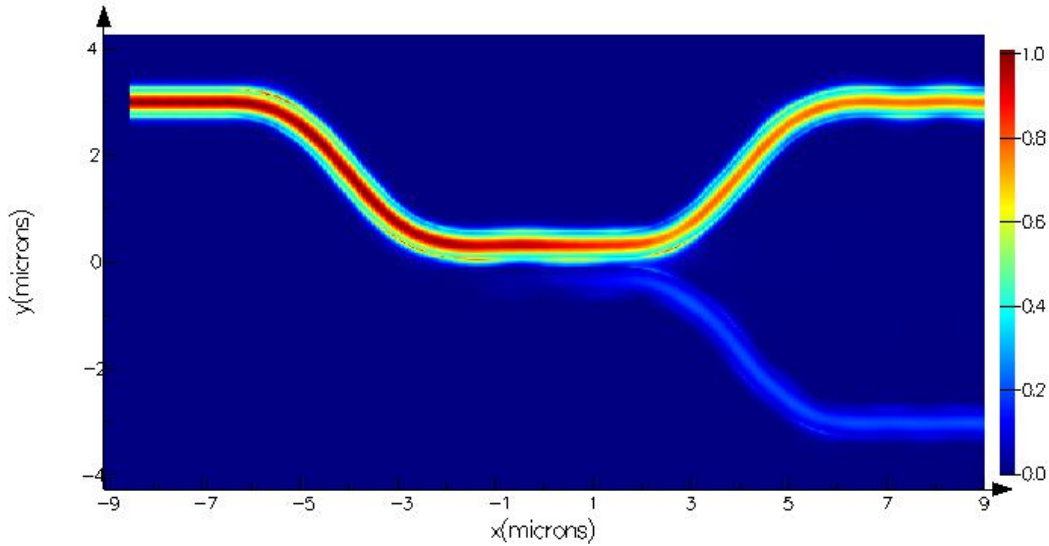


Figure 3-5: Electric field intensity in a directional coupler at 1550nm.

correlation lengths. Its response to roughness was also much more dominated by quadratic effects, as illustrated in figure 3-6. Furthermore, instead of causing loss, at finer correlation lengths, the quadratic effects increased the coupling coefficient. (At longer correlation lengths, they caused loss and an increased coupling coefficient.) A possible explanation is that while the quadratic effects are likely the effects of randomly scattering into other modes, the most nearby mode to the lowest TE mode in the upper waveguide is simply the lowest TE mode in the lower waveguide, so the majority of “random scattering” manifests as coupling. This could potentially suggest a robust design strategy for directional couplers which is simply to design for a smaller coupling coefficient than desired, on the assumption that most roughness instances have approximately the same effect.

Despite using the same mesh grid override, there was a discontinuity between the S-parameters from the base design simulation and the S-parameters of the roughened simulations, likely due to the perfectly  $x$ -aligned surfaces in the base design. Since the base simulation was used both for centering the linear approximations and for determining the shape derivatives, it is not surprising that the prediction errors shown in Figure 3-7 are large, especially for small perturbations where the overall variance

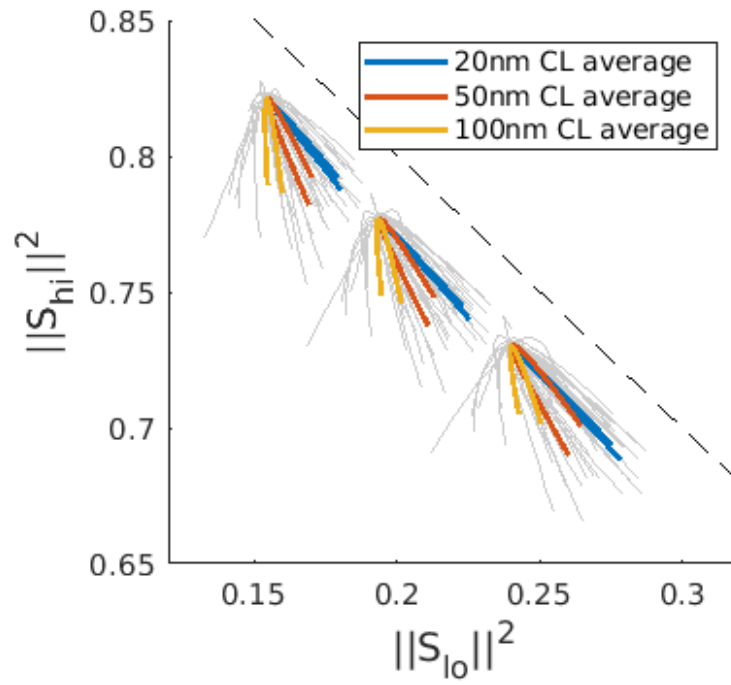


Figure 3-6: Transmitted power at 1500nm, 1550nm, and 1600nm through ports of a directional coupler for all 600 simulations, with averages for each correlation length of roughness, and a zero-loss reference line.

is small. The crossover points between best-fit linear prediction error and best-fit quadratic prediction error vary unsystematically with correlation length, but overall, the linear regime for predicting insertion loss extends up to roughness amplitudes of 10-15nm, but only up to 5-10nm for predicting coupling coefficients.

### 3.4 Effect of roughness on a smooth ring resonator

The same 600-simulation experiment was run on an  $8.7\mu\text{m}$ -diameter circular ring resonator, coupled between two waveguides with a 150nm coupling gap on each side. Unperturbed, the ring has a narrow resonance near 1550nm and a free spectral range of approximately 30nm. Roughness (again of three correlation lengths, 20nm, 50nm, and 100nm) was applied to the inside and outside edges of the ring, except in the close proximity of the coupling waveguides. The intent was to isolate the effects of roughness on the ring's resonances, not the effects of changing the coupling coefficients on the overall spectrum. Simulations were run with 51 frequency points between 1540nm and 1560nm. The geometry of the resonator was discretized using points 20nm apart, and the simulation mesh cells were 20nm by 20nm.

