Ultra-fast Simulation and Design of Nanophotonic Devices

by

Ahmet Onur Daşdemir

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

August 6, 2024

Ultra-fast Simulation and Design of Nanophotonic Devices

Koç University

Graduate School of Sciences and Engineering

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

Ahmet Onur Daşdemir

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:

____

Assist. Prof. Emir Salih Mağden (Advisor)

____

Prof. Dr. Alper Kiraz

____

Assoc. Prof. Muhammed Fatih Toy

Date: ______________________________

To my beloved mom, dad, sister and Hera...

ABSTRACT

Ultra-fast Simulation and Design of Nanophotonic Devices
Ahmet Onur Daşdemir
Master of Science in Electrical and Electronics Engineering
August 6, 2024

This thesis explores advanced methodologies for the ultra-fast simulation and design of nanophotonic devices, emphasizing the efficiency and accuracy of inverse design processes. The study introduces a multi-scale hierarchical inverse design approach that optimizes the fabrication compatibility and performance of nanophotonic components. A significant contribution is the development of a factorization caching method that dramatically reduces computational overhead in the gradient computation phase of inverse design, thereby accelerating the overall optimization process. This method proves particularly effective for two-dimensional finite-difference frequency-domain (FDFD) simulations, offering substantial time savings without compromising physical accuracy.

Furthermore, the integration of deep learning techniques into nanophotonic simulations is examined. A dual-stage model trained with physics-based losses is presented, achieving simulation speeds up to 3408 times faster than traditional three-dimensional finite-difference time-domain (FDTD) methods. This model ensures high fidelity in the simulation results, maintaining robustness across various device geometries and sizes. The advancements detailed in this thesis not only enhance the computational efficiency of photonic device design but also expand the feasible design space, paving the way for the development of innovative and practical nanophotonic devices.

ÖZETÇE

Nanofotonik Cihazların Ultra Hızlı Simülasyonu ve Tasarımı
Ahmet Onur Daşdemir
Elektrik ve Elektronik Mühendisliği, Yüksek Lisans
Ağustos 2024

Bu tez, nanofotonik cihazların ultra hızlı simülasyonu ve tasarımı için gelişmiş metodolojileri incelemekte ve ters tasarım süreçlerinin verimliliği ve doğruluğuna vurgu yapmaktadır. Çalışma, nanofotonik bileşenlerin üretim uyumluluğunu ve performansını optimize eden çok ölçekli hiyerarşik bir ters tasarım yaklaşımı sunmaktadır. Bu tezde önemli bir katkı, ters tasarımın gradyan hesaplama aşamasındaki hesaplama yükünü önemli ölçüde azaltan ve genel optimizasyon sürecini hızlandıran faktorizasyon önbellekleme yönteminin geliştirilmesidir. Bu yöntem, fiziksel doğruluğu koruyarak iki boyutlu FDFD simülasyonları için özellikle etkili olup, önemli zaman tasarrufu sağlamaktadır.

Ayrıca, nanofotonik simülasyonlarda derin öğrenme tekniklerinin entegrasyonu incelenmektedir. Fizik temelli hedef fonksiyonlarıyla eğitilmiş iki aşamalı bir model sunulmakta olup, bu model geleneksel üç boyutlu FDTD yöntemlerinden 3408 kat daha hızlı simülasyon hızlarına ulaşmaktadır. Bu model, çeşitli cihaz geometrileri ve boyutları arasında yüksek güvenilirlikte simülasyon sonuçları sağlamakta ve sağlamlığını korumaktadır. Bu tezde detaylandırılan yöntemler, fotonik cihaz tasarımının hesaplama verimliliğini artırmakla kalmayıp, aynı zamanda yenilikçi ve pratik nanofotonik cihazların geliştirilmesine olanak tanıyan uygun tasarım alanını genişletmektedir.

ACKNOWLEDGMENTS

I would like to thank my advisor, Asst. Prof. Emir Salih Mağden, for his guidance and support. Over the years, starting from my undergraduate studies, I have gained valuable knowledge in silicon photonics, computational electromagnetics, and artificial intelligence. Moreover, from Professor Mağden, I have acquired valuable skills that will benefit me throughout my life, such as how to conduct research, approach problems, and express ideas. I am very grateful to him for providing us with a healthy environment for discussion and opportunities for continuous learning.

I also would like to thank my thesis defense jury members Prof. Alper Kiraz and Assoc. Prof. Muhammed Fatih Toy.

I acknowledge the financial support from the BIDEB 2210-A program of the Scientific and Technological Research Council of Turkey (TUBITAK).

I would also like to express my gratitude to my lab mates, especially Ayşemine, Ali, and Aycan for their exceptional support and friendship. The shared experiences, collaborative efforts, and memorable moments we enjoyed together made the lab a space with truly enjoyable and rewarding experiences. I would also like to thank my dear friend İlayda. Our friendship, which began during our undergraduate years and continued as we supported each other in neighboring laboratories, has greatly enriched my life. The many things we have shared, both in life and academia, have had a significant impact on me.

I extend my heartfelt appreciation to my family for their love, encouragement, and patience throughout this journey. Their unwavering belief in me has been my greatest motivation, and I am deeply thankful for their constant support.

Finally, I would like to thank Melike for her patience and unwavering support. Throughout this challenging process, her endless encouragement and the enjoyable moments we shared have always been a motivating factor for me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

2D	Two Dimensional
3D	Three Dimensional
CMOS	Complementary Metal-Oxide-Semiconductor
FDFD	Finite Difference Frequency Domain
FDTD	Finite Difference Time Domain
FFT	Fast Fourier Transform
PSD	Power Spectral Density
Si	Silicon
SiO_2	Silicon Dioxide

Chapter 1

INTRODUCTION

1.1 Silicon Photonics

Photonic systems find wide applications across many sectors, ranging from communication and healthcare to autonomous vehicles and next-generation technologies such as quantum communication. These photonic systems, where data is generated, processed, and stored through optical signals, hold advantages over their electronic counterparts regarding power consumption, superior noise performance, and high energy efficiency [Xiang et al., 2021]. Over the past 50 years, the use of silicon-based "complementary metal-oxide-semiconductor" (CMOS) technology, which has become the standard for electronic circuit production, has enabled the integrated production of many optical systems. In a way, a revolution similar to the one experienced in integrated electronic systems such as transistors, memory, and processors after 1950 has begun to be seen in photonic systems thanks to scalable production. This technology, known as silicon photonics, allows integrated photonic systems produced at scale to perform the functions that were carried out by optical equipment spanning meters on laboratory tables just ten years ago, now at a much higher level with optical chips that fit on a fingertip and can be mass-produced in the same production line. The optical confinement provided by the high index contrast of core and cladding material ($n_{Si} = 3.45$, $n_{SiO_2} = 1.45$) enables the silicon photonic devices to be densely integrated into smaller areas with CMOS technology [Shi et al., 2022]. Another advantage of silicon is its low cost due to its abundant availability on Earth. This abundance allows for the fabrication of millions of silicon chips at a low cost [Jalali and Fathpour, 2006]. In addition to the manufacturing-related benefits, silicon photonics devices support high data transmission rates and high bandwidth which is a significant feature for telecommunication applications. This advantage has led to the development of modern products like QSFP modules [Mazzini et al., 2015], which are essential to the optical communication standard with speeds reaching up to 400 Gigabits per second [Talebzadeh et al., 2017]. These devices form the backbone of all internet communication infrastructure in data centers worldwide.

1.2 Computer Aided Photonic Design

In these new generation silicon photonics applications, the need for integrated photonic systems used for communication, sensing, and imaging to precisely obtain, access, and process data is continuously increasing. As a result, the bandwidths supported by integrated photonic devices, the types of optical modes and polarizations they can operate with, the ability to control the phase of light, the data processing speeds they can provide, and their energy efficiencies need to be continuously improved to meet current demands. The increasing number and complexity of optical operations required in communications, sensing, and computing applications continue to drive the development of compact and efficient integrated photonic components [Yang et al., 2022, Luan et al., 2018, Vandoorne et al., 2014]. Demonstration of these high-performance on-chip devices with multi-wavelength and multi-mode capabilities has significant implications for the throughput and scalability of next-generation optical systems [Piggott et al., 2015, Lu and Vučković, 2012, Magden et al., 2018, Singh et al., 2020].

As the demand for such high-performance, multi-function, and compact on-chip devices continues to grow, traditional design approaches that rely on physical intuition and trial-and-error have become insufficient in meeting these stringent requirements. Consequently, automated design algorithms have recently attracted significant attention and are rapidly proving to be indispensable tools for photonic design. To this end, inverse design techniques have emerged as a powerful approach to design photonic devices with unprecedented performance and functionality by exploring a large set of design variables [Piggott et al., 2015, Molesky et al., 2018, Hughes et al., 2018a, Lalau-Keraly et al., 2013, Lu and Vučković, 2012]. Previous research on inverse design has employed two main techniques to achieve optimal device structures: shape optimization [Lalau-Keraly et al., 2013, Michaels and Yablonovitch, 2018, Michaels et al., 2019, Piggott et al., 2017], which involves modifying geometrical boundaries within a specified design region, and topology optimization [Piggott et al., 2015, Hughes et al., 2018a, Elesin et al., 2014, Hammond et al., 2022, Zhang et al., 2021], which updates the device in a pixel-by-pixel manner. In both of these methods, the device structure is modified iteratively by optimizing an objective function that is calculated from an electromagnetic simulator. The general design procedure of the topological photonic inverse design is detailed in Figure 1.1.

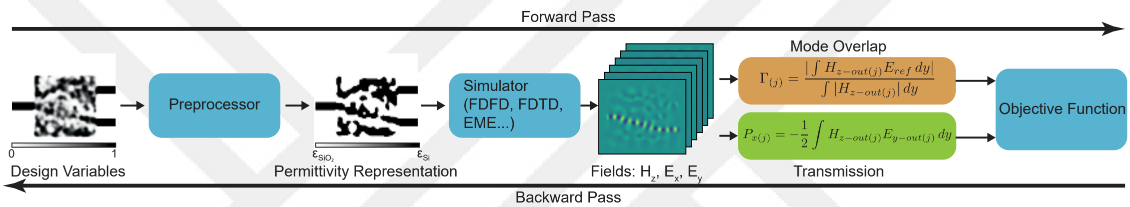


Figure 1.1: Photonic inverse design scheme. The displayed fields correspond to the TE polarization in 2D simulations, where only nonzero fields are z-component of magnetic field (H_z) and x-y components of electric field (E_x, E_y).

First, a geometry is initialized with pixel values between 0 and 1 that correspond to the design variables. When converted to the permittivity representation, 1 corresponds to the permittivity of the core material while 0 represents the cladding material. During the optimization, these design variables are allowed to vary between the permittivity values of core and cladding materials. In order to meet the fabrication compatibility requirements, a preprocessing step is applied before the simulation. The design variables between 0 and 1 are first applied a smoothing filter and then a threshold. In this way, the small structures are eliminated and the geometry is converged to a binary state. The device is then simulated and the corresponding electric field distribution is obtained. This electric field distribution is then used to compute mode overlap and transmission metrics. The objective function to be optimized is formed by mode overlap and transmission. Mode overlap quantifies the similarity between the spatial profile of the electric field in the output waveguide and the desired mode profile. The formulation of the mode overlap metric is given in Eq. 1.1 where (j) corresponds to the index of the output waveguide.

$$\Gamma_{(j)} = \left| \frac{\int H_{Z-\mathrm{out}(j)}(y) E_{\mathrm{ref}}(y)\, dy}{\int |H_{Z-\mathrm{out}(j)}(y)|^2\, dy} \right|^2 \tag{1.1}$$

Transmission metric measures the ratio of power in the output waveguide compared to the input power by relying on the Poynting vector calculation shown in Eq. 1.2.

$$P_{x(j)} = -\frac{1}{2} \int \Re\{E_{y-\mathrm{out}(j)}(y) H_{Z-\mathrm{out}(j)}(y)\}\, dy \tag{1.2}$$

The mode overlap is expected to converge to 1 which indicates all of the power is confined in the desired output mode. By also comparing the transmission metric with the desired target value, the objective function is calculated. Then, the gradient of this objective with respect to the design variables is obtained and then used to gradually modify the photonic device through gradient-based optimization methods.

1.3 Organization

The organization of this thesis is as follows: In the first part, we demonstrate a multi-scale hierarchical topology optimization approach to address the computational time of photonic inverse design, showcasing rapid convergence through a refined design process [Dasdemir et al., 2020]. In the second part, we introduce a multi-faceted factorization caching method that enhances the computational efficiency of electromagnetic simulations by leveraging the symbolic and numerical structure of the system matrix [Dasdemir et al., 2023]. In the third part, we present a dual-stage model trained with physics-based losses, achieving highly accurate and rapid simulation outcomes, significantly faster than the traditional 3D FDTD method [Dasdemir et al., 2024].

Chapter 2

MULTI-SCALE HIERARCHICAL INVERSE DESIGN

2.1 Fabrication Compatibility Concerns in Inverse Design

Inverse design has revolutionized the creation of photonic devices, enabling the realization of non-intuitive designs that were previously unattainable using traditional methods [Lalau-Keraly et al., 2013, Piggott et al., 2015, Molesky et al., 2018, Hughes et al., 2018a]. Traditional design approaches in optics often involve manual, iterative processes, which can be time-consuming and limited in their capacity to explore design spaces with many degrees of freedom. In contrast, photonic inverse design harnesses algorithmic approaches and machine learning techniques to optimize optical devices according to specific desired functionalities.

The details of inverse design scheme are given in Section 1.2. It was mentioned that the design variables are allowed to have values between core and cladding permittivity. For instance, the geometry in Figure 1.1 has pixel values between ϵ_{Si} and ϵ_{SiO_2}. However, final designs that will be fabricated are required to have binary structures with only SiO_2 or Si permittivity for any pixel. Moreover, the geometry cannot include small features that violate the minimum feature size constraint of state-of-the-art fabrication methods. In order to meet these fabrication compatibility requirements, some image processing operations such as smoothing filter and projection are applied which eliminates the small structures and binarizes the geometry. An example of the projection operation with a sigmoid function is given in Eq. 2.1. Here, $\bar{\alpha}$ is the map of the filtered design variables $\tilde{\alpha}$ after the projection operation. In addition, β represents the strength of the projection by adjusting the slope of the sigmoid and η determines the threshold value.

$$\bar{\alpha} = \frac{\tanh(\beta\eta) + \tanh(\beta(\tilde{\alpha} - \eta))}{\tanh(\beta\eta) + \tanh(\beta(1 - \eta))} \tag{2.1}$$

After the processing operations to meet the fabrication compatibility requirements, an electromagnetic simulation of the structure is run to compute the corresponding electric and magnetic fields. Then, performance metrics such as mode overlap and transmission are calculated to find the objective value by comparing the computed performance metrics with the targets. Here, transmission corresponds to the ratio of the output and input power while mode overlap measures the amount of power confined in the desired optical mode. In the final step, an optimization algorithm uses the gradient of the objective function w.r.t the design variables to modify the geometry.

2.2 Multi-scale Hierarchical Design Method

Among the optimization methods used in inverse design, iterative quasi-Newton minimization with the adjoint method has become popular, as the adjoint method yields the objective function gradient with respect to all design variables by using only two propagation simulations [Hughes et al., 2018a, Lalau-Keraly et al., 2013, Michaels and Yablonovitch, 2018, Piggott et al., 2015]. Despite reduced convergence time in comparison to heuristic techniques or finite difference-based gradient calculation, the computational speeds are still limited by the sparse linear system solutions in FDFD simulations. As these solutions range from $\mathcal{O}(n^{1.5})$ to $\mathcal{O}(n^2)$ in complexity [Schenk and Gärtner, 2004], reducing the system size n by using coarser grids greatly improves simulation speeds.

Conversely, simulation accuracies suffer with significantly coarse simulation grids, potentially preventing convergence at high efficiencies. Moreover, binary and smooth material boundaries necessary for practical fabrication are challenging to represent on such coarse grids and typically require separate parametrizations [Vercruysse et al., 2019] or averaging techniques [Michaels and Yablonovitch, 2018]. However, while a binary geometry and a fine grid are necessary for the final device, initial optimization iterations can be performed with coarser grids and continuous permittivity distributions. This idea has been investigated for shape optimization, but only with semi-manually chosen design variables [Michaels et al., 2020]. Here we demonstrate a generalized and fully-autonomous hierarchical design approach for multi-scale topology optimization by rapidly performing initial iterations to approach an approximate solution, and gradually oversample this topology to obtain the final device. We design a mode multiplexer using our approach and show record-fast convergence under 50 seconds with a transmission efficiency of over 98.1%.

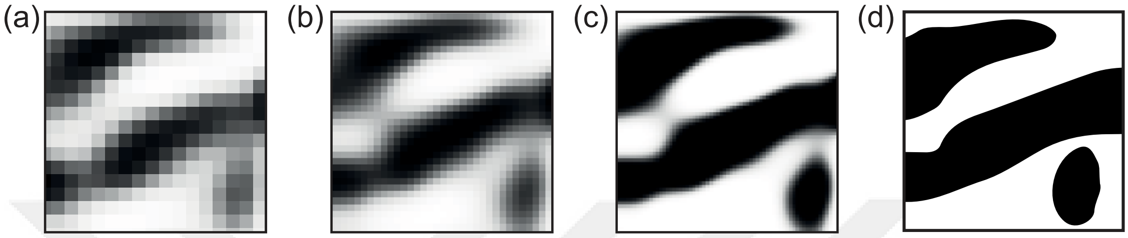


Figure 2.1: Multi-scale hierarchical topology optimization concept throughout iterations with (a) 90 nm, (b) 45 nm, (c) 22.5 nm grids, and (d) final device with a smooth boundary contour.

