Deep Photonic Networks with Arbitrary and Broadband Functionality

by

Ali Najjar Amiri

A Dissertation Submitted to the
Graduate School of Sciences and Engineering
in Partial Fulfillment of the Requirements for
the Degree of

Master of Science

in

Electrical and Electronics Engineering

KOÇ ÜNİVERSİTESİ

June 12, 2023

Deep Photonic Networks with Arbitrary and Broadband Functionality

Koç University

Graduate School of Sciences and Engineering

This is to certify that I have examined this copy of a master's thesis by

Ali Najjar Amiri

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Committee Members:

_____

Asst. Prof. Emir Salih Mağden (Advisor)

_____

Assoc. Prof. Sedat Nizamoğlu

_____

Assoc. Prof. Onur Ferhanoğlu

Date: _____

Dedicated to my parents and lovely sister

ABSTRACT

Deep Photonic Networks with Arbitrary and Broadband Functionality
Ali Najjar Amiri
Master of Science in Electrical and Electronics Engineering
June 12, 2023

The increasing application space in optical communications, computing, and sensing drives the demand for high-performance integrated photonic components. Designing on-chip systems with complex and application-specific functionality goes beyond the limits of physical intuition alone. Consequently, machine learning-based design methods have gained popularity in recent times. However, the computational requirements for accurate device simulations are expensive, posing a critical challenge. As a result, these methods often have limitations in terms of scalability and the degrees of freedom they offer for optical design in application-specific and arbitrary photonic integrated circuits. To address these challenges, in this thesis, a highly scalable physics-informed framework for designing on-chip optical systems with arbitrary functionality is proposed. The framework is based on a deep photonic network comprising custom-designed Mach-Zehnder interferometers. By utilizing this framework, ultra-broadband power splitters and a spectral duplexer are successfully designed within a remarkably short span of fewer than two minutes. Furthermore, state-of-the-art experimental performance for all devices is demonstrated, achieving insertion loss of less than 0.66 dB and a 1-dB bandwidth exceeding 120 nm. The presented framework serves as an essential tool, offering a feasible approach towards systematically designing large-scale photonic systems. It enables the customization of power, phase, and dispersion profiles, thus catering to the diverse requirements of multi-band optical applications. Such applications include high-throughput communications, quantum information processing, and medical/biological sensing.

ÖZETÇE

Rastgele ve Geniş Bantlı İşlevselliğe Sahip Derin Fotonik Ağlar
Ali Najjar Amiri
Elektrik ve Elektronik Mühendisliği, Yüksek Lisans
12 Haziran 2023

Optik iletişim, hesaplama ve algılama alanlarında artan uygulama çeşitliliği, yüksek performanslı entegre fotoniğe dayalı bileşenlere olan talebi artırmaktadır. Karmaşık ve uygulamaya özel işlevselliğe sahip çip-üzeri sistemlerin tasarımı fiziksel sezgilerin ötesine geçmektedir. Sonuç olarak, makine öğrenmesi tabanlı tasarım yöntemleri son zamanlarda popülerlik kazanmıştır. Fakat, hassas aygıt simülasyonlarının hesaplama gereksinimlerinin maliyetli olması çok önemli bir zorluk oluşturmaktadır. Bunun sonucu olarak, özellikle uygulamaya özel ve rastgele işlevselliğe sahip fotonik entegre devrelerin tasarımında, bu yöntemler genellikle ölçeklenebilirlik ve serbestlik açısından sınırlı kalmaktadır. Bu tezde, bu zorluklara çözüm olarak, rastgele işlevselliğe sahip optik sistemlerin tasarımı için yüksek ölçeklenebilirliğe sahip, fizik destekli bir çerçeve önerilmektedir. Bu çerçeve, özel tasarlanmış Mach-Zehnder interferometrelerden oluşan bir derin fotonik ağa dayanmaktadır. Bu çerçevenin kullanılmasıyla, ultra geniş bantlı güç bölücüler ve spektral filtreler, iki dakikadan bile daha kısa bir sürede başarıyla tasarlanmıştır. Ayrıca, ek kayıpların 0.66 dB'den az ve 1 dB bant genişliğinin 120 nm'den fazla olduğu bu zamana kadar tasarlanmış en gelişmiş aygıtlar deneysel olarak gösterilmiştir. Sunulan çerçeve, büyük ölçekli fotonik sistemlerin sistematik olarak tasarlanmasında uygulanabilir bir yaklaşım sunan önemli bir araçtır. Güç, faz ve dağılım profillerinin özelleştirilmesine olanak tanır ve bu da çoklu bant optik uygulamaların çeşitli gereksinimlerini karşılamayı sağlar. Bu tür uygulamalar arasında yüksek hızlı iletişim, kuantum bilgi işlem ve tıbbi/biyolojik algılama bulunmaktadır.

ACKNOWLEDGMENTS

I would like to take this opportunity to express my heartfelt appreciation and gratitude to the individuals who have supported and contributed to the successful completion of my master's thesis.

First and foremost, I extend my deepest gratitude to my advisor, Asst. Prof. Emir Salih Mağden, for his continuous encouragement, invaluable contribution, unwavering support, and guidance throughout this research endeavor. Working under his supervision has not only enhanced my knowledge in the field but has also helped me develop crucial skills in communication, teamwork, organization, and discipline. I am truly grateful for the opportunity to work with him and be a part of the Photonic Architecture Laboratories.

I would also like to extend my gratitude to my esteemed committee members, Assoc. Prof. Sedat Nizamoğlu and Assoc. Prof. Onur Ferhanoğlu, for their valuable insights, guidance, and constructive feedback throughout the thesis process.

I would like to acknowledge the Scientific and Technological Research Council of Turkey (TÜBİTAK) for their financial support during my studies.

I would like to extend my sincere appreciation to all my colleagues in the PAL group for the enriching experiences we shared and their unwavering support throughout my master's thesis project. In particular, I am deeply grateful to my teammates, Kazım and Aycan, for their outstanding collaboration, guidance, support, and the enjoyable moments we had together. Their contributions and camaraderie have been invaluable, and I am truly thankful for their involvement in this project.

Finally, I would like to express my deepest gratitude to my family for their unending love, encouragement, and belief in me. Their unwavering support and faith have been the driving force behind my accomplishments, and I am forever grateful for their presence in my life.

TABLE OF CONTENTS

LIST OF FIGURES

ABBREVIATIONS

MZI	Mach-Zehnder Interferometer
DC	Directional Coupler
CMOS	Complementary Metal-Oxide-Semiconductor
FDTD	Finite Difference Time Domain
IC	Integrated Circuit
PIC	Photonic Integrated Circuit
3D	Three Dimensional
MSE	Mean Squared Error
SOI	silicon-on-insulator
MPW	multi-project-wafer

Chapter 1

INTRODUCTION

This chapter covers developments and concepts of photonic integrated circuit (PIC), PICs design methods, Programmable PICs based on MZI networks, this thesis summary, and the organization of the thesis.

1.1 Photonic Integrated Circuit and Design Methods

Photonic integrated circuits (PICs) [Chrostowski and Hochberg, 2015, Bogaerts and Chrostowski, 2018] have significantly evolved over the last decade and are now essential technological components with critical importance in optical communications [Zhuang et al., 2015], sensing [Hu et al., 2017, Poulton et al., 2019], and computing [Zhang and Yao, 2020, Pérez et al., 2017, Carolan et al., 2015]. With the growing diversity and complexity of photonic applications, designing custom PICs with state-of-the-art performance metrics has become one of the most critical drivers of advancement in photonic systems.

In traditional design of photonic components, the designer relies on prior knowledge of relevant system architectures, fundamental electromagnetic principles, and physical intuition to determine key parameters such as waveguide widths, lengths, or gaps [Molesky et al., 2018, Piggott et al., 2015, Lu and Vučković, 2013]. These parameters are then typically tuned either manually or by optimization algorithms, according to the results of repeated electromagnetic simulations, to achieve the required optical functionality. This approach yields a limited library of known devices and severely restricts the potential capabilities of the resulting photonic systems. More general approaches have emerged under the broad category of inverse/machine-optimized design [Piggott et al., 2015, Lu and Vučković, 2013, Qu et al., 2020, Taher-

sima et al., 2019, Molesky et al., 2018, Zhang et al., 2022], allowing for greater design flexibility than manual tuning of waveguide parameters. Generally, these approaches perform more comprehensive searches over the complete domain of fabrication-compatible devices through heuristic [Zhang et al., 2013, Sanchis et al., 2009] or gradient-based [Tahersima et al., 2019, Jiang et al., 2021, So et al., 2020, Hegde, 2020, Jensen and Sigmund, 2011] computational optimization tools. Using these methods, devices such as couplers [Piggott et al., 2020], polarization splitters [Shen et al., 2015, Lu and Vučković, 2013], and spectral filters [Piggott et al., 2015, Zhang et al., 2022] have been proposed and demonstrated. However, in these "free-form" design approaches, the degrees of design freedom are effectively controlled by the specified device footprint, which has key implications on the final device performance and the associated computational cost. While larger device footprints inherently provide the necessary design flexibility for complex and arbitrary optical functionality, they also rapidly scale the computational complexity of the iterative optimization process due to the physically-accurate electromagnetic simulations required [Piggott et al., 2015, Lu and Vučković, 2013, Jia et al., 2018, Zhang et al., 2022]. These requirements preclude the design of arbitrarily complex, ultra-broadband, or wavelength-specific photonic devices for the increasing number and variety of on-chip use cases and their custom application requirements.

1.2 Programmable PICs based on MZI networks

The ideal approach to photonic design must allow for arbitrarily-specified photonic functionality while maintaining low computational cost. In recent years, programmable PICs made from Mach-Zehnder interferometers (MZIs) have been proposed as a potential solution to this problem [Bogaerts et al., 2020, Xu et al., 2022, Pérez-López et al., 2020, Zhuang et al., 2015, Capmany and Pérez, 2020]. These systems enable tuning of optical responses through thermal or electro-optic phase shifters in order to achieve wavelength-specific linear mappings between the input/output pairs [Capmany and Pérez, 2020, Shen et al., 2017, Harris et al., 2018, Miller, 2013]. Several example applications including optical signal routing

[Marpaung et al., 2019, Zhuang et al., 2015, Pérez-López et al., 2020], image/signal classification [Shen et al., 2017, Ashtiani et al., 2022, Shastri et al., 2021, Feldmann et al., 2021, Lin et al., 2018], and quantum computing [Carolan et al., 2015, Harris et al., 2016] have already been demonstrated with numerous advantages such as post-fabrication reconfigurability, superior power efficiency, and lower latency over their electronic counterparts. However, the potential utility of photonic interferometer networks extends well beyond these demonstrated capabilities, with critical implications towards the design of photonic devices and systems with arbitrarily complex transfer functions.

1.3 Summary of This Thesis

In this thesis, we introduce and experimentally demonstrate a highly-scalable framework for the design of photonic systems with arbitrarily-specified functionality, based on a deep photonic network architecture of custom-designed MZIs. Our architecture consists of a mesh of individually designed interferometers and is modeled by an equivalent computational network equipped with ultra-fast and physically-accurate simulation capabilities. In this network, each MZI is constructed from unique waveguide tapers, allowing for specific wavelength-dependent phase profiles to be achieved according to the target photonic functionality specified. The exact geometry of the individual interferometers is optimized by leveraging physics-informed machine learning capabilities in our design framework through a combination of rapid lookup of waveguide parameters and successive evaluation of photonic transfer matrices. Using this framework, we design ultra-broadband 50/50 and 75/25 power splitters and a spectral combiner/splitter, each in less than two minutes, with inherent fabrication compatibility on the 220-nm-thick silicon-on-insulator platform, and experimentally demonstrate state-of-the-art performance for all three devices. Our presented framework provides a path towards the systematic design of large-scale photonic systems with arbitrarily-specified, wavelength-dependent, or ultra-broadband responses.

1.4 Organization

The thesis is structured as follows:

Chapter 2 delves into the deep photonic network architecture and its building blocks. It provides a comprehensive understanding of the underlying concepts and components involved in the network design.

Chapter 3 focuses on the simulation and optimization procedure employed for the network. It explores the methodologies and techniques used to optimize the network's parameters and ensure its effectiveness.

Chapter 4 centers around the design and characterization of three proof-of-concept devices. It presents detailed insights into the design process, experimental results, and performance evaluation of these devices.

Chapter 5 delves into the analysis of the deep photonic network's capabilities and its robustness in fabrication. It examines the network's performance in different scenarios and explores its resilience to fabrication variations.

Chapter 6 focuses on computational performance and metrics. It investigates the efficiency and effectiveness of the computational techniques employed in the network design and optimization process.

Chapter 7 serves as a conclusion and discussion section, where key findings, implications, and future directions are elaborated upon. It provides a reflective overview of the research conducted and offers insights into potential avenues for further exploration.

The Appendix provides a detailed description of the methods used in the preceding chapters. It serves as a supplementary section that elucidates the specific techniques, algorithms, software tools, and experimental procedures employed throughout the research.

Chapter 2

DEEP PHOTONIC NETWORK ARCHITECTURE

2.1 Network Architecture

The schematic diagram in Figure 2.1(a) illustrates the architecture of our deep photonic network, which comprises an input layer, multiple layers of Mach-Zehnder interferometers (MZIs), and an output layer. This MZI-based architecture is capable of implementing any linear $N \times N$ input-output mapping, enabling the achievement of arbitrary optical functionality [Reck et al., 1994, Miller, 2013, Miller, 2015]. The input optical signal is introduced to the network either externally through a series of couplers, as depicted, or via waveguides connected to preceding on-chip devices. The input optical signal undergoes unidirectional processing through successive layers of customized MZI interferometers. Each MZI possesses its unique 2×2 mapping function denoted as T_{ij}. This modular network is characterized using the transfer matrix representation of its constituent building blocks, configured in a modular manner. Specifically, each MZI is composed of two pairs of waveguide tapers with custom geometries and two directional couplers, as illustrated in Figure 1(b). The transfer matrix for each Mach-Zehnder interferometer (MZI) in the network can be expressed as the product of the transfer matrices of its constituent blocks, such as:

$$\begin{aligned} T(\lambda)=& e^{-j\varphi(\lambda)} \begin{bmatrix} t(\lambda) & -jq(\lambda) \\ -jq(\lambda) & t(\lambda) \end{bmatrix} \begin{bmatrix} e^{-j\theta_{21}(\lambda)} & 0 \\ 0 & e^{-j\theta_{22}(\lambda)} \end{bmatrix} \\ & \times e^{-j\varphi(\lambda)} \begin{bmatrix} t(\lambda) & -jq(\lambda) \\ -jq(\lambda) & t(\lambda) \end{bmatrix} \begin{bmatrix} e^{-j\theta_{11}(\lambda)} & 0 \\ 0 & e^{-j\theta_{12}(\lambda)} \end{bmatrix} \end{aligned} \tag{2.1}$$

where $t(\lambda)$, $q(\lambda)$, and $\varphi(\lambda)$ represent the wavelength-dependent through-port and cross-port amplitude coefficients, as well as the phase response of the directional couplers, respectively. On the other hand, $\theta_{11}(\lambda)$ through $\theta_{22}(\lambda)$ represent the phase accumulations that occur in the corresponding waveguide tapers. These wavelength-

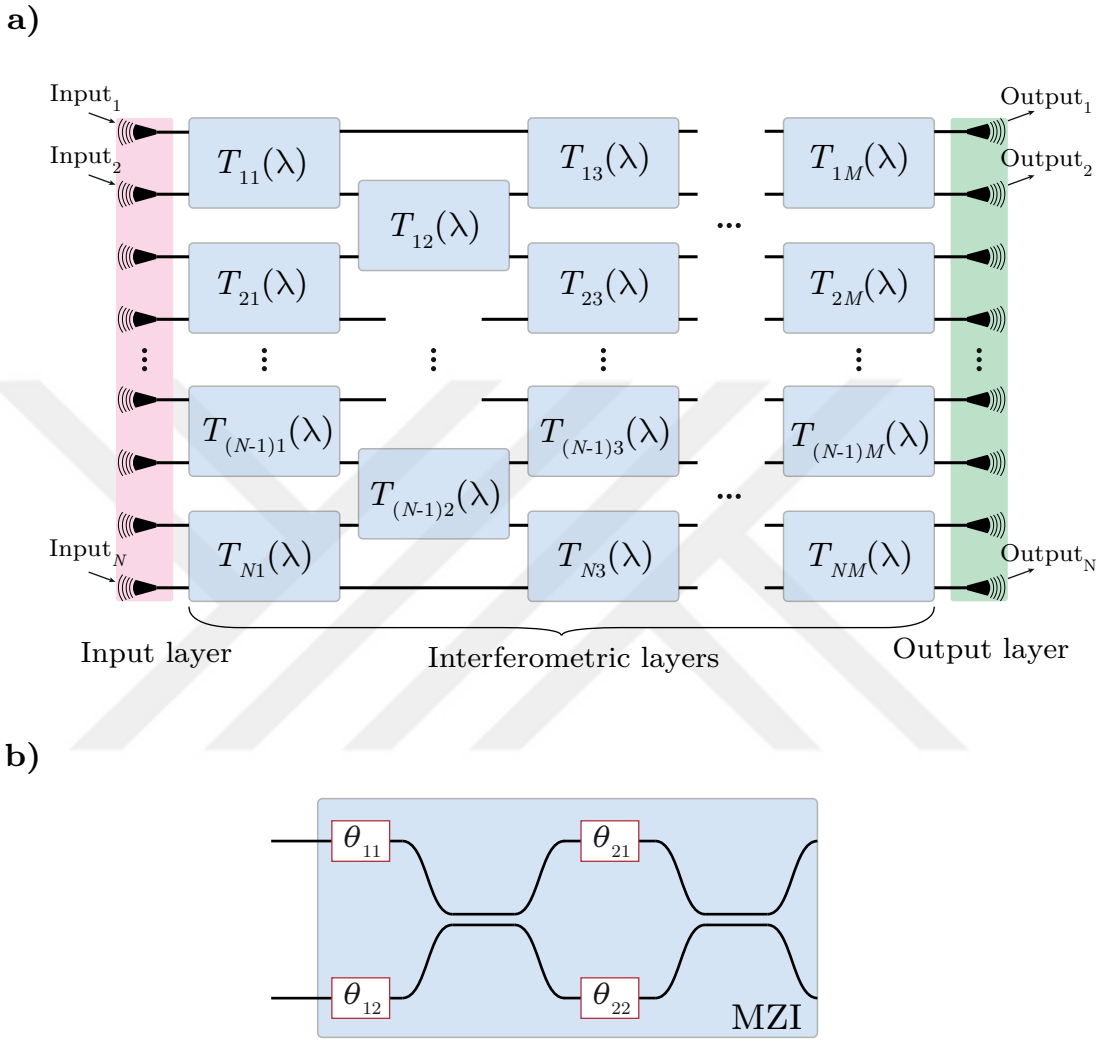

Figure 2.1: Deep photonic network architecture. (a) The network architecture is composed of the input stage, horizontally-cascaded and vertically-repeated custom interferometric layers, and the output couplers. Each interferometric layer consists of a combination of Mach-Zehnder interferometers and individually-optimized waveguide structures. (b) Block diagram of a Mach-Zehnder interferometer with two pairs of waveguide tapers of custom geometries and two directional couplers. θ_{11} through θ_{22} indicate the phases accumulated through each custom waveguide taper.

dependent parameters are crucial in enabling the attainment of arbitrary optical functionality in our networks.