The effects of roughness are shown in Figure 3-8; while the parameters of the resonance were affected by roughness, they varied smoothly, and the overall function of the device was never impaired. When resonant frequencies were fit to each spectrum, the overall trend was strongly quadratic, relatively independent of correlation length, and always towards shorter wavelengths. This is consistent with the strongly negative phase dependence from the Y-splitter results; if the optical path length of a waveguide is decreased, the wavelength of light that is a specific fraction of the optical path length will also decrease.

However, the adjoint method operates on each frequency separately, and the change in transmitted power at a single specific frequency is highly nonlinear as a resonance travels over it. The adjoint predictions for the ring resonator were unstable and incorrect by orders of magnitude. Conceptually, the adjoint method treats the effect of a change in electric permittivity at a location as if it were equivalent to

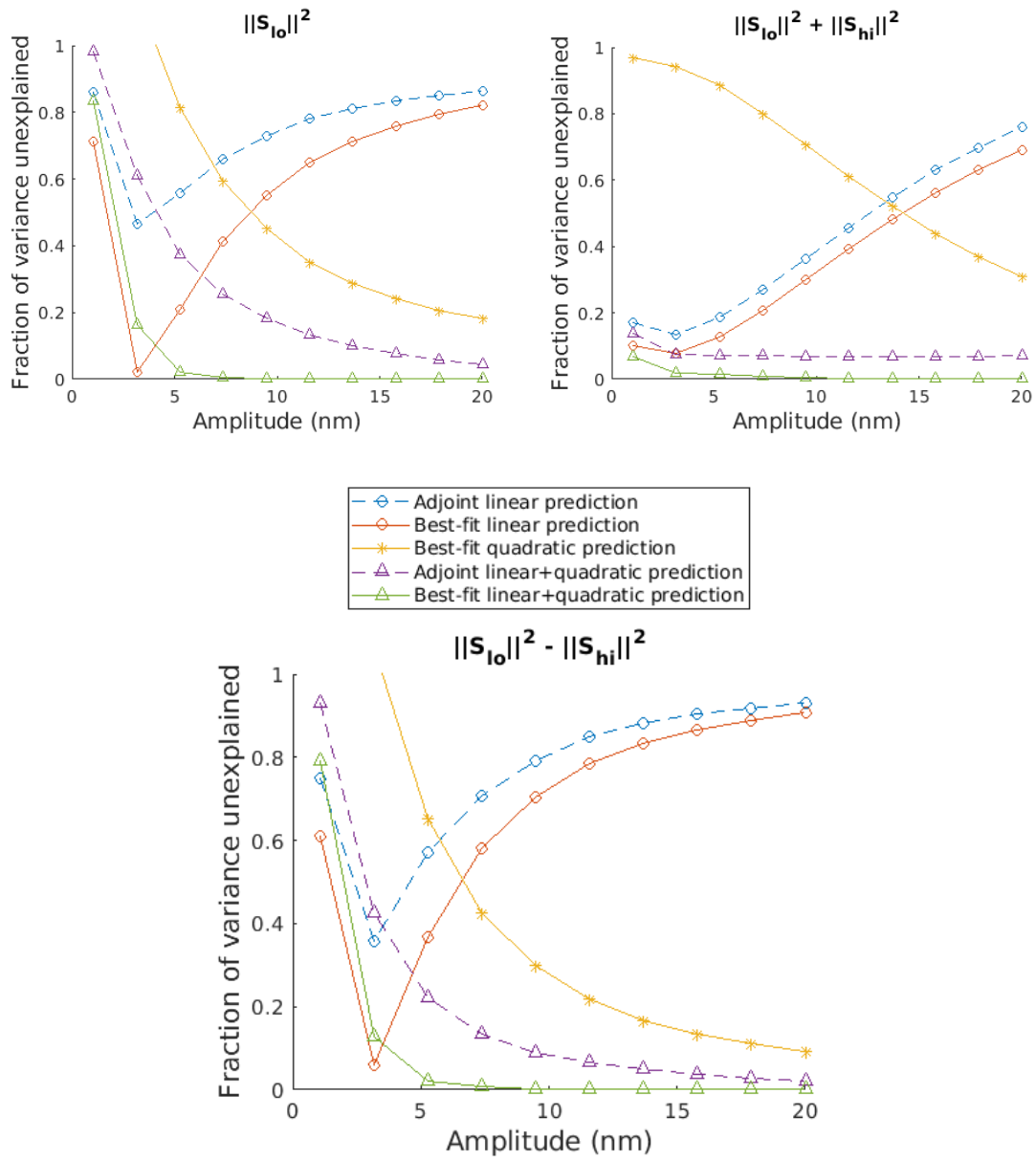


Figure 3-7: Normalized prediction errors, averaged across frequencies and profiles, for directional coupler transmission quantities under 50nm-correlated roughness.

