Thermal-based Photonic Integrated Switches using Inverse Design Methodologies

Rebecca Rogers
Department of Electrical and Computer Engineering
McGill University, Montreal

Submitted on November, 2023

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of

Master of Electrical Engineering

Abstract

With the rise of 5G, artificial intelligence and streaming services, there has been an exponential increase in data traffic. With the ever-growing amount of data being transmitted and stored, we need our communications systems to grow in capacity. Traditional electronic data processing and transmission systems are not fast enough to keep up with these demands. Not only do electronics fail to keep up in data rates, but also they are energy costly which only increases with data consumption. Alternatively, we can turn to photonics, which offers high data transfer rates at low loss and minimal crosstalk. Given the orthogonality of light, photonics offers the ability for multiple lines of data to transverse down the same path. The data then needs to be separated and directed to different paths within the data centers using switches and de-multiplexers. However, many integrated devices developed in SOI (silicon-on-insulator) technology take up significant design space and are susceptible to fabrication errors. To maintain the benefits of silicon photonics and decrease faults in some intuitively-based designs, we turn to the help of inverse design.

Inverse design utilizes modern artificial intelligence algorithms to generate devices that are more compact, energy efficient, and robust to fabrication errors. In this work, we utilize an open-source inverse design package to present a complete account of how to generate novel inverse designed devices that are tested across various software. By comparing meshing capabilities, the differences between 2-D and 3-D simulations, and conversion from matrices to layout files between simulations, we discuss the challenges and solutions to generating novel inverse designed devices. By discussing these issues,

we gain a fundamentally deeper understanding of the processes within these software which allows us to generate more robust and well performing devices compared to conventionally designed devices.

In recent years, silicon photonic inverse design has focused greatly on multiplexing technologies as there is a linear relation between refractive index of materials and wavelength. Changing the refractive index of a material also changes the path of the light and inverse design explores designs where light changes output ports from wavelength dependent refractive index changes. Additionally, the temperature of the material changes the refractive index and this relation has been explored in terms of temperature robustness. Temperature sensitivity has been scarcely studied in inverse design as the temperature dependence of many optical materials is significantly smaller compared to wavelength refractive index dependence. However, to utilize the full capabilities of inverse design, it is important to explore this temperature relation.

In this thesis, we look to increase photonic integrated switching capabilities in photonic inverse design by producing a novel inverse design one by two space switch that is thermally tuneable. We first demonstrate this thermal switch in 2-D simulations, showing a switching capability for an optical wavelength channel at 1550 nm with a 90% transmission at one of the two output ports switched to 80% at the other output port for a 200 K temperature change. In the 3-D realm for more realistic simulation at the cost of more demanding computation, we produce a device that reaches 70% transmission at T_1 in one output port and 80% transmission in the opposite output port for a 200 K temperature change. This device in SOI technology has a 7.5 μm by 2.5 μm footprint, drastically decreasing the size for switching in on-chip silicon switches which typically take up is greater than 100 μm in length. As silicon photonics increases its role in data communications, it is imperative to further explore the capabilities of inverse design and produce novel while compact devices such as the thermal-based optical space switch presented in this thesis.

Abrégé

Avec l'essor de la 5G, de l'intelligence artificielle et des services de streaming, le trafic de données a connu une augmentation exponentielle. Avec la quantité toujours croissante de données transmises et stockées, nous avons besoin que nos systèmes de communication augmentent en capacité. Les systèmes électroniques traditionnels de traitement et de transmission des données ne sont pas assez rapides pour répondre à ces demandes. Non seulement l'électronique ne parvient pas à suivre le rythme des débits de données, mais elle est également coûteuse en énergie et ne fait qu'augmenter avec la consommation de données. Nous pouvons également nous tourner vers la photonique, qui offre des taux de transfert de données élevés avec de faibles pertes et une diaphonie minimale. Compte tenu de l'orthogonalité de la lumière, la photonique offre la possibilité à plusieurs lignes de données de suivre le même chemin. Les données doivent ensuite être séparées et dirigées vers différents chemins au sein des centres de données à l'aide de commutateurs et de démultiplexeurs. Cependant, de nombreux dispositifs intégrés développés dans la technologie SOI (silicium sur isolant) occupent un espace de conception important et sont susceptibles aux déviations dans leur fabrication. Pour conserver les avantages de la photonique sur silicium et réduire les défauts de certaines conceptions intuitives, nous nous tournons vers la conception inverse.

La conception inversée utilise des algorithmes d'intelligence artificielle et modernes pour générer des dispositifs plus compacts, plus économes en énergie et plus robustes aux erreurs de fabrication. Dans ce travail, nous utilisons un logiciel de conception inverse *open source* pour présenter un compte rendu complet de la manière de générer de nouveaux dispositifs de conception inverse qui sont testés sur des logiciels commerciaux reconnus. En comparant les capacités de maillage, les différences entre les simulations 2D et 3D et la conversion des matrices en fichiers de mise en page entre les simulations, nous discutons des difficultés et des solutions pour générer de nouveaux circuits photoniques de conception inverse. En discutant de ces problèmes, nous acquérons une compréhension fondamentalement plus approfondie des processus au sein de ces logiciels, ce qui nous permet de générer des circuits photoniques plus robustes et plus performants que les circuits photoniques de conception conventionnelle.

Ces dernières années, la conception inverse photonique au silicium s'est fortement concentrée sur les technologies de (dé)multiplexage, car il existe une relation linéaire entre l'indice de réfraction des matériaux et la longueur d'onde. La modification de l'indice de réfraction d'un matériau modifie également le trajet de la lumière et la conception inverse explore les conceptions dans lesquelles la lumière change de ports de sortie en fonction des changements d'indice de réfraction en fonction de la longueur d'onde. De plus, la température du matériau modifie l'indice de réfraction et cette relation a été explorée en termes de robustesse à la température. La sensibilité à la température a peu été étudiée dans le cadre d'une conception inverse, car la dépendance à la température de nombreux matériaux optiques est nettement inférieure à la dépendance à l'indice de réfraction de la longueur d'onde. Cependant, pour utiliser toutes les capacités de la conception inverse, il est important d'explorer cette relation de température.

Dans cette thèse, nous cherchons à augmenter les capacités de commutation photonique intégrée dans la conception photonique inverse en concervant un nouveau commutateur spatial un par deux thermiquement accordable. Nous démontrons d'abord ce commutateur thermique dans des simulations 2D, montrant une capacité de commutation pour un canal de longueur d'onde optique à 1550 nm avec une transmission de 90% sur l'un des deux ports de sortie commutée à 80% sur l'autre port de sortie pour un changement de température de 200 K. Dans une simulation 3D plus réaliste au prix de calculs plus exigeants, nous produisons un dispositif qui atteint 70% de transmission dans

un port et 80% de transmission dans le port opposé pour un changement de température de 200 K. Ce dispositif en technologie SOI a une empreinte de 7,5 µm sur 2,5 µm, réduisant considérablement la taille de commutation des commutateurs en silicium sur puce qui occupent généralement plus de 100 µm de longueur. Alors que la photonique sur silicium augmente son rôle dans les communications de données, il est impératif d'explorer davantage les capacités de la conception inverse et de produire des dispositifs nouveaux et compacts tels que le commutateur spatial optique thermique présenté dans cette thèse.

Acknowledgements

I would first like to thank my advisor, Professor Odile Liboiron-Ladouceur, for her endless support. She has helped me as I went through the transition of physics to photonics engineering and guided me to a research path I genuinely enjoyed. Throughout my graduate experience, she has been incredibly patient with me while I searched for a thesis topic. I appreciate all the opportunities she has given me throughout this experience. Next, I would like to thank our NRC Collaborator: Dr. Dan-Xia Xu and Dr. Yuri Grinberg. Without their guidance and help I would have been completely lost. I want to thank them sincerely for their input and feedback, as it was incredibly valuable to my work and helped shape this thesis.

The DataComm team at McGill has been enormously helpful and welcoming when I joined the group. The knowledge of the group in photonics engineering has helped me immensely in my work, guiding me with even the simplest questions. I would like to thank Dr. Dusan Gostimirovic and Dr. Guowu Zhang for their help with my simulation set ups and optimization set ups and for their patience with my millions of emails. A very large gratitude to Dr. Kaveh Rahbardar Mojaver, who guided me in GDS layout. Additionally, I would like to thank Md Mahadi Masnad, Sunami Morrison, Rifat Nazneen, and José Garcia Echeverria for their help. Lastly, I would like to thank Eilidh Jurus for helping edit this thesis, I'm glad we both agree on the Oxford comma.

Lastly, I want to thank my family and friends for their support throughout my masters. I thank my parents, grandmother and sister for their endless support and my brother for his endless competition.

Table of Contents

List of Figures

List of Acronyms

AI	Artificial Intelligence
EM	Electro-magnetic
FOM	Figure of Merit
FDFD	Finite Difference Frequency Domain
FDTD	Finite Difference Time Domain
GDS	Graphic Design System
MEMS	Micro-electromechanical systems
MDM	Mode Division Multiplexing
MMI	Multi-Mode Interferometer
MRR	Micro-Ring Resonator
MZI	Mach Zehnder Interferometer
PCM	Phase Change Material
PCB	Printed Circuit Board
PIC	Photonic Integrated Circuit
SDM	Space Division Multiplexing
SOI	Silicon on Insulator
SPINS	Stanford Photonics INverse design Software
TE	Transverse Electric
TM	Transverse Magnetic
WDM	Wavelength Division Multiplexing

Chapter 1

Introduction

“Our fastest and most reliable technology yet”

You may see this phrase written on billboards boasting 5G technology or in a cellular network storefront. This sends a message to consumers that they may use more data, send more data, and store more data. This win for the consumer presents a challenge to the communications engineer. The electronics-based communications from the 20th century can simply not keep up with this increased data rate, hence we must turn to photonics. Photonics, very generally speaking, is the study of light and how photons interact with matter. In the world of communications, photonics can be used as a reliable and faster alternative to electrical signaling. Using photons, we can increase the speed and capacity of data transmission which allows us to keep up with modern demands for higher data rates. Based on some of the core principles of physics, more specifically Maxwell’s equations, we can design and model devices that transmit, contain, manipulate, and process light. Photonic integrated circuits (PICs) have been increasingly helpful in the transfer and modulation of data at incredibly high rates. Long haul data communication has been almost completely reliant on optical fibers to transmit information. As technology continues to grow in capacity, from artificial intelligence (AI) to quantum computing, photonics continues to play a significant role in this growth.

1.1 Rising Data Traffic

One of the biggest advancements in modern data communications is the implementation of photonics. From long haul fiber-optics that run deep in the ocean to short reach interconnects within data centers, photonics has revolutionized the field of communications. Contrasting with copper cabling, which has at best a data transfer rate of 54 Gb/s (Gigabits per second) [1], photonics can transfer hundreds of gigabits per second [2]. Additionally, photonics offers low loss with large bandwidth compared with conventional electronic circuits and copper wiring. However, with incredibly compact components that are vastly smaller than photonic components, electronics remains an integral part of data communications. In figure 1.1, the role of copper and optical components in data centers is illustrated and the extensive use of copper interconnects is shown. The electronics built into modern data communications systems cannot keep up with the data transmitted from its photonic counterparts. This poses a challenge for data centers as they rely heavily on photonic/electronic interfaces.

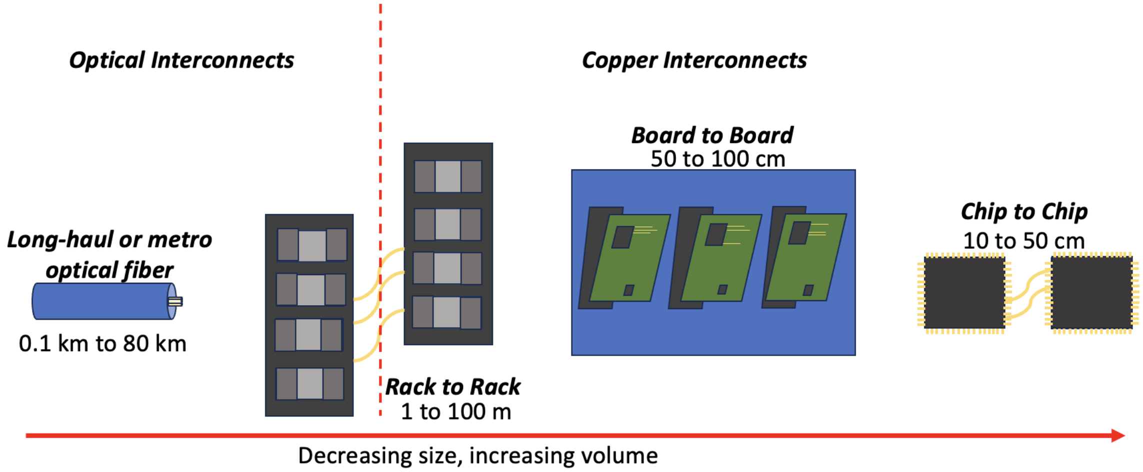


Figure 1.1: Optical interconnects for data communications as a relative function of link distances and the number of components per surface area.

Entering into a new era of data communications, we have significant amounts of data moving not only in and out of data centers, but an increasing amount of internal traffic as well. The internal traffic, also known as east-west traffic, is approximately 75% of the total communications traffic [3]. Internal traffic, responsible for processing and storage, is becoming increasingly complex and incredibly costly. With the rise of AI, the internal calculations that must be performed and processed have increased exponentially. As the computational power becomes greater, the electronics that we rely on so heavily within our processors and interconnects are limited.

Optical interconnects offer great capacity for data transmission in terms of both data transport and speed. Optical links have the benefit of light transfer over fiber, which is significantly faster than the transfer of electrons over copper wires. Additionally, we benefit from both Wavelength Division Multiplexing (WDM) and Mode Division Multiplexing (MDM) to increase our optical bandwidth. A WDM system transmits multiple data lines on the same waveguide simultaneously as each wavelength carries independent data. Given that the wavelength signals are orthogonal to one another, several lines of data can transverse down the same single mode waveguide resulting in high data transmission with minimal crosstalk (around -23 dB in more recent studies [4]). Moreover, MDM also provides similar data transmission capacities with multiple data lines transmitted on different optical modes instead of multiple wavelengths. Optical modes are the orthogonal discrete solutions to Maxwell's equations. A major advantage of utilizing MDM systems is the ability to transmit multiple lines of data all deriving from a singular source. In more recent studies, MDM systems have been implemented with as low as -30 dB crosstalk over the entirety of C-band region [5] making them broadband with little signal degradation. With multiple signals being sent down the same waveguide, we can effectively create interconnects that have much higher data capacity compared to their electronic counterparts.

With data traffic rising, there are increasing concerns about the sustainability of electronics. With data centers consuming approximately 1% of the total energy consumption globally, there is a need to generate greener devices. There is an increasing amount of data consumption worldwide, by 2026 we expect to see 175 ZB/yr with an electricity consumption of 4500 TWh/yr [6]. Optical interconnects offer a possible solution to this energy bottleneck. Compared to electronic interconnects, which consume 10W for a 10 Gb/s link, silicon photonic interconnects only require 0.2 W for a 10 Gb/s link [7]. Additionally, 50% of the energy consumption in data centers is caused by cooling systems for electronic systems [8]. Photonics do not generate as much heat as electronics leading to a reduced need for cooling systems within data centers. Utilizing photonics at a greater capacity in data centers will create a much less energy hungry system, therefore addressing cost and environmental problems that arise from increased data consumption.

Also creating a roadblock for the future of electronics is the end of Moore's law [9]. Moore's law states that the total number of transistors within a circuit doubles about every year [10]. Increasing the number of transistors on electronic chips increases our computational power, allowing for the rise of AI, streaming, and other applications. With greater computational algorithms comes greater internal and external data traffic, as users and companies have access to advanced computational processes. However, we will soon reach a physical limit to transistors as they shrink down to the size of a few atoms [10]. As we cannot break the fundamental laws of physics to continuously shrink down transistors, we must come up with new and innovative strategies to handle the exponentially increasing data.

Current photonic technology cannot replace electronics completely, despite the benefits that silicon photonics offers. For one, photonics cannot store or process data, making electronic chips imperative for computational functionality. Additionally, the wavelength of electrons is only a few nanometers compared to the wavelength of photonics at hundreds to thousands of nanometers. Such a large wavelength does not allow us to scale photonics to the same ability as electronics, even with advanced lithography. Due to this significant size difference, electronics have extremely small features and compact wiring that photonics can simply not compete with. For example, a single transistor at a few nanometers can act as a switch, whereas a photonic ring modulator can take up to 10,000 times the space for the same function [11]. Regardless, silicon photonics still offers great benefits in terms of efficiency and broadband capabilities, making it integral for the future of these systems. To increase the hand of silicon photonics in data communication, devices must decrease their footprint and the device library must be increased in non-conventional ways.

1.2 Inverse Design in Photonics

Photonics engineering is reaching a bottleneck in performance, with many devices being designed using physics-backed intuition and theory. As we put a greater emphasis on the role of photonic interconnects within data centers to address growing data consumption, we must create not only smaller and more energy efficient photonic devices, but novel photonic devices as well. Current device libraries are limited in terms of design, with large devices used for simple functions. By generating novel devices, we can further the capabilities of photonic devices as well as expand our design library to include smaller, compact and energy efficient devices. With conventional design methodologies based on theory, there are limited degrees of freedom that the engineer can access within the design space. With greater degrees of freedom, we can increase design robustness, increase efficiency, and explore optical functionalities. To access these benefits, we turn to the practice of inverse design.

Inverse design allows us to explore the entirety of the design space by utilizing advanced intelligence algorithms, such as iterative design and deep learning [12]. By treating the desired functionality of the device as an optimization problem, inverse design explores the design space to generate devices. In figure 1.2, the methodology differences between inverse design and conventional device design are shown. There are many degrees of freedom within photonics design and manipulating them through mathematical methods is a tedious task. By using inverse design, photonic devices with various geometries can be researched with greater ease.

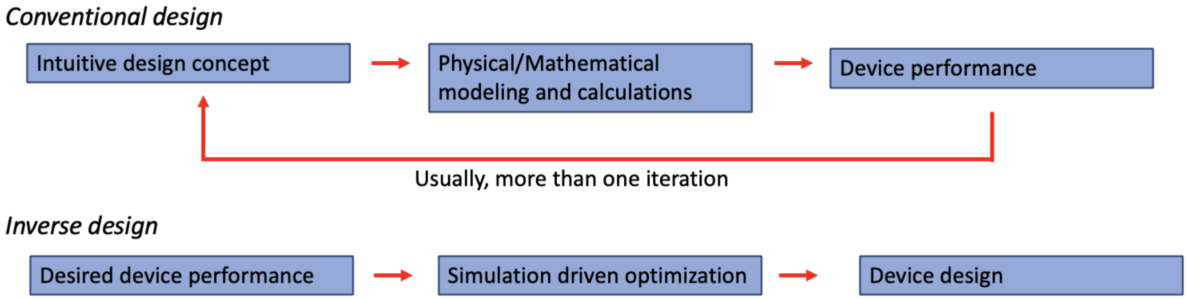


Figure 1.2: Conventional design methodology based on theory compared to inverse design methodology.

Inverse design began in the 90's within the field of chemistry, while searching for different atomic configurations for desired material properties [13]. From there, inverse design has had a hand in various fields, from mechanical engineering to material sciences to photonics. Photonic inverse design began with simpler algorithms, such as particle swarming [14] and parameter sweeping [15], where given a previously designed device, the parameters and boundaries of the device were explored and optimized for peak performance. In the late 90's, inverse design was introduced to photonics with the innovative work of Cox and Dobson [16], who utilized the gradient descent methodology to optimize the band-gap with changes to the dielectric. In the early 2000's, level-set and density topology were introduced which drastically increased our ability to explore design geometries within a given space. Level-set design utilizes the level-set function to explore the design space, which can greatly vary the geometry of the design while easing computational cost [17]. Density topology optimization also allows a great amount of variation in the design space, as this methodology relaxes the parameters of the components within a certain range [18]. To converge on a solution, many inverse design methodologies rely on a gradient descent that, when given a series of parameters, a local minimum can be reached. The local minimum of these descents often performs incredibly well, making the gradient descent a reliable method. Additionally, to decrease computational cost, the adjoint method is implemented, which uses linear algebraic processes to calculate the gradient [19]. Utilizing the extensive research and well-defined design methods of past inverse design research, inverse design has taken off in recent years for nanophotonics.

With a large bottleneck of photonic interconnect development being limited design libraries and the large size of our devices, we look to utilize inverse design methodologies to put a heavier emphasis on nanophotonic devices for interconnects. Within data centers, we rely greatly on multiplexing and switching technology to transfer data throughout the system and to perform calculations [20]. To increase broadband and decrease device size, inverse design has been investigated for mode converters and multiplexing technology [21]. In recent years, there have been mode multiplexers, wavelength multiplexers, mode converters, and Y-branch devices optimized using inverse design methods. Extending this work and coming up with novel devices allow us to increase the role photonics has in data centers, hence keeping in line with the ever increasing data demands.

1.3 Motivation

Although photonics offers many benefits for data communications, the main being speed at which data is transmitted, it still has some shortcomings compared to its electronic counterparts. For one, electronics have the capability to be much smaller than photonic devices. Electronic chips are mainly composed of transistors, which have become increasingly small and compact, with 30 billion on a die the size of a fingernail [22]. In many electronic chips, each transistor measures less than 5 nm [23]. This is incredibly compact compared to photonic devices, with many straight rectangular waveguides used in single mode transmission being approximately 450 nm in width. Along with the compactness of transistors, electronic wiring also provides many benefits that traditional photonics cannot keep up with such as the ability to create tight curves with smaller bending radii at lower signal loss. Additionally, electronic wires lack coupling from other nearby wiring, unlike photonic waveguides, which if in close proximity can have evanescent wave overlap. Electronic interconnects are still not optimal: at data rates close to 25 Gb/s they have significant signal degradation, requiring special materials that are much more costly than standard copper for printed circuit boards (PCB). Additionally, electronic interconnects are power consuming outside of short connections, taking as much as three times the power consumption (i.e., mW per Gb/s) [24]. Silicon photonics offers higher data transfer rate with lower energy cost making it a competitive candidate for processing within data centers. To reduce our footprint to keep up with increasingly small electronics, photonics can turn to inverse design.

By replacing some of our older devices with newer and more compact inverse designed devices, we can decrease the size requirements of the silicon dies in data centers and increase the capacity of data transmission. Using ring resonators for wavelength multiplexing takes up a large amount of space, with each ring resonator supporting a single wavelength. A single ring resonator can be as small as 4 μm in diameter [25], but when there are multiple cascaded rings in series to perform multiplexing tasks, this area becomes increasingly large. For example, for a ring-based device that de-multiplexes three wavelengths, each ring takes up 16π μm^2, with the total required area being larger than 48π μm^2 if we neglect the additional space required between the rings. However, a three wavelength de-multiplexer designed using inverse design has an area of approximately 4 μm by 4 μm [26], significantly reducing the size requirements of transceivers for WDM transmission systems. Additionally, an inverse designed de-multiplexer can be robust to fabrication errors, as well as temperature changes within the environment [27]. WDM technology has been heavily investigated using inverse design, allowing us to broaden our scope of what inverse design is capable of.

Photonic switches are growing in interest as there is a need for higher broadband interconnects as east-west traffic increases. Within a data center, optical switches can be used for routing, cross-connection, switching processes and add/drop functions. The optical switch has been presented in a wide range of technologies, from free space switching to III-V based optical switches. For compactness and efficiency, it is important to utilize on-chip technology for switching as on-chip switching reduces the risk of high optical insertion loss. On-chip switching in photonics has generally relied on Mach-Zehnder Interferometers (MZI), integrated Micro-electromechanical systems (MEMs) switches, Micro-Ring Resonators (MRRs) [28] and directional couplers [29]. However, these switches require large footprint on the die, with MZI often being 100 μm [30] in length and MEMS switches taking upwards of 100 μm [31]. For a 2×2 thermal-based MZI-based switch, there is a significant area of the chip to be heated, with the length of the heater being approximately 100 μm in length, hence increasing the power requirements for optical switching [32]. Additionally, directional couplers and MRRs are sensitive to fabrication errors, which can have significant impact on their optical transmission performance and their wavelength spectrum. Decreasing the footprint of our optical switches, increasing the efficiency of optical interconnects and having fabrication robustness can be further explored using inverse design. On-chip switching using inverse design has not been previously explored, and similar to the thermo-optic MZI-based switch, we can utilize silicon temperature dependence to accomplish switching behaviour in a novel inverse designed device.

In this thesis, inverse design methodologies are used to create a thermally sensitive 1×2 optical space-based switch. Due to the nature of silicon and its large temperature coefficient, we generate a silicon-on-insulator based design, using silica as the insulator. Similar to the thermo-optic MZI-based switch in the work of Chen, *et al.* [32], this thermal switch achieves a 1 dB insertion loss at its two output ports dependent on temperature changes. However, in terms of footprint efficiency, the optical switch presented in this thesis has a 7.5 μm by 2.5 μm footprint, where the MZI-based switch is upwards of 100 μm in length [32]. Throughout this thesis, we present not only the novel thermal-based optical switch in SOI technology, but the complete design methodology and the used open-source code to create thermally sensitive novel devices, as well as the limitations to fabrication and testing across various Finite Domain Time Domain (FDTD) simulations. By generating a thermal switch using inverse design, we present a novel approach to optical switching as well as push performance capabilities in the realm of inverse design.

Chapter 2

Background

2.1 Optical Modes and Waveguides

In photonic design, it is imperative that the engineer understands and can model the propagation of light through a medium. Although it is straightforward to describe light using ray optics at large scales ($>>$ λ, where λ is the optical wavelength), a more complete description of light is given by Maxwell's equations, especially for sub-wavelength devices where the wave nature of light and polarization are accounted. A non-conducting material with no free charges can be described by Maxwell's equations:

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

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

$$\nabla \cdot \mathbf{D} = 0 \tag{2.3}$$

$$\nabla \cdot \mathbf{B} = 0 \tag{2.4}$$

where **E** is the electric field vector, **H** is the magnetic field vector, **D** is the electric displacement vector, and **B** is the magnetic flux density vector. Additionally, ϵ is the permittivity of the material and μ is the permeability of the material [33]. With some manipulation, we derive a second ordered equation known as the wave equation:

$$\nabla^2 \mathbf{E} - \mu\epsilon \frac{\partial^2 \mathbf{E}}{\partial t^2} = 0 \tag{2.5}$$

Of which, we seek to find a solution of the following form:

$$\mathbf{E}(x, y, z, t) = \mathbf{E}(x, y) e^{j(\omega t - \beta z)} \tag{2.6}$$

where β is the propagation constant with units radians/meter and is described as the z-component of the wave vector. The wave vector describes the physical properties of a wave such as propagation direction and cycles per distance. ω is the carrier frequency and **E**(x,y) is the transverse field distribution.

To understand how these equations relate to photonic device design, we can utilize the infinite slab waveguide as an example. The slab waveguide extends infinitely in the y-z direction and is demonstrated in fig. 2.1.

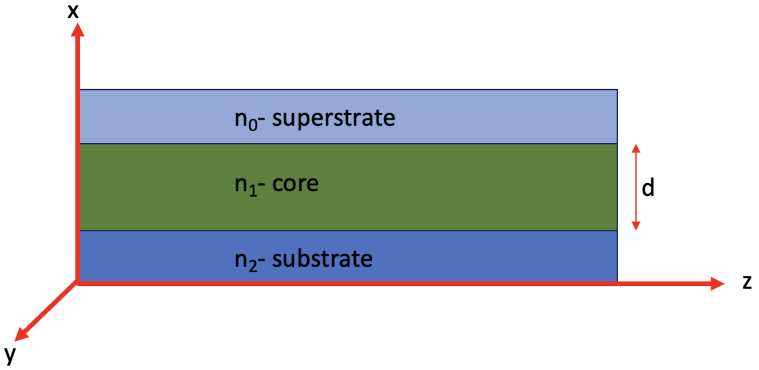


Figure 2.1: Infinite slab waveguide with a thickness of d. The indices of refraction of each material are n_0 for the superstrate or top cladding, n_1 for the waveguide core and n_2 for the substrate or bottom cladding of the waveguide. For guided light in the waveguide $n_1 >> n_2, n_0$.

As light propagates along the z-axis, we can solve for the transverse electrical (TE) modes. TE modes are the solutions to Maxwell's equation where the E-field is transverse to the plane of incidence, knowing that $\mathbf{E} = E_x\hat{x} + E_y\hat{y} + E_z\hat{z} = E_y\hat{y}$ by defining $E_x =$

$E_z = 0$ since the electric field is assumed to be perpendicular to the x and z-axis. For TE modes, the scalar wave equation can be defined as:

$$\frac{d^2E_y}{dt^2} + (k_0^2n^2 - \beta^2)E_y = 0 \tag{2.7}$$

where k_0 = $\frac{2\pi}{\lambda}$ and *n* is the refractive index. All guided modes will have a bounded propagation constant, i.e., $k_0^2n_1^2 \geq \beta^2 \geq k_0^2n_0^2$. Given these relations, the possible modes that are supported within a waveguide with a thickness of *d*, substrate index value of n_2, superstrate index value of n_0 and core index value of n_1, are defined by the dispersion relation which is an eigenvalue function:

$$tan(\kappa d - m\pi) = \frac{\kappa d(\delta d - \gamma d)}{\kappa^2d^2 - \gamma\kappa d^2} \tag{2.8}$$

where $\kappa^2 = k_0^2n_1^2 - \beta^2$, $\gamma^2 = \beta^2 - k_0^2n_2^2$ and $\delta^2 = \beta^2 - k_0^2n_1^2$. For each value of *m* (*m*=0,1,...), a discrete β is found corresponding to solutions of the TE modes.