2.2 Building Blocks and 3D Schematic

2.2.1 Directional Coupler

Throughout the network, identical directional couplers are employed, designed to function as approximately 50% couplers at a wavelength of 1550 nm. To ensure the physical accuracy of the transfer matrix model, the complete optical transmission, including losses through S-bends and directional couplers, is incorporated into the transfer matrix $T(\lambda)$. This is achieved by utilizing the optical response of the directional coupler, which is obtained from 3D-FDTD simulations, as depicted in Figure 2.2. In contrast to the custom waveguide tapers present in the interferometers, the directional couplers used in our networks are constructed to be identical and do not contain any trainable parameters. The modular nature of our design framework allows for the substitution of this specific directional coupler with alternative components. For instance, it is possible to replace the existing directional coupler with one having a different coupling length, bends with varying geometries, or an entirely different coupler with a transfer function tailored to a specific task. This flexibility enables customization and adaptation of the network to meet specific design requirements.

2.2.2 Waveguide Tapers

In contrast to the identical directional couplers, all waveguide tapers in the network are individually designed and tailored. The design process involves optimizing a set of width and length parameters, using an iterative optimization algorithm. The phase accumulated through each custom waveguide taper is calculated as a differentiable function of the custom widths (w_i), taper length (L_θ), and input wavelength (λ). This calculation incorporates the waveguide effective index $n_{\text{eff}}(w, \lambda)$ and is described as follows:

$$\theta(\lambda) = \int_0^{L_\theta} \frac{2\pi}{\lambda}\, n_{\text{eff}}(w(z), \lambda)\, dz \tag{2.2}$$

This unique implementation allows the network to achieve wavelength-dependent phase profiles that differ from those of a straight waveguide. Figure 2.3 illustrates this capability, demonstrating the network's ability to provide significantly higher degrees of freedom while maintaining the same device footprint. Consequently, this design approach enables greater flexibility in achieving desired phase characteristics for various wavelengths, enhancing the functionality and versatility of the network.

2.2.3 3D Schematic of Deep Photonic Network

A schematic of the directional coupler illustrating its structure and the simulated through-port transmission, an arbitrary waveguide taper along with its transmission characteristics, and the design platform for constructing the entire photonic integrated circuit are depicted in Figure 2.4. This framework is capable of accommodating an arbitrary number of interferometric layers, enabling the creation of complex and customizable integrated circuits. The schematic showcases the overall structure and connectivity of the system, emphasizing its scalability and versatility.

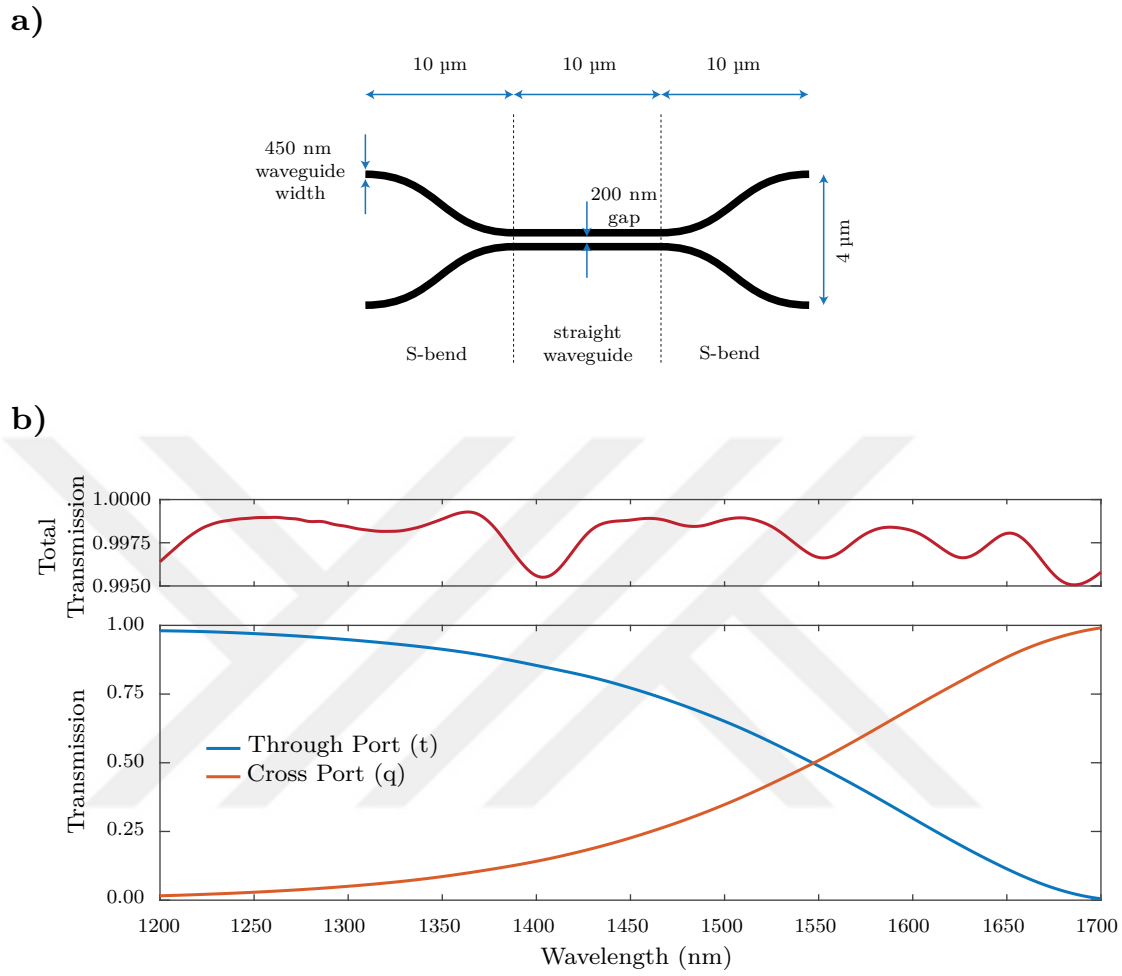

Figure 2.2: Diagram and optical response of the directional coupler. (a) Diagram of the directional coupler indicating important geometrical parameters including 10 µm-long straight coupling section with 200 nm gap, 10 µm-long S-bends, 4 µm center-to-center waveguide separation between the top and the bottom arm at the start and end of the directional coupler, and 450 nm waveguide width through the entire directional coupler. (b) Optical responses at the cross- and through-port of the directional coupler used in our deep photonic networks. Results were obtained from 3D-FDTD simulations performed with a maximum spatial discretization of 17 nm in all three dimensions. Amplitude and phase information between 1.2 µm and 1.7 µm were implemented as automatic differentiation-compatible interpolations.

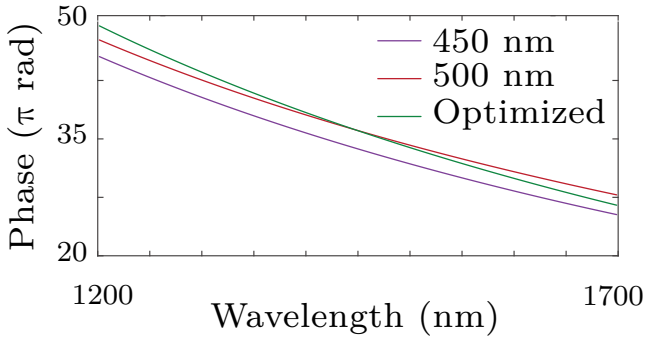

Figure 2.3: Phase profile of a sample optimized waveguide taper. The optimized custom designed tapers enable unique spectral phase profiles different from those in straight waveguides with the same footprint.

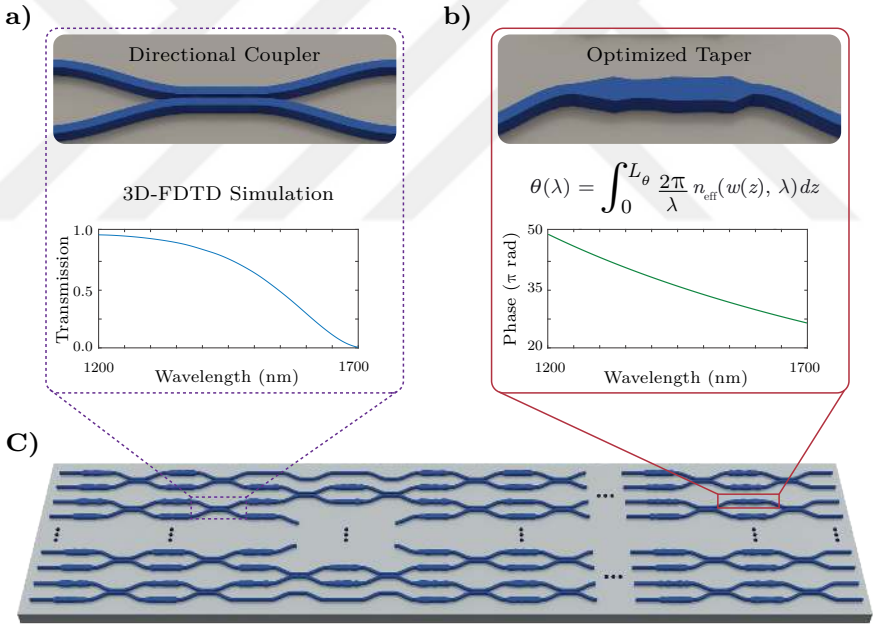

Figure 2.4: Deep photonic network 3D schematic. (a) Schematic of the directional coupler with two S-bends and a 10 µm-long coupling section, and its 3D-FDTD simulated transmission response. (b) Schematic of an example custom waveguide taper constructed from a set of optimizable width parameters, from which the accumulated phase is calculated as a function of wavelength using the effective index. (e) Overall structure of an example deep photonic network with cascaded interferometric layers of directional couplers and individually optimized waveguide tapers.

Chapter 3

SIMULATION AND OPTIMIZATION OF THE OPTICAL RESPONSE THROUGH THE NETWORK

3.1 Optimization Procedure

The propagation of the complex optical amplitude through the network is achieved by utilizing a computational graph that mimics the physical architecture of the network. Within this computational graph, the optical transformation performed by the mesh of interferometers between N input channels and N output channels is represented. The overall architecture calculates the wavelength-dependent linear scattering matrix, denoted as $S(\lambda)$, of the deep photonic network network according to

$$S(\lambda) = \prod_{(i,j)\ \epsilon\ \Gamma} T_{i,j}(\lambda) \tag{3.1}$$

where Γ indicates the specific ordering of 2×2 transfer matrices in the network. This computation involves integrating the waveguide effective index using the custom widths and lengths of each waveguide taper. Furthermore, the through-port and cross-port coefficients, as well as the phase response of the directional coupler, are extracted from 3D-FDTD (Finite-Difference Time-Domain) simulation results. To optimize the custom photonic networks for user-defined optical functionality, these operations are implemented using a differentiable programming construct [Bradbury et al., 2018]. This allows for efficient parameter lookups and automatic calculation of relevant derivatives. The calculation of $\theta(\lambda)$, representing the phase accumulation, involves numerically integrating the effective index along the length of the custom tapers using data obtained from the Silicon Photonics Toolkit [Vit et al., 2022], an open-source software package that provides access to important propagation-related parameters in silicon waveguides as functions of wavelength and waveguide width.

Similarly, the directional coupler coefficients are obtained through differentiable interpolation of its 3D-FDTD simulation results. By performing these computations, the complete transfer function of the network is obtained with a high level of physical accuracy, incorporating the wavelength-dependent mappings for each input-output pair.

This ability to rapidly calculate the optical response of a network as a differentiable function of its design parameters is crucial from an optimization perspective. Using the computational capabilities described earlier, we establish an optimization procedure to iteratively modify the waveguide tapers and create application-specific photonic networks with user-defined transfer functions. This procedure is exemplified in Figure 3.1 for a 1-input 4-output network. To begin, we initialize a network with the desired number of interferometric layers and input-output ports. We define the target optical transfer function for the input-output pairs, denoted as $(T_{\text{target}}(\lambda))$, and semi-random width and length parameters are assigned to the custom waveguide tapers within the network. The optical response of the network is evaluated at various wavelengths using the aforementioned procedure. This response is then compared with the target transfer function. The difference between the calculated and target transfer functions is quantified as the mean squared error, $J(x) = \frac{1}{Q}\sum_{\lambda}|T_{\text{calculated}}(\lambda, x) - T_{\text{target}}(\lambda)|^2$, where Q is the number of wavelengths and x represents the design parameters, including the widths and lengths of the custom tapers. To optimize the network, the gradient of $J(x)$ with respect to the design parameters, $\nabla_x J$, is computed using a back-propagation procedure. The objective is to minimize this error by iteratively adjusting the widths and lengths of the waveguide tapers. Figure 3.2 illustrates this iterative modification process for a sample taper using a gradient-based optimization algorithm [Kingma and Ba, 2014].

3.2 Initialization and Optimization of Custom Waveguide Tapers

To ensure fabrication compatibility, several regularization schemes are implemented in addition to the error minimization. These schemes restrict extreme changes in waveguide widths during the optimization procedure.

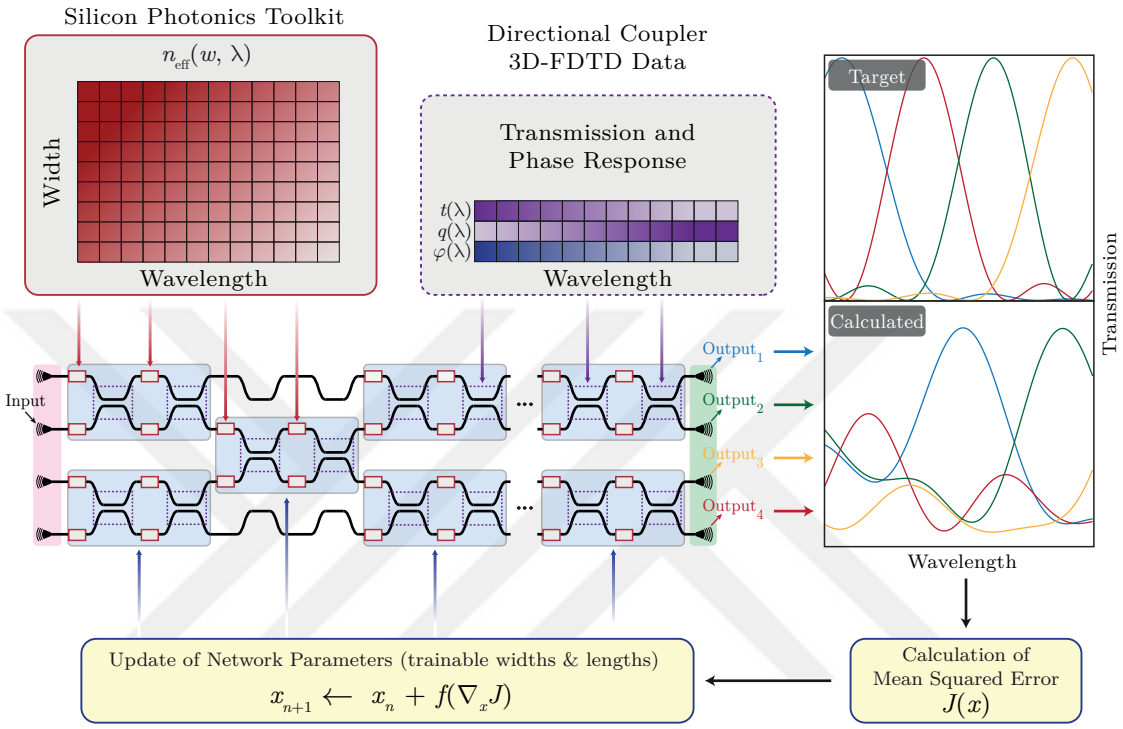

Figure 3.1: Optimization of an example 1-input 4-output photonic network. The 1×4 network structure is created with the desired number of layers, randomly-intialized custom waveguide tapers (red rectangles), and a target transmission response for input-output pairs. The mean squared error is computed from the difference between the calculated and target transfer functions, by summing over the specified wavelength range. The network parameters are trained iteratively through a backpropagation algorithm using the gradient of this error with respect to the design parameters denoted by x in the custom waveguide tapers. Other network components including the directional couplers and input/output layers are not trainable.

