Compact, Fabrication-Resilient Inverse-Designed Wavelength Demultiplexers in Silicon-on-Insulator Technology

Andy S. Li

McGill

Department of Electrical & Computer Engineering

McGill University

Montréal, Québec, Canada

August 2024

A thesis presented for the degree of Master of Science.

Abstract

Silicon photonics enables the design and fabrication of photonic circuits using methods similar to those of complementary metal–oxide–semiconductor (CMOS) electronic circuits, allowing lower fabrication costs and smaller device footprints by reusing mature design methods that have been refined over decades in the electronic field. Silicon photonics can be combined with the modern method of inverse design, which leverages new software tools that allow for fast, accurate physical simulations to design unintuitively shaped devices that have high performance despite their lack of resemblance to most common photonic devices. The combination of these two technologies enables the fabrication of physical photonic devices that are orders of magnitude smaller than their counterparts based on other technologies, yet can be fabricated quickly and accurately in dense photonic circuits.

At extremely small scales, though, even the accuracy of common silicon photonics fabrication equipment is insufficient to maintain the performance of these inverse-designed devices. In a device with a footprint of only a few square microns, even differences of nanometers between an intended device design and the actual fabricated device can

significantly affect device performance. To reduce the effect of these fabrication errors, we use an in-house machine-learning toolkit, Prefab, to design a device while correcting for expected fabrication errors, ensuring its resulting fabricated counterpart will have high performance even with the effect of these errors.

In this thesis, we study the design and fabrication methods that play a role in the performance of silicon photonic devices and explore how inverse design interacts with these methods. We then use these technologies to design and fabricate a set of C-band (wavelengths near 1550 nm) wavelength demultiplexers with footprints of 3×3 μm^2 and 3×5 μm^2. We then develop a method to accurately simulate these devices before they are fabricated. These demultiplexers, with output channels at a wavelength spacing of 20 nm, are then fabricated both with and without the error corrections of Prefab. While the uncorrected devices showed marginal performance due to a significant wavelength shift in the transmission spectrum from the simulated devices, the corrected devices reduced this shift from a median value of 19.6 nm to -3.5 nm for a 3×3 μm^2 device. A 3×5 μm^2 device showed similar improvements, with the wavelength shift falling from 23.9 nm to -2.6 nm following our error corrections. Finally, we examine methods to extend inverse design of silicon photonic devices to devices with more complicated requirements, such as athermal (temperature-insensitive) photonic devices, and explore methods to reduce optimization resources by designing a device in multiple steps.

Abrégé

La technologie photonique sur silicium (SiPh) est utilisée pour la fabrication des circuits optiques avec des méthodes semblables à celles utilisées pour la fabrication des circuits électroniques avec la technologie de semiconducteurs à oxide de métal complémentaire (CMOS). Cette similarité avec une technologie utilisée depuis plus de 50 ans réduit les coûts de fabrication et la taille ultime d'un dispositif photonique. Il est possible de combiner la technologie SiPh et la technologie moderne de rétro-ingénierie photonique, qui utilise un logiciel de simulation de photonique afin de concevoir un dispositif performant de forme géométrique arbitraire. Les deux technologies produisent ensemble un dispositif plus petit qu'un dispositif photonique conventionnel.

À très petit echelle, la précision des machines de lithographie SiPh est insuiffisante pour maintenir les performances d'un dispositif photonique. Lorsque la largeur d'un dispositif est de quelques micromètres, un erreur d'une dizaine de nanomètres peut nuire à ces performances. Dans notre recherche, nous avons utilisé un logiciel d'apprentissage automatique créé dans notre groupe de recherche. Ce logicel, appellé Prefab, permet de

concevoir un dispositif performant même si les erreurs de fabrication modifient sa géométrie.

Dans cette thèse, nous analysons les méthodes de conception et fabrication qui influencent les performances des dispositifs SiPh et l'effet de la rétro-ingénierie sur ces méthodes. Ensuite, nous utilisons ces technologies pour concevoir des démultiplexeurs de longueur d'onde avec un encombrement de 3×3 µm^2 et 3×5 µm^2. Nous développons une méthode de simulation précise avant la fabrication. Après, nous fabriquons les démultiplexeurs avec et sans correction. Bien que les dispositifs non corrigés soient non performants en raison d'un décalage de leurs canaux de transmission, les dispositifs corrigés réduisent ce décalage de 19.6 nm à -3.5 nm pour le démultiplexeur 3×3 µm^2. L'amélioration du démultiplexeur 3×5 µm^2 est similaire, de 23.9 nm à -2.6 nm, après les corrections de Prefab. Enfin, nous étudions l'application de la rétro-ingénierie sur des dispositifs ayant des exigences plus compliquées, tels que les dispositifs athermiques (avec une large plage de température) et développons une méthode de rétro-ingénierie en plusieurs étages.

Acknowledgements

The past two years have been an exciting, challenging time, and it has been a great and fulfilling experience to complete my master's research and present this thesis as I approach the end of my program. I could not have done it without my supervisor, Prof. Odile Liboiron-Ladouceur, who has supported me and always pushed me to improve while being accommodating of the engagements and outside obligations that can arise in the life of a student. I consider myself lucky to have been able to work on my project with such an encouraging and understanding supervisor, and it would not have been possible to have finished my project otherwise.

The members of my research group have been just as supportive of my work as my supervisor has, always ready to provide help and share their experiences when I run into a problem. Special thanks go to Dr. Dusan Gostimirovic, who developed and helped me use the Prefab corrector that forms an integral part of my work; Kevin Kong, who helped me with SEM imaging of our fabricated devices; Sunami Morrison, who helped me with formatting and style as I wrote this thesis; and Rebecca Rogers, who reviewed the first draft of my

thesis.

I would also like to thank our colleagues at the National Research Council Canada, Dr. Dan-Xia Xu and Dr. Yuri Grinberg, who have provided invaluable technical help with designing and fabricating our devices and helped me interpret the resulting data afterward. It was a joy to work with them through our online meetings and in person in Ottawa. In addition, I am grateful to NRC's Martin Vachon, who both measured our devices using automated test equipment and helped me interpret the results; and Dr. Siegfried Janz, who provided space on his fabrication run for our designs at short notice.

I have also been lucky enough to count on the support of my friends, who have shared with me the ups and downs of my journey in this program. Anya, Jana, Kaan, Kelly, Priya, and Sneha have made my time here in Montreal more than just work and research. Thank you for your constant support, and I wish the best to those of you who are making the same academic journey as I am.

Of course, I would never have gotten here without counting on the unconditional love and support of my family. Now and always, you have been supportive of what I do and proud of my accomplishments. Though I can hardly convey it appropriately in such a short acknowledgement, I am proud to be part of our family and I love each of you. I hope to support you in the future as much as you have supported me as I grew up and reached where I am today.

Contents

List of Figures

List of Tables

List of Acronyms

ANT	Applied Nanotools.
ATE	automated test equipment.
CMOS	complementary metal–oxide–semiconductor.
CPU	central processing unit.
DUT	device under test.
DWDM	dense wavelength-division multiplexing.
EBL	electron-beam lithography.
ER	extinction ratio.
FDFD	finite-difference frequency domain.
FDTD	finite-difference time domain.
FOM	figure of merit.
GC	grating coupler.
GDS	graphic design system.
GPU	graphics processing unit.

HPC	high-performance computing.
IL	insertion loss.
MZI	Mach-Zehnder interferometer.
OOB	out-of-band.
SEM	scanning electron microscope.
SiPh	silicon photonics.
SOI	silicon-on-insulator.
TE	transverse electric.
TM	transverse magnetic.
WDM	wavelength-division multiplexing.
XT	crosstalk.

Chapter 1

Introduction

As the amount of data transmitted and stored has increased exponentially in recent years, the systems used to transmit data, both locally and globally, must operate at increasing data rates without using more power or more physical space. In recent decades, this data transmission has moved from copper cables to optical fibers as data rates have increased from megabits to petabits per second. This use of optical fiber greatly increases the throughput of data-transmission links but requires optical components, like (de-)multiplexers, to be developed to control and manage these significant flows. The design of these components, and the overcoming of design challenges that complicate the design of smaller optical components, are the focus of our work and of this thesis.

1.1 Optical data transmission

Optical data transmission has emerged to accommodate the transmission of unprecedented amounts of data due to its much larger bandwidth than electrical data transmission. Thus, optical systems can transmit more data over longer distances than similar electronic systems [1]. These advantages have become more relevant in recent years, as the continued growth of datacenters and the internet have increased annual data flows to thousands of exabytes, with growth continuing at a rate of 30 percent per year [2]. The transition to optical data transmission has been enabled by the commercialization of photonic devices, such as transmitters and multiplexers, which enable the electrical signals from a computer to be converted into optical signals, aggregated, and routed to their destinations.

1.2 Multiplexing

While the simplest optical transmission model involves a single transmitter and receiver, real-world optical systems often must share a single transmission line with other data transmissions. This sharing, known as multiplexing, is crucial for making use of high-bandwidth optical transmission and for enabling optical switching and networking. To implement multiplexing, individual optical signals (in separate channels) enter an optical multiplexer, where the channels are combined into a single output channel. This output is then coupled to one end of an optical waveguide (such as a long optical fiber), which is

connected to the input channel of a demultiplexer (which splits the multiplexed transmission back into individual signals). To make full use of the bandwidth of the optical waveguide and ensure that the multiplexed channels can be easily separated, each multiplexed signal is slightly different from other signals, commonly by being transmitted on a different wavelength. This system, known as wavelength-division multiplexing (WDM) [3], uses input channels carried by different wavelengths of light, each chosen so the resulting spectrum does not overlap with those of the other channels (Fig. 1.1).

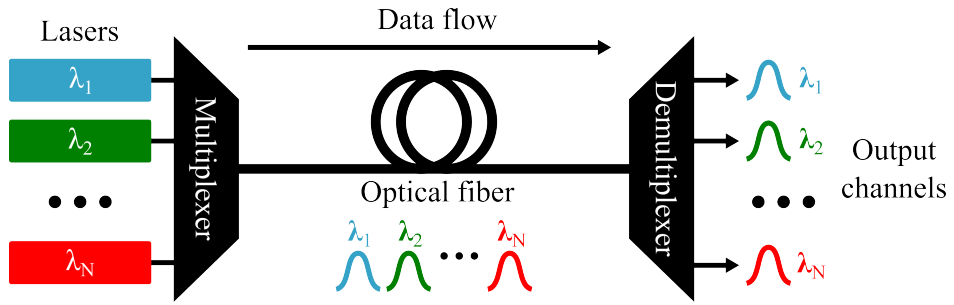


Figure 1.1: High-level schematic view of a WDM data-transmission system with N optical data channels.

Wavelength-division multiplexing has become a part of many optical data transmission systems used today, with increased channel density allowing for more data to be transmitted on a single waveguide. It is thus useful for WDM demultiplexers to have narrow channel spacings to allow for high channel density; for example, modern dense wavelength-division multiplexing (DWDM) systems use channel spacings of under 1 nm [4].

1.3 Motivation

Modern photonic devices allow data transfer at a rate far higher than with their electronic counterparts, but the persistent growth in data transmission has continued to demand higher-throughput, more efficient devices. A notable limitation on the ability of photonic devices to transmit large amounts of data is the density of connections, as higher-density systems allow for more efficient use of datacenter space and resources. Copper-based solutions can use thin, sharply bent wires both within integrated circuits and inside cables connecting computers; however, photonic solutions have traditionally used optical cables and discrete devices. These must be far larger than their copper counterparts to accommodate the inability of most optical waveguides to maintain light confinement inside sharp bends. Naturally, this limits the flexibility and functionality of photonic devices for data transmission by reducing the number of possible links inside a space-constrained datacenter setting.

In the past few years, silicon photonics has emerged as a technology that may allow for new photonic components that are much smaller than previous components, allowing for denser interconnections at lower cost. In parallel, the expansion of computer-assisted design techniques like inverse design has increased the possibilities of these components by unlocking new device topologies with functionalities not previously possible in a single device [5].

With photonic optical design in a silicon-on-insulator (SOI) process, we explore methodologies to address the problem of device density by designing compact wavelength demultiplexers using inverse design. We focus on two-channel C-band demultiplexers with a

footprint of 3×5 µm^2 and 3×3 µm^2. In addition, we designed demultiplexers with suppressed transmission outside the wavelength channels. These demultiplexers were fabricated in a SOI process, which places waveguides on a silicon die similar to the substrate used for silicon microchips. To reduce the effect on performance of fabrication errors in a small device, we used an in-house machine-learning software tool called Prefab to predict and correct for changes in the device shape caused by fabrication error.

Existing works have discussed the inverse design of compact demultiplexers, but many of these works [6–10] have done so based on simulated device layouts rather than fabricated devices. Among those which have discussed fabricated devices, these devices tend to have wide channel spacings (e.g., 250 nm in [11]) or larger footprints (e.g., 5.4×6.2 µm^2 in [12]), which reduces the channel and device density possible in a photonic system. We have thus designed our devices to have a relatively narrow channel spacing of 20 nm in a small footprint of either 3×3 µm^2 or 3×5 µm^2, while also fabricating and measuring our devices to determine the effect of imperfections in the fabrication process.

1.4 Research contributions

The research presented in this thesis was conducted by Andy Li, the candidate and author of this thesis, under the supervision of Prof. Odile Liboiron-Ladouceur. This research was undertaken in collaboration with the National Research Council Canada (NRC) as part of the Artificial Intelligence for Design Challenge program. The devices and procedures described

in this thesis have been developed and tested by the author using simulation and design tools developed by others. The fabricated designs created as part of this research were fabricated by Applied Nanotools (ANT). The measurements presented in this thesis were taken by NRC, with additional measurements taken by the author. This work and thesis are expected to form part of a future manuscript discussing inverse design and demultiplexers, with the remainder of the manuscript to be drafted upon receipt of further data.

1.5 Thesis organization

This thesis is divided into five chapters with subsections. In chapter 2, the background of our project is explained, with a discussion of the physical phenomena that enable the use of silicon photonics and the effect of temperature on these phenomena. In addition, the finite-difference time domain (FDTD) algorithm used in simulations and photonic inverse design is explained. We discuss the mathematical basis of inverse design for photonic devices and our methodology for designing our devices. We explain gradient descent, figures of merit, design spaces, topological inverse design, and superoptimizations. We further discuss the choices we made for our designs and discuss the computational and time requirements of these inverse-designed devices.

In chapter 3, we further examine the simulation of photonic devices with FDTD software tools. We discuss the simulation accuracy of inverse-designed devices and our methods to improve these simulations. We also discuss the errors most commonly seen in

the fabrication of silicon photonic devices and their resulting effects on the performance of the devices. In addition, we detail two methods to reduce these effects: design-for-manufacturing optimization (the default method for the optimization tools we used) and the Prefab machine-learning method. We further explain Prefab and its training data in this chapter.

In chapter 4, we show the inverse-designed wavelength demultiplexers that we have designed, fabricated, and measured. The first device (section 4.1) is a 3×3 μm^2 two-channel, C-band (1550-nm) demultiplexer. The second device (section 4.2) increases the device footprint to 3×5 μm^2 and reduces optical transmission outside the wavelength ranges of the two output channels. The third device (section 4.3) similarly reduces out-of-band (OOB) transmission, but reduces the device footprint to 3×3 μm^2.

In section 4.5, we explore the use of inverse design to create athermal photonic devices by optimizing for a figure of merit (FOM) in multiple simulations at different temperatures (superoptimization). We simulate a device that has been optimized for multiple temperatures and compare it with a device that has been optimized at a single temperature. We then measure the effect of temperature and the benefit of athermal superoptimization.

In section 4.6, we attempt to reduce the time and resources necessary for device optimization by examining a multi-step optimization process that begins with a simpler FOM before adding additional components to the FOM and continuing device optimization. We optimize a device using both this multi-step optimization and a standard,

single-step optimization and characterize the resulting devices' performance and optimization time.

Finally, we conclude in chapter 5 by discussing the impact of this work and its potential application in broader environments. We discuss possible next steps and future works that could be based on the techniques and results discussed in this thesis.

Chapter 2

Background

2.1 Silicon photonics and the SOI process

The basis of an optical data-transmission system is the use of a waveguide to direct light between a transmitter and a receiver. In most long-distance systems, the waveguide is a silica-based optical fiber, which exhibits low attenuation but cannot be bent sharply due to low confinement [1]. As optical transceivers become smaller to increase data density, these bends would take up more space than the components they connect. For photonic devices such as demultiplexers, a recent technology is the silicon-on-insulator (SOI) process, which uses rectangular silicon waveguides with a silicon dioxide (silica) cladding to confine light. A key advantage of the SOI process is the large contrast between the refractive indices of the two materials ($n = 3.48$ for silicon and $n = 1.44$ for silica [13, 14]), which increases optical

mode confinement and allows for tight bends and small features in waveguides with minimal loss of light. These features are crucial for practical photonic integrated circuits, by reducing the distance between individual components that must be devoted to connecting waveguides. Furthermore, SOI devices can generally be fabricated using existing technology, notably the electron-beam lithography (EBL) process for patterning devices onto silicon wafers similar to that already used in electronic CMOS integrated circuits [15]. These technologies have been used for decades in the manufacturing of electronic devices, increasing fabrication accuracy and reducing cost through the reuse of mature technology. The similarity between these photonic and electronic circuits even allows both types of components to be combined on a single substrate, allowing for electronic-photonic integration at the scale of a single die [16].

2.2 SOI technology: mode confinement in rectangular waveguides

Any electromagnetic wave propagating through a material must satisfy Maxwell's equations, which model the relationship between electric fields ($\bar{E}$) and magnetic fields ($\bar{H}$) in time (t) and space based on the magnetic permeability (μ) and electric permittivity (ϵ) of the material in which the wave is propagating [17]:

$$\nabla \times \bar{E} = -\mu \frac{\partial \bar{H}}{\partial t} \tag{2.1}$$

$$\nabla \times \bar{H} = -\epsilon \frac{\partial \bar{E}}{\partial t} \tag{2.2}$$

These equations model propagation in a material with defined values for μ and ϵ. The vector equations can be combined into a simpler scalar equation by setting the coordinate system to define propagation along a coordinate axis (in direction i), usually defined to be the z-axis for a waveguide:

$$\nabla^2 E_i - \mu\epsilon \frac{\partial^2 E_i}{\partial t^2} = 0 \tag{2.3}$$

Silicon photonic waveguides have a rectangular cross-section because of their fabrication process, which patterns a two-dimensional layout on a uniform layer of silicon (Si) between layers of silica (SiO_2), as shown in Fig. 2.1. Mode propagation in these waveguides can be modelled by an infinite slab waveguide consisting of a layer of high-permittivity material (e.g., silicon) surrounded by lower-permittivity material (e.g., silica) (Fig. 2.2), allowing for significant conceptual simplification of propagation calculations [17].