The slab waveguide is a simple structure that can be computed numerically. Solving Maxwell's equations on more complex structures becomes increasingly difficult as the length and width changes multiple times in a device, with device boundaries changing to reflect different functionalities such as in a Y-branch. With structures such as buried straight waveguides and ridge waveguides, we can use mathematical methods such as Marcatili's method and the effective index method. However, many devices are much more intricate, composing of turns and small features that make these methodologies nearly impossible to use as the *d* value is constantly changing. Hence, we must turn to simulation models to accurately predict how the electric field will propagate within these structures.

2.2 Simulations based on Finite Difference

When performing Maxwell's equations on more complex structures, it is nearly impossible to solve using mathematical methods such as the effective index method [34]. However, when looking at smaller components of the design and solving Maxwell's equations at these discrete points, we can come up with a great approximation of how light propagates through the structure. Finite Difference Time Domain (FDTD) and Finite Difference Frequency Domain (FDFD) simulations have been used quite extensively in recent years for predicting and modeling photonic structures. These simulations overlay a grid like structure, known as a mesh, over the structure and solve Maxwell's equations at each mesh point, hence the term finite difference. This type of numerical method is used quite extensively in electronic and photonic simulations.

FDFD and FDTD simulations are quite similar in nature, relying on the Finite Difference method to solve Maxwell's equations for accurate models of light propagation throughout a simulation region. Despite this, there are still some significant differences that must be taken into account when deciding which simulation is best for device simulation. In an FDFD simulation, the entirety of the structure is solved at once, whereas in an FDTD simulation the fields are solved iteratively in small time steps and updated accordingly. FDFD always returns a steady state solution with one wavelength being observed, whereas FDTD always returns a transient solution over a spectrum of wavelengths [35]. Additionally, they are comparable in terms of computation time scale, with both returning a result in the 2-D domain in a matter of seconds. However, the FDTD simulation has an advantage in result times, where we can see a spectral response in the same time it takes a FDFD simulation to return a singular wavelength solution. Indeed, for performance validation, the spectral response is an importance metric.

Beyond the equation derivation that the simulation obtains at each mesh point, the simulation also gives second-order accuracy by utilizing the Yee method. The Yee method, as illustrated in figure 2.2, is implemented on a photonic structure by taking each mesh cell and computing all components of the electric and magnetic field on a slight offset from each other in the same Yee cell, which returns a highly accurate estimate for the E and H fields [36].

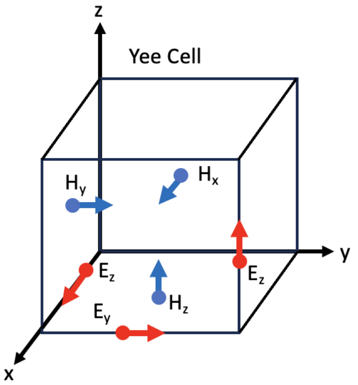


Figure 2.2: A Yee cell with the evaluation points for the H and E field marked. The H field is offset by $+\frac{1}{2}$ of the E field coordinates.

The Yee methodology, which is intrinsic to most FDTD simulations, assumes that the material within each cell is uniform. However, when we have a uniform finite mesh, we can assume that unless we are working with a rectangular waveguide, there will be discontinuities at the boundaries of the device. One way to compensate for these discontinuities is to reduce the mesh size. Reducing the mesh size offers a more accurate view of the EM wave as it propagates through the medium, but is significantly more costly in computation. In the commercial design tool by Lumerical, a service offered by Ansys, when using a uniform mesh the computing time is proportional to $1/dx^4$, where *dx* is the mesh size. Hence, this creates a trade-off between the computational cost and the mesh size in these simulations.

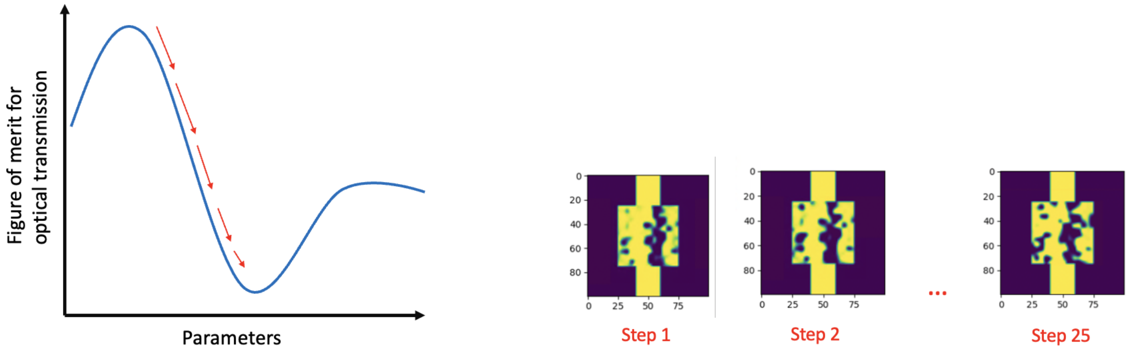
Discontinuities also lead us to averaging methodologies and non-uniform meshing within different FDTD simulation. Meshing and averaging at the boundaries can be important when determining reflection and loss, specifically at those boundaries with significant refractive index changes. Many FDTD simulations offer various ways to increase accuracy for electric and magnetic field estimations near these permittivity borders. Lumerical offers non-uniform meshing that reduces the size of the mesh near the borders to get a more precise shape estimate in the meshing. Additionally, Lumerical offers various types of averaging for refractive index changes, from as simple as volume averaging of the materials at the border to more complex averaging schemes [37].

Overall, simulations for EM waves have greatly increased in computational capacity over recent years. By utilizing these instruments in our design of photonic devices, we can accurately test these devices before fabrication to optimize their design and produce the best devices possible. Beyond pre-fabrication testing, we can utilize FDTD and FDFD simulations in the field of inverse design, which can generate non-intuitive designs for novel and compact devices.

2.3 Inverse Design

Machine learning, which utilizes statistical methods and numerical algorithms to predict patterns in data or perform parameter optimization in design work, has the ability to transform engineering design methodologies. Inverse design, a sub-sect of machine learning, can be found over a considerable number of fields, offering streamlined design processes. Inverse design in photonics is a design methodology that began with parameter sweeping in physics-backed designed devices to find the optimal device performance. Now, photonic inverse design has evolved to use gradient descents and deep-learning for device design. In this thesis, we will focus on iterative optimization processes, more specifically the gradient method, to produce device designs which fulfill our target functionalities. Iterative optimization processes typically rely on gradient descent methods which are heavily utilized across machine learning algorithms and systems.

The gradient method can take on two forms: 1) classification which produces prediction outputs that are categorized, and 2) regression which produces numerical predictions [38]. The regression form of gradient method is utilized for photonic inverse design as our device parameters (such as permittivity values of the material or boundary coor-



(a) (b)

Figure 2.3: (a) Illustration of a generic gradient descent to find the local minimum corresponding to the maximum optical transmission for a given design. The arrows show the steps taken to find the minimum of the function. **(b)** Device example showing how the device boundaries and permittivity change with each gradient descent step.

dinates) are numerical. To perform the gradient method, we must define a figure of merit (FOM) function which defines the objective of our optimization. For photonic inverse design, the FOM can represent desired performance at the output of the device such as optical transmission or optical mode overlap. To find the best parameters for the desired performance, we must find the FOM maxima or minima based on changes in parameters. To find a maximum or minimum of the FOM, we take the gradient of the function to move along the slope in small steps (fig. 2.3) such that we converge to a solution when the size of the step, the magnitude of the gradient, nears zero. To calculate the gradient of a function, we find the derivative of the FOM with respect to the parameters of the function, the permittivity values of the device and cladding. For example, we have the permittivity values altered at each step such that the device is optimized to a local minimum that represents the maximum transmission at the output. Additionally at each step, to evaluate the performance of the device, an electromagnetic (EM) simulation, typically an FDTD or FDFD simulation which solves the Maxwell's equations for the simulation region, is performed which evaluates the device performance against the desired performance.

To perform the optimization process for the desired functionality, we must first define our FOM. The figure of merit for most inverse designed photonic devices can be described generally as:

$$\min_{p} f_{obj}(\mathbf{E}(\epsilon(p))) \tag{2.9}$$

subject to:

$$p \in S_{fab} \tag{2.10}$$

where f_{obj} is the objective function of the optimization, **E** is the electric field and ϵ is the permittivity values of the design space, which is subject to p, the parameter that defines fabricable devices. Many devices will have small features ($<$ 100 nm) which cannot be fabricated given modern photonic foundry limitations. S_{fab} is a subset of all devices that have features greater than 100 nm such that we only produce devices in our optimization that can be fabricated [27]. This 100 nm minimum feature limitation is built into the optimization process such that all design solutions have features larger than the minimum feature size.

In this thesis, I will focus on topological optimization which breaks the design region into a series of pixels, similar to a mesh, and allows the permittivity of each pixel to vary [39]. What this implies is that within each pixel, the permittivity can exist in a range between 0 and 1, which is the normalized permittivity of the two materials of the device representing the core and cladding. In figure 2.4, the design space with various permittivity values that fall between silicon and silica is illustrated. To begin the optimization process, each pixel is assigned a randomized value between 0 and 1.

The randomized values can be generated by the code for each pixel and are generally pushed towards a central value. The central value of each pixel assures that we have an unbiased device, with the device not initialized to include more silicon or silica due to the random number generator. For example, in the image shown in figure 2.4, the

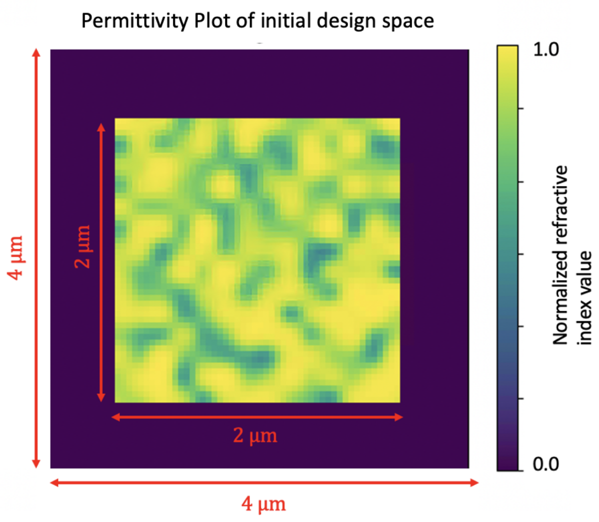


Figure 2.4: The initial design region permittivity values, where the permittivity values range between silicon and silica but are generally close to the central value between the two.

randomized values are pushed towards a more central value by using this equation:

$$rand \times 0.2 + 0.5 = \epsilon \tag{2.11}$$

where *rand* is a random number between 0 and 1 generated by the code. The 0.2 for this structure pushes the values between 0 and 0.2, with the addition of 0.5 pushing the initial values to 0.5 to 0.7 for initial normalized permittivity ranges. Having the initial values fall between the 0.5 to 0.7 range creates a slight bias towards silicon, as the values are closer to silicon (1) than silica (0). This initial condition can be changed and altered by the users, as higher bias towards silicon or silica can yield different device performances given the different material properties. For example, the refractive index of silica is less impacted by temperature so for a temperature robust device, one might want to increase the amount of silica in the device, hence having a silica bias at the initial conditions. Once the initial design is constructed, input and output waveguides can be added and a source can be set up. In figure 2.5, an input waveguide and output waveguide of 800 nm in width are added to the permittivity array.

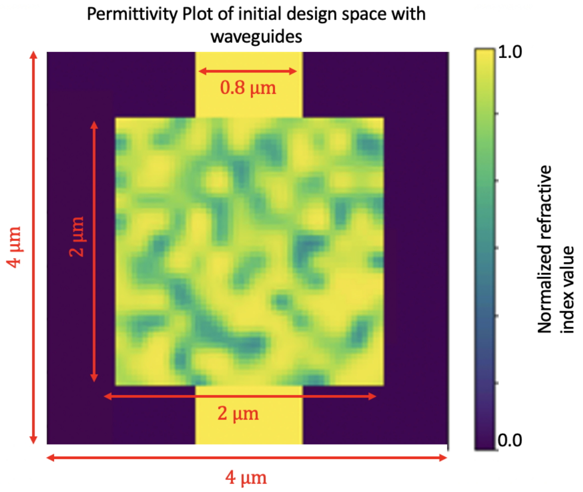


Figure 2.5: The randomized initial design region, with the input and output waveguides added.

Once the permittivity values in the simulation have been set, the simulation and the objectives of the simulation can be defined within our code. A single optical mode source is placed at the input waveguide (the top waveguide in figure 2.5) and given a specified wavelength (generally 1310 or 1550 nm for our inverse designed communication devices). The objectives of the device are what drives our optimization to a solution. The objectives are commonly mode or transmission based, with the optical modal overlap function being evaluated at the output to evaluate the transmission of the target mode. The modal overlap function for a specific mode at the output is defined as:

$$F_{EM} = -|\mathbf{E}^{T}_{\textit{target mode}}\mathbf{E}|^2 \tag{2.12}$$

where T is the conjugate transpose, with $\mathbf{E}^{T}_{\textit{target mode}}\mathbf{E}$ defined as the electric field (**E**) overlap with the target modal overlap ($\mathbf{E}_{\textit{target mode}}$). The minus sign is utilized as we are minimizing the FOM during the optimization.

For many of our inverse designed devices, we use transmission overlap for a specified mode (e.g., the fundamental mode TE0) as our FOM, which is defined as:

$$F_{EM} = (1 - |\mathbf{E}_{TE0}^{T}\mathbf{E}|)^2 \tag{2.13}$$

With the overlap term being subtracted from 1 because we are minimizing it during the optimization. Hence, the closer the modal overlap gets to 1 the smaller the value of the function will become, converging close to 0.

Utilizing inverse design for photonic design methodology increases our capacities for device performance at small scales by exploring the entire design space. Knowing material properties, such as wavelength refractive index dependence and temperature refractive index dependence, the optimization can be set up to be robust or sensitive to these proprieties. Additionally, inverse design can generate devices that are robust to fabrication errors, making it a helpful tool for designers. To further explore the capabilities of inverse design, we look to the temperature dependence of silicon.

2.4 Temperature and Refractive Index

The relation between phase and velocity through the refractive index of the medium is well defined in the field of optics. The relation of the refractive index with light velocity in a material and light velocity in free-space is defined below as:

$$n = \frac{c}{v} \tag{2.14}$$

where *n* is the refractive index, *c* is the velocity of light in free-space and v is the velocity of light in a given material. Additionally, we know that the refractive index is dependent on wavelength as described by the Sellmeier equation:

$$n(\lambda)^2 = 1 + \sum_{k} \lambda^2 - \lambda_k^2 \tag{2.15}$$

where λ_k is a Sellmeier coefficient that corresponds to the resonant wavelength [40]. Hence, any changes in the wavelength will cause a change in index of refraction. By utilizing the wavelength dependence of the refractive index, the phase and velocity of the light is changed causing a light path difference [41]. We can utilize this phenomenon to design multiplexers using inverse design by running an optimization at two different wavelengths.

Similarly, we can change the refractive index of a material by altering the temperature, which is known as the thermo-optic effect. The amount by which the refractive index changes is quantified by the material temperature coefficient, which is present in all optical materials [42]. The thermo-optic effect has a wide range of applications, from thermally tuning ring resonators to generating thermal optical switches. Thermally tuned ring modulators utilize the fact that ring modulators are sensitive to temperature fluctuations (as low as ΔT = 1 Kelvin) which can cause a significant shift to the resonating wavelength. By utilizing the thermo-optic effect, the ring can be tuned such that it operates at a specific temperature that corresponds with the desired operating wavelength [43]. Additionally, thermo-optical switches, such as ones based on directional couplers, use the change in the propagation constant from temperature to cause light switching between two waveguides. Within the field of inverse design, there has been an emphasis on creating temperature robust devices. Temperature robustness is especially important in terms of wavelength dependent devices, with temperature changes in the surrounding environment capable of creating transmission discrepancies and wavelength spectrum shifts. To take into account temperature dependence and create devices that are temperature robust, we typically use the FOM below [27]:

$$\min_{p} f_{obj}(\mathbf{E}(\epsilon_1(p))) + f_{obj}(\mathbf{E}(\epsilon_2(p))) \tag{2.16}$$

where f_{obj} is the transmission FOM defined in equation 2.13. The permittivity, which is described as $\epsilon = n^2$, has two different values corresponding to their temperature change.

ϵ_1 is the permittivity at one temperature and ϵ_2 is the permittivity at a different temperature. In general, the refractive index change due to temperature can be described as:

$$\Delta n = \tau \Delta T \tag{2.17}$$

where Δn is the change of refractive index, τ is the temperature coefficient of the material, and ΔT is the change in temperature from room temperature (which is assumed to be approximately 297 K). τ is dependent on a few properties of the material and is defined below [41]:

$$\tau = \frac{(n^2-1)(n^2+2)}{6n}\left(\frac{1}{\alpha}\frac{d\alpha}{dT} - 3\gamma\right) \tag{2.18}$$

where α is the materials polarizability, with units $\mathrm{C} \cdot \mathrm{m}^2 \cdot \mathrm{V}^{-1}$, and 3γ is the thermal expansion factor, with units K^{-1}. The polarizability of a material describes how a materials molecular makeup changes given external electric fields. Additionally, the thermal expansion factor describes how the material density changes with a given temperature change. This equation defines the shift in refractive index given temperature, change of density, and polarizability of the material.

Knowing the changes in the refractive index (Δn) of the material allows us to optimize a given FOM over a given range of ΔT. For example, we can generate a temperature robust wavelength de-multiplexer by optimizing the device at both low and high temperatures such that the changes in refractive index do not impact device transmission. However, using thermal sensitivity in the field of inverse design has been scarcely studied and will be discussed and utilized in the design of a novel device within this thesis. By studying the thermo-optic effect in waveguides and utilizing this within inverse design methodologies, we can generate new devices and push the limits of modern day inverse design within silicon photonics.

Chapter 3

Simulations Methodologies

There are a handful of inverse design numerical tools, ranging from open to closed source, user friendly to code intensive. Ansys/Lumerical offers an inverse design tool suite, named LumOpt, but is generally a more closed source code [44]. By utilizing an open-source numerical tool, the user has many more degrees of freedom, such as the ability to choose an EM simulation for FOM evaluation during optimization or the ability to vary the objectives and initial conditions of the optimization process. Stanford University's Nanoscale and Quantum Photonics Laboratory offers a free open-source design framework for photonic inverse design [45]. The Stanford Photonics INverse design Software (SPINS) allows the user to explore various design parameters such as initial conditions and the formulation of FOMs. A gradient-based form of inverse design, the SPINS framework allows the user to customize the design framework (e.g., design nodes that build up the entirety of the SPINS framework) as well as the initial conditions. Due to ease of modification for users of SPINS as a numerical tool, it was used in this research to generate novel inverse designed photonic devices.

In figure 3.1, I present a design methodology for generating a device structure, exporting a device, and testing across FDTD and FDFD simulations. Using SPINS FDFD in this work was imperative as it offered an open-source code that could be easily modified. However, it is equally important to utilize the FDTD simulation offered by Ansys/Lumerical as it offers a better simulation interface and a spectral output rather than an output at a singular wavelength. Seeing the transmission spectrum of the device allows one to view any wavelength shifts that can occur during the binarization step in the exportation from SPINS to the layout file. Additionally, comparing Ansys/Lumerical FDTD to SPINS FDFD gives a final check of the device, as to see that any modifications made to the SPINS code does not impact the integrity of the FDFD results. It is important to understand fully each step of the optimization, device exportation to layout, and simulation testing such that we can be ensured that the devices perform well.

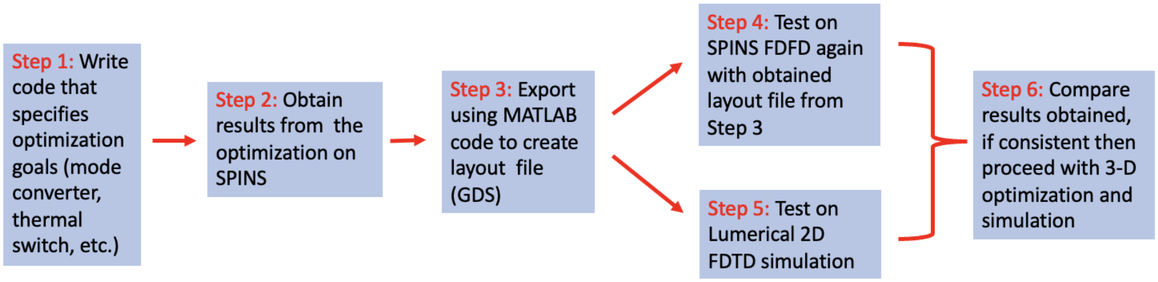
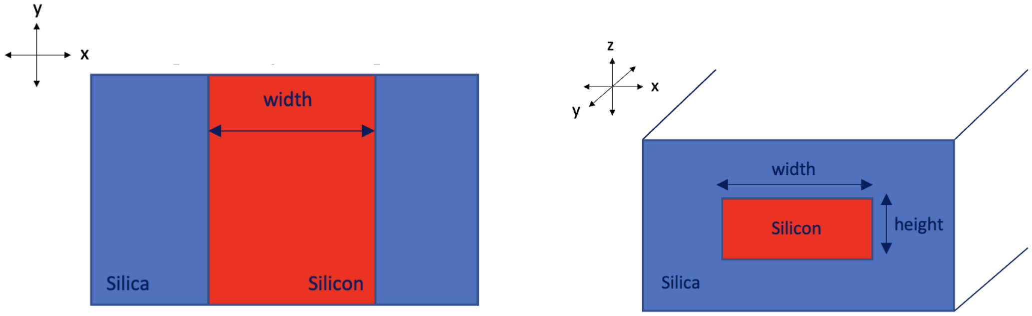


Figure 3.1: A well-defined methodology for exporting from open-source SPINS simulation platform to commercial Ansys/Lumerical tool to obtain simulation results.

3.1 Two versus Three Dimensional Simulations

To initiate the optimization for a new device, it is typical to use a methodology that is first based on a two-dimension (2-D) optimization. A 2-D optimization relies on a 2-D FDTD/FDFD simulation to determine how close the optimizer is to realizing the desired function. 2-D simulations, where only two dimensions of the waveguide are considered, are much faster to run compared to when all three dimensions (3-D) of the waveguide structure are considered in simulation. Indeed, the computational time is cut when solving Maxwell's equations only has to be performed on a two-dimensional device where only the permittivity of the x-y dimensions of the device (height of the waveguide and propagation direction) are considered as opposed to a three-dimensional device which includes the thickness of the device (z-direction). From chapter 3 onward, the z-axis will no longer be the axis of propagation, instead the y-axis will be used. When optimizing fabricable devices, it is important to run the optimization over the three dimensions. As a proof of concept, two-dimension simulation can be incredibly helpful in assessing the device but comes with drawbacks in its simulation accuracy as the depth of the device is not taken into account. When optimizing a device, the figure of merit is evaluated at each step utilizing the FDFD simulation. This simulation being in 2-D or 3-D can have significant consequences on how the fundamental mode is defined, the propagation of the light, and the modal overlap. In this section, there is an in-depth discussion on how the infinite z-component of the device assumed in 2-D simulation significantly impacts simulation results which in turn impacts the optimization of the device at each step.

The fundamental mode of the device in SPINS is defined as the optical mode (i.e., one of the discrete solutions to the wave equation) with the largest propagation constant. To find the propagation constant, we first derive the effective index, which is defined as the phase delay per unit distance in a material compared to the phase delay in a vacuum (or free-space). Using the effective index method [46], we derive the propagation constant for supported transverse electrical/magnetic (TE/TM) modes at the input waveguide. For a standard SiPh single-mode waveguide, the input waveguide can be defined as being 0.5 μm wide by 0.22 μm thick. However, when calculating the effective index in two dimensions, the thickness of the device is assumed to be infinite due to the removal of the z-direction in the simulation. The infinite thickness of 2-D simulations generates a completely separate problem from the 3-D simulation, which has a thickness of 0.22 μm. To determine the supported modes in 2-D simulations and 3-D simulations, one must calculate the effective index for the waveguide with 0.22 μm thickness and with infinite thickness. From these calculations, one can see that there are significant consequences from assuming infinite thickness for a waveguide and how the propagation constants of the modes are impacted. The 2-D waveguide is illustrated in figure 3.2**(a)** while the 3-D buried waveguide is illustrated in figure 3.2**(b)**.



(a) **(b)**

Figure 3.2: **(a)** The top view of a rectangular buried silicon 2-D waveguide. **(b)** The cross-section of a rectangular buried silicon 3-D waveguide.

The width for both the 3-D and 2-D simulations is defined as 0.5 µm. In the 2-D simulation (fig. 3.2**(a)**), the depth is not defined as it is assumed to be infinite. From these depth variations, the propagation constants can be calculated for infinite and 0.22 µm depth by first solving the 3-D waveguide. Using the effective index method, the 3-D waveguide can be broken down into two 2-D problems as seen in figure 3.3.

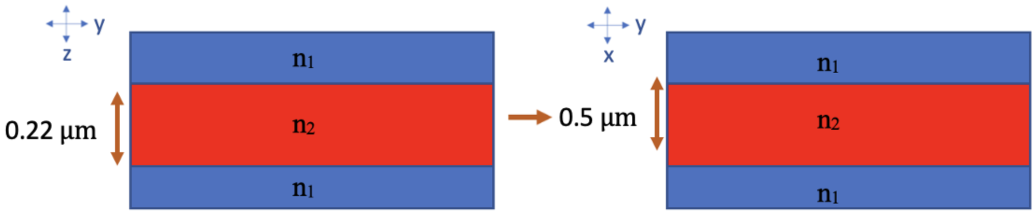


Figure 3.3: The process of dimension changes used in the effective index method where the problem is broken down into two 2-D problems.

Where the y-axis is the direction of the propagation. The refractive index of silicon (n_2) is 3.45 and the refractive index of silica (n_1) is 1.44. Additionally, the wavelength is defined as 1.55 μm which is standard in the telecommunications spectrum [47]. From these material specifications, the effective index can be solved for the buried waveguide using the effective index method. By solving for the effective index, the propagation constant and coefficient of the waveguide can be solved to determine the transverse field fundamental mode of the waveguide.

To solve for the effective index, we first calculate the normalized frequency of the waveguide, which is defined as:

$$V = k_0 d\sqrt{n_2^2 - n_1^2} \tag{3.1}$$

where k_0 is defined as the wavenumber:

$$k_0 = \frac{2\pi}{\lambda} \tag{3.2}$$

where λ was previously defined as 1.55 μm. Using V, the normalized propagation constants for TE modes is computed by using the normalized b-V curves that relates the propagation constant to the waveguide dimensions from the eigenfunction derived from the wave equation. The b-V curves illustrated in figure 3.4 represent the following dispersion relation function:

$$V^2 = k_0^2 d^2 (n_1^2 - n_2^2) \tag{3.3}$$

which can be re-written in terms of the b-parameter as follows:

$$\tan[V\sqrt{1-b} - m\pi] = \frac{\sqrt{1-b}(\sqrt{b} + \sqrt{b+c})}{1 - b - \sqrt{b}\sqrt{b+c}} \tag{3.4}$$

where *m* is the integer corresponding to the mode number and c is the asymmetry parameter. As this is a symmetrical waveguide, the value of c is equal to zero.

From the obtained b-V curves, the corresponding normalized propagation constant *b* is found where the normalized frequency intersects with the b-V curve, where *V* is equal to 6.423 obtained from equation 3.3. Given *b*, the effective index can be found which is

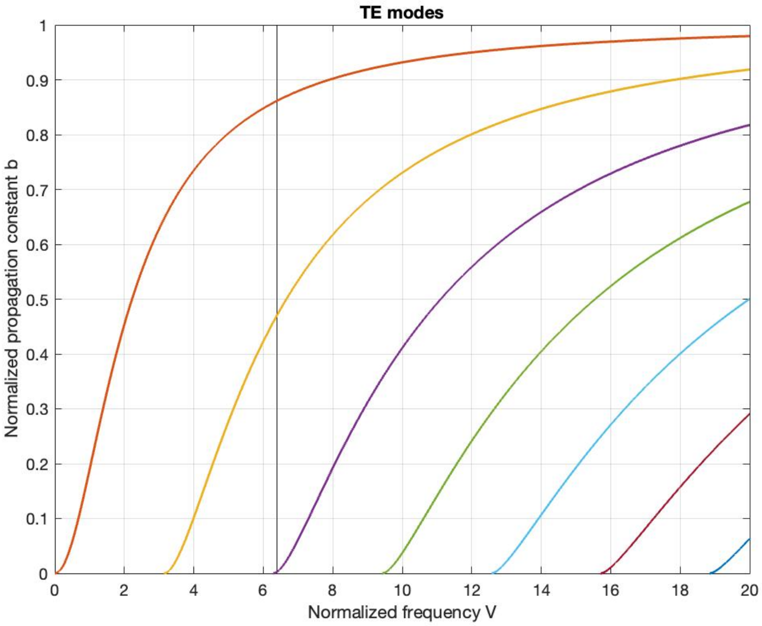


Figure 3.4: The b-V relation for the 3-D buried waveguide for TE modes. The vertical line corresponds to the normalized frequency V value of the rectangular waveguide given the known values of k_0, d, n_1 and n_2 for the given waveguide.

then used to calculate the propagation coefficient (rad/m).

$$b = \frac{n_{eff}^2 - n_1^2}{n_2^2 - n_1^2} \tag{3.5}$$

$$\beta = k_0 n_{eff} \tag{3.6}$$

where n_{eff} is 3.266 for the fundamental TE mode. Making the propagation coefficient, β, equal to 13.240 rad/m. For the TM modes, we use the corresponding eigenfunction to obtain a b-V graph and find the propagation coefficient, β, similarly, which is 7.805 rad/m. Hence, for 3-D optimization and simulation in the SPINS open-source platform, the TE mode is the dominant optical mode since it is faster and therefore is designated as mode number m=0, i.e., TE0.