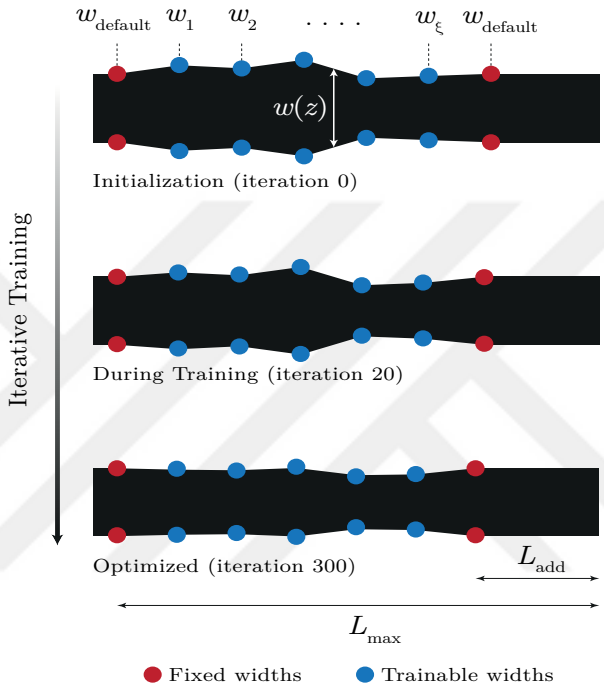

Figure 3.2: Evolution of a custom waveguide taper throughout optimization. Evolution of a custom waveguide taper throughout optimization of the deep photonic network, where its geometry is shown at random initialization, at iteration 20, and at the end of optimization. Fixed widths (w_{default}) and trainable widths ($w_1, w_2, \ldots, w_\xi$) are marked with red and blue circles along the taper, respectively. At each iteration, an additional straight waveguide of length L_{add} is inserted at the end of the custom taper in order to achieve matching L_{max} lengths for all tapers.

By incorporating these regularization techniques, the resulting photonic networks maintain compatibility with fabrication processes, enhancing the practicality of the design approach. In order to design practical photonic devices, it is important to consider the physical constraints and limitations of the fabrication process. While the optimization procedure aims to optimize the geometry of custom waveguide tapers and the resulting $\theta(\lambda)$ according to the specified objective function, it is crucial to ensure that the final widths and lengths of the waveguide tapers are suitable for an actual photonic device and the specific network architecture. To address this, our physics-informed network models are constructed with specific initialization considerations, physical boundaries, and regularization schemes imposed on the trainable parameters, as detailed below.

3.2.1 Initialization and Constraints for Trainable Widths and Lengths

In our demonstrations on the 220 nm SOI platform, we adopt a default waveguide width (w_{default}) of 450 nm for the entire network, encompassing input/output waveguides, directional couplers, and bends. This choice is made to maintain fabrication compatibility of the designed devices and minimize the excitation of higher-order modes during light propagation through the tapers. Consequently, the resulting widths in the custom tapers are designed to be relatively close to the default width, ensuring compatibility with the fabrication process and promoting optimal mode behavior.

To ensure appropriate initialization of the custom waveguide widths, we follow a procedure that maintains their proximity to w_{default}, which we implement by $w_{\text{initial}} = w_{\text{default}} + p_w \cdot \Delta w + w_{\text{offset}}$ at the beginning of optimization. Here, p_w is a uniform random variable between -1 and 1, Δw is a user-chosen maximum deviation amplitude, and w_{offset} is a constant initial offset. The option to specify an initial offset can be valuable when optimizing the parameters, particularly in cases where the updates tend to exhibit monotonic tendencies during the optimization process. This consideration is particularly relevant for certain types of artificial neural networks [Sutskever et al., 2013, Alom et al., 2019, Hanin and Rolnick, 2018], where

the choice of initial offset can help mitigate the potential convergence towards a specific pattern or direction. By allowing the designer to adjust the initial offset, they have greater control over the optimization trajectory and can effectively address the challenges posed by monotonic tendencies. During the optimization process and the computation of the transfer functions, it is important to maintain appropriate mode confinement within the waveguide core and prevent the development of strongly guided higher-order modes. To achieve this, we apply a constraint on the trainable widths, limiting them to the range of $w_{\min} \leq w \leq w_{\max}$. By setting these bounds, we ensure that the waveguide widths remain within the desired range for effective mode confinement. Specifically, for the networks we demonstrated, the parameters were typically defined as follows: 5 nm $\leq \Delta w \leq$ 20 nm, 30 nm $\leq w_{\text{offset}} \leq$ 50 nm, $w_{\min} = 400$ nm, and $w_{\max} = 520$ nm. These ranges were chosen to strike a balance between optimizing the performance of the photonic devices and maintaining the desired mode properties.

A similar approach is followed for initializing the custom taper lengths, with a few minor differences. Initially, the lengths are initialized as $L_{\text{initial}} = L_{\max} - p_L \cdot \Delta L$, where p_L is a random variable ranging between 0 and 1, and ΔL is a user-defined maximum deviation amplitude. Similarly, throughout the optimization process, the trainable lengths are constrained to fall within the range of $L_{\min} \leq L \leq L_{\max}$. Moreover, if the resulting taper length is shorter than $L_{\max}$, a straight waveguide is added at the end of the taper in each iteration of the optimization procedure. The length of this additional waveguide is set to $L_{\text{add}} = L_{\max} - L$, ensuring the alignment of successive layers of MZIs in the direction of propagation. For the networks that were demonstrated, the parameter values were selected as follows: $L_{\max} = 10$ µm, $L_{\min} = 6$ µm, and the directional couplers had a length of 30 µm. These choices resulted in an overall length of 80 µm for each MZI in our networks.

In addition to the initialization and imposed boundaries, The choice of the number of trainable widths per waveguide taper is a crucial design consideration. In our approach, the ends of the custom tapers have fixed widths of w_{default}, and the trainable widths are placed with equal spacing between them. The number of trainable widths, along with the $L_{\min}$ and $L_{\max}$ parameters, directly impacts the resulting

taper angle. If the taper has insufficient length, it can lead to potential propagation losses and the excitation of higher-order modes [Milton and Burns, 1977, Fu et al., 2014, Zou et al., 2014]. In our designs, we have opted to include 5 uniformly-spaced trainable widths within each waveguide taper, ensuring a minimum spacing of 1 µm between them. Based on our analysis using 3D-FDTD results, we have found that a minimum taper length of $L_{\min} = 1\ \mu\text{m} \times (\xi + 1)$ is sufficient for achieving low-loss operation of tapers within the range of $w_{\min} = 400$ nm and $w_{\max} = 520$ nm, where ξ represents the number of trainable widths in a single taper. However, the designer has the flexibility to choose a different number of trainable widths and adjust the $L_{\min}$ parameter accordingly.

3.2.2 Regularizers

In addition to the previously discussed initialization and constraints, we employ two major regularization schemes to control the trainable widths during the iterative optimization process. The first regularization focuses on minimizing the differences between consecutive widths within a single waveguide taper. This helps avoid sudden changes in width that could lead to propagation losses. The second regularization aims to limit the difference between each trainable width and a reference width w_{ref}, which is typically chosen to be the same as the default width (w_{default}). This regularization ensures that the optimized variables remain close to the reference width. Both regularization schemes are calculated as L2-norms of the corresponding error vectors

$$P_1 = \sum_{\text{all tapers}} \sum_{i=1}^{\xi-1} (w_i - w_{i+1})^2 \tag{3.2}$$

and

$$P_2 = \sum_{\text{all tapers}} \sum_{i=1}^{\xi} (w_i - w_{\text{ref}})^2 \tag{3.3}$$

where w_i represents the individual trainable widths in each custom taper within a specific photonic network. To incorporate the regularization schemes, an overall regularization term P is computed as $P = \alpha_1 P_1 + \alpha_2 P_2$, where α_1 and α_2 are

coefficients used to independently control the contribution strength of the two regularizers. This accumulated contribution of the regularizers is then added as an artificially-introduced loss at the end of the network. The overall objective function is subsequently calculated as

$$J(x) = \frac{1}{Q} \sum_{\lambda} \mid T_{\text{calculated}}(\lambda, x) e^{-P} - T_{\text{target}}(\lambda) \mid^2 \tag{3.4}$$

where Q is the number of wavelengths and x are design parameters including widths and lengths of the custom tapers. As default, the coefficients α_1 and α_2 are set to 3×10^{-4} and 1×10^{-4}, respectively. Depending on the photonic capabilities required, the designers may freely experiment with various contribution strengths and examine resulting taper profiles. Designers have the flexibility to explore different contribution strengths and analyze the resulting taper profiles based on the desired photonic capabilities. It is worth noting that stronger regularizations lead to waveguides with more gradually changing widths but also limit the photonic network's ability to replicate arbitrary transfer functions due to reduced degrees of freedom. With the computational efficiency of our network implementation, designers can easily experiment with different regularization contributions and compare the resulting taper structures.

Chapter 4

DESIGN AND CHARACTERIZATION OF ARBITRARY PHOTONIC DEVICES

4.1 Arbitrary Optical Functionality with Custom Deep Photonic Networks

Our proposed deep photonic network offers a significant advantage in designing photonic devices with versatile spectral specifications. This capability allows for a universal design procedure that accommodates devices with ultra-broadband responses as well as devices with specific spectral features. To demonstrate this functionality, we present three distinct devices as proof of concept. These include two broadband power splitters with splitting ratios of 50/50 and 75/25, operating within the wavelength range of 1400-1600 nm. Additionally, we showcase a 1×2 spectral duplexer, operating between 1450 nm and 1630 nm. These examples highlight the flexibility and effectiveness of our approach in designing photonic devices with various spectral characteristics.

To ensure optimal performance, the selection of hyperparameters in our deep photonic network framework depends on the desired functionality complexity. Key hyperparameters include the number of interferometric layers and the number of custom widths in each waveguide taper. Chapter 5 provides detailed information on choosing these hyperparameters. In the case of the power splitters, we designed networks with three layers each, while the duplexer utilized a network with six layers. For each custom waveguide taper in our devices, we employed five trainable widths and a trainable length, resulting in a total of 24 parameters for each Mach-Zehnder Interferometer (MZI) in our photonic networks. These hyperparameter selections ensure efficient and effective design of the devices within our framework.

The evolution of the resulting mean squared errors throughout the optimization

processes are plotted in the following sections, where convergence is achieved in several hundred iterations and, at most, a few minutes on a single Tesla V100 GPU. Details regarding the optimization time of the photonic networks and their scalability can be found in Chapter 6.

4.1.1 50/50 Power Splitter

Figure 4.1(a) shows the evolution of the mean squared error during the optimization process for the design. It demonstrates how the output state gradually approaches the desired functionality as the optical responses converge towards the target, as seen by the transmission at one of the output ports (50%) as a function of wavelength in Figure 4.1(b). The transmission spectra at both output ports of the 50/50 power splitter are plotted in Figure 4.1(c), showcasing the randomly-initialized state at the beginning of optimization, an intermediate state after partial training, and the final optimized state. The optimized devices exhibit a near-perfect match with the desired target functionality. To further validate these responses, the propagation of the optical input in the final optimized device is visualized in Figure 4.2, using the electric field intensity obtained from 3D-FDTD simulations. The achieved broadband operation of the 50/50 power splitter, spanning 200 nm, is in agreement with our expectation.

4.1.2 75/25 Power Splitter

Similar to the 50/50 power splitter, for the 75/25 power splitter, Figure 4.3(a) presents the evolution of the mean squared error during the optimization process. It shows how the output state gradually approaches the desired functionality as the optical responses converge towards the target. The transmission at one of the output ports (25%) as a function of wavelength is plotted in Figure 4.3(b), demonstrating the convergence of the optical response to the desired values. Figure Figure 4.3(c) displays the transmission spectra at both output ports of the 75/25 power splitter, showcasing the randomly-initialized state at the beginning of optimization, an intermediate state after partial training, and the final optimized state.

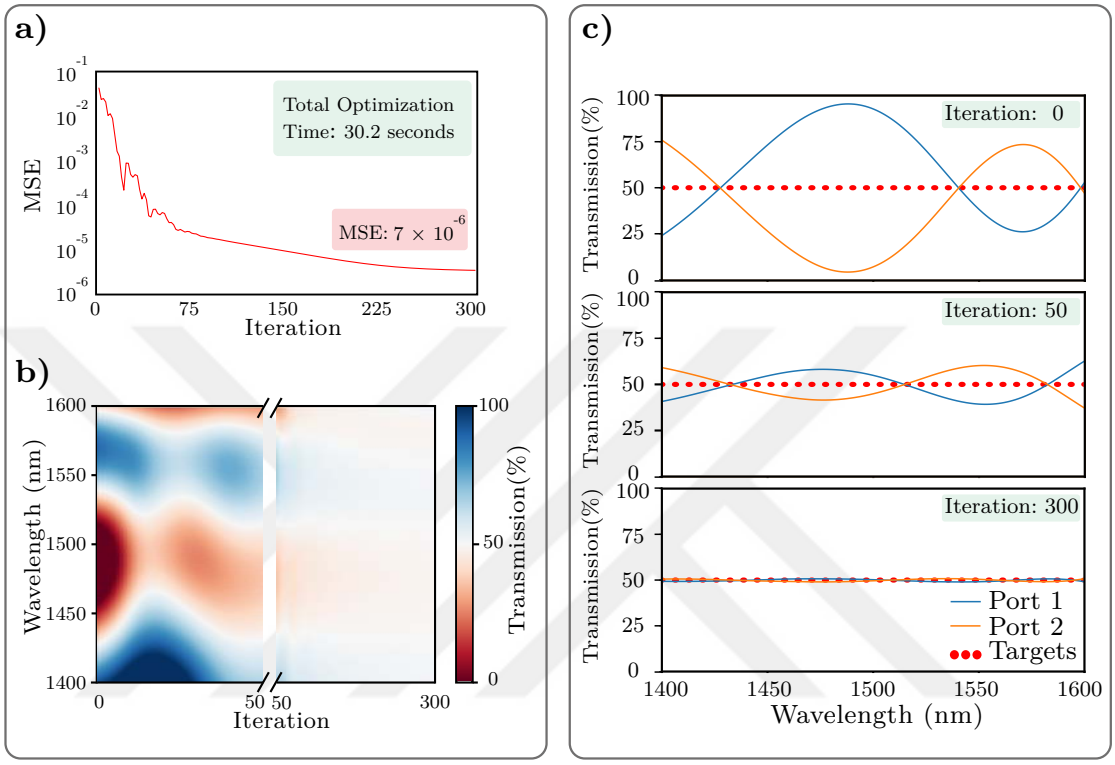

Figure 4.1: Optimization result of the 50/50 power splitter. (a) The mean squared error (MSE) versus iteration throughout optimization of a 50/50 power splitter with 3 layers of MZIs (72 trainable parameters, 240 µm device length). The device converges in three hundred iterations, within 30.2 seconds. (b) Transmission at the designated output port of the 50/50 power splitter as a function of wavelength. The evolution of this transmission through the iterative training process enables the device to achieve near-perfect transfer function by the end of optimization. (c) Transmission spectra for each output during optimizations show gradual convergence to the target transfer functions indicated by the circles. The power splitter is optimized with 32 evenly-spaced wavelengths between 1400-1600 nm.

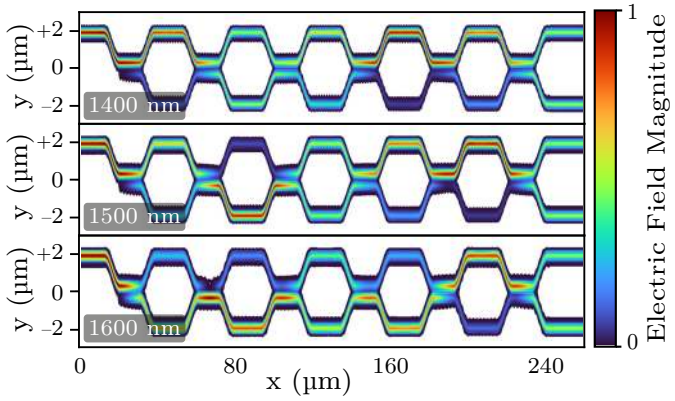

Figure 4.2: 3D-FDTD final simulation result of the 50/50 power splitter. Magnitude of the electric field at three different wavelengths obtained from 3D-FDTD simulations confirming broadband and flat-top operation for the 50/50 power splitter.