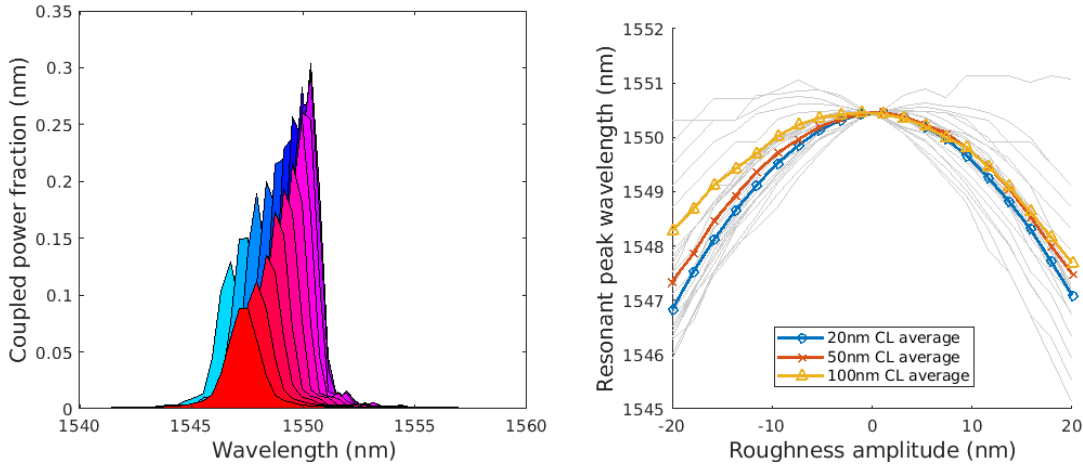


Figure 3-8: Left: the ring resonator spectra produced by a single 50nm-correlation family of roughness profiles with amplitudes from -20nm to 20nm. Right: All 600 extracted resonant frequencies, with the average for each correlation length highlighted.

the introduction of a dipole source at that location, an approach which assumes the location is only traversed once by an optical pulse - something that is deliberately incorrect in a feedback system such as a ring resonator. It is possible that the adjoint method could be made useful for resonant systems if it could analyze them in open-loop and apply feedback equations after-the-fact, but this would require substantial changes to the simulation framework.

### 3.5 Effect of smooth design changes on a roughened Y-splitter

One concern about the adjoint method is that it relies on field and tangent direction information at the material boundary, both of which are erratic in the presence of roughness. To test this, I generated five different roughness scenarios (10nm amplitude, 40nm correlation length), and ran forward and adjoint simulations under each of the scenarios. Then, I generated five different “smooth design changes” by simulating roughness with a 10nm amplitude and a 400nm correlation length, but restricted to the input waveguide and the cavity (i.e. no changes on the arms). Scaling the amplitude of the five design changes to twenty points each between -10nm and

10nm, and applying each scaled design change to each roughness scenario, produced an experiment with 500 simulations.

Figure 3-9 shows in solid lines the power imbalances and losses produced by the design changes through one of the roughness scenarios, and in dashed lines, for comparison, the same effects produced by 400nm-correlated roughness (truncated to 10nm amplitude) on a smooth Y-splitter from the experiments in Section 3.2. While the linear effects on imbalance are largely the same, the quadratic effects on loss are substantially weaker, likely because the design changes were restricted from the Y-splitter arms. In addition, the linear effects on loss are much stronger, likely because even before design changes, there was already a loss mechanism - the roughness - which could be enhanced or diminished by the design change “agreeing” or “disagreeing” with it. (This is consistent with the Payne-Lacey model, which postulates that scattering loss is primarily due to fluctuations in a single narrow spatial frequency band. If there were multiple ways to cause scattering loss, then we would expect a stronger quadratic effect, as fine-grained roughness and smooth design changes would not constructively or destructively interfere.) These same observations were seen in the other four roughness scenarios - in fact, regardless of the roughness scenario, the same smooth design change tended to produce the same S-parameter changes.

Figure 3-10 shows the prediction error for insertion loss from linear, quadratic, and adjoint models, with errors averaged over the 25 trajectories. The adjoint method closely tracks the best-fit linear model, and because of the stronger linear effects, the crossover point is beyond the 10nm maximum amplitude. This means that on a roughened Y-splitter, the adjoint method is actually a *more* reliable tool than it is on a smooth Y-splitter - it can be trusted even on objective functions which include loss, so long as design changes stay within 10nm in amplitude.

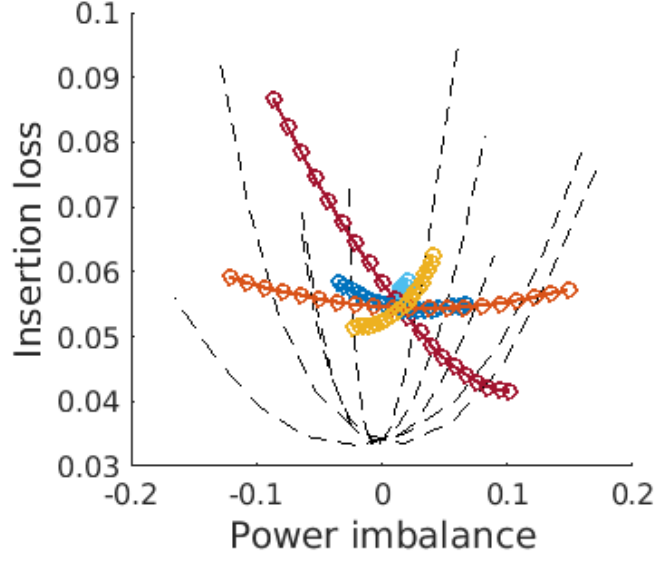


Figure 3-9: Trajectories of smooth design changes on a roughened Y-splitter, contrasted with the trajectories of the effects of smooth perturbations on a smooth Y-splitter.