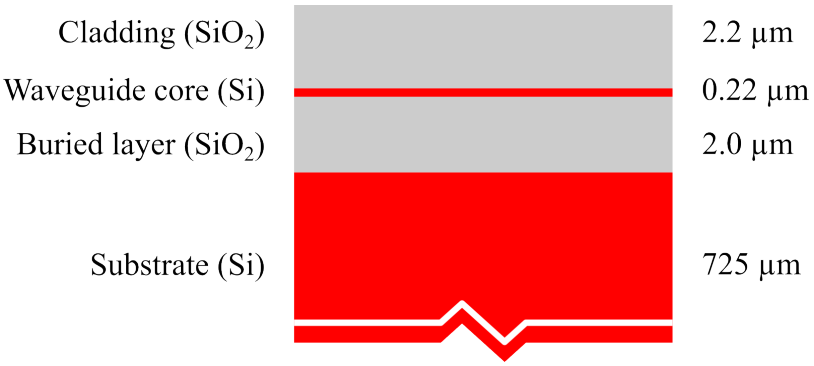


Figure 2.1: Cross-section of an SOI wafer as used in our design.

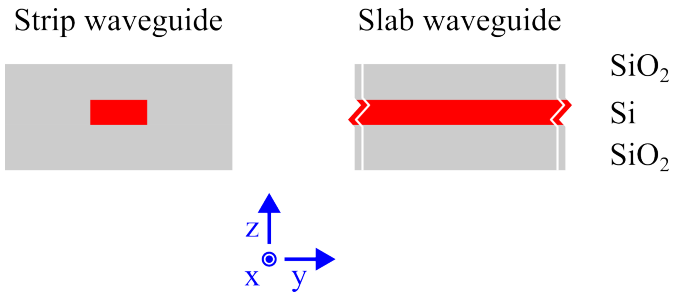


Figure 2.2: Cross-section of a strip waveguide (left) and a slab waveguide (right) fabricated on an SOI wafer, with the axis directions labeled in blue.

The solutions to Maxwell's equations inside a layer of a slab waveguide can be split into two polarizations. In the first, transverse electric (TE) polarization, the electric field is considered to be along the y-axis (transverse to the boundary between layers and to the propagation direction). This means that the electric field ($E_y(x,z)$) propagating in the core or cladding must have the form of equations 2.4 and 2.5, with n_i as the refractive index of the corresponding material, β the z-component of the wavevector, and E_0 the magnitude of the electric field at $x = 0$.

$$E_y(x,z) = E_y(x)e^{-j\beta z} \tag{2.4}$$

$$E_y(x) = E_0 e^{-\sqrt{\beta^2 - k_0^2 n_i^2}x} \tag{2.5}$$

in order to satisfy the scalar TE version of the wave equation in a slab waveguide, with k_0 set as the value of the vacuum wavevector ($k_0 = \omega_0/c$, where $\omega_0 = 2\pi f$ and c is the speed of light in free space):

$$\nabla^2 E_y + k_0^2 n_i^2 E_y = 0 \tag{2.6}$$

From the square-root component in the exponent of equation 2.5, it can be seen that β must be less than $k_0 n_i$ for the wave to propagate, as the energy would otherwise be unconfined, attenuated exponentially in a way similar to that of a wave traveling through a lossy bulk material. If β is between $k_0 n_1$ and $k_0 n_2$ (where n_1 and n_2 are respectively the refractive indices of the core and cladding material), however, the wave will propagate solely in the core. In the ray-optic model, this corresponds to total internal reflection, the basis of propagation through any dielectric waveguide.

A transverse magnetic (TM)-polarized wave[1] propagates similarly to a TE-polarized wave in a rectangular waveguide, except that the magnetic-field component is transverse to the material boundary. This means that the behaviour of a propagating TM-polarized electromagnetic wave differs from that of a TE-polarized one; this behaviour is calculated using equations reflecting the difference in field-component orientation. Because rectangular waveguides have a clearly defined orientation and maintain the polarization of propagating waves, it is useful to classify waves based on their polarization, as differently polarized waves have significantly different propagation properties in rectangular waveguides.

While an electromagnetic wave will remain confined within a waveguide if its β value is within a certain range, phase matching is also required for a wave to propagate through the waveguide. Because a propagating wave is reflected along the interface between materials, its travel can be conceptualized as in a zig-zag pattern inside the core of the waveguide, as

[1] In strip waveguides, polarizations are known as quasi-TE and quasi-TM (or TE/TM-like) because the finite width of the waveguide causes waves to behave slightly differently compared to an infinite slab waveguide [15]. For simplicity, these polarizations are referred to as TE or TM in this thesis.

in Fig. 2.3.

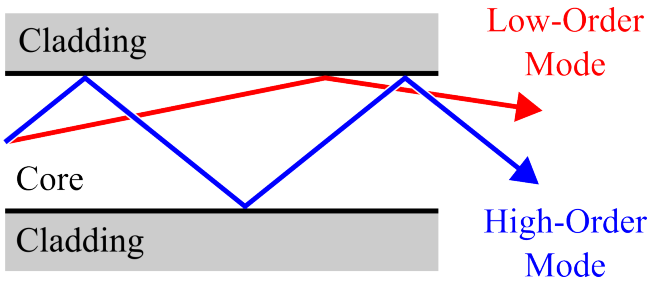


Figure 2.3: Paths taken by light as they propagate and reflect off the material interface inside an optical waveguide.

At each reflection, the phase of the reflected light shifts by an amount depending on the angle at which it meets the material interface, and thus the value of β. To allow propagation over long distances, the value of β must be set so that this phase – known as the Goos-Hänchen shift – is an integer multiple of 2π after two reflections [17]. This allows light to have a consistent phase shift at each point along the axis of a waveguide, preventing destructive interference that would quickly attenuate the propagating light. The phase-matching requirement limits the possible values of β to a small set of discrete values known as optical modes. These are numbered from 0 upward (starting with the highest allowed value of β), and are different between TE and TM polarizations.

As the value of beta decreases and the mode number increases, it eventually reaches a cutoff point ($\beta = k_0 n_2$) where the waveguide no longer confines an electromagnetic wave. At this cutoff point, the angle at which the wave meets the material surface (Fig. 2.4) is known as the critical angle because only waves at shallower angles can experience total internal

reflection. This value is important in integrated devices because it determines the loss at a waveguide bend, where the angle at which the propagating wave reflects is increased by the physical bending of the waveguide, with significant losses if the sum of the two exceeds the critical angle. To allow sharp angles while maintaining total internal reflection, it is useful to maximize the difference between the refractive indices of the core and cladding material, as this increases the cutoff angle and thus the number of modes that can be guided through a sharp waveguide bend.

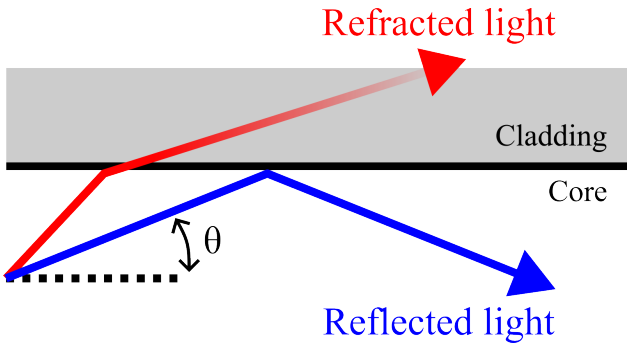


Figure 2.4: Illustration of the angle at which light reflects off a material interface. Light with an angle (θ) above the critical angle will not experience total internal reflection (red line).

Silicon photonics, in particular, allows for tight bends with low loss because of the high contrast between silicon ($n = 3.48$) and silica ($n = 1.44$). For example, a right-angle bend with a radius of 3 µm causes a loss of less than 0.02 dB [18]; in comparison, a standard optical fiber with a bend radius of 1 cm has a loss of 100 dB [1]. Because of the large number of waveguide bends in a photonic integrated circuit, the small bend radius of silicon photonics

allows for much smaller circuits without severe losses.

2.3 Demultiplexers

When a multiplexed optical signal reaches its destination, it contains a large number of individual signals that must be directed into separate output channels. Thus, the photonic demultiplexers which perform this separation must be present in optical data-transmission systems. Contemporary designs of wavelength-division demultiplexers in silicon photonics are generally based on structures with well-known, manually designed layouts; prominent examples of these include Mach-Zehnder interferometer (MZI)-based devices, echelle gratings, and arrayed waveguide gratings [19, 20], whose shapes and dimensions are based on a physical understanding of how waves behave in them. The dimensions of these devices can be tuned to affect the transmission spectrum of the demultiplexer in predictable ways based on concepts such as mode coupling and phase shifting.

While these contemporary devices are commonly used in silicon photonics, they suffer from drawbacks inherent to their design. Most notably, these devices use coupled waveguides with a defined dimensions that have been chosen based on desired wavelength characteristics. To fit these lengths of waveguide, a large device footprint is required even for silicon-on-waveguide devices. Thus, a single demultiplexer fabricated in the SOI process can be over 1 mm square [21], increasing the expense of device fabrication and reducing the efficiency of the resulting photonic device.

In addition to their large size, conventional demultiplexers based on MZI topologies have limited out-of-band (OOB) suppression because of the small range of wavelengths for which destructive interference can be maintained [15]. In the case of wavelength-division multiplexing, it may be necessary for these devices to be combined with filters or additional demultiplexers to limit the transmission of out-of-band light, further increasing the footprint and insertion loss of these devices.

Demultiplexers are generally characterized based on key performance criteria that determine their suitability for use in multiplexed systems. Important criteria include the crosstalk (at each output, how much stronger the desired wavelengths are compared to non-desired wavelengths) and insertion loss (how much the demultiplexer attenuates a signal passing through), as shown in Fig. 2.5. In addition, the channel width measures the wavelength range within which each channel maintains its performance. The crossover point between channels, where the transmission is the same from both channels, is useful for measuring any shift in the transmission spectrum and ensuring that other performance measurements are taken inside the shifted channels.

A representative design based on a silicon-on-insulator echelle grating, for example, has an insertion loss of about 3 dB and crosstalk of 19 dB in an eight-channel demultiplexer [22]. Like most other forward-designed demultiplexers, this device has a large footprint, in this case 250×200 µm^2.

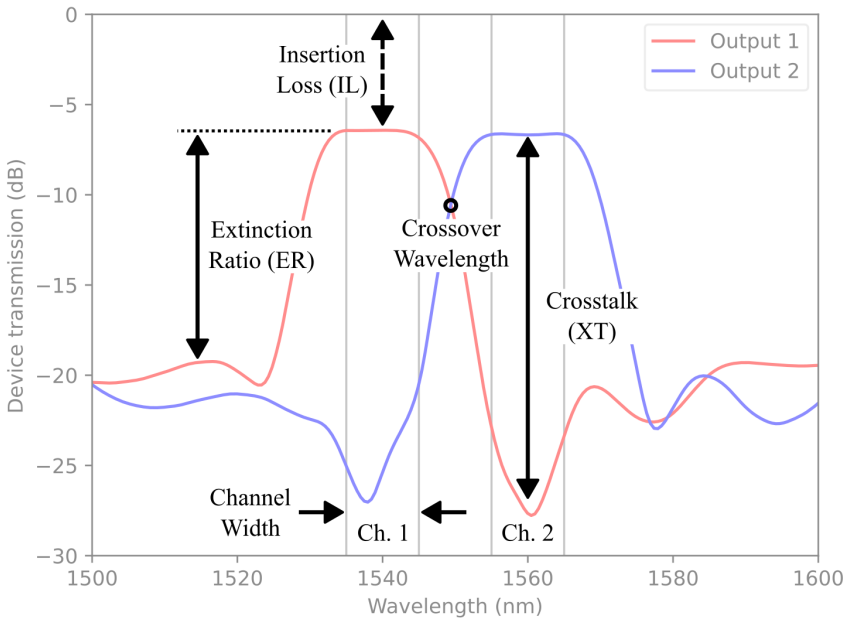


Figure 2.5: Example demultiplexer transmission spectrum with channel bandwidth, insertion loss (IL), extinction ratio (ER), and crosstalk (XT) labeled. Note that a two-channel demultiplexer is shown here; the definitions of ER and XT become more complex when more than two channels are measured.

2.4 Thermal effects and device performance

The performance of any device fabricated with silicon-on-insulator technology is affected by the temperature dependence of the materials involved. The refractive indices of both silicon and silica change with temperature, affecting the propagation of optical modes inside a waveguide. In the case of a silicon-on-insulator device near 1550 nm, the change in refractive index caused by temperature is much larger for silicon ($dn/dT = 1.87 \times 10^{-4}$ per K) [23] than for silica ($dn/dT = 1.16 \times 10^{-5}$ per K) [14],

increasing mode confinement at higher temperatures. While this affects both forward-designed devices (those based on a known topology, such as MZI-based devices) and inverse-designed devices (those designed to maximize a FOM, such as transmission), it is particularly noticeable with small inverse-designed devices because of their small features and unintuitive designs. For inverse-designed demultiplexers like those designed in this thesis, we found that temperature changes cause the transmission spectrum to shift significantly, reducing the performance of our devices. The effect of the change in refractive index on waveguide propagation is more prominent than that of physical thermal expansion (α), which differs less between silicon ($\alpha = 2.35 \times 10^{-6}$ per K) and silica ($\alpha = 0.44 \times 10^{-6}$ per K) while also having a smaller magnitude [24].

Maintaining device performance across a range of temperatures is thus a problem that must be addressed for practical photonic devices, which are often used in environments near heat-generating electronics which would cause significant wavelength shifts in a SOI inverse-designed demultiplexer. One method of maintaining performance is to attach a thermal stage that keeps the temperature of a photonic device within a small range [25]. While this system allows device performance to remain relatively constant across a larger temperature range, it adds significant complexity as a thermal stage requires heaters and control logic to maintain a set temperature range, consuming power and physical space in environments where both are at a premium.

Thus, it would be beneficial to design a device that can maintain its performance across

a temperature range large enough that a thermal stage would not be necessary in real-world conditions. Photonic devices with this property, known as athermal devices, have been designed using standard forward-design methods such as tuning the lengths of waveguides and MZIs based on known relations between waveguide characteristics and temperature variances; however, such forward-designed devices can be very large, with a footprint larger than 1 mm square, because they cascade multiple MZIs to achieve a single athermal device [21]. Inverse design may be able to design smaller devices with similar athermal performance by extending its current algorithms to account for the effects of temperature on mode confinement.

2.5 Optical simulations

Before fabricating an optical device, such as a demultiplexer, it is useful to predict its performance using a simulation tool. These software tools simulate the performance of a given device layout by solving for Maxwell's equations to predict how light will travel within the materials in optical devices and waveguides.

One common simulation method used in this project is the finite-difference time domain (FDTD) method, which calculates the behaviour of light in a device by splitting the device into individual cells and solving differential equations that correspond to the physical behaviour of light inside each cell [26]. This method is especially beneficial to the design of devices like multiplexers because it allows for a device's performance across a range of

input wavelengths to be calculated in one step (a broadband simulation), whereas many other methods, like the similar finite-difference frequency domain (FDFD) method, often require separate simulations for each individual wavelength [27].

A significant choice when simulating photonic devices with an FDTD solver is the use of two-dimensional or three-dimensional FDTD simulations. While both types of FDTD solvers split a device into individual cells, a two-dimensional solver only does so along two dimensions: usually, along the length and width of a device but not along its depth (thickness). This reduces the time and computational resources necessary to simulate a device by assuming uniform energy propagation along the thickness of a device, greatly reducing the number of cells and calculations necessary. A three-dimensional solver also splits the device into cells along the depth of a device, increasing simulation time but generating a more accurate result by taking into account non-uniform energy distribution along the thickness of a silicon photonic device. However, this increases the number of cells significantly, as each 2D cell must be split along the thickness of the device into multiple 3D cells; this increases computation time and resource usage, especially memory usage, proportionally. This tradeoff means that it is often useful to simulate and optimize devices primarily using 2D FDTD simulation and limit 3D simulations to refinement of 2D designs before they are fabricated (and thus where increased accuracy is most useful).

In addition to allowing the characterization of a device design before it is fabricated, optical simulation is key to the inverse design of optical devices because the inverse-design

process optimizes a device by simulating a large number of possible designs. At each design step, the algorithm attempts to improve the performance of the simulated device by calculating a gradient, which requires at least two simulations per step with the adjoint method (section 2.7). Inverse design as used in our project depends on accurate and fast simulation tools, without which it would be impossible to design devices as we have.

2.6 Inverse design

The ability to simulate a device enables a new design method, that of inverse design. Whereas a photonic device is usually designed based on a known device layout with dimensions tuned to achieve a desired transmission spectrum (forward design), inverse design searches a large design space to find a suitable design with the goal of maximizing a given FOM. Devices designed with this method can exhibit high performance with an unintuitive design, like that shown in Fig. 2.6. While there are several software packages that enable inverse design of photonic devices, we use the package Lumopt, which implements gradient descent in conjunction with FDTD simulations from Ansys Lumerical.

The unintuitive design of an inverse-designed device like in Fig. 2.6 generally functions on a physical level by causing input light at different wavelengths to take different paths through the design region. This causes the light to be physically directed to one of the two output ports, as shown in Fig. 2.7. However, the nature of the inverse-design optimization means that this behaviour is not specifically designed for, and it may be possible that other

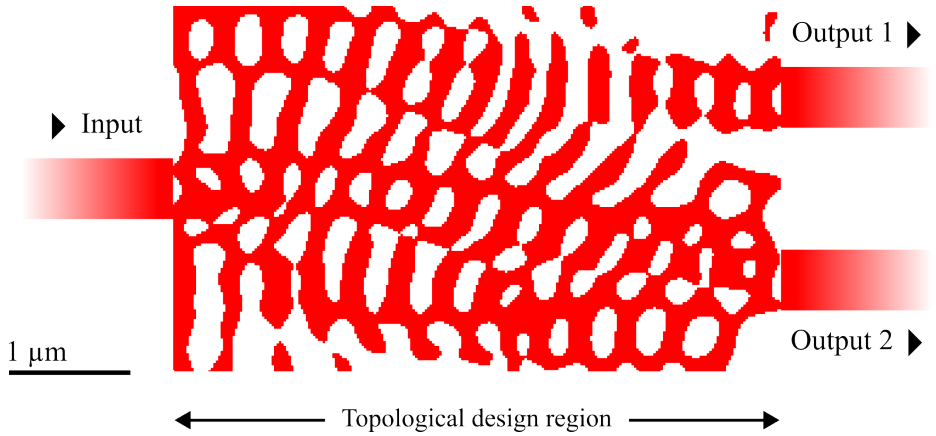


Figure 2.6: An example of an inverse-designed two-channel demultiplexer with waveguides connecting to its right and left edges. The red areas shown in the figure correspond to the regions of silicon in the fabricated device.

devices operating on different physical principles can be generated by the optimizer.