The optimized devices exhibit a near-perfect match with the desired target functionality. To further validate these responses, Figure 4.4 visualizes the propagation of the optical input in the final optimized device using the electric field intensity obtained from 3D-FDTD simulations. Indeed, the achievement of broadband operation spanning 200 nm by the 75/25 power splitter is in line with our expectation.

4.1.3 Spectral Duplexer

The construction of the power splitters involves 3 interferometric layers, while the duplexer consists of 6 interferometric layers. Throughout the optimization process, the mean squared error gradually decreases, indicating the convergence of the output state towards the desired functionality, plotted in Figure 4.5(a). The transmission at the spectrally duplexed output ports, plotted in Figure 4.5(b), demonstrates how the optical response aligns with the desired values across different wavelengths. Figure 4.5(c) provides a comprehensive view of the transmission spectra at both output ports of the spectral duplexer. It illustrates the evolution from the randomly-initialized state at the beginning of optimization to an intermediate state after partial training, and finally, to the optimized state.

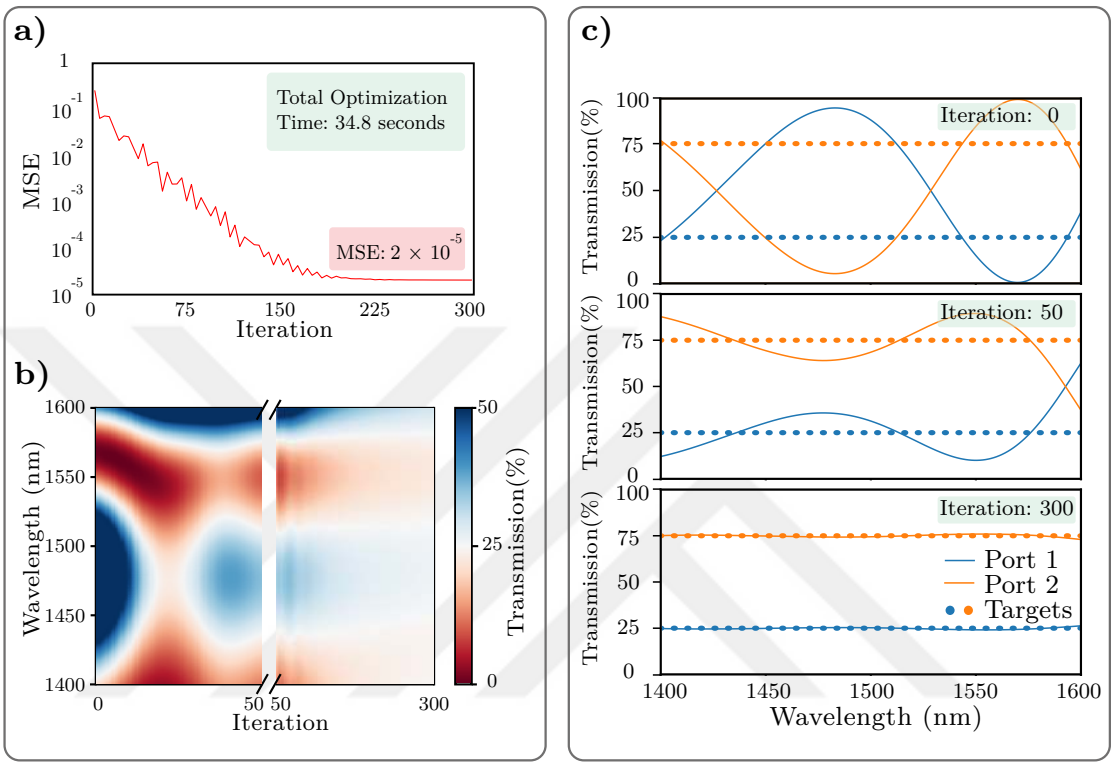

Figure 4.3: Optimization result of the 75/25 power splitter. (a) The mean squared error (MSE) versus iteration throughout optimization of a 75/25 power splitter with 3 layers of MZIs (72 trainable parameters, 240 µm device length). The device converges in three hundred iterations, within 34.8 seconds. (b) Transmission at the designated output port of the 75/25 power splitter as a function of wavelength. The evolution of this transmission through the iterative training process enables the device to achieve near-perfect transfer function by the end of optimization. (c) Transmission spectra for each output during optimizations show gradual convergence to the target transfer functions indicated by the circles. The power splitter is optimized with 32 evenly-spaced wavelengths between 1400-1600 nm.

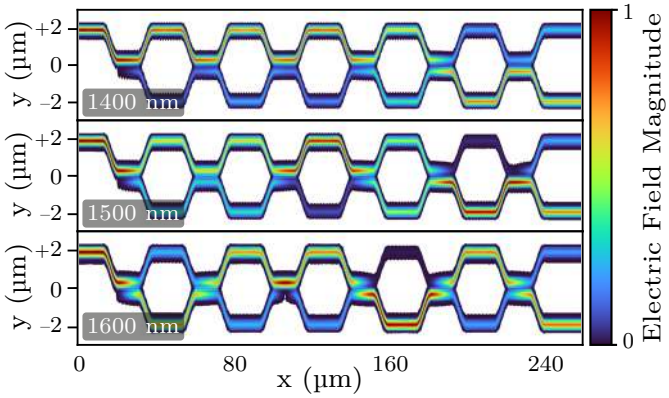

Figure 4.4: 3D-FDTD final simulation result of the 75/25 power splitter. Magnitude of the electric field at three different wavelengths obtained from 3D-FDTD simulations confirming broadband and flat-top operation for the 75/25 power splitter.

The optimized devices exhibit a remarkable match with the intended target functionality, validating the effectiveness of the design process. To further validate the performance of the optimized devices, Figure 4.6 presents a visualization of the optical input propagation in the final optimized device. The visualization is achieved using the electric field intensity obtained from 3D-FDTD simulations, providing additional confirmation of the successful design and performance of the optimized spectral duplexer. As anticipated, the duplexer effectively functions as a spectral splitter within its specified spectral design range, providing separate long-pass and short-pass outputs. This outcome aligns with our expectations and demonstrates the successful design and functionality of the optimized duplexer device. The results of the three devices serve as strong confirmation of the successful design and performance of the optimized devices. They demonstrate the effectiveness of our approach in designing photonic devices with wavelength-dependent, or ultra-broadband responses, further validating the capabilities of our framework.

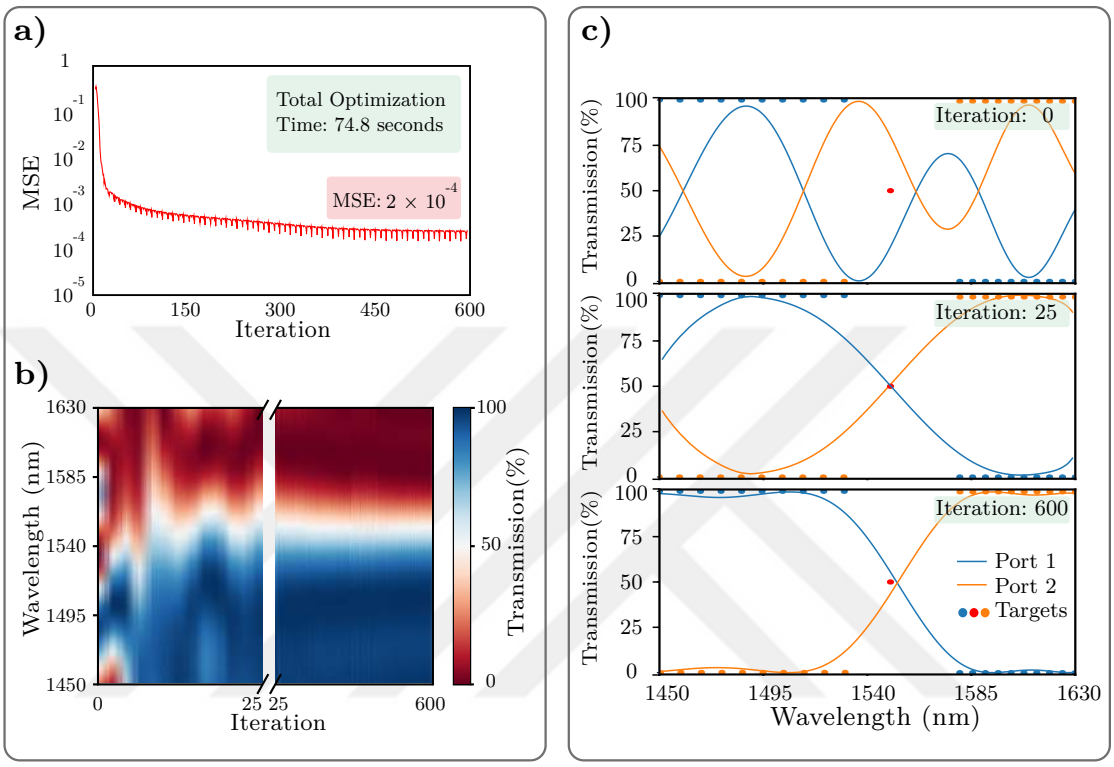

Figure 4.5: Optimization result of the spectral duplexer. (a) The mean squared error (MSE) versus iteration throughout optimization of a spectral duplexer with 6 layers of MZIs (144 trainable parameters, 480 µm device length). The device converges in six hundred iterations, within 74.8 seconds. (b) Transmission at the designated output port of the spectral duplexer as a function of wavelength. The evolution of this transmission through the iterative training process enables the device to achieve near-perfect transfer function by the end of optimization. (c) Transmission spectra for each output during optimizations show gradual convergence to the target transfer functions indicated by the circles. The spectral duplexer is optimized with 21 wavelengths between 1450-1630 nm with a target cutoff at 1550 nm.

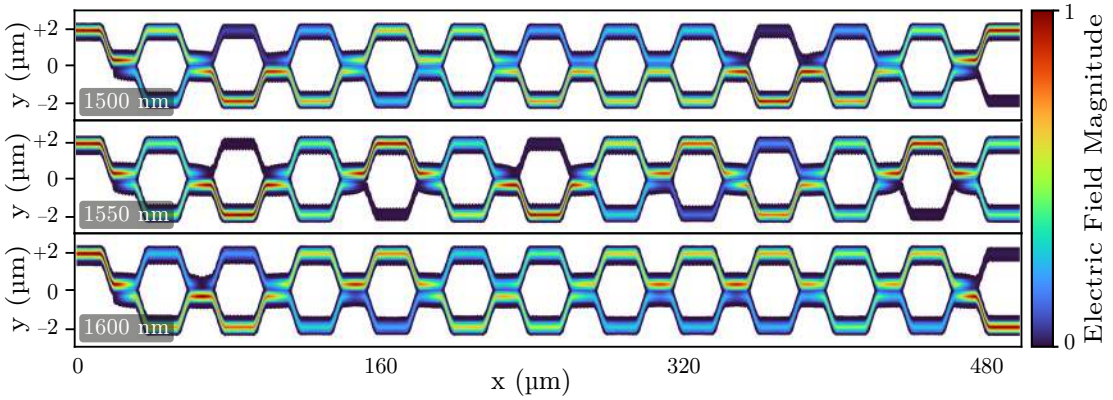

Figure 4.6: 3D-FDTD final simulation result of the spectral duplexer. Magnitude of the electric field at three different wavelengths obtained from 3D-FDTD simulations confirming broadband and flat-top operation for the spectral duplexer.

4.2 Experimental Demonstration and Analysis of Deep Photonic Networks

The performance evaluation of the power splitters and spectral duplexer was conducted using a continuous-wave tunable laser source (Santec TSL-710), an optical power meter (Santec MPM-210), on-chip grating couplers, and a polarization controller. Three plots are provided in Appendix to show the fabricated chip and measurement setups. Through the subsequent sections, all three devices showcased exceptional performance in experimental tests, closely aligning with the desired transmission responses as defined during the training phase. These findings serve as a practical confirmation of the broad applicability and effectiveness of our design methodology.

Next, in order to assess the robustness of our deep photonic networks against variations in the fabrication process, we conducted an analysis. Figures 4.7(b), 4.8(b), and 4.9(b) illustrate the resulting transmission responses obtained from transfer matrix calculations when considering potential over-etch and under-etch scenarios with waveguide widths and gaps deviating by up to 20 nm. To determine the responses of devices, simulations were performed on network structures incorporating updated

waveguide tapers and directional couplers to account for the specified etch offsets. Our observations reveal minimal deviation in the transmission response from the ideal case when employing 10 nm over- and under-etch. However, with a 20 nm deviation, more significant changes in the simulated transmission responses occur as a consequence of alterations in the wavelength-dependent phase profiles of the waveguide tapers and the shifted responses of the directional couplers, as expected. This is further illustrated in Figure 4.7(c), 4.8(c), and 4.9(c), where the mean squared error of the resulting transmission with varying over- and under-etch amounts is plotted. It is evident that the calculated error increases with larger over-/under-etch amounts, indicating a deterioration in device performance. Nevertheless, it is worth noting that all three devices can still function as intended, albeit with slightly diminished performance metrics, even with simulated 20 nm etch offsets.

4.2.1 50/50 power splitter

The experimental characterization results of the 50/50 power splitter are presented in Figure 4.7(a). The 50/50 splitter demonstrates excellent performance with a maximum deviation from the target transmission of only 6.42% for both output ports. Furthermore, the excess loss is measured to be less than 0.5 dB. This remarkable achievement validates the effectiveness of our network-based power splitter, which maintains a deviation of at most 0.6 dB across a measured bandwidth of 120 nm. Consequently, the 1-dB bandwidth exceeds this range, showcasing the wide operating bandwidth of our device.

4.2.2 75/25 power splitter

The experimental characterization results of the 75/25 power splitter are presented in Figure 4.8(a). In the case of the 75/25 splitter, the deviations from the target transmission are found to be within 5.49% (equivalent to 0.86 dB) for output port number one, and within 8.88% (equivalent to 0.55 dB) for output port number two. Additionally, the excess loss measured for both output ports is less than 0.61 dB. These results confirm that the 75/25 splitter demonstrates a wide operational

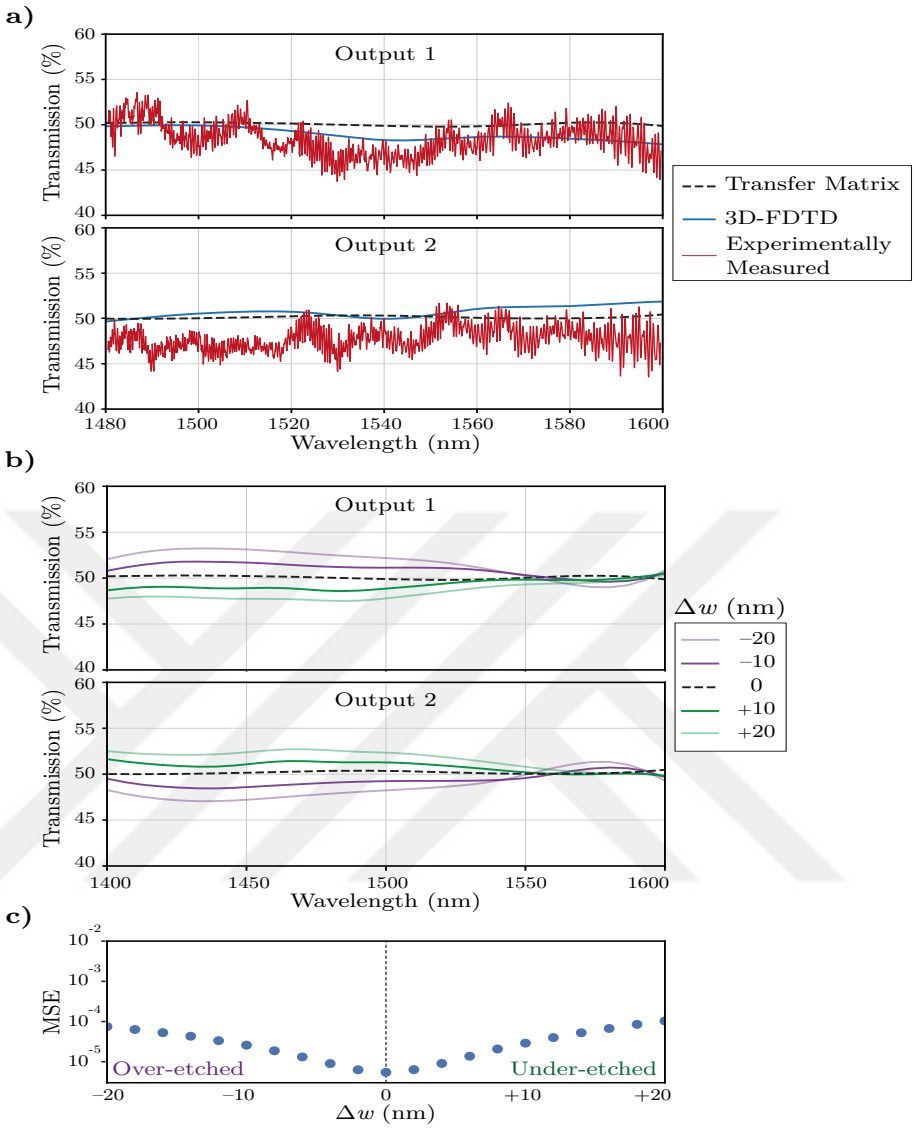

Figure 4.7: Experimental measurements and fabrication tolerance analysis of the 50/50 power splitter. (a) Measured transmission result together with transfer matrix and 3D-FDTD simulation at the output ports of the power splitter. The device demonstrates agreement with simulation results over wide bandwidths with flat-top and low-loss transmission responses. (b) Transfer-matrix analysis of robustness against fabrication-induced variations for 10 nm and 20 nm over-etch and under-etch cases. All components including directional couplers, S-bends, and waveguide tapers, are uniformly modified in simulation with the indicated etch offsets. (c) Resulting mean squared error in the device subject to over-etch and under-etch variations. With 20 nm modification of the waveguide widths, the resulting error typically increases by 1-2 orders of magnitude, corresponding to the changes in the simulated transfer function of the respective device.

bandwidth of at least 120 nm, which represents the maximum measurement range achievable in our experiments.