The 3-D simulation approach was discussed first to show that in the 3-D realm, the fundamental TE mode is dominant, as is standard for 0.5 μm width, 0.22 μm depth waveguides. When calculating the propagation constant for the 2-D simulation, however, the depth of the waveguide changes from 0.22 μm to infinity. When increasing the depth of the waveguide, the expected dominant optical mode changes and the TE0 optical mode is no longer dominant. To confirm this change, we change the depth from 0.22 to 1 μm in the effective index calculation using the method described above, sufficiently large for these calculations to be an approximation to infinity in the 2-D SPINS. This shows any inconsistencies in outputs from 2-D to 3-D simulations, where 2-D has been approximated with the 1 μm depth, concerning transmission and modal properties given the increased depth of the device. The propagation coefficient/constant remains the same for the TE optical mode since this change in depth has no impact. However, for TM modes, the light polarization experiences the impact of the depth dimension and there is a much different calculation than before. The normalization frequency equation becomes:

$$V = k_0 w \sqrt{n_{eff}^2 - n_1^2}, \tag{3.7}$$

where w = 1 μm. The corresponding b-V curves for TM modes are illustrated in figure 3.5. Where b= 0.934, when calculating the propagation coefficient, there is a higher value for the effective index for TM modes than for TE. Hence, the propagation constant for the TM modes is significantly larger than for TE modes and the TM fundamental mode will become the dominant mode.

One can further demonstrate the differences in the propagation constant obtained for the fundamental mode between the 2-D and 3-D using Ansys/Lumerical simulation tools which calculate the fundamental mode for a modal source automatically. A modal source in simulation is a light source that injects supported optical modes, chosen by the user, into the waveguide. For this exercise, a well-tested Y-branch device is used as an example which splits the optical signal by two at its output ports. This device shows that the infinite-z assumption is consistent across simulations and that the obtained fundamental

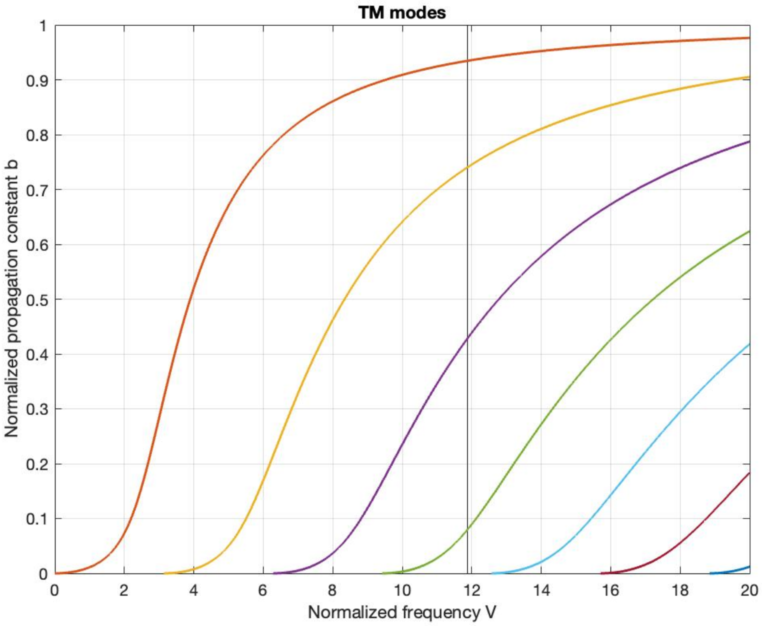


Figure 3.5: The b-V curves for the approximated 2-D simulation for TM modes. The vertical line corresponds to the normalized frequency V value of the rectangular waveguide given the known values of k_0, d, n_1 and n_2 for the given waveguide.

mode in Ansys/Lumerical and SPINS are the same. The structure layout of the Y-branch is offered on the Ansys/Lumerical website (fig. 3.6). It is a compact, low-loss, wavelength insensitive device that performs well experimentally. This particular case of inverse design deals with parameter sweeping at the border, where each width value, w, is a different point of optimization. Additionally, each w point is connected with a smoothing algorithm. An inverse designed Y-branch is discussed as this thesis covers inverse designed devices and this Y-branch has been experimentally validated extensively [48].

Quantifying differences for more complex structures, such as the Y-branch and other inverse designed devices, between 2-D and 3-D simulations is more straightforward through FDTD simulations. For future considerations of these inverse designed devices, it is more efficient to use the Ansys/Lumerical FDTD simulation tool. For SPINS FDFD, 3-D simulations requires the use of a GPU server due to the intensive computation required. The

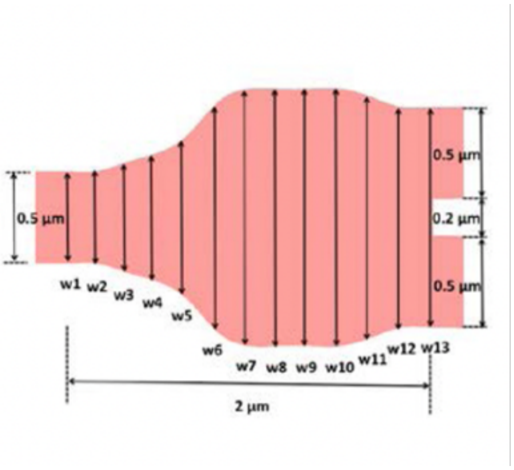
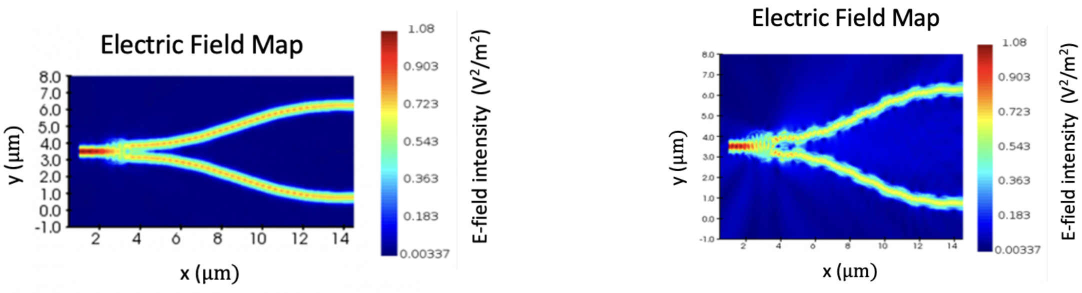


Figure 3.6: Schematic of the inverse-designed Y-branch provided. Reprinted with permission from [48] © The Optical Society.

FDTD simulation offered through Ansys/Lumerical can more smoothly go from 2-D to 3-D simulations, sufficient for providing the difference in results. Ansys/Lumerical not only offers a computationally efficient 3-D simulation, but provides calculations of the supported optical modes in the waveguide with their corresponding propagation constants and optical modal field cross-sections. Understanding which mode is the first fundamental mode is critical to this work and Ansys/Lumerical's modal calculations offers an additional confirmation of the modal calculations. In the results obtained through 2-D simulation, the field is very uneven and choppy, however in the 3-D range the field is smooth (fig. 3.7).



(a)

(b)

Figure 3.7: **(a)** Field intensity of the Y-branch obtained through 3-D simulation in Ansys/Lumerical. **(b)** The same structure simulated in 2-D.

The noticeable differences in the results obtained is due to the existence of higher order optical modes in the 2-D simulation from the infinite z-component. The increase in the waveguide dimensions is increasing the modal crosstalk between all existing optical modes in the device leading to the visually choppy field profile in fig. 3.7**(b)**. To confirm that this is caused by the existence of higher ordered modes, the depth dimension of the waveguide in the z-direction is increased for the 3-D simulation to simulate something approximately similar to an infinite depth. Note that it is important to discuss the existence of higher ordered modes in 2-D simulations as it has a significant impact on the field profile and the optimization of the inverse-design structure. Knowing that the figure of merit is evaluated using this field profile, one can see through this work that 2-D and 3-D are operating in vastly different conditions. In the 2-D optimization problem, the design will not optimize in a similar manner to the 3-D as the respective fundamental modes are not the same and the 2-D simulation has the presence of higher order modes. Understanding this further led to the conclusion that 2-D optimizations lead to good proof of concept in this investigation, but are not reliable for fabrication.

For the infinite-z approximation, the depth of the waveguide on Ansys/Lumerical is changed from 0.22 μm to 1 μm. Changing the depth of the structure lead to the existence of higher ordered modes, given that larger structures lead to more discrete solutions to the wave equation, i.e., more supported optical modes by the waveguide. In fig. 3.8, the 3-D simulated structure now has a noisy and non-uniform electric field intensity profile. Looking at the normalized optical transmission in figure 3.9 of each optical mode in the Y-branch, observe that the presence of the TM0 mode is likely causing the interference in the field graph.

Having the fundamental mode change between TE0 and TM0 due to dimension variations in the 2-D and 3-D simulations is an important distinction that must be taken into account. In the design of the device when comparing the optical transmission between 2-D optimized designs and 3-D optimized designs, it is important to define the same mode from a light source across simulations hence not to compare a TM0 to a TE0 device.

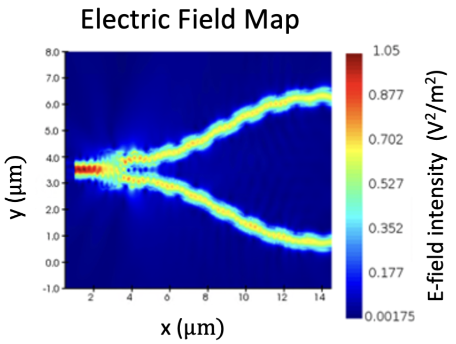


Figure 3.8: Results from 3-D simulation using Ansys/Lumerical of the Y-branch with the depth of the waveguide set to 1 µm instead of the standard 0.22 µm.

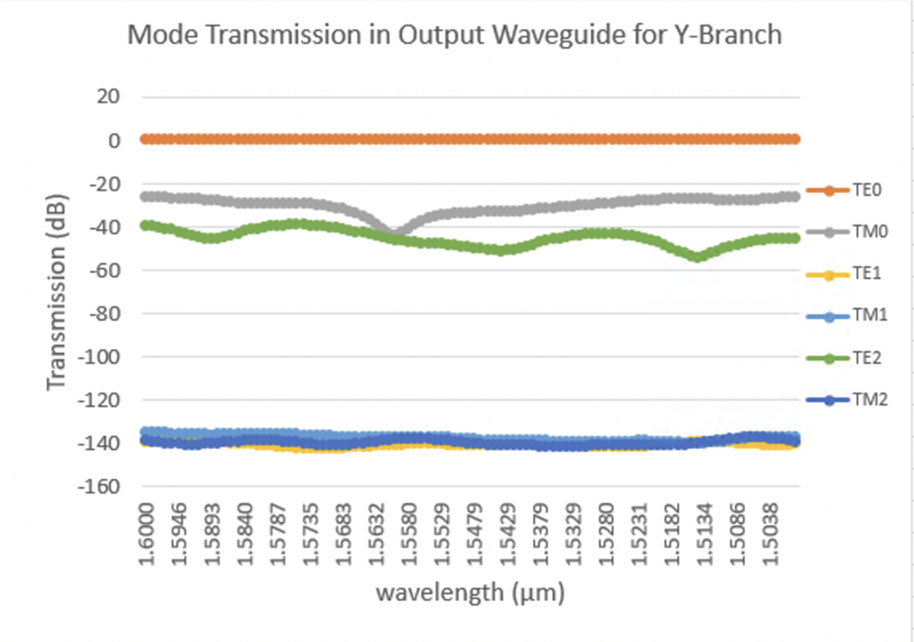


Figure 3.9: The normalized optical transmission of each existing mode in the Y-branch structure.

Additionally, to reduce computational cost, the 2-D simulation for the optimization can be used in a low binarized device before using the 3-D optimization with higher binarization. These concepts will be further discussed in chapter 5. Understanding this concept also ensures the accuracy of the optical mode source settings across multiple simulations so that the usage of TM or TE modes on the modal source stays consistent throughout simulations. It is also important to note that when going from 2-D to 3-D simulations, the

2-D optimized device may look significantly different than a 3-D optimized device due to not only the difference in dimension but the transition of the fundamental mode polarization from TM and TE. Finally, understanding the differences between 2-D and 3-D simulations for FDFD and FDTD simulations is imperative to simulation testing which is verified across SPINS and Ansys/Lumerical simulation platforms.

3.2 Optimized Design to its Layout

Grey-scale optimization is an optimization that allows continuous change between silicon and silica in terms of permittivity values at each pixel. The resulting structures exhibit low binarization with several pixels having in-between permittivity values. These permittivity values are mathematical values that cannot be fabricated since they are not representative of real optical materials. As such, an inverse designed device with non-binarized permittivity values must be made into a design that is completely composed of silicon and silica and therefore fabricable. To do this, the device can be exported to a layout file with a matrix that represents the permittivity values of the device at each pixel point and establishes a hard level-set between silicon and silica. The resulting binarized device must then be validated through simulation on both the Ansys/Lumerical and SPINS framework to confirm its performance.

In the grey-scale optimization step, the matrix that represents the device permittivity consists of a series of values continuous between 0 and 1. Grey-scale optimization begins with the initial conditions seen in figure 2.4, in the previous chapter, with each pixel point generated randomly between 0 and 1. The 1 represents silicon and the 0 represents silica, and all other values in that range are considered partly silicon and partly silica in terms of permittivity. As described in Chapter 2, the initial conditions push the values of each pixel towards a central value that allows more uniform continuous values between silicon and silica with little to no bias towards either. By using this optimization method and slowly converging to a binarized design within the optimization process, the design space can be more fully explored as the electric field does not face sharp changing permittivity boundaries or small features. Binarization can be increased within the design optimization process by pushing the values of the permittivity towards silicon or silica. In the SPINS algorithm, the level of binarization is enforced using the sigmoid function:

$$f(x) = \frac{1}{1 + e^{-c_1(x-c_2)}} \tag{3.8}$$

where the c_1 value increases the steepness of the sigmoid function while the c_2 value shifts the sigmoid function along the x axis where the values of the permittivity fall in the 0 to 1 range to represent silicon and silica. The sigmoid function forces this permittivity towards 1 or 0 based illustrated in figure 3.10.

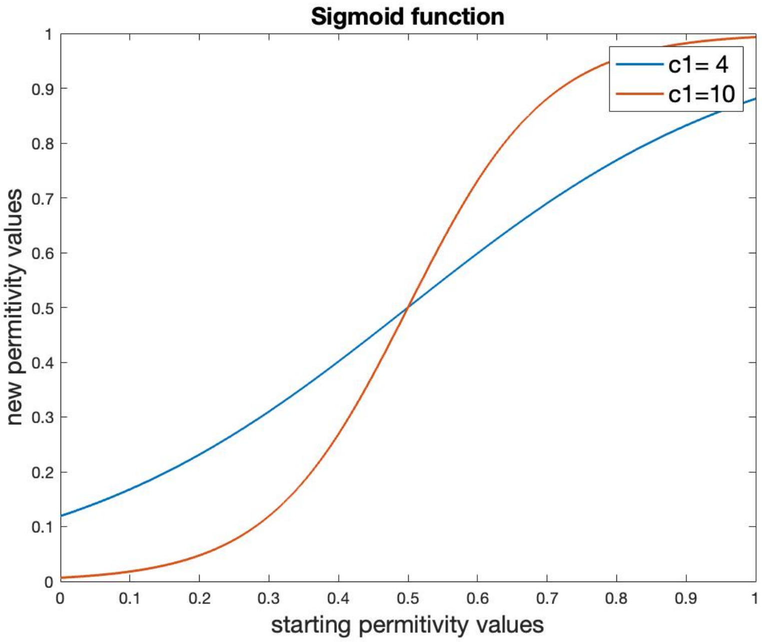


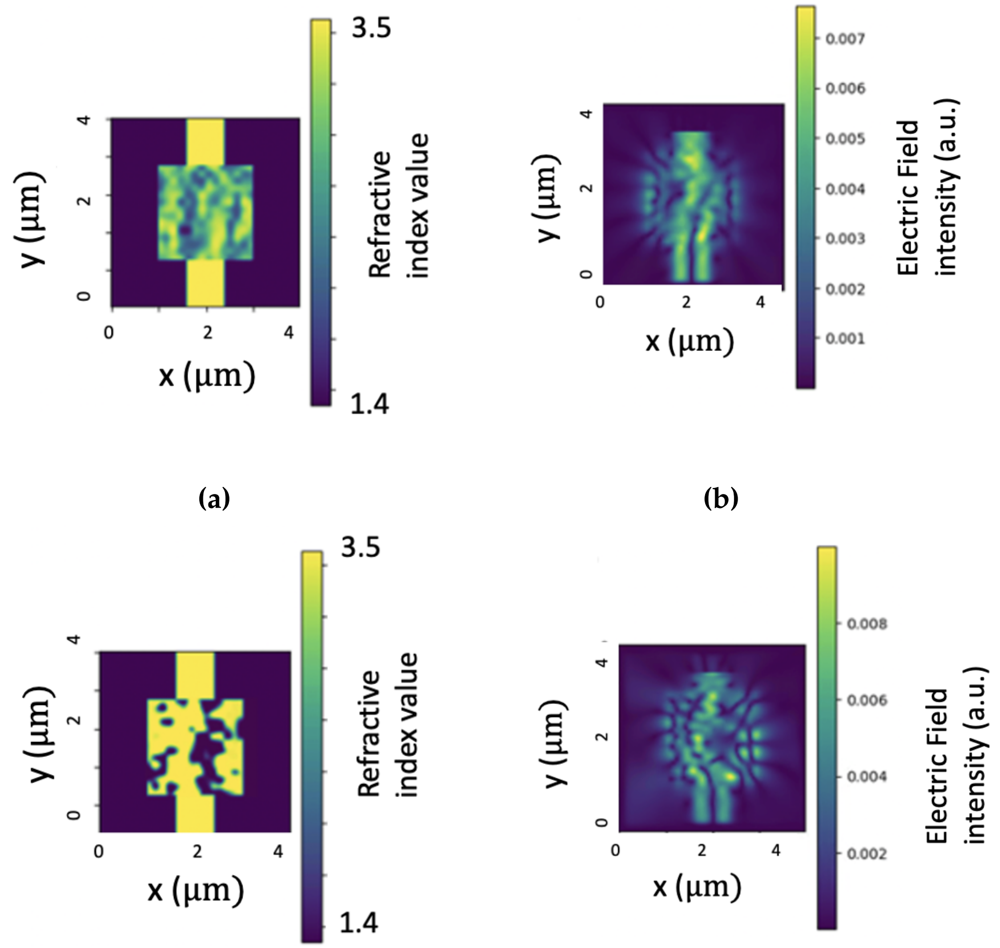
Figure 3.10: The sigmoid function for two c_1 values. c_2 would shifts the central value along the x-axis.

The value of c_1, also known as the sigmoid factor, pushes the device to be more binarized when increased. To show the impact of the sigmoid function on the optimization process, an optical mode converter is studied. Mode converters take in one optical mode and then convert the optical mode to another optical mode at the output waveguide. The mode converter is simple to set up on the SPINS optimization process and is a relatively small device (2 μm by 2 μm design space) hence it is computationally efficient to simulate. When looking at a mode converter that converts TM0 to TM1, the sigmoid factor is changed to demonstrate the difference in figure 3.11. The simulation results in SPINs provide two types of results from its algorithm: 1) a permittivity map showing the refractive index obtained in the design space seen in figure 3.11**(a)**, and 2) the corresponding electric-field map seen in 3.11**(b)**.

Once a device has been optimized with good performance and is mostly binarized, usually there is a trade-off between binarization and performance, any remaining in-between permittivity values must be dealt with when exporting the device to a layout file. When exporting, a hard boundary point can be defined, where any pixels above a certain value are silicon and any pixels below are silica, called the level-set. This is when the level-set must be explored and discussed, as even well-binarized devices may have some blurring at the permittivity boundaries. The treatment of this blurring, changing it to silicon or silica, may have significant impacts on the optical transmission and spectral profile of the device.

Similarly to fabrication errors, which can cause over-etching and under-etching that changes the device feature sizes, the binarization process can change the borders and edges of the device when selecting a level-set. This generates a need for more fabrication and simulation robust devices, which is addressed in the work of Zhang *et al*. [26]. However, when using simply the SPINS optimization process, one can see the importance of a good level-set choice. Using a similar mode converter, the process of moving from SPINS to layout to Ansys/Lumerical is shown in figure 3.12.

In figure 3.13, there is a spectral shift from lower wavelengths to higher wavelengths with a decreasing level-set value when tested on Ansys/Lumerical's FDTD simulation. A decreased level-set value increases the amount of silicon in the device, creating a better confinement for higher wavelengths and hence changing the peak optical transmission.



(a)

(b)

(c)

(d)

Figure 3.11: Refractive index and electric field maps. **(a)** and **(b)**: Resulting device with low binarization obtained from a low sigmoid factor (c_1=1). **(c)** and **(d)**: resulting design with better binarization from a high sigmoid factor (c_1=10). For both simulations, c_2 is set to the central normalized permittivity value ($c_2 = 0.5$).

As the amount of silicon in the device increases with a higher level-set, the effective index is increased. Note that the effective index is proportional with wavelength, hence longer wavelengths yield a peak optical transmission. There is a higher level of confinement for longer wavelengths at higher effective index values, hence a peak optical transmission at these wavelengths.

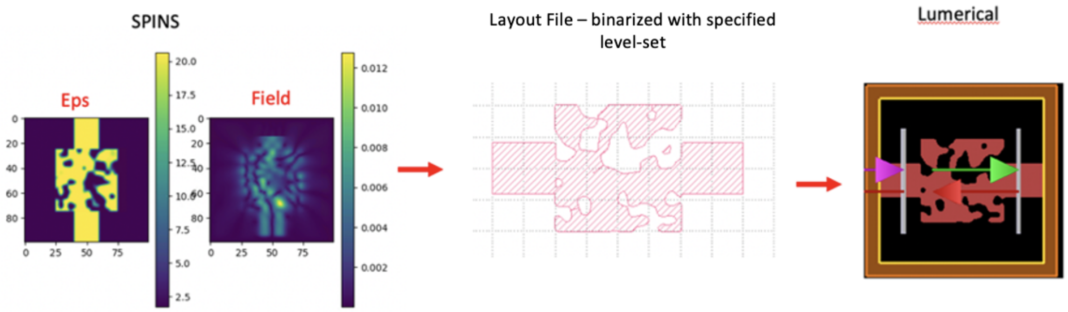
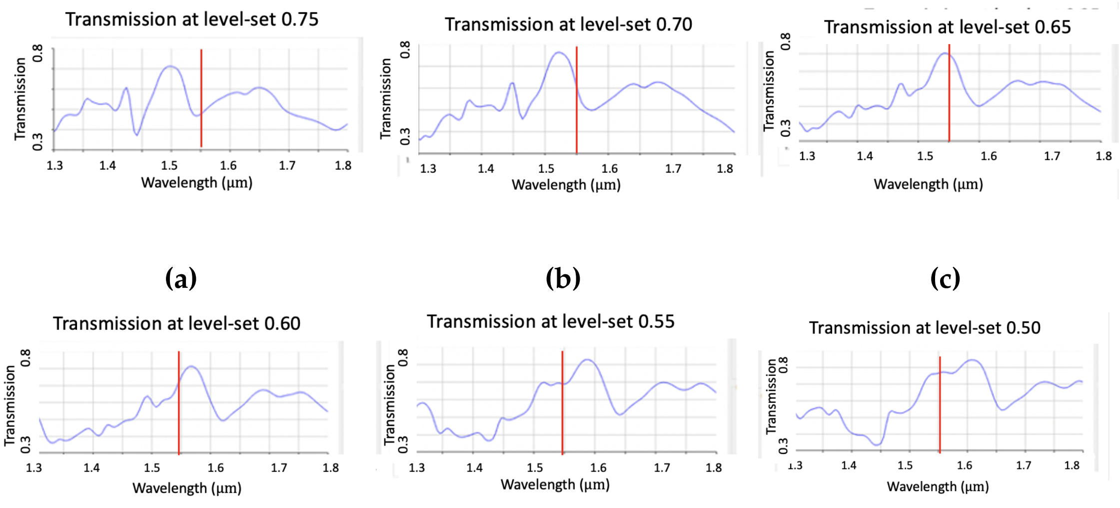


Figure 3.12: Transferring steps from SPINS results to the corresponding layout file and then imported into commercial simulation tools by Ansys/Lumerical. In the latter, the source is labeled with pink and red arrows and the monitor is labeled with green and red arrows.



(a) (b) (c)

(d) (e) (f)

Figure 3.13: Transmission results for different level-set values over a wavelength spectrum for a TM0 to TM1 mode converter. **(a)** A level-set of 0.75 with peak wavelength at 1.5 µm. **(b)** A level-set of 0.70 with peak wavelength at 1.525 µm. **(c)** A level-set of 0.65 with peak wavelength at 1.550 µm. **(d)** A level-set of 0.60 with peak wavelength at 1.575 µm. **(e)** A level-set of 0.55 with peak wavelength at 1.6 µm. **(f)** A level-set of 0.50 with peak wavelength at 1.625 µm.

Across these varied level-sets in figure 3.13, the spectral shift offers some insight into what the chosen level-set should be. For wavelength dependent devices, it is important to have the level-set with the peak transmission value close to the optimized value. For example, these devices are optimized for 1.55 µm wavelength and the level-set value of 0.65 offers the best performance at this wavelength.

Understanding how to transform the permittivity matrix from SPINS to layout to Ansys/Lumerical offers important insight into how these devices work and how incredibly sensitive they can be to small shifts at permittivity borders. By delving deeper into this topic, one can see how important small features are within inverse design and how flaws in fabrication can deeply impact these structures.

3.3 Binarization Process for 2-D Optimization

When optimizing a device, it is important to not only consider the figure of merit during optimization, but the level of binarization at each stage. As the optimization progresses, a low sigmoid factor can lead to a device being poorly binarized since pixels are allowed to range more freely in the 0 to 1 range. Ultimately, this will generate a poor performing design when fully binarized in the exportation step. Moving from low binarization to complete binarization is detrimental to the device performance as the fully optimized device should not rely on these in-between permittivity values for transmission at the outputs. Hence, it is important that the binarization process is incorporated into the optimization process. On Lumerical Lumopt, this function is built into the process and they use the following optimization stepping for their sigmoid function:

$$c_{1new} = c_{1old} \times c_1 \tag{3.9}$$

where c_{1new} is the sigmoid factor for the sigmoid function and is changed according to previous sigmoid factor values, c_{1old} and c_1. Lumerical starts with an initial sigmoid value of 1.4, which increases quickly in the process. The binarization process used in this thesis is similar, but instead follows an exponential growth with a starting sigmoid factor of 1, and at each binarization point this value is multiplied by 2. Figure 3.14 illustrates the given values for 10 binarization steps within the process. The first two binarization steps occur after 50 iterations of the optimization process and the last eight binarization steps occur every 30 iterations as the device becomes more binarized.

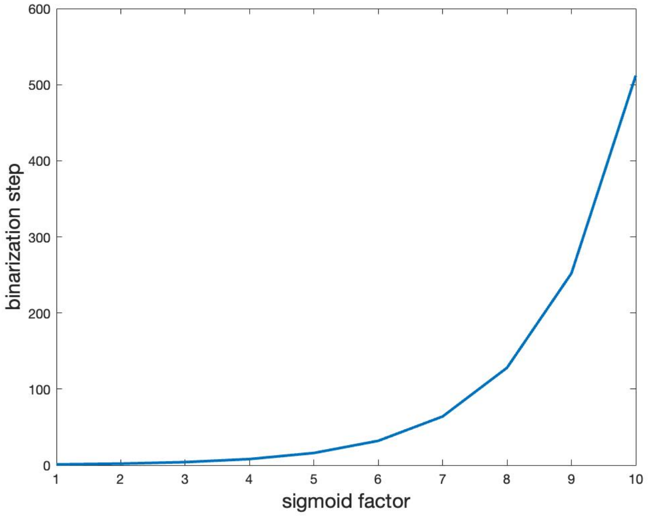


Figure 3.14: Sigmoid value, c_1, at each binarization step.