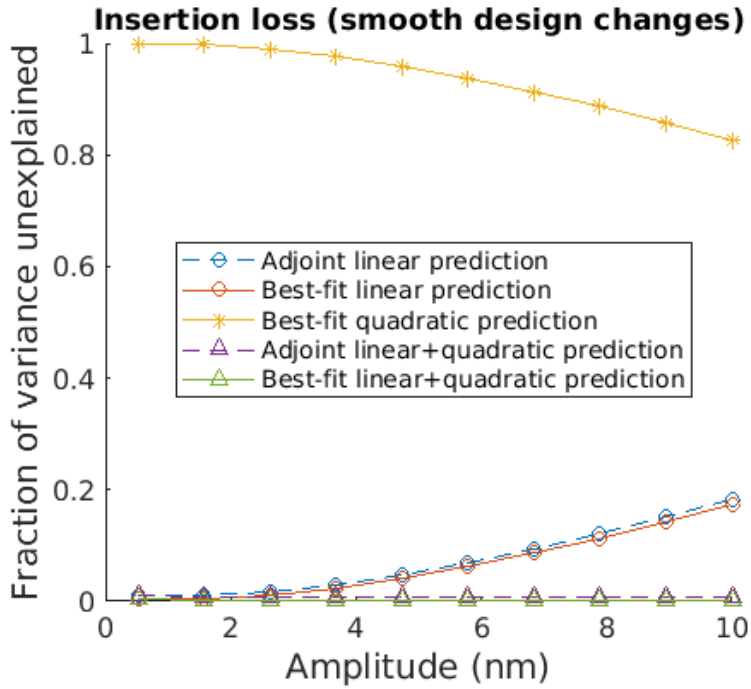


Figure 3-10: Prediction error from linear and quadratic models on the effect of smooth design changes on the total loss from a roughened Y-splitter.



# Chapter 4

## Robust shape optimization of Y-splitter against line edge roughness

### 4.1 Algorithm

The approach to robust optimization taken here is largely inspired by the work in [20] and its successor [22] on the robust taper-shape problem for the transition between two different types of optical waveguide. In those works, as in this one, the overall strategy goes as follows:

1. Have a current design  $\mathbf{p}^{(i)}$ .
2. Formulate a finite set  $\mathcal{U}^{(i)} \subset \mathcal{V}$  of roughness scenarios  $\nu_j$  that will likely lead to high penalty function values  $F(\mathbf{p}^{(i)}, \nu_j)$ .
3. Compute or estimate the gradients  $\partial F(\mathbf{p}, \nu_j)/\partial \mathbf{p}$  at  $(\mathbf{p}^{(i)}, \nu_j)$  for each  $\nu_j$ .
4. Solve a multi-scenario optimization problem<sup>1</sup> to find the step  $\delta \mathbf{p}$  (within trust-

---

<sup>1</sup>This is a convex linear problem and can be solved with an out-of-the-box linear programming toolkit.

region step size constraints) minimizing the estimated maximum error function:

$$\min_{\delta \mathbf{p}} \max_j \left[ F(\mathbf{p}^{(i)}, \nu_i) + \frac{\partial F(\mathbf{p}, \nu_i)}{\partial \mathbf{p}} \cdot \delta \mathbf{p} \right] \quad (4.1)$$

5. Update the design  $\mathbf{p}^{(i+1)} = \mathbf{p}^{(i)} + \delta \mathbf{p}$ .

6. Repeat until a stopping criterion is satisfied.

But there are key differences in the specifics, particularly in the selection of the scenario set  $\mathcal{U}^{(i)}$ , which in [20] is done by using the adjoint method to pick worst-case scenarios limited by an  $\ell_\infty$  norm. Because responses to roughness are so often dominated by quadratic effects, using the adjoint method is unlikely to correctly guess the worst-case scenarios. Further, because responses to roughness are so dependent on the correlation length of the roughness, using purely the adjoint method would produce highly unlikely scenarios with qualitatively different responses - and potentially incentivize qualitatively different design decisions - than the vast majority of actual roughness instances. (This is particularly true because of the high dimensionality of the roughness space.) This is a tradeoff inherent to any worst-case optimization; if the set of allowable scenarios  $\mathcal{V}$  is large, there is a risk of over-optimizing for unlikely edge cases at the expense of overall group performance, but if it is small, there is a risk of over-optimizing for high-performance scenarios without truly robustifying the design.

The set selection procedure used here is a hybrid approach, focused on robustifying to 2 or 3 standard deviations rather than to 10 or 15 like a true worst-case approach might be. At each iteration of the algorithm,  $N_{c_1}$  candidate scenarios are generated according to the usual roughness probability distribution. The adjoint sensitivity at  $(\mathbf{p}^{(i)}, 0)$  is used to predict and select the  $N_{c_2} < N_{c_1}$  of those candidates likely to be the most disruptive. Then, of the scenarios in  $\mathcal{U}^{(i-1)}$ , one of them (a rotating slot, so that each scenario is only used for  $|\mathcal{U}|$  iterations) is replaced with the remaining candidate most different from the members of  $\mathcal{U}^{(i-1)}$  (as measured by maximum dot product). The intent is to ensure that in every iteration, the scenario set is sufficiently diverse

to be useful for worst-case aggregation (i.e. they will be affected in different ways by the same design change), but also sufficiently damaging to ensure that design updates are in the direction of robustification.

## 4.2 Implementation on Y-splitter

The robust algorithm described above was run twice on the Y-splitter design problem, with identical implementation details. The first run had as its initial design the nominal design in [31], which is already a high-performing Y-splitter design. The second run started from a crude triangular cavity design, in the hopes of avoiding getting caught in a local minimum. The penalty function used was quadratic for both loss and imbalance:

$$F = (1 - \|S_{lo}\|^2 - \|S_{hi}\|^2)^2 + w(\|S_{lo}\|^2 - \|S_{hi}\|^2)^2 \quad (4.2)$$

with an imbalance weight of  $w = 0.3$ , to compensate for the fact that imbalances are typically much larger than losses. Roughness scenarios were generated with an amplitude of 10nm, and a correlation length of 50nm. Three roughness scenarios were considered at each iteration, with replacements chosen using  $N_{c_1} = 20$ ,  $N_{c_2} = 5$ .