Our optimization process starts with a continuous structure on a coarse grid. This structure is then iteratively refined, oversampled by factors of 2 throughout multiple stages of optimization, and gradually binarized as illustrated in Figure 2.1. In each iteration, FDFD simulations are used to calculate the objective function and its gradient. Then, a 2D Gaussian low-pass filter and a sigmoid-like projection function [Hughes et al., 2018b] are utilized for gradual binarization of the resulting permittivity distribution. The Gaussian filter radius is specified according to the minimum feature size desired; and the binarization strength is controlled with the slope of the projection function, which is gradually incremented throughout the optimization. The combination of the smoothing filter and the binarization projection allows for rapid convergence while ensuring fabrication compatibility with smooth device boundaries and sufficiently large features.

2.3 *Mode Multiplexer Design with Multi-scale Optimization*

Using our multi-scale optimization method, we designed a 1×2 spatial mode multiplexer at $\lambda = 1.55\,\mu\text{m}$, a common building block in silicon photonics and a popular choice for inverse design algorithms. Here we restrict the device to a 2-dimensional

3 µm × 3 µm geometry, and only use TM-polarized inputs and outputs (Figure 2.2a). We set the $1/e^2$ radius for the Gaussian filter at $\sim$ 200 nm for compatibility with 80 nm allowed minimum features in modern e-beam lithography. The final device shown in Figure 2.2b has a minimum feature size of 129.5 nm and smooth material boundaries as expected. The resulting electric field intensities are plotted in Figure 2.2c and Figure 2.2d for the two input modes, where each polarization is mapped to the desired output mode.

For this multiplexer design, the convergence of our multi-scale hierarchical optimization method is detailed in Figure 2.2e where the simulation was performed using 16 cores on an Intel Xeon CPU. The first phase of optimization (shaded blue) is performed on a $\lambda/5 \approx 90$ nm grid and with a continuous permittivity distribution to quickly find an approximate initial guess for the next optimization stages. This initial phase completes extremely quickly due to the relatively small number of FDFD cells. The drop in efficiency at iteration 37 results from an increase in the projection slope which is repeated every time transmission efficiency exceeds 98%. The amount of change in this slope is determined by how continuous the current device is when this 98% efficiency is reached. The optimization algorithm then oversamples the device topology to a $\lambda/10 \approx 45$ nm grid (shaded orange) after the permittivity distribution exceeds a pre-determined "discreteness" threshold, $\Gamma = 0.5$, calculated as a statistical parameter from the permittivity values. The drops in efficiency continue as binarization strength increases, but are easily recovered within a few iterations. The iterative oversampling continues until desired convergence is reached at a sufficiently fine grid ($\lambda/20$), or when a smooth boundary contour can be drawn with negligible changes in efficiency when simulated on the same fine grid. Here, the multiplexer achieves 98.1% (-0.083 dB) efficiency and a final discreteness of $\Gamma \approx 0.997$ within 104 iterations in a total of 49.7 seconds, without needing further optimization at a $\lambda/20$ grid. For comparison, a fully binary and single-scale optimization (dashed gray) with 45 nm grid only converges to a much worse solution at 85% efficiency while taking over 50% longer than our multi-scale method. These results demonstrate the potential for multi-scale hierarchical topology optimization in rapid design of efficient devices for widespread use in nanophotonics.

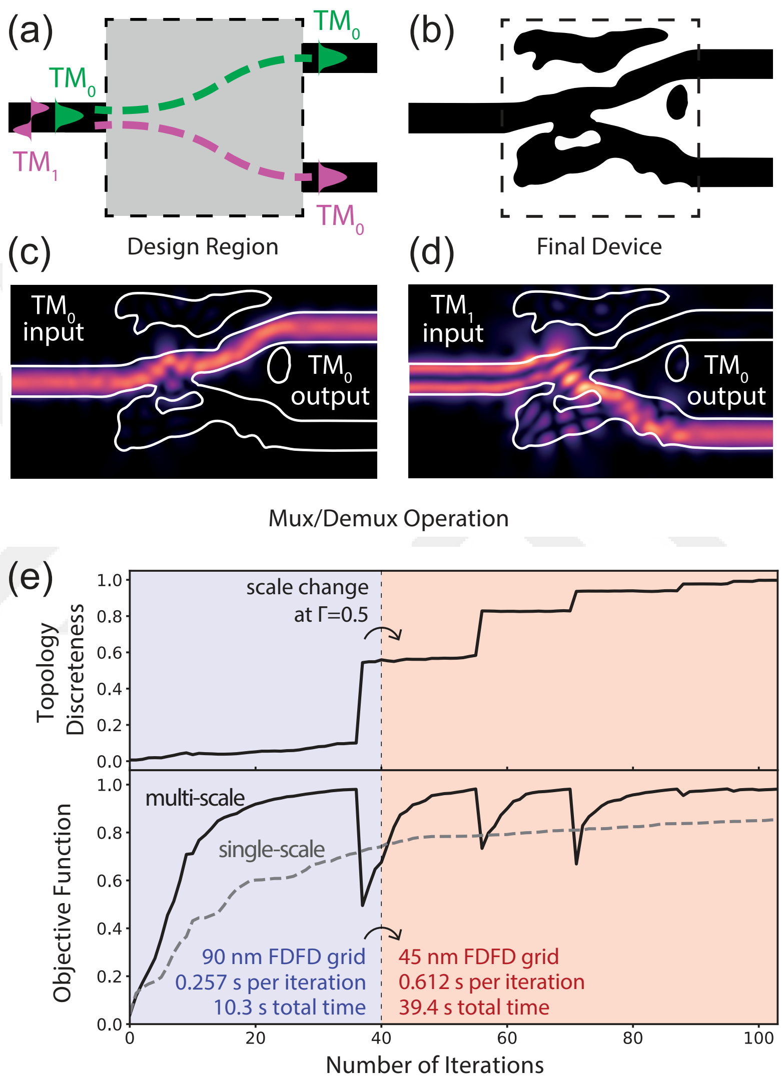


Figure 2.2: (a) Desired device functionality. (b) Device design after optimization. (c) and (d) —E_z— showing mux/demux operation. (e) Convergence of objective function and topology discreteness Γ.

Chapter 3

COMPUTATIONAL SCALING IN INVERSE PHOTONIC DESIGN THROUGH FACTORIZATION CACHING

3.1 Factorization Caching Method

In the inverse design procedure given in Section 1.2 and Figure 1.1, the gradient computation has been shown as a necessary step to gradually modify the design variables towards a device that will perform the desired functionality. A finite difference method for computing the gradient involves perturbing each design variable and running a simulation to observe its impact on the objective function. Since this approach necessitates thousands of simulations per optimization iteration, scaling linearly with the number of design variables, the total design time is significantly constrained by the lengthy computation time required for gradients. To this end, the adjoint method has been demonstrated as a powerful approach to compute the gradient of the objective function with respect to design variables by running only two simulations [Hughes et al., 2018a, Lalau-Keraly et al., 2013, Michaels and Yablonovitch, 2018, Piggott et al., 2015].

Although only two simulations are required for each iteration of optimization, these "forward" and "adjoint" simulations in photonic design are known to be computationally intensive, making up the vast majority of the total optimization time [Hammond and Camacho, 2019]. To address this issue, prior studies accelerated simulations by reducing the system matrix size in FDFD via Schur complement decomposition [Zhao et al., 2019]. Even though this method works well when the optimizable design region constitutes a relatively small portion of the simulation domain; it is less effective for devices with larger footprints. Moreover, the initial system solution overhead can be computationally prohibitive, limiting this method's applicability to only a subset of photonic design problems. On the other hand, deep learning-enabled approaches replacing Maxwell's solvers have also been proposed [Ma et al., 2021, Chen et al., 2022, Liu et al., 2018, Liu et al., 2021]. However, despite their computational efficiency, these methods can suffer from the lack of physical accuracy due to the loss of generalization when designing devices significantly different from those used in their training datasets [Wiecha et al., 2021]. Moreover, the use of such deep learning models poses inherent challenges in photonic design due to multiple distinct geometries potentially yielding the same optical response, a challenge known as the degeneracy problem [Liu et al., 2021]. Hence, addressing these computational requirements for accurate device simulations remains a critical challenge, as larger device footprints and more iterations for convergence collectively increase total optimization runtimes significantly, hindering the design of large-scale and multi-purpose photonic devices for next-generation photonic systems. Therefore, significant progress in computational capabilities is necessary to enable and streamline physically accurate and data-independent procedures for the design of large-scale integrated optical devices with previously-elusive, and multi-purpose functionalities.

Here we demonstrate a set of factorization caching approaches that drastically improve the computational efficiency of FDFD simulations in photonic inverse design. Specifically, we utilize the patterns in discretized Maxwell's equations, cache factorizations of the resulting system matrices, and re-use them throughout the design process. With this method, we demonstrate a substantial speedup in FDFD simulations, enabling faster and more efficient optimizations for large-scale and multi-purpose photonic device design. As proof-of-concept devices, we design a wavelength duplexer and a mode multiplexer using our caching methods and demonstrate enhancements of up to 9.2-fold in simulation speeds over conventional FDFD techniques. Our approach demonstrates how computational enhancements can be leveraged to accelerate physical simulations and enable exciting future possibilities for large-scale on-chip photonic design.

We first begin with a brief overview of adjoint optimization and specifically its use in nanophotonics. The goal of inverse photonic design is to create an ideal device geometry by optimizing an objective function $F(\mathbf{E})$ using the electric field $\mathbf{E}$

computed from Maxwell's equation

$$\left(-\omega^2\varepsilon_0\varepsilon_r(\mathbf{r}) + {\mu_0}^{-1}\nabla\times\nabla\times\right)\mathbf{E}(\mathbf{r}) = -i\omega\mathbf{J}(\mathbf{r}) \tag{3.1}$$

where $\varepsilon_r(\mathbf{r})$ is the relative permittivity distribution representing the design variables θ_i, and $\mathbf{J}(\mathbf{r})$ represents the input excitation to the electromagnetic simulation, such as a dipole, waveguide mode, Gaussian beam, or a plane wave. FDFD simulations solve a discretized version of Eq. 3.1 as a linear system of equations in the form of

$$A\mathbf{E} = -i\omega\mathbf{J} \tag{3.2}$$

where A is the system matrix including terms with permittivity distribution and spatial derivatives. Once the linear system in Eq. 3.2 is solved and the design performance is calculated by the objective function $F(\mathbf{E})$, the gradient of this objective with respect to the design variables θ_i is computed using the adjoint system [Hughes et al., 2018a, Lalau-Keraly et al., 2013, Michaels and Yablonovitch, 2018]

$$\frac{\partial F}{\partial \theta_i} = -2\mathrm{Re}\{{\mathbf{E_{adj}}}^T\frac{\partial A}{\partial \theta_i}\mathbf{E}\} \tag{3.3}$$

and $\mathbf{E_{adj}}$ is obtained by solving the linear set of equations given by

$$A^T\mathbf{E_{adj}} = \left(\frac{\partial F}{\partial \mathbf{E}}\right)^T \tag{3.4}$$

The solution of Eq. 3.2 and Eq. 3.4 is performed through a sparse linear system solver. Specifically, we use a direct solution method called LU factorization[Schenk et al., 2000] to decompose A into the product of lower and upper triangular matrices such that $A = LU$. After this factorization, the linear system is solved with a back-substitution where equations in the rows of LU representation are solved in consecutive steps. The solutions can be formulated as

$$\mathbf{E} = -i\omega U^{-1}L^{-1}\mathbf{J} \tag{3.5}$$

and

$$\mathbf{E_{adj}} = (L^T)^{-1}(U^T)^{-1}\left(\frac{\partial F}{\partial \mathbf{E}}\right)^T \tag{3.6}$$

Considering the structure of the individual terms in Eq. 3.1, we note that A is a sparse matrix, and its LU factorization introduces fill-ins (new nonzero entries) in the L and U factors. Symbolic factorization is the process in which the locations of these fill-ins and their analytical expressions are determined, according to the sparsity pattern of A. Following symbolic factorization, the numerical values of these entries are calculated via a subsequent numerical factorization [Davis et al., 2016]. Once these factorizations have been completed, the final solution to the sparse linear system is obtained through back substitution as shown in Eq. 3.5.

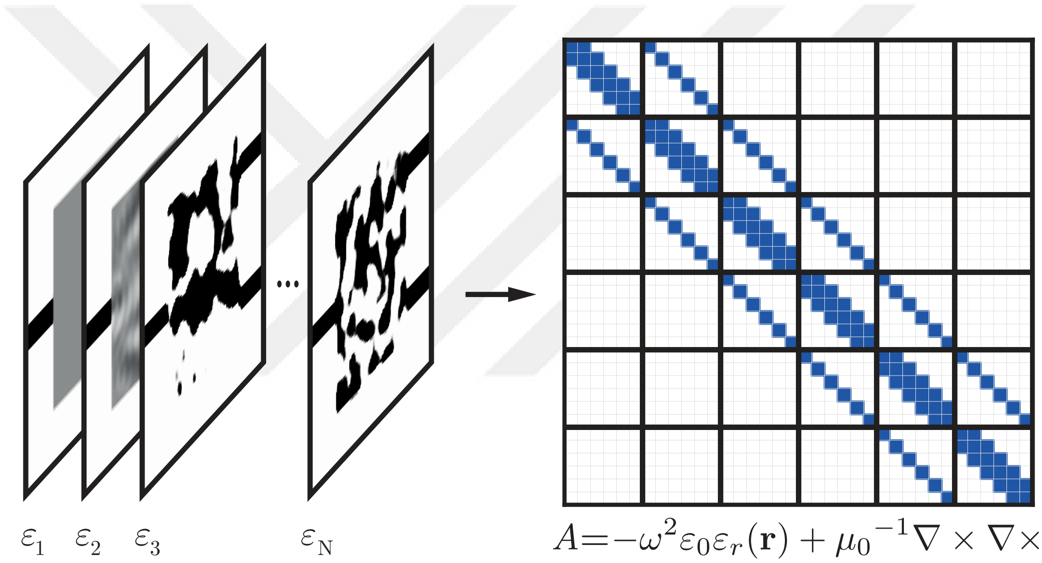


Figure 3.1: Sparsity pattern of the system matrix A is independent of the spatial distribution of materials, and remains the same throughout the entire design procedure.

Our factorization caching approach is motivated by the structure of and the updates on the system matrix A, and its corresponding L and U factors. First of all, during a single iteration of optimization, numerical factors of A^T, namely U^T and L^T are required to solve the adjoint system in Eq. 3.4. Fortunately, once L and U are numerically calculated, L^T and U^T are already known, making the solution of Eq. 3.6 trivial through back substitution. As such, no separate numerical factorizations are necessary for the adjoint system solution in the same iteration. This is extremely advantageous since solving a system through back substitution can be over 10 times faster than re-factoring A^T from scratch [Zhao et al., 2022]. Secondly, as depicted in Fig. 3.1, the sparsity pattern of A is dictated by the spatial derivative terms in Eq. 3.1. Consequently, as the device geometry is iteratively modified during optimization, while A is updated numerically, its sparsity pattern remains the same. Therefore, even though consecutive system solutions need to be performed using numerically different A matrices, they all share the same symbolic factorization which needs to be computed only once at the beginning. This symbolic factorization can then be reused throughout the rest of the optimization process. Moreover, as wavelength dependence in permittivity $\varepsilon_r(\mathbf{r})$ and in ω only alter A numerically, this stored symbolic factorization can be reused for FDFD simulations at different wavelengths, which is especially beneficial for designing multi-wavelength photonic devices. Finally, we note that these advantages scale well to devices processing multiple inputs such as spatial mode multiplexers. Systems with different right-hand sides (i.e. different $\mathbf{J}$ inputs) can be solved using the existing L and U factors, as the device geometry and the system matrix A remain unchanged across different inputs during a single iteration. The combination of these three separate factorization caching approaches can enable significant computational scaling for designing large-scale and multi-purpose integrated photonic devices.

3.2 Design of Broadband Wavelength Duplexer

Our factorization caching approach is first applied in the design of a broadband silicon photonic wavelength duplexer by using the effective index method [Buus, 1982] and ideal 90-degree sidewall angles with a standard 220-nm-thick silicon-on-insulator platform. The duplexer is designed with a 7 µm×7 µm footprint, to separate the input light into short-pass (1.50-1.55 µm) and long-pass (1.55-1.60 µm) output ports. The objective function $F(\mathbf{E})$ is defined as the difference between the calculated and desired transmissions at 10 target wavelengths and minimized with an L-BFGS-B optimizer [Zhu et al., 1997]. Fig. 3.2(a) shows the final device design, for which a smoothing filter with a 200 nm radius and a sigmoid-like projection was used to binarize the device geometry to $\varepsilon_{\mathrm{Si}}$ and $\varepsilon_{\mathrm{SiO_2}}$ permittivities [Dasdemir et al., 2020].

