



Silicon Photonic Bragg-Based Devices : Hardware and Software

Mémoire

Jonathan Cauchon

Maîtrise en génie électrique - avec mémoire
Maître ès sciences (M. Sc.)

Québec, Canada

Silicon Photonic Bragg-Based Devices: Hardware and Software

Mémoire

Jonathan Cauchon

Sous la direction de:

Wei Shi, directeur de recherche

Résumé

L'avènement de la photonique intégrée a attiré beaucoup de recherche et d'attention industrielle au cours des deux dernières décennies, plusieurs croyant qu'il s'agit d'une révolution équivalente à la microélectronique. Tout en tirant parti des procédés de fabrication de masse hérités de la microélectronique, la photonique sur silicium est compacte, éconergétique et permet l'intégration complète de dispositifs et de circuits photoniques à l'échelle nanométrique pour des applications cruciales dans les télécommunications, la détection et le calcul optique.

À l'instar des débuts de la microélectronique, les efforts de recherche actuels en photonique sur silicium sont principalement consacrés à la proposition, à la conception et la caractérisation de composants standardisés en vue d'une éventuelle intégration de masse dans des circuits photoniques. Les principaux défis associés à ce développement comprennent la complexité de la théorie électromagnétique dans le fonctionnement des dispositifs, les variations et les non-uniformités du procédé de fabrication limitant les performances, et les ressources informatiques considérables nécessaires pour modéliser avec précision des circuits photoniques complexes. Dans ce mémoire, ces trois limitations sont abordées sous forme de contributions de recherche originales.

Basées sur des dispositifs photoniques sur silicium et l'apprentissage machine, les contributions de ce mémoire concernent toutes les réseaux de Bragg intégrés, dont le principe de fonctionnement de base est la réflexion optique sélective en fréquence. Premièrement, un nouveau filtre optique double-bande basé sur les réseaux de Bragg multimodes est introduit pour des applications dans les télécommunications. Deuxièmement, une nouvelle architecture de filtre accordable basée sur un coupleur contra-directionnel à étage unique avec un dispositif de micro-chauffage segmenté permettant des profils de température arbitraires démontre une accordabilité de la bande passante record et des capacités de compensation des erreurs de fabrication lorsqu'opérée par un algorithme de contrôle. Troisièmement, un modèle d'apprentissage machine basé sur un réseau de neurones artificiels est introduit et démontré pour la conception de coupleurs contra-directionnels et le diagnostic de fabrication, ouvrant la voie à la production de masse de systèmes photoniques intégrés basée sur les données.

Abstract

The advent of integrated photonics has attracted a lot of research and industrial attention in the last two decades, as it is believed to be a hardware revolution similar to microelectronics. While leveraging microelectronics-inherited mass-production-grade fabrication processes for full scalability, the silicon photonic paradigm is compact, energy efficient and allows the full integration of nano-scale optical devices and circuits for crucial applications in telecommunications, sensing, and optical computing.

Similar to early-day microelectronics, current research efforts in silicon photonics are put toward the proposal, design and characterization of standardized components in sights of eventual black-box building block circuit design. The main challenges associated with this development include the complexity of electromagnetic theory in device operation, the performance-limiting fabrication process variations and non-uniformities, and the considerable computing resources required to accurately model complex photonic circuitry. In this work, these three bottlenecks are addressed in the form of original research contributions.

Based on silicon photonic devices and machine learning, the contributions of this thesis pertain to integrated Bragg gratings, whose basic operating principle is frequency-selective optical transmission. First, a novel dual-band optical filter based on multimode Bragg gratings is introduced for applications in telecommunications. Second, a novel tunable filter architecture based on a single-stage contra-directional coupler with a segmented micro-heating device allowing arbitrary temperature profiles demonstrates record-breaking bandwidth tunability and on-chip fabrication error compensation capabilities when operated by a control algorithm. Third, an artificial neural network-based machine learning model is introduced and demonstrated for large-parameter-space contra-directional coupler inverse design and fabrication diagnostics, paving the way for the data-driven mass production of integrated photonic systems.

Contents

| | |
|--|-----|
| Résumé | ii |
| Abstract | iii |
| Contents | iv |
| List of Tables | vi |
| List of Figures | vii |
| List of Acronyms | xi |
| Foreword | xii |
| Introduction | 1 |
| 1 The Silicon-on-Insulator Platform | 3 |
| 1.1 Motivation | 3 |
| 1.2 Waveguides | 4 |
| 1.3 Directional Couplers | 6 |
| 1.4 Bragg Gratings | 8 |
| 1.5 Optical Input and Output | 11 |
| 2 Contra-Directional Couplers | 14 |
| 2.1 Introduction | 14 |
| 2.2 Operation Principle and Theory | 14 |
| 2.3 Simulation | 18 |
| 2.4 Apodization | 23 |
| 2.5 Chirp | 24 |
| 2.6 Flexible WDM Filters | 27 |
| 3 Machine Learning | 34 |
| 3.1 Introduction | 34 |
| 3.2 Concepts and Methods | 34 |
| 3.3 Artificial Neural Networks | 37 |
| 3.4 Applications in Photonics | 39 |
| 4 Dual-Band Optical Filters Using Integrated Multimode Bragg Gratings | 41 |
| Résumé | 41 |

| | |
|--|-----------|
| Abstract | 41 |
| 4.1 Introduction | 41 |
| 4.2 Design and fabrication | 42 |
| 4.3 Results | 43 |
| 4.4 Conclusion | 44 |
| 5 Thermally-Chirped Contra-Directional Couplers for Residue-Less, Bandwidth-Tunable Bragg Filters with Fabrication Error Compensation | 45 |
| Résumé | 45 |
| Abstract | 45 |
| 5.1 Introduction | 46 |
| 5.2 Design | 47 |
| 5.3 Results | 50 |
| 5.4 Conclusion | 53 |
| 6 BraggNet: Complex Photonic Integrated Circuit Reconstruction Using Deep Learning | 54 |
| Résumé | 54 |
| Abstract | 55 |
| 6.1 Introduction | 55 |
| 6.2 Problem Definition | 57 |
| 6.3 Deep Learning Model | 60 |
| 6.4 Results | 61 |
| 6.5 Conclusion | 69 |
| Conclusion | 71 |
| Bibliography | 72 |

List of Tables

| | | |
|-----|--|----|
| 1.1 | The required phase shift between the left- and right-side corrugations to employ on an integrated Bragg grating to achieve coupling between TE modes. The same entries also apply to TM modes. | 11 |
| 2.1 | Summary of the model's input parameters | 21 |
| 6.1 | Summary of all trained models' number of parameters and performance on the test dataset. | 62 |

List of Figures

| | | |
|-----|--|----|
| 1.1 | (a) Cross-section schematic of an SOI strip waveguide. (b) Electric field intensity distributions and key parameters of the fundamental TE and TM modes of a 500-nm-wide waveguide, at a wavelength of 1550 nm. | 5 |
| 1.2 | Simulated Effective indices of supported modes for the SOI strip waveguide, as a function of the waveguide width. Source: [12]. | 6 |
| 1.3 | (a) Mode localization along the device. (b) Power splitting of the device as a function of the coupler length, at an operation wavelength of 1550 nm, for two 500-nm-wide waveguides separated by a gap of 200 nm. Source: [12]. | 7 |
| 1.4 | Schematic of an integrated Bragg grating filter. Source: [12]. | 8 |
| 1.5 | Example spectral response of an integrated Bragg grating, showing the central wavelength and bandwidth. Source: [12]. | 9 |
| 1.6 | Schematic and field distributions of the first three TE modes of a multimode waveguide. | 10 |
| 1.7 | (a) Fibre-chip focusing grating coupler SEM image [8]. (b) Grating coupler TE insertion loss spectrum [18]. (c) Nano-taper edge coupler and its polarization-dependent insertion loss spectrum [8]. (d) SWG edge coupler and its polarization-dependent insertion loss spectrum [8]. | 12 |
| 2.1 | Scheme of a uniform contra-directional coupler grating showing the amplitude of the electric fields coming in and out of each port. | 15 |
| 2.2 | CDC simulation flow. Some input parameters are used to perform 3D interpolation using the pre-simulated waveguide database. The resulting propagation constants are fed to the transfer matrix model which carries out the computation, and outputs the spectral response. | 18 |
| 2.3 | Schematic drawing of the contra-directional coupler device, showing the four ports and a sample grating section. The inset shows a detailed view of the grating. | 19 |
| 2.4 | Lumerical MODE simulation result showing the two-waveguide system's first two supported optical TE modes' electric field distributions at 1550 nm, with effective indices of 2.535 and 2.344 respectively. | 19 |
| 2.5 | Eigenmode simulation database for two-waveguide systems with a gap of 100 nm, with varying waveguide 1 and 2 widths and wavelength. Only the first mode's effective index is shown. | 20 |
| 2.6 | Sample simulation result and performance interface. "BW" represents the 3-dB bandwidth. "Avg. ref." and "Std dev." respectively represent the average and standard deviation reflection inside the 1-dB bandwidth | 22 |

| | | |
|------|--|----|
| 2.7 | Spatial profile of the uniform, gaussian and tanh apodization profiles, along with the drop response produced by each apodization type. The Gaussian profile has a gaussian constant $a = 10$, while the tanh profile has parameters $\alpha = 2$ and $\beta = 3$ | 24 |
| 2.8 | Simulated individual effect of the three types of chirp on the filter's drop spectrum. The hyperbolic tangent apodization profile is used in the simulations. | 25 |
| 2.9 | Difference between gaussian and tanh apodization profiles, including chirp-less spectrum and the effect of a 300 K/mm temperature chirp on the device's response. | 26 |
| 2.10 | Simulator result window for the nominal CDC filter design centred at 1530 nm. | 28 |
| 2.11 | Schematic and spectral response of the considered CDC-based WDM system. | 29 |
| 2.12 | Schematic of the tunable CDC filter presented in [4], where two CDCs are cascaded and heated independently to shift their responses and tune the system's bandwidth. | 30 |
| 2.13 | Summary of the reconnected waveguide solution and its experimental results on fabricated devices. | 31 |
| 2.14 | Summary of the three-CDC solution . Two key measurements are shown for the drop and thru spectra of the filter during tuning. | 32 |
| 3.1 | Training process of a supervised machine learning model. | 37 |
| 3.2 | Schematic representation of (a) a neuron and its synaptic connections, and (b) a neural network comprising three hidden layers (5 layer total) and its synaptic connections (weights). | 38 |
| 4.1 | (a) Simulated TE mode profiles, (b) simulated mode-dependent effective indices and corresponding Bragg wavelengths at 1550 nm as a function of waveguide width. (c) Reflection bandwidth as a function of corrugation width. | 43 |
| 4.2 | (a) Apodization profile of the fabricated design, (b) schematic of the fabricated device including apodized multimode Bragg grating and asymmetric y-branches used for characterization purposes, and (c) scanning electron microscope image of the fabricated device. | 44 |
| 4.3 | Measured (a) thru-port and (b) drop-port responses. | 44 |
| 5.1 | Operational scheme of fully-tunable CDC-based filters. (a) Cascaded filters as proposed in [4]. (b) The segmented micro-heater solution proposed in this work. The heater dimensions are not at scale. | 47 |
| 5.2 | (a) Difference between gaussian and tanh apodization profiles. Effect of a 300 K/mm chirp on a (b) Gaussian-apodized and a (c) tanh-apodized device. | 48 |
| 5.3 | Simulated demonstration of tunability, showing how the application of electrical powers (a-b) creates a temperature profile (c-d), which in turn tunes the response's central wavelength (e) and bandwidth (f). The light blue region shows the spatial span of the CDC being tuned. | 49 |
| 5.4 | Passive drop response of the CDC device. In-band notches limit the bandwidth to 10.7 nm, while the out-band rejection band is 13.2 nm wide. | 50 |
| 5.5 | Visual representation of the cost criterion used for algorithmic tuning. The cost is calculated as the root-mean-squared error in the blue-highlighted region. This measurement has a cost of 1.43. | 51 |

| | | |
|-----|---|----|
| 5.6 | Experimental demonstration of the Nelder-Mead algorithm used to tune the CDC response, with target center wavelength and bandwidth of 1587 and 14 nm respectively. (a) The convergence of the algorithm showing the cost (defined as per Fig. 5.5) at each iteration. Sample states of the filter tuning during the optimization process, showing the 8 voltages applied to the segmented micro-heater's elements and the measured spectra after 0 (b,f), 2 (c,g), 6 (d,h) and 20 (e,i) iterations. | 52 |
| 5.7 | Experimental demonstration of the left- and right-edge band tunability of the device, shown on the drop port response. Various right edge locations are shown for left edges located around (a) 1580 nm, (b) 1582 nm, (c) 1584 nm, and (d) 1586 nm. All responses were measured from the same device. | 53 |
| 6.1 | Schematic of a non-uniform CDC device and one of its segments. | 58 |
| 6.2 | Three samples from the created dataset. (a) Pseudo-randomly-generated grating physical parameters including the number of grating periods (N), the Gaussian apodization profile along the grating ($\lambda_B(z)$), and the linear coupling power profile along the grating ($\kappa(z)$), where z is the normalized propagation axis of light traveling inside the device. (b) The simulated spectral responses of the CDC devices with added noise, including the drop response ($ r ^2$) and the drop group delay (T_g). | 60 |
| 6.3 | (a) A naive approach to the problem, in which all outputs are decoupled. (b) A more complete approach, in which outputs are considered interdependent. The direction of the arrows indicates the flow of relevant information that can be used to reach an accurate solution. | 61 |
| 6.4 | General architecture of the BraggNet model. Six sub-models are trained to achieve either high prior or posterior performance. The drop amplitude and group delay spectra first enter the naive (prior) models in order to obtain zeroth-order estimations. Those estimations are then fed into the complete (posterior) models to obtain higher-accuracy estimations. | 62 |
| 6.5 | Random examples of BraggNet's reconstructions (predictions) of test data and comparison with the actual design values. N and \tilde{N} are the design number of periods and reconstructed number of periods, respectively. | 63 |
| 6.6 | Effect of the measuring instrument's noise floor on the reconstruction performance of BraggNet. The dashed line shows the best-obtained performance without any noise applied to the dataset. | 64 |
| 6.7 | Samples of the experimental-oriented dataset showing (a) the coupling and phase-matching condition profiles of the device along the propagation direction, and (b) the amplitude and group delay spectral responses of the fabrication-distorted devices. | 66 |
| 6.8 | Adaptation of BraggNet for applications on experimental data. The number of grating period remains unhindered by fabrication, which makes it known from the design. N -models can hence be replaced by the known value of 1000. | 66 |

- 6.9 Demonstration of BraggNet applied on experimental data. (a) three conditioned measurement of a CDC device under different thermal chirp conditions, including the measured drop and weighted group delay spectra. (b) BraggNet prediction of the apodization and chirp profile of the experimental data. (c) Comparison between the original measurements and the output of the transfer matrix model fed with BraggNet’s profile prediction for the three measurements, including the drop and weighted group delay spectra. The central wavelength was offset manually to match the experimental and reconstructed spectra. . . . 68

List of Acronyms

| | |
|---------------|---|
| OFC | Optical Fiber Communications |
| IEEE | Institute of Electrical and Electronics Engineers |
| CMOS | Complementary Metal Oxide Semiconductor |
| SOI | Silicon-On-Insulator |
| BOX | Buried Oxide |
| E-beam | Electron Beam |
| TE | Transverse Electric |
| TM | Transverse Magnetic |
| CMT | Coupled-Mode Theory |
| I/O | Input/Output |
| SWG | Sub-Wavelength Grating |
| WG | Waveguide |
| 2D | Two-Dimensional |
| RAM | Random Access Memory |
| FBG | Fibre Bragg Grating |
| SLSR | Side-Lobe Suppression Ratio |
| CDC | Contra-Directional Coupler |
| UV | Ultraviolet |
| WDM | Wavelength-Division Multiplexing |
| ML | Machine Learning |
| GAN | Generative Adversarial Networks |
| DL | Deep Learning |
| MMBG | Multimode Bragg Grating |
| LAN | Local Area Network |
| AMF | Advanced Micro-foundry |
| RMSE | Root-Mean-square Error |
| GPU | Graphics Processing Unit |
| LP | Layer Peeling |
| AI | Artificial Intelligence |
| TMM | Transfer Matrix Method |
| NN | Neural Network |
| MAE | Mean absolute error |

Foreword

This work contains the integral reproduction of one published conference paper, one published journal article, and one submitted journal article. The entirety of the publications were reprinted in this work, with slight modifications to one journal article.

Chapter 4 was published in the proceedings of The Optical Fiber Communications Conference (OFC) in June 2020 [1] and is co-authored by Wei Shi. It is entitled "Dual-Band Optical Filters Using Integrated Multimode Bragg Gratings." I am responsible for the concept, device design, characterization, and manuscript. Wei Shi contributed to the manuscript. No modifications were made in the reproduction.