Design points were separated by less than 200nm, with the geometry spline-interpolated by a factor of 10 between design points. During each iteration, design points were only allowed to move by at most 10nm, and the difference between neighboring design points' motion could not exceed 1.6nm, to encourage smoothness. Most importantly, designs were required to be symmetric between the upper and lower edges, to avoid overreacting in the event that all roughness scenarios were imbalanced in the same direction. Lastly, although roughness was still applied to the arms leading away from the cavity, they were not part of the design optimization problem, which only dealt with the cavity and the waveguide at its input.

The algorithm ran for 30 iterations in each run, producing a design evolution shown in Figure 4-1. The step size and smoothness constraints were highly active,

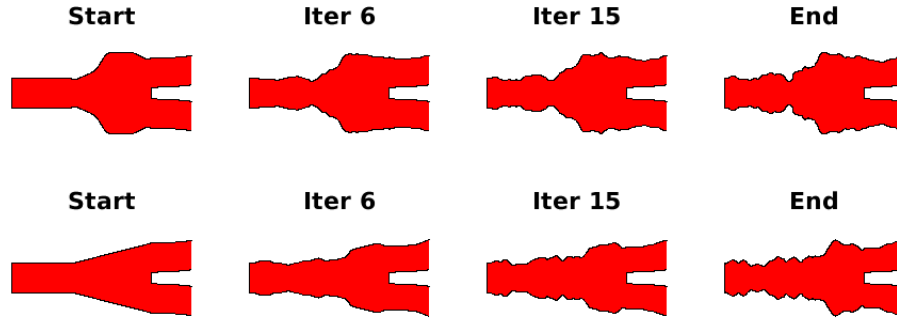


Figure 4-1: Snapshots of the designs produced by the algorithm as it progressed, for both its nominal-seeded run (top) and its triangle-seeded run (bottom).

and in each iteration, the majority of design points tended to move the maximum amount allowed, sometimes only to move the maximum amount allowed in the opposite direction the following iteration. The resulting designs consequently do not have smooth, pleasant-looking shapes, and some of their features are likely superfluous and hard to manufacture.

The starting and ending designs, as well as the designs produced halfway through after 15 iterations, were evaluated with 100 random roughness scenarios (also with an amplitude of 10nm and a correlation length of 50nm). Their performance is tabulated in Table 4.1 and, in the nominal-seeded case, displayed in Figure 4-2, with ‘T1’, ‘T15’, and ‘T31’ referring to the starting, intermediate, and final designs of the triangle-seeded run, and ‘N1’, ‘N15’, and ‘N31’ referring to the nominal-seeded run. In both runs, the intermediate and final designs outperformed the starting designs on every metric. Even in the best-case, no-roughness scenario (including when measured broadband), the optimized designs had less loss than the nominal design, demonstrating that robustification need not necessarily come at the expense of best-case performance. (On the flip side, the results cannot be taken as a direct comparison between a “deterministic optimum” and a “robust optimum”, because the nominal design was not a deterministic optimum.)

However, the intermediate designs slightly outperformed the final designs on every metric except RMS imbalance, including the penalty function. That is, during the

Design	Mean loss	Best loss	RMS imbalance	Mean penalty	90th %-ile penalty
T1	0.2756	0.2138	0.0467	0.0774	0.0972
T15	0.0377	0.0127	0.0431	0.0020	0.0031
T31	0.0401	0.0147	0.0409	0.0022	0.0036
N1	0.0580	0.0337	0.0453	0.0040	0.0058
N15	0.0306	0.0056	0.0395	0.0014	0.0023
N31	0.0353	0.0108	0.0349	0.0017	0.0028

Table 4.1: Performance metrics of the starting, intermediate, and final Y-splitter designs under 100 Monte Carlo roughenings.

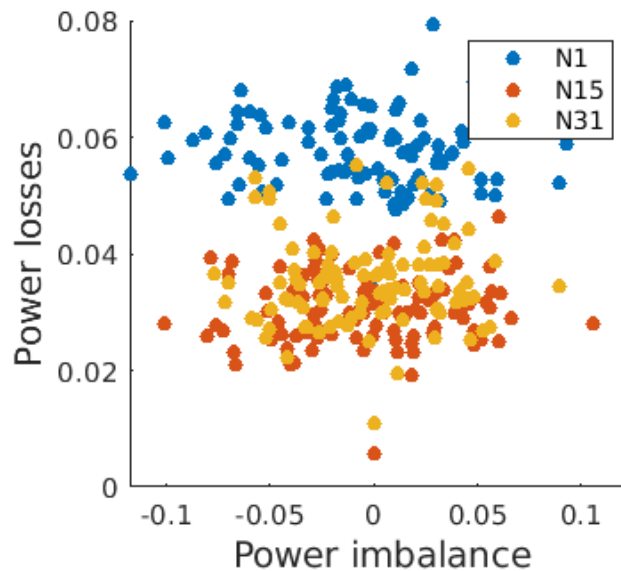


Figure 4-2: Performance of the starting, intermediate, and final Y-splitter designs from the nominal-seeded optimization, under 100 Monte Carlo roughenings.

second half of its progress, the optimization was counter-productive. This is consistent with the increasingly erratic boundary shapes seen in the later designs and suggests that overfitting occurred.