By using this methodology, it ensures that there is a slow increase in the binarization of the device within the first few steps which will allow the optimizer to explore more of the design space without the restrictions of harsh refractive index boundaries between silicon and silica. As the optimization progresses, it undergoes these binarization steps until a device is produced that is mostly binarized at the output (with the sigmoid factor reaching 500 after 10 binarization steps throughout the optimization). Some of the optimization processes do not reach full binarization as there is sometimes a trade-off between performance and high binarization with the optimizer prioritizing performance. Regardless, significant levels of binarization are achieved where only the permittivity boundaries between silicon and silica remain blurry with intermediate permittivity values. To look at the relation between high binarization and the figure of merit, the trend of the FOM can be observed. The figure of merit is measured at each iteration, and at certain iterations a binarization step occurs which causes a significant drop in device performance (peak in

FOM). For a de-multiplexer device optimized in SPINS, the figure of merit is mapped in figure 3.15.

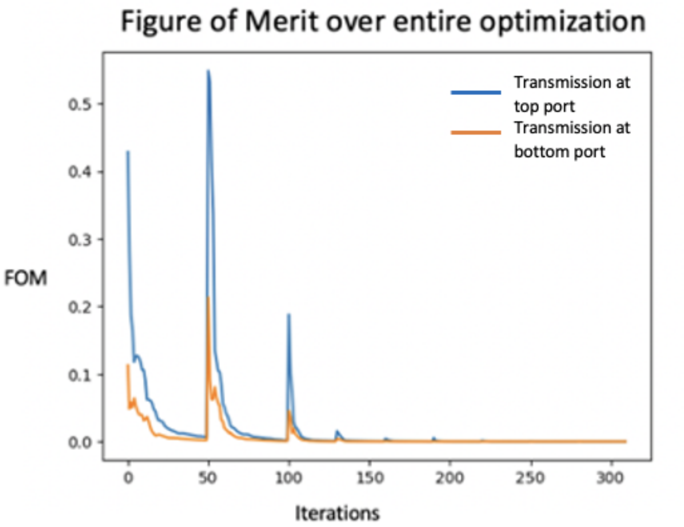


Figure 3.15: Figure of Merit evaluated at each iteration step with the peaks occurring at different binarization steps.

These very sharp defined peaks in figure 3.15 are where the points of increased binarization occur from the sigmoid function within the optimization. These peaks decrease throughout the optimization as with higher binarization there are less permittivity changes resulting in less impact to the FOM. This clearly shows the importance of using binarization steps throughout the optimization, as optimizing with a low sigmoid function throughout the entirety of the process would have negative consequences in performance when the device is exported. These devices with low binarization rely on in-between permittivity values that cannot be fabricated and will be pushed towards silicon or silica with a level-set function during exportation to a layout file. Moving a device from low binarization to high binarization in one step with no further optimization would result in poor performance and hence, these binarization steps are integral to the optimization process.

3.4 Meshing

Photonic inverse designs are often very sensitive to different meshing schemes across various EM simulations. Meshing can have a significant impact on the performance of the device on a given numerical tool. The complexity of these devices makes evaluation of the electric field difficult, with feature sizes significantly small on some inverse designed devices. The ideal mesh is small to account for small feature sizes and can accurately calculate the electric field at silicon/silica boundaries. However, meshing algorithms can be computationally costly with decreasing mesh sizes. When verifying the results from one EM simulation, it is important to see similar results across different FDFD and FDTD simulators. Consistency across simulation platforms is integral to this work as any changes to the SPINS code could impact the performance of the FDFD simulations unbeknownst to us. By confirming that the device generated by the optimization is utilizing a properly functioning FDFD simulation, assurance can be offered that not only any changes made to the optimization process work, but that the devices are fabricable. Further, it is important to have a deep understanding of any differences between FDTD/FDFD simulations during optimization and validation to guarantee a lack of error in changes to the FDFD simulation built into SPINS. SPINS and Ansys/Lumerical both have significantly different meshing default settings that must be changed to align closely with one another.

In Ansys/Lumerical, developers use advanced mesh settings to simulate devices. However, SPINS uses a much simpler mesh which makes simulation results between the two significantly different in both spectral shifts and optical transmission at the outputs. Ansys/Lumerical offers a variety of mesh settings that are both uniform and non-uniform. A uniform mesh stays consistent across the entirety of the simulation, while a non-uniform mesh can change size depending on the boundaries and the refractive index of the device. Generally, a mesh of around 20 nm to 40 nm is used for these inverse designed devices depending on the size of the device. Smaller devices (less than 9 μm^2) can utilize a finer mesh as it is not as computationally costly as larger devices.

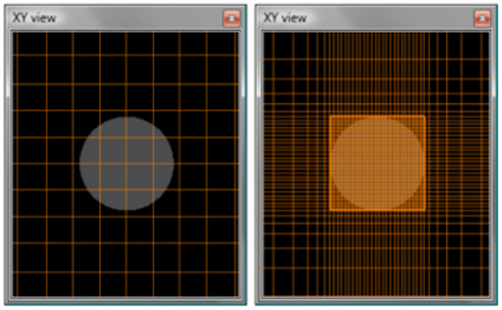
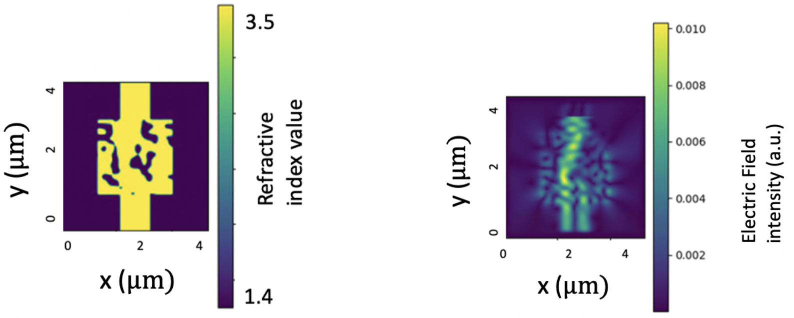


Figure 3.16: On the left there is a uniform meshing on the Ansys/Lumerical software. The right has a non-uniform mesh which changes due to material and due to the boundaries of the device. Close to the boundaries the mesh gets smaller for a more precise representation of the device feature. Reprinted with permission from [37] © ANSYS.

In figure 3.16, there is an example of a uniform mesh and a non-uniform mesh as used by Ansys/Lumerical. The non-uniform mesh varies over the two materials to take into account different refractive indices as well as the border of the device with the cladding. By reducing the size of the mesh at the borders of the device or within small features of a device, the simulation has a more realistic sense of the design shape as a greater number of pixels in a given area has much better resolution. On SPINS, however, the optimization and simulations utilize a uniform meshing, which in turn must be used in Ansys/Lumerical to give consistent results. For this design methodology, it is important to utilize uniform meshing due to the fact that Ansys/Lumerical is a closed source FDTD solver. Without knowing the specifications of their non-uniform meshing, such as refinement on curves of non-rectangular features and meshing across different materials, similar meshing algorithms may not be seen if a non-uniform meshing on SPINS was imposed on the device. Additionally, this would pose a problem during the optimization phase as a non-uniform meshing can change size for different refractive indices. Grey-scale optimization utilizes in-between permittivity values which would be constantly varying in mesh size due to the nature of grey-scale optimization. It is imperative that there are consistent results across simulations, as the FOM relies on the accuracy of the FDFD simulation in SPINS. As access to the simulation code is preferred, to view meshing calculations, permittivity calculations, and so on, it is better to use the FDFD simulation on SPINS than to have the FOM measured on Ansys/Lumerical's FDTD.

To determine what are the best settings on Ansys/Lumerical for inversely designed devices optimized on SPINS, a previously designed device by our group is looked at. In the paper, Topological Inverse Design using Energy Constraint [26], an optimized device which acts as a mode converter is designed and fabricated. Using this methodology on SPINS, this device is optimized and returns results as below in figure 3.17. The permittivity of the device (fig. 3.17 **(a)**) shows a well-binarized device for TM0 to TM1 mode conversion. This permittivity graph is then converted from a matrix to a workable layout file with a level-set of 0.6 as 0.6 is the standard level-set. Note that this device was generated before the level-set study discussed previously.



(a) **(b)**

Figure 3.17: Results from SPINS of a TM0 to TM1 mode converter. **(a)** Permittivity map output from SPINS **(b)** Electric field map output from SPINS.

This device is 2 µm by 2 µm and is placed in a simulation space of 4 µm by 4 µm. The input and output waveguides are 800 nm in width and since this is a 2-D simulation, the depth is assumed to be infinite. This device is optimized for 1550 nm wavelength. In Ansys/Lumerical, this device can be tested using different mesh settings and different mesh sizes to compare the simulations. After changing the meshing type to uniform as

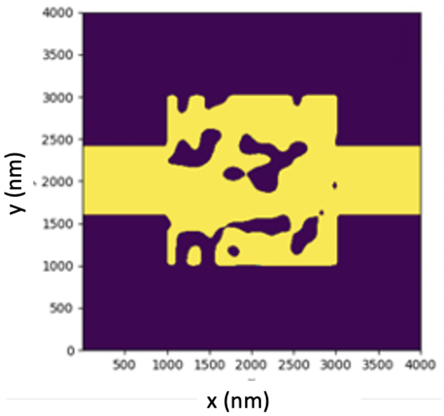


Figure 3.18: Resulting permittivity map of the 2 µm by 2 µm design area exported from SPINS.

done in SPINS, the mesh settings can be varied in Ansys/Lumerical and the most similar meshing type to SPINS is investigated.

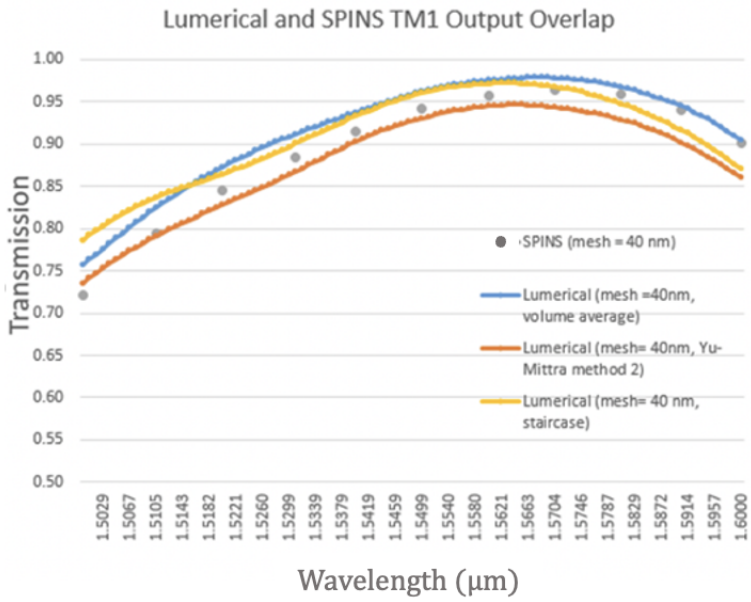


Figure 3.19: Normalized optical spectrum transmission (in linear scale) of modal overlap for TM1 at the output of the device for different meshes in Ansys/Lumerical compared to the SPINS output with a uniform mesh size of 40 nm.

Since SPINS is a FDFD based numerical solver, only one frequency point at a time can be measured hence the 10 nm interval between points in wavelength. This allows one to view the trend and save on computational efficiency compared to measurements at smaller wavelength intervals. In figure 3.19, there are three different meshing types that are all compared with the SPINS 40 nm mesh. The volume average mesh has no spectral shift as opposed to the Yu-Mittra method 2 and the staircase method, proving to be quite similar to SPINS meshing in terms of transmission and spectral position.

The staircase methodology is the simplest of all the methodologies. This methodology takes the refractive index with the highest ratio in the Yee cell and solves the Maxwell's equations using that refractive index value. What this implies is that at the boundaries of the material there is no smoothing in the mesh, but strict boundaries at each 40 nm mesh point. This creates a harsh boundary (staircase looking) that can be detrimental to the device performance, especially when the meshing is relatively large (i.e., 40 nm). A 40 nm meshing is considered large for inverse designed devices, but sometimes is necessary when designing large devices in 3-D simulations due to the computational cost of lower mesh sizes [49].

The Yu-Mittra method-2 meshing is one of the more advanced meshing methodologies offered by Ansys/Lumerical. The Yu-Mittra method is a good method for calculating Maxwell's equations at the interfaces of the structure. The permittivity of the structure is evaluated at different dimensions of the Yee Cell (fig. 3.20) and then these permittivity values are weighted with the ratio of two materials that is in the Yee cell. This advanced meshing algorithm is displayed below with the usage of a Yee Cell (Fig. 3.20).
$\mathcal{E}_1$ is the permittivity of the material in dark blue and $\mathcal{E}_2$ is the permittivity of the material in white. $\Delta\mathbf{x}$ and $\Delta\mathbf{y}$ are the dimensions of the Yee cell while $\delta\mathbf{y}$ and $\delta\mathbf{x}$ are the dimensions of the different permittivity values within the cell.

Volume averaging meshing was tested against the SPINS simulation meshing in addition to the two previously mentioned meshes. This meshing is a simple algorithm, comprised of volume averaging the two permittivity values within the mesh cell. At the

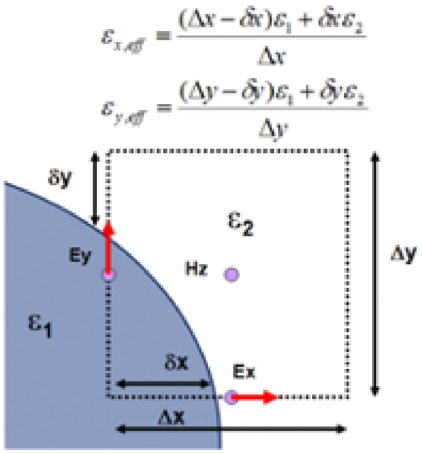
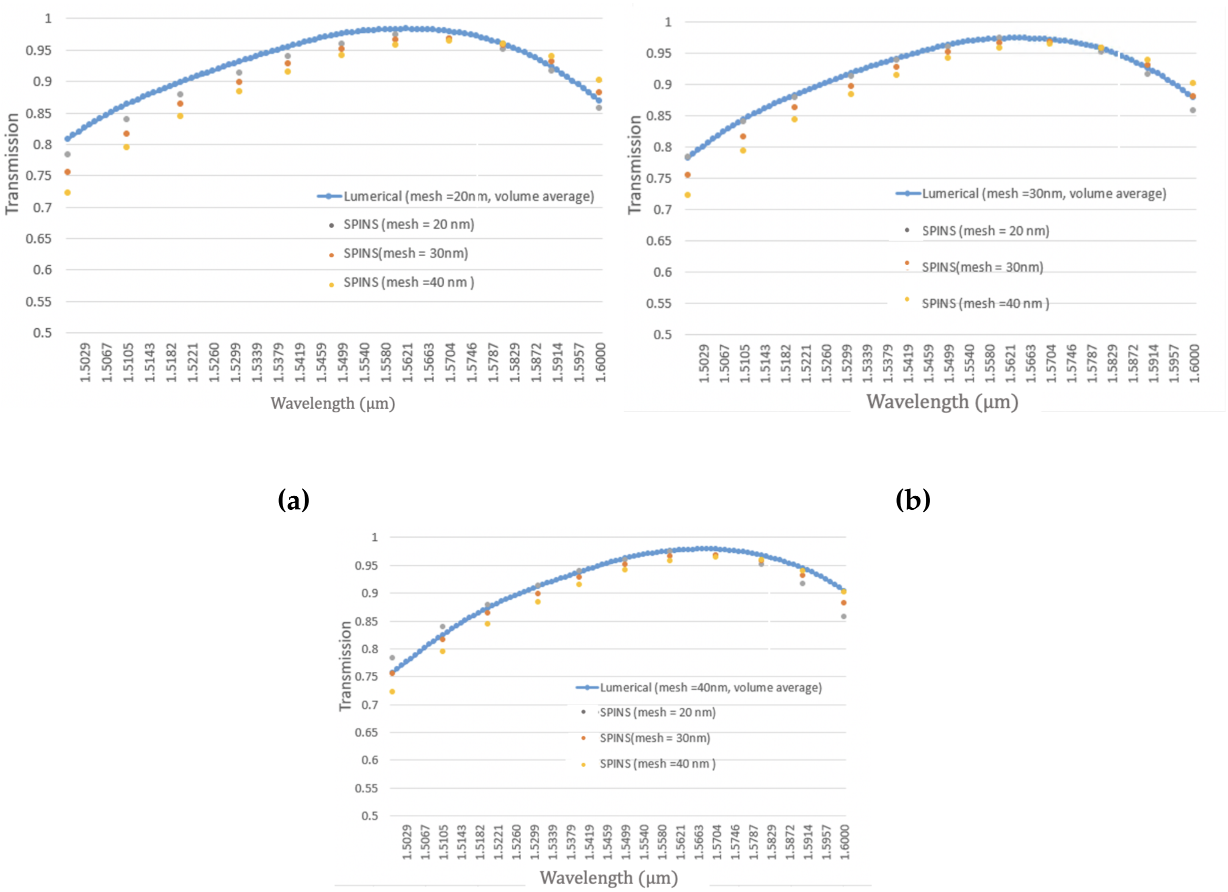


Figure 3.20: Permittivity evaluated inside the Yee Cell. Reprinted with permission from [36] © ANSYS.

boundaries, this creates a blurred like effect where the permittivity is a value somewhere in-between silicon and silica based on the averaging. When looking at the results of the different meshes versus the mesh in SPINS, the volume average offers the closest transmission results without a spectral shift (Fig. 3.21). From this, a meshing can be chosen that is a good approximation for SPINS simulations.

Along with the type of meshing, different mesh sizes on Ansys/Lumerical and SPINS can be compared to see which closely align. There is a certain amount of trade-off between mesh sizing and computational cost, with finer meshes taking a significantly longer time. Simulation time is proportional to $(\frac{\lambda}{dx})^4$ for 3-D simulations and $(\frac{\lambda}{dx})^3$ for 2-D simulations. The mesh sizing between SPINS and Ansys/Lumerical is compared using the mode converter that was detailed above.

From these findings, one can see that the 30 nm volume average meshing on Ansys/Lumerical offers the closest results to the SPINS 20 nm and 30 nm meshing. When optimizing for a device in 2-D, a 30 nm meshing will be used since the devices that have been generated are significantly larger than the 2 µm by 2 µm mode converter.



(a) **(b)**

(c)

Figure 3.21: Optical transmission spectra (linear scale) comparing mesh sizes in Ansys/Lumerical and SPINS. Optical transmission at different mesh sizes in SPINS compared to optical transmission in Ansys/Lumerical at **(a)** 20 nm mesh, **(b)** 30 nm mesh, and **(c)** 40 nm mesh.

To verify these results, another device is used, the Y-branch, and tested on the two simulations and the results compared. On Ansys/Lumerical, a uniform, volume average 30 nm mesh is applied to the structure, and on SPINS the standard meshing is used. In figure 3.22, one can see that the meshing settings are quite similar on the FDFD and FDTD simulations. This is important as it validates that a volume averaging mesh at 30 nm is very closely aligned between SPINS and Ansys/Lumerical.

Given these changes in the simulations, a simulation based testing plan is created as seen in figure 3.1. This testing plan outlines the steps necessary when converting from

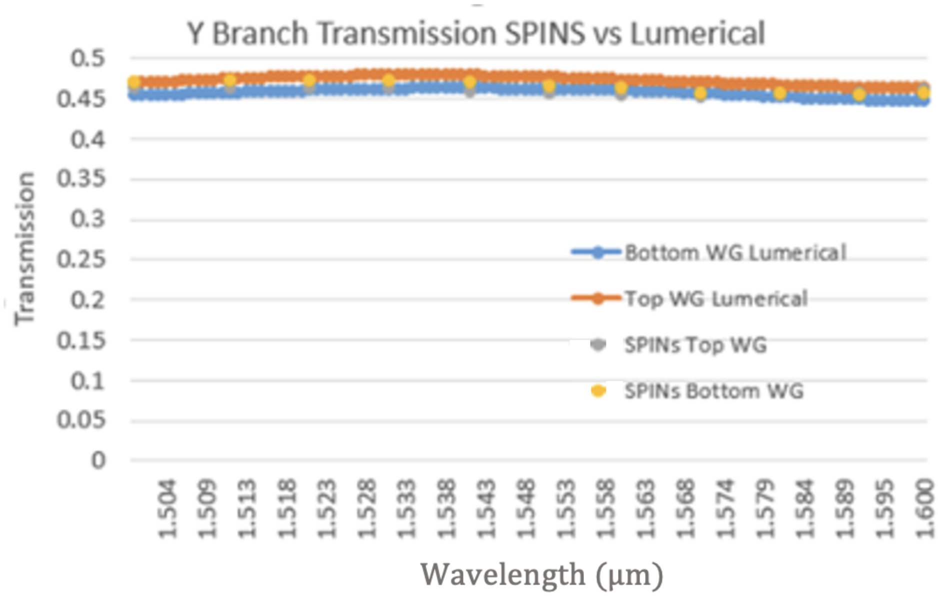


Figure 3.22: A Y-branch optical transmission (linear scale) output compared to SPINS with a specified mesh setting of 30 nm.

one simulation to the next with an optimized design and how to check for consistency across the platforms.

3.5 Summary

As the figure of merit of inverse designed device is generally measured from FDTD/FDFD simulation results, it is critical to understand the setup of these EM simulations and possible variations across simulation platforms. Even within FDTD simulations, there are significant differences between 2-D and 3-D simulations that must be considered. The transition from 2-D to 3-D simulation not only changes the fundamental mode polarization, but also can cause the rise of higher order modes. The change of fundamental mode along with increased modal crosstalk will have an impact on the FOM as these have a direct impact on optical transmission. Additionally, devices can be impacted by the file exportation from the permittivity graph on SPINS to the layout file. Increasing the binarization of the device during optimization can mitigate those effects in the file exportation. Binarization has a level-set value that defines silicon versus silica and removes all intermediate permittivity values. Hence, to maintain device performance, the device is binarized during optimization which generates a device that is more consistent across FDTD and FDFD tools. Lastly, meshing can have significant consequences on validating these devices. Differences in meshing can generate different results, hence it is important that the same meshing style and size across Ansys/Lumerical and SPINS is used. Ultimately, uniform meshing is used as it remains constant through grey-scale optimization and is consistent in closed-source Ansys/Lumerical and open-source SPINS. It is important that the set up of the simulations is accurate and consistent throughout platforms for consistent, reliable results for inverse designed devices.

Chapter 4

Devices - Thermal-based Optical Switch

Optical switching is an integral part of data communications as add/drop systems and data routing networks control the traffic within data centers. Conventionally-designed photonic integrated switches take up significant design space. For example, Mach Zehnder Interferometer with Multi-Mode Interferometer (MZI-MMI) switches can be easily 100 μm in length. To address this problem, an inverse designed thermal-based optical switch is proposed using the SPINS framework discussed in the previous chapter. Similarly in literature, devices such as mode converters and multiplexers have been studied extensively using inverse design to decrease their design footprint. This novel inverse designed thermal switch is intended to decrease the design space requirements for on-chip switching platforms by leveraging permittivity changes due to temperature. In this chapter, a thermal-based switch is generated using the open-source framework SPINS while a deeper understanding is gained of how changing the initial conditions, refractive index, design parameters (e.g., device length and width) can impact performance across FDFD and FDTD simulations.

4.1 Basic Principles

For wavelength multiplexers, the change in the refractive index is linear with wavelength in the C-band. This change in refractive index causes the light path to shift due to changes in the wave vector, allowing the wavelengths to propagate through different output ports. These light paths change with non-conventional structures generated through optimization since the refractive index changes per wavelength. Given that inverse designed wavelength multiplexers utilize the refractive index dependence of wavelength and have been validated through simulation and experimentation, this novel switch looks to control light movement with another form of refractive index change: thermal dependence. Rather than utilizing a phase shift from wavelength change, the thermally dependent refractive index can be leveraged to generate a thermal-based inverse designed optical switch. The general design set up for optimization is proposed in figure 4.1 for a 1 by 2 optical switch.

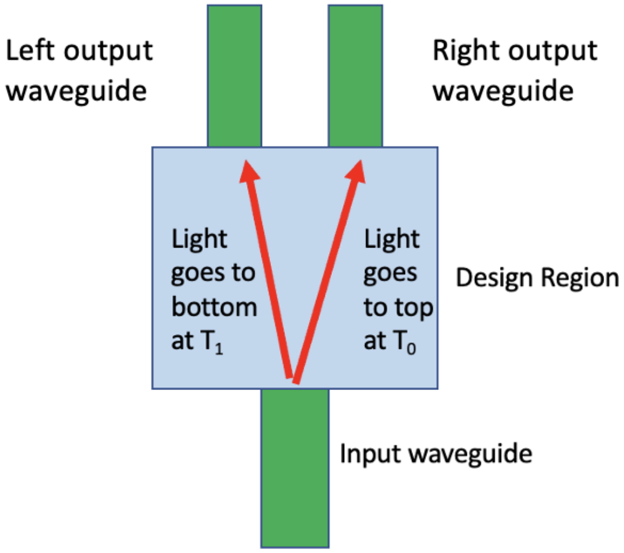


Figure 4.1: Proposed thermal switch operation using temperature-induced change in the refractive index to reroute the optical signal from one output port to the another. T_0 is low temperature and T_1 is high temperature.

The refractive index of optical materials is both temperature and wavelength dependent. The permittivity changes due to temperature change are significantly less than permittivity changes due to wavelength changes. Wavelength dependence is described as:

$$\frac{\lambda_0}{\lambda} = n \tag{4.1}$$

where λ_0 is the wavelength in free space and λ is the wavelength in a material. Thermal refractive index changes are dependent on the polarizability of the optical material as well as the thermal expansion factor, as described earlier by equation 2.18. The refractive index thermal dependence of optical materials is defined in chapter 2 section 4 as:

$$n_{new} = n + \tau \Delta T \tag{4.2}$$

where n_{new} is defined as the refractive index as changed with temperature and *n* is the refractive index at room temperature. τ is the temperature coefficient of the material and ΔT is the change in temperature from room temperature (which is assumed to be 297 K).

For silicon and silica at a wavelength of 1550 nm, the temperature coefficients of the materials are seen in table 4.1. In this thesis, these refractive index changes are incorporated into the simulation and used to generate a thermal switch.

Material	*n* at 297 K	Temperature Coefficient (τ)	*n* at 500 K (see eqn. 4.2)
Si	3.4757	1.8600×10^{-4} K^{-1}	3.5134
Si02	1.4440	1.50×10^{-5} K^{-1}	1.4470

Table 4.1: Temperature coefficients of silicon and silica based on their refractive indices (*n*) at 1.55 μm wavelength [40].

4.2 Design Description

When setting up a design problem that reflects temperature dependence, the figure of merit must be changed to reflect two optimization goals. The first optimization goal being maximum optical transmission to one output port of the device at room temperature (297 K). The second optimization goal should then reflect maximum transmission in the opposite output port when the refractive index is changed to reflect higher temperature (500 K for our devices). This figure of merit must be optimized for both transmission terms, reflecting light through the top output port at low temperature and light through the bottom output port at high temperature, creating a two term figure of merit similar to equation 2.16 in chapter 2 for temperature robustness. The optical path change can be reflected in the code by changing the objective function to reflect transmission at a different physical point (i.e., from one output port to the other output port). However, modifying the temperature in simulation is challenging because the FDFD simulation in SPINS has been constructed with no built-in temperature dependence. In this thesis, implementing the appropriate simulation methodology for incorporating temperature dependence in SPINS is resolved.

To determine how to alter the simulation to reflect temperature change, first the set up of the grey-scale initial conditions must be taken into account. As discussed previously, in section 2.3, the initial design consists of a permittivity matrix of values between 0 and 1 that are all pushed towards an intermediate normalized value. This creates a difficult problem for the optimization process, as the temperature coefficients for silicon and silica are not the same, hence the in-between permittivity values cannot be uniform in their temperature coefficient value. Another problem that arises from the set up of the design space is that the coefficients are based on the refractive indices of 1.4440 and 3.4757 for silica and silicon, respectively, hence the coefficients in table 4.1 will not properly scale the normalized values of refractive index which range from 0 to 1. To address these problems, the values of the temperature coefficient must be normalized and a linear scale for in-between permittivity values must be created to map values to an in-between temperature coefficient.