When using inverse design to design a device, the transmission spectrum of the device is controlled by defining a FOM that the optimizer will try to maximize or minimize. In our project, the FOM is defined by calculating the mode matching (transmission for a given mode) between the input and output waveguides of a device at a set of given wavelengths. At each wavelength, the mode matching is either maximized or minimized by defining the ideal mode matching (either zero or one) and finding the difference between the simulated value for mode matching and the ideal value. These differences, multiplied by a weight value, are then added together across all target wavelengths and the resulting value – the FOM – is then minimized by the inverse-design algorithm [28]. By controlling the ideal mode-matching values and the wavelengths at which they are measured, different types of devices can be

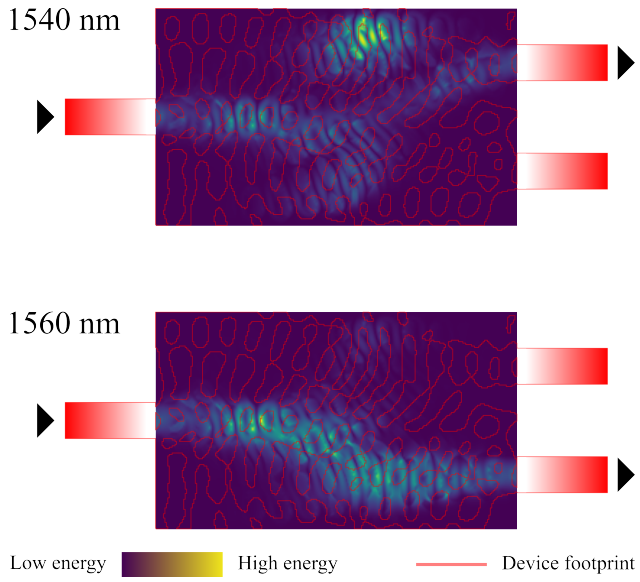


Figure 2.7: Paths taken by 1540 nm and 1560 nm light through the example inverse-designed demultiplexer.

designed with a single algorithm.

2.7 Gradient descent and the adjoint method

Photonic inverse-design algorithms rely on an optimization method known as gradient descent, in which device performance is optimized by calculating the gradient of a FOM with respect to a set of device parameters. In this process, a device with a set of parameters is simulated using a photonic simulation tool such as Lumerical and a FOM is calculated based on the device's performance. The gradient of the FOM is calculated using the adjoint method discussed below in this section; this gradient defines how changes in the

device parameters will affect the value of the FOM [29]. Once the algorithm calculates these effects, it changes the parameters of the device in a way predicted by the gradient calculation to maximize the FOM.

To optimize a device using gradient descent, it is necessary to efficiently calculate the gradient of the FOM with respect to the parameters in a matrix that defines the device. This determines how the FOM would change if an individual parameter or set of parameters changes and allows the optimizing algorithm to find the optimal parameters for the next step of the optimization. In the case of a FOM that is a sum of transmission values at multiple wavelengths, the gradient corresponds to a sum of partial derivatives which measure the change in transmission at each frequency when device parameters are changed.

While the gradient can be calculated at each step by calculating each derivative, by changing each parameter slightly and simulating to measure the resulting change in the FOM, this would require hundreds of computationally expensive simulations at each optimization step. To reduce this number, it is possible to algebraically calculate these derivatives (and the resulting gradient) using the adjoint method [30]. The adjoint method takes advantage of how a small change in the permittivity of any pixel changes electric-field propagation through a device, and thus its figure of merit, only slightly. With this assumption, an adjoint field can be calculated by simulating a device normally (forward simulation) and then with the source and output measurements reversed (adjoint simulation) for each output, in order to measure the propagation of a small electric field from the device's output port to its

input port. Using this adjoint field information, it is then possible to calculate the forward gradient without directly calculating the derivative of each parameter. This reduces the resources needed for gradient calculation, as only one adjoint simulation per output and one forward simulation per device are needed. For a device optimization like ours, this reduces the number of simulations per step from tens of thousands, the number of parameters to optimize across, to just three: one forward simulation and two adjoint simulations (one for each output).

2.8 Design space

In inverse design, the design space of a device consists of the possible area that can be designed by a given inverse-design algorithm. Selection of this design space has important effects on the performance and fabrication of the device because it controls how an algorithm can change a device that it designs.

One common method of device optimization, and the one used for our devices, is topological optimization, where the design space is defined by a grid of pixels, each with a parameter value (Fig. 2.8). The optimization algorithm attempts to find the combination of parameter values that provides maximum performance. The grid is then binarized (section 2.11) to create a device layout, with parameter values above a threshold value considered as silicon and other pixels as silica.

Other methods of optimization with different design spaces exist, such as shape

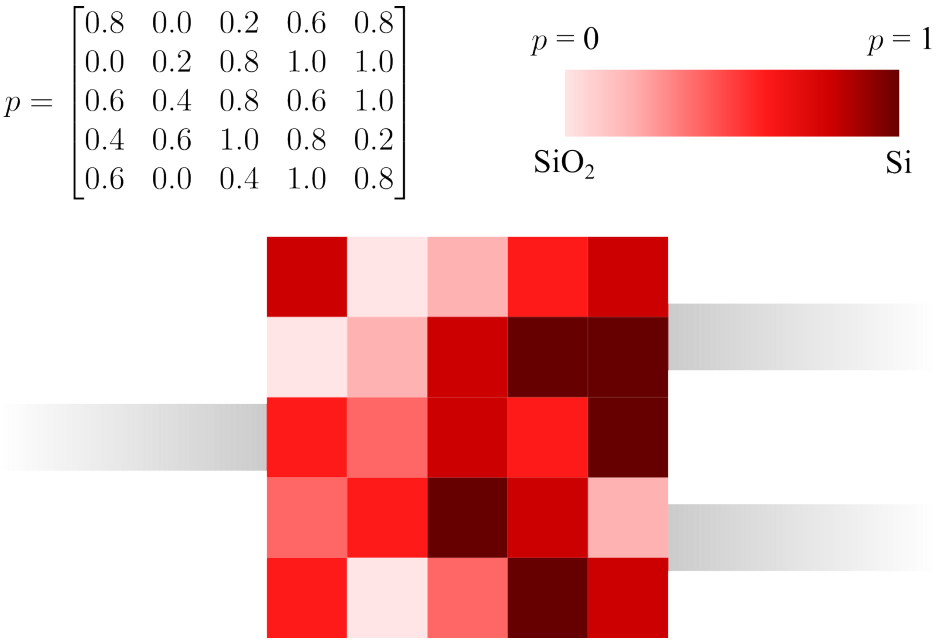


Figure 2.8: Example of the design space of a two-output topological device optimization, such as for a demultiplexer. p represents a parameter value between 0 (silica) and 1 (silicon) for each pixel.

optimization. Here, the device is defined as a single mass of silicon with a shape defined by a set of values that define its boundary (Fig. 2.9). This reduces the occurrence of small features, as only a single piece of silicon is present. However, the smaller design space can reduce the performance of devices, especially those with a complex FOM like demultiplexers. It is thus better suited for further optimization of designs that already have good performance than for designing a fully inverse-designed device [31].

Other, more specialized methods exist for inverse design, such as tuning the size and location of openings in a waveguide [32]. Like with shape optimization, these methods often

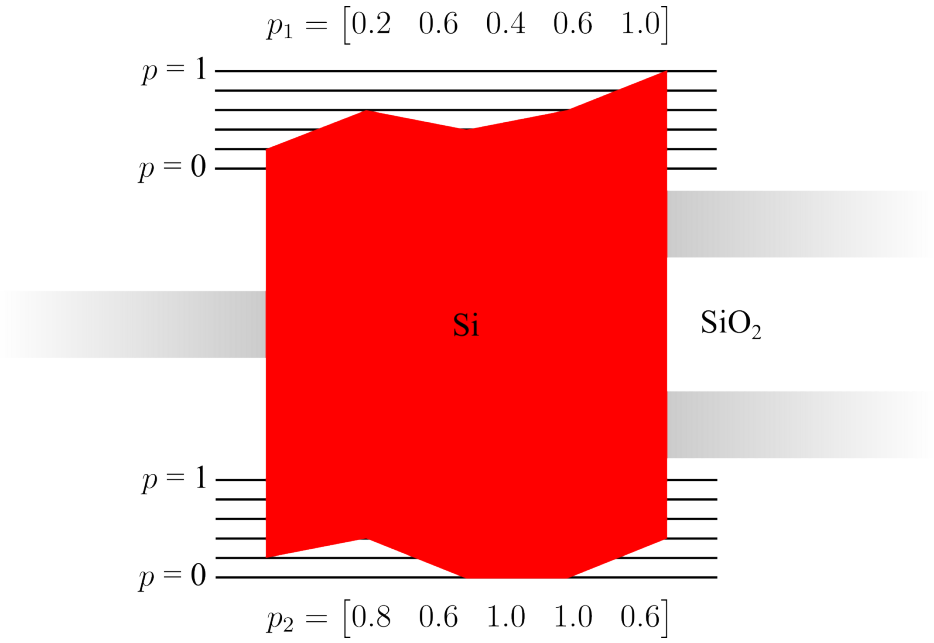


Figure 2.9: Design space of an example shape-based device optimization. p represents the shape of the boundary between silicon and silica.

work best for optimizing devices based on an existing design, due to the size and nature of the design space available to the inverse-design algorithm.

2.9 Optimizers and algorithms

After a design space and FOM are selected, the inverse-design algorithm must use an optimizer to find a design that maximizes the FOM. These optimizers use algorithms that solve the problem largely by using the gradients calculated with the adjoint method, thus acting as the key component of the gradient-descent algorithm used to optimize devices. In our project, we used the L-BFGS-B algorithm from the Python package SciPy. This

algorithm, which optimizes by calculating and updating an approximate Hessian matrix [33], is well suited for photonic devices because of its low memory use and compatibility with the bounded parameters of a permittivity matrix.

Because of the difficulty in finding a single optimal design for a given FOM, the optimization algorithm will usually find a local maximum of the FOM, where other layouts with better performance may exist but are hard to find. We found that the layout of the optimized devices depends on the starting conditions of the parameter matrix, as the implementation of L-BFGS-B acts in a largely deterministic manner (with small variations caused by differences in simulation results and processing between simulations). For the devices we optimized, we started with a uniform permittivity matrix with values halfway between the permittivities of silicon and silica. We found that this allowed for increased performance of the optimized device compared to starting points consisting of uniform silicon or silica. In addition, we also tested a randomized starting matrix, which sometimes increased performance slightly but – given the randomness of the starting point – made the performance of the final device more variable across optimizations.

2.10 Machine learning

Machine learning, in its most general sense, is simply the use of algorithms to find patterns in data [34]. Whereas the gradient descent algorithm we used for inverse design optimizes a figure of merit using a set of device simulations, a machine-learning algorithm directly

learns data patterns from a large training dataset. A key method to implement machine learning in photonics is deep learning, which uses neural networks (large groups of connected nodes) trained on layout and performance data from previous photonic devices. Recent research has used these neural networks to generate photonic devices without the time and resource requirements of gradient-descent algorithms [35]. Our group has used machine-learning algorithms to create the Prefab corrector (section 3.4.3), which uses a neural network to improve the performance of our fabricated devices by learning from a training dataset of fabricated photonic devices [36].

2.11 Binarization of topologically designed devices

For a topologically designed device, the gradient-descent algorithm runs in two steps. First, the material permittivity in each pixel in the design region is optimized within a continuous range between that of the background material (silica) and the waveguide material (silicon). A continuous, or greyscale, range is used to facilitate the gradient-descent algorithm by preventing discontinuities (steps) between the permittivity values of adjacent pixels [37]. Then, the device is binarized in the second design stage so that each pixel has a permittivity value equal to one of the two materials used. This converts the greyscale device optimized in the first design step to one that is fully binarized, which is necessary to achieve a design that can be fabricated in the silicon-on-insulator process.

The binarization step implemented in Lumopt binarizes a greyscale device by using a

Heaviside filter to convert the permittivity value of each pixel to one that is closer to that of one of the two design materials using a β (beta) value[2]; as this value increases from 1 to infinity, the mapping between a device parameter and the permittivity of its corresponding cell transitions from a linear function to a step function. The binarization algorithm increases this beta value in steps, which increases binarization but reduces device performance, then re-optimizes the device to increase its performance while maintaining its higher binarization; this process is then repeated. The final level of binarization (the proportion of pixels that are either silicon or silica) can be indirectly set using the beta parameter controlling the level of device binarization; usually it is set so that binarization is about 99.5% after device optimization. The permittivity values in this binarized parameter matrix generally correspond to silicon or silica, but the pixels on the edges between the two materials can have in-between values.

2.12 Computation time and resource requirements

A significant constraint on the use of inverse design to optimize devices is the time and compute power required. These resource requirements are significant because of the large number of steps (often over 600) that the gradient-descent algorithm takes to optimize a device. While the gradient descent itself can be performed quickly once the FOM and gradients are known for a given set of parameters, the calculation of these values is much

[2]This beta value is unrelated to that used in propagation calculations in Chapter 2.

more computationally difficult. Each step requires at least one forward device simulation (to calculate the FOM) and one adjoint simulation (to calculate the gradients). The computation time varies significantly for topological inverse design based on the device size and size of each pixel, which determines the space of possible devices to search using gradient descent. In addition, a physically larger device with more pixels will take longer to simulate using a photonic simulation tool because of the proportionally larger number of physical equations to be solved in each step.

In addition to device size, the algorithm, or solver, used by the photonic simulation tool affects the time and resources used in optimization. Notably, there is a tradeoff between the use of a 2D FDTD solver, which is faster but slightly less accurate, and a slower 3D FDTD solver that more accurately predicts the performance of a physical device. In this project, we generally started optimizing a device with 2D FDTD solvers but used 3D solvers for the final layouts that were fabricated.

As FDTD simulations are computationally expensive, the time to complete a single optimization step also depends heavily on the computational power available for simulation. For example, a 3D optimization of a 3×5 µm^2 with 20 nm pixels takes about 6 minutes per optimization step on the high-performance computing cluster detailed in section 3.6, while the same optimization on a desktop computer takes about 15 minutes per step.

When optimizing a device in a design space as complex as that of parameter optimization,

it is difficult to find the single most optimal design (i.e., to reach the global maximum of the FOM) within a reasonable time, as performance tends to increase more slowly as the optimization process continues. To balance between increasing device performance (by further searching the space) and optimizing a device quickly, a cutoff value is specified at which the algorithm will cease optimizing the device further. For our optimizations, we stopped optimization in the first phase when the FOM stopped improving by more than 10^{-6} of the difference between the maximum and minimum values of the FOM between iterations. In addition, a separate cutoff stops the optimization after 400 iterations have been calculated in the grayscale stage or 20 iterations have been calculated in a binarization step. We found experimentally that further optimization after this point did not noticeably improve the final device.

Chapter 3

Maximization of simulation and fabrication accuracy

3.1 Output of inverse-design tools

The basis of topological inverse design is the parameter matrix that defines the layout of a photonic device's optimization region; it is this matrix that an inverse-design algorithm seeks to optimize and that it simulates in each optimization step. After the matrix is optimized, though, it must still be translated to a physical device for fabrication.

A complication is that the matrix used by Lumopt, and thus the output of the optimization, is slightly larger than would be expected based on the sizes of the design region and the pixels; specifically, there is one extra pixel in each dimension. Thus, for

example, a 3×3 µm^2 design region with 20-nm pixels would have a permittivity matrix of 151×151 instead of the expected 150×150 pixels. The Lumerical (FSP) file deals with this extra pixel by treating the outside perimeter of pixels as being half as wide as the other pixels. This difference between the sizes is caused by the representation Lumopt uses to define the permittivity of a device, as the permittivity matrix includes both the beginning and end edges of a device as valid locations to define permittivity values. Because the difference in sizes can cause errors in resimulation and fabrication by making the device larger than intended, our tracing methods also reduce the size of the pixels on the perimeter of the device.

3.2 Conversion to a GDS layout

To fabricate an inverse-designed device, the parameter matrix must be converted into the layout of a physical device that can be read during the fabrication process. The most common file format for device layouts is the graphic design system (GDS) format, which is a vector file containing layers that describe locations to be etched by fabrication equipment [15]. There are multiple ways to create a GDS layout from the output of a device optimization. In the case of Lumopt, the most common method is to trace the device using a contour script, which reads the refractive index values of a device from a Lumerical file and creates a GDS layout based on whether each point is higher or lower than a given threshold. However, by default the contour method smooths the GDS output by up to the width of half a pixel

(e.g., when the pixel width/mesh size is 10 nm, the error can be up to 5 nm), which affects performance because of the significant effects of small errors at the edges of inverse-designed devices (Fig. 3.1).

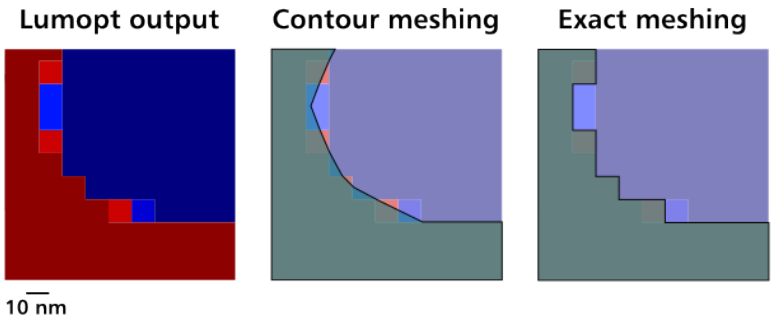


Figure 3.1: Comparison between GDS layouts from different tracing methods, overlaid on a graphical representation of the original parameter matrix. The color of each pixel represents its refractive index.