Output transmissions are plotted in Fig. 3.2(b), demonstrating an insertion loss as low as 0.26dB (94.1% transmission); and the resulting electric field profiles are shown in Fig. 3.2(c-e).

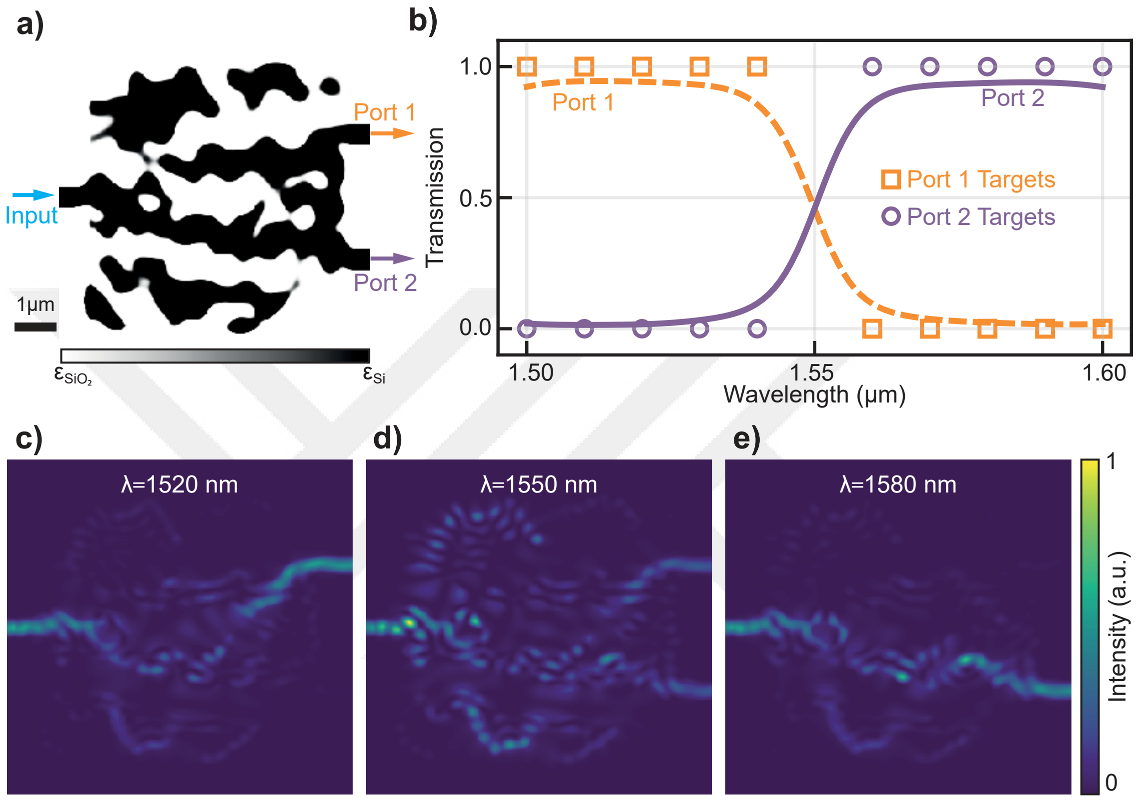
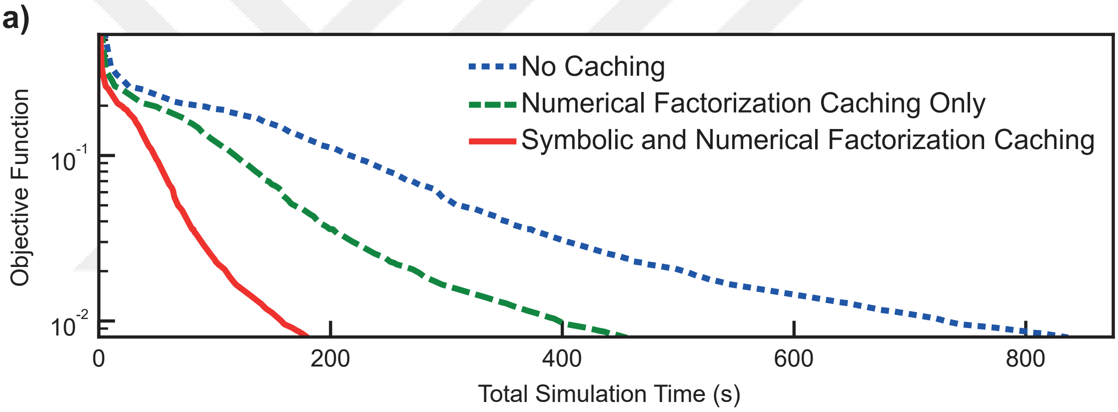


Figure 3.2: **(a)** Wavelength duplexer geometry. **(b)** Transmission targets (colored circles) and the transmission at output ports. **(c-e)** Electric field intensity distribution of device at 1520 nm, 1550 nm and 1580 nm.

In order to demonstrate the utility and the computational advantage of our factorization caching approach, the same device design is repeated three times, with progressive applications of numerical and symbolic factorization caching. All simulations were performed on a PC with a 2.4 GHz Intel Xeon Gold processor with 8 cores, using Intel oneMKL PARDISO package [int,] with a Python interface [Marchant, 2017, Minden, 2022]. In Fig. 3.3(a), we plot the objective function with respect to the total simulation time spent through 164 iterations, for a target objective of 8×10^{-3}. Shown in blue is the standard approach as reference where Eq. 3.2 and Eq. 3.4 are solved from scratch every iteration by repeating symbolic and numerical factorizations. The green curve demonstrates our improved approach where we compute and cache the numerical factorizations of the forward simulations at different wavelengths, and reuse them in the adjoint simulations. Hence, the solution of the adjoint system in Eq. 3.4 is reduced to only a back substitution operation. Finally, the red curve shows optimization performance when both the numerical and symbolic factorizations are cached. Specifically, the additional caching of the symbolic factorization enables its reuse for all subsequent iterations in the optimization process. As a result of these cached factors, while the standard method reaches our reference objective in 833 s of simulation time, the factorization-cached approach achieves the same objective in 179 s, presenting an improvement of 4.7 times.

a)



b)

	Number of Operations						
	Forward			Adjoint			Total Time Required (s)
Simulation Method	SF	NF	BS	SF	NF	BS	0 – 5.09
Standard each iteration	10	10	10	10	10	10	
Factorization-Cached							
iteration 1	1	10	10	0	0	10	
iteration 2	0	10	10	0	0	10	
iteration 3	0	10	10	0	0	10	
⋮							
iteration N	0	10	10	0	0	10	

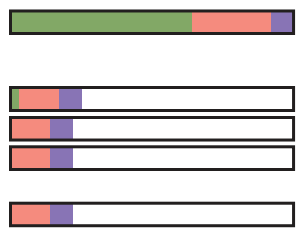


Figure 3.3: **(a)** Objective function evolution of a wavelength duplexer as a function of total simulation time with different factorization caching approaches. **(b)** Numbers of symbolic factorization (SF), numerical factorization (NF), and back substitution (BS) operations in each iteration for the standard state-of-the-art method and our factorization caching method. Bar plots visualize the total simulation and the corresponding SF, NF, and BS runtimes per iteration.

The computational advantage is detailed in Fig. 3.3(b) with a comparison of the number of factorization and back substitution operations for non-cached and cached optimizations. In the optimization process of this device, each iteration involves 20 simulations, with 10 target wavelengths requiring 1 forward and 1 adjoint simulation each. Without caching, this entails performing 20 symbolic factorizations, 20 numerical factorizations, and 20 back substitutions as these factorizations are re-computed in every iteration. In contrast, when factorizations are cached, the symbolic factorization is only computed once during the first iteration. On the other hand, the numerical factorizations are computed anew for each wavelength in the forward simulation but then are cached and reused in the adjoint simulations of the same iteration. Therefore, our caching approach requires only 10 numerical factorizations and 20 back substitutions for all iterations after the first one. As a result, while the standard approach takes 5.09 s of computational time at each one of 164 iterations, our factorization caching approach only takes 1.28 s for the first iteration, and 1.12 s for all following iterations. This enhancement illustrates the significant computational capabilities enabled by factorization caching in the design of multi-wavelength photonic devices.

3.3 Design of Broadband Mode Multiplexer

The computational scaling achieved by factorization caching is even more significant in photonic devices that process multiple optical inputs. To demonstrate this, we design a 1×3 broadband silicon photonic mode multiplexer with a $10\,\mu m \times 10\,\mu m$ footprint as shown in Fig. 3.4(a), using the same 220-nm-thick silicon-on-insulator platform. In this device, a multi-mode optical input (TE_0, TE_1, and TE_2) is demultiplexed into fundamental modes at three outputs, for a broad wavelength range from 1500 nm to 1600 nm. The transmission for all three inputs is plotted in Fig. 3.4(b), demonstrating an insertion loss of at most 0.35dB (92.2% transmission). The corresponding electric fields are plotted in Fig. 3.4(c-e), after an objective of 2.5×10^{-3} is reached in 333 iterations.

We perform a similar analysis on this broadband mode multiplexer by repeating

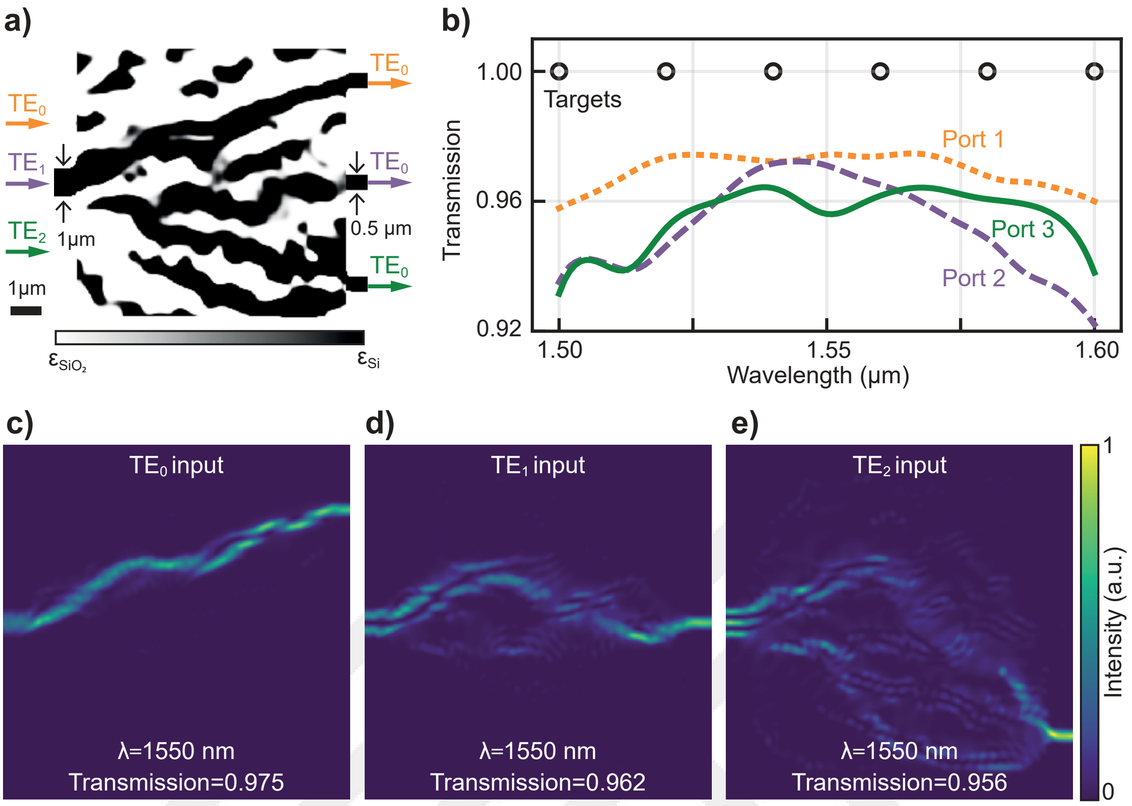
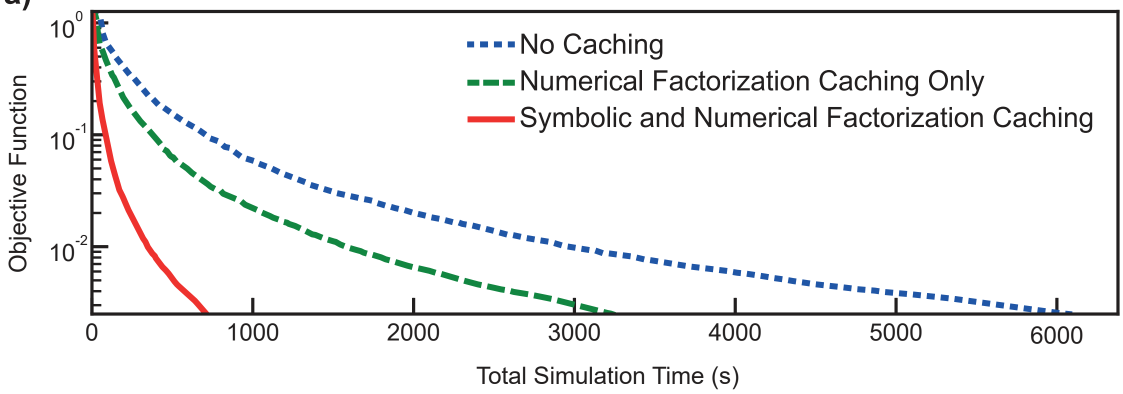


Figure 3.4: **(a)** Mode multiplexer geometry. **(b)** Transmission targets and the resulting output transmission for TE_0 mode. **(c-e)** Electric field intensity distribution of mode multiplexer for each input mode.

the design process with and without factorization caching and plot the resulting objective function in Fig. 3.5(a). Here, while the standard method (blue) reaches our reference objective in 6076 s of simulation time, the same objective is achieved in only 712 s when both the symbolic and numerical factorizations are cached (red). This result represents a computational speedup of over 8.5 times over the standard approach, achieving even better scaling than the spectral duplexer we demonstrated above.

The enhanced computational scaling in this device stems from the further advantages enabled by our cached factorizations for the solutions of linear systems with multiple inputs. These advantages are illustrated by the number of specific operations shown in Fig. 3.5(b) for standard and factorization-cached methods. The standard approach requires $2 \times 6 \times 3 = 36$ system solutions by re-solving Eq. 3.2

a)



b)

Simulation Method	Forward SF	Forward NF	Forward BS	Adjoint SF	Adjoint NF	Adjoint BS	Total Time Required (s) (0 – 18.12)
Standard							
each iteration	18	18	18	18	18	18	
Factorization-Cached							
iteration 1	1	6	18	0	0	18	
iteration 2	0	6	18	0	0	18	
iteration 3	0	6	18	0	0	18	
⋮							
iteration N	0	6	18	0	0	18	

Figure 3.5: **(a)** Objective function evolution of a mode multiplexer as a function of total simulation time with different factorization caching approaches. **(b)** Numbers of symbolic factorization (SF), numerical factorization (NF), and back substitution (BS) operations in each iteration for the standard state-of-the-art method and factorization caching method.

and Eq. 3.4 at each wavelength and input mode, for every iteration. This involves performing 36 symbolic factorizations, 36 numerical factorizations, and 36 back substitutions, totaling 18.12 s per iteration. In contrast, caching these factorizations allows for the stored symbolic factorization to be reused throughout, and also for the stored numerical factorizations to be reused across multiple different inputs at each wavelength. As a result, after the first iteration, only 6 numerical factorizations and 36 back substitutions are required per iteration. Compared to the no-caching approach, our method eliminates over 83% of numerical factorizations and all subsequent symbolic factorizations, resulting in computational times of 2.40 s for the first iteration, and only 2.08 s for each subsequent iteration. This clear computational advantage highlights how cached factorizations significantly enhance the design of multi-input and multi-wavelength optical devices.

We note that for both of the results demonstrated in Fig. 3.3(b) and Fig. 3.5(b), the runtime of symbolic and numerical factorizations may vary based on specific factorization algorithms or the hardware resources used. However, while the computational time inherently depends on these factors, the number of factorizations performed for device optimization remains a fundamental quantity independent of implementation or hardware details. Hence, our demonstrations with significantly fewer symbolic and numerical factorizations, coupled with the accelerated optimization time, underscores the efficiency of our caching method independent of hardware resources and algorithm implementation.

3.4 Computational Analysis of Factorization Caching

We also analyze how these demonstrated computational advantages scale with the device footprint, as the footprint is a crucial parameter influencing the design degrees of freedom and functional capabilities of a photonic device. To do so, we run simulations on devices with footprints ranging from 4 µm × 4 µm to 85 µm × 85 µm, with an additional 1 µm on all four sides for the simulation boundary. Using a 25 nm spatial discretization, these footprints result in system matrices of size from $240^2 \times 240^2$ to $3480^2 \times 3480^2$. In Fig. 3.6(a) and Fig. 3.6(b) we plot the runtime of three subprocedures for a single simulation and the ratio of these runtimes to a single simulation time. From these results, we conclude that while symbolic factorization takes approximately twice as long as numerical factorization with smaller devices (due to smaller system matrices), numerical factorization requires progressively longer times for larger devices. This result reveals that caching the symbolic factorization alone can eliminate over half of the required simulation time, using typical device footprints in inverse photonic design (<100 µm^2) [Piggott et al., 2015, Molesky et al., 2018, Hughes et al., 2018a, Lalau-Keraly et al., 2013]. Furthermore, by caching both factorizations, it is possible to reduce the necessary computational time to the point where back substitution is the only remaining operation, which corresponds to only

4-8% of the simulation time.