To achieve the linear scale, the following generic normalization equation can be revisited:

$$norm(x) = \frac{x - x_{min}}{x_{max} - x_{min}}, \tag{4.3}$$

which, with the refractive index values, is now:

$$norm(n) = \frac{n - n_{SiO_2low}}{n_{Sihigh} - n_{SiO_2low}} \tag{4.4}$$

where the higher and lower temperature refractive indices are taken into account. As such, n_{SiO_2} is equal to 1.4440 representing the lower bound of the refractive index (i.e., silica index at low temperature). n_{Sihigh} is equal to $3.4757 + \tau_{Si}\Delta T$ which is the upper bound of the refractive index (silicon at high temperature), where τ_{Si} is the temperature coefficient of silicon defined in table 4.1. In this work, the temperature difference ΔT is set to 200 K with the corresponding refractive index value of silicon at high temperature equal to 3.5134 (n_{Sihigh}). Now, the intermediate normalized permittivity values can be calculated, $norm(n_{Si02high})$ and $norm(n_{Silow})$, when silica is at high temperature and when silicon is at low temperature.

$$norm(n_{Si02high}) = 0.001465, \tag{4.5}$$

$$norm(n_{Silow}) = 0.981800, \tag{4.6}$$

Then, both normalized coefficients can be solved by calculating the normalized refractive index change for both silica and silicon:

$$0 + \Delta T\tau_{SiO2norm} = 0.001465, \tag{4.7}$$

$$0.9818 + \Delta T\tau_{Sinorm} = 1, \tag{4.8}$$

where the normalized thermal coefficient is solved for silica ($\tau_{SiO2norm} = 7.325 \times 10^{-6} K^{-1}$) and for silicon ($\tau_{Sinorm} = 9.1 \times 10^{-5} K^{-1}$). The developed code has the design permittivity defined in a range of normalized values between 0 and 1. To change these values to higher normalized permittivity values occurring with temperature, the slope between $\tau_{SiO2norm}$ and τ_{Sinorm} is found to compute the normalized temperature coefficients for intermediate permittivity values. The lower bound temperature coefficient is defined as $\tau_{SiO2norm} = 7.325 \times 10^{-6} K^{-1}$ for all points with a normalized permittivity of 0. The upper bound temperature coefficient is $\tau_{Sinorm} = 9.1 \times 10^{-5} K^{-1}$ with a normalized permittivity of 1. Hence, through simple algebra, the slope between the two temperature coefficients is found to be $8.3665 \times 10^{-5} K^{-1}$. Thus, the equation for changes in all normalized refractive index values (including those values in between silicon and silica) due to temperature will be:

$$n_{norm\_high} = n_{norm\_low} + \Delta T (8.3665 \times 10^{-5} \times n_{norm\_low} + 7.325 \times 10^{-6}), \tag{4.9}$$

By multiplying values of *n* by this equation for the second simulation with higher temperature, the refractive index values are therefore changed and now vary when temperature is applied in simulation to the device.

4.3 2-D Device Simulations

To test the operation of the inverse design thermal-based optical switch, a 2-D optimization is performed using SPINS. The optical switch operates on the first fundamental mode, which is the transverse magnetic mode (TM0), as discussed in section 3.1. A 1550 nm wavelength is applied for the defined TM0 optical mode source in SPINS. The width of the input and output waveguides is 480 nm for single-mode propagation. The design space area is 4 µm by 4 µm and the two output port waveguides are 0.5 µm apart from one another (waveguide center to waveguide center). The simulation results provide two permittivity maps and two electric field maps: one at the higher temperature (500 K) and one at the lower temperature (297 K, room temperature). The permittivity maps only vary in their refractive index values, corresponding to temperature, and the energy maps illustrate how the light propagates from one output port to the other with refractive index changes from the temperature difference.

As a first attempt, a significant temperature change ($\Delta T = 600\text{K}$) was used. Having a large temperature change reduces computational cost as a smaller design space can be used. Indeed, a larger refractive index due to temperature change generates more significant optical path changes, reducing the need for larger design space area. Hence, this offers insight into if thermal switching is possible with inverse design. The results obtained for the first designed device are reported in figures 4.2 and 4.3.

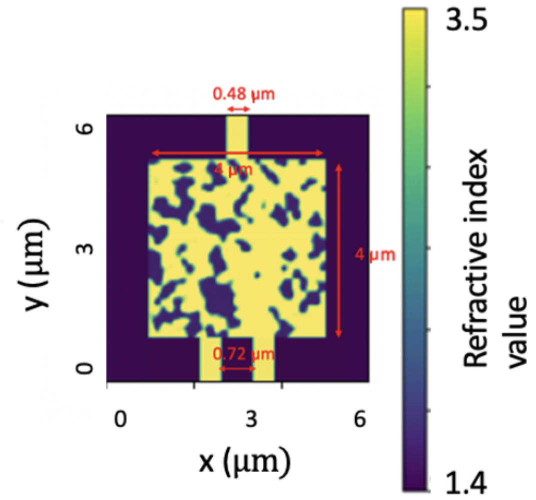


Figure 4.2: Permittivity map obtained for a thermal-based optical switch optimized on SPINS for a 600 K temperature difference.

The permittivity map is then converted to a layout file using MATLAB code which takes the matrix of permittivity values representing the design and converts it to a layout file that can be imported into Ansys/Lumerical tools. The binarization that occurs when exporting to a layout file ensures that there are no in-between permittivity values for the

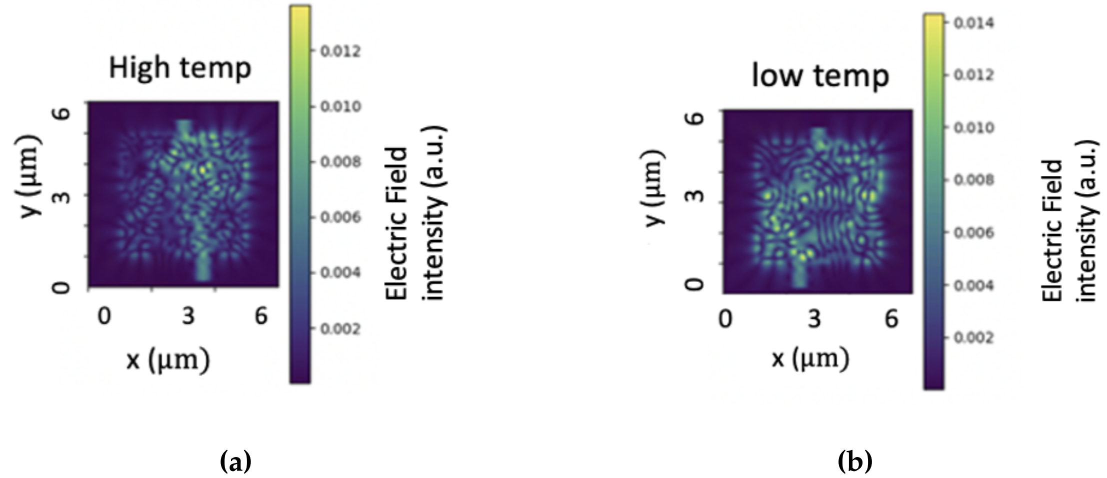


Figure 4.3: Corresponding electric field maps of the device in fig. 4.2 at **(a)** high temperature and **(b)** low temperature.

final device design. For these designs, a binarization level set value of 0.65 is used. The corresponding layout file is shown in figure 4.4.

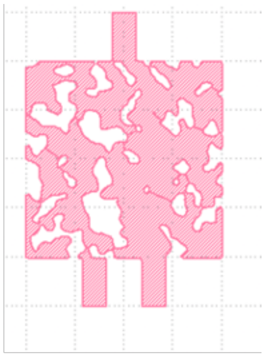


Figure 4.4: Layout file for device in figure 4.3.

From the layout file, the design can be simulated using Ansys/Lumerical to be further validated. The utilization of Ansys/Lumerical is imperative to the validation process as it: a) ensures that changing the device with temperature does not degrade the performance obtained in SPINS FDFD, b) allows for the testing of a fully binarized device as any intermediate permittivity values are removed when converted to a layout file, and c) shows a spectral response with low computational cost for the device instead of optical transmission at a singular wavelength. In Lumerical, two new materials are created that are temperature dependent with refractive indices of 3.47 and 1.44. These new materials represent silicon and silica since temperature dependence is not built into the materials database for the previously defined silicon and silica on Lumerical. The design is imported from the layout file and is given a temperature dependent silicon refractive index with a silica substrate that is also temperature dependent. Once in Lumerical, all of the monitors and sources are placed at the same points that had been defined in SPINS, with the source usually 1 µm away from the device at the input and the monitors 1 µm away at the outputs. Similarly to the monitor settings, the meshing was changed to uniform meshing with a volume average at 30 nm to match the SPINS mesh. From there, low temperature results are produced as seen in figure 4.5 where the field is radiating from the input, then interacting with the design and ultimately transmitting through the bottom output port. At high temperature, the following results are obtained in figure 4.7.

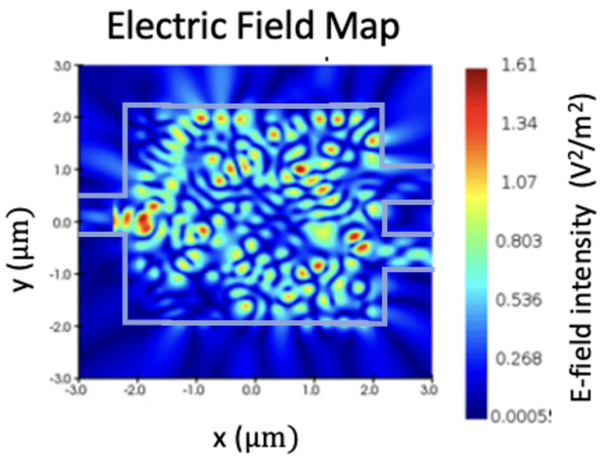
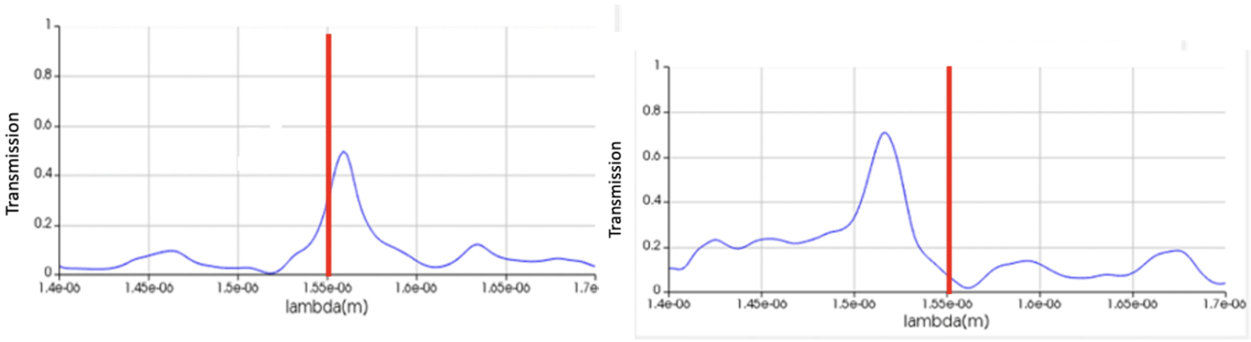


Figure 4.5: The electric field map obtained in Ansys/Lumerical for the thermal device at low temperature for a wavelength at 1.55 µm.

When looking at the transmission at the output at high and low temperature of the optimized structure on Ansys/Lumerical (fig. 4.8), the optical transmission is not optimal, with the transmission for low temperature being as low as 50% whereas on SPINS the optical transmission was at approximately 90% at the output. The performance degradation seen here is due to the exportation from SPINS to Ansys/Lumerical. Beyond slight



(a) **(b)**

Figure 4.6: Optical transmission of the device in linear scale **(a)** at low temperatures at the bottom output port and **(b)** high temperatures at the top output port.

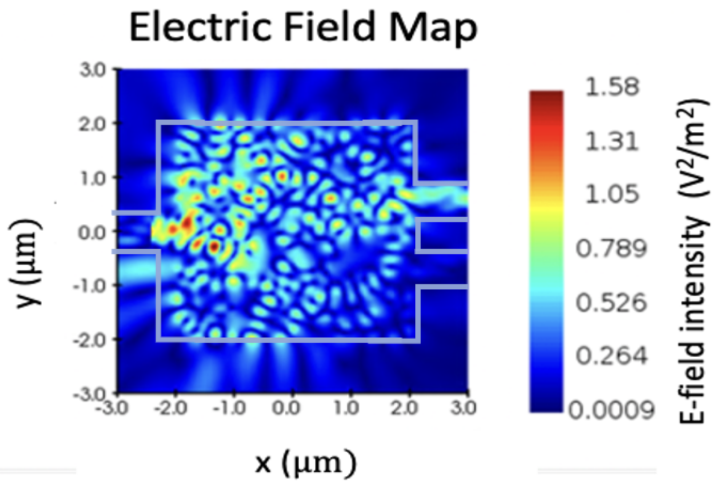
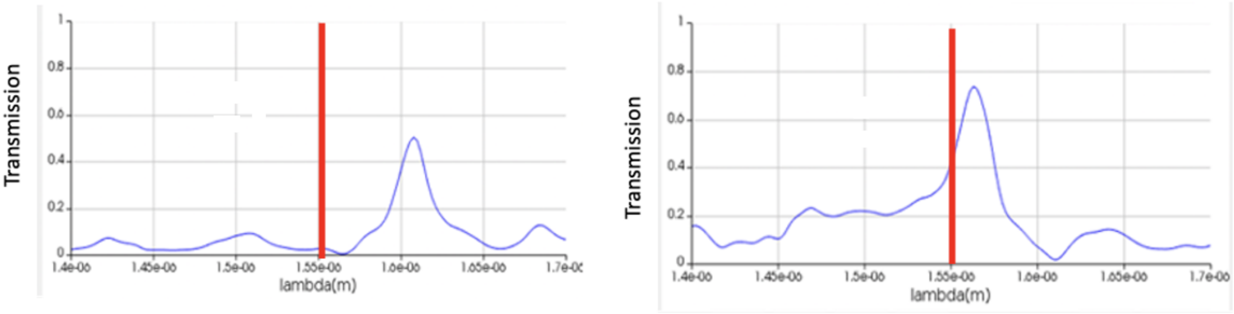


Figure 4.7: The electric field map for the thermal device on Lumerical at high temperatures for 1.55 µm wavelength.

meshing differences and impact from level-set choices, the design layout export from SPINS to Lumerical will cause a loss in transmission and/or a spectral shift in the devices. When the device is exported, there are some remnants of in-between permittivity values at the borders of the device, from both mesh averaging schemes and binarization limitations during optimization, which get binarized when transferring to a layout file. Inverse designed devices can be significantly impacted in terms of performance when it



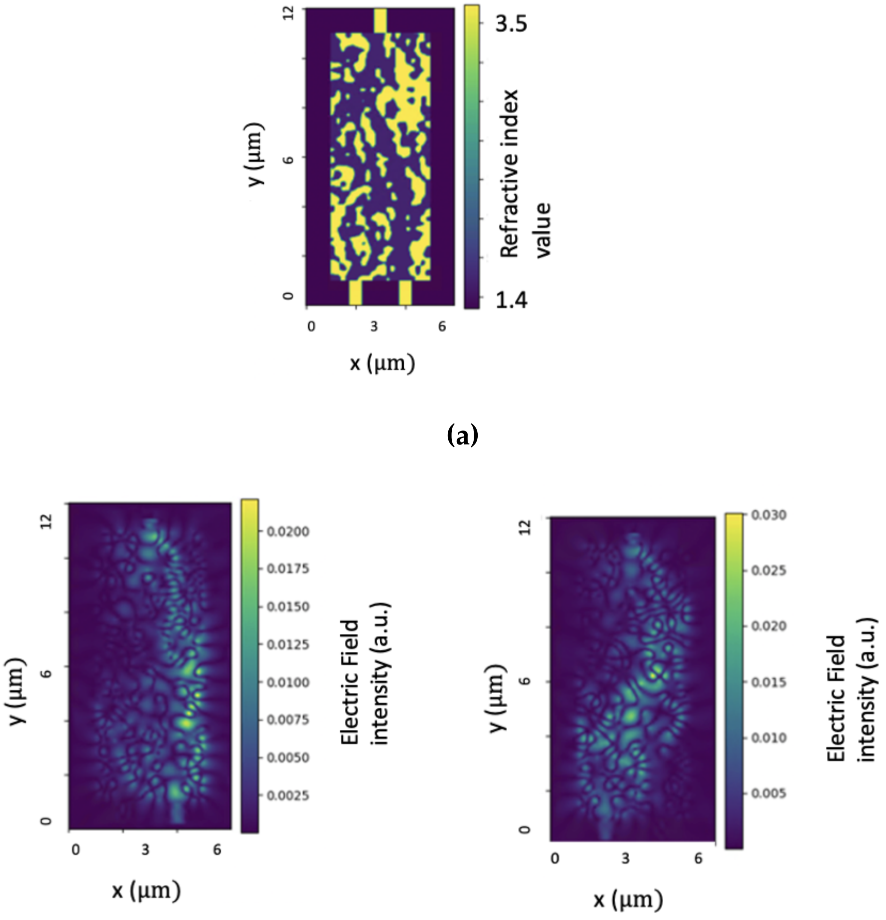
(a) **(b)**

Figure 4.8: The transmission of the device **(a)** at low temperatures at the bottom output port and **(b)** high temperatures at the top output port.

comes to changes in layout, whether that be in fabrication or simulation exportation and binarization, due to their very small features.

To resolve these meshing issues, a device must be found that is more robust to these changes in the layout, such that the layout file performs well in both simulation tools. With this thermal switch, there are several parameters that can be changed to generate different device designs, such as the widths of the input/output waveguides, the dimensions of the design area, the distance between the output waveguides, the initial conditions, and others. When moving the outputs further apart from each other, it creates a more difficult optimization problem as the light must spatially move a greater distance hence needing a larger phase shift for a more significant path change. When testing how well devices would work with a large distance between the waveguides (~ 1 µm), SPINS produced a very silica heavy device, a device with a higher ratio of silica-to-silicon, as seen in figure 4.9.

In this design (fig. 4.9), the light is given a greater length. Approximately 6 µm is added to the length of the device compared to the first device to help with this optical path change. The design width is not changed to make up for computational cost. This design is 10 µm by 4 µm with a 1 µm distance between the two output waveguides. In this device, a 150 K temperature difference was utilized to test the limitations of device



(a)

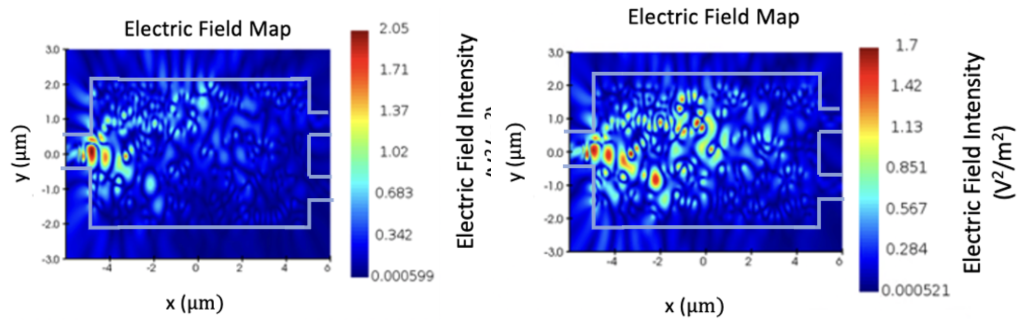
(b)

(c)

Figure 4.9: Resulting output maps from SPINS with a high silica-to-silicon ratio in the design space. **(a)** Permittivity map at low temperature, **(b)** electric field map at low temperature, **(c)** electric field map at high temperature.

optimization for smaller temperature changes that are more easily achievable in practice. When importing the layout into Lumerical, the field graphs shown in figure 4.10 are obtained.

With the results yielded in Ansys/Lumerical, one can see that this design, which performs poorly in SPINS, completely degrades in performance after importation into Ansys/Lumerical. These results yield very low transmission, but show the impact of some device parameters that need to be varied to produce better performing thermal switches.



(a) **(b)**

Figure 4.10: The field graphs obtained using Lumerical with a high silica ratio in the design space for **(a)** low temperature and **(b)** high temperature.

Reducing the number of variables cuts down on design time, with only a few design parameters needing tuning instead of many.

Initially, it was hypothesized that the thermal switch with high levels of silica, meaning a higher ratio of silica-to-silicon, performs poorly compared to the silicon heavy devices because of silica's significantly lower temperature coefficient. As mentioned previously, the temperature coefficient for silicon is 1.86×10^{-4} K^{-1} while the temperature coefficient for silica is only 1.55×10^{-5} K^{-1}. Hence, a better performing device will be obtained with a more silicon heavy design, as the effective index will change more significantly with temperature. Additionally, a decision was made to work with devices that had output ports closer apart ($\sim$ 500 nm apart). Compared to wavelength multiplexers, refractive index changes due to temperature are significantly smaller such that changes in light path occur with more difficulty. Generating compact and computationally efficient designs is more easily achievable when the light does not have to shift significantly along the width of the device, as with a device with two output ports at 1 µm distance from each other. Hence, most of the following designs will have: 1) output waveguides close together, but not so close such that they couple with one another, and 2) a higher silicon-to-silica ratio.

To increase the bias towards silicon, the initial conditions are altered to create an optimization that has more pixels in the initial design space closer to silicon than silica. As previously discussed in section 2.3, all the values in the 0 to 1 grey-scale matrix of the design space are pushed towards a central value as it is standard in the optimization process to not accidentally bias towards silicon or silica. However, to increase a bias towards silicon, the equation that has previously pushed for a central value can be changed to one that favors silicon as an initial condition. Hence, the equation is changed from

$$\rho = random \times 0.2 + 0.5 \tag{4.10}$$

to

$$\rho = random \times 0.4 + 0.5 \tag{4.11}$$

for an initial silicon heavy device. In figure 4.11, a device with a high silicon-to-silica ratio is produced and the layout is shown compared to that of a device with no silicon bias.

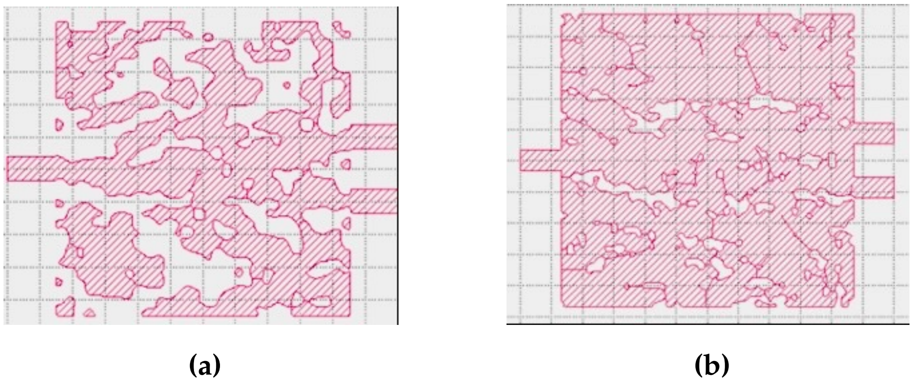

(a) (b)

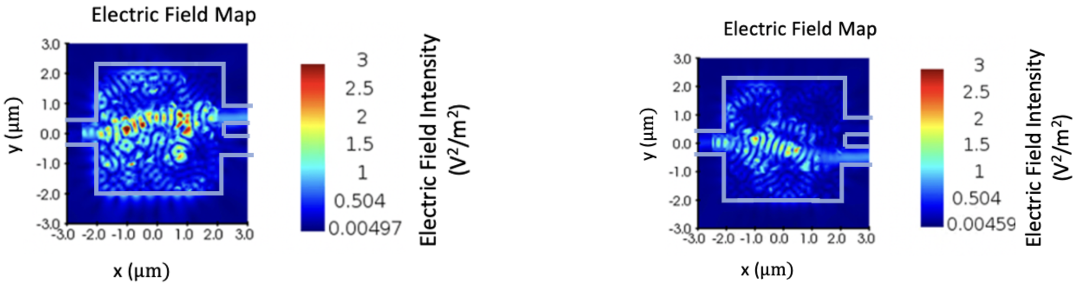
Figure 4.11: Two optimized devices, **(a)** with a high silicon bias and **(b)** with a low silicon bias.

When designing a device using inverse design, a trade off occurs between computational time and device size. The larger the device size, the more functionalities for inverse designed devices is explored. Despite a large device being realistic in the 2-D realm, it becomes incredibly complex in the 3-D realm. For this research, small devices can be defined as devices that have a width and length less than 8 µm. Computational cost of inverse design, even when utilizing a GPU server, is incredibly high when simulating 3-D devices. When designing a device in 3-D optimization, it can take multiple days to weeks for a singular device to be produced compared to 2-D optimization which takes at most a few hours.

By decreasing the size of the device, the computational cost of the simulation is decreased. However, there are some limitations on the size of the devices beyond computational cost. In smaller scales, the chance of generating small features is increased and some of these small features cannot be fabricated using typical foundries. Additionally, light confinement, the ability of the device to confine the light into the silicon region, can be limited. Another significant reason as to why these thermal devices must find balance between computationally efficient small devices and well performing large devices, is their temperature dependency. Given that the refractive index change with wavelength is significantly higher than that of temperature, multiplexer devices have been made on a smaller scale. However, this is not the case with temperature dependent devices as the temperature coefficient changes the refractive index in much smaller increments. Hence, a balance must be found between the computational cost of size and the experimental allowances of temperature.

With a large temperature difference (approximately 600 K) small devices can be generated, at a scale of approximately 4 μm by 4 μm. Using a ΔT of 600 K for small devices simulated in the 2-D is an acceptable proof of concept, however, it is impractical for fabricated devices as generating that much heat may damage the system and is energy costly. However, when the temperature difference is reduced, there is a need to generate larger device sizes to give the light more length and width for path shifts due to refractive index changes. Given these limitations, the best approach for temperature and device size tradeoff is using long devices (length of approximately 7-10 μm and width of approximately 3 μm) for a 150 to 200 K temperature difference.

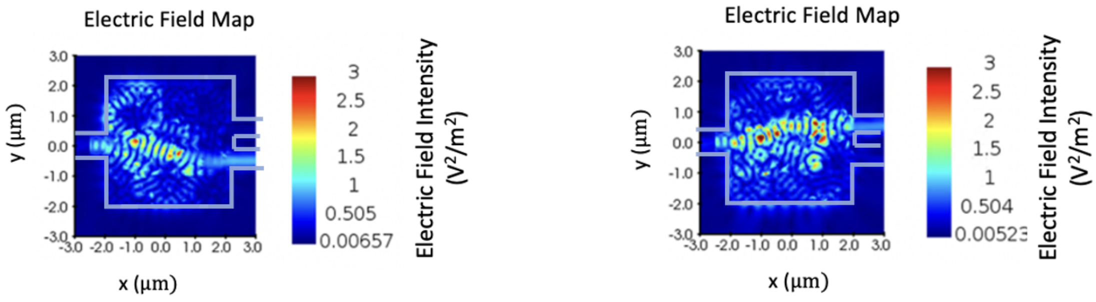
Another challenge that is seen within 2-D optimization of these thermal switch devices is that they act as a tunable thermal de-multiplexer. Indeed, one can see that when monitoring one of the output ports of the device, the peak transmission wavelength will shift with temperature. This is displayed in figure 4.12 on a 6 µm by 6 µm device with a 200 K temperature difference in 2-D optimization.



(a) **(b)**

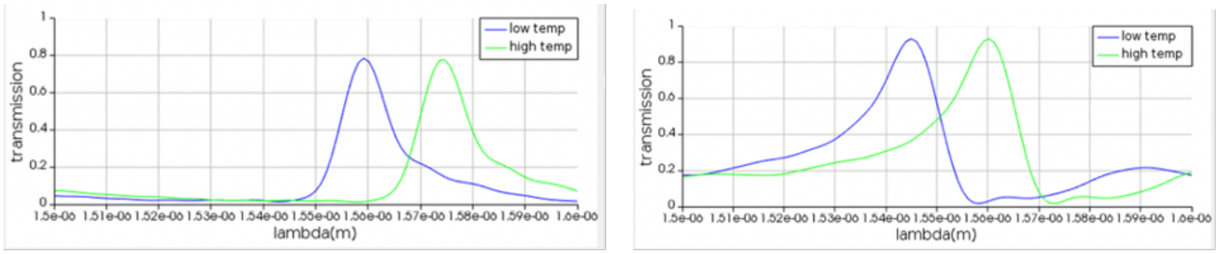
Figure 4.12: Optimized device at low temperature at two different wavelengths: **(a)** 1.54 µm and **(b)** 1.56 µm.

At low temperature, after optimization for the top output port at 1.55 µm, the light reaches the top output port at 1.56 µm. This is due to a wavelength shift caused by the binarization in the exportation from SPINS to the layout file, hence the device peak transmission is at 1.56 µm rather than at 1.55 µm. The light peak transmission at the bottom output port is at 1.54 µm. Similarly, at high temperature (fig. 4.13), 1.54 µm wavelength light reaches the bottom output port while 1.56 µm reaches to the top output port. The transmission spectrum profiles are shown in fig. 4.14.



(a) **(b)**

Figure 4.13: Results obtained for the optimized device at high temperature at two different wavelengths: **(a)** 1.54 µm and **(b)** 1.56 µm.



(a) **(b)**

Figure 4.14: Optical transmission spectra of the device: **(a)** bottom output port and **(b)** top output port at two different temperatures showing a spectral shift of approximately 15 nm for a 200 K temperature difference.