It is possible to increase the accuracy of a contoured GDS layout by reducing the mesh (pixel) size of a device after it is optimized. This method uses the meshing algorithm in Lumerical to create a denser permittivity matrix based on conformal mesh generation, then uses the same contour algorithm to trace the resulting mesh more precisely. This requires a mesh calculation at the smaller scale, increasing memory use and calculation time. While this reduces the error caused by contour tracing, it does not completely eliminate it, as seen in Fig. 3.2.

To create a GDS file from a parameter matrix without any contour-based smoothing, we used a Python script to import the most recent NPZ file from Lumopt and run

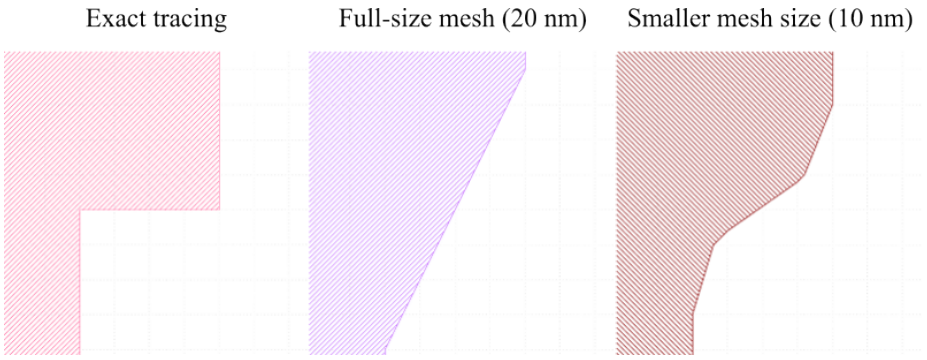


Figure 3.2: Comparison between GDS layouts traced on different mesh sizes. The length of the horizontal line segment is 20 nm.

conversions on it. This script imports the parameter matrix from the output of Lumerical before upconverting it to represent the shape of a device using 1×1 nm^2 pixels, thus multiplying each dimension of the matrix by the original mesh size (in nm). The script then crops the matrix by trimming the outside perimeter of pixels (similar to what Lumerical does, as described above) and adds input and output waveguides to the device layout (to avoid a misalignment between the optimization area and the waveguides). Finally, it creates a GDS file by iterating through the matrix and copying pixels to a GDS cell when the permittivity matrix has a value corresponding to silicon (i.e., above a threshold set between the permittivities of silica and silicon). Note that the output GDS file at this stage can be very large (hundreds of megabytes, even for simple devices).

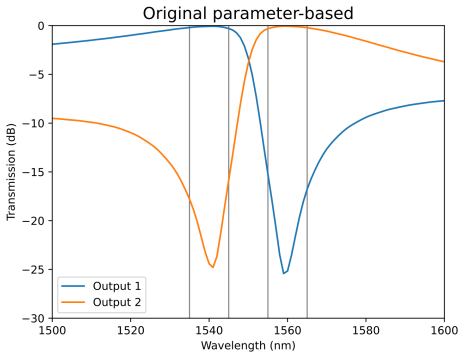
The output of this script is then opened in a GDS layout editor, such as the CAD program Klayout, to simplify it from an array of pixels to a vector file by merging adjacent pixels into larger shapes. By completing this simplification, the file size can be reduced to a much

smaller size (about 1 MB). After being simplified, the output GDS file can be imported into Lumerical to be simulated again. The properties of the imported GDS structure are then modified to ensure that the thickness and material are correct, as the GDS layout does not store material or depth information in a way that can be read by Lumerical, and the device is simulated with the same wavelength settings as used in inverse-design simulations.

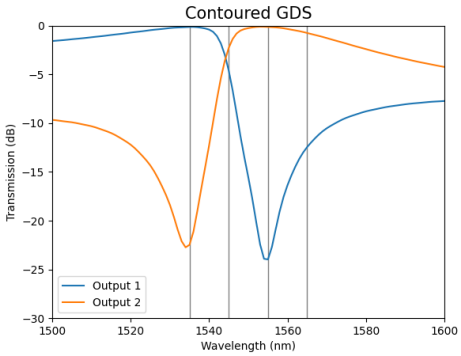
3.3 Meshing and simulated device performance

When a GDS file created from a Lumopt output is imported into Lumerical and simulated, its performance differs significantly from the expected performance at first. Here, the GDS file for an example device (a two-channel, C-band wavelength demultiplexer optimized for maximum transmission across each channel) made by exactly tracing the pixels of the output shows only a moderate improvement over the contoured GDS file, reducing the blue shift from 5.8 nm to 3.5 nm (Fig. 3.3).

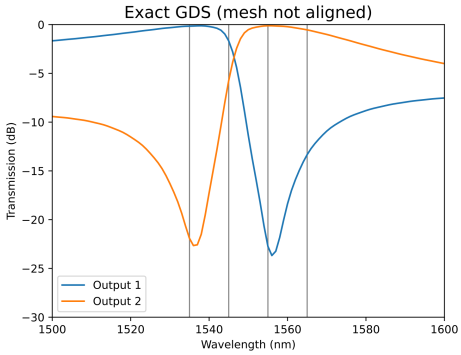
This difference is caused by the meshing methods that Lumerical uses. If the mesh is not aligned, the permittivity of the mesh pixels at the edges of the device will be calculated by averaging across multiple pixels. This causes the parameters of the device as simulated by Lumerical to differ from those of the device as generated by Lumopt or traced into the GDS file. Because this effectively changes the shape of the device, the output spectrum of the simulated device exhibits a blue shift and a reduction in performance. Thus, to avoid a change in simulated performance, it is necessary to align the mesh in Lumerical with the



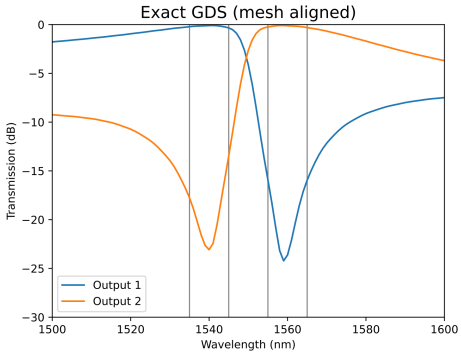
(a) Before GDS conversion (crossover = 1550 nm).



(b) Contour GDS conversion (crossover = 1544.2 nm).



(c) Exact conversion before mesh alignment (crossover = 1546.5 nm).

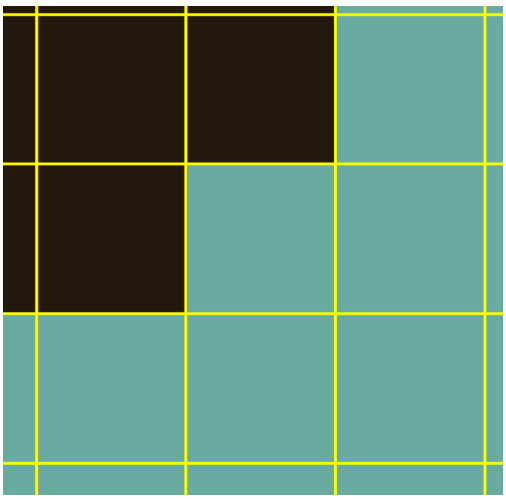


(d) Exact conversion after mesh alignment (crossover = 1551.5 nm).

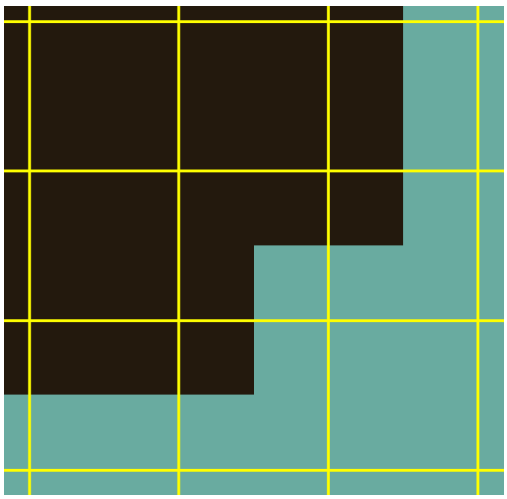
Figure 3.3: Transmission spectrum of example device before and after GDS conversion.

pixels of the GDS structure, as shown in Fig. 3.4.

To align the simulation mesh, the FDTD meshing settings must be changed to be a uniform mesh with maximum mesh step equal to the mesh size used in Lumopt. It may also be necessary to change the minimum mesh step; we set it to be half the maximum mesh step. In addition, we changed the mesh refinement algorithm from conformal to precise volume averaging, as this allows a direct conversion from the material inside each mesh cell



(a) Aligned mesh: each pixel is entirely silicon or silica.



(b) Unaligned mesh: some pixels contain both materials.

Figure 3.4: Mesh cells on the perimeter of a device designed using topological inverse design.

to its corresponding permittivity value [38]. After this, the simulation can be run and the results compared to those of the original Lumopt output. As Fig. 3.3d shows, the blue shift is reduced to less than 0.5 nm, maximizing the accuracy of the simulated device performance within the two target channels.

3.4 Limits of fabrication processes

A device that has been optimized with topological inverse design typically does not resemble a normal device based on a known topology (i.e., a forward-designed device). It tends to have small, non-intuitive features that nonetheless perform well in simulation. While these features allow for high device performance in a small footprint, they also pose issues for

device fabrication, as errors in the fabrication process can cause unpredictable changes in device performance.

As an example of the limits of device fabrication, the topological inverse-designed devices optimized in our project have pixels that are 20 nm square, while the fabrication process we used – the NanoSOI process from Applied Nanotools (ANT) – has a minimum feature size of 60 nm [39]. Thus, it is possible that small features in our devices as optimized would be entirely absent after they have been fabricated. While increasing the pixel size to be as large as the minimum feature size would reduce the effects of fabrication error, device performance would be significantly impacted by the smaller design space for a given device footprint. Even features larger than the minimum feature size are affected by fabrication errors, as sharp corners on device features can be rounded and straight lines roughened by the fabrication process. To reduce the effects of these errors on topological inverse-designed devices, the design process can be modified to enlarge small features, by adjusting the parameters of a low-pass filter or using design for manufacturing, or to directly increase the fabrication resiliency of these features using the Prefab corrector.

3.4.1 Filter size adjustment

During the grayscale-optimization stage of a device in Lumopt, a key method to improve manufacturability of a device is the removal of very small elements that would likely disappear in a fabricated device. To do so, Lumopt uses a top-hat filter with a defined radius, which

acts as a convolutional low-pass filter and "denoises" the device layout by removing features like small holes and single-pixel pieces of silicon [40]. Accordingly, the definition of the filter radius partially determines the minimum feature size in the design area of a device. As we discovered experimentally, this radius controls the general feature sizing but is not an absolute limit on the minimum feature size. For example, many details, such as sharp corners, are as small as 20 nm even on a device with a 100-nm filter radius. In addition, many features are slightly larger than the filter radius but have significant edge detail, which may be lost during fabrication. To fully remove features and edge details below a threshold size, another method – such as design for manufacturing – is needed.

3.4.2 Design for manufacturing in the optimization process

In topological inverse design, a simple method to avoid the small features that are particularly vulnerable to fabrication error is to modify the FOM to penalize these small features. Lumopt, the algorithm we used for inverse design, implements this method by calculating an indicator function that is minimized when all features and voids are larger than a threshold size [41]. This indicator function is then calculated by adding it to the FOM calculations for the gradient-descent algorithm. This causes the optimization function used in Lumopt to search for device layouts that have larger feature sizes, as these are less likely to disappear or be significantly changed in fabrication. By adding the indicator function to the existing FOM for gradient descent, it does so while attempting to minimize

performance losses.

While the design-for-manufacturing process increases the minimum feature size, this optimization increases the time necessary for the design of a device by adding an additional optimization stage after binarization. This process cannot be integrated with the optimization stages before or during binarization, as feature sizes are not yet known until binarization can be completed and the device shape is known. In addition to increasing optimization time and resource needs, performance of the optimized device is often reduced because of the more limited design space of a device after the binarization step; as only two parameter values are allowed (the permittivities of silicon and silica), the gradient becomes harder to solve due to the large steps between the permittivity values of adjacent areas.

3.4.3 The Prefab corrector tool for manufacturing

As an alternative to the design-for-manufacturing algorithm described above, our group has developed a machine-learning algorithm, Prefab, that is intended to minimize the effects of fabrication error on the physical layout of a device. Instead of further optimizing an already binarized device with the goal of removing small features, Prefab modifies, or corrects, the binarized output of the topological binarization step of the optimizer so that the fabricated device closely resembles the output of the optimization algorithm, including small features.

To accurately correct a device layout that will be fabricated, Prefab has been previously trained by fabricating large, complex designs using the same fabrication process as the

device to be fabricated. These fabricated layouts are then imaged with a scanning electron microscope (SEM) and the resulting images traced to show the layout after fabrication. These traced images, with their fabrication errors, are used to train a neural network that learns the relation between a designed GDS layout and its corresponding fabricated equivalent.

Using these trained networks, it is possible both to predict and correct for expected fabrication error before a layout is fabricated. Using a model trained with designed layouts as inputs and fabricated layouts as outputs, a GDS layout can be provided to the model and a corresponding fabricated layout generated [36]. We can then simulate this output layout to characterize a device before it is fabricated and understand how expected fabrication errors will affect device performance, saving the time and resources otherwise necessary to fabricate and measure a device. Conversely, another model can be trained with fabricated device layouts as the input and designed (nominal) layouts as the output. This corrector model can then be used to generate a corrected layout for a given input; when a nominal layout is provided as the input, the resulting output will be corrected so that the fabricated layout more closely matches the nominal layout after the effect of fabrication error [42]. By using the output of this corrector as the device layout to be fabricated, the resulting fabricated layout will have much lower fabrication error than if the nominal layout were to be directly fabricated.

3.5 Performance measurements

After the devices were fabricated, we tested the device layouts with automated test equipment (ATE), which measures the transmission through the device across a wavelength range between 1500 and 1600 nm. This equipment measures the transmission of each device by connecting a tunable laser to the device input through its corresponding on-chip grating coupler (GC), which connects an on-chip waveguide to an external fiber placed slightly above the chip surface. The test equipment then measures the output power with an optical power meter connected through another grating coupler to the device outputs. By scanning through the wavelength range, we obtained a transmission spectrum for each device. We used an ATE located at NRC consisting of a Keysight 8164B measurement system controlling a Keysight 81606A tunable laser module and 81624B InGaAs photodetector.

As each test circuit contains a pair of vertical grating couplers to allow light to enter and leave the device, along with a silicon waveguide, we remove the effect of these components from the measured transmission spectra from the test equipment. These components do not have a flat transmission spectrum over the full wavelength range, so they significantly change the measured transmission spectra. We corrected for these effects by fabricating and measuring a loopback circuit, which consists solely of a waveguide connecting two grating couplers, on the same die as the circuits to measure the effect of these components. By subtracting this effect (in dB) from the transmission spectrum for each measured test circuit, we obtained a transmission spectrum for each device on the layout, as shown in Fig. 3.5.

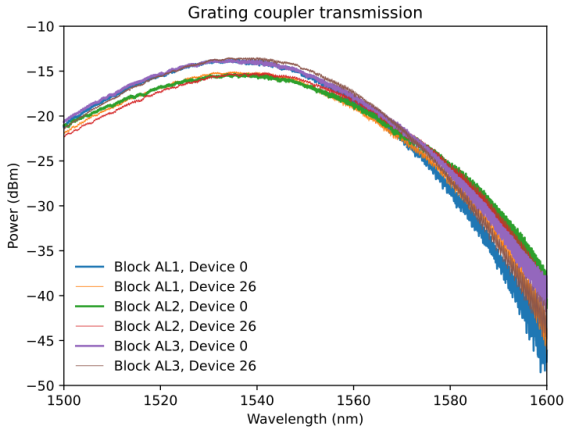


Figure 3.5: Transmission spectrum of a grating coupler loopback circuit.

As the spectrum for the grating couplers shows, transmission is maximized at about -15 dB near 1540 nm and decreases significantly above this point, with transmission falling below -20 dB above 1560 nm. While device transmission measurements can be corrected for the effects of grating coupler transmission, as they are for the device measurements shown below, the precision of the transmission measurements is reduced above 1560 nm because the lower transmission values approach the noise level of our test equipment. Device measurements above 1580 nm become dominated by noise, which causes measurement artifacts like insertion loss above 0 dB at certain wavelengths. For more accurate measurements within this range, it may be necessary to use different grating couplers (centered at a higher wavelength) or to use edge couplers, which have lower coupling loss and a wider spectral range but are more challenging to measure [15].

In addition to the measurements of the transmission spectrum, the devices are

characterized through SEM images taken by the fabrication service after the devices are fabricated. These images show the fabricated device with a resolution of about 4.5 nm, which is sufficiently detailed to show small details and, when compared to the original GDS layout, show the physical effect of fabrication error. Due to the cost and resource requirements of SEM images, only selected devices were imaged, particularly the nominal and Prefab-corrected (threshold = 0.5) devices.

3.6 Inverse-designed wavelength multiplexers: design methodology

The devices we designed are optimized using the topological inverse-design algorithm in Lumopt on a high-performance computing (HPC) cluster located within and operated by NRC. The optimization process used four computers running calculations in parallel, each with two Intel Xeon 6130 CPUs (32 cores each). While the physical computers contain graphics processing units (GPUs) in addition to CPUs, they were not used in the optimization or simulation of this device. The optimization process used Lumopt running on Ansys Lumerical FDTD with HPC extensions. To allocate resources on the HPC cluster, the Slurm workload manager is used. Device optimizations are split into sequential optimization jobs of 12 hours each; each job begins with the partially optimized result of the previous workload. The splitting of the optimization into individual jobs, which is

required due to a limit set on the computing cluster, may have increased the total optimization time because the partially optimized result that is transferred between jobs does not contain information related to the state of the gradient-descent optimizer and the binarization algorithm. Thus, each new job restarts the gradient-descent step first and only continues the binarization step when the gradient-descent step reaches a cut-off point. If the 12-hour time limit could be extended, it would be possible to significantly reduce the amount of time spent in the binarization process by reducing the number of optimizations at each binarization level.

After optimization is complete, we simulate the output of the resulting device in Lumerical and create a fully binarized GDS layout of the optimized device; we set the permittivity threshold to $\epsilon = 6.25$, close to halfway between the permittivities of silicon ($\epsilon = 12.08$) and silica ($\epsilon = 2.08$). We then import this GDS layout into another Lumerical simulation to ensure the binarization process did not significantly affect the performance of the device.