4.2.3 Spectral Duplexer

The experimental characterization results of the spectral duplexer are presented in Figure 4.9(a). Within the pass-bands, the measured maximum loss for the short-pass output is 11.45% (equivalent to 0.52 dB), while for the long-pass output it is 15.30% (equivalent to 0.72 dB). The excess insertion loss, occurring at 1590 nm, is measured to be approximately 0.66 dB. The measured cutoff wavelength is around 1555.2 nm, slightly deviating from the specified target cutoff wavelength of 1550 nm. The extinction ratio between the two outputs is better than 15 dB across most of the characterized wavelength range, with a slight decrease to 13.6 dB at the edge of the measured spectrum (1600 nm).

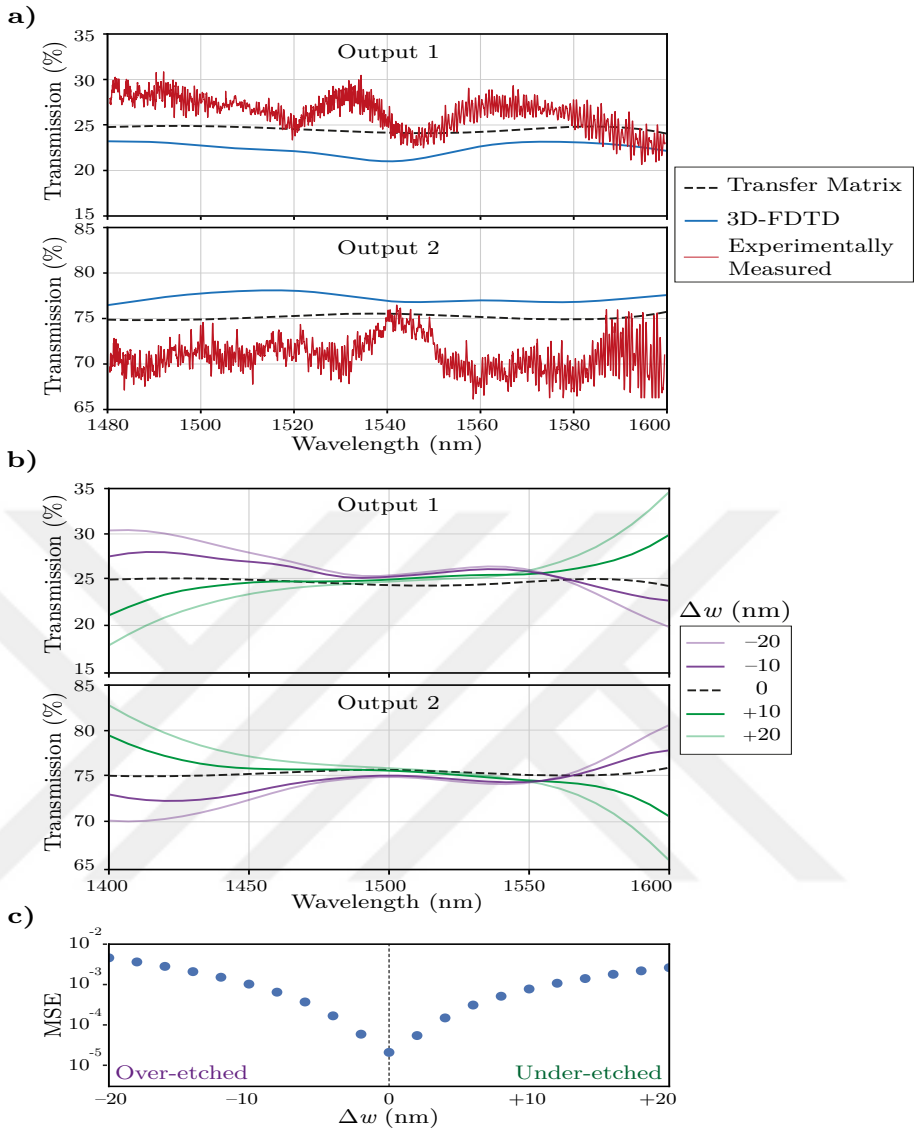

Figure 4.8: Experimental measurements and fabrication tolerance analysis of the 75/25 power splitter. (a) Measured transmission result together with transfer matrix and 3D-FDTD simulation at the output ports of the power splitter. The device demonstrates agreement with simulation results over wide bandwidths with flat-top and low-loss transmission responses. (b) Transfer-matrix analysis of robustness against fabrication-induced variations for 10 nm and 20 nm over-etch and under-etch cases. All components including directional couplers, S-bends, and waveguide tapers, are uniformly modified in simulation with the indicated etch offsets. (c) Resulting mean squared error in the device subject to over-etch and under-etch variations. With 20 nm modification of the waveguide widths, the resulting error typically increases by 1-2 orders of magnitude, corresponding to the changes in the simulated transfer function of the respective device.

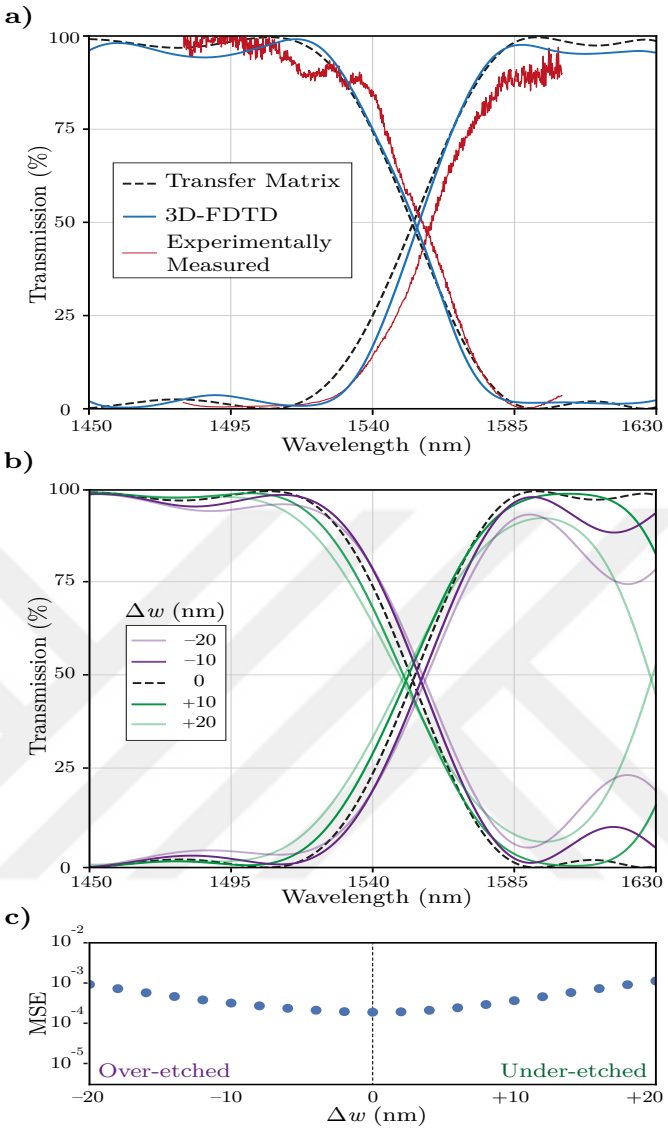

Figure 4.9: Experimental measurements and fabrication tolerance analysis of the spectral duplexer. (a) Measured transmission result together with transfer matrix and 3D-FDTD simulation at the output ports of the spectral duplexer. The device demonstrates agreement with simulation results over wide bandwidths with flat-top and low-loss transmission responses. (b) Transfer-matrix analysis of robustness against fabrication-induced variations for 10 nm and 20 nm over-etch and under-etch cases. All components including directional couplers, S-bends, and waveguide tapers, are uniformly modified in simulation with the indicated etch offsets. (c) Resulting mean squared error in the device subject to over-etch and under-etch variations. With 20 nm modification of the waveguide widths, the resulting error typically increases by 1-2 orders of magnitude, corresponding to the changes in the simulated transfer function of the respective device.

Chapter 5

DEEP PHOTONIC NETWORK CAPABILITY AND FABRICATION ROBUSTNESS

5.1 Number of Interferometric Layers

The scalability of our deep photonic networks, combined with the computational efficiency of our simulation and optimization framework, enables the design of highly capable networks with a large number of degrees of freedom for creating complex optical devices. In our architecture, the choice of the number of interferometric layers plays a crucial role in determining the network's degrees of freedom. Increasing the number of layers initially enhances the trainability and capability of the network. However, it also introduces additional propagation loss due to the waveguide bends added with each layer. This tradeoff between device capability and excess loss can be evaluated by analyzing devices with varying numbers of layers trained to achieve the same desired functionality.

Moreover, it should be noted that longer networks with more interferometric layers, while providing increased degrees of freedom and complex optical capabilities, are also more susceptible to fabrication variations. The accumulated errors in the phase profiles propagate through the additional layers, leading to a negative impact on device performance. This analysis serves as a valuable guideline for determining the optimal number of layers in the design of specific structures using our custom networks. In our approach, we take into consideration both the final error and the fabrication tolerance of the resulting photonic networks. When devices with different numbers of layers exhibit similar performances in terms of calculated errors, we prioritize selecting the device with a smaller number of layers. This choice ensures a higher level of robustness against fabrication variations while maintaining a compact device footprint.

5.1.1 50/50 power splitter

In Figure 5.1, we present the final mean squared error in the simulated transmission responses for different 50/50 power splitters designed with varying numbers of layers, ranging from $M = 2$ to $M = 60$. As expected, the simulated error initially decreases and reaches a minimum with networks consisting of 3 and 4 layers. However, as the number of layers increases beyond this point, the accumulation of excess loss in the additional layers outweighs the benefits of enhanced network capability. Consequently, a larger calculated error and an inferior transmission response are observed.

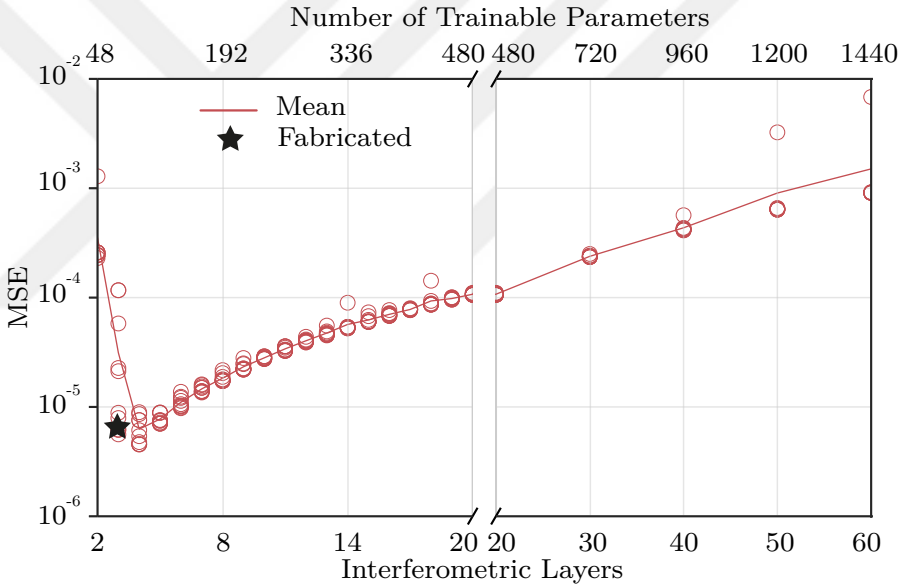

Figure 5.1: Influence of network size on the final 50/50 power splitter performance. Performance of the 50/50 power splitter deep photonic network as a function of the number of interferometric layers, which directly controls the number of trainable parameters. The plotted mean squared error includes propagation loss in the directional couplers and the S-bends extracted from their 3D-FDTD simulations. For each network size, ten different randomly-initialized devices are optimized and depicted with red circles.

To further investigate the impact of fabrication variations, we analyze the fabrication tolerance of 50/50 splitters constructed with different numbers of layers in Figure 5.2. In this analysis, we plot the mean squared error as a function of the etch offset. The results clearly indicate that longer networks with a higher number of layers are more susceptible to changes induced during fabrication. This heightened sensitivity arises from the cumulative effect of phase and coupling errors within consecutive MZIs. For example, although the minimum error calculated is similar for 3-layer and 4-layer splitters, the 4-layer network exhibits significantly poorer performance when subjected to etch offsets reaching 20 nm.

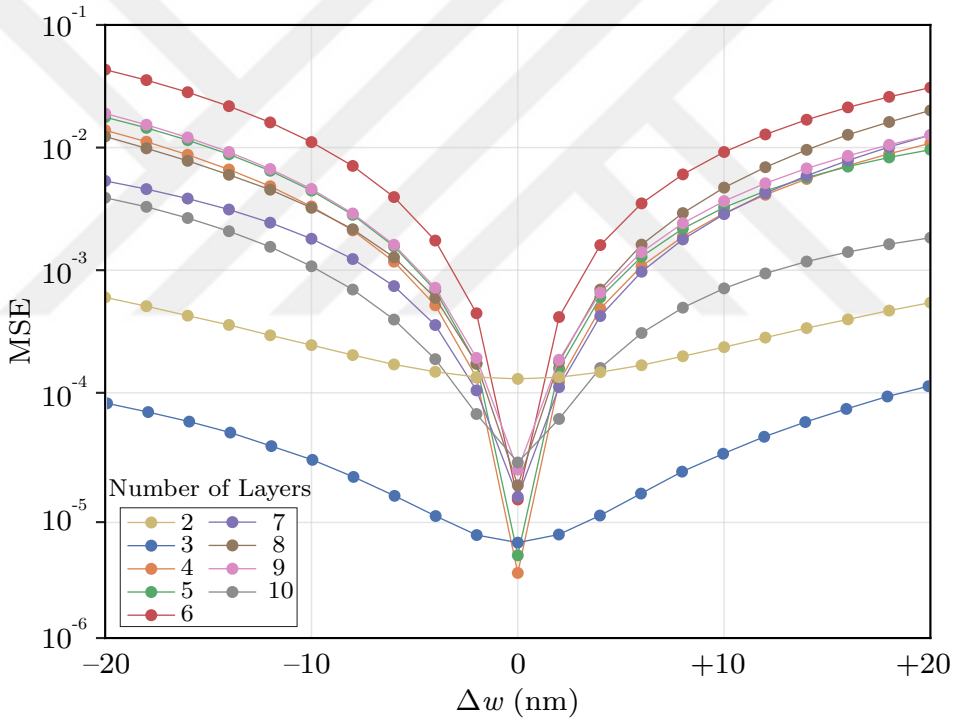

Figure 5.2: Robustness against fabrication variations for 50/50 power splitter. Robustness of device performance against fabrication-induced variations with number of layers from $M = 2$ through $M = 10$. While increasing the number of interferometric layers initially provides better-performing devices under ideal fabrication conditions ($\Delta w = 0$), longer devices perform worse under significant fabrication variations due to accumulating phase errors.

5.1.2 75/25 power splitter

In Figure 5.3, we extend our investigations to the 75/25 power splitter by comparing the final performance and robustness of 10 different devices designed with varying numbers of layers, ranging from $M = 2$ to $M = 60$. The resulting number of trainable parameters falls within the range of 48 to 1440. In Figure 5.3, we plot the calculated errors for the 75/25 power splitter and observe an initial decrease in error as the number of layers increases, indicating improved network capability. The minimum errors are typically achieved with networks consisting of around $M = 3$ to 5 layers for the 75/25 splitter. It is important to note that deeper networks, with more layers, offer greater degrees of freedom and enhanced capabilities.