4.4 Generating Broadband Thermal Switches

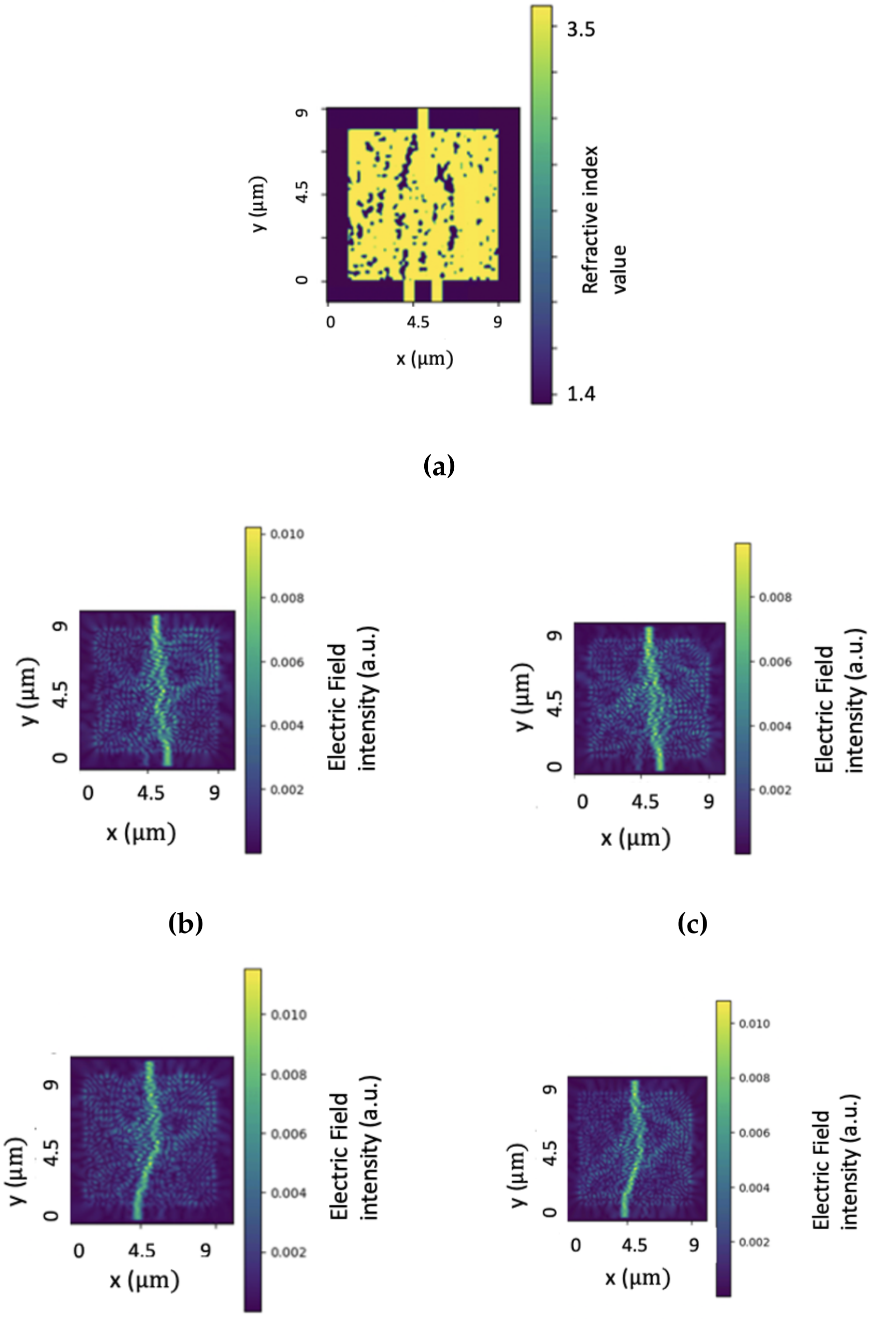
The previous thermal switch is a narrow-band device because of the wavelength demultiplexer properties the device exhibits. It is an incredibly wavelength sensitive device and is likely to have further wavelength shifts given the chosen level-set value during the layout export binarization process and errors in the fabrication process.

To generate a broadband device, the objective function must be improved within the simulations to incorporate more wavelengths for each FDFD simulation. This is one of the major drawbacks in the SPINS framework, as the FOM is measured using FDFD simulations at one wavelength. When testing the devices in Lumerical, one can see that they do not perform well over a range of optical wavelength spectrum and that the output port changes based on wavelength. To circumvent this problem, more objective functions are added to the optimization process to generate a broadband switch over a specified wavelength spectrum range.

To test the upper limits of how optically broadband the devices can be made, a broad frequency spectrum is chosen (40 nm wavelength range) between the two wavelength simulations and optimized with a figure of merit that is the same as the previous thermal switches, which optimized for transmission through the two output ports based on temperature. In this optimization, at 1.53 μm and 1.57 μm, the light propagates to the top output port for low temperature and light propagates to the bottom output port for high temperature. Hence, the device must be optimized for two separate wavelengths, 20 nm apart, with the assumed goal that the central wavelength (1.55 μm) will follow the same propagation pattern when temperature is changed. The results from SPINS are as shown in figure 4.15 for both high and low temperature.

This device has a 500 nm separation between the output ports, a 7 μm by 7 μm design space, and is optimized for TM0 mode. When the layout is exported to Lumerical with a 0.65 level-set, the broadband capabilities of the device are tested. The electric field graphs for the three different wavelengths (1.53 μm, 1.55 μm, and 1.57 μm) at high and low temperature are shown in fig. 4.16. From figure 4.16, the optimization process achieved peak optical transmission for the two wavelengths (1.53 μm and 1.57 μm), but the 1.55 μm optical transmission falls to a minimum. In figure 4.17, it is clearly seen that transmission at the designated output has peaked at the two optimized wavelengths but does not at 1.55 μm. Instead, the optical transmission at the designated port falls to a minimum at 1.55 μm and propagates through the other output port.

A 20 nm wavelength difference is thus too wide for the optimization process. A smaller wavelength difference, 6 nm, is then used to achieve a broadband device. The optimizer now uses the following simulations to drive the optimization: two simulations at low temperature with both 1.553 μm and 1.547 μm wavelength and another two at

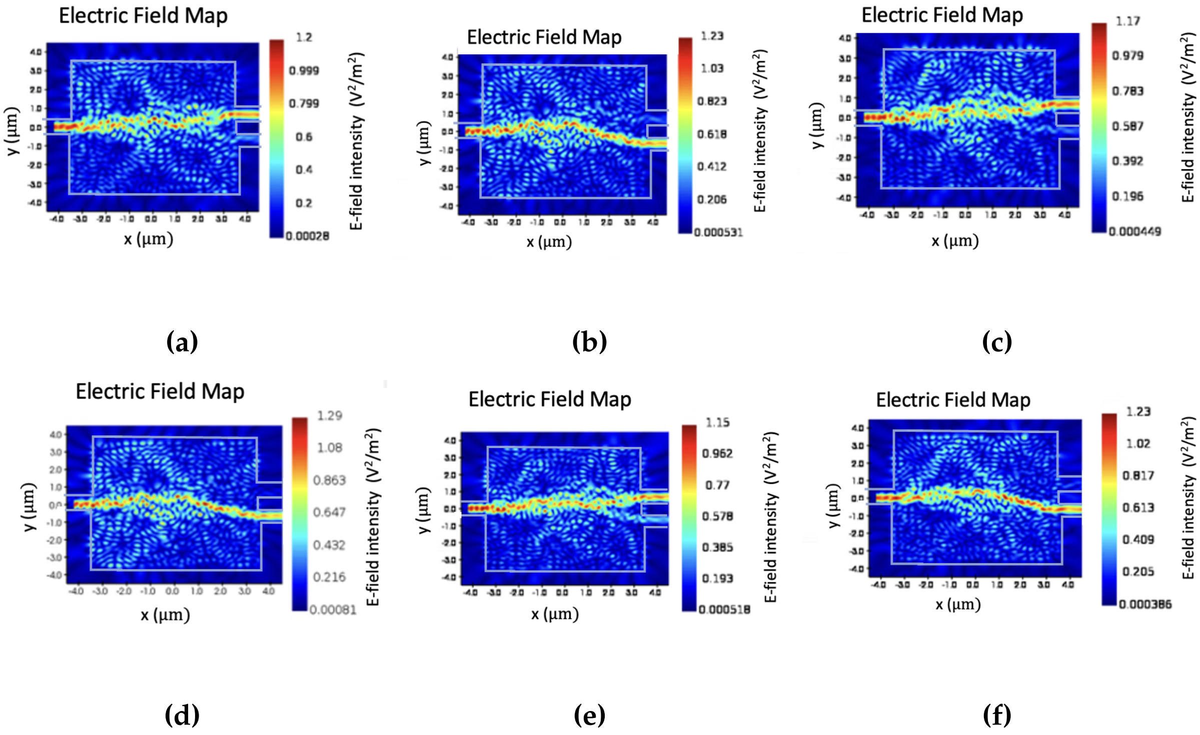


(a)

(b) **(c)**

(d) **(e)**

Figure 4.15: A device performing at low temperatures at 1530 and 1570 nm. **(a)** The SPINS permittivity map, **(b)** the SPINS field map at 1530 nm, **(c)** the SPINS field map at 1570 nm. The SPINS field map at high temperature for **(d)** 1530 nm, **(e)** 1570 nm.

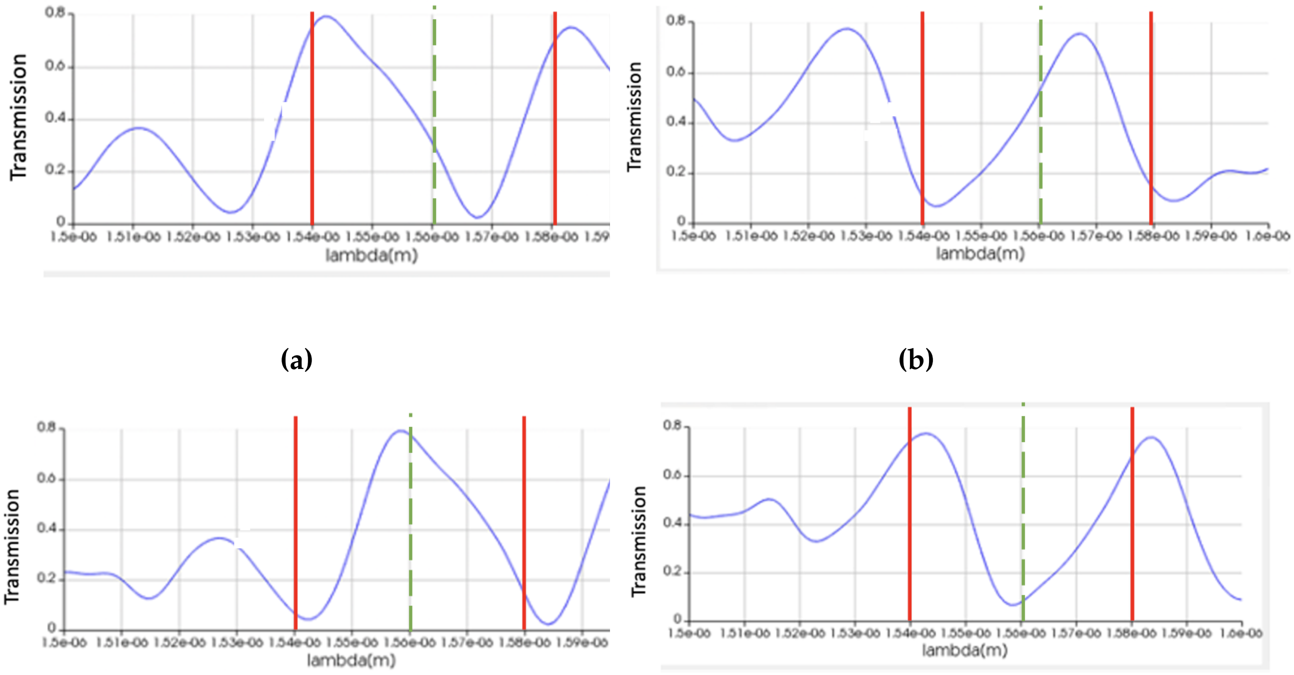


(a) (b) (c)

(d) (e) (f)

Figure 4.16: Resulting device performing at low and high temperature in Lumerical at wavelength: (**a, d**) 1.53 µm, (**b, e**) 1.55 µm, (**c, f**) 1.57 µm. The top row represents operation at low temperature (**a, b, c**). The bottom row (**d, e, f**) represents high temperature operation.

high temperature with both 1.553 µm and 1.547 µm wavelength. For 1.553 µm wavelength at high and low temperature, SPINS yields the electric field maps in figure 4.18 **(b)** and **(c)**. Then, for 1.547 µm wavelength at high and low temperature, SPINS yields the field graphs in figure 4.18 **(d)** and **(e)**. When exported to Lumerical, the electric field map at 1.55 µm is produced for low and high temperature in figure 4.19. From these electric field maps, it is seen that the device causes light to propagate through the intended output waveguides at 1.550 µm despite being optimized at 1.547 µm and 1.553 µm. Then, when looking at the transmission for the top and bottom output ports, one can see a flat and wide transmission in the transmission plots in figure 4.20. This device has achieved a broader optical band for the design by optimizing with a 6 nm difference between the

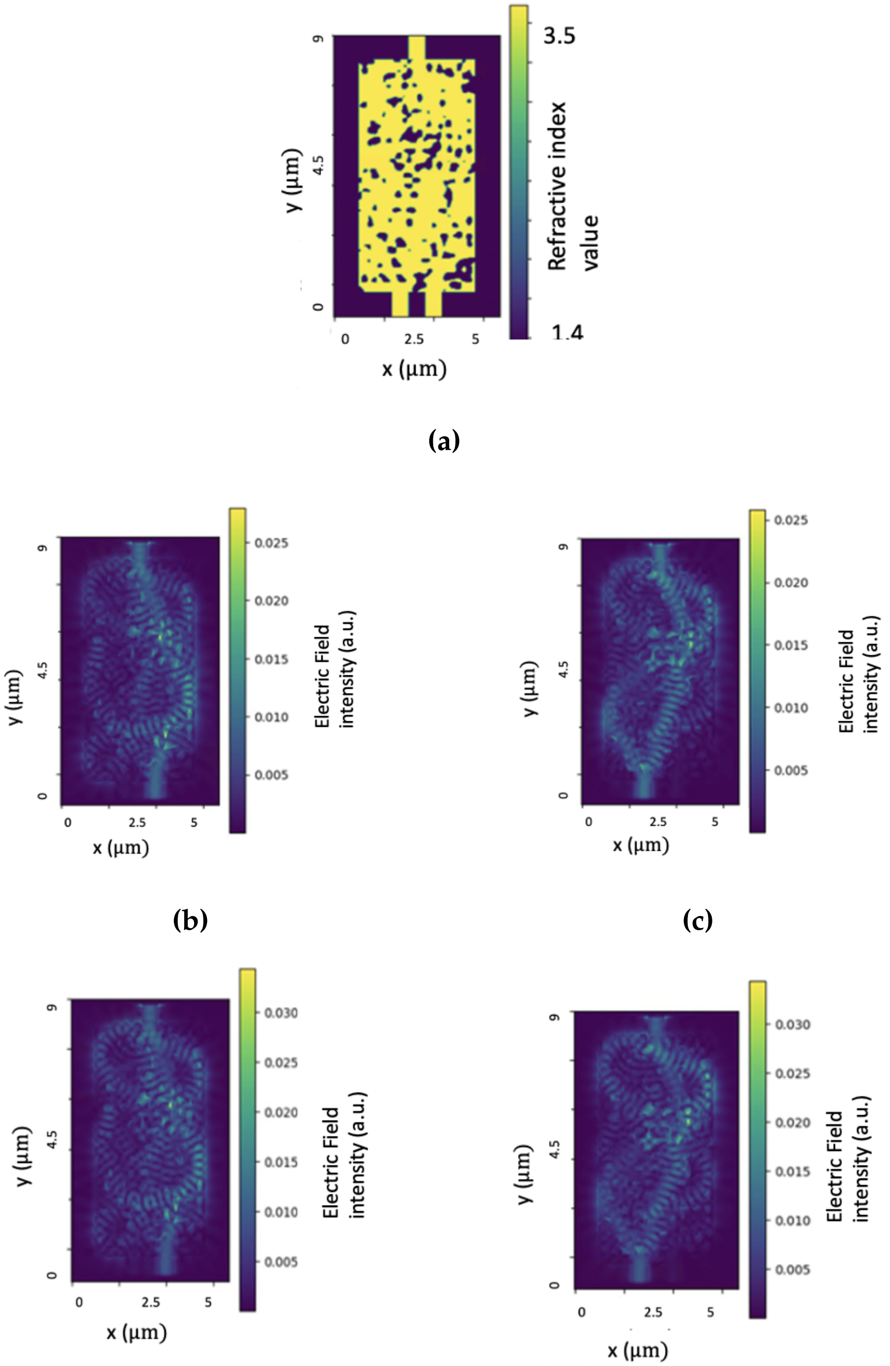


(a) **(b)**

(c) **(d)**

Figure 4.17: The top (left) and bottom (right) output port optical transmissions in the linear scale at **(a, c)** high and **(b, d)** low temperature. The red line represents the optimized wavelength with a spectral shift due to SPINS exportation. The green dashed line represents the central wavelength between the two points with a spectral shift due to SPINS exportation.

two optimized wavelengths. Indeed, the transmission map at low and high temperature (fig. 4.20) shows a thicker optical band for peak optical transmission.



(a)

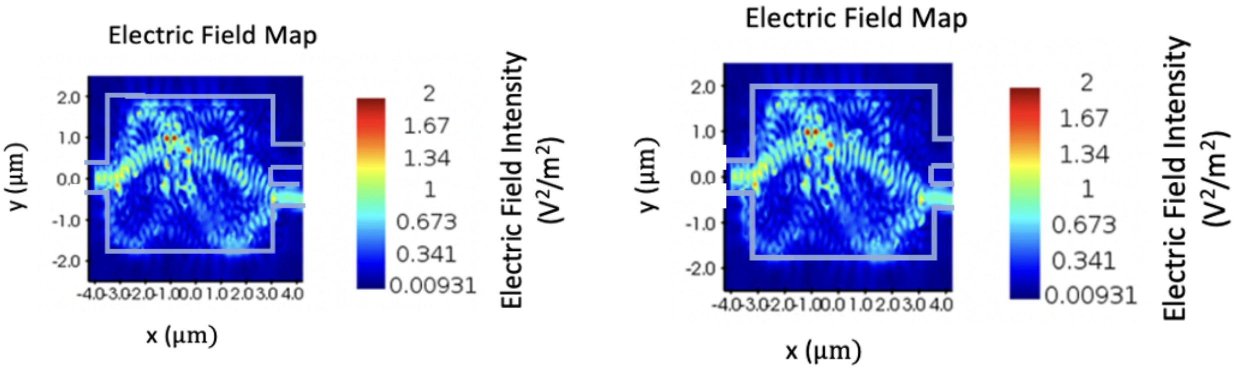
(b)

(c)

(d)

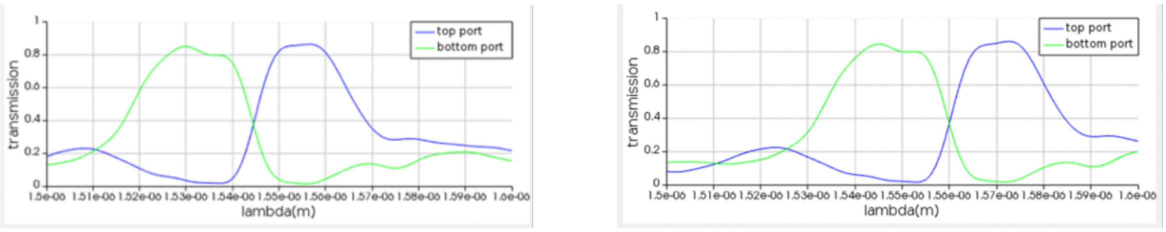
(e)

Figure 4.18: Results obtained in SPINS for an optimized device obtained for a wavelength range of 6 nm. **(a)** The permittivity map of the device. The electric field map at 1.553 µm for **(b)** low temperature, **(c)** at high temperature. The electric field map at 1.547 µm for **(d)** low temperature, **(e)** high temperature.



(a) **(b)**

Figure 4.19: Lumerical field results of the optimized device in 4.18 at 1.55 µm at **(a)** low temperature and at **(b)** high temperature for device in figure 4.18.



(a) **(b)**

Figure 4.20: Lumerical results showing the spectral profile at **(a)** low temperature for top and bottom output ports and, **(b)** high temperature at top and bottom output ports.

4.5 Two-Mode Thermal Switch

To further investigate the data transfer capabilities with inverse designed thermal switching, a two-mode thermal switch is implemented in the 2-D optimizer (fig. 4.21). A two-mode thermal switch would be incredibly practical for MDM systems. These systems rely on the propagation of multiple data lines on different optical modes, and the function of routing and splitting modes is imperative for the transfer of these data lines. In this section, an initial simulation for such a device is presented based on the same principals as the thermal switch.

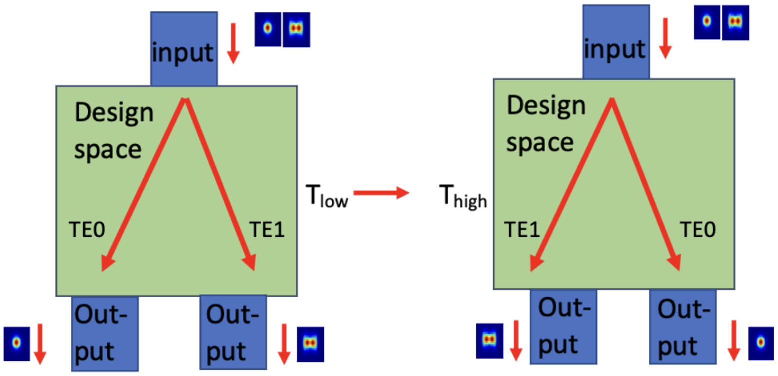


Figure 4.21: Illustration of a two-mode switch where TE0 and TE1 switch output ports with a temperature change.

The figure of merit for low temperature is defined as:

$$F_{EM\_low} = (1 - |\mathbf{E}_{TE0}^{T}\mathbf{E}|)_{bottom}^{2} + (1 - |\mathbf{E}_{TE1}^{T}\mathbf{E}|)_{top}^{2} \tag{4.12}$$

with transmission for TE0 through the bottom output port and transmission for TE1 at the top output port. T is the conjugate transpose with $\mathbf{E}_{TE0/1}^{T}\mathbf{E}$ defined as the electric field ($\mathbf{E}$) overlap with the target mode ($\mathbf{E}_{TE0/1}$), with the first term representing the overlap of TE0 at the bottom output port and the second term representing TE1 overlap at the top output port. The figure of merit for high temperature is defined as:

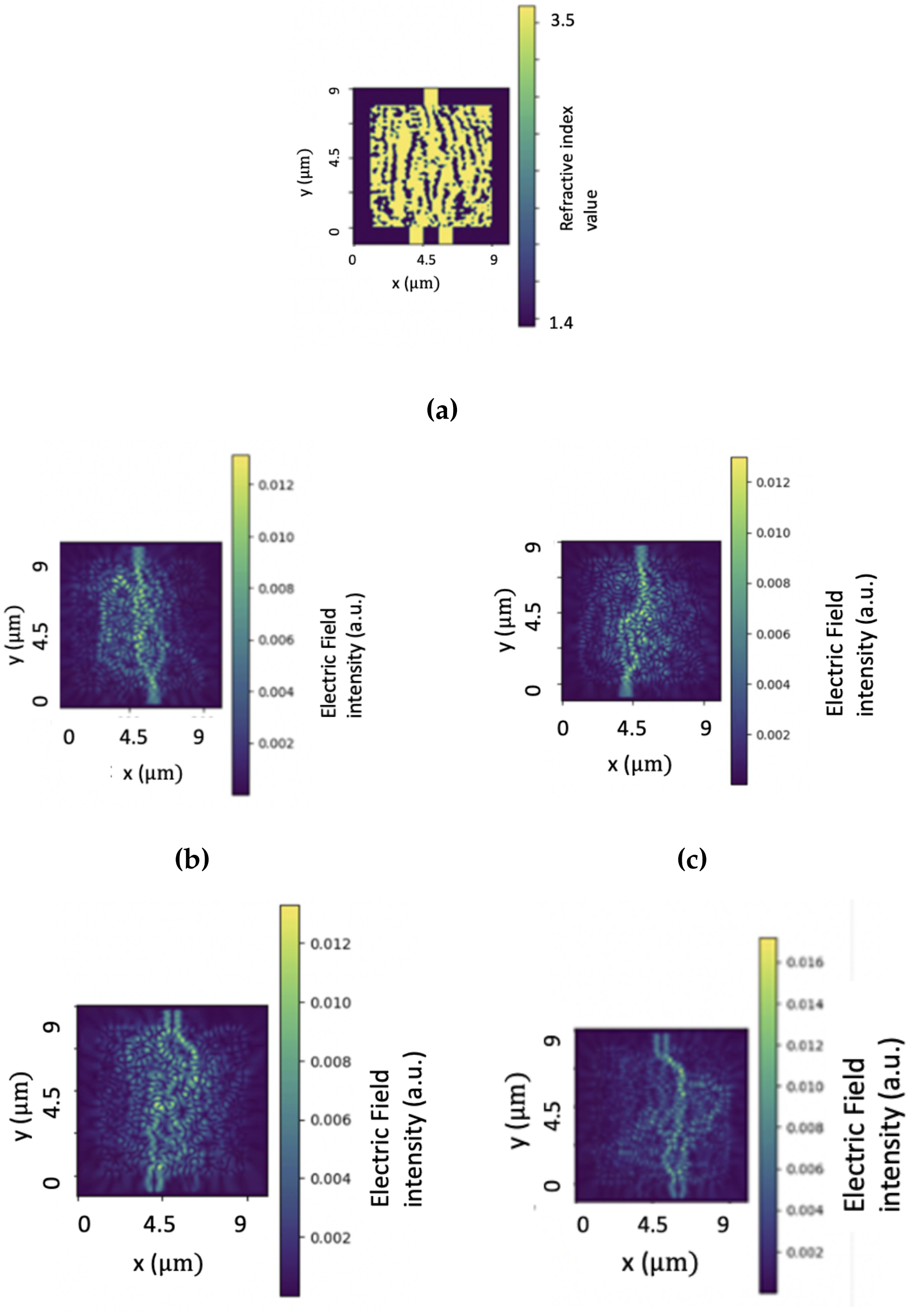
$$F_{EM\_high} = (1 - |\mathbf{E}_{TE0}^{T}\mathbf{E}|)_{top}^{2} + (1 - |\mathbf{E}_{TE1}^{T}\mathbf{E}|)_{bottom}^{2} \tag{4.13}$$

with TE0 propagating through the top output port and TE1 propagating through the bottom output port. The entire figure of merit for the optimization problem is the summation of the high and low temperature FOM.

When generating a design for higher order modes, it must be determined how large the input/output waveguides should be to support higher order modes. Using the same methodology as described in chapter 3, using the effective index method and b-V curves, a 800 nm waveguide width is determined to support the first two optical modes (in this case, TM0 and TM1). To set up this optimization, two different sources are created, one at TM0 and one at TM1, which are run on different simulations, and monitors are placed at the output waveguides to measure modal overlap with the designated modes. SPINS results are shown in figure 4.22.

In figure 4.22, the design has a 200 K temperature difference, a 7 µm by 7 µm design space, and operates at 1.55 µm wavelength for both TM0 and TM1 modes. When exported to Lumerical, the results are shown in figure 4.23. When looking at the optical transmission spectra for the top and bottom output ports, at different temperatures, the device does cause the two modes to switch from one output port to the other. This is verified through the modal overlap measurements, seen for low temperature and high temperature (fig. 4.24).

To become more broadband, this device would need for more simulations to run in parallel with wavelengths spanning 6 nm. As well as the need to become broadband, the two-mode thermal switch will also have to be run in 3-D optimizations to become fabricable. This device is a 7 µm by 7 µm sized switch with four terms in its FOM. Each term in the FOM must be run on a separate FDFD simulation, which greatly increases the time for each iteration. The total time for one iteration equals the time of a singular simulation multiplied by the number of simulations. With each iteration taking over two hours in 3-D FDFD simulation, this device must be further explored for ways in which computational cost can be decreased, such as meshing, device size and temperature change. Regardless, this device still shows promise within the 2-D realm and can be further explored with higher computational capacities. This device offers thermal modal switching that would increase the capacity of modal data transmission, allowing for greater broadband in our data transmission systems.