Chapter 5 was published in The Optical Society of America's Optics Letters Vol. 46, issue 3 (2021) [2] and it is entitled "Thermally-Chirped Contra-Directional Couplers for Residue-Less, Bandwidth-Tunable Bragg Filters with Fabrication Error Compensation." It was written in collaboration with Jonathan St-Yves and Wei Shi. I am responsible for the concept, device design, characterization, and manuscript. Jonathan St-Yves offered advice for the micro-heater geometry design based on previous simulations of his, and Wei Shi contributed to the problem definition and manuscript. Section titles were added to the original version of the article to improve the readability.

Chapter 6 was submitted for publication at IEEE Photonics Society's Journal of Selected Topics in Quantum Electronics in a call for papers for a special edition on Industry 4.0 in January 2021 [3]. It is entitled "BraggNet: Complex Photonic Integrated Circuit Reconstruction Using Deep Learning" and was written in collaboration with Jean-Michel Vallée, Jonathan St-Yves and Wei Shi. Jean-Michel Vallée, Jonathan St-Yves and myself are responsible for the concept idea. I am responsible for building and training the model and writing the article. Wei Shi contributed to the concept, problem definition, and writing the manuscript. No modifications were made in the reproduction.

Introduction

Similar to microelectronics, the goal of silicon photonics is to miniaturize and integrate nano-scale components on chips, allowing to create complex photonic circuits exhibiting small footprints, low energy consumption and low production costs for applications in telecommunications [4], bio-sensing [5], quantum information [6] and optical computing [7]. This great potential has led to tremendous research efforts in the last decades, where research groups all around the globe are working on record-performance components and circuits. The main challenges associated with the development of a standardized integrated platform have to do with the complex wave nature of electromagnetic theory making novel device design somewhat counter-intuitive, the fabrication processes bringing about feature size limitations and chip-wide fabrication errors and variability [8], and the computing and time resources required for the thorough simulation of complex components and circuits.

This work addresses all three of these challenges in the form of original contributions to enrich engineers' expertise in the design, simulation and fabrication assessment of Bragg-based silicon photonic devices. Bragg gratings offer flat-top, broadband frequency-selective filtering functionalities, which are crucial to many applications ranging from telecommunications [9] to quantum information [6]. This thesis is organized as follows. Three introductory chapters lay some conceptual foundation and the three subsequent chapters present original contributions to the study field of integrated Bragg gratings.

Chapter 1 introduces the research field of silicon photonics, exposing the motivation behind the integrated paradigm and introducing a handful of basic building-block components serving as a foundation for the more complex devices comprised in the original contributions. In Chapter 2, the contra-directional coupler is introduced, its underlying mathematical model is thoroughly carried out, and its numerical simulation software is presented. Additionally, an application of bandwidth-tunable contra-directional couplers in telecommunications is presented, along with its main limitation – band residue loss. Chapter 3 introduces the field of machine learning by defining fundamental concepts and methods, and more particularly the architecture of artificial neural networks.

Chapter 4 demonstrates, in concept and in experiment, the use of the previously-defined building-block devices to create a frequency-selective dual-stop/pass-band filter for applica-

tions in telecommunications. Chapter 5 demonstrates a broadband add-drop filter by employing a single-stage contra-directional coupler device and a novel segmented micro-heating architecture allowing arbitrary temperature profiles, achieving high bandwidth tunability and fabrication error compensation when operated by a control algorithm. In Chapter 6, a deep-learning, neural network-based machine learning model is presented, in an effort to rapidly reconstruct contra-directional coupler designs from their frequency response. The deep learning model is shown suitable for inverse design for a wide variety of design parameters, and shows promising performance for fabrication quality assessment when fed with experimentally-measured data.

Chapter 1

The Silicon-on-Insulator Platform

1.1 Motivation

Modern-day communication has been highly impacted by the advent of the Internet and the large amount of data transfer it requires. The evolution of Internet has caused a gradual paradigm shift starting from simple communications network to what is known as the Internet of Things, where many of our day-to-day objects are connected to the Internet and require constant data transmission [10]. The high increase in connectivity requirements, in this big-data era where large companies seek to extract and store every last piece of information, calls for massive scaling of the optical hardware required for data transmission.

Similar to the microelectronics-enabled historical technological development in the 1970s, which allowed the massive integration of electronic components onto chips [11], the integration of photonic devices onto silicon chips has the advantage of being low-cost, energy efficient and suitable for mass production by leveraging the complementary metal oxide semi-conductor (CMOS) fabrication processes inherited from the microelectronic industry.

The promise of photonic-electronic co-integration has driven massive efforts in the last decade, and engineers and scientists have been working to develop standardized integrated photonic components capable of accomplishing key optical tasks. In this chapter, a few of these building-block devices are presented. Notably, a device called Bragg grating. Integrated Bragg gratings are created by applying periodic dielectric perturbations to a waveguide, in the form of sidewall corrugations. This effectively causes partial reflections at every local effective index change. As a result, some optical frequencies are reflected back to the grating input, while others pass through the grating. This operating principle proves useful for filtering applications. Chapters 4, 5 and 6 all utilize the basic concept of Bragg gratings to offer state-of-the-art performance in key applications. For this reason, this chapter introduces operating principle of Bragg gratings, along with other key building-block devices that are essential to the use of Bragg grating.

1.2 Waveguides

Waveguides are the fundamental building block of the SOI platform and integrated photonics in general. As opposed to optical fibres, integrated waveguides do not have a circular cross section, but rather a rectangular one. Waveguides and other optical devices are integrated on a silicon wafer, which is later diced into individual centimetre-sized photonic chips. Figure 1.1a shows a schematic view of the cross section of SOI waveguides, where the bottom layer is a thick silicon substrate whose main use is mechanical: it is almost a millimetre thick and holds the wafer together. A 2-micron SiO₂ buried oxide (BOX) layer sits on top of the silicon substrate to isolate the waveguide core from the substrate. The waveguide (or core layer) sits on top of the BOX layer and is made of silicon. During the design process, the photonic designer specifies the wafer etching pattern and during fabrication, this pattern is reproduced on the wafer to remove the waveguide layer at the desired locations. As a result, the remaining (un-etched) silicon sections can be shaped as waveguides, as displayed in Fig. 1.1a. The last fabrication step is to deposit a SiO₂ cladding, which fills all the areas where the silicon waveguide layer has been etched off. As a result, it is possible to integrate arbitrary-width waveguides and optical devices on a single SOI chip. The only fixed parameters of this process are the wafer thicknesses and some foundry-specific minimum feature size and spacing that can be achieved during the etching process.

The simplest waveguide geometry uses a single etch layer, which in turn creates rectangular waveguides commonly called strip waveguides. Depending on the strip waveguide dimensions, optical modes are allowed to propagate inside the waveguide. Due to their rectangular geometry, two main mode groups are supported by such a waveguide. Transverse electric (TE) modes have their electric field mainly oscillating along the horizontal axis, while transverse magnetic (TM) modes have their electric field mainly oscillating along the vertical axis. Due to the waveguide's finite width, those modes are also referred to as quasi-TE and quasi-TM because they bear a small orthogonal field component. Figure 1.1b shows the electric field intensity distribution of the first two fundamental modes (TE₀ and TM₀) of a 500-nm-wide strip waveguide, at a wavelength of 1550 nm.

Inside the strip waveguide, a given mode is supported if it is constantly confined inside the core and can propagate freely without suffering from substantial propagation loss. The optical field propagates inside a waveguide with a strong index contrast ($n_{\text{Si}} \approx 3.47$, $n_{\text{SiO}_2} \approx 1.44$). A given mode has a given confinement inside the waveguide core and waveguide cladding, and thus sees a weighted combination of the two refractive indices. As a result, a given mode is attributed an effective index n_{eff} valued between n_{SiO_2} and n_{Si} . This effective index is directly linked to the mode's propagation velocity inside the waveguide:

$$v_p = \frac{c}{n_{\text{eff}}}, \quad (1.1)$$

where v_p is the phase velocity and c is the speed of light in free space (299,792,458 m/s). The

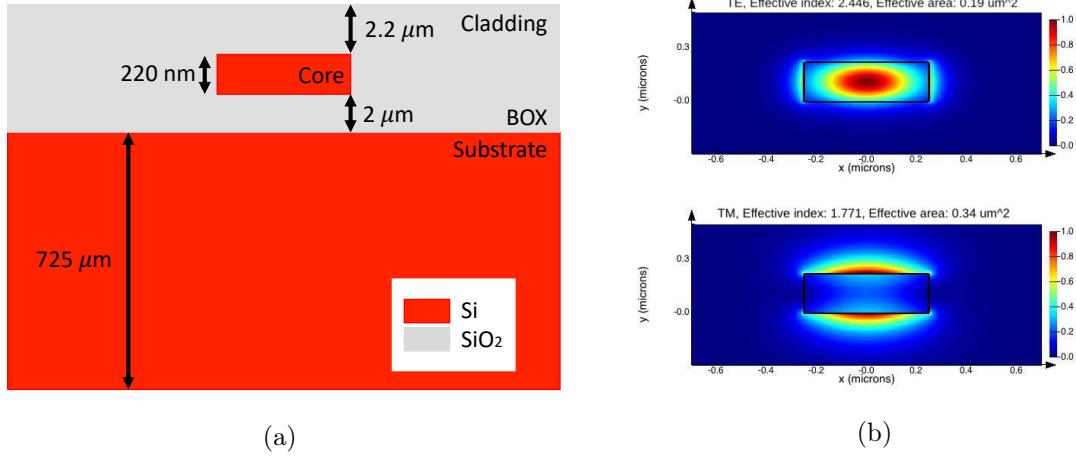


Figure 1.1 – (a) Cross-section schematic of an SOI strip waveguide. (b) Electric field intensity distributions and key parameters of the fundamental TE and TM modes of a 500-nm-wide waveguide, at a wavelength of 1550 nm.

effective index is hence a measure of a mode's propagation speed, but also its confinement. That is, the more confined the mode, the more silicon constitutes its field cross-section and hence, the higher resulting index it experiences. On Fig. 1.1b, the fundamental TE and TM modes are shown along with their effective indices and areas. As it can be seen, the TE₀ mode has the highest effective index and the smallest effective area because it is the most confined. Generally, optical modes are divided into TE and TM modes, and numbered from 0 onward with decreasing effective indices. That is, a waveguide's most confined mode is the fundamental mode, and all higher-order modes have decreasingly lower effective indices and confinement.

Figure 1.2 shows the simulated effective indices of the first four optical modes of an SOI strip waveguide at a wavelength of 1550 nm, as a function of the waveguide width. A mode is considered to be supported when its effective index is greater than the cladding's (1.44). As can be seen, increasing the waveguide width gradually increases the supported modes' effective indices, as well as the number of modes that are supported. This is once again due to confinement. Based on such simulations, strip waveguides can be designed to support a desired number of optical modes that have desired effective indices to accomplish the required tasks. The most common strip waveguide is the one used for single-mode routing. It is commonly designed to have a width of 500 nm. Under those conditions, it only supports one TE mode and one TM mode. While this width value is slightly over the strict single-mode cutoff seen in Fig. 1.2, it is in practice single mode for the most part, and using a slightly wider strip geometry is useful in avoiding the excess optical loss caused by the core side-wall roughness by offering more optical confinement.

While the phase velocity previously introduced gives a good idea of the optical propagation

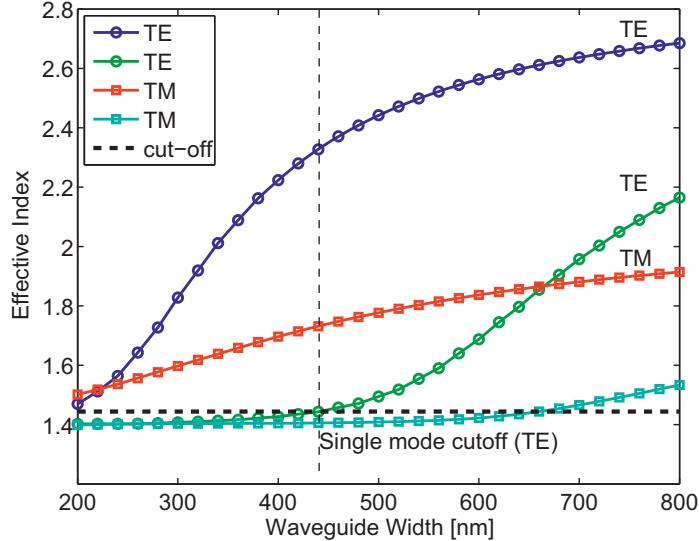


Figure 1.2 – Simulated Effective indices of supported modes for the SOI strip waveguide, as a function of the waveguide width. Source: [12].

speed at a given operating wavelength, monochromatic propagation never occurs in practice. It is thus useful to define a group velocity v_g that expresses a given mode's wave packet propagation speed:

$$v_g = \frac{\partial \omega}{\partial \beta} = \frac{c}{n - \lambda \frac{\partial n}{\partial \lambda}}, \quad (1.2)$$

where ω is the angular frequency ($\omega = 2\pi\nu = 2\pi c/\lambda$) and β is the propagation constant:

$$\beta = \frac{2\pi n_{\text{eff}}}{\lambda}. \quad (1.3)$$

With these physical quantities properly defined, SOI waveguides can be used as building blocks to create complex integrated devices and circuits to accomplish complex tasks.

1.3 Directional Couplers

Directional couplers are the most basic integrated devices that utilize mode coupling, originally described in a theory called Coupled Mode Theory (CMT) [13]. CMT studies how the spatial overlap of two optical modes can translate into optical power transfer from one mode to the other. The directional coupler is the simplest manifestation of this phenomenon. An integrated directional coupler is created by placing two parallel-run waveguides close to each other, so as to allow each individual waveguide optical mode field distribution to overlap with that of the other waveguide. This overlap results in a coupled power $|\kappa|^2$ from one waveguide to the other and an uncoupled power $|t|^2$ that makes it through the device, given by [14, 12]:

$$|\kappa|^2 = \frac{P_{\text{cross}}}{P_{\text{in}}} = \sin^2(CL), \quad |t|^2 = \frac{P_{\text{through}}}{P_{\text{in}}} = \cos^2(CL), \quad (1.4)$$

where C is the coupling coefficient and L is the length of the directional coupler. For strip waveguides, the coupling coefficient is given by

$$C = \frac{\pi \Delta n_{\text{eff}}}{\lambda}, \quad (1.5)$$

where Δn_{eff} is the difference in effective indices of the two first modes of the two-waveguide directional coupler system, commonly called super-modes.

Figure 1.3 shows the operating principle of the integrated directional coupler, where the optical mode is seen to transition from the bottom waveguide's confinement to the top waveguide's during propagation. The power splitting ratio between the through and cross ports is also shown as a function of the directional coupler length.

Based on the sinusoidal behaviour of the power transfer, the half period of the oscillation is commonly called the cross-over length l_x , defined as the coupler length allowing all the optical power to be coupled into the second waveguide. It is given by

$$l_x = \frac{\lambda}{2\Delta n_{\text{eff}}}. \quad (1.6)$$

Through the proper engineering of the individual waveguides, the gap between them and the length of the coupling region, it is hence possible to design directional couplers to achieve arbitrary wavelength-dependent splitting ratios on chip. Directional couplers leveraging unidirectional inter-mode coupling hence prove useful for many on-chip tasks requiring power splitting and combining.

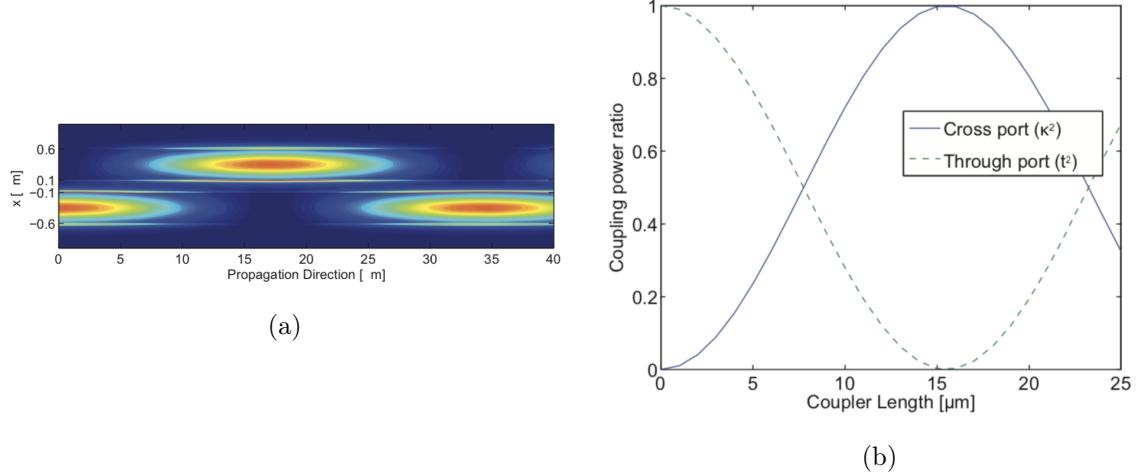


Figure 1.3 – (a) Mode localization along the device. (b) Power splitting of the device as a function of the coupler length, at an operation wavelength of 1550 nm, for two 500-nm-wide waveguides separated by a gap of 200 nm. Source: [12].