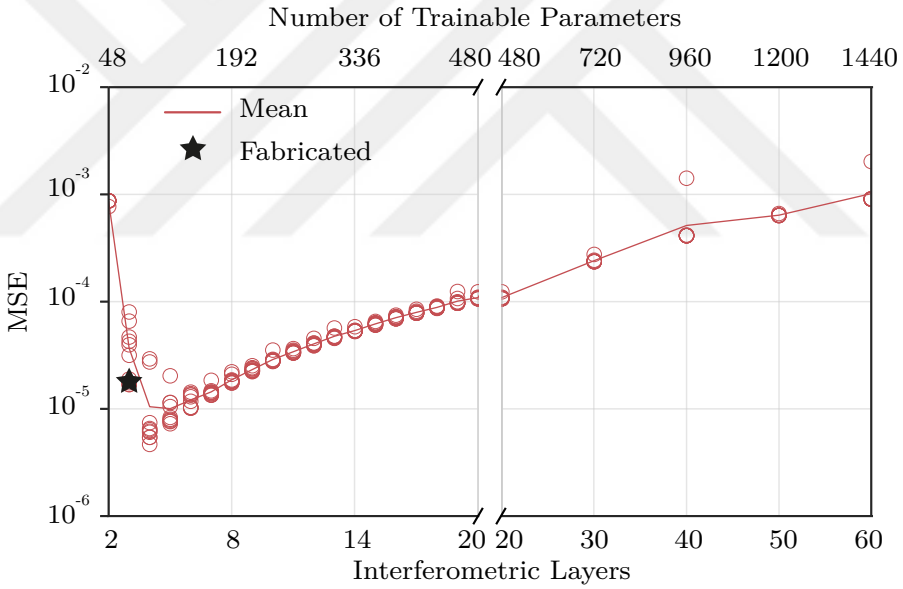

Figure 5.3: Influence of network size on the final 75/25 power splitter performance. Mean squared error for 75/25 power splitter deep photonic network as a function of the number of interferometric layers. The plotted error includes propagation loss in the directional couplers and the S-bends extracted from their 3D-FDTD simulations. For each network size, ten different randomly-initialized devices are optimized and depicted with red circles.

However, they are also more susceptible to fabrication imperfections, including phase errors and losses in directional couplers and S-bends. Figure 5.4 presents the profile of errors as a function of the etch offset Δw for different over-etch and under-etch scenarios. Generally, shorter networks exhibit better performance (flatter error profile) under strong etch offsets reaching 20 nm. While the minimum error is achieved with a 4-layer network for the 75/25 splitter, this network shows significantly inferior fabrication tolerance under 20 nm etch offsets compared to the 3-layer network. This emphasizes the trade-off between network capability and robustness against fabrication variations.

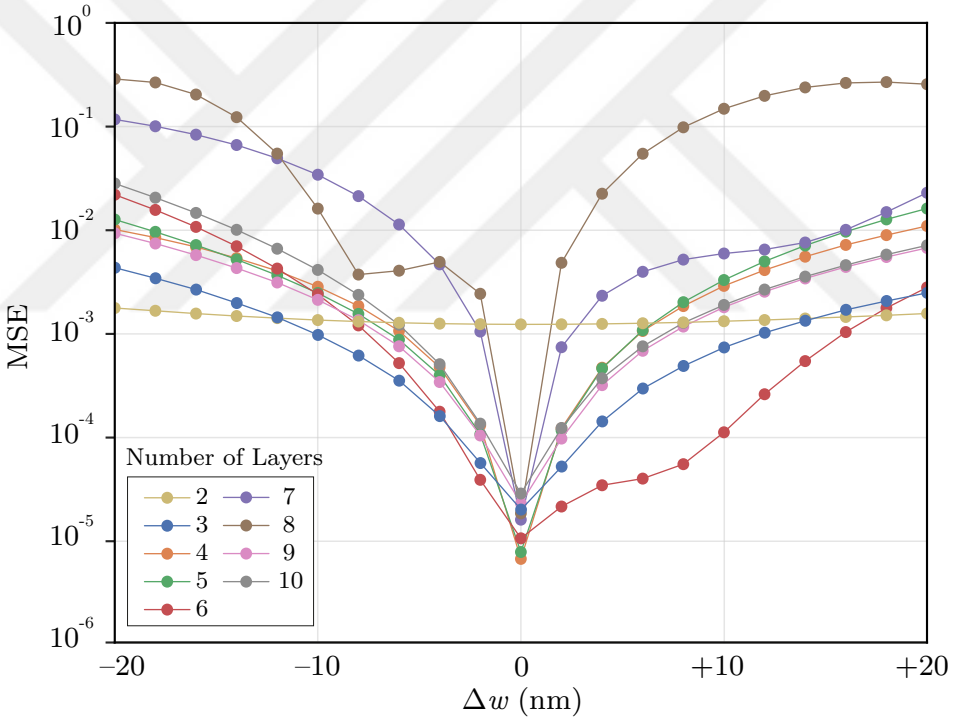

Figure 5.4: Robustness against fabrication variations for 75/25 power splitter. A greater number of interferometric layers initially reduces the calculated errors. However, longer devices suffer from stronger deviations from target functionality under fabrication variations.

5.1.3 Spectral Duplexer

In Figure 5.5, we extend our investigations to the spectral duplexer by comparing the final performance and robustness of 10 different devices designed with varying numbers of layers, ranging from $M = 2$ to $M = 60$. The resulting number of trainable parameters falls within the range of 48 to 1440. In Figure 5.5, we plot the calculated errors for the spectral duplexer and observe an initial decrease in error as the number of layers increases. This decrease is due to the increased network capability provided by deeper networks with more layers. The minimum errors are typically achieved with networks consisting of around $M = 9$ to 14 layers for the duplexer. As the spectral duplexer performs a more complex spectral functionality, it requires a deeper and more capable network with greater degrees of freedom to achieve optimal performance.

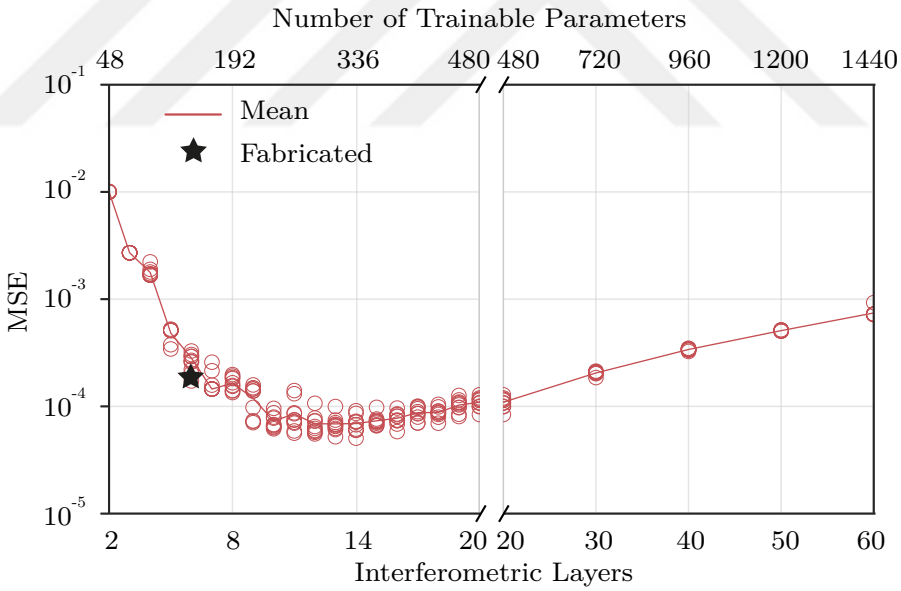

Figure 5.5: Influence of network size on the final spectral duplexer performance. Mean squared error for spectral duplexer deep photonic network as functions of the number of interferometric layers. The plotted error includes propagation loss in the directional couplers and the S-bends extracted from their 3D-FDTD simulations. For each network size, ten different randomly-initialized devices are optimized and depicted with red circles.

However, similar to the power splitters, deeper networks are more susceptible to fabrication imperfections, such as phase errors and losses in directional couplers and S-bends. This is evident from the general profile of errors plotted as a function of the etch offset Δw in Figure 5.6, where shorter networks generally demonstrate better (flatter) performance under strong etch offsets reaching 20 nm. For the spectral duplexer, the minimum error is reached with a 10-layer network. However, the fabrication tolerance of this 10-layer network is significantly inferior compared to the 6-layer and 7-layer networks. It is worth noting that the spectral duplexer exhibits better fabrication tolerance in these simulations compared to both power splitters. This difference can be attributed to the final error of the duplexer, which is already 1-2 orders of magnitude greater (2×10^{-4}) than the final errors for the power splitters (between 7×10^{-6} and 2×10^{-5}).

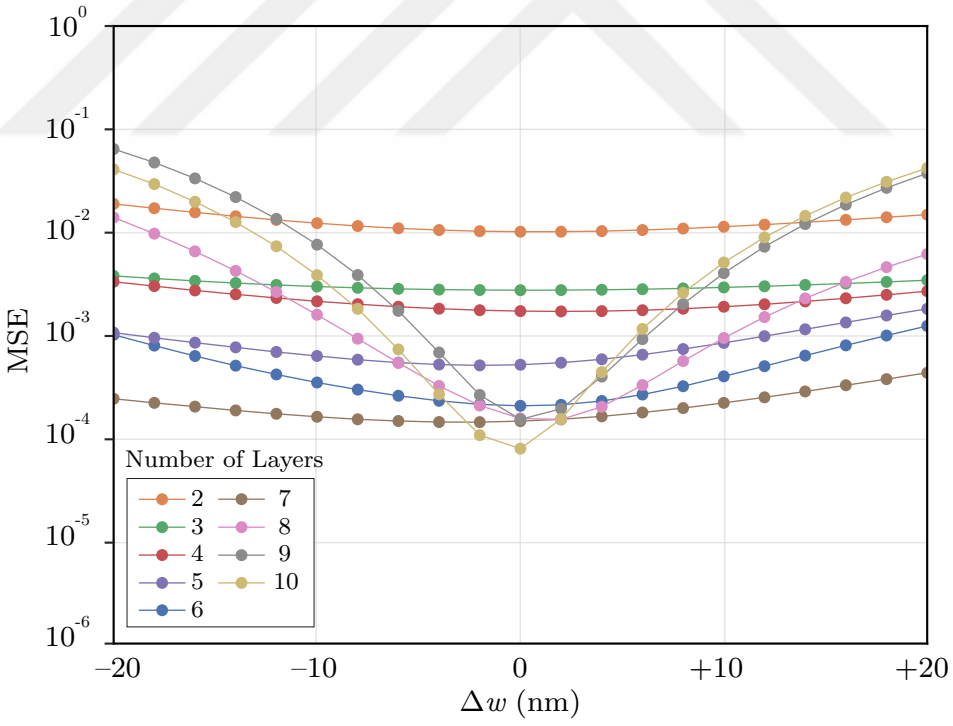

Figure 5.6: Robustness against fabrication variations for the spectral duplexer. A greater number of interferometric layers initially reduces the calculated errors. However, longer devices suffer from stronger deviations from target functionality under fabrication variations.

5.2 Number of Trainable Parameters

The trainable parameters in our photonic networks consist of widths and lengths for each custom taper, as depicted in Figure 3.2. Specifically, each Mach-Zehnder Interferometer (MZI) in the network is constructed from four waveguide tapers with trainable widths, four trainable lengths (one for each taper), and two fixed (non-trainable) directional couplers. The network depth directly determines the number of trainable lengths in the design. However, the number of trainable widths per custom taper is a parameter specified by the designer. In our presented devices, we utilized five trainable widths for each taper, resulting in a total of $4 \times (5 + 1) = 24$ trainable parameters per MZI. The constituent widths and the taper length for each taper are used to calculate the accumulated phase, as discussed before. We generally limit the number of trainable widths to $\xi \leq (L_{\text{min}} - 1\ \mu\text{m})/1\ \mu\text{m}$, as discussed in Chapter 3. This limitation ensures slowly-varying waveguide geometries in the propagation direction, minimizing unwanted loss. In our devices with a maximum taper length of 10 µm, this guideline suggests a recommended number of trainable widths of up to 9. However, within this constraint, the number of trainable widths remains a design choice specified by the user. For the devices demonstrated in our work, we selected $\xi = 5$ trainable widths per taper. We found that beyond five trainable widths, there was a negligible additional benefit, as the required optimized phase profiles could already be achieved in the custom tapers. Moreover, choosing a higher number of trainable widths ($\xi > 5$) may lead to slightly longer optimization times without improving device performance.

Chapter 6

COMPUTATIONAL PERFORMANCE AND TIME REQUIREMENTS FOR NETWORK OPTIMIZATION

One of the key benefits of our deep photonic network design framework is its capability to replace computationally demanding 3D-FDTD simulations with physically accurate and computationally efficient scattering matrix calculations. This is particularly advantageous in design tasks that involve iterative modifications of device parameters and repeated simulations of device responses. By leveraging state-of-the-art machine learning software and hardware infrastructure, our design framework achieves both computational efficiency and scalability, effectively addressing the computational requirements associated with these simulations. This enables faster and more streamlined design iterations, enhancing the overall efficiency of the design process.

In Figures 6.1, we demonstrate this computational performance for the design of several power splitters with various numbers and configurations of output ports, using a single Tesla V100 GPU. Specifically, we consider 2-port, 4-port, and 8-port broadband power splitters in evenly distributed and randomly distributed output power configurations, optimized at 32 evenly-spaced wavelengths between 1400-1600 nm. In Figure 6.1(a), we plot the total time required for the convergence of each device as a function of the network depth. A relative convergence criterion of

$$\frac{\mid J(x)_{\text{current}} - J(x)_{\text{previous}} \mid}{\max(J(x)_{\text{current}}, J(x)_{\text{previous}})} < 10^{-3} \tag{6.1}$$

was used for all devices where $J(x)$ was defined in Eq.(3.4). As expected, the total optimization time scales with the number of interferometric layers (M) and the number of network outputs (N). As a function of these two parameters and the number of trainable widths per custom waveguide taper, the total number of

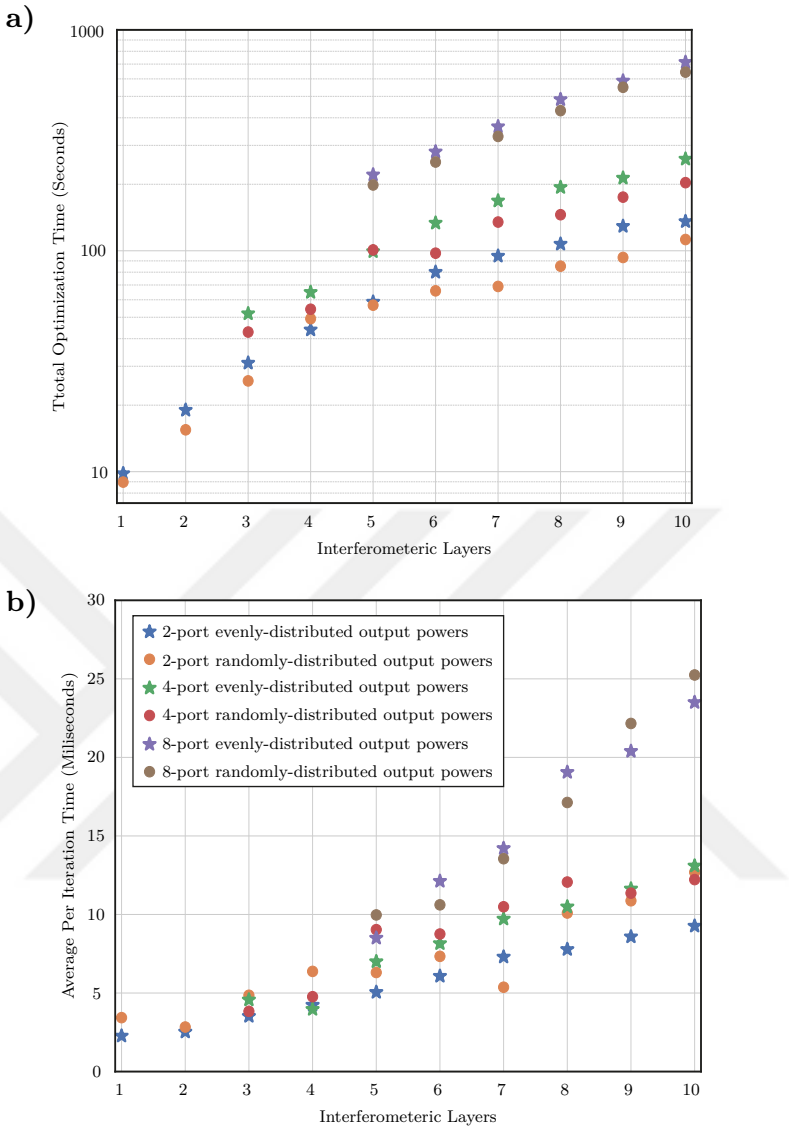

Figure 6.1: Computational performance of the design framework for deep photonic networks as arbitrary power splitters. (a) Total optimization time for networks with different number of interferometric layers. (b) Average time per optimization iteration. Networks up to 10-layers deep with the number of output ports $N = 2$, 4, and 8 are optimized for broadband power splitting using 32 wavelengths between 1400 nm and 1600 nm. Separate photonic networks are designed with evenly distributed output powers, and randomly distributed output powers. Light is injected at the input number $\lfloor \frac{N}{2} \rfloor$ for all devices. All optimizations were performed using an open-source, end-to-end deep learning software library [Bradbury et al., 2018, Tra, 2020], adaptive moment estimation optimizer [Kingma and Ba, 2014], and a single Tesla V100 GPU. Optimizations were completed in several hundred iterations for each device, using a convergence criterion of 10^{-3} relative change in the overall objective $J(x)$.

trainable network parameters is given by

$$\begin{cases} 4(\xi+1)M, & N=2 \\ 4(\xi+1)(\lceil \frac{M}{2} \rceil \lfloor \frac{N}{2} \rfloor + \lfloor \frac{M}{2} \rfloor \lfloor \frac{N-1}{2} \rfloor), & N \geq 3. \end{cases} \tag{6.2}$$

The computational efficiency of our design framework is evident in the optimization procedures for deep photonic networks. Even for networks with a depth of 10 layers, corresponding to a device length of 800 µm, and complex functionalities such as random and broadband output power splitting with 8 separate output ports, the entire optimization process is completed in less than 12 minutes. This achievement represents a significant improvement in scalability and computational speed compared to optimization procedures based on 3D-FDTD simulations, as reported in previous studies [Piggott et al., 2015, Lu and Vučković, 2013, Jia et al., 2018, Zhang and Liboiron-Ladouceur, 2021, Zhang et al., 2022]. Our framework offers multiple orders of magnitude advancements in terms of efficiency and computational performance. Our design framework offers remarkable efficiency for smaller photonic networks with simpler functionalities. For networks of up to four layers, the optimization process is completed in less than approximately 1 minute. This rapid turnaround time empowers photonic designers to iterate through different versions of their networks and explore device performance with various hyperparameters. The ability to quickly simulate and evaluate the performance of photonic devices using physically accurate simulations within minutes is a crucial advantage, enabling designers to efficiently explore design variations and make informed decisions.