(a)

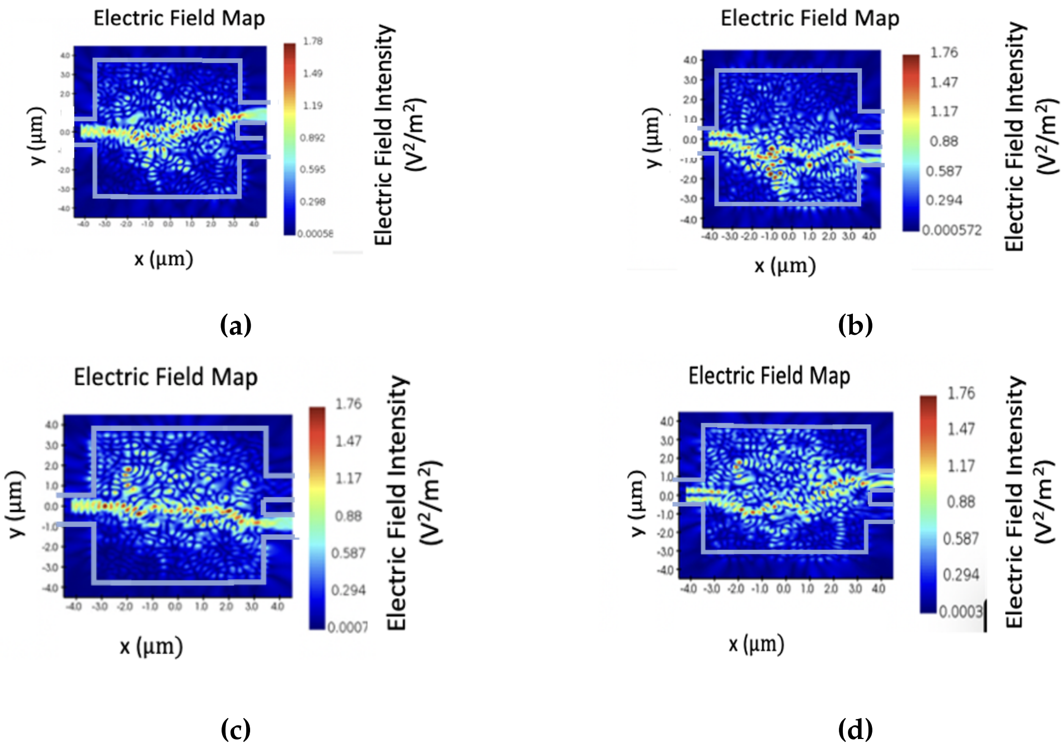
(b)

(c)

(d)

(e)

Figure 4.22: SPINS results for the two-mode thermal switch, **(a)** the permittivity map, **(b)** the electric field map at low temperature for TE0, and **(c)** the electric field map at high temperature for TEO. **(b)** The electric field map at low temperature for TE1 and **(c)** the electric field map at high temperature for TE1.



(a) (b)

(c) (d)

Figure 4.23: Lumerical outputs for the two-mode thermal switch at (**a** , **b**) high temperature and (**c** , **d**) low temperature at 1.55 μm.

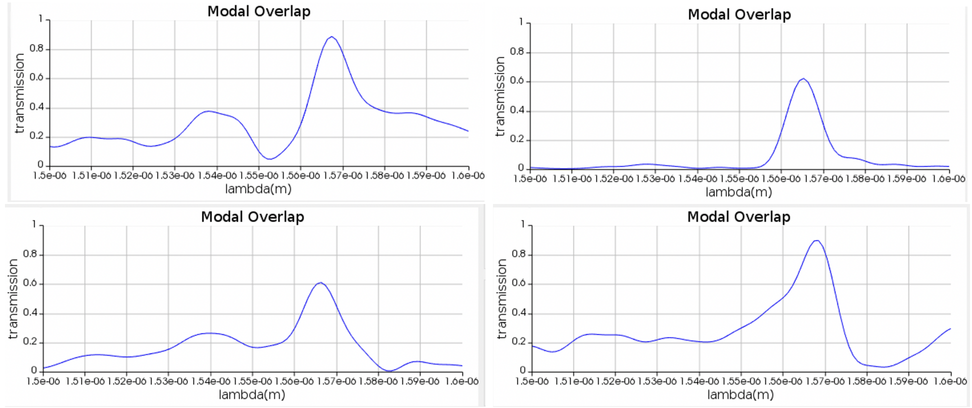


Figure 4.24: Optical transmission results in Ansys/Lumerical (linear scale). Modal overlap at the two outputs for the two-mode thermal switch at low temperatures for (**a**) TE1 and for (**b**) TE0. At high temperature for (**c**) TE1 and for (**d**) TE0

4.6 Summary

In this chapter, a 2-D optimized thermal switch is presented. By working only in 2-D simulation, one can further improve the optimization process while reducing computational cost. In this chapter, we present an initial condition that pushes the devices towards a higher silicon-to-silica ratio causing greater refractive index changes per unit area. These devices with more silicon perform better than the unbiased or silica heavy devices, which have 40 to 70% transmission at the output ports. Additionally, the length of the devices is increased for a larger phase change. With these improvements, a device is generated that has 80% transmission at one output port and 90% transmission at the opposite output port. Using the temperature coefficient of silicon, a novel device is generated that increases the switching capabilities within integrated silicon chips by decreasing the footprint size to approximately 8 μm by 3 μm. Further, it is seen that the switch also acts as a de-multiplexer with a spectral shift at higher temperatures, making this device multifunctional. Another device that is proposed in this chapter is a two-mode thermal switch, which acts as a de-multiplexer for different optical modes as well as a thermal switch for the two modes. Given the novelty of these devices, they should be further studied and offer insight into the future of inverse design utilizing the thermal material properties of silicon.

Chapter 5

3-D Simulation Results of the Thermal Switch

Using 2-D optimization for inverse design is a useful proof of concept. It allows designers to explore initial conditions and device parameters while acquiring intuition to improve device optimization at low computational cost. However, to get an accurate sense of the device performance, the optimizer must be simulated in 3-D using the same figure of merit. This methodology will generate devices that are fabricable. Indeed, the infinite z-component assumption used in 2-D simulations is no longer valid. The simulations will now calculate Maxwell's equations on a device with a thickness of 0.22 μm, standard for silicon-on-insulator photonic chips. These simulations are more computationally expensive and must be run on a GPU server. Running these 3-D optimizations ensures that the devices sent for fabrication are more functional and reliable.

5.1 Device Design in 3-D Simulation

With a general sense of performance from the 2-D optimization, the optimization with 3-D simulations is more easily set up. Compared to 2-D simulations, 3-D simulations can take multiple days to optimize on a GPU server, with one iteration taking anytime from three minutes up to an hour. Hence, it is important that these devices are realized in 2-D beforehand such that any detrimental mistakes in the initial conditions (e.g., using TM0 modes instead of TE0) can be realized and corrected early before progressing with a computational heavy step. The first completely 3-D optimized thermal switch device has the following layout in Lumerical (fig. 5.1).

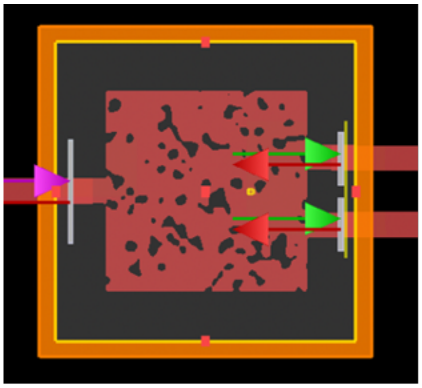
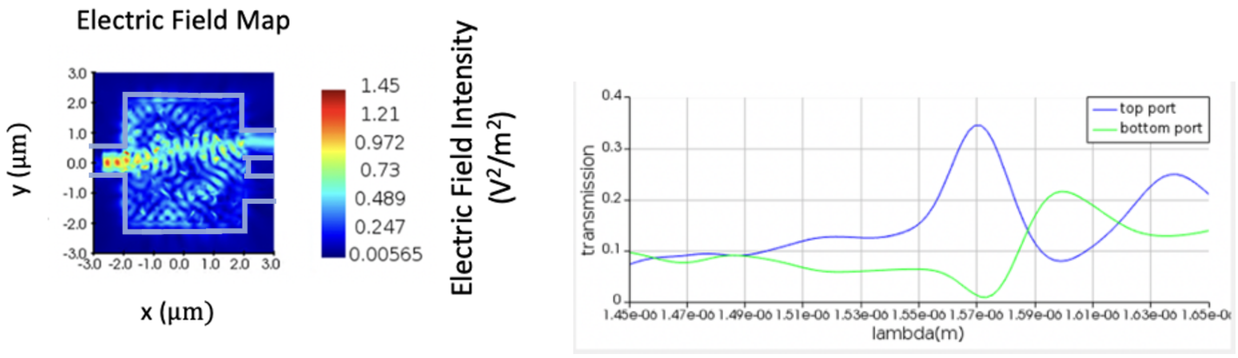


Figure 5.1: Permittivity map of a 3-D optimized device in Lumerical after exportation from SPINS.

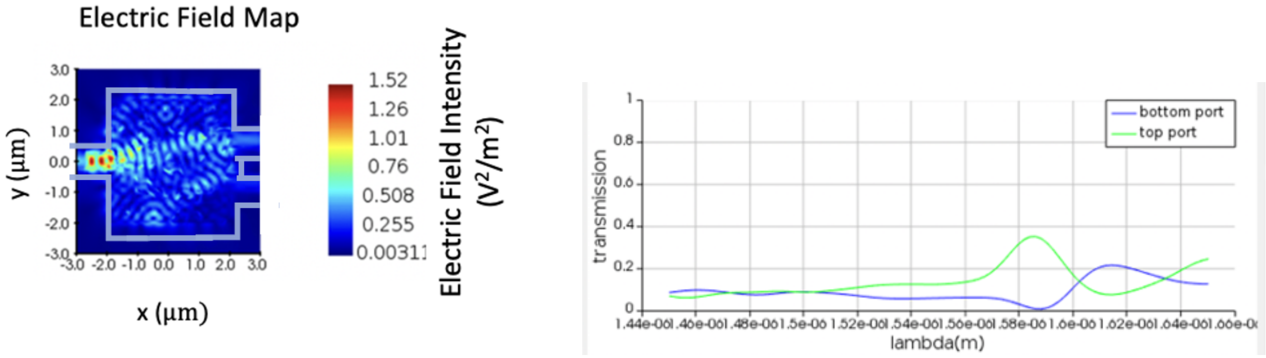
The design space for the device in figure 5.1 is 4 µm by 4 µm. The output waveguide separation is 0.8 µm. The temperature difference is 200 K. The initial optimization results are poor as they are obtained before the changes, detailed in chapter 4, are implemented into the optimization framework, e.g., closer output waveguides, higher silicon bias, a thinner and longer design space. Indeed, at low temperature, the device optical transmission results in figure 5.2 show poor performance. At high temperature, poor results are also obtained with large losses and weak optical transmission at the designated output ports (fig. 5.3).

The 4 µm by 4 µm device performs poorly. This situation is similar to the structures optimized in 2-D before implementing the device improvements in sections 4.3 and 4.4: thinner width with longer length design spaces, high silicon bias and closer distance between output waveguides. Optimally, the 3-D optimization would perform similarly to the best performing 2-D device, which had at most a 1 dB insertion loss. To further the case for thin-width long optimized structures, which is utilized in the best performing



(a) **(b)**

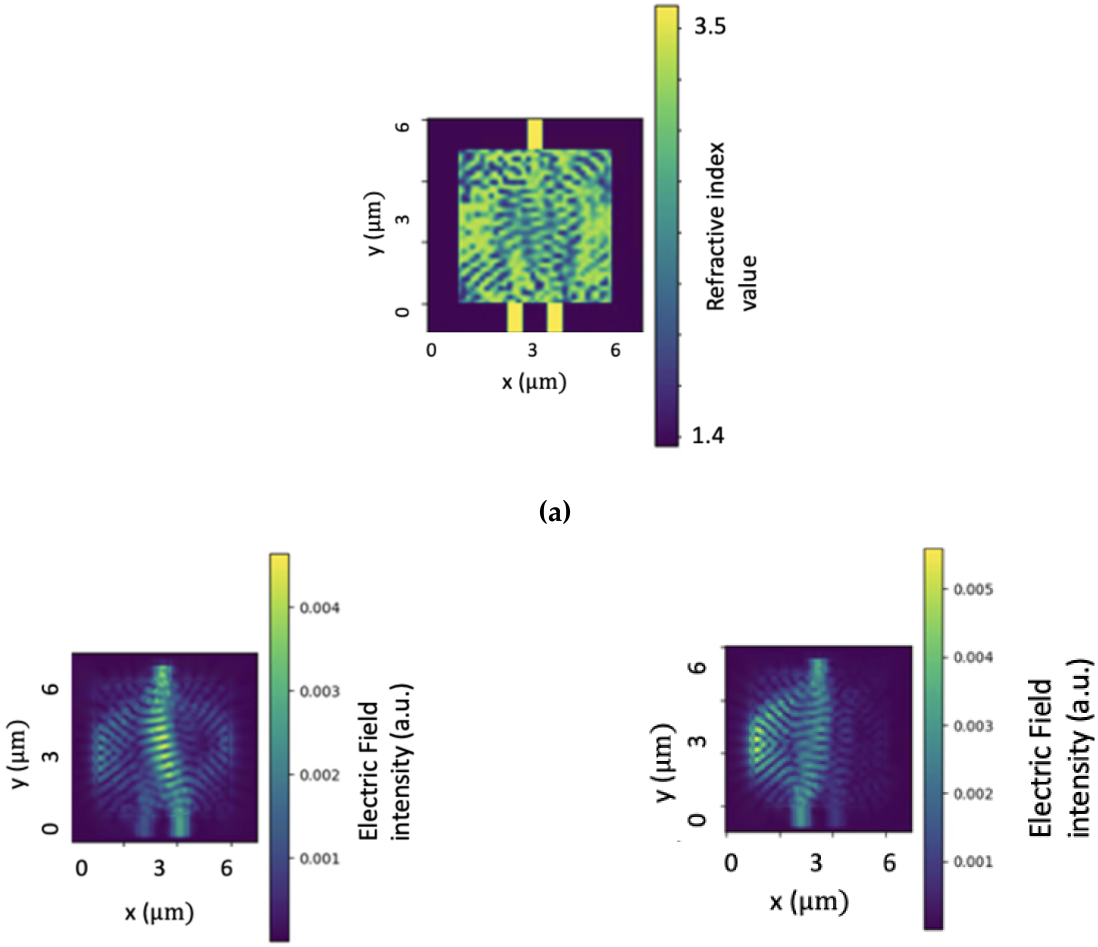
Figure 5.2: Lumerical results for a 3-D optimized device at low temperature: **(a)** the electric field map and **(b)** the optical transmission (in the linear scale) at both output ports.



(a) **(b)**

Figure 5.3: Lumerical results for a 3-D optimized device at high temperature;**(a)** the electric field map and **(b)** the optical transmission (in linear scale) at both output ports.

2-D device (described in section 4.4), the electric field maps at various stages of binarization can be monitored. This allows one to see patterns and possible pathways to a better optimized device based on how the field moves in the structure with low and high binarization. In the first binarization step of the 3-D optimized thermal switch, where the sigmoid factor, c_1, is equal to 1, the resulting outputs from SPINS are shown in figure 5.4. In this figure, the field propagates across the entire design space, close to the



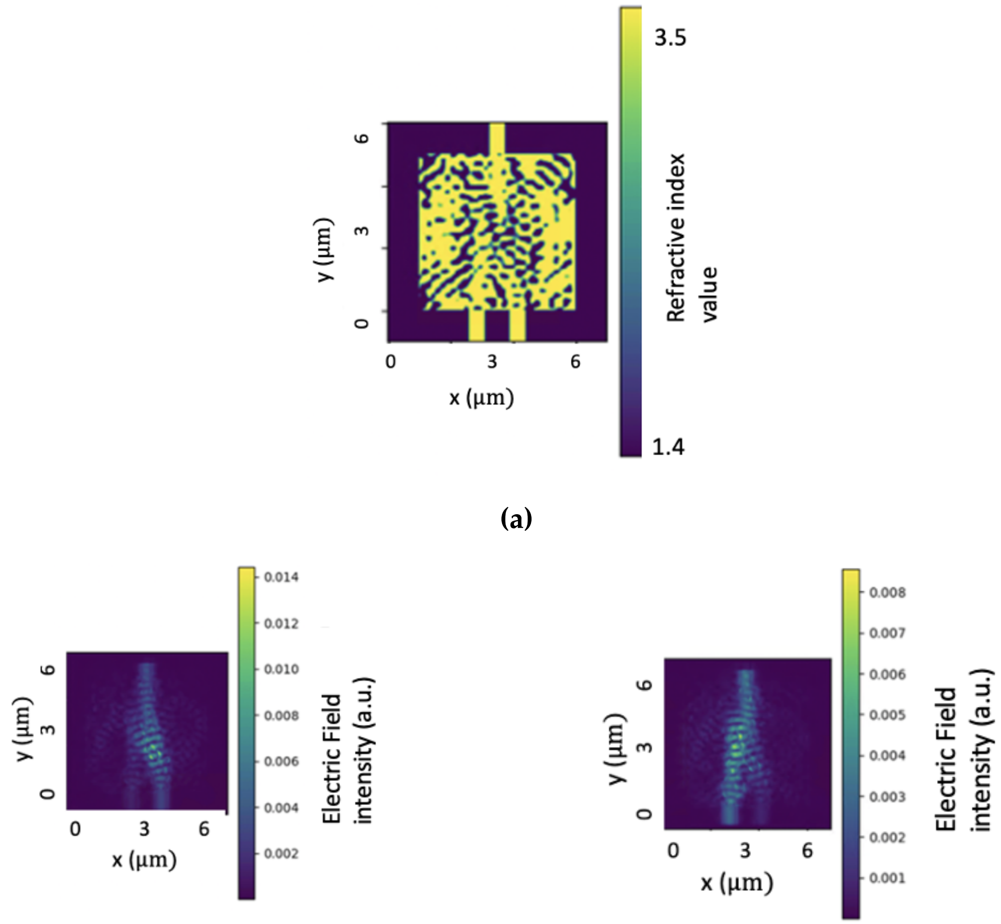
(a)

(b)

(c)

Figure 5.4: SPINS results for the 3-D optimized device with low binarization; (**a**) the permittivity map, and field maps at (**b**) low temperature and (**c**) high temperature.

boundaries between the design region and the silica cladding. With higher binarization for the same device, a different field propagation is observed. The field seen in figure 5.5 is more confined to the central region of the device due to the high confinement of silicon at the silica/silicon boundaries. To further understand this, note that the boundaries in a low binarized device result in smaller refractive index changes (between one material with an in-between permittivity and another) while a high binarized device results in



(a)

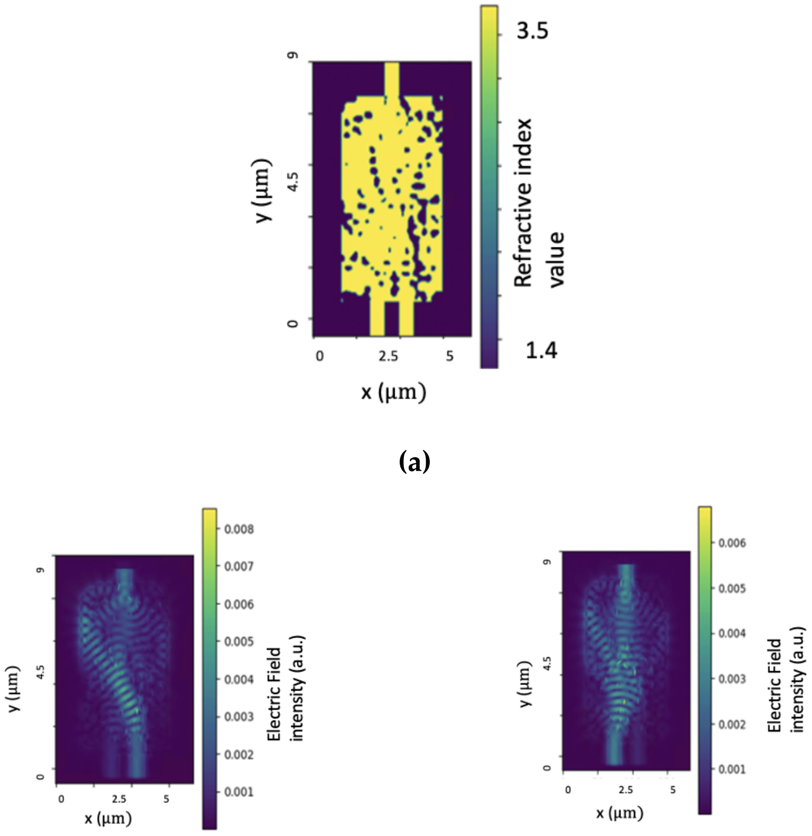
(b) **(c)**

Figure 5.5: SPINS results for the 3-D optimized device with high binarization; (**a**) permittivity map, and field maps at (**b**) low temperature and (**c**) high temperature.

larger refractive index changes (between silicon, 3.45, and silica, 1.44). Remember that a high propagation constant corresponds to higher light confinement. With the propagation constant, β, being proportional to n_{eff} and higher n_{eff} correlating with higher refractive index changes between core, cladding, and substrate, one can conclude that high binarization, with the largest refractive index change, corresponds to a higher propagation constant and, hence, more confinement of light in the silicon. Knowing this about the device, the optimization is altered to reduce the design space width, which will allow the optimizer to explore longer and narrower design spaces. This leads to similar light paths in both low and high binarization such that there is no decay in device performance when moving to higher binarization due to significant path changes from more abrupt boundaries. Having this restriction makes it more straightforward for the optimizer to create well-binarized devices and cut down on the computational cost of simulating larger devices. In all following the 3-D devices, only devices with a smaller width and longer length will be presented.

A device with a design region of 3 μm by 6 μm is initially looked at. To reduce the computational cost of larger devices in 3-D optimization, since the 3 μm by 6 μm increases the design area by 2 μm^2 from the previous 4 μm by 4 μm design, 2-D simulations are first done. To incorporate and benefit from 2-D simulation speed, a 2-D optimization can run during the first binarization step, when the design is still in its grey-scale optimization phase. In this approach, the designer leverages the quickness of the 2-D optimization and then changes to 3-D simulations for all following binarization stages. However, the source must be carefully considered when using this methodology as the fundamental mode changes from TM0 to TE0 when moving from 2-D to 3-D. Hence, when optimizing in the grey-scale within the 2-D simulation, the mode with the second highest propagation constant (TE0) must be injected given the infinite z-component (section 3.1). Then in 3-D optimization, the simulation must use the mode with the highest propagation constant (TE0). When moving from 2-D to 3-D in the optimization process, maintaining consistency in source definition is important to obtain a well performing device. Incorporating both 2-D to 3-D optimization with an increased length yields the following results from SPINS as demonstrated in figure 5.6.

An important distinction between this device and the previous 4 μm by 4 μm device, is that another modification to the optimization process is used here. Here, the initial conditions have been changed for silicon bias, driving the device to have a higher silicon-to-silica ratio. On Lumerical, the field profiles are obtained at low and high temperature, respectively, in figure 5.7. The optical transmission for high and low temperatures for the top and bottom output ports are shown in figures 5.8 and 5.9. At low temperature

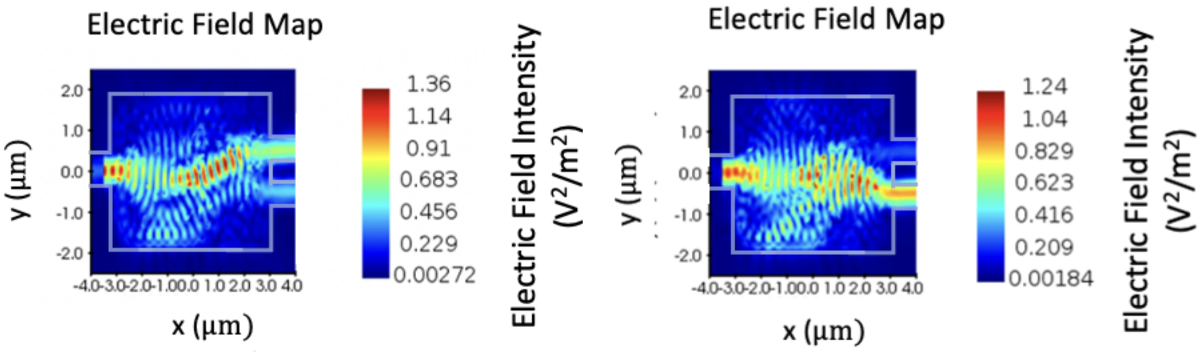


(a)

(b) **(c)**

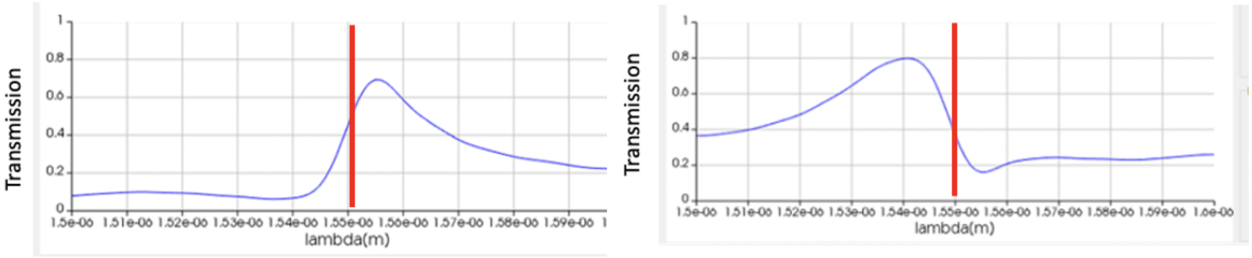
Figure 5.6: SPINS results for a 3-D optimized device with the (**a**) permittivity map and field maps at (**b**) low temperature and (**c**) high temperature at 1.55 μm wavelength.

(fig. 5.8), there is an insertion loss of approximately 1.5 dB and at high temperature (fig. 5.9), the insertion loss is approximately 1 dB. What is also important to note is that at peak transmission in one output port, the opposite output port falls to a minimum in transmission. This is important, as one goal of the optimization is to minimize the output at the opposite output port while maximizing the transmission through the designated output port. Leakage transmission in the wrong output port is detrimental to switching as it reduces the transmission through the target output port and impacts the data signal integrity within an optical system.



(a) **(b)**

Figure 5.7: Lumerical electric field maps for the 3-D optimized device at (**a**) low temperature and (**b**) high temperature at 1.550 µm wavelength.

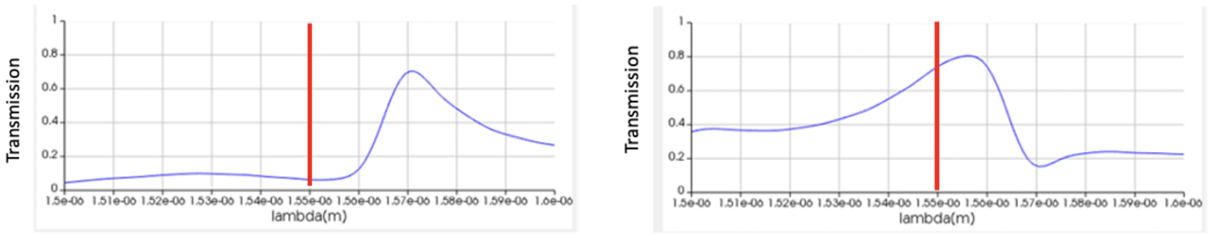


(a) **(b)**

Figure 5.8: Optical transmission (linear scale) for the 3-D optimized device in figure 5.7 at low temperature for (**a**) the top output port and (**b**) the bottom output port. The red line is the wavelength of optimization.

Similarly to the device detailed in section 4.4, increasing range of the peak optical transmission spectrum of these devices allows them to operate in a broadband capacity. The same methodology is carried out as with the 2-D optimized device: two more objectives are added to the optimizer with a 6 nm wavelength spacing between the objectives at each temperature.