1.4 Bragg Gratings

While directional couplers are the simplest devices utilizing inter-modal coupling, Bragg gratings are more sophisticated devices that also use inter-modal coupling, but propagating inside the same waveguide in opposite directions. That is, forward-propagating modes are coupled to backward-propagating modes and vice-versa, which allows to design wavelength-dependent integrated reflective devices.

The design principle of a Bragg grating is the application of periodic dielectric perturbations on a waveguide. These dielectric perturbations change the waveguide local effective index. When applied periodically, the slight reflections undergone at every perturbation interfere constructively following a certain phase matching condition which sums up to important reflection bands in the spectral response. In the case of optical fibres, the dielectric perturbation is created by locally changing the core index by utilizing the material's photo-sensitivity [15]. In the case of SOI Bragg gratings, another strategy is employed: the waveguide core width is changed periodically, resulting in a periodic effective index profile. These core width variations are commonly called corrugations.

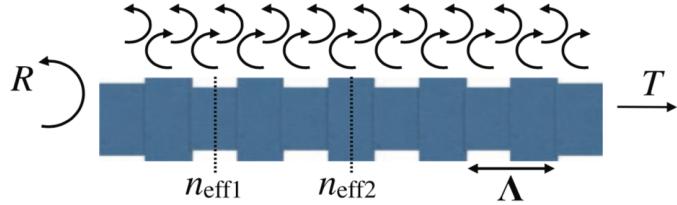


Figure 1.4 – Schematic of an integrated Bragg grating filter. Source: [12].

Figure 1.4 shows the schematic of such an integrated corrugation-enabled Bragg grating. The waveguide's width is periodically changed between a smaller and a larger width value, and the length of this periodicity (the pitch) is Λ . The two waveguide widths result in two different effective indices $n_{\text{eff},1}$ and $n_{\text{eff},2}$. As the optical mode propagates inside the grating, every interface between the two sections brings partial reflections and transmissions, which over a large number of grating periods allow for constructive and destructive interference at certain wavelengths. The constructive interference of the reflections results in a strongly reflected spectral band, while the spectral content outside of this band undergoes partial or total destructive interference which results in the power being transmitted to the through port T . Figure 1.5 shows a typical Bragg grating spectral response, where it is seen that a certain spectral band is strongly reflected while the spectral content outside of this reflection band is transmitted through the grating. Bragg gratings are hence suitable stop-band filters, and can be used for on-chip filtering tasks.

The reflection band is centred at the first-order phase-matching (or Bragg) condition, given

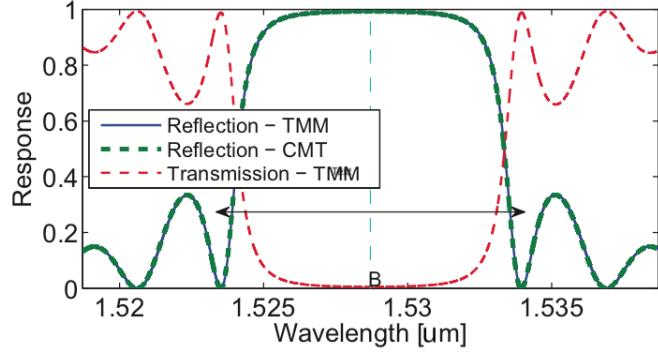


Figure 1.5 – Example spectral response of an integrated Bragg grating, showing the central wavelength and bandwidth. Source: [12].

by

$$\lambda_B = 2\bar{n}_{\text{eff}}\Lambda, \quad (1.7)$$

where \bar{n}_{eff} is the average effective index of the grating and λ_B is the central wavelength of the grating's reflection spectrum. The spectral width of the grating's reflection band is given by [16]:

$$\Delta\lambda = \frac{\lambda_B^2}{\pi n_g} \sqrt{\kappa^2 + (\pi/L)^2}, \quad (1.8)$$

where $\Delta\lambda$ is the spectral distance between the two first nulls of the reflection spectrum, κ is the coupling coefficient of the grating and L is the length of the grating. Once again referring to CMT, the coupling coefficient κ is given by the following coupling integral [17]:

$$\kappa_{ij} = \frac{\omega}{4} \iint \varphi_i^*(x,y) \Delta\varepsilon(x,y) \varphi_j(x,y) dydx, \quad (1.9)$$

giving the coupling coefficient between two modes i and j. $\varphi_i(x,y)$ and $\varphi_j(x,y)$ are the electric field spatial distributions of the two modes (similar to Fig. 1.1b). $\Delta\varepsilon(x,y)$ describes the shape of the periodic dielectric perturbation creating the grating (the corrugations). Beyond its numerical value, the coupling integral in Eq. 1.9 can be analyzed qualitatively to understand how the mode coupling mechanism operates inside the grating. In general, a single-mode Bragg grating achieves mode coupling between the forward-propagating fundamental mode and the backward-propagating fundamental mode. If the waveguide is wide enough to support more than a single mode, higher-order mode coupling is possible. Equation 1.9 can rapidly be estimated by considering symmetry. Figure 1.6 shows the spatial field distributions of the first three TE modes supported by a 1500-nm-wide strip waveguide.

As can be seen, the field distributions of TE_0 and TE_2 are symmetric, while the field distribution of TE_1 is antisymmetric. These symmetry properties can be used to assess if the coupling coefficient integrates to a non-zero value or not; the integral of an antisymmetric function is null while the integral of a symmetric function is not. In the case of the grating

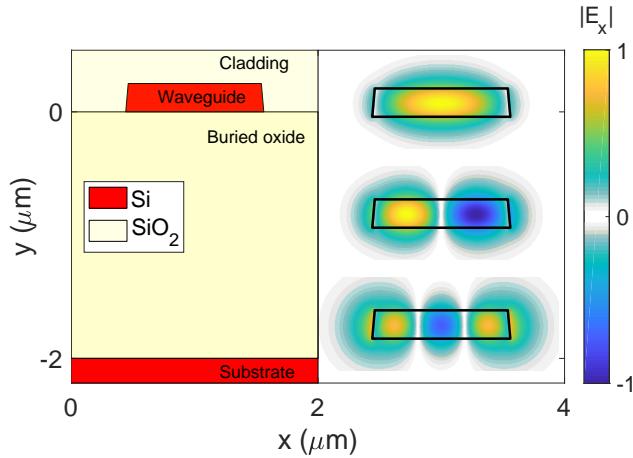


Figure 1.6 – Schematic and field distributions of the first three TE modes of a multimode waveguide.

presented in Fig. 1.4, the dielectric perturbation $\Delta\epsilon$ is symmetric. If the Bragg coupling of the forward-propagating TE_0 mode with the backward-propagating TE_0 mode is considered, then the integrand is symmetric:

$$[TE_0 \text{ (symmetric)}] \cdot [\Delta\epsilon \text{ (symmetric)}] \cdot [TE_0 \text{ (symmetric)}] = \text{symmetric integrand},$$

which in turn integrates to a non-zero coupling coefficient. This non-zero coupling coefficient makes it possible to create coupling between these two modes, which allows this Bragg grating to act as a stop-band reflecting filter. If the grating were operated in a multimode regime, and the required operation was to allow coupling between the forward propagating TE_0 mode and the backward-propagating TE_1 mode, the integrand would be antisymmetric:

$$[TE_0 \text{ (symmetric)}] \cdot [\Delta\epsilon \text{ (symmetric)}] \cdot [TE_1 \text{ (antisymmetric)}] = \text{antisymmetric integrand},$$

which would result in a null coupling coefficient, and thus no filtering reflection occurring. If the coupling between these two modes was necessary, then the grating corrugations would need to be shifted by π . That is, the left-side corrugations would be made out of phase with respect to the right-side corrugations on the waveguide. This would effectively create an antisymmetric dielectric perturbation $\Delta\epsilon$, and the coupling coefficient integrand would be symmetric:

$$[TE_0 \text{ (symmetric)}] \cdot [\Delta\epsilon \text{ (antisymmetric)}] \cdot [TE_1 \text{ (antisymmetric)}] = \text{symmetric integrand},$$

and the desired higher-mode coupling would take place. Table 1.1 shows a summary of the relative phase between the left and right corrugations that must be employed in order to achieve Bragg coupling between the different modes. Note that the same treatment applies to TM modes.

| Modes | TE ₀ | TE ₁ | TE ₂ |
|------------------------|------------------------|------------------------|------------------------|
| TE ₀ | 0 | π | 0 |
| TE ₁ | π | 0 | π |
| TE ₂ | 0 | π | 0 |

Table 1.1 – The required phase shift between the left- and right-side corrugations to employ on an integrated Bragg grating to achieve coupling between TE modes. The same entries also apply to TM modes.

Generally, the corrugation size employed on a Bragg grating is the primary factor affecting the filter bandwidth because bigger corrugations cause greater coupling coefficients κ , which in turn cause broader reflection bands. The design of integrated Bragg gratings starts from the waveguide width, where the number of modes and their effective indices can be controlled. Then, the correct corrugation symmetry must be employed for the desired regime (single-mode or multimode). Finally, the choice of grating pitch Λ and corrugation size allows the design of a device with desired central reflection wavelength and bandwidth, which finds useful applications in telecommunications and quantum information [6]. Nevertheless, there remains an issue with integrated Bragg gratings. Since they are two-port devices, the reflected optical power is returned to the input, which can be problematic for the operating laser. Moreover, if the reflected spectral content needs to be used, this reflected optical power can only be extracted by using an off-chip optical circulator, which limits the compactness of the system. In Chapter 2, the operation principle of the Bragg grating is combined with that of the directional coupler to offer a four-port add-drop filter called a contra-directional coupler, which solves this issue.

1.5 Optical Input and Output

Now that a handful of useful SOI devices allowing to route, split and filter light on the silicon chip have been introduced, it is useful to consider devices allowing to couple light in and out of the chip itself. Optical I/O is challenging because SOI waveguides are typically sub-micrometre wide and their index contrast makes for optical modes confined within a few hundred nanometres. Prior to chip injection, light is usually guided in single-mode fibres with core diameters between 8 and 10 μm . That is, a mode area reduction of a factor 600 is required to couple light from single-mode fibre to the chip [12]. This drastic difference requires creative solutions. Moreover, the inherent difference between the cylindrical geometry of the fibre and the rectangular geometry of SOI waveguides makes the task even more difficult because

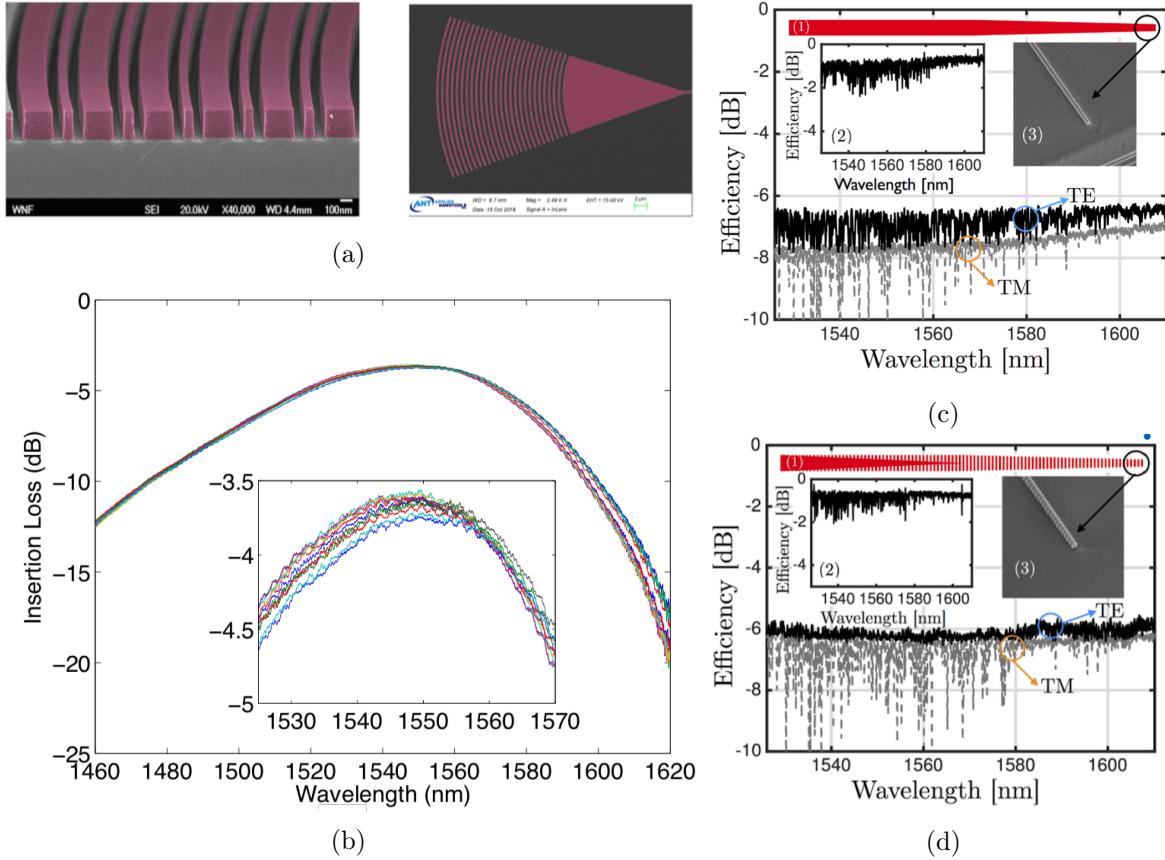


Figure 1.7 – (a) Fibre-chip focusing grating coupler SEM image [8]. (b) Grating coupler TE insertion loss spectrum [18]. (c) Nano-taper edge coupler and its polarization-dependent insertion loss spectrum [8]. (d) SWG edge coupler and its polarization-dependent insertion loss spectrum [8].

mode polarization properties are different. Consequently, broadband polarization-independent fibre-chip coupling is a tedious task. Two approaches have emerged as suitable to accomplish fibre-to-chip coupling, and are summarized in Fig. 1.7. First, vertical coupling was made possible by using fibre-chip grating couplers. These sub-wavelength-scale structures either use single-etch or dual-etch silicon features in a focusing geometry, so that a beam of light coming from an optical fibre with the right injection angle is mainly diffracted and focused inside the SOI waveguide. The advantages of this device are the fact that it can be placed anywhere on the chip, and that it allows easy, alignment-tolerant power injection. The drawbacks are that they are large devices using up substantial chip real estate, and they are highly polarization-dependent. Additionally, they have a relatively narrow injection bandwidth and a relatively high insertion loss, as seen on Fig. 1.7b.

The second approach employs edge-coupling devices, where the optical fibre is brought close to the chip edge and a lensed fibre can be used to couple directly into the chip. To ensure the modal conversion between the fibre and the 500-nm-wide silicon waveguides, a tapered section

as displayed in Fig. 1.7 is employed. Simple nano-taper devices serve to transition from a fibre's large low-confinement modes (narrow waveguide core) to narrow, highly-confined modes (wide waveguide core) found in the SOI strip waveguides. This transition is achieved either using simple waveguide inverted tapers, or using sub-wavelength grating (SWG) tapers, which allow to achieve a better mode match with the fibre. The advantages of edge couplers are that they allow broadband and polarization-independent optical insertion, as displayed in Fig. 1.7. The drawbacks of edge-coupling devices are that they require being placed on a chip edge and hence limit chip space organization, and that they are more sensitive to the fibre alignment both in position and angle.

A basic set of devices has been introduced and analyzed quantitatively as well and qualitatively. This paves the way for the following chapters that are built on the basic devices presented in this chapter to create more complex applications in the form of original contributions. The principles of operation of the basic devices presented in this chapter are essential to understanding the operation of the more complex devices presented in the subsequent chapters. The next chapter is dedicated to contra-directional couplers, which are devices that integrate waveguides, directional couplers and Bragg gratings for versatile pass/stop-band filtering.

Chapter 2

Contra-Directional Couplers

2.1 Introduction

This chapter introduces the concept of a contra-directional coupler (CDC) device, which is a 4-port add-drop filter capable of broadband, flat-top filtering and offers many degrees of freedom in its design. The chapter is organized as follows. First, the device's operating principle and physical parameters are presented. Second, the theory used to model the device is introduced and carried out. Third, a detailed explanation of the programmatic implementation of the model in the Python programming language is explained and demonstrated. Apodization and chirp are then introduced and analyzed as practical design strategies. Finally, a key application is shown in the design of a 4-channel wavelength division multiplexing system where the grid channels are flexible and bandwidth can be reallocated dynamically among the channels. A recently-demonstrated approach for such tunability is demonstrated, along with its performance and main limitation. Two potential solutions to this limitation are presented. However, conceptual and experimental evidence shows that they are not suitable for practical applications. These conclusions pave the way for Chapter 5 which proposes a novel tuning approach showing important upsides over its predecessor.