The scalability of our design framework is further demonstrated by the per-iteration time for each network optimization, as shown in Figure 6.1(b). In this context, an iteration refers to the computational operation of updating the network parameters once, after calculating the objective function $J(x)$ and its gradient $\nabla_x J$ using all 32 wavelengths in a parallelized manner. The average time required for each iteration ranges from a few milliseconds to a few tens of milliseconds, depending on the network depth (and therefore the number of trainable parameters). In addition to the modularity of our photonic network construction, the efficient performance of our design framework is facilitated by the use of open-source deep learning software

libraries such as JAX [Bradbury et al., 2018]and Trax [Tra, 2020]. These libraries enable automatic differentiation-compatible computations to be executed on various hardware accelerators (CPU, GPU, or TPU) in a parallelized manner through just-in-time (JIT) compilation [Aycock, 2003] and execution of required function calls. It is important to note that after creating a photonic network model, there is a certain preparation time required for JIT compilation to complete before iterative optimization can commence. For the specific optimizations shown in Figure 6.1(a), the preparation time ranged between 8 and 650 seconds for 1-layer to 10-layer deep networks. This preparation time is accounted for in the total optimization durations plotted in Figure 6.1(a) but excluded from the per-iteration time plotted in Figure 6.1(b).

Chapter 7

CONCLUSION AND DISCUSSION

Our design framework offers a computationally efficient and physically accurate approach to designing deep photonic network architectures for on-chip optical systems. Although we have specifically demonstrated its application in silicon-based devices, the methodology can be extended to various material platforms and spectral applications. Despite having been designed by an arbitrarily capable photonic optimization framework, all three of our proof-of-concept devices experimentally demonstrate state-of-the-art performance. Even in comparison to structures with specific power splitter geometries [Kim et al., 2022, Lin and Shi, 2019, Chang et al., 2018, Chen et al., 2017, Wang et al., 2016, Lu et al., 2015], our network-type power splitters exhibit better experimental performance metrics with 1-dB bandwidths reaching as wide as the entire measured spectrum of 120 nm and over 200 nm in simulation. Similarly, our duplexer demonstrates better experimental performance than devices with similar functionality [Zhang et al., 2022, Xu et al., 2019, Magden et al., 2018, Piggott et al., 2015], with less than 0.66 dB insertion loss, flat-top transmissions at both outputs, and a cutoff wavelength shift of only 5 nm. Even though our networks at their current state do not implement reconfigurable photonic capabilities, they also do not require the use of electronic driving circuitry for detuning or calibration purposes. Consequently, our devices are not prone to long-term stability issues and excessive power draw requirements of systems with thermal, MEMS-based, or electro-optic phase shifters [Xu et al., 2022, Quack et al., 2023]. The absence of electronic driving circuitry simplifies the device architecture and eliminates the need for complex control mechanisms. This leads to improved device reliability and reduced power consumption, making our networks more practical and efficient for on-chip applications. By leveraging the inherent properties of the photonic network design, such as waveguide geometries and coupling coefficients,

our devices achieve their desired functionality without the need for additional electronic components. This advantage contributes to the scalability and robustness of our photonic systems.

In summary, our design framework offers a highly scalable approach for implementing arbitrary transfer functions on-chip. By formulating photonic design as a constrained optimization problem, we leverage the power of physics-informed machine learning and accurate waveguide parameters to enable efficient and accurate simulations of photonic devices. The integration of 3D-FDTD simulations within our framework allows us to capture the intricate behavior of light propagation and interactions with various components. This ensures the accuracy and reliability of our simulations, which in turn inform the optimization process. Through this methodology, we achieve rapid design iterations and scalability, enabling the creation of complex and high-performance photonic devices. By considering fabrication compatibility as an inherent constraint, our approach ensures that the designed devices can be feasibly realized using available fabrication techniques. Our modular network design approach offers a significant number of degrees of freedom, allowing for the implementation of complex photonic functionalities. This modularity enables the design of large-scale integrated photonic systems by combining multiple custom layers of MZIs. The ability to incorporate specific phase and dispersion profiles into the target functionality of the photonic networks further enhances their versatility and customization. Additionally, our computational design framework accurately captures and maintains complete phase information throughout the individual network components. This opens up possibilities for designing photonic networks with integrated on-chip amplifiers and lasers [Zhou et al., 2023, Li et al., 2018], modulators and detectors [Liu et al., 2015], and other components. By leveraging the rapid individual device simulations enabled by our framework, future designs can explore various integrated functionalities and robustness against fabrication-induced variations [Pérez and Capmany, 2019]. These capabilities hold great promise for advancing optical communication applications, neuromorphic photonic information processors, and medical/biological sensing. The ability to design photonic components with arbitrary transfer functions and tailored functionalities paves the way for

innovative advancements in these fields.

BIBLIOGRAPHY

[Tra, 2020] (2020). Trax: an end-to-end library for deep learning that focuses on clear code and speed, https://github.com/google/trax.

[Alom et al., 2019] Alom, M. Z., Taha, T. M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M. S., Hasan, M., Van Essen, B. C., Awwal, A. A., and Asari, V. K. (2019). A state-of-the-art survey on deep learning theory and architectures. *electronics*, 8(3):292.

[Ashtiani et al., 2022] Ashtiani, F., Geers, A. J., and Aflatouni, F. (2022). An on-chip photonic deep neural network for image classification. *Nature*, pages 1–6.

[Aycock, 2003] Aycock, J. (2003). A brief history of just-in-time. *ACM Computing Surveys (CSUR)*, 35(2):97–113.

[Bogaerts and Chrostowski, 2018] Bogaerts, W. and Chrostowski, L. (2018). Silicon photonics circuit design: methods, tools and challenges. *Laser & Photonics Reviews*, 12(4):1700237.

[Bogaerts et al., 2020] Bogaerts, W., Pérez, D., Capmany, J., Miller, D. A., Poon, J., Englund, D., Morichetti, F., and Melloni, A. (2020). Programmable photonic circuits. *Nature*, 586(7828):207–216.

[Bradbury et al., 2018] Bradbury, J., Frostig, R., Hawkins, P., Johnson, M. J., Leary, C., Maclaurin, D., Necula, G., Paszke, A., VanderPlas, J., Wanderman-Milne, S., and Zhang, Q. (2018). JAX: composable transformations of Python+NumPy programs, http://github.com/google/jax.

[Capmany and Pérez, 2020] Capmany, J. and Pérez, D. (2020). *Programmable integrated photonics*. Oxford University Press.

[Carolan et al., 2015] Carolan, J., Harrold, C., Sparrow, C., Martín-López, E., Russell, N. J., Silverstone, J. W., Shadbolt, P. J., Matsuda, N., Oguma, M., Itoh, M., et al. (2015). Universal linear optics. *Science*, 349(6249):711–716.

[Chang et al., 2018] Chang, W., Ren, X., Ao, Y., Lu, L., Cheng, M., Deng, L., Liu, D., and Zhang, M. (2018). Inverse design and demonstration of an ultracompact broadband dual-mode 3 db power splitter. *Opt. Express*, 26(18):24135–24144.

[Chen et al., 2017] Chen, G. F., Ong, J. R., Ang, T. Y., Lim, S. T., Png, C. E., and Tan, D. T. (2017). Broadband silicon-on-insulator directional couplers using a combination of straight and curved waveguide sections. *Scientific reports*, 7(1):1–8.

[Chrostowski and Hochberg, 2015] Chrostowski, L. and Hochberg, M. (2015). *Silicon photonics design: from devices to systems*. Cambridge University Press.

[Feldmann et al., 2021] Feldmann, J., Youngblood, N., Karpov, M., Gehring, H., Li, X., Stappers, M., Le Gallo, M., Fu, X., Lukashchuk, A., Raja, A. S., et al. (2021). Parallel convolutional processing using an integrated photonic tensor core. *Nature*, 589(7840):52–58.

[Fu et al., 2014] Fu, Y., Ye, T., Tang, W., and Chu, T. (2014). Efficient adiabatic silicon-on-insulator waveguide taper. *Photonics Research*, 2(3):A41–A44.

[Gabrielli, 2020] Gabrielli, L. H. (2020). Gdstk (GDSII Tool Kit) a C++ library for creation and manipulation of GDSII and OASIS files, https://github.com/heitzmann/gdstk.

[Hanin and Rolnick, 2018] Hanin, B. and Rolnick, D. (2018). How to start training: The effect of initialization and architecture. *Advances in Neural Information Processing Systems*, 31.

[Harris et al., 2016] Harris, N. C., Bunandar, D., Pant, M., Steinbrecher, G. R.,

Mower, J., Prabhu, M., Baehr-Jones, T., Hochberg, M., and Englund, D. (2016). Large-scale quantum photonic circuits in silicon. *Nanophotonics*, 5(3):456–468.

[Harris et al., 2018] Harris, N. C., Carolan, J., Bunandar, D., Prabhu, M., Hochberg, M., Baehr-Jones, T., Fanto, M. L., Smith, A. M., Tison, C. C., Alsing, P. M., et al. (2018). Linear programmable nanophotonic processors. *Optica*, 5(12):1623–1631.

[Hegde, 2020] Hegde, R. S. (2020). Deep learning: a new tool for photonic nanostructure design. *Nanoscale Advances*, 2(3):1007–1023.

[Hu et al., 2017] Hu, T., Dong, B., Luo, X., Liow, T.-Y., Song, J., Lee, C., and Lo, G.-Q. (2017). Silicon photonic platforms for mid-infrared applications. *Photonics Research*, 5(5):417–430.

[Jensen and Sigmund, 2011] Jensen, J. S. and Sigmund, O. (2011). Topology optimization for nano-photonics. *Laser & Photonics Reviews*, 5(2):308–321.

[Jia et al., 2018] Jia, H., Zhou, T., Fu, X., Ding, J., and Yang, L. (2018). Inverse-design and demonstration of ultracompact silicon meta-structure mode exchange device. *Acs Photonics*, 5(5):1833–1838.

[Jiang et al., 2021] Jiang, J., Chen, M., and Fan, J. A. (2021). Deep neural networks for the evaluation and design of photonic devices. *Nature Reviews Materials*, 6(8):679–700.

[Kim et al., 2022] Kim, J., Kim, J.-Y., Yoon, J., Yoon, H., Park, H.-H., and Kurt, H. (2022). Experimental demonstration of inverse-designed silicon integrated photonic power splitters. *Nanophotonics*, 11(20):4581–4590.

[Kingma and Ba, 2014] Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

[Li et al., 2018] Li, N., Vermeulen, D., Su, Z., Magden, E. S., Xin, M., Singh, N., Ruocco, A., Notaros, J., Poulton, C. V., Timurdogan, E., et al. (2018). Monolithically integrated erbium-doped tunable laser on a cmos-compatible silicon photonics platform. *Optics Express*, 26(13):16200–16211.

[Lin et al., 2018] Lin, X., Rivenson, Y., Yardimci, N. T., Veli, M., Luo, Y., Jarrahi, M., and Ozcan, A. (2018). All-optical machine learning using diffractive deep neural networks. *Science*, 361(6406):1004–1008.

[Lin and Shi, 2019] Lin, Z. and Shi, W. (2019). Broadband, low-loss silicon photonic y-junction with an arbitrary power splitting ratio. *Opt. Express*, 27(10):14338–14343.

[Liu et al., 2015] Liu, K., Ye, C. R., Khan, S., and Sorger, V. J. (2015). Review and perspective on ultrafast wavelength-size electro-optic modulators. *Laser & Photonics Reviews*, 9(2):172–194.

[Lu and Vučković, 2013] Lu, J. and Vučković, J. (2013). Nanophotonic computational design. *Optics express*, 21(11):13351–13367.

[Lu et al., 2015] Lu, Z., Yun, H., Wang, Y., Chen, Z., Zhang, F., Jaeger, N. A. F., and Chrostowski, L. (2015). Broadband silicon photonic directional coupler using asymmetric-waveguide based phase control. *Opt. Express*, 23(3):3795–3808.

[Magden et al., 2018] Magden, E. S., Li, N., Raval, M., Poulton, C. V., Ruocco, A., Singh, N., Vermeulen, D., Ippen, E. P., Kolodziejski, L. A., and Watts, M. R. (2018). Transmissive silicon photonic dichroic filters with spectrally selective waveguides. *Nature communications*, 9(1):3009.

[Marpaung et al., 2019] Marpaung, D., Yao, J., and Capmany, J. (2019). Integrated microwave photonics. *Nature photonics*, 13(2):80–90.

[Miller, 2013] Miller, D. A. (2013). Self-configuring universal linear optical component. *Photonics Research*, 1(1):1–15.

[Miller, 2015] Miller, D. A. (2015). Perfect optics with imperfect components. *Optica*, 2(8):747–750.

[Milton and Burns, 1977] Milton, A. and Burns, W. (1977). Mode coupling in optical waveguide horns. *IEEE Journal of Quantum Electronics*, 13(10):828–835.

[Molesky et al., 2018] Molesky, S., Lin, Z., Piggott, A. Y., Jin, W., Vucković, J., and Rodriguez, A. W. (2018). Inverse design in nanophotonics. *Nature Photonics*, 12(11):659–670.

[Pérez and Capmany, 2019] Pérez, D. and Capmany, J. (2019). Scalable analysis for arbitrary photonic integrated waveguide meshes. *Optica*, 6(1):19–27.

[Pérez et al., 2017] Pérez, D., Gasulla, I., Crudgington, L., Thomson, D. J., Khokhar, A. Z., Li, K., Cao, W., Mashanovich, G. Z., and Capmany, J. (2017). Multipurpose silicon photonics signal processor core. *Nature communications*, 8(1):1–9.

[Pérez-López et al., 2020] Pérez-López, D., López, A., DasMahapatra, P., and Capmany, J. (2020). Multipurpose self-configuration of programmable photonic circuits. *Nature communications*, 11(1):1–11.

[Piggott et al., 2015] Piggott, A. Y., Lu, J., Lagoudakis, K. G., Petykiewicz, J., Babinec, T. M., and Vučković, J. (2015). Inverse design and demonstration of a compact and broadband on-chip wavelength demultiplexer. *Nature Photonics*, 9(6):374–377.

[Piggott et al., 2020] Piggott, A. Y., Ma, E. Y., Su, L., Ahn, G. H., Sapra, N. V., Vercruysse, D., Netherton, A. M., Khope, A. S., Bowers, J. E., and Vuckovic, J. (2020). Inverse-designed photonics for semiconductor foundries. *ACS Photonics*, 7(3):569–575.

[Poulton et al., 2019] Poulton, C. V., Byrd, M. J., Russo, P., Timurdogan, E., Khandaker, M., Vermeulen, D., and Watts, M. R. (2019). Long-range lidar

and free-space data communication with high-performance optical phased arrays. *IEEE Journal of Selected Topics in Quantum Electronics*, 25(5):1–8.

[Qu et al., 2020] Qu, Y., Zhu, H., Shen, Y., Zhang, J., Tao, C., Ghosh, P., and Qiu, M. (2020). Inverse design of an integrated-nanophotonics optical neural network. *Science Bulletin*, 65(14):1177–1183.

[Quack et al., 2023] Quack, N., Takabayashi, A. Y., Sattari, H., Edinger, P., Jo, G., Bleiker, S. J., Errando-Herranz, C., Gylfason, K. B., Niklaus, F., Khan, U., et al. (2023). Integrated silicon photonic mems. *Microsystems & Nanoengineering*, 9(1):27.

[Reck et al., 1994] Reck, M., Zeilinger, A., Bernstein, H. J., and Bertani, P. (1994). Experimental realization of any discrete unitary operator. *Physical review letters*, 73(1):58.