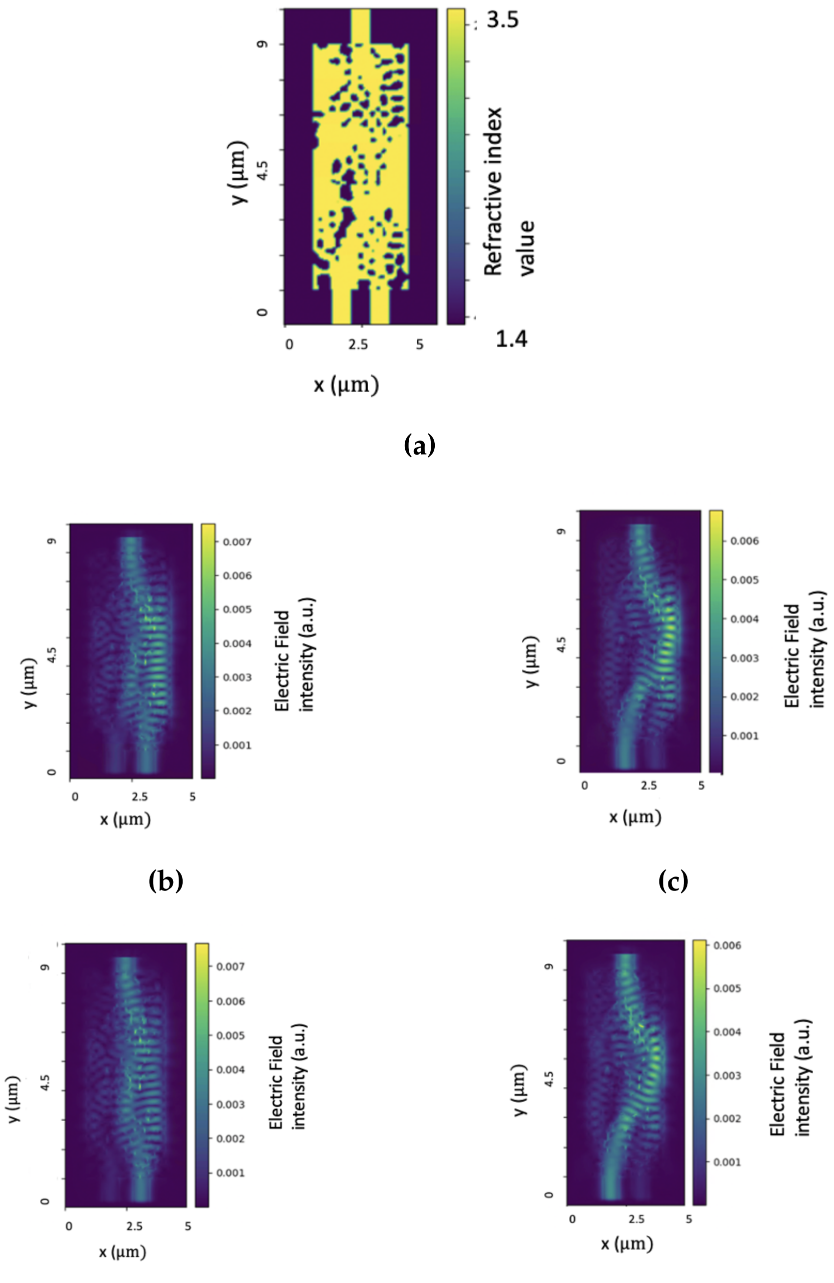
The device in figures 5.10 is 7.5 µm by 2.5 µm and is optimized at two different wavelength points (1.547 µm and 1.553 µm) for broadband operation. The computational time



(a) **(b)**

Figure 5.9: Optical transmission (linear scale) for the optimized 3-D device in figure 5.7 at high temperature for (**a**) the top output port and (**b**) the bottom output port. The red line is the wavelength of optimization.

increases greatly with two additional 3-D simulations doubling the computational time. Hence, the mesh is set to 40 nm to cut down on computational cost. The output waveguides are 500 nm apart and 500 nm in width, with the input waveguide also 500 nm in width. When exported to Lumerical, the field map is obtained at 1.55 µm wavelength and the following results are shown in figure 5.11. Additionally, when observing the optical transmission in figure 5.12, a characteristic plateau can be seen in the optical transmission peaks over a 6 nm wavelength range. This device still exhibits de-multiplexing properties, as detailed in section 4.3, with the light propagating through the top output port for longer wavelengths and the bottom output port for shorter wavelengths. This device is much better performing than the initial 4 µm by 4 µm device generated in 3-D optimization. The device modifications in the initial conditions, such as a smaller width, higher silicon bias, and an optimization at two wavelengths, produces a well-performing device that can be later used for fabrication.



(a)

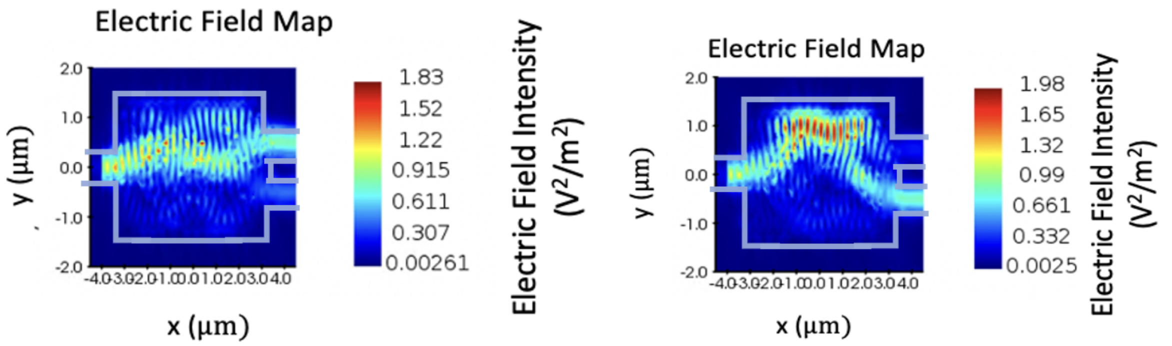
(b)

(c)

(d)

(e)

Figure 5.10: SPINS results for a 3-D optimized device. (**a**) Permittivity map and field maps at (**b**) low temperature and (**c**) high temperature at 1.547 μm wavelength. Field maps at (**d**) low temperature and (**e**) high temperature at 1.553 μm wavelength



(a) **(b)**

Figure 5.11: Lumerical electric field maps for the 3-D optimized device for broader spectral transmission at (**a**) low temperature and (**b**) high temperature at 1550 µm wavelength.

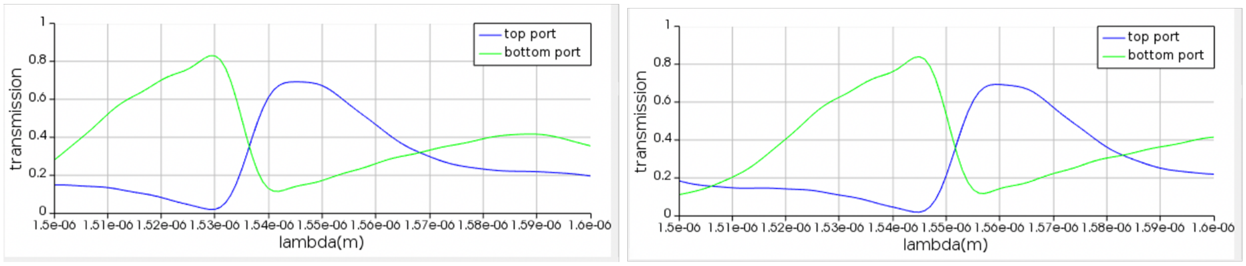


Figure 5.12: Lumerical optical transmission (linear scale) for the 3-D optimized device at (**a**) the top output port and (**b**) the bottom output port at high temperatures.

5.2 Comparison of conventional optical switches to inverse designed thermal switch

In this thesis, an inverse designed optical thermal switch is presented for optical interconnects with a reduced device footprint. A thermal optical switch is produced that operates at a ΔT of approximately 200 K with a device size of 7.5 µm by 2.5 µm. In the past, many different forms of optical switching have been used, from off-chip free space switching, such as a MEMS-based switch, to using materials with significant phase changes from applied voltage or temperature changes such as PCM (phase change material) switches.

Additionally, in optical switching, silicon devices such as MRR or MZI switches, which use the thermo-optic and electro-optic properties of silicon for switching, have been used. In this section, a comparison of modern optical switches with the inverse designed thermal switch is provided.

To reduce the large footprints needed for optical switching the usage of off-chip switching should reduced for optical interconnects since it relies on fiber-to-chip and chip-to-fiber interfaces such as grating couplers, which can be hundreds of microns in length and width. Additionally grating couplers have large insertion losses with a loss of approximately 3 dB at 1550 nm. It is important to minimize the utilization of off-chip functions as they generate a need for amplifiers within optical systems to make up for such losses. Additionally, a free space MEMS switch for optical switching can take up to tens of millimeters for switching functions [50]. The inverse designed thermal switch not only reduces the need for off-chip switching which would generate a need for additional grating couplers, but is significantly smaller.

Similar to off-chip switching, the inverse designed thermal switch can drastically reduce the size requirements of on-chip switching. For a MRR silicon-on-insulator switch, the footprint is around 2.5 mm^2 for a 4×4 to 8×8 switch [51]. Additionally, this switch operates at a significantly smaller bandwidth, with optical transmission peaking in only a 0.5 nm wavelength span. The inverse designed thermal switch offers a 6 nm wavelength span for peak optical transmission through the two output ports. MZI-based switches typically have a footprint area of 60×30 μm^2, while the inverse designed thermal switch is about 100× smaller in terms of area [52]. Another option for on-chip switching relies on PCMs for MZI and MRR structures, which has the capability to reduce size requirements, due to higher phase changes in PCMs, to sizes of around 100 μm^2. However, PCMs such as vanadium dioxide are fairly new to fabrication processes for silicon chips which can result in detrimental fabrication errors on the device. Additionally, PCMs have significantly high losses, with up to 2 dB of loss per 1×2 or 2×2 switch whereas the inverse designed thermal switch has a 1 dB insertion loss.

For the inverse designed thermal switch presented in this thesis, the footprint of the device is significantly reduced by as much as 100× less for on-chip switching relative to other alternatives. Additionally, there is an increased broadband capacity in the inverse designed thermal switched, 10× larger than that of modern MRR switches. With PCM technology vastly behind that of well-established silicon foundries, it is clear that silicon on-chip switching could be the future of data routing. The inverse designed thermal switch offers a new pathway to decrease size requirements while maintaining data capacities of silicon devices.

5.3 Summary

The 3-D simulations offer an in-depth analysis into how the thermal switch may work experimentally. By optimizing the thermal switch using 3-D simulations, assurance can be offered that the device geometry works in 3-D without the infinite z-component assumption in 2-D. Normally computationally costly, it is found that by starting the optimization process in 2-D for low binarization and then moving to 3-D optimization for high binarization, some of the computational cost can be reduced. Further, the improved specifications found in the 2-D simulations are implemented, such as high silicon bias and longer devices, to obtain devices that operate at 70% transmission at low temperature and 80% transmission at high temperature. Compared to other modern switches, the thermal switch presented in this thesis offers a footprint that is 100× smaller. Additionally, the inverse designed thermal switch offers a completely on-chip silicon device, which reduces foundry process variations and the losses seen with off-chip switching. The 3-D results serve as an important proof of concept, showing that not only is temperature dependence possible, but helpful for reducing the footprint of switching devices.

Chapter 6

Conclusion

6.1 Summary of Work

In this thesis, the following was accomplished; 1) an in-depth discussion on simulation methodologies used in inverse design, and 2) a novel inverse designed thermal switch. The thermal optical switch was validated in 2-D and 3-D simulation. Additionally, improvements made to the initial conditions of the optimization framework were presented and discussed.

The inverse design process detailed in this thesis uses FDFD simulations for FOM measurements in the SPINS framework and then FDTD simulations for device testing using Ansys/Lumerical. Discussed in chapter 3, there are several simulation conditions to consider when running an optimization including 2-D versus 3-D simulations, meshing, and the binarization process. In 2-D and 3-D simulations, mode calculations results differ due to the assumption made in the 2-D domain of an infinite z-component. This infinite z-component gives rise to a higher propagation constant for TM0 such that TM0 is defined as the fundamental mode. In 3-D, the TE0 mode has the highest propagation constant and therefore the fundamental mode changes between 2-D and 3-D simulations from TM0 to TE0. Additionally, it is important to consider meshing choices during optimization and device validation. In this thesis, a uniform volume average mesh in Lumerical is used for device validation as this mesh choice leads to similar spectral optical transmission results compared to results in SPINS. Finally, it is important to consider the exportation of the device and how a level-set binarization will impact the device. As seen in section 3.2, when exporting a device from SPINS, any non-binarized points of the device will be completely binarized based upon how the level-set is defined. A spectral shift is seen at different level-sets, hence it is important to choose a level set that maintains the optical transmission peak at the optimized wavelength. The peak optical transmission remains most stable at level-set values between 0.65 to 0.60. With these concepts thoroughly explored, the optimization for the thermal optical switch can then be set up and observed.

The inverse designed thermal optical switch is a novel device that has not yet been proposed nor studied in the literature. Utilizing 2-D simulations, a significant proof of concept is constructed. These 2-D optimizations and simulations not only offer a proof of concept, but also offer a computationally efficient way to explore the initial conditions of the device to generate a better performing design. The first device that was explored had a 4 μm by 4 μm design space with temperature difference of 600 K. The temperature difference was reduced down to 200 K for all further devices. This was decided upon to generate a device that could be fabricated and validated experimentally but also had a large enough temperature difference for a significant refractive index change in the silicon. Further, to increase device performance, the initial conditions of the device were explored. The design space was increased such that light propagating within the device could undergo a larger phase shift. The device was lengthened to achieve this, but the width was decreased to cut down on computational cost. Additionally, the two output waveguides were placed closer together for a decreased distance that the light path would need to change from temperature variations. To increase the transmission of the device, changing the initial conditions of the permittivity matrix were explored in SPINS. By changing the initial conditions to have a silicon bias, devices with higher silicon-to-silica ratios are produced. Silicon offers a much higher temperature coefficient compared to silica, as discussed in section 4.1. This is used to the advantage of the device, as the greater change in index values generates a more significant phase change in the device. Additionally, the optical band of the device can be increased by adding additional objectives to the FOM at different wavelengths with a 6 nm difference. Using these changes, a 2-D optimized thermal optical switch is generated with an optical transmission of 80-90%. Another device is also proposed in chapter 4, a two-mode thermal optical switch. This device acts as both a mode de-multiplexer and a switch, where the optical modes TM0 and TM1 switch output ports dependent on temperature. This device is yet to be validated in 3-D, as it has a high computational cost due to the amount of objectives for the FOM and significant design space.

3-D optimization has a much higher computational cost compared to 2-D optimization and is hence run on a GPU server. To reduce the computational cost of 3-D optimization, 2-D optimization is used at the starting point of the optimization at low binarization, then 3-D optimization is used at higher binarization levels. In chapter 5, the benefits of the changes to the initial conditions proposed in chapter 4 are further explored in the 3-D realm, ultimately resulting in a broadband device with a 7.5 μm by 2.5 μm design space and a 200 K temperature difference. The 3-D inverse designed optical thermal switch has a 1.5 dB insertion loss at low temperature and 1 dB insertion loss for high temperature in simulation. Compared to other optical switches, such as the MEMS based and MZI-MMI switches, the insertion loss is similar except that the inverse designed thermal switch is much more compact than these switches, which can take up to 100 μm in length. The 3-D inverse design optical thermal switch offers promising results in both 2-D and 3-D simulations and should be further explored.

6.2 Future Work

The thermal switch yields promising results in simulation. For further validation of the thermal optical switch device, the device must be sent for fabrication at a photonic foundry and tested experimentally. In the experimental set up, it is important to consider how the temperature difference can be applied to the device. There are two options for applying a temperature difference: 1) a thermal stage on which the chip is placed upon or 2) a metal cladding within the chip. A thermal stage is not practical for application in photonic systems, as it takes a significant amount of space within the set up. However, the thermal stage is easy to use and offers a good level of control without risk of damaging the chip. For metal cladding on the chip, the dimensions of the metal piece must be considered when applying a current, as to not damage the metal pads and routing. For both, the thermal stage and metal layering, a thermal map must be generated in simulation to estimate the amount of heat that will be applied to the device. Once the thermal component of the experimental set up is decided upon, the design can be fabricated and tested. When tested, a C-band laser should be used such that any spectral shifts in peak wavelength due to fabrication errors can be accounted for and discussed. Ultimately, we hope that the experimental results will be able to validate the insertion loss and broadband behaviour that has been achieved in simulation.

The thermal switch has many avenues for future simulation experimentation. Given that there is higher insertion loss in 3-D optimization than in 2-D optimization, there is a case for future work exploring these discrepancies. Additionally, work can be done to increase the transmission in both 2-D and 3-D optimization, exploring the initial conditions of the device and the binarization steps. Further, methodologies on cutting down on the computational cost should be explored. Inverse designed devices using FDFD simulations for the optimization process require a significant amount of computation for broadband optical transmission. Increasing the design space for the optimization also increases the computational cost. For exploring devices such as the two-mode thermal switch, discussed in section 4.6, the computational cost must be reduced.

Exploring the benefits and limitations of inverse design is fundamental for the future of photonics. As we garner a need for smaller broadband devices, it is important to further utilize inverse design for increasing the design library. The inverse designed thermal optical switch offers an insight into how inverse design can be used for novel devices and how future devices can be created.

Bibliography

[1] J. Shalf, "The future of computing beyond moore's law," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 378, no. 2166, p. 20190061, 2020.

[2] Y. Shi, Y. Zhang, Y. Wan, Y. Yu, Y. Zhang, X. Hu, X. Xiao, H. Xu, L. Zhang, and B. Pan, "Silicon photonics for high-capacity data communications," *Photon. Res.*, vol. 10, no. 9, pp. A106–A134, Sep 2022. [Online]. Available: https://opg.optica.org/prj/abstract.cfm?URI=prj-10-9-A106

[3] C. Fiandrino, D. Kliazovich, P. Bouvry, and A. Zomaya, "Performance and energy efficiency metrics for communication systems of cloud computing data centers," *IEEE Transactions on Cloud Computing*, vol. 5, pp. 738 – 750, 12 2017.

[4] D. Munk, M. Katzman, Y. Kaganovskii, N. Inbar, A. Misra, M. Hen, M. Priel, M. Feldberg, M. Tkachev, A. Bergman, M. Vofsi, M. Rosenbluh, T. Schneider, and A. Zadok, "Eight-channel silicon-photonic wavelength division multiplexer with 17 ghz spacing," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 25, no. 5, pp. 1–10, 2019.

[5] Y. Ding, V. Kamchevska, K. Dalgaard, F. Ye, R. Asif, S. Gross, M. J. Withford, M. Galili, T. Morioka, and L. K. Oxenløwe, "Reconfigurable SDM switching using novel silicon photonic integrated circuit," *Scientific Reports*, vol. 6, no. 1, 2016.

[6] H. Wang, H. Chai, Z. Lv, Z. Zhang, L. Meng, X. Yang, and T. Yang, “Silicon photonic transceivers for application in data centers,” *Journal of Semiconductors*, vol. 41, no. 10, p. 101301, 2020.

[7] M. Duranton, D. Dutoit, and S. Menezo, “Key requirements for optical interconnects within data centers,” in *Optical Interconnects for Data Centers*, ser. Woodhead Publishing Series in Electronic and Optical Materials, T. Tekin, R. Pitwon, A. Håkansson, and N. Pleros, Eds. Woodhead Publishing, 2017, pp. 75–94. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780081005125000036

[8] M. Dayarathna, Y. Wen, and R. Fan, “Data center energy consumption modeling: A survey,” *IEEE Communications Surveys amp;amp; Tutorials*, vol. 18, no. 1, p. 732–794, 2016.

[9] M. Tan, Y. Wang, K. X. Wang, Y. Yu, and X. Zhang, “Circuit-level convergence of electronics and photonics: basic concepts and recent advances,” *Frontiers of Optoelectronics*, vol. 15, no. 1, p. 16, Apr. 2022. [Online]. Available: https://doi.org/10.1007/s12200-022-00013-8

[10] T. Tamplin, “Moores law: Definition, history, and limitations,” Mar 2023. [Online]. Available: https://www.financestrategists.com/wealth-management/moores-law/

[11] A. F. J. Levi, “Silicon photonics’ last-meter problem: Economics and physics still pose challenges to ”fiber to the processor” tech,” *IEEE Spectrum*, vol. 55, no. 9, pp. 38–43, 2018.

[12] S. Mao, L. Cheng, C. Zhao, F. N. Khan, Q. Li, and H. Y. Fu, “Inverse design for silicon photonics: From iterative optimization algorithms to deep neural networks,” *Applied Sciences*, vol. 11, no. 9, 2021. [Online]. Available: https://www.mdpi.com/2076-3417/11/9/3822

[13] R. C. Loonen, S. de Vries, and F. Goia, “15 - inverse design for advanced building envelope materials, systems and operation,” in *Rethinking Building Skins*, ser. Woodhead Publishing Series in Civil and Structural Engineering, E. Gasparri, A. Brambilla, G. Lobaccaro, F. Goia, A. Andaloro, and A. Sangiorgio, Eds. Woodhead Publishing, 2022, pp. 377–402. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B978012822477900022X

[14] R. Shiratori, M. Nakata, K. Hayashi, and T. Baba, “Particle swarm optimization of silicon photonic crystal waveguide transition,” *Opt. Lett.*, vol. 46, no. 8, pp. 1904–1907, Apr 2021. [Online]. Available: https://opg.optica.org/ol/abstract.cfm?URI=ol-46-8-1904

[15] X. Li, M. Zhang, H. Guo, Z. Shi, Y. Guo, T. Zhao, and A. Wang, “Parameter optimization for modulation-enhanced external cavity resonant frequency in fiber fault detection,” *Photonics*, vol. 10, no. 7, 2023. [Online]. Available: https://www.mdpi.com/2304-6732/10/7/822

[16] S. J. Cox and D. C. Dobson, “Maximizing band gaps in two-dimensional photonic crystals,” *SIAM Journal on Applied Mathematics*, vol. 59, no. 6, pp. 2108–2120, 1999. [Online]. Available: http://www.jstor.org/stable/118418

[17] M. Y. Wang, X. Wang, and D. Guo, “A level set method for structural topology optimization,” *Computer Methods in Applied Mechanics and Engineering*, vol. 192, no. 1, pp. 227–246, 2003. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0045782502005595

[18] J. Jensen and O. Sigmund, “Topology optimization of photonic crystal structures: A high-bandwidth low-loss T-junction waveguide,” *Journal of The Optical Society of America B-optical Physics*, vol. 22, 06 2005.

[19] M. B. Giles and N. A. Pierce, “An Introduction to the Adjoint Approach to Design,” *Flow, Turbulence and Combustion*, vol. 65, no. 3, pp. 393–415, Dec. 2000. [Online]. Available: https://doi.org/10.1023/A:1011430410075

[20] K. Y. Yang, C. Shirpurkar, A. D. White, J. Zang, L. Chang, F. Ashtiani, M. A. Guidry, D. M. Lukin, S. V. Pericherla, J. Yang, H. Kwon, J. Lu, G. H. Ahn, K. Van Gasse, Y. Jin, S.-P. Yu, T. C. Briles, J. R. Stone, D. R. Carlson, H. Song, K. Zou, H. Zhou, K. Pang, H. Hao, L. Trask, M. Li, A. Netherton, L. Rechtman, J. S. Stone, J. L. Skarda, L. Su, D. Vercruysse, J.-P. W. MacLean, S. Aghaeimeibodi, M.-J. Li, D. A. B. Miller, D. M. Marom, A. E. Willner, J. E. Bowers, S. B. Papp, P. J. Delfyett, F. Aflatouni, and J. Vučković, “Multi-dimensional data transmission using inverse-designed silicon photonics and microcombs,” *Nature Communications*, vol. 13, no. 1, p. 7862, Dec. 2022. [Online]. Available: https://doi.org/10.1038/s41467-022-35446-4

[21] 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, Nov. 2018. [Online]. Available: https://doi.org/10.1038/s41566-018-0246-9

[22] “How to squeeze billions of transistors onto a computer chip,” Accessed: August 2023. [Online]. Available: https://www.ibm.com/thought-leadership/innovation-explanations/mukesh-khare-on-smaller-transistors-analytics

[23] H. Mekawey, M. Elsayed, Y. Ismail, and M. A. Swillam, “Optical interconnects finally seeing the light in silicon photonics: Past the hype,” *Nanomaterials*, vol. 12, no. 3, p. 485, 2022.

[24] M. Duranton, D. Dutoit, and S. Menezo, “3 - key requirements for optical interconnects within data centers,” in *Optical Interconnects for Data Centers*, ser. Woodhead Publishing Series in Electronic and Optical Materials, T. Tekin, R. Pitwon, A. Håkansson, and N. Pleros, Eds. Woodhead Publishing, 2017, pp.

75–94. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780081005125000036

[25] M. S. Nawrocka, T. Liu, X. Wang, and R. R. Panepucci, “Tunable silicon microring resonator with wide free spectral range,” *Applied Physics Letters*, vol. 89, no. 7, p. 071110, Aug. 2006. [Online]. Available: https://doi.org/10.1063/1.2337162

[26] 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

[27] 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, 2020.

[28] Q. Cheng, S. Rumley, M. Bahadori, and K. Bergman, “Photonic switching in high performance datacenters,” *Opt. Express*, vol. 26, no. 12, pp. 16 022–16 043, Jun 2018. [Online]. Available: https://opg.optica.org/oe/abstract.cfm?URI=oe-26-12-16022

[29] A. M. Al-Hetar, A. S. M. Supa’at, A. Mohammad, and I. Yulianti, “Crosstalk improvement of a thermo-optic polymer waveguide MZI–MMI switch,” *Optics Communications*, vol. 281, no. 23, pp. 5764–5767, 2008. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0030401808008122

[30] G. Kim, J. W. Park, I. G. Kim, S. Kim, K.-S. Jang, S. A. Kim, J. H. Oh, J. Joo, and S. Kim, “Compact-sized high-modulation-efficiency silicon Mach-Zehnder modulator based on a vertically dipped depletion junction phase shifter for chip-level integration,” *Opt. Lett.*, vol. 39, no. 8, pp. 2310–2313, Apr 2014. [Online]. Available: https://opg.optica.org/ol/abstract.cfm?URI=ol-39-8-2310

[31] L. Fan, S. Gloeckner, P. Dobblelaere, S. Patra, D. Reiley, C. King, T. Yeh, J. Gritters, S. Gutierrez, Y. Loke, M. Harburn, R. Chen, E. Kruglick, M. Wu, and A. Husain, “Dig-

ital mems switch for planar photonic crossconnects," in *Optical Fiber Communication Conference and Exhibit*, 2002, pp. 93–94.

[32] S. Chen, Y. Shi, S. He, and D. Dai, "Low-loss and broadband silicon thermo-optic Mach-Zehnder switch with bent directional couplers," *Opt. Lett.*, vol. 41, no. 4, pp. 836–839, Feb 2016.

[33] D. J. Griffiths, *Introduction to electrodynamics*. Cambridge University Press, 2023.

[34] R. Scarmozzino, A. Gopinath, R. Pregla, and S. Helfert, "Numerical techniques for modeling guided-wave photonic devices," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 6, no. 1, pp. 150–162, 2000.

[35] R. Rumpf, "Electromagnetic analysis using finite-difference time-domain," Jan 2023. [Online]. Available: https://empossible.net/academics/emp5304/

[36] "Finite-Difference Time Domain FDTD solver introduction," Accessed: Aug. 2023. [Online]. Available: https://optics.ansys.com/hc/en-us/articles/360034914633-Finite-Difference-Time-Domain-FDTD-solver-introduction

[37] "Understanding mesh refinement and conformal mesh in FDTD," accessed: Aug. 2023. [Online]. Available: https://optics.ansys.com/hc/en-us/articles/360034382594-Understanding-Mesh-Refinement-and-Conformal-Mesh-in-FDTD

[38] E. Alpaydin, *Introduction to machine learning*. The Mit Press, 2020.

[39] J. Jensen and O. Sigmund, "Topology optimization for nano-photonics," *Laser & Photonics Reviews*, vol. 5, no. 2, pp. 308–321, 2011. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/lpor.201000014

[40] C. Pollock and M. Lipson, "Integrated photonics," 01 2003.

[41] R. Hauffe and K. Petermann, *Thermo-Optic Switching*. Boston, MA: Springer US, 2006, pp. 111–139. [Online]. Available: https://doi.org/10.1007/0-387-29159-8_4

[42] D.-X. Xu, A. Delâge, P. Verly, S. Janz, S. Wang, M. Vachon, P. Ma, J. Lapointe, D. Melati, P. Cheben, and J. H. Schmid, "Empirical model for the temperature dependence of silicon refractive index from O to C band based on waveguide measurements," *Opt. Express*, vol. 27, no. 19, pp. 27 229–27 241, Sep 2019. [Online]. Available: https://opg.optica.org/oe/abstract.cfm?URI=oe-27-19-27229

[43] K. Padmaraju, J. Chan, L. Chen, M. Lipson, and K. Bergman, "Thermal stabilization of a microring modulator using feedback control," *Opt. Express*, vol. 20, no. 27, pp. 27 999–28 008, Dec 2012. [Online]. Available: https://opg.optica.org/oe/abstract.cfm?URI=oe-20-27-27999

[44] [Online]. Available: https://www.lumerical.com/solutions/inverse-design/

[45] J. Lu, "Stanfordnqp/spins-b: Photonic optimization library." [Online]. Available: https://github.com/stanfordnqp/spins-b

[46] G. T. Reed, *An Introduction to Silicon Photonics*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 161–203. [Online]. Available: https://doi.org/10.1007/978-3-540-28912-8_7

[47] T. Hu, B. Dong, X. Luo, T.-Y. Liow, J. Song, C. Lee, and G.-Q. Lo, "Silicon photonic platforms for mid-infrared applications," *Photon. Res.*, vol. 5, no. 5, pp. 417–430, Oct 2017. [Online]. Available: https://opg.optica.org/prj/abstract.cfm?URI=prj-5-5-417

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

[49] 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, Jun. 2019. [Online]. Available: https://doi.org/10.1038/s41598-019-45026-0

[50] L. Lin and E. Goldstein, “MEMS for free-space optical switching,” in *1999 IEEE LEOS Annual Meeting Conference Proceedings. LEOS’99. 12th Annual Meeting. IEEE Lasers and Electro-Optics Society 1999 Annual Meeting (Cat. No.99CH37009)*, vol. 2, 1999, pp. 483–484 vol.2.

[51] X. Guo, T. Dai, B. Chen, Y. Wang, H. Yu, and J. Yang, “An ultra-compact 4 × 4 and 8 × 8 optical switch based on dual-microring resonators,” *IEEE Photonics Technology Letters*, vol. 32, no. 21, pp. 1365–1368, 2020.

[52] T. Kita and M. Mendez-Astudillo, “Ultrafast silicon MZI optical switch with periodic electrodes and integrated heat sink,” *Journal of Lightwave Technology*, vol. 39, no. 15, pp. 5054–5060, 2021.