2.2 Operation Principle and Theory

As indicated by its name, the contra-directional coupler offers the combined functionalities of a directional coupler and a Bragg grating. It allows for wavelength selectivity without the need for an optical circulator. As such, it can be described as a 4-port add-drop stop/pass band filter. To accomplish this functionality, two antisymmetric Bragg gratings are run parallel to one another to allow for their respective optical modes to experience some overlap. Figure 2.1 shows a basic scheme of a contra-directional coupler grating, where the four ports (input, drop, thru and add) are displayed. Inside the grating, multiple optical coupling mechanisms occur. First, there is optical coupling between the two waveguides, allowing to transfer optical power

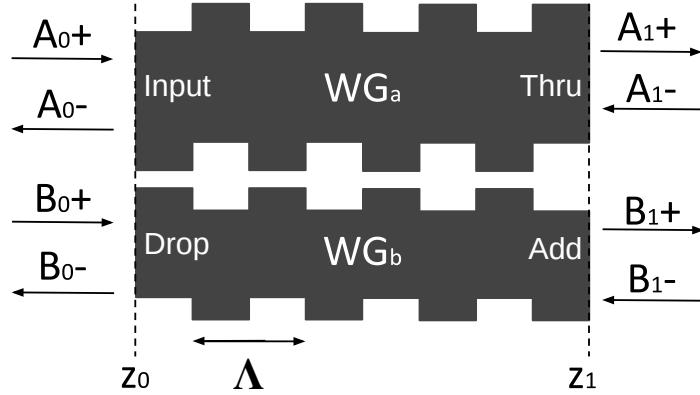


Figure 2.1 – Scheme of a uniform contra-directional coupler grating showing the amplitude of the electric fields coming in and out of each port.

from one to the other. Second, Bragg coupling between forward- and backward-propagating modes in the system create a wavelength selectivity phenomenon.

To model the coupling mechanisms occurring inside contra-directional couplers, coupled-mode theory can be used to describe the system and lay the coupled differential equations that need to be solved. A uniform contra-directional coupler is first considered, where the grating geometry is constant along the whole length of the device. Figure 2.1 shows the naming convention used for the model. Any optical waves located in the first waveguide (WG_a) has a field amplitude A , and any wave located in the second waveguide (WG_b) has a field amplitude B . Additionally, optical waves travelling from left to right are tagged as positive (+), while all waves travelling from right to left are tagged negative (-). The dynamics for the wave propagation between points z_0 and z_1 is considered.

The transverse modes of each waveguide are considered to have normalized electric field distributions $\mathbf{E}_a(x,y)$ and $\mathbf{E}_b(x,y)$. The overall electric field distribution inside the grating is hence fully described by

$$\begin{aligned} \mathbf{E}(x,y,z) = & \left[A^+(z)e^{-j\hat{\beta}_a z} + A^-(z)e^{j\hat{\beta}_a z} \right] \mathbf{E}_a(x,y) \\ & + \left[B^+(z)e^{-j\hat{\beta}_b z} + B^-(z)e^{j\hat{\beta}_b z} \right] \mathbf{E}_b(x,y), \end{aligned} \quad (2.1)$$

where $\hat{\beta}$ is the complex propagation constant containing both the propagation constant constant and loss ($\hat{\beta} = \beta - j\alpha$) for either waveguide.

In order to properly describe the physical system, two quantities need to be defined: the detuning from the Bragg condition and the coupling coefficients. The Bragg detuning arises from the phase-matching condition of the grating with a pitch Λ . Starting from Bragg's law in terms of propagation constants ($\beta_{Bragg} = \pi/\Lambda$), the detuning from the Bragg condition in

its complex form is given by

$$\begin{aligned}\Delta\hat{\beta} &= \hat{\beta} - \beta_{Bragg} \\ &= \beta - j\alpha - \frac{\pi}{\Lambda}.\end{aligned}\tag{2.2}$$

Inside the grating, the optical modes supported in waveguides a and b can either undergo optical coupling between one another, or self coupling. The coupling coefficient of the electric field distributions of two modes m and n is given by the overlap integral [14]:

$$\kappa_{mn} = \frac{\omega}{4} \iint \varphi_m^*(x,y) \Delta\varepsilon(x,y) \varphi_n(x,y) dydx,\tag{2.3}$$

where $\Delta\varepsilon(x,y)$ is once again the dielectric perturbation profile creating the coupling (the corrugation). The detuning from the Bragg condition and coupling coefficients just defined can be used to describe the system using the following coupled differential equations [19]:

$$\begin{aligned}\frac{dA^+}{dz} &= -j\kappa_{aa} A^- e^{2j\Delta\hat{\beta}_a z} - j\kappa_{ab} B^- e^{j(\Delta\hat{\beta}_a + \Delta\hat{\beta}_b)z} \\ \frac{dB^+}{dz} &= -j\kappa_{ab} A^- e^{j(\Delta\hat{\beta}_a + \Delta\hat{\beta}_b)z} - j\kappa_{bb} B^- e^{2j\Delta\hat{\beta}_b z} \\ \frac{dA^-}{dz} &= j\kappa_{aa}^* A^+ e^{-2j\Delta\hat{\beta}_a z} + j\kappa_{ab}^* B^+ e^{-j(\Delta\hat{\beta}_a + \Delta\hat{\beta}_b)z} \\ \frac{dB^-}{dz} &= j\kappa_{ab}^* A^+ e^{-j(\Delta\hat{\beta}_a + \Delta\hat{\beta}_b)z} + j\kappa_{bb}^* B^+ e^{-2j\Delta\hat{\beta}_b z}.\end{aligned}\tag{2.4}$$

The solution to this system of coupled differential equations was obtained by [20] and makes the link between the field amplitudes at z_0 and z_1 in the following way:

$$\begin{bmatrix} A^+(z_0) \\ B^+(z_0) \\ A^-(z_0) \\ B^-(z_0) \end{bmatrix} = M(z_0, z_1) \begin{bmatrix} A^+(z_1) \\ B^+(z_1) \\ A^-(z_1) \\ B^-(z_1) \end{bmatrix},\tag{2.5}$$

Where $M(z_0, z_1)$ is the 4×4 transfer matrix linking the grating segment left side to its right side. Furthermore, [20] demonstrates that the transfer matrix is given by

$$M(z_0, z_1) = e^{S_1(z_1 - z_0)} e^{S_2(z_1 - z_0)},\tag{2.6}$$

where e^A denotes the matrix exponential of A :

$$e^A = \sum_{k=0}^{\infty} \frac{1}{k!} A^k,\tag{2.7}$$

with A^k denoting the k^{th} power of the matrix A , or simply the matrix A multiplied by itself k times. S_1 and S_2 are the scattering matrices given by

$$S_1 = \begin{bmatrix} j\Delta\beta_1 & 0 & 0 & 0 \\ 0 & j\Delta\beta_2 & 0 & 0 \\ 0 & 0 & -j\Delta\beta_1 & 0 \\ 0 & 0 & 0 & -j\Delta\beta_2 \end{bmatrix},$$

$$S_2 = \begin{bmatrix} -j\Delta\beta_1 & 0 & -j\kappa_{11}e^{2j\Delta\beta_1 z_1} & -j\kappa_{12}e^{j(\Delta\beta_1+\Delta\beta_2)z_1} \\ 0 & -j\Delta\beta_2 & -j\kappa_{12}e^{j(\Delta\beta_1+\Delta\beta_2)z_1} & -j\kappa_{22}e^{2j\Delta\beta_2 z_1} \\ j\kappa_{11}^*e^{-2j\Delta\beta_1 z_1} & j\kappa_{12}^*e^{-j(\Delta\beta_1+\Delta\beta_2)z_1} & j\Delta\beta_1 & 0 \\ j\kappa_{12}^*e^{-j(\Delta\beta_1+\Delta\beta_2)z_1} & j\kappa_{22}^*e^{-2j\Delta\beta_2 z_1} & 0 & j\Delta\beta_2 \end{bmatrix}. \quad (2.8)$$

In the case of a non-uniform contra-directional coupler, where the grating geometry varies along the z direction, a more extensive transfer matrix method must be used. In that case, the total grating is separated into N_{seg} segments. Each grating segment is considered uniform and therefore a transfer matrix M can be obtained for each one. The complete system transfer matrix is obtained by individual segment matrices multiplication:

$$M_{total} = \prod_{n=1}^{N_{seg}} M_{N_{seg}-n+1}. \quad (2.9)$$

Now that the system transfer matrix M is obtained, linking the left-end side field amplitudes to the right-end side field amplitudes, it is useful to define the transformation needed to express the transfer matrix in terms of the device input, drop, thru, and add ports. Provided a 4×4 transfer matrix of the form

$$M = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}, \quad (2.10)$$

the device port amplitudes are obtained with the following matrix operation [19]:

$$\begin{bmatrix} E_{thru} \\ E_{add} \\ E_{input} \\ E_{drop} \end{bmatrix} = \begin{bmatrix} M_{11} - M_{12}(M_{22})^{-1}M_{21} & M_{12}(M_{22})^{-1} \\ -(M_{22})^{-1}M_{21} & (M_{22})^{-1} \end{bmatrix} \begin{bmatrix} A^+(z_0) \\ B^+(z_0) \\ A^-(z_{end}) \\ B^-(z_{end}) \end{bmatrix}. \quad (2.11)$$

Once the complex electric field amplitudes E_{thru} , E_{add} , E_{input} and E_{drop} are obtained, it is easy to extract the device amplitude and phase response by simply considering the norm and the phase of these quantities. In the following section, computer code implementation of this model allowing rapid and versatile simulations of the device is described.

2.3 Simulation

Now that a complete and rigorous theoretical model using coupled-mode theory and a transfer matrix method to obtain a device's spectral response has been established, it is necessary to implement the theory into a numerical model allowing rapid simulation of various devices for design and optimization. Here I present a simulation flow from geometric waveguide dimensions to output spectral response. The numerical model combines a result database obtained from commercial software simulations, and open-source software I wrote that consumes the database to complete the simulation every time it is queried. This simulation flow is shown in Fig. 2.2.

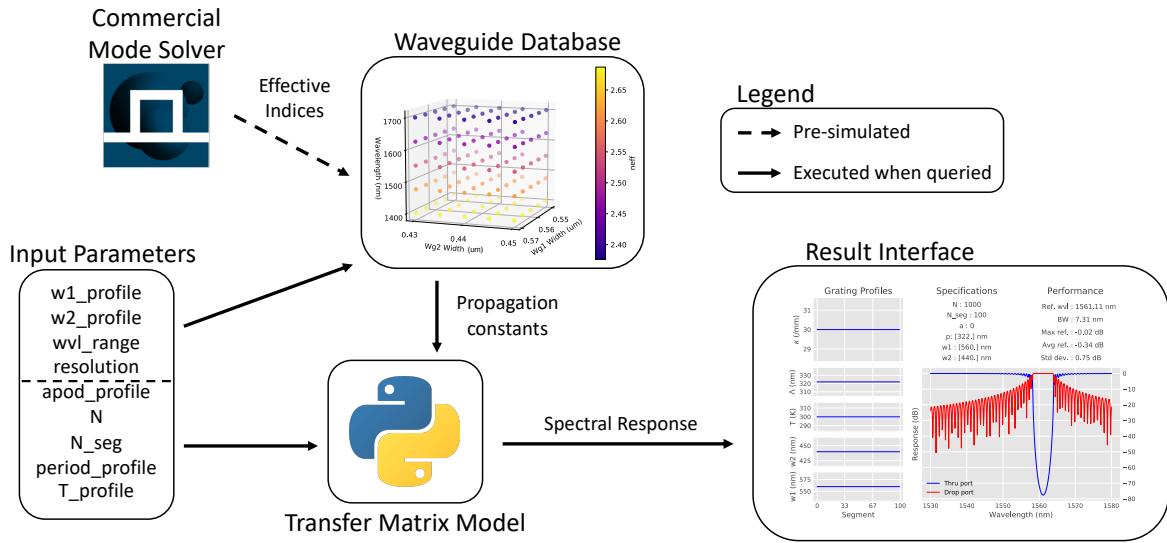


Figure 2.2 – CDC simulation flow. Some input parameters are used to perform 3D interpolation using the pre-simulated waveguide database. The resulting propagation constants are fed to the transfer matrix model which carries out the computation, and outputs the spectral response.

A schematic of the CDC device and all the parameters required for its simulation is shown in Fig. 2.3. The grating comprises N grating periods and it is sub-divided in N_{seg} grating segments, each considered uniform. The inset of Fig. 2.3 shows a close view of a segment. Each segment has its own grating pitch Λ and local temperature T , waveguide one width and corrugation size (w_1 and Δw_1), and waveguide two width and corrugation size (w_2 and Δw_2). All these quantities are individually defined for each segment forming the whole grating, and these profiles completely define the device geometry. It can then be simulated using the transfer matrix method shown in the previous section.

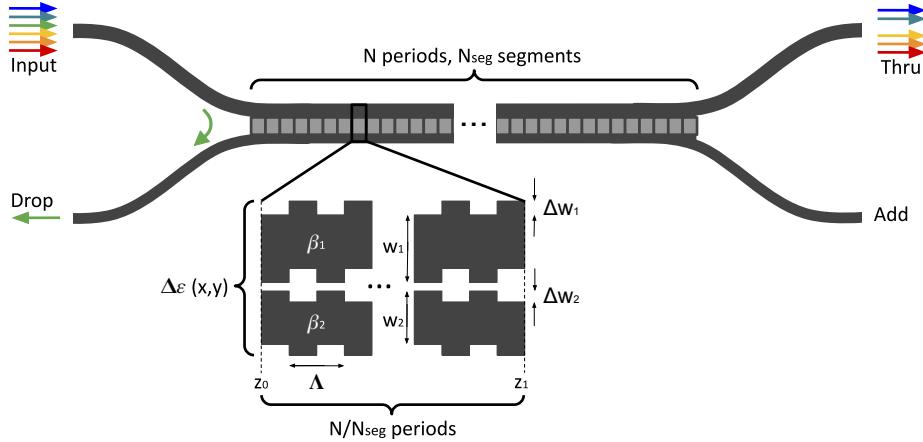


Figure 2.3 – Schematic drawing of the contra-directional coupler device, showing the four ports and a sample grating section. The inset shows a detailed view of the grating.

2.3.1 Waveguide Database

As seen in section 2.2, the inputs to the transfer matrix model are the detuning from the Bragg condition $\Delta\hat{\beta}$ of both waveguides and the coupling coefficients given by Eq. 2.3. In order to obtain the Bragg detuning, Lumerical's eigenmode solver MODE is used [21]. The 2D simulation finds all solutions to Maxwell's equations for the two-waveguide cross section presented. As shown in Fig. 2.4, the software finds the mode field distributions of all supported modes as well as their effective indices. As can be seen, the fundamental and first-higher-order modes are essentially confined in the widest and narrowest waveguide, respectively. These are usually called super-modes because they exist in a two-waveguide system but have most of their optical power in one of the waveguides. When contra-directional coupling occurs, it is between these two super-modes that the optical power is coupled, allowing certain optical wavelengths to emanate from the drop port. From the super-mode effective indices, the propagation constants

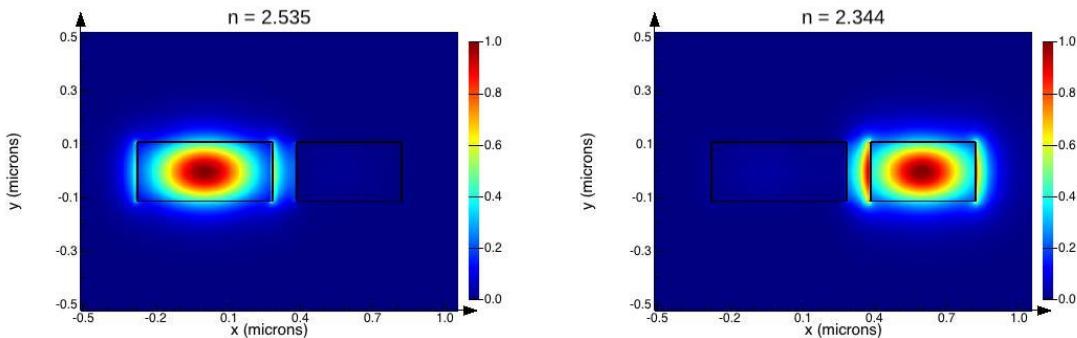


Figure 2.4 – Lumerical MODE simulation result showing the two-waveguide system's first two supported optical TE modes' electric field distributions at 1550 nm, with effective indices of 2.535 and 2.344 respectively.