We then use our Prefab corrector (section 3.4.3) to correct the GDS layout for fabrication. This ensures that the device as fabricated, with its inevitable fabrication errors, will match the intended GDS layout as closely as possible, increasing device performance. We also predict the fabricated output of both the corrected layout and the original, uncorrected GDS layout to predict the effect of Prefab's correction on a fabricated device. For comparison with Prefab, we also import the parameters from the inverse design back into Lumopt and performed the design-for-manufacturing optimization of section 3.4.2. As with the Prefab-

corrected and original devices, we then trace the resulting parameters into a GDS file that can be fabricated.

For the device fabrication, we create a device layout using the software package Klayout. This layout contains the inverse-designed devices, grating couplers for connection to external test equipment, and 500×220 nm^2 waveguides (separated by 1000 nm) connecting the devices to the grating couplers. As each demultiplexer has one input and two output ports, each device under test (DUT) is connected to three vertical grating couplers (Fig. 3.6). This layout is then fabricated on a 220-nm silicon-on-insulator die with the NanoSOI process at Applied Nanotools (ANT), an external device-fabrication company. After fabrication, SEM images of the layout are taken and the devices on the layout are measured using the automated test equipment setup.

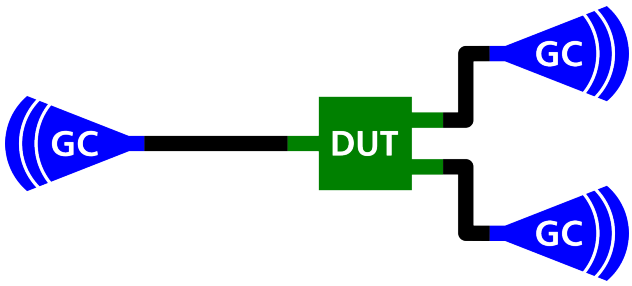


Figure 3.6: Schematic layout of each photonic circuit as fabricated. The layout shown is not to scale.

Chapter 4

Performance of inverse-designed wavelength demultiplexers

4.1 Device 1: 3×3 µm^2 wavelength demultiplexer

We designed our first device as a two-channel C-band wavelength demultiplexer with a 3×3 µm^2 topological design region, one input waveguide, and two output waveguides, as shown in Fig. 4.1. The design region has pixels that are each 20 nm wide, which means that the region theoretically contains 150×150 square pixels. The FOM of this demultiplexer measures the mode matching between the input waveguide and one of the two output waveguides for the wavelength range of each channel. For channel 1 this range is 1535-1545 nm, and for channel 2 this range is 1555-1565 nm. To ensure high mode matching throughout each

channel range, we measured mode matching (from 0 to 1) at 1-nm intervals within each channel; the FOM is then taken as the mean of these values. The optimization of this device on the cluster took a total of 84 hours over nine sequential optimization jobs.

Figure 4.1: Layout of the optimized device, with the topological optimization region in red and the input/output waveguides in gray.

4.1.1 Results

Based on the simulations conducted as part of the inverse-design process, the demultiplexer is expected to have the transmission spectrum shown in Fig. 4.2, with the crossover wavelength (the wavelength at which the transmission from both outputs is identical) at 1550 nm. The insertion loss at the center of the lower-performing channel is 0.46 dB and the crosstalk at the center of the lower-performing channel is 12.00 dB.

When this device is fabricated without any correction, performance is affected by a wavelength shift in the transmission spectrum. Between the six nominal (uncorrected) versions fabricated and measured at NRC, this shift, measured at the crossover point, ranges from 19.0 nm to 21.7 nm. As this is much larger than the 10-nm channel width, the

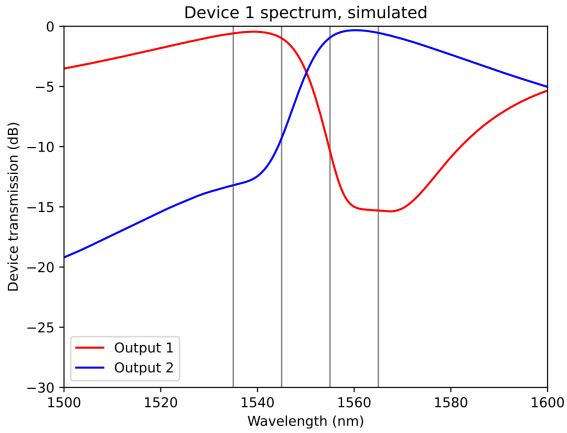


Figure 4.2: Simulated transmission spectrum for the optimized design of the 3×3 µm^2 wavelength demultiplexer.

devices function poorly as demultiplexers in the defined channels, as seen in Fig. 4.3. Note the noise and the ringing, especially above 1580 nm; this is caused by the imperfect transmission spectrum of the grating coupler as shown in Fig. 3.5. Because the transmission of the grating coupler is particularly low at higher wavelengths, the combined loss from the grating couplers and the demultiplexer reaches -40 dB, causing the signal level to approach the noise floor of the optical power meters we used and making the noise more prominent when the loss from the grating couplers is subtracted from the combined transmission.

In contrast, the devices that are corrected with Prefab have a much smaller wavelength shift when the etching threshold is set to 0.25 (over-etching) to account for under-etching by the fabrication process. When the device is corrected in this way, the fabricated devices have a much smaller wavelength shift of 2.4 nm to 4.5 nm across the six devices measured

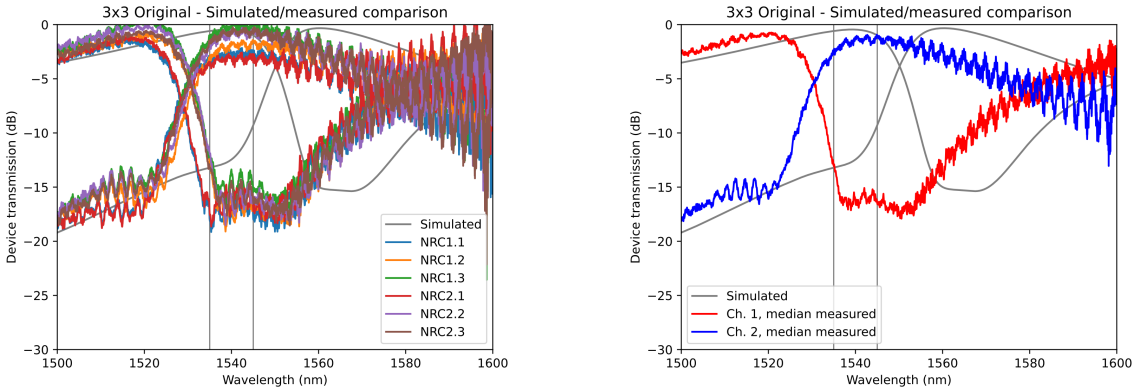


Figure 4.3: Transmission spectra of fabricated nominal versions of the 3×3 µm^2 wavelength demultiplexer (left) and median values across versions (right).

at NRC. The resulting devices thus perform much better as demultiplexers when channel wavelengths are fixed, as shown in Fig. 4.4.

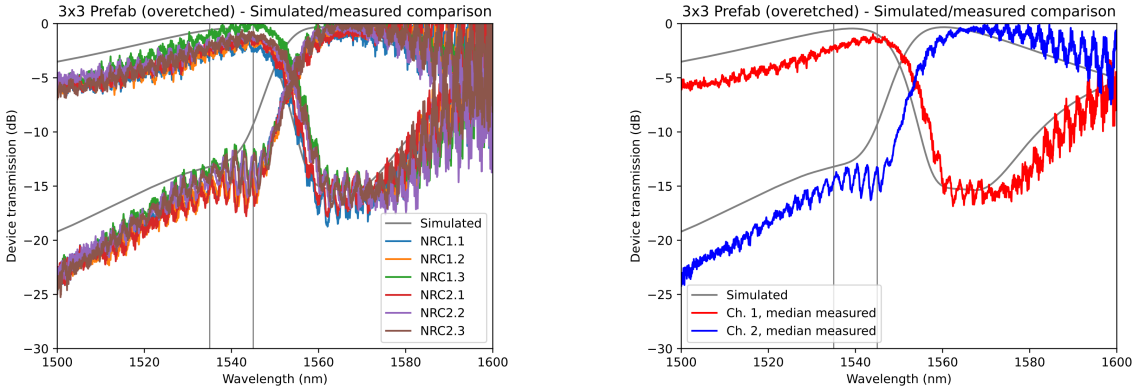


Figure 4.4: Transmission spectra of Prefab-corrected (over-etched) versions of the 3×3 µm^2 wavelength demultiplexer (left) and median values across versions (right).

The importance of the over-etching threshold parameter can be seen in the corrected fabricated devices where the threshold is set to the default 0.5. These devices, which have

been corrected based on a fabrication process with an average amount of over-etching/under-etching error (Fig. 4.5), have a similar magnitude of wavelength shift as the nominal devices, between 23.2 nm and 24.2 nm.

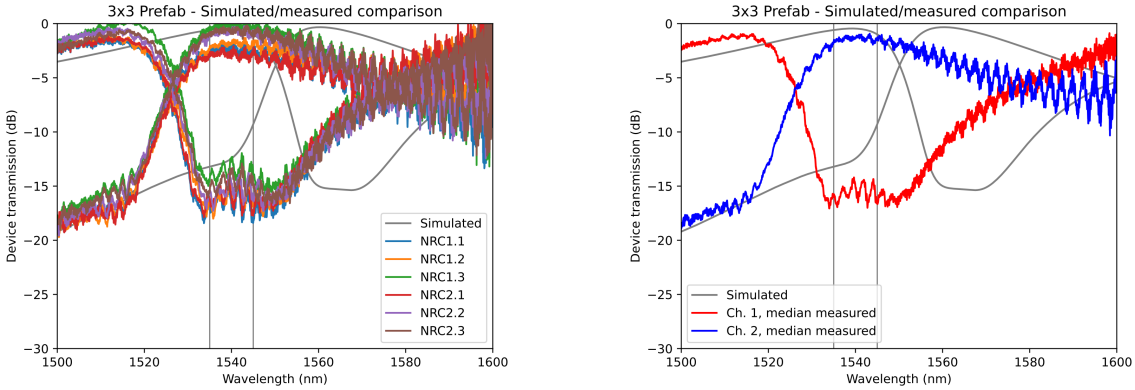


Figure 4.5: Transmission spectra of Prefab-corrected versions of the 3×3 µm^2 wavelength demultiplexer with the default over-etching threshold (left) and median values across versions (right).

While the magnitude of the wavelength shift for these devices is significant, it is notable that the variation between the wavelength shifts between the six fabricated devices is much smaller than for the nominal amount, with the difference between the largest and smallest shift among six fabricated versions falling from 2.7 nm to 1.0 nm. This suggests that Prefab may increase consistency between fabrications of the same layout by correcting for small features. These may otherwise be present in some fabricated devices but not others in the same run due to the inherent variation caused by the fabrication method, causing the variation between devices in the same fabrication run.

The device variant that has been corrected with Lumopt's design-for-manufacturing

algorithm (section 3.4.2) has a similar shift as the nominal and Prefab-corrected (threshold = 0.5) devices, with wavelength shifts between 20.3 nm and 22.8 nm. The transmission spectra of the fabricated devices are shown in Fig. 4.6.

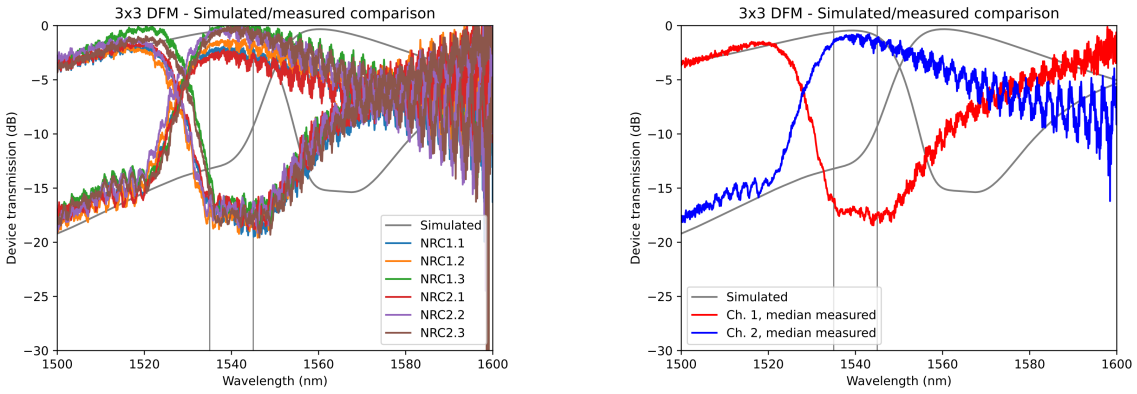


Figure 4.6: Transmission spectra of Lumopt-corrected fabricated versions of the 3×3 µm^2 demultiplexer (left) and median values across versions (right).

Notably, unlike Prefab correction, the Lumopt design-for-manufacturing correction cannot be tuned by a threshold value to account for over-etching or under-etching in the fabrication process because it only seeks to reduce the number of small features on a device layout. In contrast, the machine learning on which Prefab is based considers the variation within its training data and allows a corrected device to be more over-etched or under-etched based on these variations.

A notable difference between the wavelength shifts of the above devices is the direction of the shift. Whereas the corrected device with threshold 0.5 – like the nominal device – experienced a blue shift after fabrication (a shift toward smaller wavelengths), the device

with threshold 0.25 experienced a red shift, or a negative blue shift. This suggests that further tuning of the over-etching parameter may be useful, as it may be possible to select a value between the two thresholds that further reduces the magnitude of the wavelength shift. The similar performance between the two fabrication runs we measured, despite being fabricated months apart, suggests that the ideal threshold value may vary slowly over time (see the next section).

4.2 Device 2: Inverse-designed 3×5 µm^2 wavelength demultiplexer with reduced out-of-band transmission

In this section, we discuss the second device we designed and the changes to the optimization process that allow us to design a device with a more complex FOM. Because the first device we designed only seeks to maximize transmission within the wavelength channels, transmission outside the channels (e.g., at 1500 and 1600 nm) is similar in magnitude to transmission within the channels. This prevents more than two channels from being used because they would experience undesired transmission to one or both of the demultiplexer outputs, requiring additional filtering to ensure that each output contains only a single channel. Thus, our goal for the second device is to limit light transmission at the two outputs solely to that within defined wavelength channels; input light that is not

within the two defined channels should not be transmitted to either output waveguide. This device is thus designed to have two distinct channels with much narrower channel widths than those of the first device. This is similar to the transmission spectra of MZI-based forward-designed demultiplexers, but with the added advantage of flatter channel transmission and reduced transmission outside the channel range. These advantages would further reduce the footprint necessary to implement a WDM receiver, as the demultiplexer outputs could be used without further filtering to remove out-of-band (OOB) transmission components.

Like with the first device, we designed this device using topological inverse design with Lumopt for gradient descent and Lumerical for forward and adjoint photonic simulations. We define this device with a larger design region of 3×5 μm^2, making the device longer but keeping the same width (i.e., the waveguides connect to the design region along its shorter sides). We use the same 20-nm pixel length, leading to a 151×251-pixel design region. In addition, the filter radius (100 nm) and device thickness (220 nm) are the same as for the first device; the latter property allows both devices to be fabricated on the same die in a single production run.

To design a device with suppressed out-of-band transmission, we define a FOM to minimize transmission across wavelengths between 1500 and 1600 nm, except within the intended transmission bands within each channel (where transmission is instead maximized). Specifically, the FOM measures the mode matching between the input

waveguide and the output waveguides at the 16 wavelengths shown in Figure 4.7.

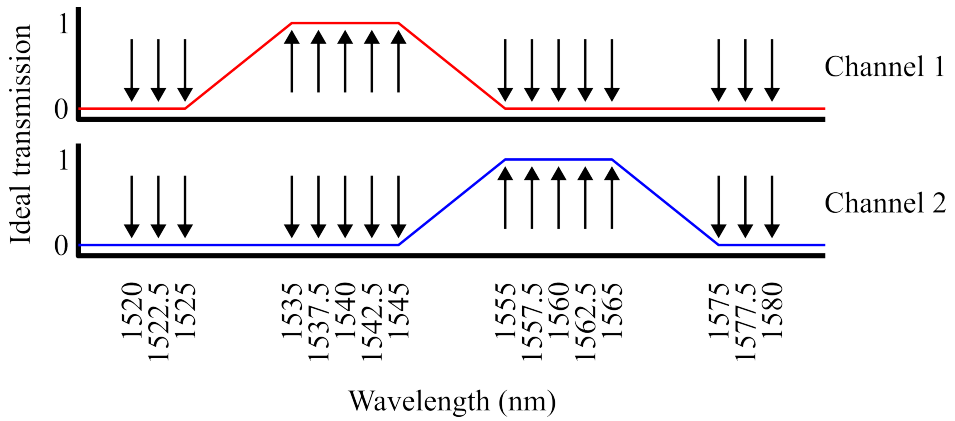


Figure 4.7: Wavelengths (shown as arrows) used in optimizing devices with suppressed out-of-band transmission.

As with the first, 3×3 µm^2 device, the optimization of the 3×5 µm^2 device took place on an HPC cluster in increments of 12 hours. For this device, the optimization used 14 sessions, with the last one being shorter (about half an hour long) than the others due to the device reaching its final binarization. Together, the optimization took about 157 hours on the cluster.

4.2.1 Results

Simulations of the device (Fig. 4.8) show the ideal transmission spectrum of the demultiplexer without fabrication error. Unlike for the 3×3 µm^2 device, the ideal 3×5 µm^2 device has a crossover point between the two output channels at 1548.8 nm due to the difference between the widths of the two transmission channels (section 4.4).

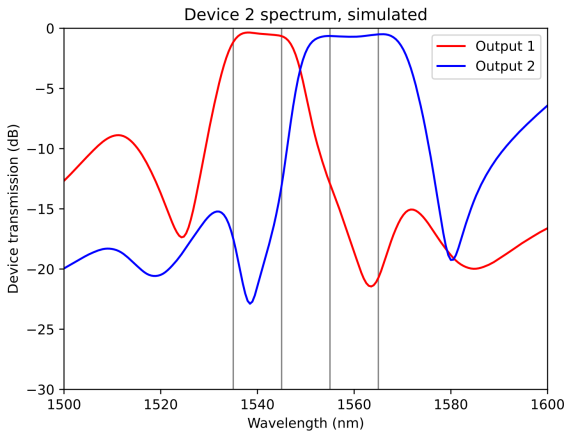


Figure 4.8: Simulated transmission spectrum for the optimized design of the 3×5 µm^2 wavelength demultiplexer.