### 4.3 Analysis

The success of the robust optimization algorithm is predicated on two main assumptions - first, that the effect of a design change on the performance of a given roughness scenario can be predicted, and second, that by improving the performance of a design on a few representative scenarios, the performance of the design can be improved for

the bulk of possible scenarios at once.

Figure 4-3 shows the evolution of the penalty function over the iterations, testing the second assumption. Nearly all of the lines are downward-sloping, particularly the worst-case ones at the top of the figure, indicating that the algorithm was effective at optimizing for the scenarios actively under its consideration. But with the exception of a brief period of significant progress in the first few iterations, the performance improvements on individual roughness scenarios are reliably erased as soon as another scenario is generated to replace it. This supports the conclusion that the algorithm overfit, effectively playing “whack-a-mole” on all the individual ways that a design could be perturbed. There is a slow long-term trend towards lower errors, so the strategy is not ineffective, but it could be considerably improved through tighter regularization constraints, or an increased number of scenarios per iteration.

The linear programming step was based on minimizing the worst-case forecast across the scenario set, which under different conditions might have meant doing a single gradient descent step based on whichever scenario was worst. Of the first 29 iterations from the nominal-seeded run, however, 17 of them were expected to produce a two-way tie for worst scenario, and 2 of them a three-way tie - meaning the linear programming step more often than not was optimizing for more than one scenario at a time, as desired. (This also means that adding more scenarios per iteration would have made a real difference.)

The first assumption was largely validated in the previous chapter, but does not entirely hold here either. Figure 4-4 shows that the forecasts were consistently more optimistic than the actual penalty function’s decreases, implying that the design steps were not actually contained within the penalty function’s linear regime. The discrepancy could be partially due to the quadratic terms in the penalty function, or perhaps also due to the design changes’ symmetry requirements (which rule out the easiest way to linearly affect imbalance). Nevertheless, the strongest forecasts consistently anticipated the strongest changes, so although the precision of the convex optimization was affected, it is unlikely that the quality of the forecasts was a limiting factor in the algorithm’s taking design steps in the right direction.

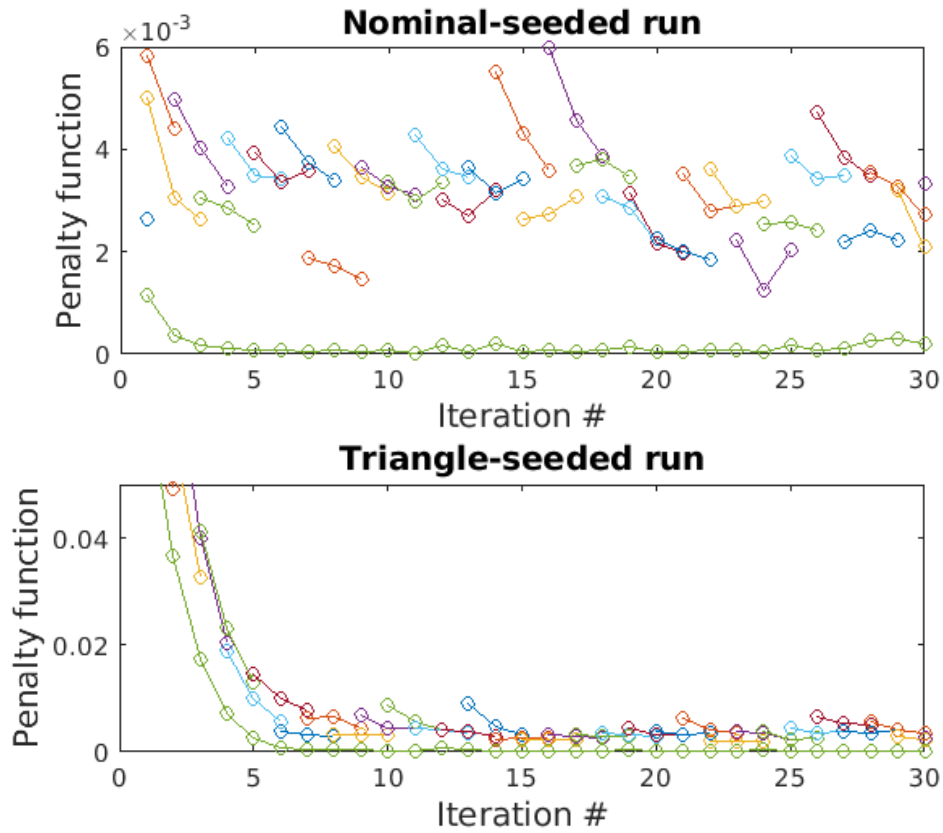


Figure 4-3: Penalty function values for the considered roughness scenarios (including “no roughness”) over the course of the optimization, with lines connecting results from the same scenario.

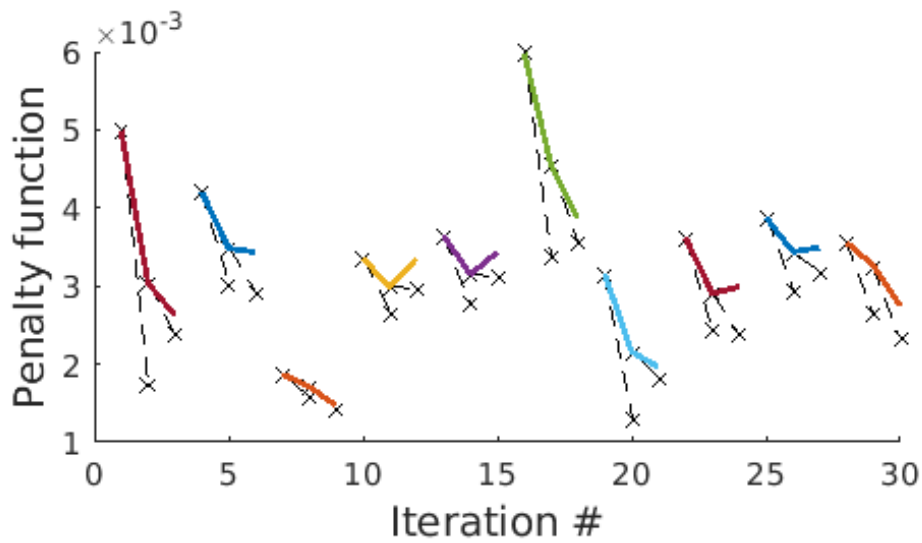


Figure 4-4: Penalty function values for some roughness scenarios over the course of the nominal-seeded optimization, with dashed lines illustrating how the penalty function was forecasted to evolve.