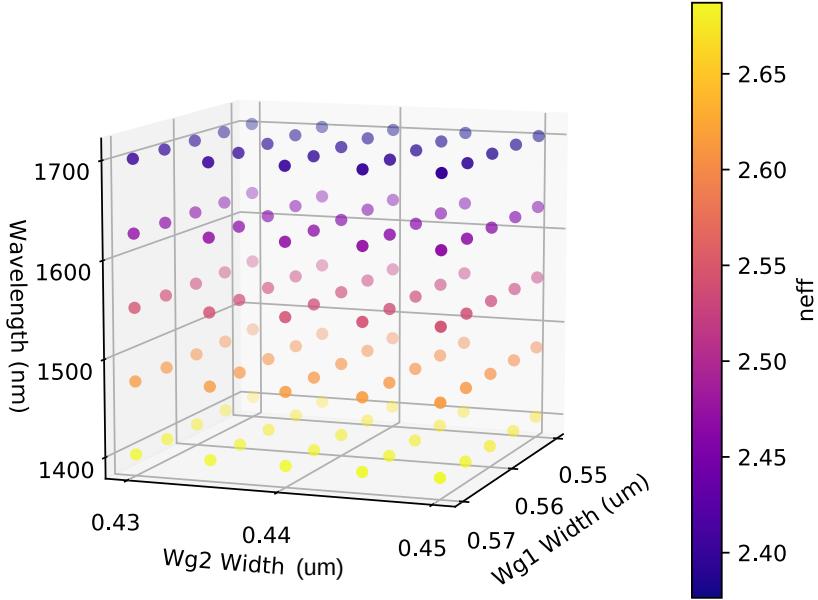


Figure 2.5 – Eigenmode simulation database for two-waveguide systems with a gap of 100 nm, with varying waveguide 1 and 2 widths and wavelength. Only the first mode’s effective index is shown.

and detunings from the Bragg condition $\Delta\hat{\beta}$ can be obtained directly.

The scripted operation of the eigenmode software allows to run parameter sweeps and obtain results for many parameter combinations. As a result, it is possible to build result databases that can later be rapidly accessed as needed. Figure 2.5 shows a sample simulation database obtained in this way, where the waveguide widths were varied from 550 to 570 nm for waveguide 1, and from 430 to 450 nm for waveguide 2. The wavelength was varied from 1400 nm to 1700 nm and the gap between the waveguides was kept constant at 100 nm.

Upon simulating many geometries, a complete database is obtained and saved in an easily-accessible file. From then on, it is easy, rapid and accurate to query the database and perform linear 3D interpolations when needed.

The second input to the transfer matrix model is the mode coupling profile $\kappa(z)$. The coupling coefficient, or coupling power, can be obtained either from eigenmode expansion simulations using Ansys-Lumerical software, or by finite-difference time domain simulations using Bloch limit conditions. Although these techniques give realistic values, they have been found to be insufficiently accurate when compared with experimental data. For this reason, The free parameter κ is manually input into the transfer matrix model. From experimental data, it is possible to correlate coupling power with experimentally-obtained filter bandwidths, which is the actual purpose of considering a coupling value to start with.

2.3.2 Model Computation

Upon obtaining a database of commercial-software-performed simulations readily producing the inputs to the transfer matrix model, the latter was implemented to complete the simulation flow. The transfer matrix model is implemented in the Python 3 programming language [22]. This open-source, object-oriented programming language offers easy integration and simple syntax. NumPy is the primary and most popular Python library for array programming [23]. It allows seamless vectorization, broadcasting and reduction of N-dimensional array objects (ndarrays). It utilizes functionalities compiled in C, and threading to parallelize operations, resulting in faster computations when compared to sequential operations (using loops).

When the user specifies the grating geometry and coupling power, the simulator queries the database to extract the super-mode effective indices $n_{\text{eff},1}$ and $n_{\text{eff},2}$, using interpolation if needed. From these, a temperature adjustment is performed based on a linear thermo-optic effect model [24]:

$$n_{\text{eff}}(T) = n_{\text{eff}, \text{300K}} + 1.87 \times 10^{-4} \cdot T, \quad (2.12)$$

where T is the specified temperature in Kelvin. From these effective indices, the propagation constants and Bragg detuning $\Delta\hat{\beta}$ of each waveguide is obtained. From this point, the transfer-matrix model is carried out exactly as explained in Section 2.2. Table 2.1 shows a summary of the simulator input parameters.

| Param. | Dimensionality | Units | Description |
|----------------|------------------------------|-----------------|--|
| N | 1×1 | - | Number of grating periods |
| N_seg | 1×1 | - | Number of grating segments |
| resolution | 1×1 | - | Number of wavelength points to compute |
| wvl_range | $\text{resolution} \times 1$ | m | Wavelength range to simulate |
| apod_profile | $\text{N_seg} \times 1$ | m^{-1} | Coupling power profile $\kappa(z)$ |
| w1_profile | $\text{N_seg} \times 1$ | m | Waveguide 1 width profile $w_1(z)$ |
| w2_profile | $\text{N_seg} \times 1$ | m | Waveguide 2 width profile $w_2(z)$ |
| period_profile | $\text{N_seg} \times 1$ | m | grating pitch profile $\Lambda(z)$ |
| T_profile | $\text{N_seg} \times 1$ | K | Temperature profile $T(z)$ |

Table 2.1 – Summary of the model’s input parameters

When all the inputs are specified by the user, the simulator creates a 4-dimensional array of shape (resolution, N_seg, 4, 4). That is, the first dimension bears wavelength dependence, the second dimension bears grating segment (z position) dependence, and the last two dimensions bear the 4×4 transfer matrices S_1 and S_2 introduced in Eq. 6.3. Numpy’s native broadcasting functionalities allow 2D matrix operations on the 4D array, where the 2D operations are performed on the last two dimensions and the remaining two dimensions are broadcast (i.e., the same 2D operation is reproduced for every slice). This avoids the use of loops and makes computations much faster.

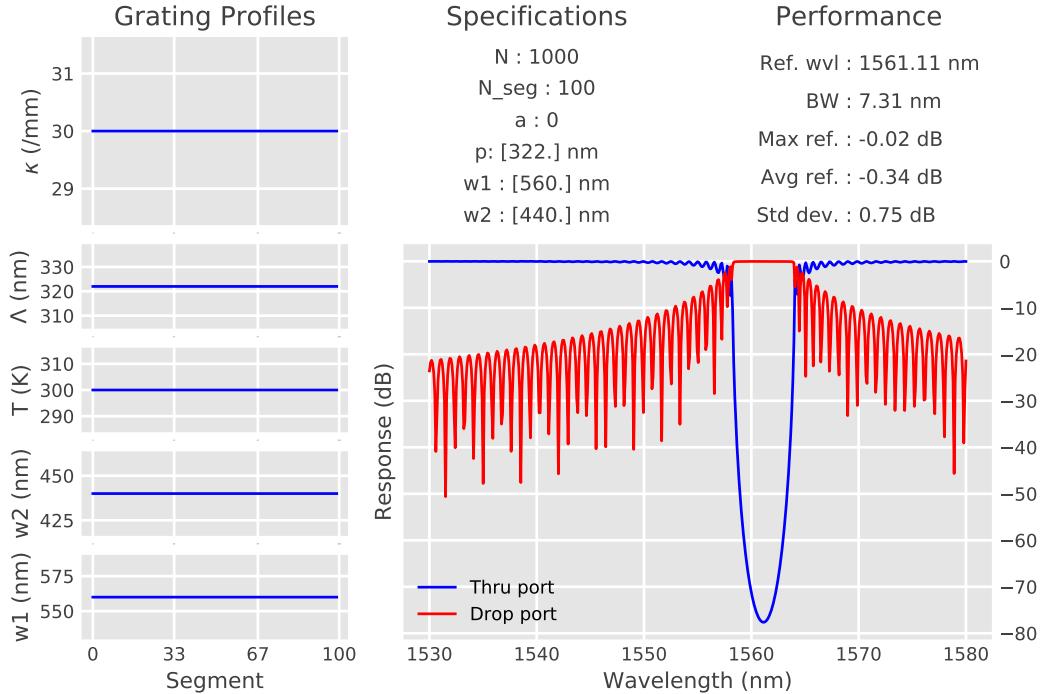


Figure 2.6 – Sample simulation result and performance interface. "BW" represents the 3-dB bandwidth. "Avg. ref." and "Std dev." respectively represent the average and standard deviation reflection inside the 1-dB bandwidth

An interface was put together to easily analyze contra-directional coupler simulations, as shown on Fig. 2.6. This interface contains all the information pertaining to the physical design (grating profiles, number of periods, etc.), the plots of the drop and thru port spectral responses, as well as some basic performance figures such as central wavelength, bandwidth and maximum reflection. In a single line of code, a user can create a CDC device with custom parameters, simulate it and obtain a visual representation of the performance. The following example uses the default device parameters:

```
device = ContraDC().simulate().displayResults()
```

This whole process takes less than two seconds on an Intel Core i5 processor with 16 GB RAM memory. Figure 2.6 shows the result of this code line, which simulates the most basic contra-directional coupler design. The model software is available open-source for anyone to use right out of the box [25]. The next two sections explain in detail what the profile parameters found in Table 2.1 represent. Namely, the design strategies of apodization and chirp are introduced and demonstrated using the numerical transfer-matrix model.

2.4 Apodization

Apodization is a design strategy which consists in slowly bringing the coupling power κ from null to a maximum value, in order to avoid an abrupt index discontinuity at the grating edge. It corresponds to the "apod_profile" model parameter. The simulation shown in 2.6 does not use apodization, as the coupling power remains constant at the maximum value:

$$\kappa_{\text{uniform}} = \kappa_{\text{max}}. \quad (2.13)$$

Consequently, the spectral response shows numerous side-lobes (out-band reflections) that affect the device's response quality. By simply making a smooth transition from 0 coupling to maximum coupling, these side-lobes can be greatly suppressed. The typical apodization profile used in FBGs and integrated gratings is the Gaussian shape, where the coupling profile is given by:

$$\kappa_{\text{Gaussian}} = \kappa_{\text{max}} e^{-a\left(\frac{z-\frac{1}{2}z_{\text{max}}}{z_{\text{max}}}\right)^2}, \quad (2.14)$$

where a is the Gaussian parameter that determines the half-maximum width of the gaussian function. It usually varies between 1 and 20, and a value of 10 typically yields sufficient post-fabrication results [26]. Additionally, another apodization profile shape can be introduced, this one following a hyperbolic tangent (tanh) function:

$$\kappa_{\text{tanh}} = \frac{\kappa_{\text{max}}}{2} \left[1 + \tanh \left(\beta \left\{ 1 - 2 \left| \frac{2z - z_{\text{max}}}{z_{\text{max}}} \right|^{\alpha} \right\} \right) \right], \quad (2.15)$$

where α and β are parameters that influence the smoothness of the function's transition near its top and near its bottom, respectively. Like the Gaussian parameter a , different sets of values can work for this type of apodization. To ensure a smooth-enough profile that is tolerant to fabrication, $\alpha = 2$ and $\beta = 3$ are typically chosen [27].

Figure 2.7 shows a visual representation of the coupling power distributions using uniform, Gaussian and tanh apodization profiles. Clearly, the mere concept of applying some type of apodization to the coupling profile greatly improves the drop response side-lobe suppression ratio (SLSR). The maximum coupling power for all three presented devices is the same (40 mm^{-1}). Employing a gaussian apodization profile drastically improves the response quality by bringing the SLSR upwards of 40 dB. While the tanh apodization profile also allows an enhanced out-of-band suppression ratio, some high side-lobes remain present near the drop band. For this reason, tanh apodization is not ideal when the phase matching condition remains uniform along the device (i.e., no chirp is present). In the presence of chirp however, the tanh apodization profile proves to be very advantageous, as will be discussed in the next section.

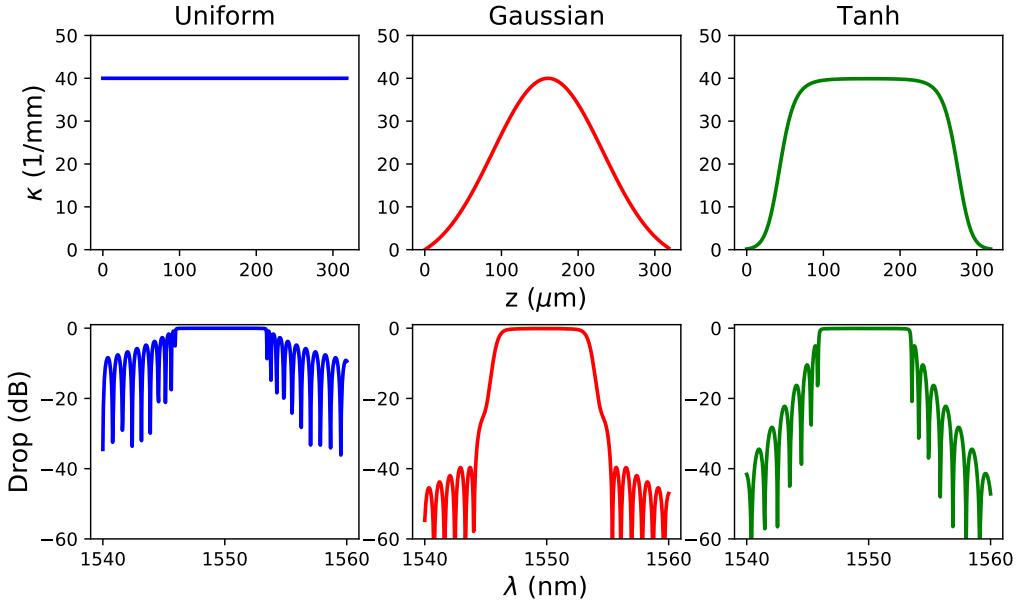


Figure 2.7 – Spatial profile of the uniform, gaussian and tanh apodization profiles, along with the drop response produced by each apodization type. The Gaussian profile has a gaussian constant $a = 10$, while the tanh profile has parameters $\alpha = 2$ and $\beta = 3$.

2.5 Chirp

The CDC bandwidth is essentially determined by its number of periods and coupling power κ , related to the structure band gap [28]. Consequently, increasing the filter bandwidth requires higher κ_{max} values, which essentially means using a smaller gap between the waveguides. However, the waveguide gap is limited by the fabrication process minimum feature spacing, making it impossible to reach bandwidth values much higher than 10-12 nm. To address this issue, sub-wavelength grating waveguide structures have been proposed to build CDCs, resulting in 3-dB bandwidths upwards of 30 nm [29, 30]. However, the small features required for sub-wavelength grating waveguides makes this approach incompatible with UV lithography and hence not suitable for production. Conversely, chirp is an approach long used in fibre Bragg gratings [27]. Since FBGs have very low index contrasts and thus small band gaps, the chirp technique is used to drastically increase FBG bandwidths at will. Chirp designates any change in the phase matching condition of a Bragg filter along its propagation axis, thereby sweeping a desired range of reflection wavelengths. It corresponds to the "w1_profile", "w2_profile", "period_profile", and "T_profile" parameters of the simulation model (see Table 2.1). As a result of this phase matching condition variation, partial reflections at different wavelengths sum up as a broadband reflection spectrum, and the bandwidth can be finely tuned by changing the phase matching condition sweep range. In the case of CDC filters, adopting a phase matching condition varying in the propagation direction z results in a z -dependent reflection

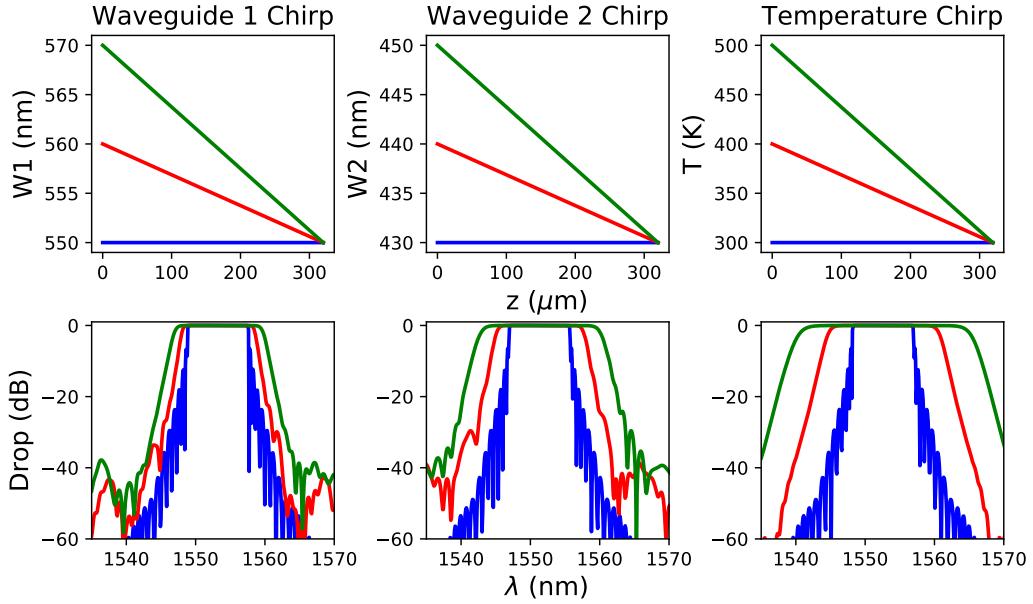


Figure 2.8 – Simulated individual effect of the three types of chirp on the filter’s drop spectrum. The hyperbolic tangent apodization profile is used in the simulations.

wavelength:

$$\lambda_B(z) = 2\Lambda(z)\bar{n}_{\text{eff}}(z), \quad (2.16)$$

which indicates that a chirp can be achieved by varying either the grating pitch Λ or the average effective index \bar{n}_{eff} . Additionally, \bar{n}_{eff} depends on a variety of factors. As shown in Fig. 2.5, it depends upon the two waveguides widths used. As per Eq. 2.12, it also depends upon the temperature because of the thermo-optic effect. This leaves us with four physical quantities that may be varied in order to achieve CDC chirp: the grating pitch, waveguide 1 width, waveguide 2 width, and temperature.