In simulation, the worst-case insertion loss at the channel centers is 0.70 dB and the worst-case crosstalk is 17.85 dB. When this device is fabricated without any correction, a wavelength shift similar to that of the first device appears. Across the six fabricated versions of this device that were not corrected, this shift – measured at the crossover wavelength where transmission from the two channels was equal – appeared as a blue shift of 20.8 to 24.4 nm. As shown in Fig. 4.9 below, this leads to low demultiplexer performance because the output channels have very low transmission within the desired wavelength ranges.

This blue shift is essentially unchanged when using Lumopt's default design-for-manufacturing method, with blue shifts between 21.5 and 24.2 nm. In addition, the insertion loss becomes larger for channel 2, even when accounting for the blue shift, as seen in Fig. 4.10. Similarly, the Prefab-corrected device has a large shift when the

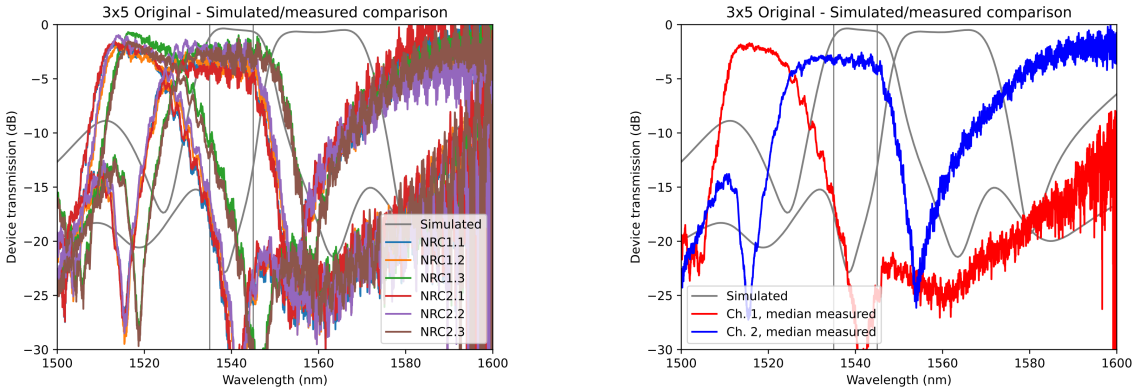


Figure 4.9: Transmission spectra of nominal fabricated versions of the 3×5 µm^2 wavelength demultiplexer (left) and median values across versions (right).

over-etching threshold is set to 0.5. Because this corresponds to the error experienced in a median fabrication run, this suggests that the fabrication runs containing this device differ from the median run. These fabricated, Prefab-corrected devices have a blue shift of between 19.9 and 24.3 nm (Fig. 4.11). However, when the threshold value of the Prefab correction is changed to 0.25, anticipating that the fabrication process will under-etch a layout (and over-etching the layout accordingly), the magnitude of the wavelength shift is greatly reduced to between 1.2 and 3.8 nm (Fig. 4.12). Like with the first device, this shift becomes a red shift (or a negative blue shift), suggesting that a value slightly higher than 0.25 may have further reduced the magnitude of the shift. Even with the small shift here, though, the device has significantly better performance as a demultiplexer relative to the uncorrected (nominal) device. The similarity between the performance of this corrected device and the previous corrected device shows a key benefit of Prefab correction, even

with tuning of the overetching parameter: while this tuning may have to be undertaken over multiple fabrication runs, it generalizes to multiple devices, whereas tuning by changing the channel wavelengths optimized in the inverse-design process must be done for each device individually.

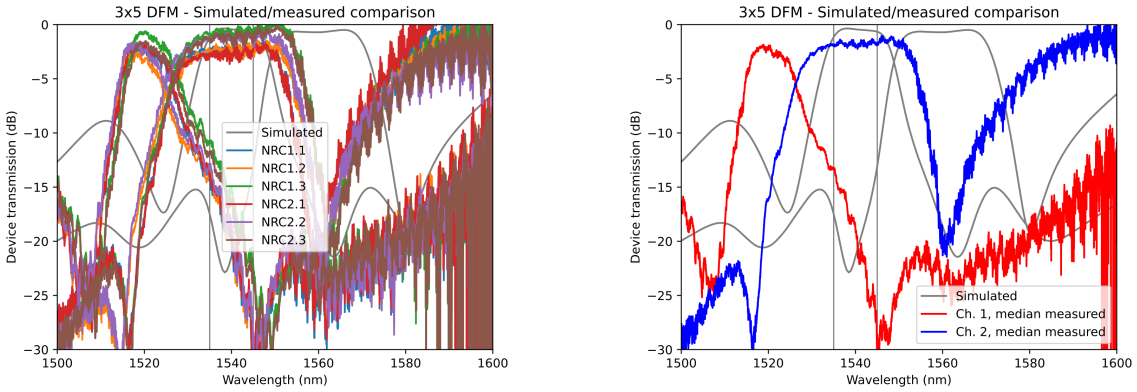


Figure 4.10: Transmission spectra of Lumopt-corrected fabricated versions of the 3×5 µm^2 wavelength demultiplexer (left) and median values across versions (right).

4.3 Device 3: Inverse-designed 3×3 µm^2 wavelength demultiplexer with reduced out-of-band transmission

Here, we explore the effect of the size of a device, and by extension the size of its design space, on that device's performance. We analyze the relation between size and performance on a

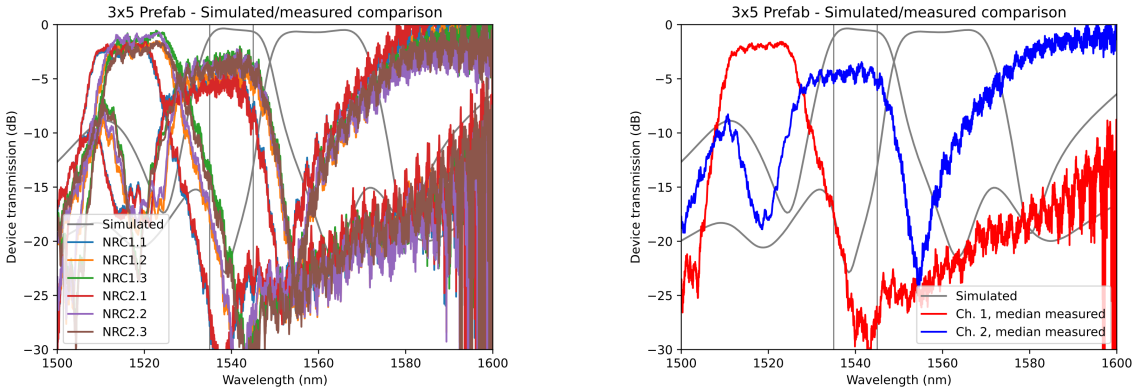


Figure 4.11: Transmission spectra of Prefab-corrected versions of the 3×5 µm^2 wavelength demultiplexer with the default over-etching threshold (left) and median values across versions (right).

device with demanding design requirements by optimizing a device similar to the previous demultiplexer but in a significantly smaller size.

This device was optimized using a process similar to that of the previous 3×5 µm^2 device. However, the optimization region is 3×3 µm^2, or 151×151 pixels, like the first device. Otherwise, the FOM is identical to that of the previous, 3×5 µm^2 demultiplexer, and the optimization, simulation, and fabrication methodologies also matched those of the previous devices. All three devices were fabricated and measured together, reducing the likely variation from different calibrations of the fabrication and measurement tools. In addition, the use of a single die reduces the variation in waveguide thickness caused by variations between silicon wafers.

On the same computing cluster used for the optimization of the previous two devices, we optimized the 3×3 device with suppressed out-of-band transmission over the course of

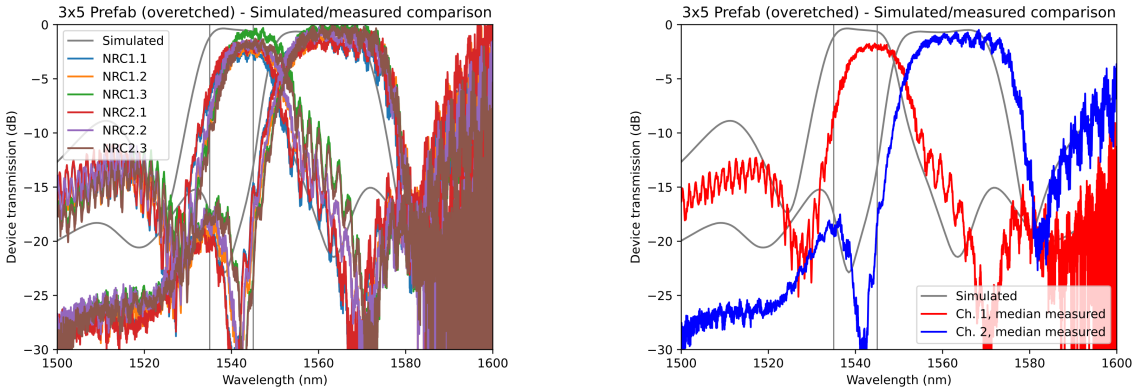


Figure 4.12: Transmission spectra of Prefab-corrected, over-etched versions of the 3×5 µm^2 wavelength demultiplexer (left) and median values across versions (right).

60 hours (five optimization sessions of 12 hours each). This was significantly less time than necessary for the 3×5 µm^2 demultiplexer with the same FOM or the 3×3 µm^2 demultiplexer with the simpler FOM. This is likely related to the nature of the gradient-descent algorithm, which requires fewer calculations to find an optimal set of parameters when fewer parameters (pixels) must be optimized.

4.3.1 Results

In the simulated version of this device (Fig. 4.13), its transmission spectrum is similar to that of the 3×5 µm^2 device, with a slightly worse insertion loss (at the center of the lowest-performing channel, 0.74 dB) but a much worse crosstalk (up to 11.74 dB). However, this complicated FOM does notably reduce out-of-band transmission compared to the first device (which also has a 3×3 µm^2 footprint), at the cost of a lower crosstalk ratio and higher

insertion loss. As with the 3×5 µm^2 device, the slight difference in the effective width of the channels means that the crossover point is not exactly between the two defined channels; here, it is at 1549.8 nm.

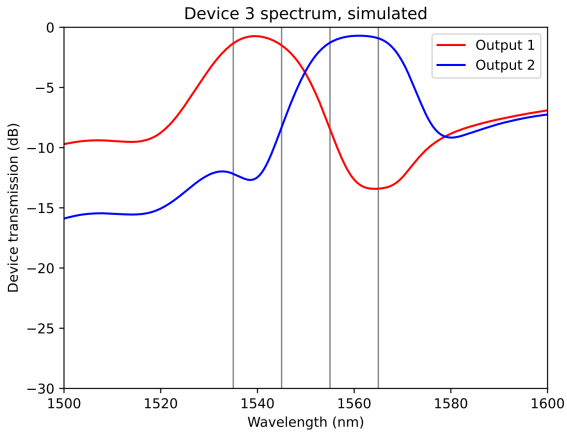


Figure 4.13: Simulated transmission spectrum of the optimized design of the 3×3 µm^2 demultiplexer with OOB suppression.

Consistent with the two previous devices that were fabricated in the same production runs, we again see a significant blue shift (19.4 nm to 21.2 nm) in the fabricated, uncorrected device (Fig. 4.14). As with the previous devices, this significant shift makes the device ineffective as a demultiplexer for the given channel ranges. Similarly, the device variants that have been corrected with Lumopt's design-for-manufacturing algorithm (Fig. 4.15) have a blue shift of 17.4 nm to 20.5 nm.

As expected based on the performance of the previous two devices, the performance of the Prefab-corrected variations differs based on the threshold used to set the amount of over-

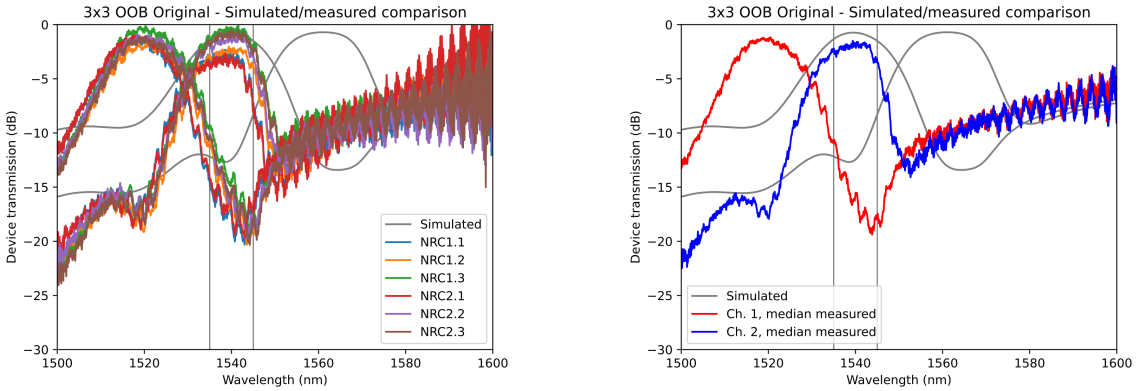


Figure 4.14: Transmission spectra of nominal fabricated versions of the 3×3 µm^2 demultiplexer with OOB suppression (left) and median values across versions (right).

or under-etching. When the threshold is set to 0.5, the device blue shift of 21.3 to 22.7 nm is similar to that of the nominal and Lumopt-corrected devices, as seen in Fig. 4.16. However, when the threshold is set to 0.25 (corresponding to fabrication errors tending toward under-etching), the wavelength shift becomes a red shift of 3.3 to 3.7 nm, making the demultiplexer much more useful within the target wavelength ranges (Fig. 4.17).

The comparison between this device and the previous one, which was optimized with the same FOM but a larger footprint, allows us to examine the effect of device size on its performance. As the transmission charts show, insertion loss is only slightly higher in a device with a 40% smaller footprint and a correspondingly smaller number of parameters to control. This small difference between the two can be explained by this smaller number of parameters, which generally results in lower performance, being partially counteracted by the smaller path through which light travels (and correspondingly lower energy loss from

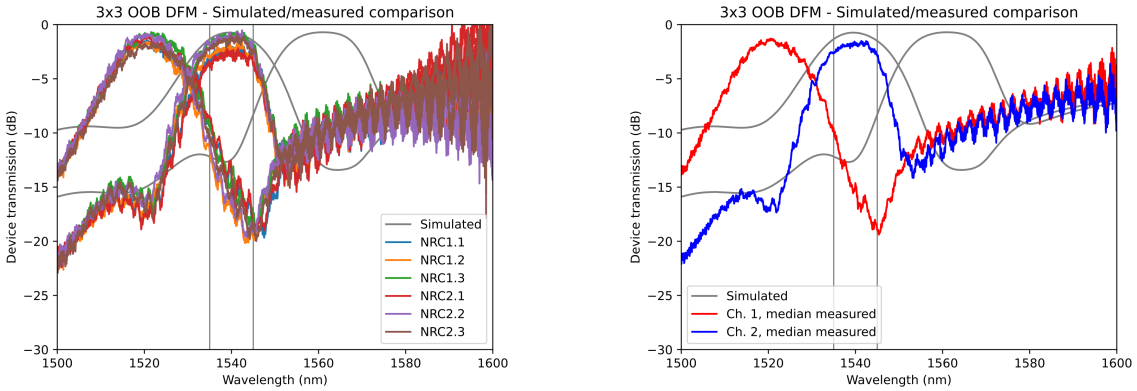


Figure 4.15: Transmission spectra of Lumopt-corrected fabricated versions of the 3×3 µm^2 demultiplexer with OOB suppression (left) and median values across versions (right).

attenuation).

While the insertion loss of this device is slightly worse, the crosstalk ratio of this smaller device is significantly worsened: this ratio falls to 11.74 dB, compared to 17.85 dB for the simulated 3×5 µm^2 device (a 34% reduction in performance). This reduced performance shows that crosstalk increases as the footprint of a device decreases, even when the FOM does not change. Thus, device applications that require a certain level of performance will require a larger inverse-designed device to maintain such a level of insertion-loss and crosstalk performance.

4.4 Device performance analysis

For all three fabricated devices, we saw a significant difference between the transmission of the simulated device and most versions of the fabricated, measured device (table 4.1). This

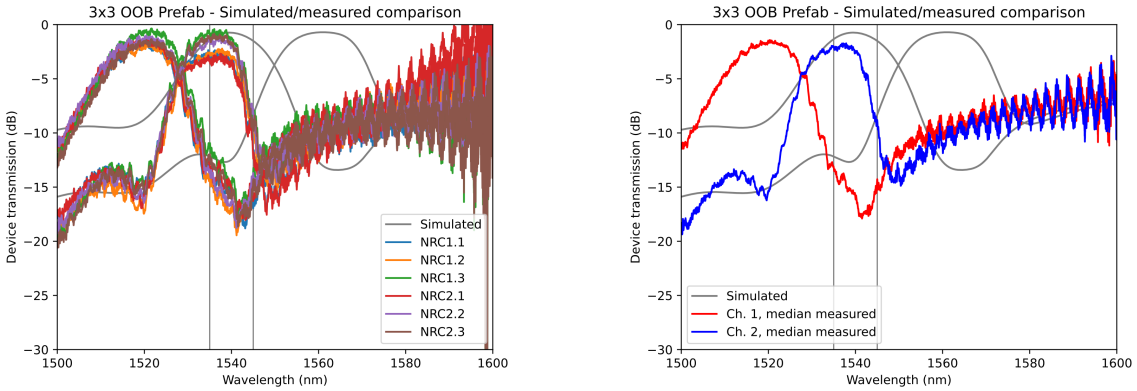


Figure 4.16: Transmission spectra of Prefab-corrected (non-over-etched) versions of the 3×3 µm^2 demultiplexer with OOB suppression (left) and median values across versions (right).

deviation, in the form of a blue shift of the transmission spectrum, appears with a similar magnitude in both the standard Prefab-corrected fabricated device (when the binarization threshold controlling over-etching is set to 0.5) and the one designed with Lumopt's design-for-manufacturing optimizations.

To explore potential causes of this spectral shift, we obtained measurements of the thickness of the silicon layer in our devices. This thickness, nominally 220 nm, has a significant effect on device performance by controlling the confinement of light in a rectangular waveguide (as discussed in the Background section). In addition, this thickness is defined by the thickness of the silicon wafer used in the NanoSOI fabrication process, so it cannot be predicted or corrected for in the same way as fabrication error caused by the process itself unless the thickness of the wafer is known before device design.