[Sanchis et al., 2009] Sanchis, P., Villalba, P., Cuesta, F., Håkansson, A., Griol, A., Galán, J. V., Brimont, A., and Martí, J. (2009). Highly efficient crossing structure for silicon-on-insulator waveguides. *Optics letters*, 34(18):2760–2762.

[Shastri et al., 2021] Shastri, B. J., Tait, A. N., Ferreira de Lima, T., Pernice, W. H., Bhaskaran, H., Wright, C. D., and Prucnal, P. R. (2021). Photonics for artificial intelligence and neuromorphic computing. *Nature Photonics*, 15(2):102–114.

[Shen et al., 2015] Shen, B., Wang, P., Polson, R., and Menon, R. (2015). An integrated-nanophotonics polarization beamsplitter with 2.4× 2.4 μm2 footprint. *Nature Photonics*, 9(6):378–382.

[Shen et al., 2017] Shen, Y., Harris, N. C., Skirlo, S., Prabhu, M., Baehr-Jones, T., Hochberg, M., Sun, X., Zhao, S., Larochelle, H., Englund, D., et al. (2017). Deep learning with coherent nanophotonic circuits. *Nature photonics*, 11(7):441–446.

[So et al., 2020] So, S., Badloe, T., Noh, J., Bravo-Abad, J., and Rho, J. (2020).

Deep learning enabled inverse design in nanophotonics. *Nanophotonics*, 9(5):1041–1057.

[Sutskever et al., 2013] Sutskever, I., Martens, J., Dahl, G., and Hinton, G. (2013). On the importance of initialization and momentum in deep learning. In *Proceedings of the 30th International Conference on Machine Learning*, number 3 in Proceedings of Machine Learning Research, pages 1139–1147. PMLR.

[Tahersima et al., 2019] Tahersima, M. H., Kojima, K., Koike-Akino, T., Jha, D., Wang, B., Lin, C., and Parsons, K. (2019). Deep neural network inverse design of integrated photonic power splitters. *Scientific reports*, 9(1):1–9.

[Vit et al., 2022] Vit, A. D., Gorgulu, K., Amiri, A. N., and Magden, E. S. (2022). Silicon photonics toolkit, https://github.com/photonic-architecture-laboratories/si-photonics-toolkit.

[Wang et al., 2016] Wang, Y., Gao, S., Wang, K., and Skafidas, E. (2016). Ultra-broadband and low-loss 3-db optical power splitter based on adiabatic tapered silicon waveguides. *Opt. Lett.*, 41(9):2053–2056.

[Xu et al., 2022] Xu, X., Ren, G., Feleppa, T., Liu, X., Boes, A., Mitchell, A., and Lowery, A. J. (2022). Self-calibrating programmable photonic integrated circuits. *Nature Photonics*, 16(8):595–602.

[Xu et al., 2019] Xu, X.-B., Guo, X., Chen, W., Tang, H. X., Dong, C.-H., Guo, G.-C., and Zou, C.-L. (2019). Flat-top optical filter via the adiabatic evolution of light in an asymmetric coupler. *Phys. Rev. A*, 100:023809.

[Zhang and Liboiron-Ladouceur, 2021] Zhang, G. and Liboiron-Ladouceur, O. (2021). Scalable and low crosstalk silicon mode exchanger for mode division multiplexing system enabled by inverse design. *IEEE Photonics Journal*, 13(2):1–13.

[Zhang et al., 2022] Zhang, G., Xu, D.-X., Grinberg, Y., and Liboiron-Ladouceur, O. (2022). Experimental demonstration of robust nanophotonic devices opti-

mized by topological inverse design with energy constraint. *Photonics Research*, 10(7):1787–1802.

[Zhang and Yao, 2020] Zhang, W. and Yao, J. (2020). Photonic integrated field-programmable disk array signal processor. *Nature communications*, 11(1):1–9.

[Zhang et al., 2013] Zhang, Y., Yang, S., Lim, A. E.-J., Lo, G.-Q., Galland, C., Baehr-Jones, T., and Hochberg, M. (2013). A compact and low loss y-junction for submicron silicon waveguide. *Optics express*, 21(1):1310–1316.

[Zhou et al., 2023] Zhou, Z., Ou, X., Fang, Y., Alkhazraji, E., Xu, R., Wan, Y., and Bowers, J. E. (2023). Prospects and applications of on-chip lasers. *Elight*, 3(1):1–25.

[Zhuang et al., 2015] Zhuang, L., Roeloffzen, C. G., Hoekman, M., Boller, K.-J., and Lowery, A. J. (2015). Programmable photonic signal processor chip for radiofrequency applications. *Optica*, 2(10):854–859.

[Zou et al., 2014] Zou, J., Yu, Y., Ye, M., Liu, L., Deng, S., Xu, X., and Zhang, X. (2014). Short and efficient mode-size converter designed by segmented-stepwise method. *Optics letters*, 39(21):6273–6276.

Appendix A

METHODS

A.1 Numerical Simulations

The effective indices of silicon strip waveguides in our design were obtained using the Silicon Photonics Toolkit [Vit et al., 2022], an open-source software package specifically developed for the design of integrated photonic structures. This toolkit supports automatic differentiation, making it compatible with our optimization framework. The toolkit provides fast and efficient calculations of waveguide parameters on the 220 nm SOI platform, which is crucial for the rapid and scalable evaluation of our optical transfer functions. For other components in our deep photonic networks, such as directional couplers and waveguide bends, their optical responses were obtained through 3D-FDTD simulations. The simulations were performed with a maximum spatial discretization of 17 nm in all three dimensions to accurately capture the behavior of these components. The resulting optical responses, which include both amplitude and phase information, were linearly interpolated at 1000 wavelengths ranging from 1.2 µm to 1.7 µm. The interpolated responses were implemented as lookup functions that are compatible with automatic differentiation. These functions were utilized during the performance evaluation of the constructed photonic networks, enabling efficient and accurate calculations of the network's optical properties.

A.2 Numerical Optimization Framework

Our deep photonic network optimization framework leverages an open-source, end-to-end deep learning library called Trax [Tra, 2020]. This library provides access to state-of-the-art machine learning software constructs and supports various hardware accelerators, including GPUs and TPUs. By utilizing Trax, we can take advantage

of advanced machine learning techniques and efficient computation on modern hardware. In our framework, each interferometric structure within the photonic network is modeled as a component of a physics-informed artificial neural network. We evaluate the amplitude and phase profiles of the transfer functions between input and output pairs using the automatic differentiation-compatible lookup functions described earlier. This modular architecture allows for flexible connections between interferometric layers, including serial, parallel, or residual connections, enabling the construction of complex network topologies. The trainable parameters of our networks are optimized iteratively using the adaptive moment estimation (Adam) algorithm [Kingma and Ba, 2014]. During the optimization process, the learning rate is progressively reduced from 3×10^{-3} to 10^{-4} to facilitate smooth and efficient convergence. For the design of power splitters and the spectral duplexer, we used batch sizes of 32 and 21, respectively, during the optimization process. All optimization tasks were performed using a single Tesla V100 GPU, which demonstrates the computational efficiency and scalability of our framework.

A.3 Device Fabrication

After the optimization process, the final designs of the photonic devices were converted into mask layouts using capabilities provided by our design framework. We utilized an open-source layout construction software library [Gabrielli, 2020] for this purpose. The mask layouts included the necessary components and structures, and additional grating couplers were added at the inputs and outputs of the network to enable efficient light coupling. Once the mask layouts were finalized, the devices were fabricated using standard 193 nm CMOS photolithography techniques. The fabrication process was carried out on the silicon-on-insulator (SOI) platform with a 220-nm-thick silicon device layer. The fabrication was performed through IMEC's multi-project-wafer (MPW) foundry service, which provides access to a shared wafer for the fabrication of multiple designs from different users. This allows for cost-effective and efficient fabrication of the photonic devices.

A.4 Measurement Setup and Optical Images of Fabricated Chips

As mentioned before, the performance evaluation of devices was conducted using a continuous-wave tunable laser source (Santec TSL-710), an optical power meter (Santec MPM-210), on-chip grating couplers, and a polarization controller. The experimental setup involved the generation of light by the laser, which then propagated through an optical fiber. The light passes through a polarization controller to control the polarization state. The tip of the optical fiber was used to couple the light with the chip through the grating coupler. At the other end of the device, a similar optical fiber was coupled to the end grating coupler and connected to the power meter to measure the transferred power. Figure A.1 provides an overview of the setup and illustrates the two optical fibers coupled to the chip through grating couplers. Additionally, Figure A.2 presents optical images of the fabricated chips. Figure A.2(a) shows several devices on a chip with their input-output connected grating couplers. These devices consist of three layers of MZIs, resulting in a device length of 240 µm. Figure A.2(b) zooms in on a section of a device, revealing the partially observable waveguide tapers and directional couplers. An MZI within a device is shown, with a length of 80 µm, resulting from two pairs of 30 µm long directional couplers and two pairs of 10 µm long waveguide tapers. To further illustrate the fiber coupling, Figure A.3(a) presents a top view of a coupled fiber to the grating coupler, while Figure A.3(b) offers a side view of the coupled fiber. These images provide a visual representation of the experimental setup and the coupling between the optical fibers and the chip.

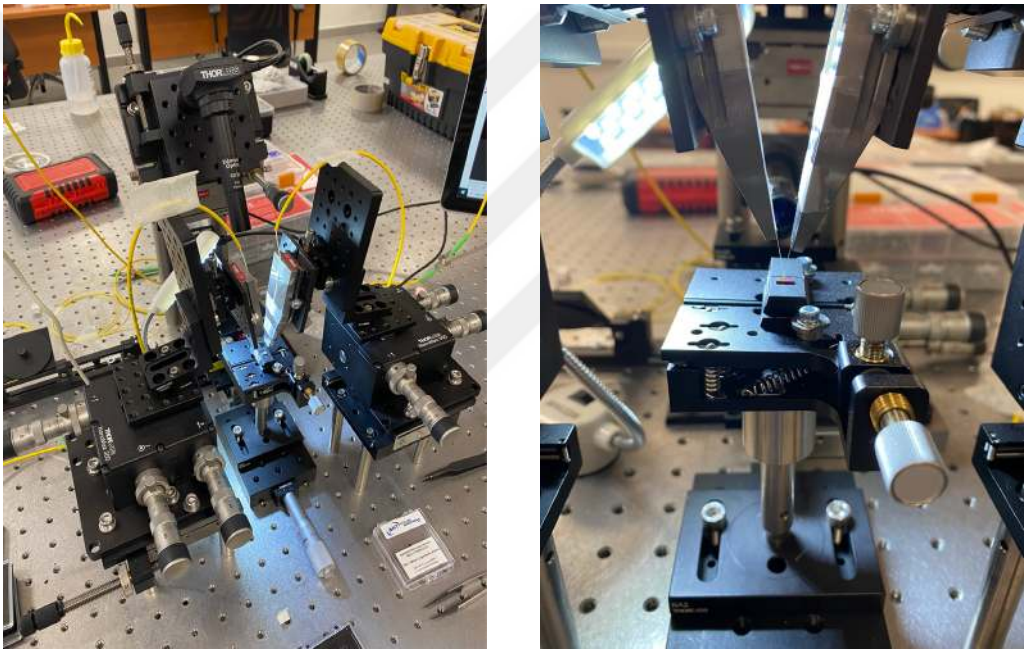

Figure A.1: An overview of the measurement setup

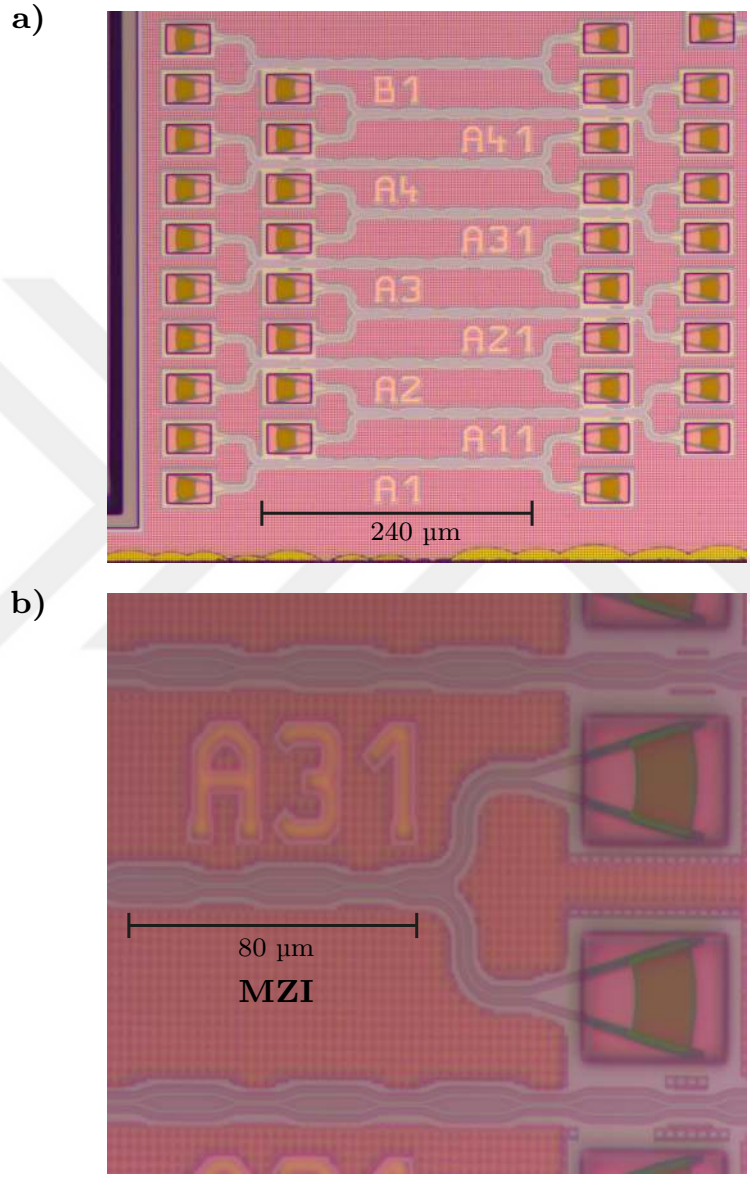

Figure A.2: Optical images of the fabricated chips

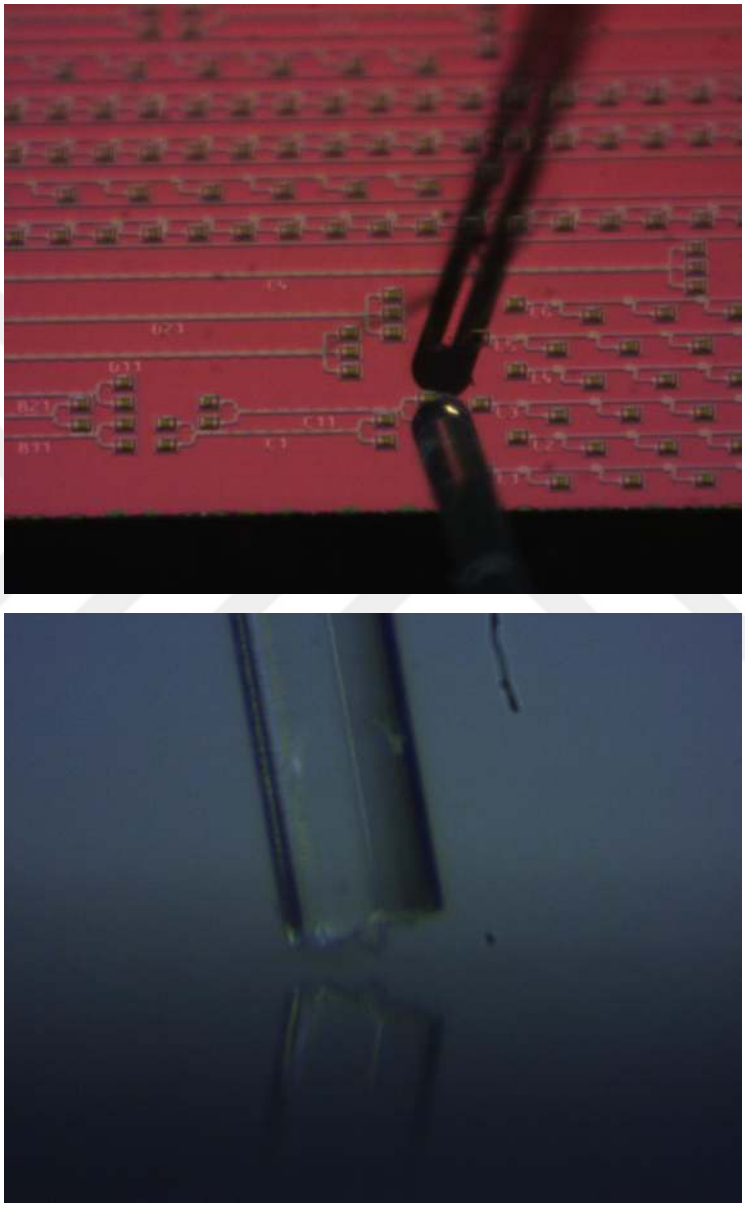

Figure A.3: Top and side view of a coupled fiber to the grating coupler