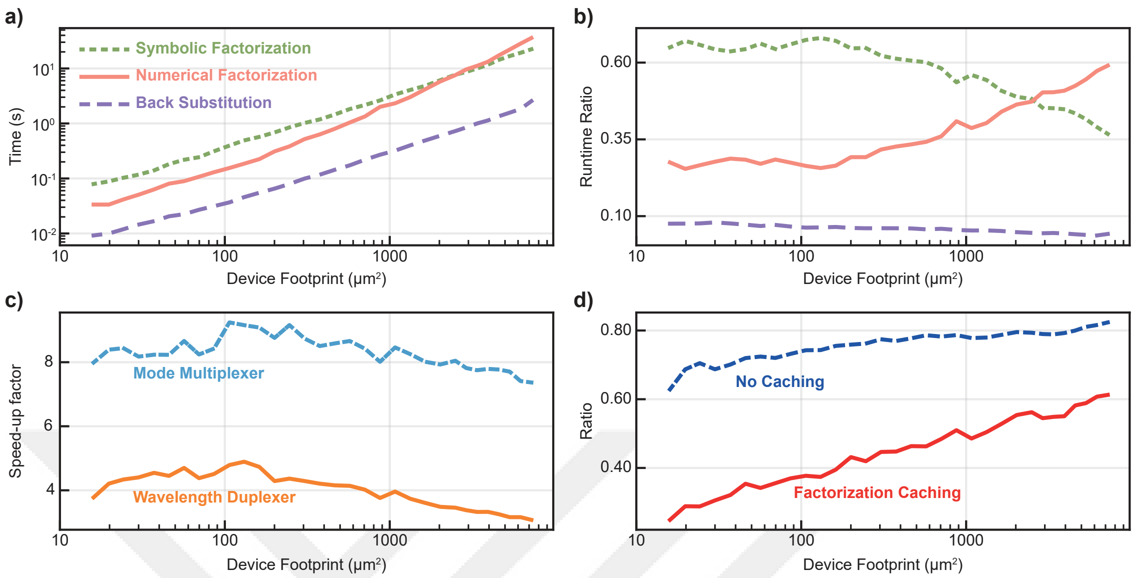


Figure 3.6: **(a)** Runtimes of symbolic factorization (SF), numerical factorization (NF), and back substitution (BS) as a function of footprint. **(b)** The ratio of runtimes to the total simulation time (SF+NF+BS). **(c)** Speed-up factor for demonstrated devices, calculated as the ratio of standard method and factorization-cached method simulation runtimes. **(d)** The portion of total optimization time used for running FDFD simulations for no-caching and factorization-cached approaches, for the design of the wavelength duplexer.

The total enhancement in simulation speedup is plotted in Fig. 3.6(c), showing that factorization cached wavelength duplexer simulations are over 4.7 times faster than the standard no-caching approach for devices smaller than 10 µm × 10 µm, and are about 3.1 times faster for much larger devices (up to 85 µm × 85 µm). Even though the symbolic factorization is completely eliminated after the first iteration, the computational enhancement decreases slightly with increasing device size, as the necessary numerical factorizations start to occupy the majority of total iteration times with larger system matrices. A similar trend is observed for the broadband mode multiplexer where the speed-up factor of 9.2 for the typical device footprint reduces to 7.3 for the largest device we tested (85 µm × 85 µm). Although the achieved speedup for these much larger footprints is lower than standard ones, Fig. 3.6(c) demonstrates that our approach still leads to a significant reduction in simulation time regardless of footprint in the scope of the inverse design applications. Lastly, in Fig. 3.6(d), we present a comparison of the simulation time to the total iteration time, which also includes operations such as smoothing and binarization of the device geometry, gradient computations, and the optimization algorithm itself. Here, when no caching is employed, over 62% of the total iteration time is spent on the simulations. This ratio reaches as high as 82% with larger devices. With our factorization caching approach , only 24% of the iteration time is used for these simulations in 4 µm × 4 µm devices, reaching up to 61% in 85 µm × 85 µm devices. As expected, our achieved speedup factors experience a slight decrease with increasing footprint since simulation time already comprises the vast majority of computational requirements for these much larger devices. These results demonstrate that with factorization caching, linear system solutions can now be completed quicker than the other operations necessary for nanophotonic device design (including image smoothing, gradient calculations, and optimization). Other enhancements through GPU-accelerated image operations or advancements in rapid optimization algorithms can further accelerate inverse design frameworks.

Chapter 4

DEEP LEARNING ENABLED DUAL-STAGE NANOPHOTONIC DEVICE SIMULATIONS FOR INVERSE DESIGN

4.1 Simulation bottleneck in inverse design

The details of the inverse design approach and usage of electromagnetic simulations have been discussed in previous chapters. Among these electromagnetic simulation algorithms utilized in photonic inverse design, Finite-Difference Frequency-Domain (FDFD) and Finite-Difference Time-Domain (FDTD) methods are widely employed. While 2D versions of these methods offer faster computations, they often sacrifice physical accuracy. Consequently, 3D simulations are essential for designing devices that exhibit agreement with experimental measurements. However, the computational requirements necessary for these 3D simulations result insignificant limitations, extending design procedures to multiple days or weeks, particularly for optimizations requiring a large number of iterations or large device footprints [Liu et al., 2018]. In response to these challenges, deep learning approaches have emerged as promising alternatives to traditional electromagnetic simulations. However, many existing models either struggle to capture underlying physical phenomena or are confined to 2D representations, resulting in a compromise of physical accuracy. Therefore, fast and physically accurate deep learning models remain a critical need for optimizing complex and non-intuitive photonic devices. To address these challenges, we propose a multi-stage physics-informed computational architecture trained using physical losses in addition to pixel-based losses obtained from electromagnetic fields. Our model accurately simulates photonic devices across various sizes, achieving physical accuracy comparable to that of 3D-FDTD simulations with a fraction of the computational requirements necessary. Specifically, our model achieves a remarkable speed-up of over 3400-fold, demonstrating its efficacy in expediting the design process for complex photonic devices. These advancements pave the way towards rapid and precise optimization of complex photonic devices, unlocking new frontiers in the field of nanophotonics.

4.2 Multi-wavelength Dataset

In order to train our model, we created a dataset consisting of geometries and their corresponding 3D FDTD simulation results as targets. As our trained model will be used in inverse design simulations, the geometries in the dataset are tailored to this application. Specifically, we formulated inverse design objectives of 1×2 devices (1 input waveguide, 2 output waveguides) for different cases, then ran these optimizations using 2D FDFD and saved the geometries generated during the iterations. As a result, our dataset includes devices that are not entirely arbitrary but rather representative of those encountered in device optimizations. The dataset includes various design region sizes, ranging from $4\,\mu m \times 4\,\mu m$ to $8\,\mu m \times 8\,\mu m$, ensuring that the model is trained to generalize across various design region dimensions. This aspect is crucial, as device optimizations may necessitate the use of larger design regions, particularly when convergence becomes challenging. The design cases also include different power splitting targets, and varying smoothing filter radii and binarization threshold levels. As the deep learning model will take the geometry as input in the form of a tensor, it is important for the training performance that all input tensors have the same shape. Therefore, we added padding to the geometries with smaller design region size than the $8\,\mu m \times 8\,\mu m$, ensuring that the overall simulation region of devices are uniform through the dataset with $11\,\mu m \times 11\,\mu m$. We also note that the devices include $1\,\mu m$ PML region on each side and the additional padding is included between the input-output waveguides and PML.

In addition to the geometries taken from inverse design cases, we also generated geometries with random pixels. This ensures that the model not only memorizes the simulations of tailored geometries in the dataset but also learns the underlying physical relationship between a geometry and its simulation result. In total, we generated 4000 geometry samples from inverse design cases and 1000 samples with random pixels. Some examples from our dataset are given in Fig. 4.1

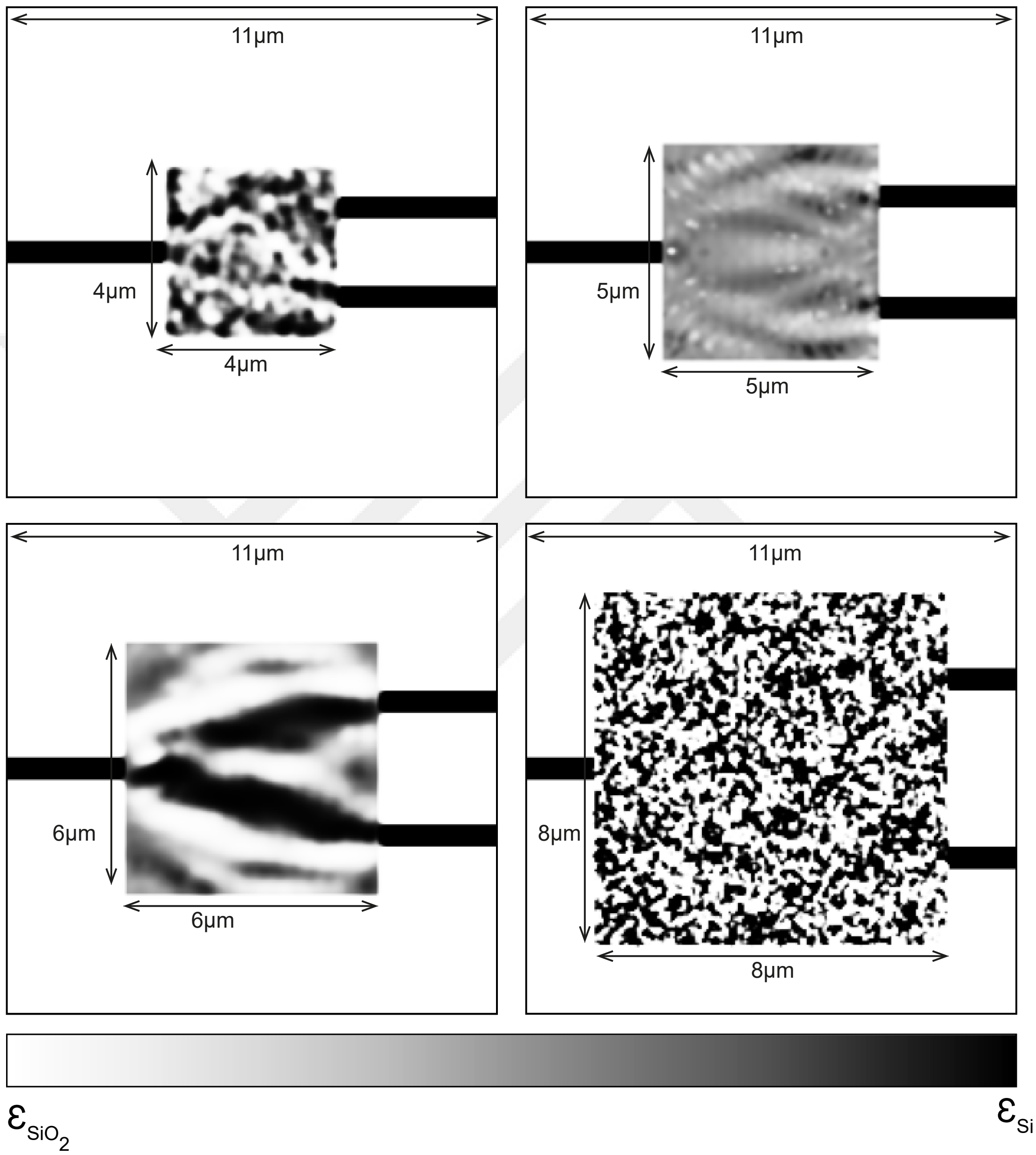


Figure 4.1: Sample geometries from training set

After the geometries are created, we run their 3D FDTD simulations to construct the target fields. The simulation region is discretized into 50 nm squares. The simulation results are extracted for 11 wavelengths with 10 nm spacing in the range of 1500 nm to 1600 nm. We also note that we included 2D slices taken from the middle of the vertical dimension of 3D FDTD simulations in the dataset. We tested the mode overlap and transmission on both the 3D results and their slices, observing that the results corresponded closely. Therefore, we used the slices as our target, since predicting a 2D region is easier than predicting a 3D region.

Finally, we generated 4000 geometries from inverse design cases and 1000 random geometries, storing them with their 3D FDTD results for 11 wavelengths. Before training, we selected the simulation results at 4 random wavelengths for each geometry and created a dataset of 20000 samples. We then split this dataset into 18000 samples for the training set, 1024 samples for the validation set, and the remainder for testing.

4.3 Physics Informed Model

The schematic diagram of our multi-stage simulation architecture and a breakdown of the constituent operations are illustrated in Fig. 4.2. Our model integrates two U-net blocks, interconnected with a 2D Finite-Difference Frequency-Domain (FDFD) simulator. Initially, an input geometry is fed into a U-net with residual blocks, akin to the architecture in [Wang et al., 2018]. The u-net architecture incorporates 7x7 convolutional layers at both the beginning and the end. The downsampling layers utilize 3x3 convolutions with a stride of 2, facilitating the reduction of spatial dimensions while preserving essential features. Similarly, upsampling layers employ transposed convolutions to upsample the feature maps to their original resolution. Residual blocks within the u-net also consist of 3x3 convolutions, designed to maintain the data shape and the number of channels while enabling the extraction of intricate features present in the device geometry. After this first u-net, the modified geometry undergoes a discretized, frequency domain Maxwell's solver calculate the magnetic and electric field components H_z, E_x, and E_y, using a fundamental TE input mode.

The primary objective of the initial stage is to selectively manipulate the device geometry, such that the result of this 2D-FDFD simulation matches as closely as possible with the 3D-FDTD fields that are accepted as ground truth. The inter-

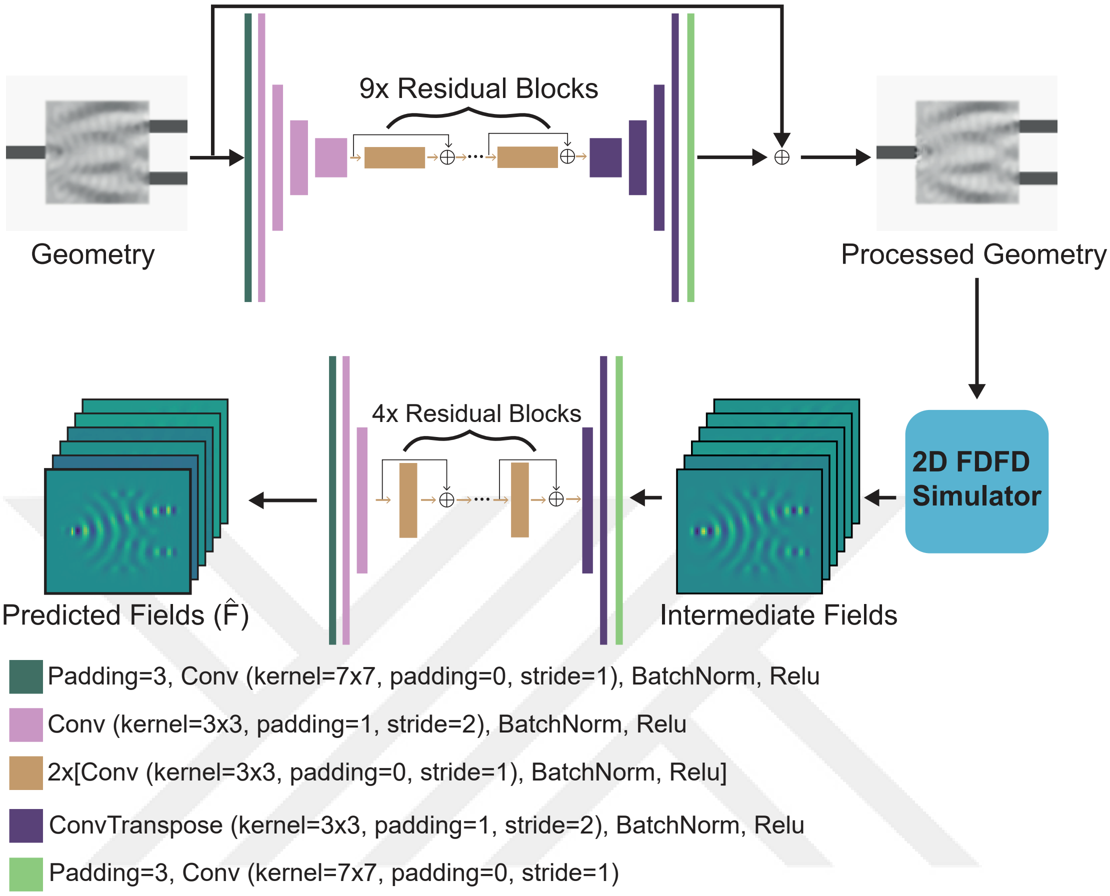


Figure 4.2: Schematic representation of the proposed configuration for the dual-stage model, comprising two U-net models and a 2D FDFD simulator.

mediate field tensor consisting of the real and imaginary parts of these three field profiles is then processed through another smaller u-net for subsequent refinement. Since the E_x and E_y fields are derived from the calculated H_z in 2D-FDFD simulations using the permittivity distribution, this next refinement stage mainly modifies the fields in order to yield more physically accurate spatial derivatives, and in turn yield a better match with the 3D-FDTD ground truth. This dual-stage network architecture with a 2D-FDFD intermediary forces physical constraints in the model, thereby enhancing its robustness and reducing dependence on training data alone.