# Chapter 5

## Conclusions and Possibilities

This work is a limited analysis. It examines the performance of one particular implementation of one particular robust optimization algorithm, using one particular implementation of the adjoint method as its underpinning gradient estimator. Furthermore, given more time and attention to parameter tuning, the accuracy and impact of the results could have been significantly enhanced - by using 3D simulations, for instance, with tighter discretizations and mesh cells, or by decreasing the optimization step size, tightening the smoothness requirements, and using more roughness scenarios per iteration. A more involved step would even be to switch from polygonal shape parameterization to a level set parameterization, as in [25], which would likely improve the algorithm's flexibility and resilience to shapes with tight bends. And to make sure that the resulting device performs well once fabricated, the optimization should consider not just line edge roughness, but also width and thickness variations.

However, it is best read as a proof of concept. Using an off-the-shelf FDTD electromagnetic simulator, I have successfully demonstrated the applicability of a relatively simple and adaptable robust optimization algorithm to the problem of line edge roughness in photonic device design. The resulting Y-splitter design, despite its imperfections, has over 20% less root-mean-square power imbalance compared to the nominal design it came from, as well as less insertion loss and equally stable broadband performance. I believe this approach is a step up from heuristic-based robust design principles, and has promise in designing other devices, such as directional couplers

and wavelength multiplexers. It could even be used at the system level, designing cascaded filters or Mach-Zehnder interferometers, although the reduced parameter dimensionality in circuits opens the door to more model-based approaches.

The adjoint method is a powerful and accurate tool for cheaply calculating the shape derivative of an S-parameter, which is useful in uncertainty quantification (UQ) as well as in design optimization. It fails, however, in three types of situations: when the underlying simulation is systematically inaccurate (e.g. when a material boundary aligns with the mesh for long stretches), when optical pulses pass the same geometric pulse multiple times (e.g. in resonant devices), and when the shape derivative is simply insignificant compared to nonlinear effects. Quadratic effects dominate in solitary waveguides (e.g. the arms of a Y-splitter), where the effect of roughness is primarily to decrease phase and partially to introduce scattering loss, but also in any symmetric or optimized device where any perturbation is a decrease in some quantity. This work has shown that in any given device, the quadratic effect of roughness tends to depend only on the correlation length of the roughness, not the nature of the perturbation itself, which suggests there are underlying principles that can be modeled. Developing a model to forecast the quadratic effects of roughness on a device would unlock a new level of accuracy in UQ and in optimization.

Lastly, the hybrid worst-case / random-case scenario generation procedure taken here seemed to produce good results, but I did not experimentally contrast it with a strictly worst-case approach, or a strictly Monte Carlo approach, or an approach that did not retain scenarios for more than one iteration. On the one hand, there did not seem to be a tradeoff between 90th percentile performance and mean performance, so it is possible that it would not make a difference what type of robustness is sought. On the other hand, different approaches react differently to the nonlinearities and dimensionality of an optimization problem - for instance, a worst-case approach may be particularly problematic for a directional coupler, where extreme fluctuations in the narrow gap could destabilize a simulation. So even if the overall result is the same, an algorithm's stability and convergence can be impacted by how a scenario set is chosen, and merits further study.

# Bibliography

- [1] K.O. Abrokwah, P. Chidambaram, and D.S. Boning. Pattern Based Prediction for Plasma Etch. In *The 17th Annual SEMI/IEEE ASMC 2006 Conference*, pages 77–82, Boston, MA, 2006. IEEE.
- [2] Tom Baehr-Jones, Thierry Pinguet, Patrick Lo Guo-Qiang, Steven Danziger, Dennis Prather, and Michael Hochberg. Myths and rumours of silicon photonics. *Nature Photonics*, 6:206–208, March 2012.
- [3] T. Barwicz and H. A. Haus. Three-dimensional analysis of scattering losses due to sidewall roughness in microphotonic waveguides. *Journal of Lightwave Technology*, 23(9):2719–2732, September 2005.
- [4] M. C. Bernardo, R. Buck, L. Liu, W. A. Nazaret, J. Sacks, and W. J. Welch. Integrated circuit design optimization using a sequential strategy. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 11(3):361–372, March 1992.
- [5] D. Cassano, F. Morichetti, and A. Melloni. Statistical Analysis of Photonic Integrated Circuits Via Polynomial-Chaos Expansion. In *Advanced Photonics 2013*, page JT3A.8. Optical Society of America, July 2013.
- [6] Lukas Chrostowski and Michael Hochberg. *Silicon Photonics Design: From Devices to Systems*. Cambridge University Press, March 2015. Google-Books-ID: sJQSBwAAQBAJ.
- [7] Vincent Dubourg. *Adaptive surrogate models for reliability analysis and reliability-based design optimization*. PhD thesis, Université Blaise Pascal - Clermont-Ferrand II, December 2011.
- [8] N. Fajraoui, S. Marelli, and B. Sudret. Sequential Design of Experiment for Sparse Polynomial Chaos Expansions. *SIAM/ASA Journal on Uncertainty Quantification*, 5(1):1061–1085, January 2017.
- [9] R. K. Gupta, S. Chandran, and B. K. Das. Wavelength-Independent Directional Couplers for Integrated Silicon Photonics. *Journal of Lightwave Technology*, 35(22):4916–4923, November 2017.
- [10] Lauren A. Hannah. Stochastic Optimization. In *International Encyclopedia of the Social & Behavioral Sciences*, pages 473–481. Elsevier, 2015.