In chirped FBGs, it is usually the grating pitch Λ that is linearly chirped [27]. Since SOI CDCs have much higher waveguide index contrast, the minimum step at which the pitch can be changed according to foundry-compliant feature sizes creates big steps in the phase matching condition. For this reason, it is best to achieve chirp by varying the other three parameters: waveguide widths, and temperature. While the waveguide widths are chirped in the passive design and cannot be changed once fabricated, applying temperature profiles with micro-heating elements is done after fabrication. This combination of passive and active chirp hence allows versatile tunability of the filter central wavelength and bandwidth, simultaneously. Figure 2.8 shows the effects of chirping these physical quantities on the resulting drop response. Clearly, applying chirp to w_1 , w_2 and T effectively broadens the filter’s pass band. Spectral broadening is greater in waveguide one than in waveguide two because it is narrower, so an equivalent absolute change in width results in a greater relative change and thus, a greater

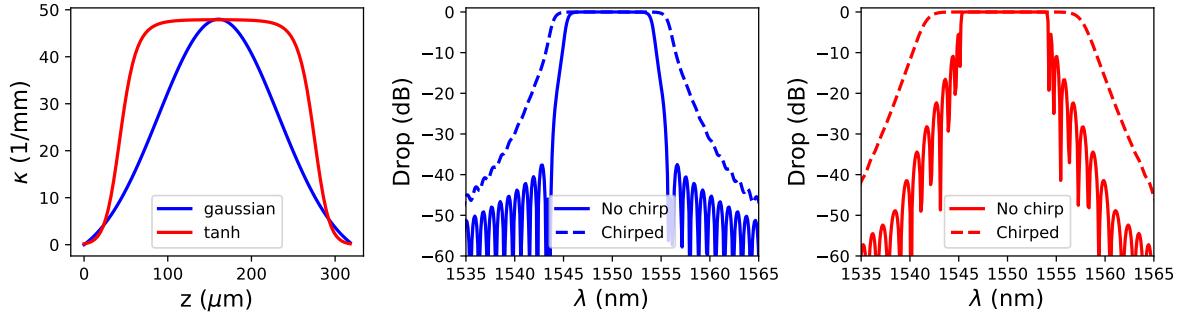


Figure 2.9 – Difference between gaussian and tanh apodization profiles, including chirp-less spectrum and the effect of a 300 K/mm temperature chirp on the device’s response.

effective index change. Thanks to the large thermo-optic coefficient of silicon, temperature chirp offers even more significant spectral broadening when reaching a maximum temperature of 500 K.

Figure 2.8 shows simulations run using a tanh apodization profile. Although the chirp-less simulations (blue curves) show considerable side-lobes, the application of chirp drastically reduces those side-lobes, yielding high-quality responses. This is an important result and constitutes an argument in favour of the tanh apodization profile for chirped designs. Figure 2.9 shows the difference between using a gaussian apodization profile and using a tanh profile on a given CDC’s drop spectrum. Two important results arise from Fig. 2.9. First, the undesirable side-lobes brought by the tanh apodization vanish when a chirp is applied (similarly to what is observed in Fig. 2.8). This is because the tanh profile has a flat top acting similar to a uniform grating. In the flat top section, coupled sub-cavities are able to interfere, creating side-lobes. Once a chirp is applied, the reflection wavelengths at the flat top region are swept, which suppresses these undesirable resonance contributions.

Second, for the same chirp slope of 300 K/mm, the tanh apodization experiences a superior band broadening than its gaussian counterpart (15.5 nm as opposed to 12.1 nm). This is once more explained by the broad flat-top shape of the tanh function. When chirp occurs, reflection wavelengths are linearly swept. In the case of the gaussian apodization, the coupling strength is at a maximum in the centre of the grating and rapidly decreases on either side. Conversely, the tanh apodization profile provides strong coupling for a wider variety of swept reflection wavelengths, resulting in an overall broader drop spectrum. It is hence possible to conclude that for chirp-less operation, the gaussian apodization profile is preferable because tanh creates side-lobes. However, in the case of chirped operation in sights of broadband filtering, the tanh apodization profile offers much more spectral broadening potential while also considerably suppressing the side-lobes.

2.6 Flexible WDM Filters

Now that the concept of contra-directional coupler has been introduced along with its theory, simulation tool and design strategies (apodization, chirp), an interesting application of CDCs for telecommunications is now presented. Wavelength division multiplexing (WDM) is very interesting for service providers because it allows to transfer the information to many clients through a single thread of optical fibre. To achieve the combination and separation of the optical signals destined to each client, quality multiplexing hardware needs to be employed so that the signals are effectively and precisely combined and separated. Contra-directional couplers have been demonstrated suitable for such applications before [31, 9]. Taken a step further, the considered WDM system can be designed to allow flexible operation. That is, the channel grid could be reprogrammed on the fly to reallocate bandwidth between clients. For instance, channel 1 could be broadened and use some of the spectral content originally used by channel 2, and channel 2 could be narrowed to leave a portion of its bandwidth to channel 1. Such tunability is highly desirable and constitutes a key solution for service providers. Although flexible operation has been proposed in fibre networks [32], their implementation remains a challenge for integrated photonics. The current state-of-the-art solution for tunable broadband filters has been demonstrated using contra-directional couplers [4]. In this section, a flexible network is demonstrated using the notions introduced previously. The design of a 4-channel WDM system with a 20-nm channel spacing and an 8-nm channel bandwidth is demonstrated, using channels centred at 1530, 1550, 1570 and 1590 nm, ± 1 nm.

The design strategy is the following. First, a nominal CDC design that offers an (8.0 ± 0.1) nm bandwidth is obtained. Then, necessary grating pitches to attain the target centre wavelengths are found and channel isolation is improved by cascading the CDC devices. Finally, an in-depth analysis of a previously-demonstrated tuning scheme for CDCs is presented, where its flaws are presented and potential solutions are explored, none of which prove satisfactory for practical applications.

For the nominal design, a Gaussian-apodized CDC is used because no chirp is required. The bandwidth is hence dependent upon the maximum coupling κ_{max} employed. Using the simulator and its result interface, it is straightforward to find that using $\kappa_{max} = 45$ 1/mm, $N = 1000$ and $\Lambda = 311$ nm yields the desired central wavelength and bandwidth for the 1530-nm channel, as shown in Fig. 2.10.

Next, the remaining three devices need to be designed to accommodate the remaining three WDM channels. The same design is employed, where only the grating pitch is changed in order to shift the filter pass band to the desired grid spots. Upon simulation, it is found that employing grating pitches of 318, 325 and 332 nm results in pass bands centred around the targeted values of 1550, 1570 and 1590 nm, respectively.

As seen in Fig. 2.10, the out-band suppression ratio offered by the filter is approximately

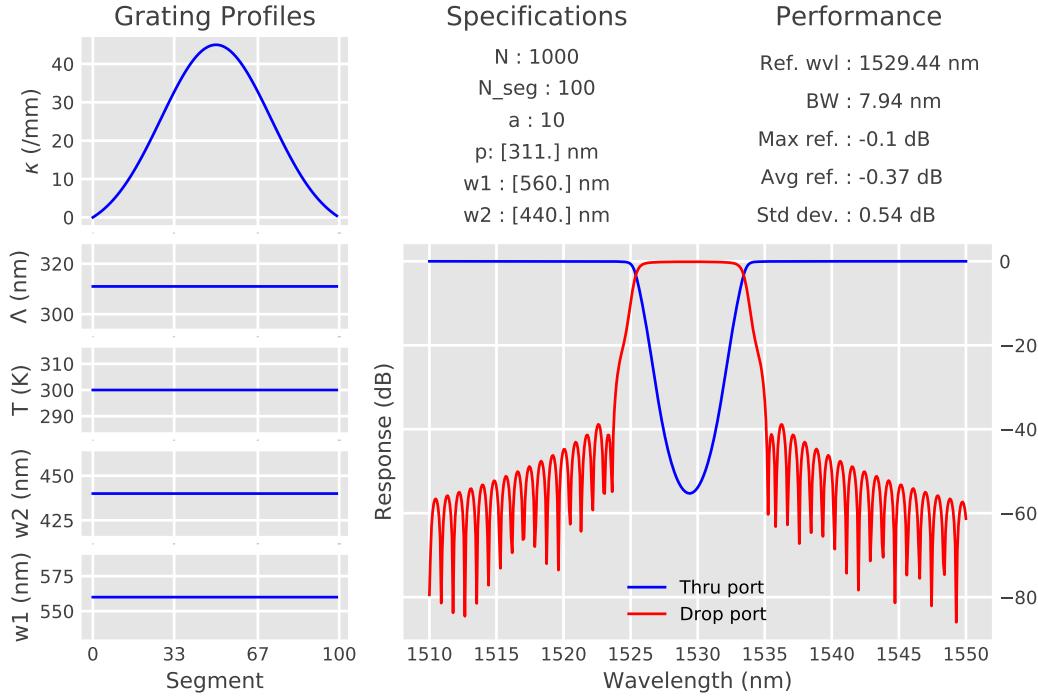
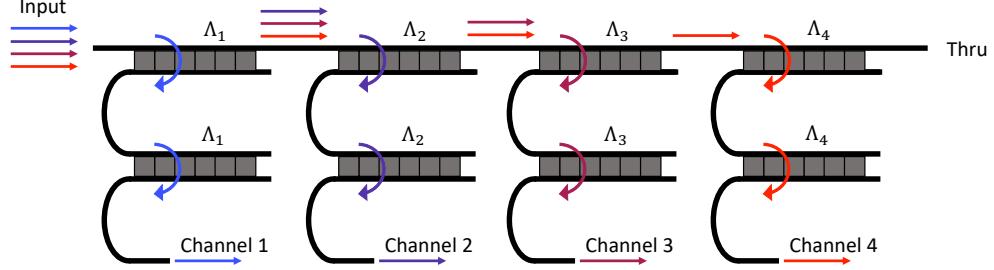


Figure 2.10 – Simulator result window for the nominal CDC filter design centred at 1530 nm.

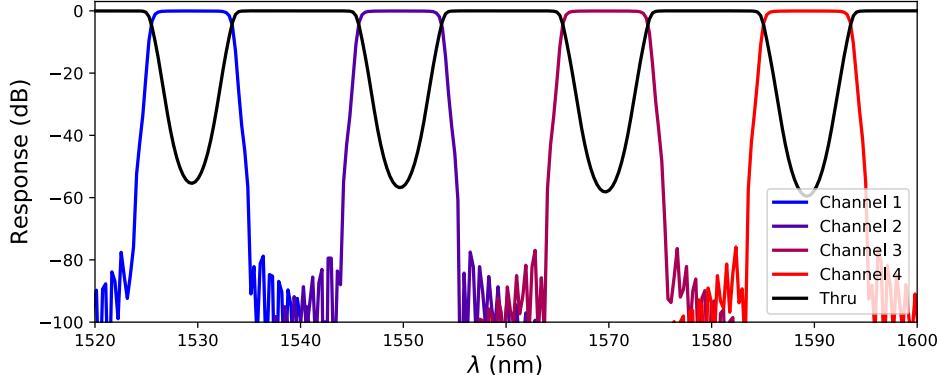
40 dB, which is satisfactory. In reality, fabrication quality often leads to poorer suppression ratios, creating crosstalk between channels. To avoid this, CDC cascading has been proposed to drastically improve the out-band suppression ratio [4, 9]. The optical signal coming out of a given CDC drop port is directly reinserted into the input port of a second, identical CDC, thereby providing dual-stage filtering and improving the response quality [4]. Figure 2.11 shows a schematic of the system and its channel-dependent spectral responses. The schematic shows the four different CDC designs, each comprising their own grating pitch Λ . The channel responses show the low-crosstalk, low-loss performance of the system.

To achieve bandwidth tunability, micro-heating elements are placed over each one of the devices, allowing to increase the local temperature of each CDC independently. The schematic of this operating principle is shown on Fig. 2.12. When one CDC is heated more than the other one, its spectrum experiences a bigger red shift than the other one, due to the thermo-optic effect. As a consequence, the overlapping spectral region between both CDCs is narrowed down and the resulting system's drop port 3-dB bandwidth can be tuned continuously. If both micro-heaters are heated equally, then the spectral overlap does not change, and the resulting system's 3-dB bandwidth remains unchanged, while the central wavelength is red-shifted.

Using this technique, a central wavelength tunability over 4 nm and a 3-dB bandwidth tunability over 670 GHz (5.4 nm) has been demonstrated, breaking the record of widest tunability



(a) Schematic of the WDM system.



(b) Spectral grid of the system.

Figure 2.11 – Schematic and spectral response of the considered CDC-based WDM system.

range demonstrated on a silicon chip [4]. While this demonstration clearly paves the way for the elastic WDM paradigm by offering a wide tunability range, a detail remains to be fixed about the approach. As shown in Fig. 2.12, the spectral region not overlapping between the two responses (namely, the band residue) ends up at the thru port of the second device. As a consequence, this spectral resource cannot be further used by a subsequent channel down the main bus. In other words, while the filter drop port is tunable in bandwidth, it's thru port bandwidth remains constant. Whatever spectral resource is filtered out of the drop spectrum is lost as band residue. This is highly inconvenient for flexible WDM applications because the whole point of dynamic bandwidth allocation is to re-allocate unused bandwidth to another channel. Now that the main issue with the tuning scheme shown in Fig. 2.12 has been identified, some simple solutions to the problem are addressed, and then proven unsuitable by arguments of theory and experimental demonstration.

First, it might seem very straightforward to simply connect the second CDC's thru port to the first CDC's add port, as shown in Fig. 2.13a. The reconnected waveguide solution introduced another problem: the resulting structure constitutes a resonant structure comparable to a ring resonator. Additionally to the solution schematic, Fig. 2.13 shows sample measurements that were taken on fabricated devices. As seen in Fig. 2.13c and 2.13d, the usually-clean response of a cascaded CDC is highly impacted by the ring-like resonant structure created by the simple waveguide reconnection. Both the drop and thru responses show massive ripples

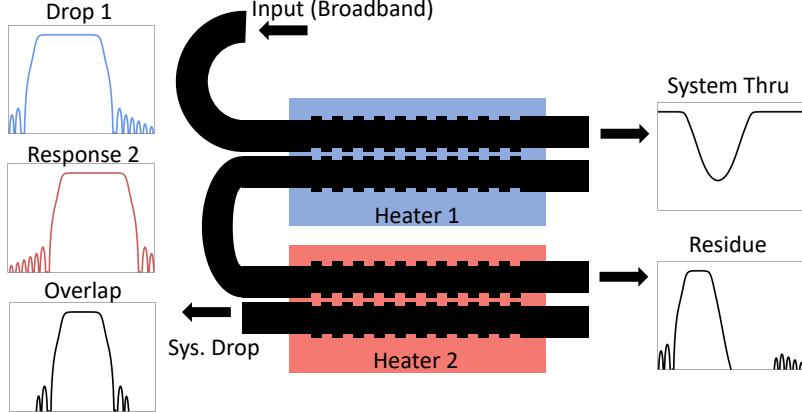


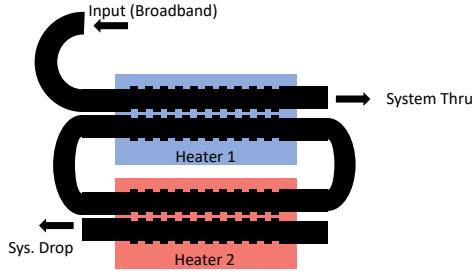
Figure 2.12 – Schematic of the tunable CDC filter presented in [4], where two CDCs are cascaded and heated independently to shift their responses and tune the system's bandwidth.

that make such a filter unusable in any practical application. To verify the resonant nature of the structure, the length of the reconnection waveguides linking the two CDCs was varied, and the free spectral range of the thru port notches was extracted from the spectral responses. In general, the free spectral range of a ring resonator is inversely proportional to the ring's waveguide length L :

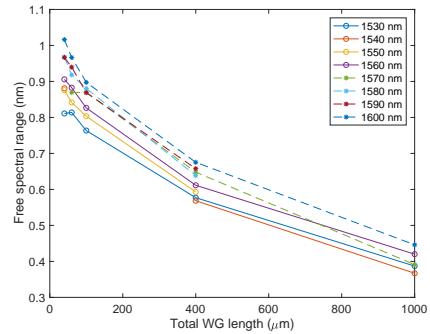
$$\Delta\lambda_{\text{FSR}} \propto \frac{\lambda^2}{L}, \quad (2.17)$$

Figure 2.13b serves to demonstrate two important conclusions. First, the free spectral range indeed varies in an inversely-proportional way with the waveguide reconnection length. Second, the shortest-possible waveguide reconnection length used still causes a free spectral range mainly below the 1-nm mark, which is far too small to ever consider designs where the notches would not interfere with spectral regions of interest (e.g., the C band). The mere existence of this structure is highly problematic, let alone tuning. For these reasons, the reconnected waveguide solution proves unfit for the considered problem. Taken a step further, some creative bent CDC design employing very short reconnection waveguide lengths could be envisioned. However, The close proximity between the two CDCs would be highly problematic for the independent thermal tuning required, bringing a lot of thermal crosstalk and some phase delay would still occur inside the CDC itself. Some potential improvements to this approach could include reducing the coupling power of the second CDC design, in order to avoid the over-coupling condition. This would affect the cavity quality factor and potentially alleviate the ripples.