4.4 Physics Informed Training

Both the dataset and multi-stage model are carefully constructed to perform accurate simulations on inverse design tasks. Consequently, it is crucial that the training process is tailored to this purpose. Therefore, the objective for this model includes both mode overlap (Eq. 1.1) and transmission (Eq. 1.2) metrics at the outputs of the nanophotonic device.

We derive the mode overlap and transmission metrics from the output ports of the predicted field and target field and formulate the training loss to minimize the mean absolute error. In addition to the physical losses, the L1-norm of the difference between the predicted and 3D-FDTD simulated fields is also separately included in this objective. While the L1 loss facilitates pixel-by-pixel matching between predictions and targets, the inclusion of mode overlap and transmission losses directs the model's attention toward the output port regions, which hold significant importance in inverse design applications. The formulation of the training objective including these three separate loss parts is given in Eq. 4.1. We introduced a coefficient to combine the mode overlap and transmission losses so that we can adjust the model's focus on output waveguide locations.

$$L = \alpha \sum_{j=1}^{M} |\Gamma_{(j)} - \Gamma^{T}_{(j)}| + \alpha \sum_{j=1}^{M} |P_{x(j)} - P^{T}_{x(j)}| + \frac{1}{N} \sum_{\text{pixels}} |I_{\text{pred}} - I_{\text{target}}| \tag{4.1}$$

4.4.1 Training Performance Comparison with Standard Model

We compared our model's training and validation performance with a standard model where the geometry processing is not used and 2D FDFD fields of the original geometry are directly given as input to the training model. The schematics of the model is given in Figure 4.3.

Both models are trained for 124 iterations through which the evolution of three components of training loss are shown in Fig. 4.4. The trainings were run with a batch size of 8 and an ADAM optimizer with a learning rate of 0.0002 and $\beta_1 = 0.5$, $\beta_2 = 0.999$. We also show the final values of all losses in Table 4.1. At the end of

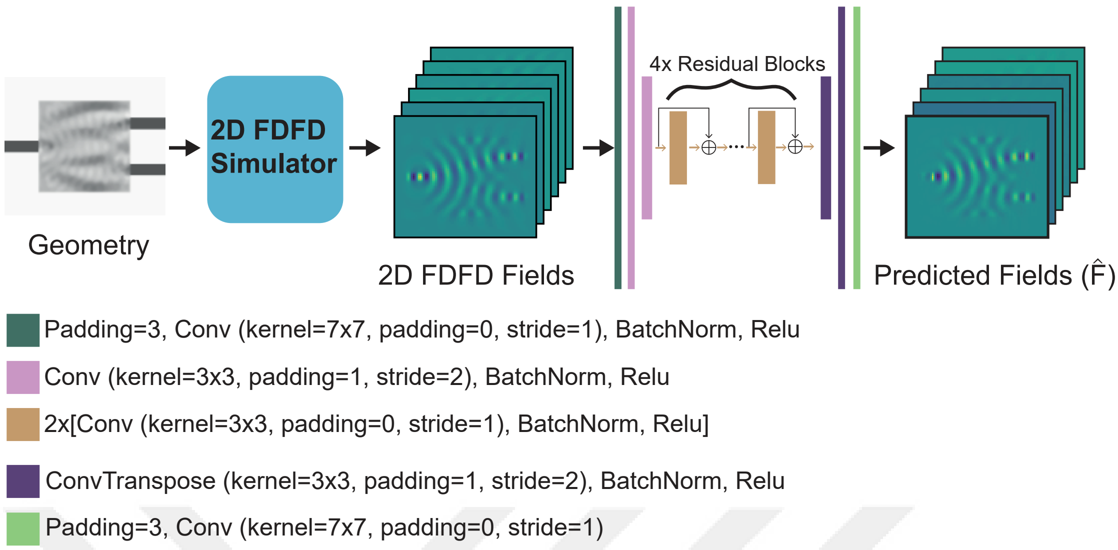


Figure 4.3: Schematic representation of the standard model with only one U-net directly processing the 2D FDFD simulator results of the original geometry.

the training, the L1 loss for our dual-stage model reaches a final value of 0.0063, indicating a low average absolute difference between the predicted and target fields. In contrast, the standard model reaches 0.0137, demonstrating a worse performance than our dual-stage model. Additionally, for the dual-stage model, the transmission loss and mode overlap loss converge to final values of 0.0053 and 0.0065, respectively which corresponds to a decrease of more than 20-fold compared to their initial values. This significant reduction underscores the model's convergence towards producing outcomes that align closely with the physics of device simulations. In comparison, the standard model's final transmission loss and mode overlap loss are 0.0069 and 0.0081, respectively, indicating that the dual-stage model performs better.

We also monitor the training performance of our model and standard model with the Structural Similarity Index (SSIM). Although SSIM is popularly used in image processing problems, it is useful also in our context as it can measure the structural similarity between two images by considering the patterns [Hore and Ziou, 2010]. This approach is particularly beneficial because a simple pixel-by-pixel comparison using L1 loss may not capture the similarity between two high-resolution images. Hence, as SSIM converges to 1, the model's predicted fields not only match the

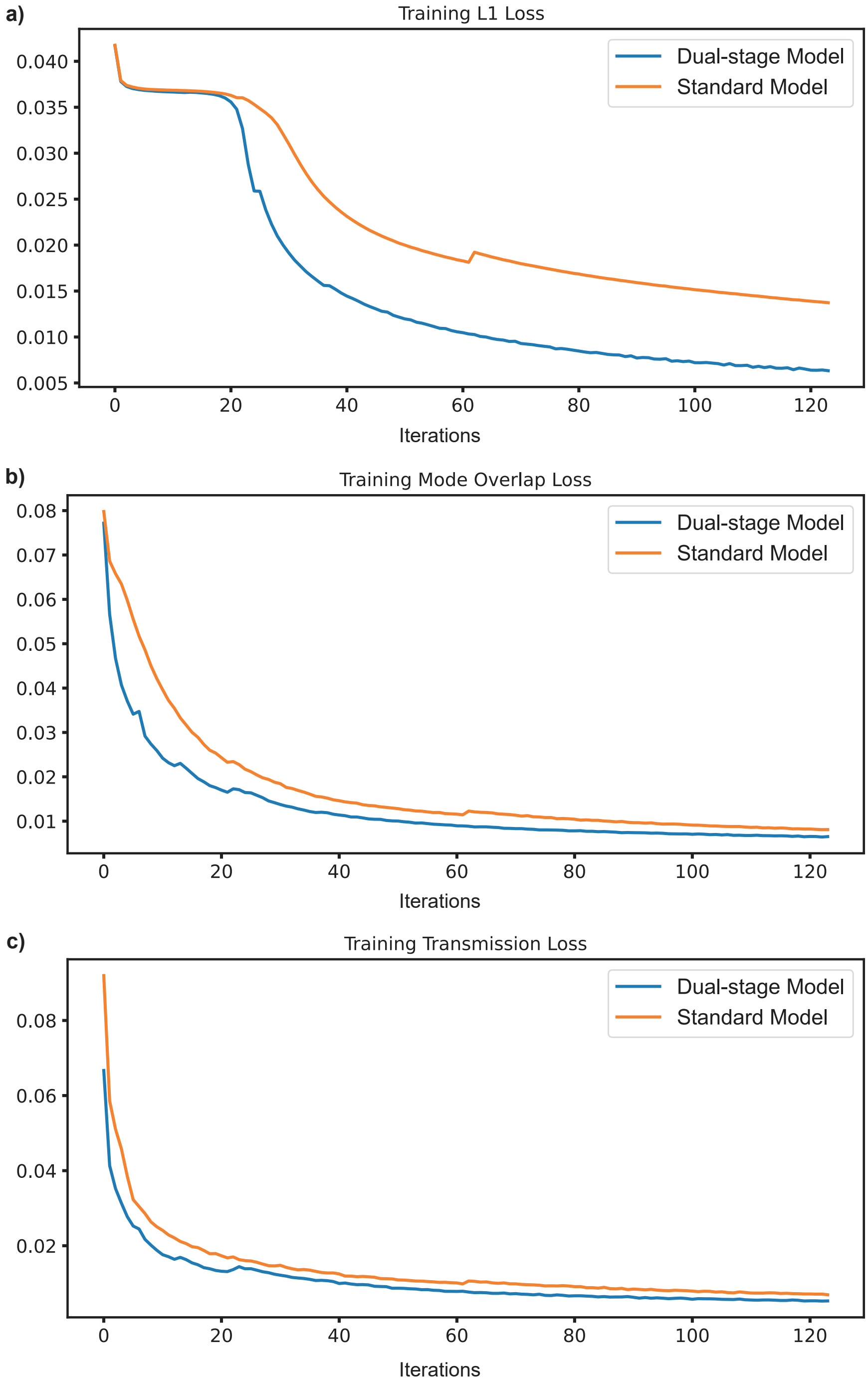


Figure 4.4: Training loss components for the dual-stage and standard model with only U-net on 2D FDFD fields.

target fields both physically and pixel by pixel but also exhibit structural similarity, indicating relative alignment of the resulting electromagnetic field patterns. This

Table 4.1: Training MAE and SSIM values obtained from standard (U-net on 2D FDFD) model and proposed dual-stage model.

Model	Field MAE	Mode Overlap MAE	Transmission MAE	Field SSIM
Standard Model	0.0137	0.0081	0.0069	0.9181
Dual-stage Model	0.0063	0.0065	0.0053	0.9801

suggests that the model accurately reproduces the electromagnetic field distributions while also preserving the finer structural details present in the target simulations.

We showed the training SSIM for our dual-stage model and standard model in Fig. 4.5. While our model reaches a very close value to 1 with 0.9801, the standard model has a maximum SSIM of 0.9181. This indicates that our model learned the structural patterns of electromagnetic fields in the training set better than the standard model.

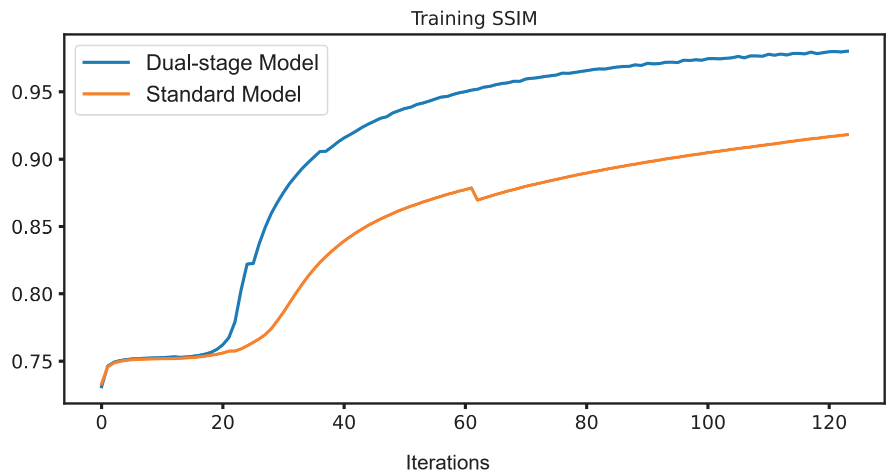


Figure 4.5: Structural similarity index measure throughout the training procedure for dual-stage and standard models.

4.4.2 Validation Performance Comparison with Standard Model

We measured the validation performance of our dual-stage model and standard model during the training on a validation set of 1028 samples which include unseen samples from different geometries as well as target fields from different wavelengths. The L1 loss for both models is plotted in Fig. 4.6 as well as the final loss values given in Table 4.2. While the standard model reaches a final value of 0.0177, our dual-stage model reaches 0.0078, showing superior performance. The performance gap is even more emphasized in transmission and mode overlap losses. Although the dual-stage model converged to a lower value for these losses, the difference was not that high. In contrast, the validation transmission loss reached 0.0061 for our dual-stage model while the standard model's transmission loss converged to 0.0092. For the mode overlap loss, the standard model showed a final value of 0.0245. In comparison, our model's mode overlap loss for validation converged to 0.1201 which is almost half of the standard model's result.

Table 4.2: Validation MAE and SSIM values obtained from standard (U-net on 2D FDFD) model and proposed dual-stage model.

Model	Field MAE	Mode Overlap MAE	Transmission MAE	Field SSIM
Standard Model	0.0177	0.0245	0.0092	0.8866
Dual-stage Model	0.0078	0.0120	0.0061	0.9676

We also monitor the SSIM of the validation set for both models. We observe that our model shows a better SSIM performance in validation by reaching a maximum value of 0.9676 in contrast to the maximum SSIM of the standard model, which is 0.8866.

The superior performance of our dual-stage model in L1 loss and SSIM metrics indicates that using a geometry processing model enhances the accuracy and structural integrity of the predicted electromagnetic fields. Moreover, the dual-stage model exhibits lower physical losses in both the training and validation phases, mak-

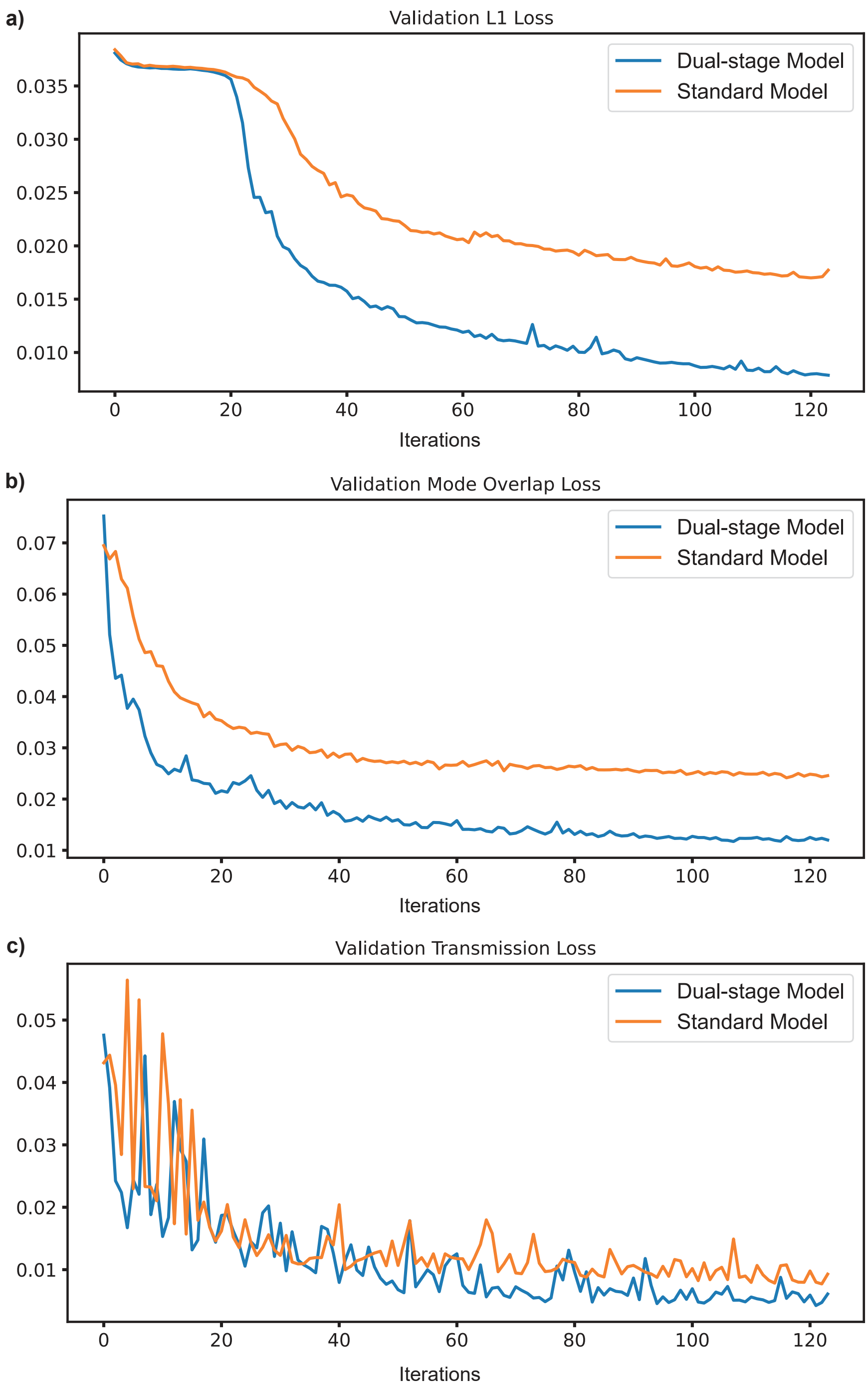


Figure 4.6: Evolution of the validation loss components for the dual-stage and standard model.

ing it particularly well-suited for the context of inverse design simulations.