We found the average thickness of the silicon layer to be 217.4 nm for the first fabrication

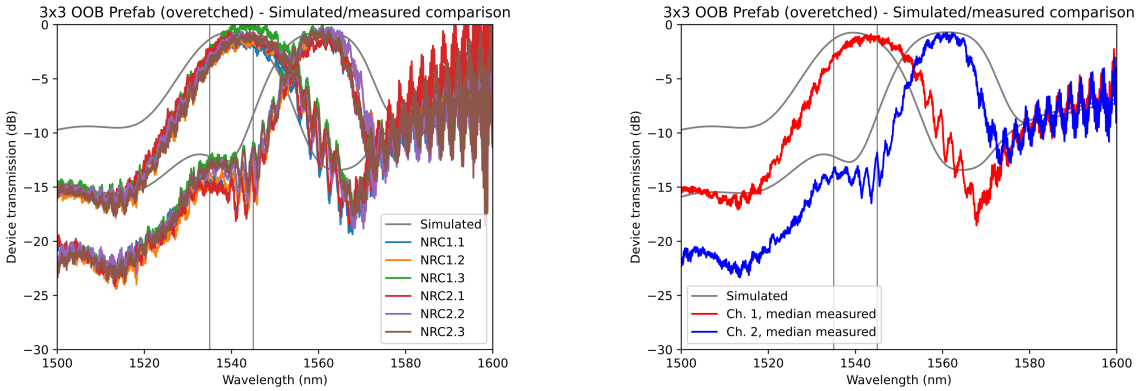


Figure 4.17: Transmission spectra of Prefab-corrected, over-etched versions of the 3×3 µm^2 demultiplexer with OOB suppression (left) and median values across versions (right).

run and 218.9 nm for the second based on measurements taken by ANT, a difference of up to 2.6 nm from the nominal thickness. We then simulated a device with the same layout as the nominal device but with a thickness of 217 nm instead of the nominal 220-nm thickness. As shown in the transmission spectrum in Fig. 4.18, the crossover wavelength shifts from 1550 nm to 1545.8 nm, partially explaining the shift we observed but suggesting that the remaining shift is caused by other effects, such as fabrication error.

To characterize this fabrication error, we examined an image of the fabricated nominal device taken with an SEM. By overlaying the intended design from the GDS layout on the image in yellow, the fabrication error can be seen in the difference between the two in Fig. 4.19. Based on the SEM image, it can be seen that the fabricated device has rounding in its corners and smoother edges. In addition, there is size variation that causes the fabricated device to be slightly smaller than intended and bumps on the edges of the device that were not

Table 4.1: Measured performance of each demultiplexer. Each value is the median of six measurements of fabricated devices.

Device	Version	Channel performance[1] IL (dB)	XT (dB)	Blue shift (nm)[2]
3×3 µm^2	Simulated	0.46	12.00	0.00
	Nominal	1.42	15.13	19.61
	DFM	1.84	16.71	21.71
	Prefab	1.50	14.70	23.62
	Prefab (over-etched)	1.48	14.98	-3.50
3×5 µm^2, OOB suppression	Simulated	0.70	17.85	0.00
	Nominal	3.20	25.31	23.90
	DFM	2.61	23.25	22.30
	Prefab	3.96	21.55	20.70
	Prefab (over-etched)	1.96	27.08	-2.60
3×3 µm^2, OOB suppression	Simulated	0.74	11.74	0.00
	Nominal	1.85	16.69	20.10
	DFM	2.12	15.13	18.60
	Prefab	2.29	13.41	21.90
	Prefab (over-etched)	1.85	14.87	-3.60

[1] Measured at the center of the lowest-performance wavelength channel, adjusted for the blue shift caused by fabrication.

[2] The shift in the crossover wavelength of the device, with positive values corresponding to lower crossover wavelengths.

part of the intended layout. These changes likely contributed to the variation seen between the simulated and fabricated (nominal) devices. This exists for the nominal fabricated device and the corrected ones, except for the one designed with Prefab correction and over-etching. Because this is similar across both fabrication runs, it suggests that the fabrication process exhibited an under-etching error that was counteracted by the over-etching of the Prefab correction.

In the two devices with out-of-band transmission suppression, the more complicated FOM

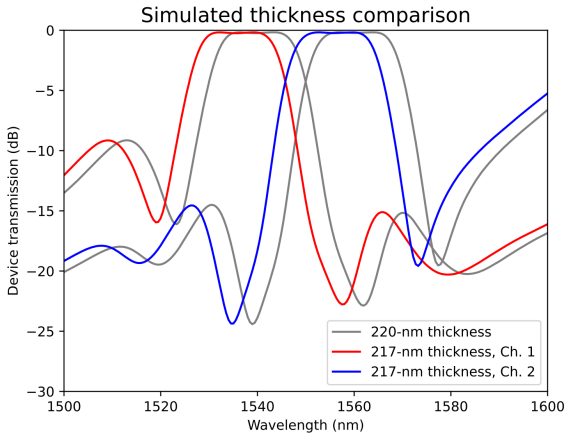


Figure 4.18: Simulated transmission spectrum for the 3×5 µm^2 wavelength demultiplexer with a thickness of 217 nm.

with distinct transmission channels is also affected by the effects of fabrication. Throughout the fabricated transmission spectra, the width of channel 1 (nominally 1535-1545 nm) is narrower than that of channel 2 (1555-1565 nm). While most visible in the device corrected with Lumopt's design-for-manufacturing optimization, where channel 1 has no flat top at all, it is present in both the nominal and corrected devices.

This variation in channel width may partly be caused by the choice of the FOM for the optimization algorithm, which optimizes for transmission within the nominal wavelength ranges corresponding to the two channels. As the transmission spectrum of the simulated devices show, both channels have high transmission within these 10-nm ranges, even though the simulated channel widths vary between the two outputs. While the widths of the channels are constrained by the 10-nm wavelength range in which transmission is maximized and the 30-nm range between wavelengths where transmission is minimized, the variation in channel

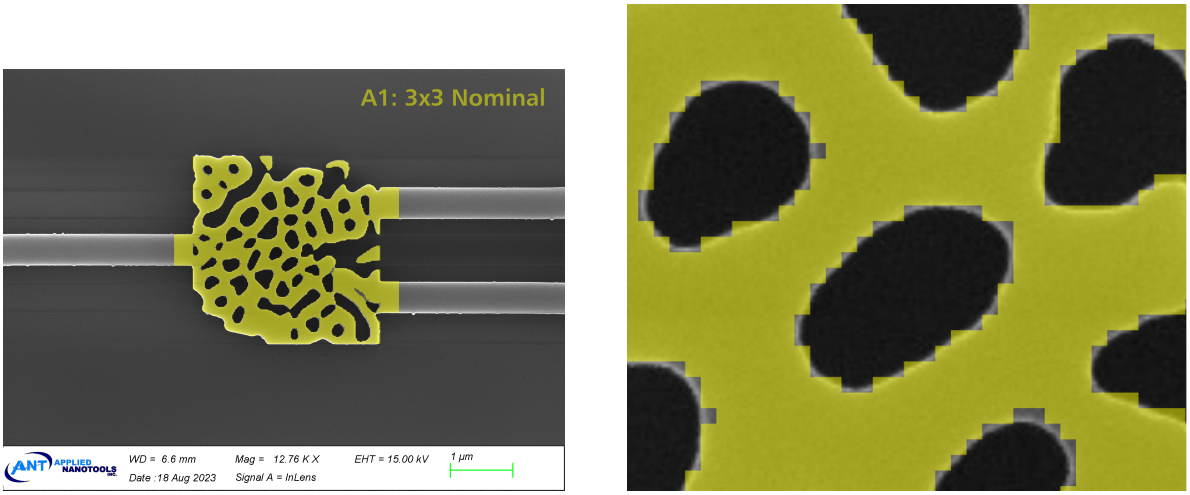


Figure 4.19: Overlay of GDS layout on SEM image of a fabricated device.

widths may become more pronounced when both channels are narrowed by fabrication error.

Compared to existing fabricated devices from previous works (Table 4.2), these devices exhibit a much smaller wavelength shift when fabricated with Prefab and over-etching. This higher performance allows for tighter channel spacings while maintaining high performance in a small footprint, increasing the number of data channels that can be transmitted within the limited bandwidth of a photonic data-transmission system and reducing the physical size of such a system. With these two advantages combined, the corrected devices that we have demonstrated have the potential to make real-world photonic systems more capable and efficient without requiring more physical space.

Table 4.2: Comparison of fabricated devices with C-band (1550 nm) wavelength demultiplexers from previous literature.

Device	Footprint (μm^2)	Ch. spacing (nm)	IL (dB)[1]	XT (dB)[1]	Fabricated blue shift (nm)
Device 2 (nominal)	15	20	3.20	25.3	24
Device 2 (Prefab over-etched)	15	20	1.96	27.1	-2.6
Device 3 (nominal)	9	20	1.85	16.7	20
Device 3 (Prefab over-etched)	9	20	1.85	14.9	-3.6
Piggott 2015 [43]	7.84	250	2.4	11	30[2]
Su 2018 [44]	24.75	40	2.29	10.7	28
Vercruysse 2019 [11]	9	250	2.3	6[2]	61
Yuan 2022 [45]	6.76	50	2	7	3
Zhang 2022 [12]	33.48	50	1.5	16.3	20[2]

[1] Performance is measured at the center of the lowest-performing channel. Channel wavelengths are adjusted for wavelength shifts caused by fabrication.
[2] Estimated value.

4.5 Toward an athermal demultiplexer

The calculation of the FOM based on device parameters can be extended to design a device that is particularly robust to changes that affect a device in predictable, defined ways. This is accomplished by calculating a FOM using a sum of the performance of multiple simulated devices. These devices would all have a design based on a single, shared set of device parameters, but convert these parameters to device design in different ways. For example, a device that is robust to waveguide thickness could be designed by defining a FOM across multiple devices with the same material layout (defined by a set of device parameters) but different thicknesses. Thus, the gradient-descent algorithm will maximize performance across all device variations, ideally creating a device that functions well at any

of the defined waveguide thicknesses.

Superoptimization can be particularly useful in designing an athermal device, as a FOM can be defined to be calculated across multiple temperatures. To design a topologically defined athermal device, a parameter matrix is used to define the design region of multiple devices. In this matrix, the value of each parameter (ρ between 0 and 1) no longer defines the permittivity of a device. Instead, the matrix is used to define a device shape, where $\rho = 0$ is silica and $\rho = 1$ is silicon. This parameter matrix is then used in two separate simulations, where the parameter matrix is converted into a matrix of permittivity values based on the linear equation $\epsilon = \epsilon_{low} + \rho(\epsilon_{high} - \epsilon_{low})$ for each temperature. The sum of the resulting figures of merit is then used in a gradient descent step similar to those used to optimize at a single temperature.

The time taken by an inverse-design algorithm increases when a superoptimization is being run because of the larger number of devices to be simulated in each step. Each device requires its own forward and adjoint simulations for each step, so the total time taken increases roughly proportionally to the number of devices.

We explored extending the optimization algorithm to design an athermal inverse-designed wavelength demultiplexer. We implement this extended algorithm as a superoptimization of two devices with different refractive indices (corresponding to silicon and silica at two different temperatures) and a layout based on a parameter matrix with values (from 0 to 1) that are converted to permittivities for each device. This allows for the optimization of a

device with a fixed layout that can be simulated at two different temperatures.

We tested this superoptimization by optimizing a device using the same materials as the previous devices (silicon and silica) at temperatures of 293K and 350K (20°C and 77°C), a temperature range likely encountered in a datacenter environment. We focus on the first device we designed (3×3 μm^2 footprint), as the FOM used in its optimization is the least complex of the devices we designed, due of its lack of out-of-band suppression, and because a spectral shift could easily be measured by finding the crossover point between channel transmission values.

To find the performance of a device that has not been specifically optimized to be athermal, we first simulated the 3×3 μm^2 demultiplexer, which was optimized at 293K, at 350K by changing the refractive indices of the material but keeping the same layout. At the higher temperature, the simulated device had a crossover point at 1553.1 nm (Fig. 4.20), a shift of 3.9 nm or 50 pm/K from the low-temperature crossover point at 1549.2 nm.

To minimize this shift, we ran a superoptimization that attempted to maximize the FOM from the same device (maximizing transmission) at both 293K and 350K. However, the resulting device, as shown in Fig. 4.20, still experienced a shift in the transmission spectrum of 4.4 nm (57 pm/K). In addition, the crosstalk (6.03 dB) and insertion loss (2.06 dB) were significantly worse at both temperatures for the superoptimized device than for the original, non-athermal device. For comparison, the non-superoptimized device had a crosstalk of 11.74 dB and insertion loss of 0.74 dB.

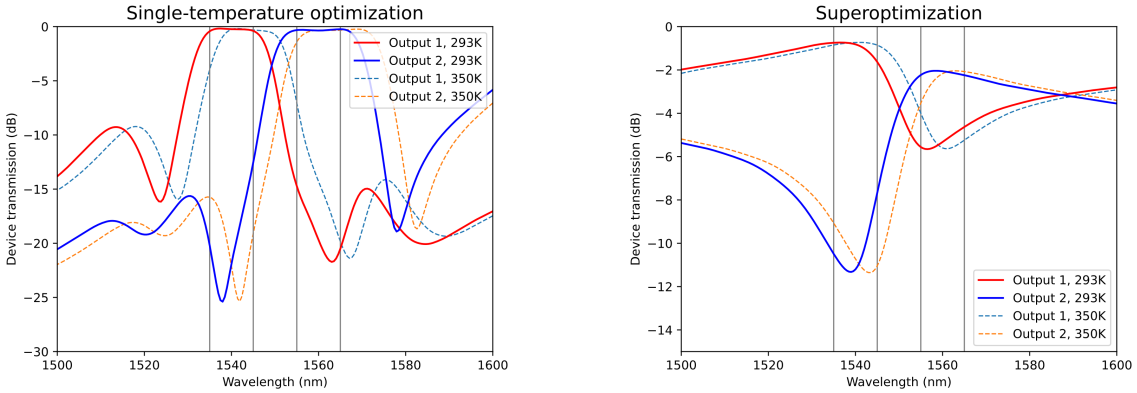


Figure 4.20: Transmission spectra of devices optimized at one temperature (left) and two when simulated at two temperatures. Note the different axis scales for transmission values.

This difference is possibly due to the more complicated combined FOM of the superoptimized device, which makes it more difficult for the optimization algorithm to find a device with particularly high performance. A similar shift also appeared for other superoptimizations we tested with different figures of merit, such as optimization only at the channel centers. Even with fewer terms in the FOM, the resulting device still had crosstalk much worse than for a non-athermal device.

Based on our attempts to design an athermal demultiplexer using inverse design, it appears that our method was unable to reduce the temperature-induced spectral shift of a demultiplexer using superoptimization alone. In the future, it would be useful to test other methods of using inverse design for athermal devices, such as by changing the optimization function or simulation algorithm used in the inverse-design process.

4.6 Multi-step optimizations

During the optimization process for our demultiplexers, we also explored the possibility of conducting the process in multiple steps with different figures of merit. In this procedure (Fig. 4.21), we first optimize a device for a simple FOM (e.g., without reducing out-of-band transmission or crosstalk) using topology optimization to obtain a parameter matrix. This matrix is then used as the input to a second optimization process, which optimizes for the full FOM (with the out-of-band and crosstalk terms). We explore this method based on the hypothesis that performing gradient descent on a simpler figure-of-merit calculation is less computationally expensive and find a better local maximum, while the second optimization is also simpler because it starts with a parameter matrix that already performs well for most of the terms in the FOM.

We tested this method by optimizing a demultiplexer with the same parameters as the second device (3×5 μm^2 footprint, with out-of-band and crosstalk suppression) in two stages and simulating the result. As we did not plan to fabricate these devices, we optimized this device, along with a one-step comparison device, using 2D FDTD simulations.

For the first step in the two-step optimization, we optimize a device based on a FOM containing only the transmission within the two channels, with a spacing of 1 nm between each point at which transmission is measured. The resulting device after this optimization, which does not include a binarization stage, is shown in Fig. 4.22.

When the device is designed only to maximize in-channel transmission, the simulated

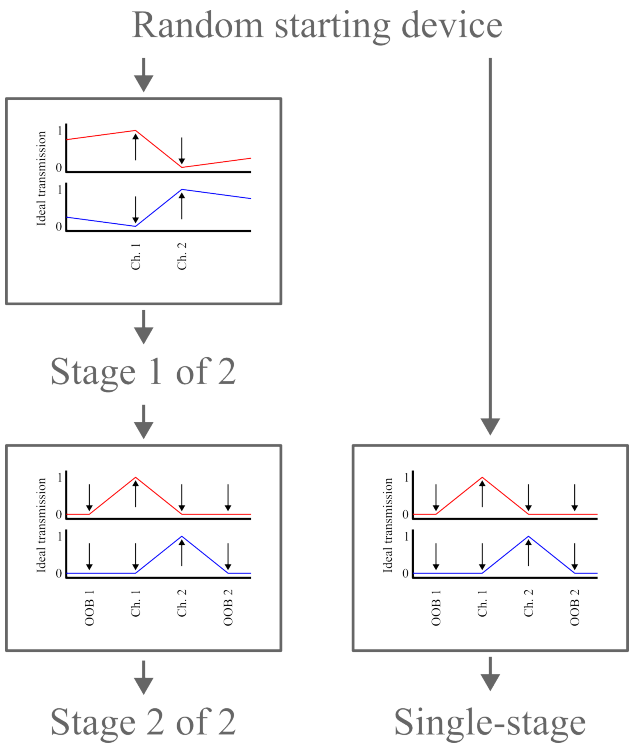


Figure 4.21: Schematic of the multi-stage optimization process (left) compared to single-stage optimization.

worst-case crosstalk at the channel centers is 23.0 dB and the insertion loss is 0.2 dB, with no out-of-band transmission suppression. The second stage of this optimization uses the end of the first optimization as a starting point. Here, the FOM includes the transmission components from the first stage but adds crosstalk reduction components (at 1-nm spacings within the channels) and out-of-band transmission reduction (at wider spacings), as with our 3×5 μm^2 demultiplexer.

The result of this optimization, including a binarization stage, is shown in Fig. 4.23. It has a crosstalk of 19.4 dB and insertion loss of 0.3 dB.