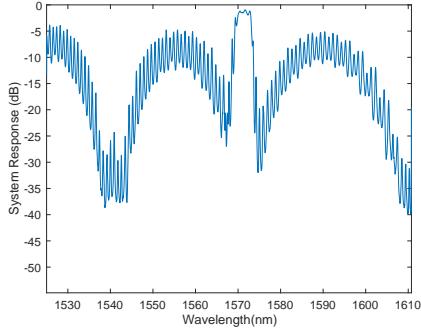
The second approach to the band residue loss problem is the use of a third CDC device, whose sole purpose is the reintroduction of the band residue back into the main waveguide bus. Figure 2.14 shows a summary of the approach of using a third CDC for band residue reintroduction. On Fig. 2.14a, it is possible to see that CDC 1's add port is not connected to CDC 3's bottom left port. This is to avoid optical power circulating inside the system in a ring-like manner,



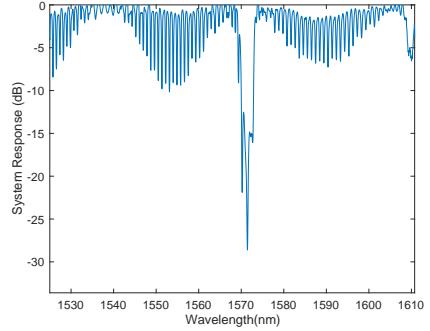
(a) Schematic of the solution



(b) Experimental FSR in function of connection waveguide length



(c) Example measured drop response

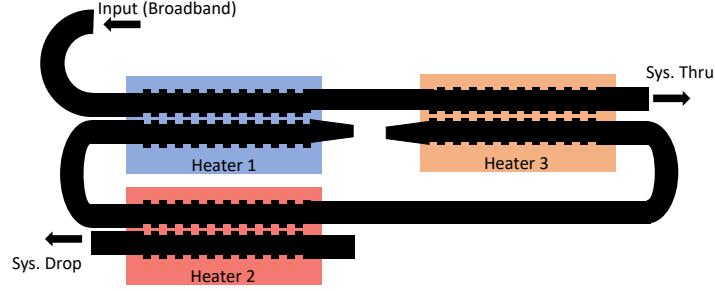


(d) Example measured thru response

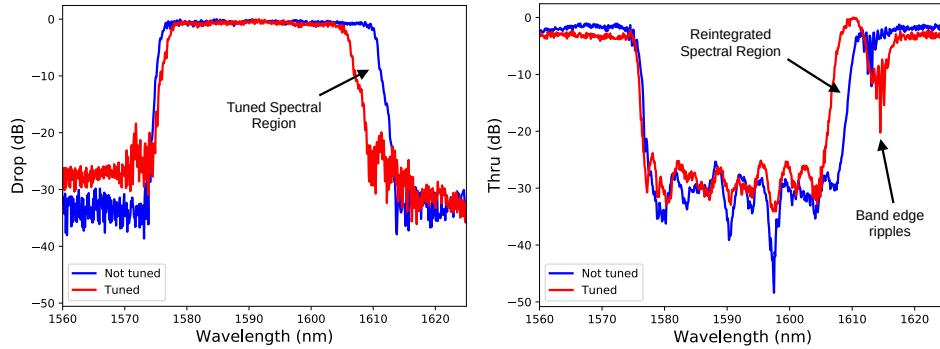
Figure 2.13 – Summary of the reconnected waveguide solution and its experimental results on fabricated devices.

which is the source of the ripples in the reconnected waveguide solution. There are now also three micro-heaters that can tune all three CDCs independently. As a result, the drop and thru port bandwidths can be tuned independently. This 3-CDC tunable filter was designed by Mustafa Hammod from The University of British Columbia, and the CDCs themselves are based on a sub-wavelength waveguide design proposed in [29]. Hammod and I have collaborated on solving the band residue problem, and he had a chip fabricated in an effort to show that using a third device would effectively avoid the spectral ripples and allow seamless band reintroduction. Figures 2.14b and 2.14c show a sample measurement of the tunable filter designed by Hammod, fabricated on an electron beam (E-beam) lithography process, and measured by myself. It can be seen that the operating principle works effectively to tune the filter bandwidth and avoid ripples in the spectral response. It is also seen in the figure that the spectral region removed from the drop port during tuning is successfully reintegrated to the system thru port. These two results show a great improvement from the previous, more simple solution proposed.

Nevertheless, there remains an issue with the solution. As shown in Fig. 2.14c, some ripples are present at the filter band edge. Those ripples are present whether tuning occurs or not,



(a) Schematic of the 3-CDC solution where a third CDC device is used to reintegrate the band residue back into the main bus.



(b) Experimental tuning of the drop spectrum.

(c) Experimental tuning of the thru spectrum.

Figure 2.14 – Summary of the three-CDC solution . Two key measurements are shown for the drop and thru spectra of the filter during tuning.

although they are more important when tuning occurs. This issue is caused by the filter roll-off, which is not perfectly vertical. When the device is not tuned, the right band roll-off is approximately -7.1 dB/nm , whereas it is around -5.8 dB/nm when it is tuned. As seen in Fig. 2.14c, straying away from an ideal rectangular filter shape makes the reintegration harder, no matter how fine the system is tuned. In other words, the task at hand is to stitch two spectral bands together. If the filter roll-off is not sharp enough, some spectral content is missing at the band edge and the stitching cannot be seamless. This is problematic because while some amount of band reintegration occurs, this band edge ripple remains present and avoids the spectral resource reallocation. In summary, the third CDC solution avoids the ripples caused by creating an optical cavity, it successfully accomplishes band residue reintegration, but the reintegration is not seamless.

Two potential solutions to the band residue problem hindering flexible WDM filters based on cascaded contra-directional couplers have been presented. While the first, most intuitive solution proves to be unsuitable for any practical application due to the resonant structure it creates, the second solution proves to succeed at a clean band residue reintegration, with its only problem being that the reintegration is not seamless (i.e. it leaves a ripple artifact at the band edge). Overall, important progress is shown in the understanding and implementation

of a residue-less cascaded-CDC filter. Nonetheless, the current solution is still not suitable for practical applications. Notably, the tuning scheme can only be used to narrow down the bandwidth, not broaden it. On production-grade fabrication processes (UV lithography), obtaining broadband CDCs is challenging due to the relatively large minimum feature size attainable. The current tuning scheme does not solve this problem because it does not allow band broadening. A new broadband filter tuning concept needs to be introduced to achieve bandwidth-broadening, residue-less, and seamless flexible WDM filtering. In Chapter 5, the idea of using thermal chirp to accomplish this task is introduced in an original contribution, where even-better tunability and fabrication error compensation are demonstrated using a single-stage contra-directional coupler and a segmented micro-heating device allowing to create arbitrary temperature profiles.

Chapter 3

Machine Learning

3.1 Introduction

Machine learning (ML) is a subset of the artificial intelligence (AI) paradigm, which aims for computers to be able to perform human tasks by imitating human brain learning processes. In particular, machine learning comprises all statistical algorithms that enable AI implementation through the use of data [33]. With the advent of readily available large amounts of data through increased Internet bandwidth and more powerful computing hardware capable of treating the data, machine learning has emerged in recent years as a powerful mathematical tool able to create highly complex models and perform even better than humans for certain tasks such as image recognition [34]. The implementation of machine learning usually requires the following steps: 1) acquiring a dataset representative of the problem considered, 2) choosing a machine learning method that fits the dimensionality and complexity of the dataset, 3) training the model on the training dataset and assessing its performance on the validation dataset, then 4) putting the model into production for real-life applications. This chapter aims to explain these steps and their underlying concepts, many of which are used in Chapter 6.

3.2 Concepts and Methods

This section aims to lay the conceptual foundation of the use of ML for any application. The goal of any ML application is to obtain a mathematical model (or function) able to accurately infer outputs based on inputs fed to it. The former can be represented in a simple equation:

$$\mathbf{y} = f(\mathbf{X} \mid \theta), \quad (3.1)$$

where \mathbf{X} is the input vector (the observables upon which inference is performed), f is the mathematical function (or model) performing the inference task upon the input, θ is the set of parameters of the model which best describe the training data, and \mathbf{y} is the model's output (its inference). For a given problem, the goal is a model f whose generic shape is complex enough

to model the problem, and a set of parameters θ such that the model accurately describes the data at hand.

In its simplest forms, Eq. 3.1 is fairly intuitive and widely used. For instance, performing a linear regression in a two-dimensional space consists in using Eq. 3.1 where $y = \theta_1 x + \theta_2$. In this simple example, f is chosen to be a first-degree polynomial function and the parameters θ_1 and θ_2 are optimized to best describe the data at hand, in an effort to model the problem as accurately as possible. The optimization process by which the θ vector is obtained is highly dependent upon the model f chosen to describe the problem. They range from the simple method of least-squares for linear regression to sophisticated stochastic gradient descent algorithms used for deep neural networks.

Given the freedom given to Eq. 3.1 in using an arbitrary function f to model any problem, a few methods are listed and briefly explained so as to give an idea of the wide variety of techniques proposed in the realm of supervised learning (where a human agent is in charge of performing the training and assessing the model performance).

- Parametric Regression is based upon a statistically-reasonable distribution shape (such as multivariate Gaussian distribution), and fitting algorithms such as maximum likelihood estimation are used to obtain the parameters.
- k-nearest neighbours is based upon a multivariate clustering assumption and assumes that a classification task can be performed by imitating known data points in the parameter-space vicinity of the queried inputs. It does not require training per say but needs to be assessed on training dataset to be deemed efficient.
- Linear discrimination assumes that there exists a linear hyper-plane allowing to discriminate between two data classes and that this hyper-plane parameters can be obtained through gradient descent.
- Multi-layer perceptrons, also called artificial neural networks (NNs) are based upon a brain-emulating mathematical architecture suitable to express non-linear behaviours and utilize gradient descent for parameter optimization.

This work focuses solely on the multi-layer perceptron model because it is best suited for the considered problems. Now that the basis of machine learning is laid, some useful concepts need be introduced to understand the training procedure.

The **dataset** at hand comprises the entirety of the empirical data either collected or acquired through real-life experience. It contains a set of organized inputs \mathbf{X} and outputs \mathbf{y} that are known to be true. In matrix form, the dataset is usually organized in the form $[\mathbf{X}, \mathbf{y}]$, a N by $(d_y + d_X)$ matrix (where N is the size of the dataset and d_y and d_X are the dimensionality

of the outputs and inputs, respectively). The size of the dataset is important in assessing possible over-fitting.

Over-fitting generally occurs when the mathematical model chosen to describe the data and its parameters are too tightly-bound to the dataset used for training. That is, the model describes the training data very well but any new input data fails to be inferred accurately. An Over-fitted model is useless for any practical application. To avoid over-fitting, it is necessary to use a sufficiently-large dataset and to employ a dataset splitting method.

The dataset is generally split at random between a **training set** and a **test set**. The splitting ratio can be chosen at will and it is best to shuffle the data first to ensure proportional representation. The training algorithms are performed entirely on the training dataset, while the test dataset is set apart and never used for training. The role of the test dataset is to use the model on it and assess its performance to then compare it against the training dataset performance.

A model training requires the definition of a **loss or cost function**, that maps a given inference event onto a scalar number quantifying the quality of the inference. For linear regression, it could be the root mean squared error. For a classification task, it could be a simple binary function for success/failure to classify properly. The cost function is central to the training process, whose goal is to obtain the set of parameters θ that best describes the dataset and hence minimizes the cost function.

The **model training** consists in performing an iterative algorithm where the entirety of the training dataset is fed to the model, and the latter performs inference tasks. The result of those inference tasks are compared against the known values of the samples in order to compute the cost associated with the model in its current state. The iterative process of training consists in adjusting the model parameters θ until the model performance cost is deemed satisfactory on both training and test data. The parameter tuning can either occur once all training samples have been fed, or once every n samples have been fed (if a large number of test samples are available). In this case, a training subset comprising those n samples is called a **batch** and n is called the **batch size**. Once all the batches have been used for training and parameter tuning, an **epoch** has passed. A given training generally comprises many epochs.

Figure 3.1 shows a schematic view of the training process of a machine learning model. The dataset is divided into training and test data, and the training data is further sub-divided into batches. For every batch, the input data is fed into the model in its current state, which performs predictions. These predictions are compared against the known values contained in the training dataset, which allows the computation of a cost value associated with the batch. This cost is fed into the optimization algorithm which updates the model parameters accordingly. Once all the batches have been fed and the parameters updated, an epoch has passed. The next section will show in detail how the model, cost and update blocks work in

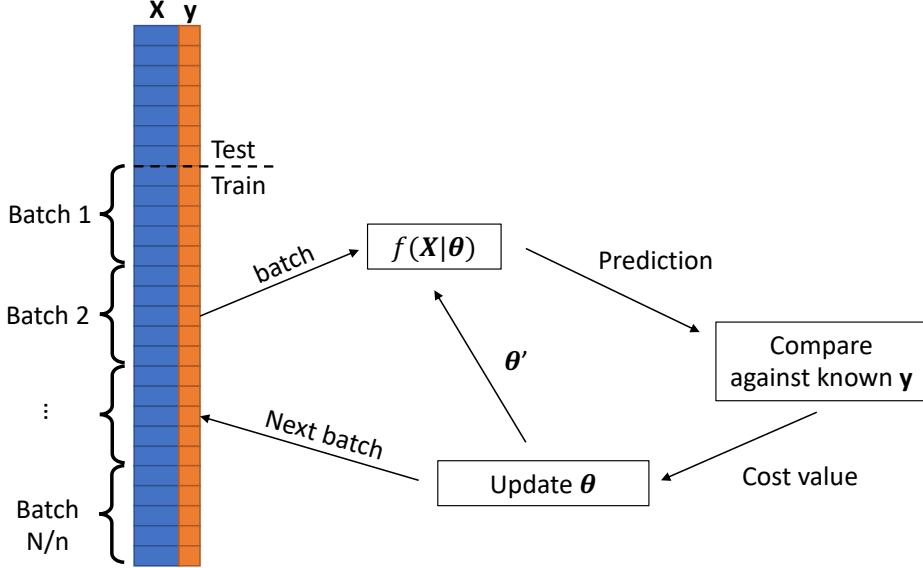


Figure 3.1 – Training process of a supervised machine learning model.

the case of artificial neural networks.

3.3 Artificial Neural Networks

Multi-layer perceptrons or artificial neural networks are machine learning models built to emulate the architecture of the human brain by adopting a network of nodes called neurons connected by synaptic links. Figure 3.2a shows a schematic representation of a neuron. the inputs \mathbf{X} are each multiplied by a given weight. The summation of those weighted inputs and a neuron bias w_0 is then performed. After this is done, the result of the summation passes through a non-linear function that maps it between 0 and 1 in the neuron output y . Overall, the mathematical representation of an artificial neuron can be expressed as [33]:

$$y = f \left(\sum_{i=1}^N w_i x_{n,i} + w_0 \right), \quad (3.2)$$

where the non-linear function can take many forms such as sigmoid, ReLu or tanh [33]. The motivation behind using a non-linear output function comes from the desire to model non-linear mathematical problems.

Artificial neurons can be organized in artificial neural networks, where many neurons are placed side-by side to form a layer, and layers are cascaded one after the other. This creates a dense network of synaptic connections effectively accomplishing complex mathematical computations as a whole. In the case of artificial neural networks, each neuron's set of weights \mathbf{w} acts as the model's parameters θ from Eq. 3.1.