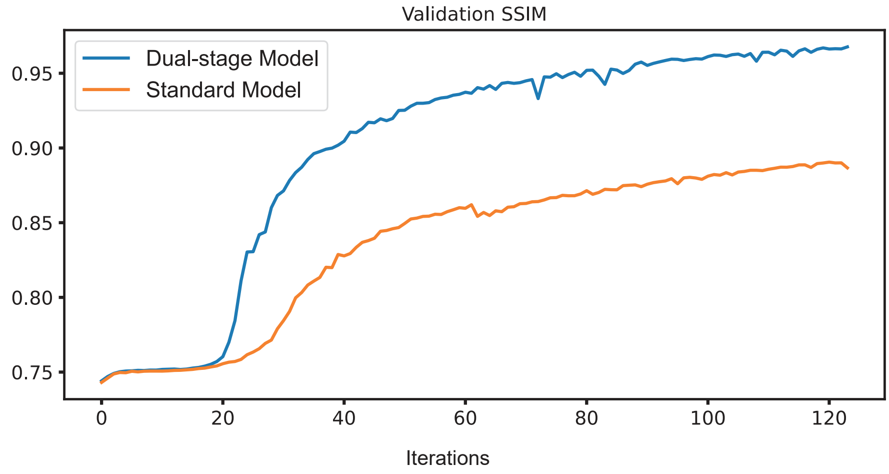


Figure 4.7: Structural similarity index measure of the validation set for dual-stage and standard models.

4.5 Comparisons with 2D FDFD

We run separate 2D FDFD simulations of validation samples using the effective index method at their respective wavelength and compare the results to our dual-stage model and standard model [Knox and Toulios, 1970]. The effective index method simplifies the 3D simulations by using the vertical refractive index of the respective mode as the refractive index of the silicon, effectively reducing it to a 2D problem. In this way, the computational complexity is significantly reduced while maintaining a reasonable degree of accuracy, especially for devices that don't have complex shapes. The performance of this representative effective index-based FDFD method and our model together with the standard method is compared against the target 3D FDTD simulations in Fig. 4.8. Here, we plot the L1 norm as a function of the design region area. We find that the effective index-based FDFD exhibits an average L1 loss of 0.0393 throughout all device sizes tested, significantly inferior to our model's performance. Furthermore, the performance of effective index-based FDFD deteriorates with larger device sizes. In contrast, our model achieves an L1 loss of 0.0238 while the standard model has an average L1 loss of 0.0255. These results show that the standard method outperforms the effective index-based 2D FDFD approach and our dual-stage model surpasses both in matching pixel-based similarities to the target 3D FDTD fields. A similar trend is observed for the SSIM results, with our model achieving an average SSIM of 0.7540 and the standard method's SSIM converging to 0.6894. Both of these approaches are again superior to effective index-based 2D FDFD as it reaches a mean SSIM of 0.5424. It is also worth emphasizing that our model consistently achieves more successful performance than 2D FDFD and standard model across various device sizes from 4×4 to 8×8 μm^2, highlighting its robustness and generalization capability. We note that the visibly better performance in 8 µm × 8 µm design region area arises from the fact that the validation set is not balanced, with the 8 µm × 8 µm has the smallest number of data. This imbalance may contribute to the improved performance observed in this specific design area. We also note that the measurements presented here are derived from the runs on the same validation set as shown in Figure 4.6 and Table 4.2. However, the values of L1 loss and SSIM are different because these metrics in Figure 4.6 were calculated using batches during training, whereas we did not use batching in this section. This discrepancy accounts for the variation in the reported values.

In addition to the tests at different device sizes, we also conducted an analysis of geometrical device complexity, and how that affects the simulator performance within our dataset. To quantify geometrical complexity, we computed the entropy of the Fast Fourier Transform (FFT) of each permittivity distribution. The details of complexity metric calculation are given in A.1. Subsequently, we plotted the SSIM results of our model, standard model, and the effective index-based FDFD method in Figure 4.9, providing insights into the performance of all approaches across varying levels of feature complexity. Specifically, we observed in Figure 4.9a that while effective index-based 2D FDFD achieves SSIM results lower than 0.7 across all levels of complexity, 46.9% of samples for the standard model reach an SSIM of higher than 0.7. In contrast, our model demonstrates superior performance (Figure 4.9b) with an SSIM of minimum 0.7 for 78.4% of devices in the dataset. This significant difference underscores the effectiveness of the presented dual-stage model in accurately predicting electromagnetic field distributions across a wide range of geometric complexities, outperforming traditional simulation methods.

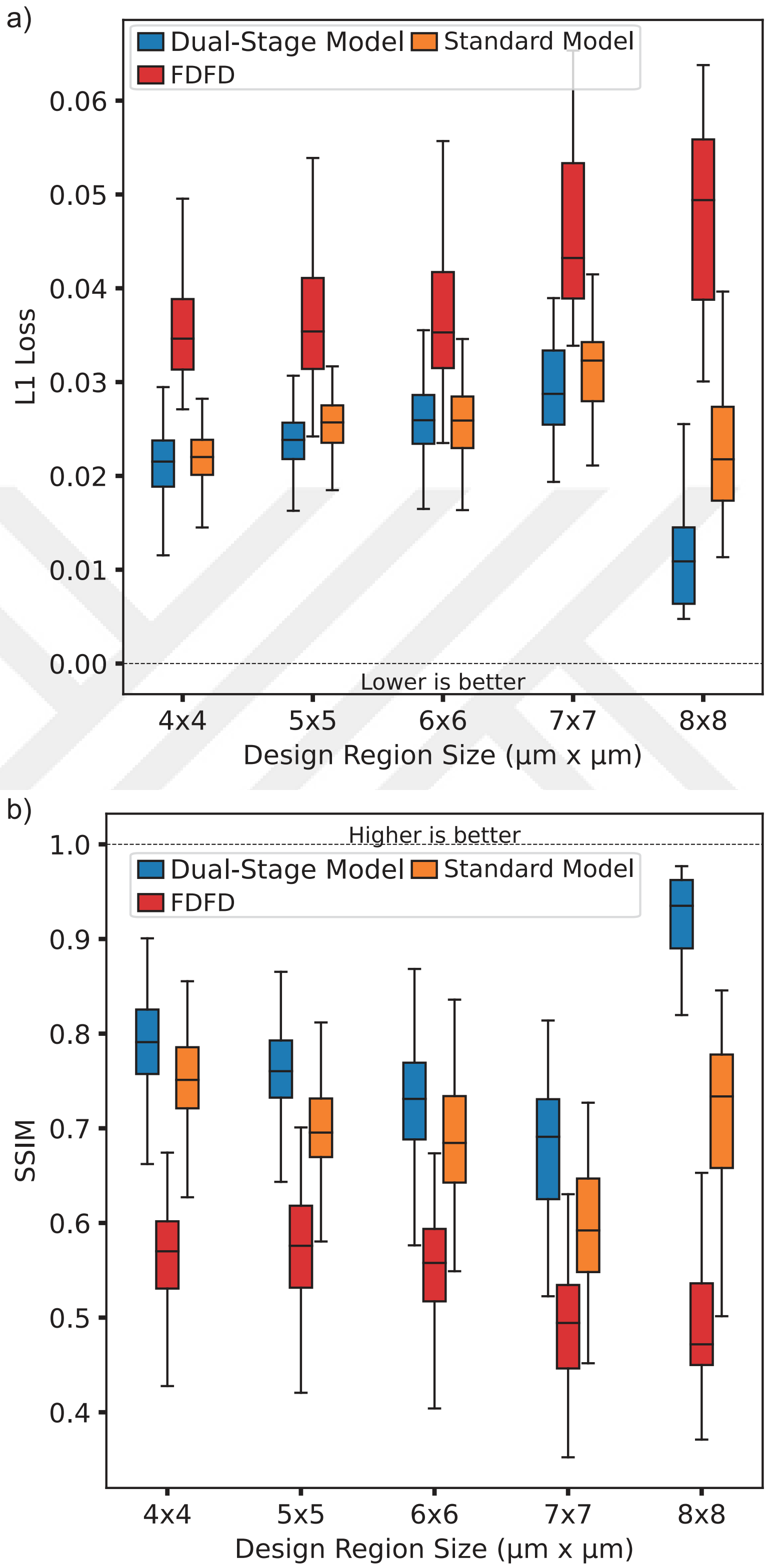


Figure 4.8: (a) L1 Loss, (b) SSIM of the validation samples for our dual-stage model, standard model with only u-net and 2D FDFD.

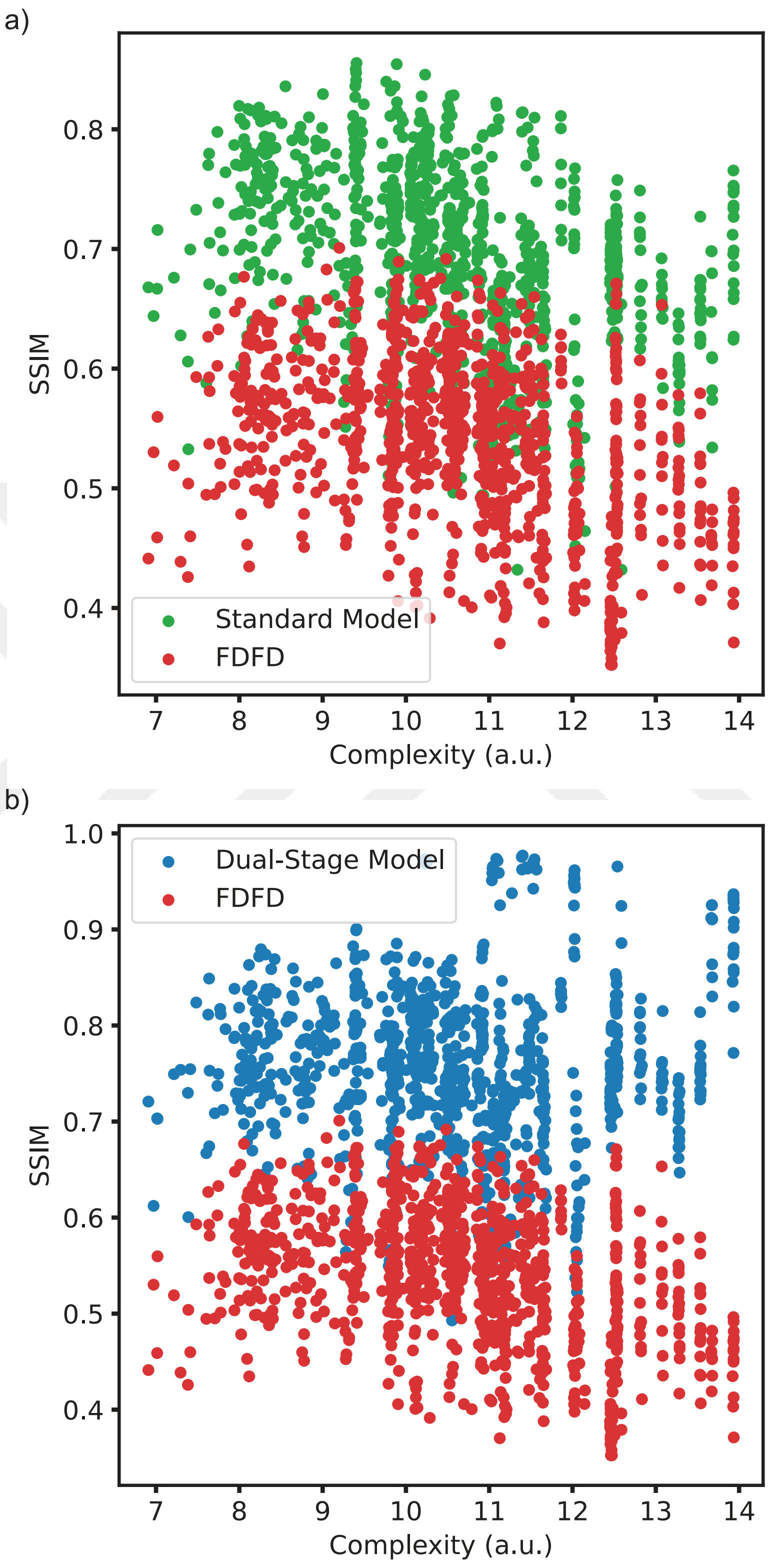


Figure 4.9: Variation of SSIM with device complexity, determined by entropy of geometry FFTs.

4.6 Application to the design of arbitrary power splitters

To show the performance of the dual-stage model with physically informed losses on the photonic device optimization tasks, we replace it with the simulator block shown in Fig. 1.1 and perform the inverse design of arbitrary power splitters at a broadband range between 1500 nm-1600 nm wavelengths. These devices divide the light into two ports, each with specified power ratios relative to the input port. We formulated the objective of the splitter optimizations to match the specified splitting target in 1500 nm, 1520 nm. 1550 nm, 1580 nm, 1600 nm wavelengths. We also run the same optimization using the standard model in Figure 4.3 and compare the results with our model. After optimizing the devices using our model as a simulator, we run a 3D FDTD simulation on the designed geometry to assess the physical accuracy of model simulations.

Our first design is a 10:90 power splitter on a design region area of 4 µm × 4 µm with 50 nm discretization where the geometry of the final designed device is shown in Figure 4.10. In order to match the full simulation region used in the training set of the model (11 µm × 11 µm), we extend the waveguides in the horizontal direction and add SiO_2 as padding.

The H_z component of 3D FDTD fields and predictions of our dual-stage model at 1550 nm for this 10:90 splitter are given in Figure 4.11. The same plots for E_y and E_x are shown in A.2. We observe that both the real and imaginary components of the predicted field $\hat{H}_z$ closely resemble the corresponding 3D FDTD results. The bottom plots, which display the absolute difference between the predicted and actual fields, reveal a periodic structure. This indicates that our dual-stage model accurately predicts the physical pattern of the electromagnetic fields. The observed error primarily arises from differences in intensity for corresponding pixels.

The transmission results of optimization and 3D FDTD simulations for the broadband wavelength range are given in Figure 4.12. Mean values of transmission results in this figure are given in Table 4.3. The overall loss, which is attributed to the sum of transmissions in both ports, for optimizations employing the standard model is 21.2% while our model outperforms this result by a value of 6.7%. This

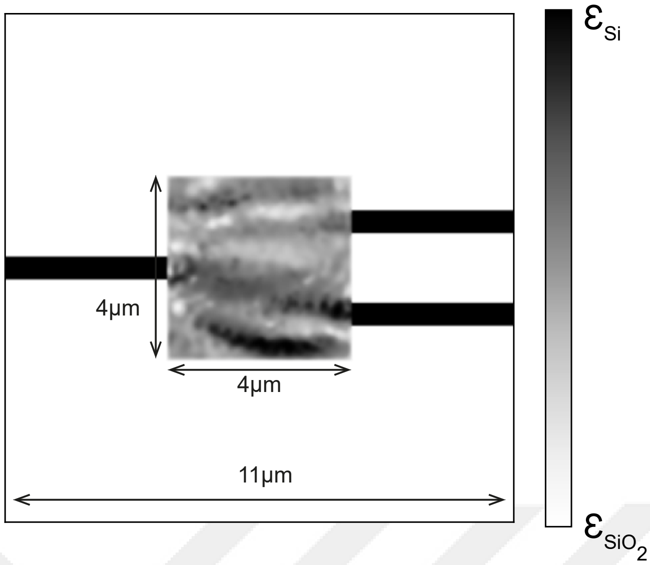


Figure 4.10: Permittivity distribution of the 10:90 splitter designed with optimization employing proposed dual-stage model.

Table 4.3: Inverse design and 3D FDTD results with 4 µm × 4 µm device dimensions

Model	Device Size (μm^2)	Splitting Target	Transmission (Optimization)	Transmission (3D FDTD)
Standard Model	4×4	0.10-0.90	0.009-0.779	0.012-0.679
Dual-Stage Model	4×4	0.10-0.90	0.051-0.882	0.069-0.828

result shows that using our model resulted in an easier convergence in the inverse design of devices. Moreover, while the difference between standard model employed optimizations and 3D FDTD transmissions is 9.7%, the dual-stage model reduces this difference to 3.6%. These results indicate that using a geometry processing u-net as the first stage of the dual-stage model translates to higher physical accuracy in the simulation of inverse design devices.