- [11] J. P. R. Lacey and F. P. Payne. Radiation loss from planar waveguides with random wall imperfections. *IEE Proceedings J - Optoelectronics*, 137(4):282–288, August 1990.
- [12] Jesse Lu and Jelena Vučković. Nanophotonic computational design. *Optics Express*, 21(11):13351–13367, June 2013.
- [13] 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):9712–9733, May 2017.
- [14] Zeqin Lu, Han Yun, Yun Wang, Zhitian Chen, Fan Zhang, Nicolas A. F. Jaeger, and Lukas Chrostowski. Broadband silicon photonic directional coupler using asymmetric-waveguide based phase control. *Optics Express*, 23(3):3795–3808, February 2015.
- [15] Lumerical, Inc. Lumerical’s 2.5d FDTD Propagation Method.
- [16] Chris A. Mack. Reducing roughness in extreme ultraviolet lithography. In *International Conference on Extreme Ultraviolet Lithography 2017*, volume 10450, page 104500P. International Society for Optics and Photonics, October 2017.
- [17] A. Melloni and M. Martinelli. Synthesis of direct-coupled-resonators bandpass filters for WDM systems. *Journal of Lightwave Technology*, 20(2):296–303, February 2002.
- [18] Owen D. Miller. *Photonic Design: From Fundamental Solar Cell Physics to Computational Inverse Design*. PhD thesis, University of California, Berkeley, August 2013. arXiv: 1308.0212.
- [19] Sean Molesky, Zin Lin, Alexander Y. Piggott, Weiliang Jin, Jelena Vučković, and Alejandro W. Rodriguez. Inverse design in nanophotonics. *Nature Photonics*, 12(11):659, November 2018.
- [20] Almir Mutapčić, Stephen Boyd, Ardavan Farjadpour, Steven G. Johnson, and Yehuda Avniel. Robust design of slow-light tapers in periodic waveguides. *Engineering Optimization*, 41(4):365–384, April 2009.
- [21] Ardavan Oskooi and Steven G. Johnson. Electromagnetic Wave Source Conditions. In *Advances in FDTD Computational Electrodynamics: Photonics and Nanotechnology*. Artech House, January 2013. arXiv: 1301.5366.
- [22] Ardavan Oskooi, Almir Mutapčić, Susumu Noda, J. D. Joannopoulos, Stephen P. Boyd, and Steven G. Johnson. Robust optimization of adiabatic tapers for coupling to slow-light photonic-crystal waveguides. *Optics Express*, 20(19):21558–21575, September 2012.

- [23] N. Owen, P. Challenor, P. Menon, and S. Bennani. Comparison of Surrogate-Based Uncertainty Quantification Methods for Computationally Expensive Simulators. *SIAM/ASA Journal on Uncertainty Quantification*, 5(1):403–435, January 2017.
- [24] Alexander Y. Piggott. *Automated design of photonic devices*. PhD thesis, Stanford University, June 2018.
- [25] Alexander Y. Piggott, Jan Petykiewicz, Logan Su, and Jelena Vučković. Fabrication-constrained nanophotonic inverse design. *Scientific Reports*, 7(1):1786, May 2017.
- [26] 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 *8th IEEE International Conference on Group IV Photonics*, pages 71–73, London, United Kingdom, September 2011. IEEE.
- [27] Sameer Rehman and Matthijs Langelaar. Efficient global robust optimization of unconstrained problems affected by parametric uncertainties. *Structural and Multidisciplinary Optimization*, 52(2):319–336, August 2015.
- [28] Jelena Vučković. From inverse design to implementation of robust photonics. In *Advanced Photonics 2018 (BGPP, IPR, NP, NOMA, Sensors, Networks, SPP-Com, SOF)*, page IM2B.5. Optical Society of America, July 2018.
- [29] Y. Wang, Z. Lu, M. Ma, H. Yun, F. Zhang, N. A. F. Jaeger, and L. Chrostowski. Compact Broadband Directional Couplers Using Subwavelength Gratings. *IEEE Photonics Journal*, 8(3):1–8, June 2016.
- [30] Tsui-Wei Weng, Daniele Melati, Andrea Melloni, and Luca Daniel. Stochastic simulation and robust design optimization of integrated photonic filters. *Nanophotonics*, 6(1):299–308, January 2017.
- [31] Yi Zhang, Shuyu Yang, Andy Eu-Jin Lim, Guo-Qiang Lo, Christophe Galland, Tom Baehr-Jones, and Michael Hochberg. A compact and low loss Y-junction for submicron silicon waveguide. *Optics Express*, 21(1):1310, January 2013.
- [32] Zhengxing Zhang, Sally I El-Henawy, Allan Sadun, Ryan Miller, Luca Daniel, Jacob K White, and Duane S Boning. Adjoint-Based Sensitivity Analysis for Silicon Photonic Variations. In *2019 IEEE MTT-S International Conference on Numerical Electromagnetic and Multiphysics Modeling and Optimization (NEMO)*, Cambridge, MA, May 2019. IEEE.