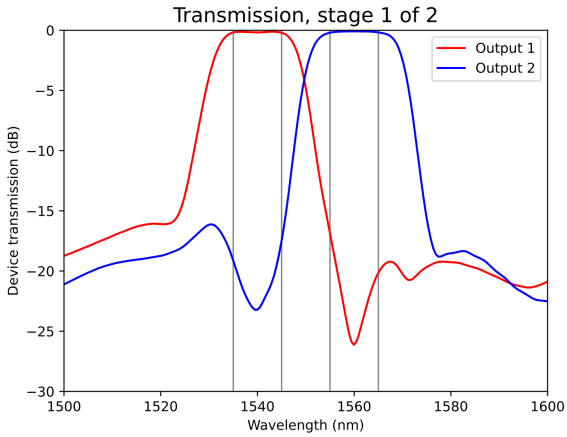


Figure 4.22: Simulated transmission spectrum after the first optimization stage, where the FOMonly includes in-channel transmission.

To compare the result of this two-stage optimization with the baseline one-stage optimization, we optimize the second device using the same FOM as the second stage of the two-stage optimization but with a uniform starting point. As shown in Fig. 4.24, this device has a crosstalk ratio of 23.0 dB and insertion loss of 0.2 dB in simulation, showing higher performance than the device optimized in two stages.

Based on the comparison between the transmission spectra of the two devices, we find that the difference in performance between two-step and one-step optimizations is small and often favors the device optimized in a single stage. In addition, we found the optimization time is not lower for the two-stage device; it took 24 hours to be optimized on a desktop computer (containing a 10-core Intel Core i7-12700K CPU and 32 GB of memory) using 2D FDTD simulations, while the optimization of the one-stage device took 11 hours. Based

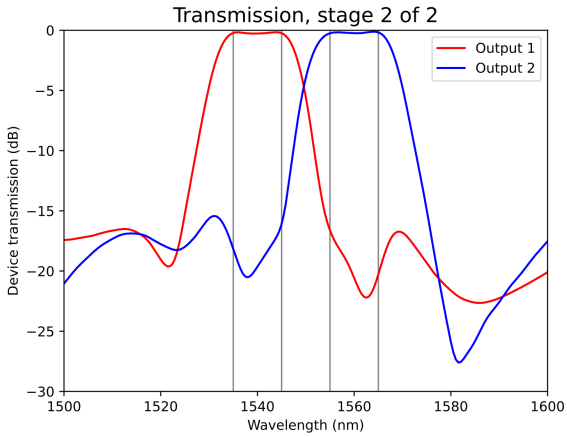


Figure 4.23: Simulated transmission spectrum after the second optimization stage and binarization.

on the results of this experiment, we found that there was limited benefit to the two-stage topology optimization method that we tested.

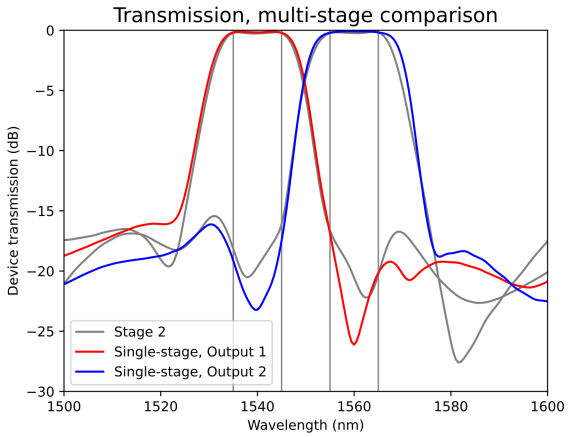


Figure 4.24: Simulated transmission spectrum of a single-stage optimized device.

Chapter 5

Conclusion

In this thesis, we used the process of designing compact wavelength demultiplexers to explore three related topics: converting from the output of an inverse-design algorithm to a physical GDS layout and vice versa; the fabrication and measurement of these devices with machine-learning tools to ensure resilience to fabrication imperfections; and the exploration of superoptimization to design an athermal inverse-designed device.

The conversion from a parameter matrix generated by an inverse-design algorithm to a GDS layout may seem like a mere intermediate step in the inverse-design process, but at the footprint sizes of our demultiplexers even small changes in the shape of a device can have a significant impact. The effects of these differences may have been less apparent in previous devices because of errors caused by fabrication imperfections, but methods (like the Prefab corrector tool) to reduce fabrication errors have made the conversion step more important

to final performance. By developing a method to trace the pixels in the design region of a device more exactly, we are able to more accurately simulate and fabricate an inverse-designed device, improving the performance of our devices and allowing us to test changes to device layouts without the expensive and time-consuming process of device fabrication.

In conjunction with our tracing method, our use of Prefab's machine learning-based correction allows us to design devices that are resilient to fabrication variations by using previous experiences of how these errors affect fabricated devices. We can then design devices that will perform well after these variations are included, even with very small device footprints. This correction even allows us to design devices with details – such as sharp pixel corners – that are significantly smaller than the usual minimum feature size of the fabrication process. By allowing these small features, the correction process allows our devices to achieve higher performance than would normally be possible in a device of its size.

In the course of designing our devices, we also found promising related areas for future research and testing. As an extension of our inverse-design process, we explored the possibility of using superoptimization to design an athermal device that would have similar performance at multiple temperatures. While the device resulting from this superoptimization did not appear to have improved athermal performance, it would be worthwhile to examine the process further to improve the athermal performance of our devices. In addition, the significant optimization time and resource requirements spurred us

to explore multi-step optimization as a potential method to design high-performance devices more quickly. In the future, we could examine other possibilities, such as changes in the optimization algorithm or simulation methods, that would enable us to increase the efficiency of our design process and allow these inverse-designed devices to make an even more significant improvement over current technologies.

The amount of data transmitted across the world is still growing, requiring data-transmission systems to keep up with this growth using new technologies. In our work, we demonstrate a version of the wavelength demultiplexers that play a key role in this data transmission. With its small footprint and high performance, it has the potential to play a role in increasing the physical density of data connections, reducing the amount of space and power needed and enabling systems to keep up with the demands placed on them today and into the future.

Bibliography

[1] G. Keiser, *Fiber optic communications*, 1st ed. Singapore: Springer, 2021. [Online]. Available: https://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=2802598

[2] International Telecommunication Union. (2023) Measuring digital development: Facts and figures 2023. Accessed: May 24, 2024. [Online]. Available: https://www.itu.int/en/ITU-D/Statistics/Pages/facts/default.aspx

[3] J. R. Kiniry, "Wavelength division multiplexing: ultra high speed fiber optics," *IEEE Internet Computing*, vol. 2, no. 2, pp. 13–15, 1998.

[4] S. V. Kartalopoulos, *Introduction to DWDM technology : data in a rainbow*, 1st ed. Bellingham, Wash.: SPIE Optical Engineering Press, 2000.

[5] S. Molesky, Z. Lin, A. Y. Piggott, W. Jin, J. Vucković, and A. W. Rodriguez, "Inverse design in nanophotonics," *Nature Photonics*, vol. 12, no. 11, pp. 659–670, 2018. [Online]. Available: https://doi.org/10.1038/s41566-018-0246-9

[6] J. Han, J. Huang, J. Wu, and J. Yang, "Inverse designed tunable four-channel wavelength demultiplexer," *Optics Communications*, vol. 465, p. 125606, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0030401820301966

[7] Y. A. Yilmaz, A. M. Alpkilic, A. Yeltik, and H. Kurt, "Inverse design of efficient and compact 1 × N wavelength demultiplexer," *Optics Communications*, vol. 454, p. 124522, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0030401819307989

[8] H. Yuan, J. Huang, Z. Wang, J. Zhang, Y. Deng, G. Lin, J. Wu, and J. Yang, "An ultra-compact dual-channel multimode wavelength demultiplexer based on inverse design," *Results in Physics*, vol. 27, p. 104489, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S221137972100601X

[9] Y. Xu, J. Huang, L. Yang, H. Ma, H. Yuan, T. Xie, J. Yang, and Z. Zhang, "Inverse-designed ultra-compact high efficiency and low crosstalk optical interconnect based on waveguide crossing and wavelength demultiplexer," *Scientific Reports*, vol. 11, no. 1, p. 12842, 2021. [Online]. Available: https://doi.org/10.1038/s41598-021-92038-w

[10] G. Zhang, D.-X. Xu, Y. Grinberg, and O. Liboiron-Ladouceur, "Topological inverse design of nanophotonic devices with energy constraint," *Opt. Express*, vol. 29, no. 8, pp. 12 681–12 695, Apr 2021. [Online]. Available: https://opg.optica.org/oe/abstract.cfm?URI=oe-29-8-12681

[11] D. Vercruysse, N. V. Sapra, L. Su, R. Trivedi, and J. Vučković, "Analytical level set fabrication constraints for inverse design," *Scientific Reports*, vol. 9, no. 1, p. 8999, 2019. [Online]. Available: https://doi.org/10.1038/s41598-019-45026-0

[12] G. Zhang, D.-X. Xu, Y. Grinberg, and O. Liboiron-Ladouceur, "Experimental demonstration of robust nanophotonic devices optimized by topological inverse design with energy constraint," *Photonics Research*, vol. 10, no. 7, pp. 1787–1802, 2022. [Online]. Available: https://opg.optica.org/prj/abstract.cfm?URI=prj-10-7-1787

[13] H. H. Li, "Refractive index of silicon and germanium and its wavelength and temperature derivatives," *Journal of Physical and Chemical Reference Data*, vol. 9, no. 3, pp. 561–658, 1980. [Online]. Available: http://dx.doi.org/10.1063/1.555624

[14] J. H. Wray and J. T. Neu, "Refractive index of several glasses as a function of wavelength and temperature," *Journal of the Optical Society of America*, vol. 59, no. 6, pp. 774–776, 1969. [Online]. Available: https://opg.optica.org/abstract.cfm?URI=josa-59-6-774

[15] L. Chrostowski and M. Hochberg, *Silicon Photonics Design: From Devices to Systems*, 1st ed. Cambridge: Cambridge University Press, 2015. [Online]. Available: https://www.cambridge.org/core/product/BF3CF13E8542BCE67FD2BBC7104ECEAB

[16] J. S. Orcutt, B. Moss, C. Sun, J. Leu, M. Georgas, J. Shainline, E. Zgraggen, H. Li, J. Sun, M. Weaver, S. Urošević, M. Popović, R. J. Ram, and V. Stojanović, "Open foundry platform for high-performance electronic-photonic

integration," *Optics Express*, vol. 20, no. 11, pp. 12 222–32, 2012. [Online]. Available: http://dx.doi.org/10.1364/OE.20.012222

[17] C. R. Pollock and M. Lipson, *Integrated photonics*, 1st ed. Boston: Kluwer Academic, 2003.

[18] E. Zhang, S. Zhang, J. Huang, X. Zhu, T. Yang, L. Chen, Y. Zhang, X. Yang, and L. Zhang, "Compact low-loss dual-mode silicon waveguide bend with simple geometry," *2023 21st International Conference on Optical Communications and Networks (ICOCN)*, pp. 1–3, 2023.

[19] W. Bogaerts, S. K. Selvaraja, P. Dumon, J. Brouckaert, K. D. Vos, D. V. Thourhout, and R. Baets, "Silicon-on-insulator spectral filters fabricated with CMOS technology," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 16, no. 1, pp. 33–44, 2010.

[20] K. Okamoto, "Progress and technical challenge for planar waveguide devices: silica and silicon waveguides," *Laser & Photonics Reviews*, vol. 6, no. 1, pp. 14–23, 2012. [Online]. Available: http://dx.doi.org/10.1002/lpor.201100003

[21] K. Hassan, C. Sciancalepore, J. Harduin, T. Ferrotti, S. Menezo, and B. B. Bakir, "Toward athermal silicon-on-insulator (de)multiplexers in the O-band," *Optics Letters*, vol. 40, no. 11, pp. 2641–2644, 2015. [Online]. Available: https://opg.optica.org/ol/abstract.cfm?URI=ol-40-11-2641

[22] F. Horst, W. M. J. Green, B. J. Offrein, and Y. A. Vlasov, "Silicon-on-insulator echelle grating WDM demultiplexers with two stigmatic points," *IEEE Photonics Technology Letters*, vol. 21, no. 23, 2009. [Online]. Available: http://dx.doi.org/10.1109/LPT.2009.2032151

[23] B. Frey, D. Leviton, and T. Madison, "Temperature-dependent refractive index of silicon and germanium - art. no. 62732j," *Proceedings of SPIE - The International Society for Optical Engineering*, vol. 6273, 06 2006.

[24] G. K. White, "Thermal expansion of reference materials: copper, silica and silicon," *Journal of Physics D: Applied Physics*, vol. 6, no. 17, pp. 2070–2078, 1973. [Online]. Available: http://dx.doi.org/10.1088/0022-3727/6/17/313

[25] F. Morichetti, S. Grillanda, and A. Melloni, "Breakthroughs in photonics 2013: Toward feedback-controlled integrated photonics," *IEEE Photonics Journal*, vol. 6, no. 2, pp. 1–6, 2014.

[26] K. Okamoto, *Fundamentals of optical waveguides*, 3rd ed. London: Academic Press, 2022. [Online]. Available: https://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=2741998

[27] S. D. Gedney, *Introduction to the finite-difference time-domain (FDTD) method for electromagnetics*, 1st ed., ser. Synthesis lectures on computational electromagnetics; #

27. Cham, Switzerland: Springer, 2011. [Online]. Available: http://public.ebookcentral.proquest.com/choice/publicfullrecord.aspx?p=881226

[28] C. Lalau-Keraly. (2018) Figures of merit - LumOpt documentation. Accessed: May 30, 2024. [Online]. Available: https://lumopt.readthedocs.io/en/latest/foms.html

[29] L. Su, D. Vercruysse, J. Skarda, N. V. Sapra, J. A. Petykiewicz, and J. Vučković, "Nanophotonic inverse design with SPINS: Software architecture and practical considerations," *Applied Physics Reviews*, vol. 7, no. 1, p. 011407, 5/13/2024 2020. [Online]. Available: https://doi.org/10.1063/1.5131263

[30] C. M. Lalau-Keraly, S. Bhargava, O. D. Miller, and E. Yablonovitch, "Adjoint shape optimization applied to electromagnetic design," *Optics Express*, vol. 21, no. 18, pp. 21 693–21 701, 2013. [Online]. Available: https://opg.optica.org/oe/abstract.cfm?URI=oe-21-18-21693

[31] ANSYS, Inc. Photonic inverse design overview - python API – Ansys Optics. Accessed: May 10, 2024. [Online]. Available: https://optics.ansys.com/hc/en-us/articles/360049853854-Photonic-Inverse-Design-Overview-Python-API

[32] G. H. Ahn, K. Y. Yang, R. Trivedi, A. D. White, L. Su, J. Skarda, and J. Vučković, "Photonic inverse design of on-chip microresonators," *ACS Photonics*, vol. 9, no. 6, pp. 1875–1881, 06 2022. [Online]. Available: https://doi.org/10.1021/acsphotonics.2c00020

[33] C. Zhu, R. H. Byrd, P. Lu, and J. Nocedal, "Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization," *ACM Trans. Math. Softw.*, vol. 23, no. 4, pp. 550–560, dec 1997. [Online]. Available: https://doi.org/10.1145/279232.279236

[34] S. Shalev-Shwartz and S. Ben-David, *Understanding machine learning : from theory to algorithms*. New York, NY: Cambridge University Press, 2014. [Online]. Available: https://doi.org/10.1017/CBO9781107298019

[35] Z. Liu, D. Zhu, L. Raju, and W. Cai, "Tackling photonic inverse design with machine learning," *Advanced Science*, vol. 8, no. 5, p. 2002923, 2021. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/advs.202002923

[36] D. Gostimirovic, D.-X. Xu, O. Liboiron-Ladouceur, and Y. Grinberg, "Deep learning-based prediction of fabrication-process-induced structural variations in nanophotonic devices," *ACS Photonics*, vol. 9, no. 8, pp. 2623–2633, 08 2022. [Online]. Available: https://doi.org/10.1021/acsphotonics.1c01973

[37] J. S. Jensen and O. Sigmund, "Topology optimization for nano-photonics," *Laser & Photonics Reviews*, vol. 5, no. 2, pp. 308–321, 2024/05/14 2011. [Online]. Available: https://doi.org/10.1002/lpor.201000014

[38] ANSYS, Inc. Selecting the best mesh refinement option in the FDTD simulation object. Accessed: May 10,

2024. [Online]. Available: https://optics.ansys.com/hc/en-us/articles/360034382614-Selecting-the-best-mesh-refinement-option-in-the-FDTD-simulation-object

[39] Applied Nanotools Inc. Silicon resources: Process specifications. Accessed: May 10, 2024. [Online]. Available: https://www.appliednt.com/nanosoi/sys/resources/specs/

[40] ANSYS, Inc. Optimizable geometry - python API. Accessed: May 13, 2024. [Online]. Available: https://optics.ansys.com/hc/en-us/articles/360052044913-Optimizable-Geometry-Python-API

[41] M. Zhou, B. S. Lazarov, F. Wang, and O. Sigmund, "Minimum length scale in topology optimization by geometric constraints," *Computer Methods in Applied Mechanics and Engineering*, vol. 293, pp. 266–282, 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0045782515001693

[42] D. Gostimirovic, Y. Grinberg, D.-X. Xu, and O. Liboiron-Ladouceur, "Improving fabrication fidelity of integrated nanophotonic devices using deep learning," *ACS Photonics*, vol. 10, no. 6, pp. 1953–1961, 06 2023. [Online]. Available: https://doi.org/10.1021/acsphotonics.3c00389

[43] A. Y. Piggott, J. Lu, K. G. Lagoudakis, J. Petykiewicz, T. M. Babinec, and J. Vučković, "Inverse design and demonstration of a compact and broadband on-chip wavelength demultiplexer," *Nature Photonics*, vol. 9, no. 6, pp. 374–377, 2015. [Online]. Available: https://doi.org/10.1038/nphoton.2015.69

[44] L. Su, A. Y. Piggott, N. V. Sapra, J. Petykiewicz, and J. Vučković, "Inverse design and demonstration of a compact on-chip narrowband three-channel wavelength demultiplexer," *ACS Photonics*, vol. 5, no. 2, pp. 301–305, 02 2018. [Online]. Available: https://doi.org/10.1021/acsphotonics.7b00987

[45] M. Yuan, G. Yang, S. Song, L. Zhou, R. Minasian, and X. Yi, "Inverse design of a nano-photonic wavelength demultiplexer with a deep neural network approach," *Optics Express*, vol. 30, no. 15, pp. 26 201–26 211, 2022. [Online]. Available: https://opg.optica.org/oe/abstract.cfm?URI=oe-30-15-26201