After demonstrating our model's success on the design of arbitrary splitters with 4 µm × 4 µm design region, we extend our application tests to a dimension never before encountered in training, 8 µm × 5 µm. Table 4.4 and Figure 4.13 reveal that the standard u-net approach yielded sub-optimal results, failing to match the results

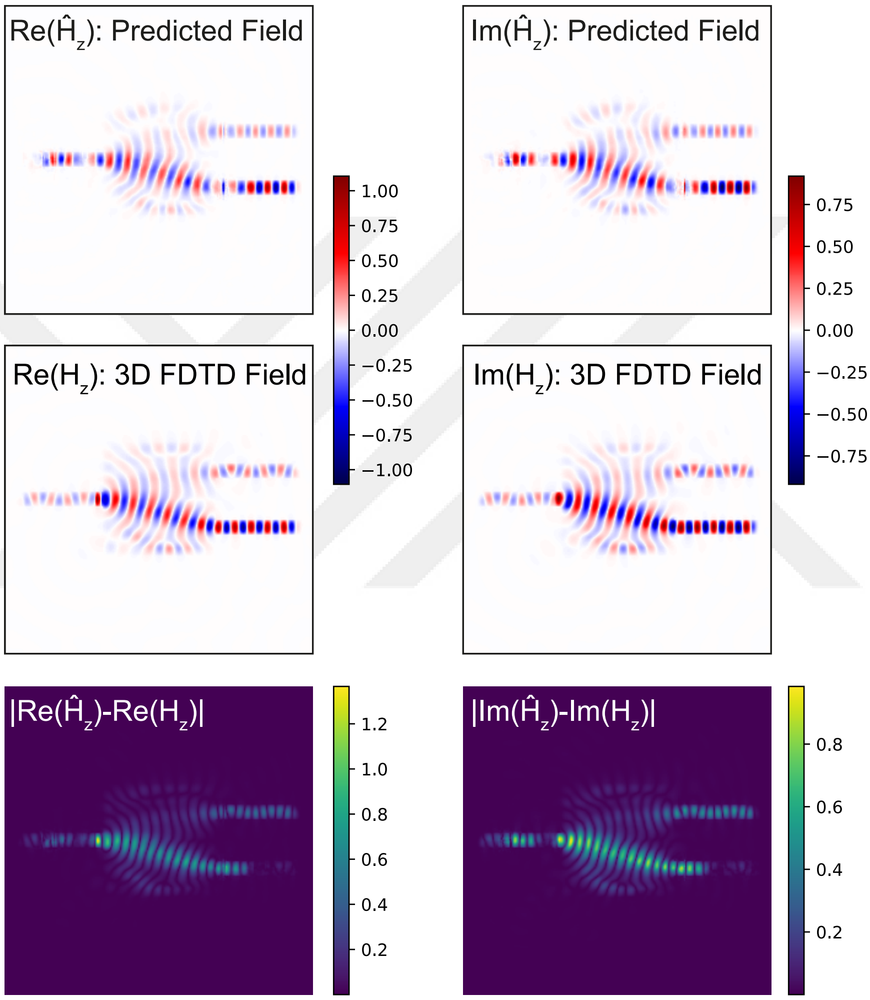


Figure 4.11: Real and imaginary parts of z-component of the predicted and 3D FDTD field for the 10:90 power splitter design. The bottom row shows the absolute difference between predicted and ground truth values.

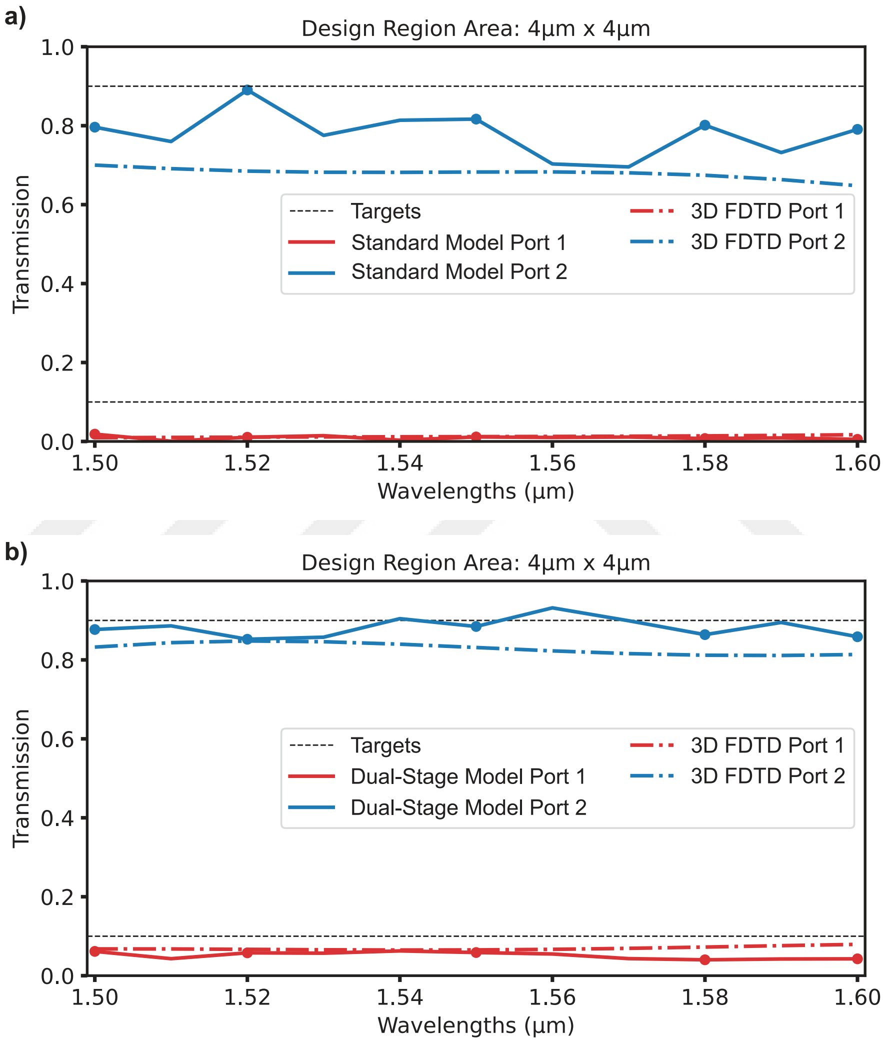


Figure 4.12: Comparison of the transmission measurements for the 4 µm × 4 µm 10:90 power splitter.

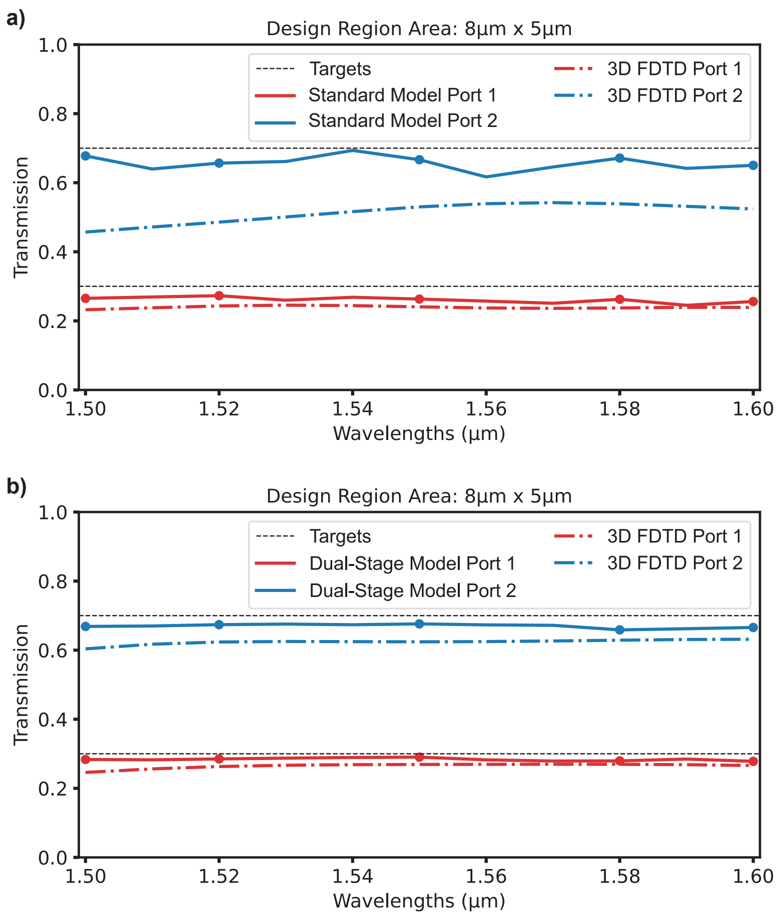


Figure 4.13: Comparison of the transmission measurements for the 8 µm × 5 µm 30:70 power splitter.

of 3D FDTD simulations by a large transmission difference of 16.5%. In contrast, our model performs better with a smaller error of 6.5%. This indicates that our model has acquired a deep understanding of the underlying physics of electromagnetic simulations, as it showed high accuracy in both data and dimensions not included in the training set.

4.7 Model Runtime Analysis

Finally, we present the inference time comparison of our model with 3D FDTD in Table(4.5). The tests were conducted on our standardized 11μm x 11μm region with a 50 nm discretization. For our model's runtime breakdown, three components are listed: the first u-net, 2D FDFD, and the second u-net. When running the model on

Table 4.4: Inverse design and 3D FDTD results with 8 µm × 5 µm device dimensions

Model	Device Size (µm^2)	Splitting Target	Transmission (Optimization)	Transmission (3D FDTD)
Standard Model	8×5	0.30-0.70	0.260-0.656	0.239-0.512
Dual-Stage Model	8×5	0.30-0.70	0.283-0.669	0.264-0.623

the CPU, we observe that 2D FDFD takes 0.193 s, while the rest of the model takes 0.506 s. However, transitioning to a GPU (Tesla V100) for the first and second u-net stages leads to a significant decrease in runtime, reducing the entire simulation to 0.214 s. Notably, in this GPU version, 2D FDFD simulations are the bottleneck, accounting for 90% of the model's total runtime. To address this bottleneck, we employed a factorization caching approach [Dasdemir et al., 2023], which results in a significant reduction of the 2D FDFD runtime by 39% for this specific device size. With this enhancement, the 2D FDFD runtime decreases to 0.122 s, while the overall model runtime drops to 0.144 s. Consequently, a device simulation using our model runs 3408 times faster than traditional 3D FDTD without sacrificing the physical accuracy of the resulting fields. It is also important to emphasize that employing factorization caching not only improves inference performance but also reduces backward propagation time by storing numerical factorizations. These factorizations are then utilized in adjoint simulations that are necessary for computing the gradient of FDFD simulations. In inverse design applications where multiple forward and backward propagations are required, this approach significantly enhances runtime performance. By leveraging our model together with factorization caching, the computational burden associated with running additional 3D FDTD simulations in the gradient phase of inverse design applications is entirely mitigated. This enhancement streamlines the iterative optimization process, making it more efficient and enabling orders of magnitude faster convergence to optimal designs. These findings streamline the application of computational models to physically accurate device simulations and underscore the significant speed-up our model offers compared to

Table 4.5: Comparing model, 2D FDFD, and 3D FDTD methods in terms of the operation time with and without Factorization Caching (FC).

Method	Time(s) (without FC)	Time(s) (with FC)
Model CPU	0.506	0.435
Model GPU	0.214	0.144
FDFD	0.193	0.122
3D FDTD	490.895	490.895

traditional simulation methods, particularly in the case of iterative processes in inverse design approaches.

Chapter 5

CONCLUSION

In the first part of this thesis, we demonstrated a multi-scale hierarchical topology optimization approach to address the long computational time of photonic inverse design. While the adjoint method enabled the computation of the gradient of the objective function with only two simulations, the computational speed of photonic optimizations remains constrained by the solution of sparse linear systems in FDFD simulations. To address this, our study introduces a multi-scale hierarchical topology optimization approach that begins with a coarse grid and continuous permittivity distribution, progressively refining and binarizing the design to achieve a high-efficiency final device. This method allows for rapid convergence, demonstrated by the design of a mode multiplexer achieving over 98.1% efficiency in under 50 seconds. Our findings underscore the potential of multi-scale hierarchical optimization in the rapid and efficient design of nanophotonic devices, paving the way for broader applications in the field.

In the second part of this thesis, we have demonstrated a multi-faceted factorization caching approach for the computational enhancement of inverse photonic design by leveraging the symbolic and numerical structure of the system matrix. With our approach, by caching and reusing the factorizations, we eliminate all symbolic factorizations in forward and adjoint simulations after the first iteration. We also reduce the numerical factorizations in the forward simulations proportionally to the different input modes used and eliminate all numerical factorizations in adjoint simulations by utilizing the stored factorizations in forward simulations. Our results highlight the effectiveness of factorization caching in reducing the computational cost associated with device simulation and optimization processes. As we introduce no changes or approximations in the underlying system of equations, our method remains generalizable to a large class of photonic design problems with the system solutions remaining exact, physically accurate, and data-independent. Currently, since we specifically focused on enhancing the computational efficiency of sparse direct solvers, our factorization caching method is particularly suitable for 2D FDFD-based optimizations, as simulations like 3D FDTD typically rely on iterative solvers due to their large memory requirements [Su et al., 2020]. However, the demonstrated caching approach fundamentally remains applicable to FDTD-based or other simulations, provided that direct approaches are employed in solving the associated linear system of equations. Moreover, our method is also applicable to any photonic platform with different choices of materials, layer thicknesses, or even sidewall angles in the case of 3D simulations, as these properties are all represented with the construction of the underlying linear system of equations. Coupled with state-of-the-art optimization algorithms, this approach can enable the design of large-scale, and multi-purpose nanophotonic devices in a variety of integrated photonics platforms. Finally, our factorization caching method can even extend beyond photonics, as it can be applied to many other computational applications requiring repetitive solutions of structurally similar system matrices.

Finally, in the last part, we have presented a dual-stage model trained with physics-based losses, resulting in highly accurate simulation outcomes within simulation short timeframes that are orders of magnitude faster than traditional methods. By integrating deep learning techniques with a 2D FDFD simulator and incorporating mode overlap and transmission losses into the training scheme, we achieved remarkable accuracy in simulations at unprecedented speeds—up to 3408 times faster than conventional 3D FDTD. The model's effectiveness in producing physically accurate simulation results that are as close as possible to the ground truth was extensively validated through comprehensive testing, demonstrating superior performance in various geometry complexities and robust generalization capabilities to a wide range of device sizes. These advancements represent significant progress in the field of inverse photonic design, enabling the optimization design of photonic devices that were previously elusive and/or impractical due to prohibitively high simulation costs.

BIBLIOGRAPHY

[int,] oneMKL PARDISO: Parallel direct sparse solver interface. `https://www.intel.com/content/www/us/en/docs/onemkl/developer-reference-c/2023-1/onemkl-pardiso-parallel-direct-sparse-solver-iface`.

[Buus, 1982] Buus, J. (1982). The effective index method and its application to semiconductor lasers. *IEEE Journal of Quantum Electronics*, 18(7):1083–1089.

[Chen et al., 2022] Chen, M., Lupoiu, R., Mao, C., Huang, D.-H., Jiang, J., Lalanne, P., and Fan, J. A. (2022). High speed simulation and freeform optimization of nanophotonic devices with physics-augmented deep learning. *ACS Photonics*, 9(9):3110–3123.

[Dasdemir et al., 2020] Dasdemir, A. O., Koral, A. K., Kiraz, B., Kiraz, A., and Magden, E. S. (2020). Multi-scale hierarchical topology optimization for nanophotonic design. In *Integrated Photonics Research, Silicon and Nanophotonics*, pages ITu4A–11. Optica Publishing Group.

[Dasdemir et al., 2023] Dasdemir, A. O., Minden, V., and Magden, E. S. (2023). Computational scaling in inverse photonic design through factorization caching. *Applied Physics Letters*, 123(22).

[Dasdemir et al., 2024] Dasdemir, A. O., Minden, V., and Magden, E. S. (2024). Accelerating simulations in inverse photonic design through factorization caching. In *Physics and Simulation of Optoelectronic Devices XXXII*, volume 12880, pages 233–237. SPIE.

[Davis et al., 2016] Davis, T. A., Rajamanickam, S., and Sid-Lakhdar, W. M. (2016). A survey of direct methods for sparse linear systems. *Acta Numerica*, 25:383–566.

[Elesin et al., 2014] Elesin, Y., Lazarov, B. S., Jensen, J. S., and Sigmund, O. (2014). Time domain topology optimization of 3d nanophotonic devices. *Photonics and Nanostructures-Fundamentals and Applications*, 12(1):23–33.

[Hammond and Camacho, 2019] Hammond, A. M. and Camacho, R. M. (2019). Designing integrated photonic devices using artificial neural networks. *Optics express*, 27(21):29620–29638.

[Hammond et al., 2022] Hammond, A. M., Oskooi, A., Chen, M., Lin, Z., Johnson, S. G., and Ralph, S. E. (2022). High-performance hybrid time/frequency-domain topology optimization for large-scale photonics inverse design. *Optics Express*, 30(3):4467–4491.

[Hore and Ziou, 2010] Hore, A. and Ziou, D. (2010). Image quality metrics: Psnr vs. ssim. In *2010 20th international conference on pattern recognition*, pages 2366–2369. IEEE.

[Hughes et al., 2018a] Hughes, T. W., Minkov, M., Williamson, I. A., and Fan, S. (2018a). Adjoint method and inverse design for nonlinear nanophotonic devices. *ACS Photonics*, 5(12):4781–4787.

[Hughes et al., 2018b] Hughes, T. W., Minkov, M., Williamson, I. A. D., and Fan, S. (2018b). Adjoint method and inverse design for nonlinear nanophotonic devices. *ACS Photonics*, 5(12):4781–4787.

[Jalali and Fathpour, 2006] Jalali, B. and Fathpour, S. (2006). Silicon photonics. *Journal of lightwave technology*, 24(12):4600–4615.

[Knox and Toulios, 1970] Knox, R. and Toulios, P. (1970). Integrated circuits for the millimeter through optical frequency range. In *Proc. Symp. Submillimeter Waves*, volume 20, pages 497–515. Brooklyn, NY.

[Lalau-Keraly et al., 2013] Lalau-Keraly, C. M., Bhargava, S., Miller, O. D., and

Yablonovitch, E. (2013). Adjoint shape optimization applied to electromagnetic design. *Optics express*, 21(18):21693–21701.

[Liu et al., 2018] Liu, D., Tan, Y., Khoram, E., and Yu, Z. (2018). Training deep neural networks for the inverse design of nanophotonic structures. *Acs Photonics*, 5(4):1365–1369.

[Liu et al., 2021] Liu, Z., Zhu, D., Raju, L., and Cai, W. (2021). Tackling photonic inverse design with machine learning. *Advanced Science*, 8(5):2002923.

[Lu and Vučković, 2012] Lu, J. and Vučković, J. (2012). Objective-first design of high-efficiency, small-footprint couplers between arbitrary nanophotonic waveguide modes. *Optics express*, 20(7):7221–7236.

[Luan et al., 2018] Luan, E., Shoman, H., Ratner, D. M., Cheung, K. C., and Chrostowski, L. (2018). Silicon photonic biosensors using label-free detection. *Sensors*, 18(10):3519.

[Ma et al., 2021] Ma, W., Liu, Z., Kudyshev, Z. A., Boltasseva, A., Cai, W., and Liu, Y. (2021). Deep learning for the design of photonic structures. *Nature Photonics*, 15(2):77–90.

[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.

[Marchant, 2017] Marchant, D. (2017). pyMKL: Python wrapper of Intel MKL routines. `https://github.com/dwfmarchant/pyMKL`.

[Mazzini et al., 2015] Mazzini, M., Traverso, M., Webster, M., Muzio, C., Anderson, S., Sun, P., Siadat, D., Conti, D., Cervasio, A., Pfnuer, S., et al. (2015). 25gbaud pam-4 error free transmission over both single mode fiber and multimode fiber in a qsfp form factor based on silicon photonics. In *2015 Optical Fiber Communications Conference and Exhibition (OFC)*, pages 1–3. IEEE.

[Michaels et al., 2019] Michaels, A., Wu, M. C., and Yablonovitch, E. (2019). Hierarchical design and optimization of silicon photonics. *IEEE Journal of Selected Topics in Quantum Electronics*, 26(2):1–12.

[Michaels et al., 2020] Michaels, A., Wu, M. C., and Yablonovitch, E. (2020). Hierarchical design and optimization of silicon photonics. *IEEE Journal of Selected Topics in Quantum Electronics*, 26(2):1–12.

[Michaels and Yablonovitch, 2018] Michaels, A. and Yablonovitch, E. (2018). Leveraging continuous material averaging for inverse electromagnetic design. *Optics Express*, 26(24):31717–31737.

[Minden, 2022] Minden, V. (2022). Linear system solver. `https://github.com/Photonic-Architecture-Laboratories/linear-system-solver`.

[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.

[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., 2017] Piggott, A. Y., Petykiewicz, J., Su, L., and Vučković, J. (2017). Fabrication-constrained nanophotonic inverse design. *Scientific reports*, 7(1):1786.

[Schenk and Gärtner, 2004] Schenk, O. and Gärtner, K. (2004). Solving unsymmetric sparse systems of linear equations with PARDISO. *Future Generation Computer Systems*, 20(3):475–487.

[Schenk et al., 2000] Schenk, O., Gärtner, K., and Fichtner, W. (2000). Efficient sparse lu factorization with left-right looking strategy on shared memory multiprocessors. *BIT Numerical Mathematics*, 40:158–176.

[Shi et al., 2022] Shi, Y., Zhang, Y., Wan, Y., Yu, Y., Zhang, Y., Hu, X., Xiao, X., Xu, H., Zhang, L., and Pan, B. (2022). Silicon photonics for high-capacity data communications. *Photonics Research*, 10(9):A106–A134.

[Singh et al., 2020] Singh, N., Xin, M., Li, N., Vermeulen, D., Ruocco, A., Magden, E. S., Shtyrkova, K., Ippen, E., Kärtner, F. X., and Watts, M. R. (2020). Silicon photonics optical frequency synthesizer. *Laser & Photonics Reviews*, 14(7):1900449.

[Su et al., 2020] Su, L., Vercruysse, D., Skarda, J., Sapra, N. V., Petykiewicz, J. A., and Vučković, J. (2020). Nanophotonic inverse design with spins: Software architecture and practical considerations. *Applied Physics Reviews*, 7(1).

[Talebzadeh et al., 2017] Talebzadeh, A., Koo, K., Vuppunutala, P. K., Nadolny, J., Li, A., Liu, Q., Sochoux, P., Pommerenke, D., and Drewniak, J. L. (2017). Si and emi performance comparison of standard qsfp and flyover qsfp connectors for 56+ gbps applications. In *2017 IEEE International Symposium on Electromagnetic Compatibility & Signal/Power Integrity (EMCSI)*, pages 776–781. IEEE.

[Vandoorne et al., 2014] Vandoorne, K., Mechet, P., Van Vaerenbergh, T., Fiers, M., Morthier, G., Verstraeten, D., Schrauwen, B., Dambre, J., and Bienstman, P. (2014). Experimental demonstration of reservoir computing on a silicon photonics chip. *Nature communications*, 5(1):3541.

[Vercruysse et al., 2019] Vercruysse, D., Sapra, N. V., Su, L., Trivedi, R., and Vučković, J. (2019). Analytical level set fabrication constraints for inverse design. *Scientific Reports*, 9(1):1–7.

[Wang et al., 2018] Wang, T.-C., Liu, M.-Y., Zhu, J.-Y., Tao, A., Kautz, J., and Catanzaro, B. (2018). High-resolution image synthesis and semantic manipulation with conditional gans. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8798–8807.

[Wiecha et al., 2021] Wiecha, P. R., Arbouet, A., Girard, C., and Muskens, O. L. (2021). Deep learning in nano-photonics: inverse design and beyond. *Photonics Research*, 9(5):B182–B200.

[Xiang et al., 2021] Xiang, C., Bowers, S. M., Bjorlin, A., Blum, R., and Bowers, J. E. (2021). Perspective on the future of silicon photonics and electronics. *Applied Physics Letters*, 118(22).

[Yang et al., 2022] Yang, K. Y., Shirpurkar, C., White, A. D., Zang, J., Chang, L., Ashtiani, F., Guidry, M. A., Lukin, D. M., Pericherla, S. V., Yang, J., et al. (2022). Multi-dimensional data transmission using inverse-designed silicon photonics and microcombs. *Nature Communications*, 13(1):7862.

[Zhang et al., 2021] Zhang, G., Xu, D.-X., Grinberg, Y., and Liboiron-Ladouceur, O. (2021). Topological inverse design of nanophotonic devices with energy constraint. *Optics Express*, 29(8):12681–12695.

[Zhao et al., 2019] Zhao, N. Z., Boutami, S., and Fan, S. (2019). Accelerating adjoint variable method based photonic optimization with schur complement domain decomposition. *Optics Express*, 27(15):20711–20719.

[Zhao et al., 2022] Zhao, N. Z., Boutami, S., and Fan, S. (2022). Efficient method for accelerating line searches in adjoint optimization of photonic devices by combining schur complement domain decomposition and born series expansions. *Optics Express*, 30(4):6413–6424.

[Zhu et al., 1997] Zhu, C., Byrd, R. H., Lu, P., and Nocedal, J. (1997). Algorithm 778: L-bfgs-b: Fortran subroutines for large-scale bound-constrained optimization. *ACM Transactions on mathematical software (TOMS)*, 23(4):550–560.

Appendix A

A.1 Complexity Metric

We showed our model's successful performance over various geometry complexities in 4.5. The procedure for calculating the complexity metric is given in A.1.

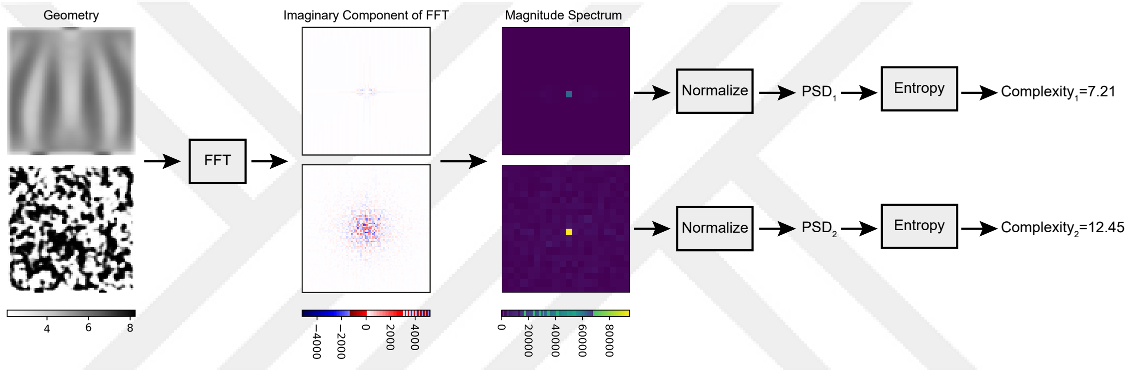


Figure A.1: Diagram for the computation of the complexity metric.

Here we considered only the design region without including the waveguides as they are common for all devices. We compared the intermediate and results of two images with a less complex one (top) and a more complex one (bottom). The FFT of the design region image is computed to convert the image into frequency domain from spatial domain and represent it in terms of its frequency components. The equation for FFT is given in A.1 where $f(x, y)$ represents the input design region image.

$$\mathcal{F}(u, v) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) e^{-i2\pi\left(\frac{ux}{M} + \frac{vy}{N}\right)} \tag{A.1}$$

As this FFT result $\mathcal{F}(u, v)$ have complex values, we compute its magnitude to find the magnitude spectrum of the design region.

$$\text{Magnitude}(u, v) = |\mathcal{F}(u, v)| \tag{A.2}$$

We then normalize the magnitude spectrum to form a probability distribution which we call normalized power spectral density.

$$\text{PSD}_{\text{normalized}}(u, v) = \frac{\text{Magnitude}(u, v)}{\sum_{u=0}^{M-1} \sum_{v=0}^{N-1} \text{Magnitude}(u, v)} \tag{A.3}$$

Finally, as shown in A.1, we calculate the entropy of the normalized PSD values and compute the complexity metric.

$$\text{Entropy (Complexity)} = -\sum_{u=0}^{M-1} \sum_{v=0}^{N-1} \text{PSD}_{\text{normalized}}(u, v) \log_2 (\text{PSD}_{\text{normalized}}(u, v)) \tag{A.4}$$

We observe that the imaginary component of FFT for the bottom design region shows a dispersed behavior which indicates the existence of higher frequency components. Moreover, the magnitude spectrum of the top design region has only a bright spot in the middle. In contrast, the bottom magnitude spectrum has nonzero pixel values around the bright point in the center. As a result of this, the uncertainty of the bottom design region is computed as higher. This high entropy indicates that the bottom design region has a more complex pattern compared to the upper one with a simpler and more uniform shape.

A.2 Predicted and 3D FDTD Fields

The visuals of the predicted field and 3D FDTD field as well as their difference are plotted in the following figures.

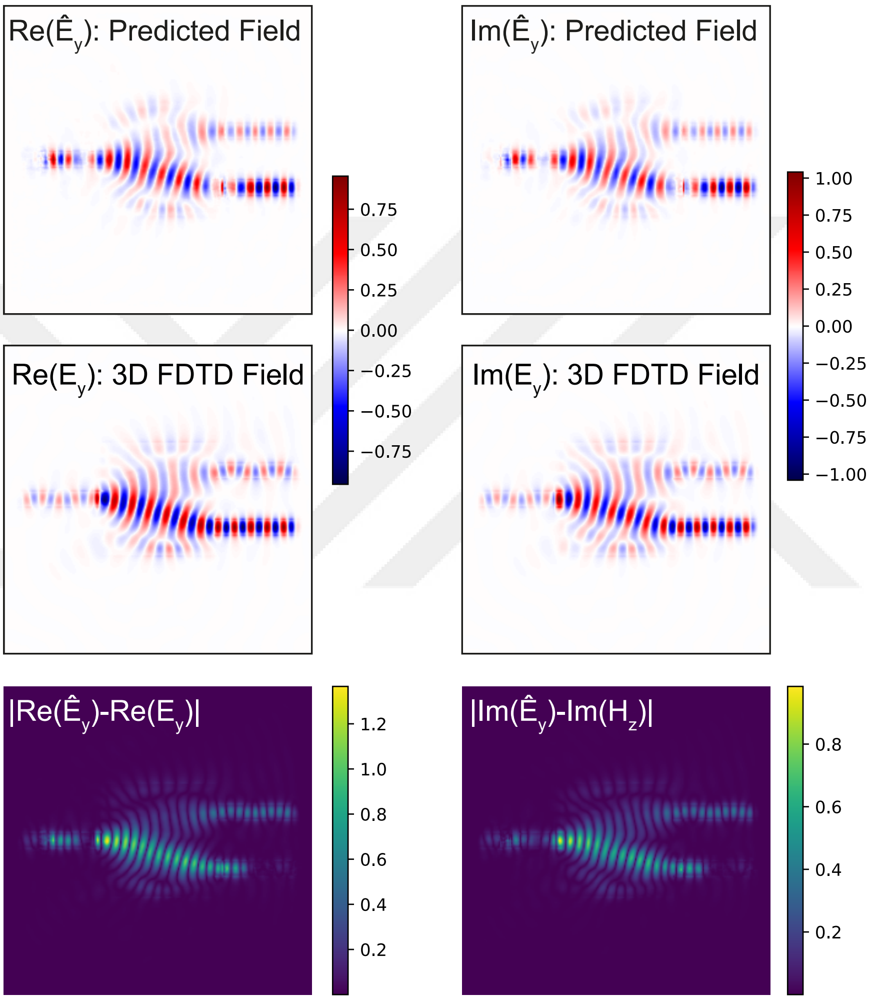


Figure A.2: Real and imaginary parts of y-component of the predicted and 3D FDTD field for the 10:90 power splitter design. The bottom row shows the absolute difference between predicted and ground truth values.

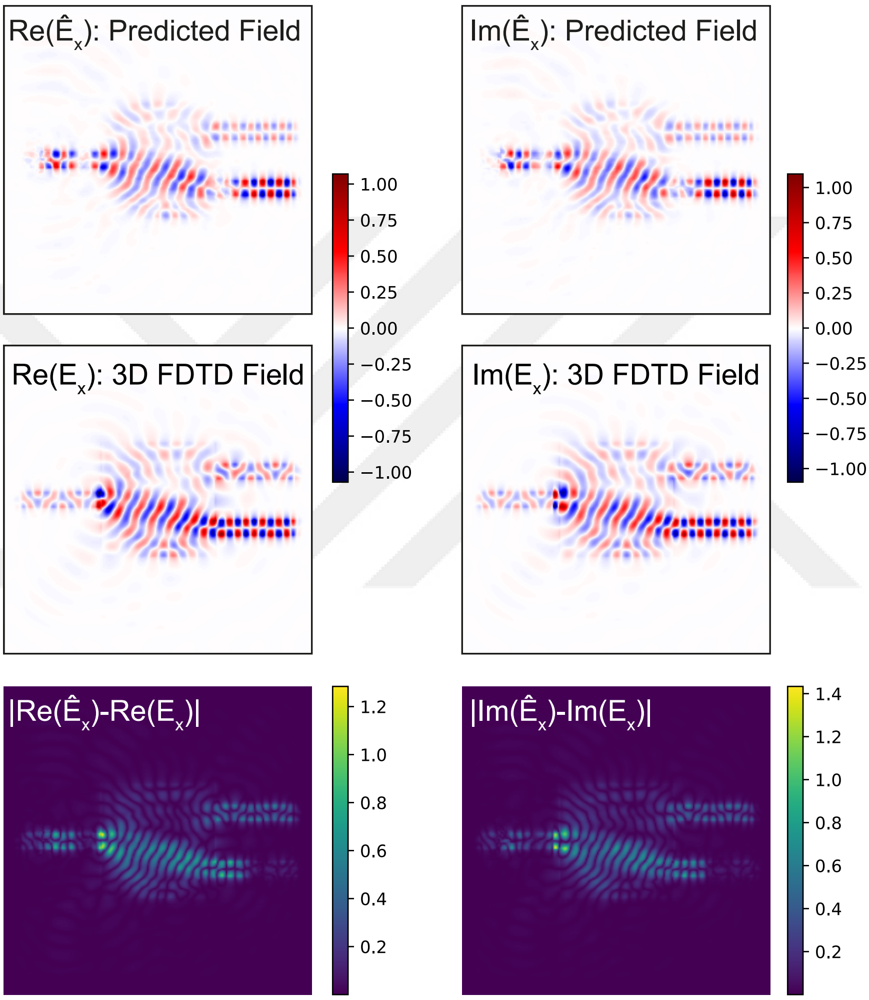


Figure A.3: Real and imaginary parts of x-component of the predicted and 3D FDTD field for the 10:90 power splitter design. The bottom row shows the absolute difference between predicted and ground truth values.