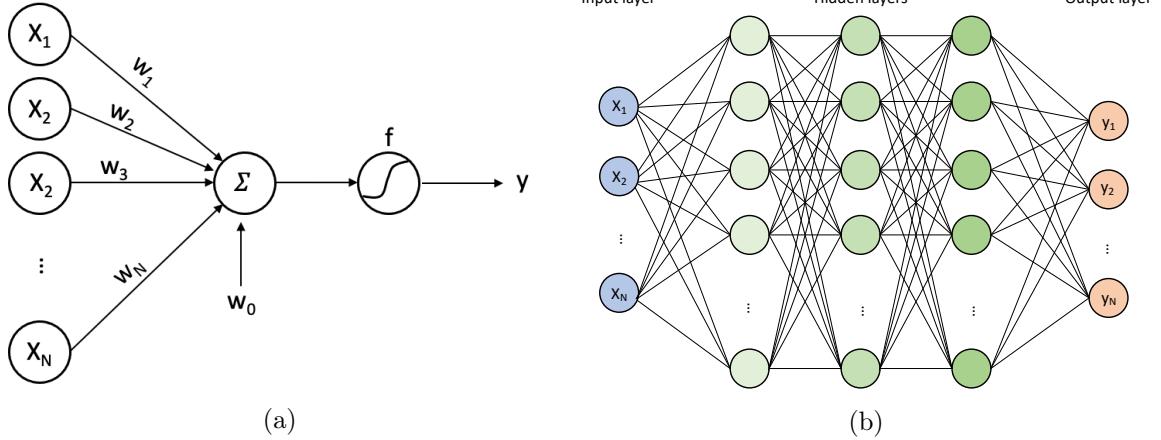


Figure 3.2 – Schematic representation of (a) a neuron and its synaptic connections, and (b) a neural network comprising three hidden layers (5 layer total) and its synaptic connections (weights).

Using this unit architecture, arbitrarily-sized artificial neural networks can be assembled by creating multiple layers of synapse-connected neurons. Figure 3.2b shows a schematic representation of such a neural network, where the inputs are connected to subsequent layers. The inside layers are called hidden layers and are eventually connected to the output layer. Such layers with synaptic connections between all adjacent neurons are called fully-connected layers.

In order to train a neural network, the backpropagation algorithm is employed using the training data at hand. The basic principle behind backpropagation is the desire to update every weight w of all neurons with values computed via the error gradient [33]:

$$w_i \leftarrow w_i - \eta \frac{\partial E}{\partial w_i}, \quad (3.3)$$

where E is the user-defined error (or lost, cost) criterion and η is the **learning rate**, a scalar number typically between 0 and 1, responsible for the step size of the optimization at each iteration. Starting from the output of the network and gradually making its way to the input, all neurons have their weights updated using Eq. 3.3. Of course, computing the partial derivative requires knowledge of the activation function used in the network and the employment of the chain rule for differentiation. Algorithm 1 shows the typical procedure employed in order to train a neural network using training and test data. The basic idea is to feed the network all the training data, while periodically (after every batch) making parameter adjustments to the model (updating its weights). This algorithm is commonly known as the gradient descent algorithm. It consists in exploring the multidimensional parameter space and assessing the local gradient in order to converge to a potential absolute minimum of the cost function E .

In Algorithm 1, it is indicated that the learning rate η should be decreased as the model gets

```

while Test data error is too large do
    for every batch in training dataset do
        Propagate forward (Eq. 3.2);
        Compute error  $E$ ;
        Perform backpropagation (Eq. 3.3);
    end
    Compute error on test data;
    if test error is plateauing then
        | decrease learning rate;
    end
end

```

Algorithm 1 : Typical neural network training procedure using gradient descent.

better and better trained. This can be understood intuitively because as the model slowly approaches the absolute minimum of the optimization space, reducing the step employed at each optimization iteration serves to examine the parameter space in more detail and essentially converge more finely. This technique is commonly called adaptive learning rate, or learning rate scheduling. Further methods might be employed, such as introducing a momentum to Eq. 3.3 in order to avoid abrupt gradient reversal, or even more sophisticated algorithms such as adaptive moment estimation [35].

3.4 Applications in Photonics

The field of machine learning has been proven to be highly compatible with photonics, each discipline being of great service to its counterpart. First, integrated photonics has been demonstrated as a suitable platform for neural network hardware implementations. That is, the photonic components are able to intrinsically accomplish many tasks required by neural networks. For instance, microring resonators can be put in a wavelength-division multiplexing scheme to serve as neural weight banks [7], germanium photodiodes can accomplish the task of summation, and Mach-Zehnder modulators can be used as activation functions. The wave nature of light can be used to multiplex signals and increase the number of vector operations while decreasing the required energy by implementing photonic tensor core units [36].

On the flip side, machine learning techniques can greatly serve photonics, especially for simulation and design optimization. Electromagnetic waves are often complex to simulate and require substantial time and computing resources. Sophisticated non-AI optimization algorithms make use of parameter-space navigation in order to find local or global extrema [37]. While such algorithms are effective, they require to navigate the parameter space every time a design is queried. This often requires the model to be run many times, which can be computationally intensive and time consuming. Machine learning, on the other hand, relies on a first step of model training and optimization using a training dataset. While the dataset

creation and training processes can be just as time-consuming, they yield a trained model that can handle any query in a matter of microseconds. This is highly desirable for inverse design and fabrication diagnostics purposes. Artificial intelligence has proven to offer highly-effective photonics design tools [38]. This task has generally been tackled in one of three ways.

First, inverse design has been demonstrated using generative adversarial networks (GANs). This unsupervised learning technique aims to train a model to generate artificial input data that displays statistical properties similar to the training dataset's. Notable examples of this approach include the synthetic generation of realistic human portraits [39]. In photonics, such an approach has been demonstrated to design high-dimensionality devices such as free-form metasurfaces [40, 41].

Second, parameter-based regression models offer forward, inverse or bidirectional design of photonic components. Parameter-based models are built on the assumption that the device physical properties and/or its spectral response follow a certain shape, usually based on theoretical models. The presumed shape is usually a function fully defined by a handful of parameters. This creates a low-dimensionality input/output parameter space that can be learned by a ML model using multivariate regression. For instance, a forward deep learning- (DL-) based model has been demonstrated to predict the Lorentzian oscillator-shaped optical response of a nanoslit array [42]. Inverse design demonstrations include core-shell nanoparticles [43], materials targeting certain properties [44], and topological photonics [45]. Bidirectional modelling has also been demonstrated by training separate models for forward and inverse design like integrated Bragg gratings [46]. Some demonstrations also include coupled bidirectional models that use an iterative, back-and-forth training process [47] [48].

A third approach lies in the use of deep learning for free-form parameter space design. A good example of this is the design of integrated power splitters targeting specific splitting ratios [49]. A pixel grid of $2.6 \times 2.6 \mu\text{m}^2$ is defined as parameter space where each pixel is independent and can either contain silicon or not. While parameter-based methods assume a certain shape to which it seeks to fit parameters, a free-form approach allows for an infinite design parameter space. It then relies on the provided dataset's parameter space contents to learn possible output shapes. This approach is more versatile because the training dataset can be supplemented at any time with further design parameter shapes. In this sense, it constitutes an architecture-independent approach. This approach will be proposed in Chapter 6, where complex photonic circuits in the shape of Bragg grating will be considered. Each grating segment is treated as a pixel where three physical parameters of the segment completely define the local design. A synthetic dataset is generated by a transfer matrix model, which allows to train a neural network to accurately perform inverse design tasks. Taken a step further, an experimental-emulating dataset is also created and trained upon in an effort to accomplish fabrication diagnostics tasks on actual fabricated devices.

Chapter 4

Dual-Band Optical Filters Using Integrated Multimode Bragg Gratings

Résumé

Nous démontrons l'utilisation d'un réseau de Bragg multimode permettant le filtrage optique double bande dans la région spectrale de 1.5 à 1.6 μm . Des largeurs de bande de 4.4 et 7.5 nm ainsi qu'un espacement inter-bande de 42 nm sont atteints.

Abstract

We demonstrate a multimode integrated Bragg grating allowing dual-band filtering in the 1.5-1.6 μm region. Bandwidths of 4.4 and 7.5 nm and a band separation of 42 nm are achieved.

4.1 Introduction

Optical networks provide the physical foundation for high-speed data exchange through optical fibers. With the ever-increasing volume of data traffic, data centers and service providers need to keep up with novel modulating and multiplexing devices. The silicon-on-insulator platform has proven to be promising in that regard, due to its compatibility with established CMOS manufacturing processes, small footprint, and remarkably low energy consumption. Along with the integrated-optics paradigm comes the need for various building blocks of such systems. Passive filters are among those devices crucial for the evolution of optical networks. Single-band filtering is typically achieved using either microring-based or grating-based devices. In the same way, dual-band optical filters find use in telecommunications, namely for dual-band WDM [50]. Dual-band operation has been demonstrated using multimode microdisk resonators [51]. This approach showed narrow-band, small-FSR and non-flat-top spectral characteristics.

In this work, we demonstrate a silicon-on-insulator multimode Bragg gratings (MMBG) to achieve dual broadband filtering in the 1.5-1.6 μm spectral region, towards dual-band CWDM or LAN WDM applications. The MMBG is designed such that the self-coupling of odd modes and the inter-coupling of even modes create two well-defined pass bands in the drop port. The grating is formed using corrugated sidewalls similar to [4]. A simple asymmetric y-branch is used as mode converter and for measurement purposes.

4.2 Design and fabrication

The proposed design uses a Bragg grating in a multimode strip silicon waveguide. The three first TE modes are considered. Both the TE_0 and the TE_2 modes have symmetric electric field distributions, while the TE_1 mode's electric field shows antisymmetry, as shown in Fig. 4.1 (a). According to the coupled-mode theory (CMT), a dielectric perturbation causes the coupling between two given modes having field distributions ϕ_m and ϕ_n . The coupling coefficient is given by the overlap integral between the fields and the perturbation profile [14]:

$$\kappa_{mn} = \frac{\omega}{4} \iint \varphi_m^*(x,y) \Delta\varepsilon(x,y) \varphi_n(x,y) dy dx, \quad (4.1)$$

where $\Delta\varepsilon(x,y)$ describes the shape of the periodic perturbation imposed upon the waveguide. From Eq. 4.1, it is possible to deduce what modes are allowed to couple into each other. In [52], an asymmetric Bragg grating allows the forward TE_0 mode to couple into the backward TE_1 mode. In our case, a symmetrical Bragg grating is used, i.e. $\Delta\varepsilon(x,y)$ is symmetric with respect to the longitudinal axis. This strictly allows for even-to-even mode coupling as well as self-coupling.

The phase-matching condition of the first-order Bragg reflection imposes $\beta_m - \beta_n = 2\pi/\Lambda$, where $\beta = 2\pi n_{eff}/\lambda$ is the propagation constant and Λ the grating pitch. The sign of β must account for the direction of propagation. In the case of self-coupling, $\beta_n = -\beta_m \equiv \beta$ and the Bragg reflection occurs at $\lambda_B = 2\Lambda n_{eff}$. When the coupling from a mode to another occurs, the Bragg reflection is located at $\lambda_B = 2\Lambda \bar{n}_{eff}$, i.e. the reflection wavelength is the arithmetic mean of the reflection wavelengths of each individual mode, as demonstrated in [53]. Fig. 4.1 (b) shows the simulated effective indices and Bragg wavelengths based on a 320-nm pitch for all supported modes. The result of the coupling between TE_0 and TE_2 as a function of the waveguide width used is also shown. While the self-coupling of the different modes shows considerable spacing in terms of effective indices, for a waveguide width around 1.1 μm , the coupling of forward TE_0 into backward TE_2 yields a Bragg reflection close enough to that of the TE_1 self-coupling, such that both coupling conditions exhibit a reflection peak in the 1.5-1.6 μm spectral region. The principle of operation of this dual-band filter is hence based on the combined Bragg reflections of the TE_1 mode into itself and the contradirectional coupling between the TE_0 and TE_2 modes to obtain two distinct reflection bands.

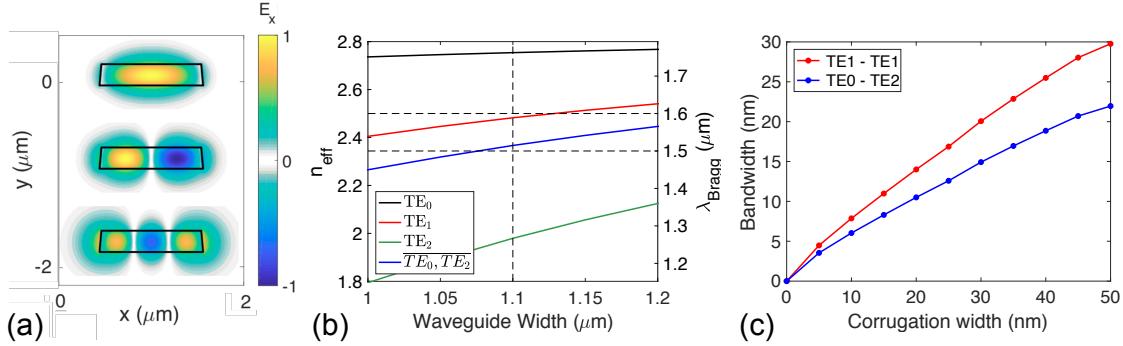


Figure 4.1 – (a) Simulated TE mode profiles, (b) simulated mode-dependent effective indices and corresponding Bragg wavelengths at 1550 nm as a function of waveguide width. (c) Reflection bandwidth as a function of corrugation width.

Fig. 4.1 (c) shows the simulated filter bandwidths as functions of corrugation width, based on the mode overlap (Eq. 4.1). The bandwidths increase with the corrugations and so does the difference in bandwidths. This is due to the limited spatial overlap between the TE₀ and the TE₂ mode distributions, while the overlap between the TE₁ mode and itself is unsurprisingly higher. Hence, dual-band filtering with a wider most-right band is expected. The Bragg grating is also apodized according to the equation shown in Fig. 4.2 (a) in order to avoid sidelobes in the reflection spectrum.

For measurement purposes, a y-branch was used. Y-branch responses depend on the considered modes. It was designed to be asymmetric such that both reflection bands have similar loss. By using branch widths of 600 and 475 nm, the simulated device showed both reflections bands having a total intrinsic loss of 12.5 dB due to the two-way passage through the device. Fig. 4.2 shows the grating Gaussian apodization profile with a Gaussian parameter of 12, which was discretized for fabrication requirements. It also displays the scheme of the suggested device featuring the asymmetric y-branches placed on either side of the MMBG so as to allow add-drop-like operation.

The device was fabricated using Advanced Micro Foundry (AMF)'s 193-nm deep UV lithography process. An electron microscope image is displayed on Fig. 4.2 (c), showing the corrugations near the center of the grating. The designed MMBG featured a 1.1-μm-wide multimode waveguide with 30-nm-wide sidewall corrugations and a 320-nm pitch. Due to lithography smoothing, the fabricated design actually displays a corrugation width of 10 nm. This effect had been accounted for in the design as wider-than-needed corrugations were chosen.

4.3 Results

The fabricated device was tested using off-chip optical spectral analyzing and polarization controlling equipment.

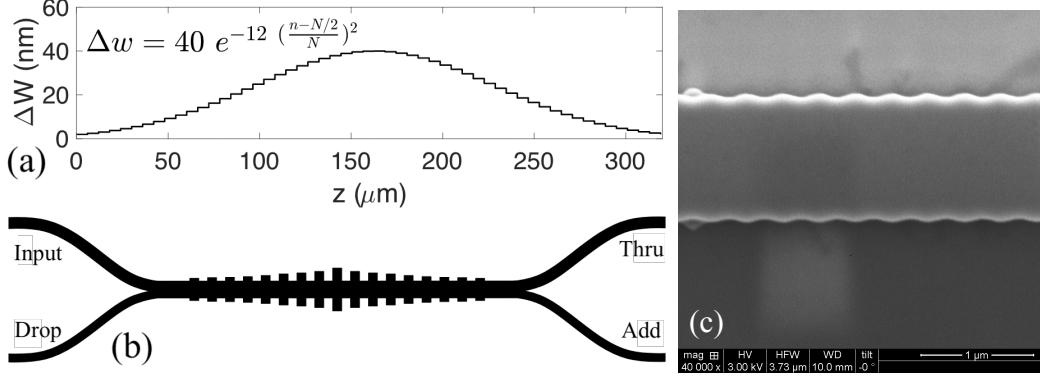


Figure 4.2 – (a) Apodization profile of the fabricated design, (b) schematic of the fabricated device including apodized multimode Bragg grating and asymmetric y-branches used for characterization purposes, and (c) scanning electron microscope image of the fabricated device.

Fig. 4.3 shows the measured transmissivity (thru-response) and reflectivity (drop-response) spectra. The drop-port response shows two well-defined reflection bands, centered at 1543 and 1585 nm, hence showing 42-nm band spacing. The reflection bandwidths are 4.4 and 7.5 nm, which agrees well with the simulation results shown on Fig. 4.1 considering the as-fabricated 10-nm-wide sidewall corrugations. An out-of-band rejection ratio of 9.2 dB is also shown experimentally. This filter shows an experimental loss of 13 dB, which is considerable, although mainly caused by the asymmetric y-branch's intrinsic loss (12.5 dB simulated). This could be avoided by placing the device in a Mach-Zehnder configuration [54].

4.4 Conclusion

We demonstrated dual-band optical filtering using multimode integrated Bragg gratings, achieving bandwidths of 4.4 and 7.5 nm and a 42-nm band separation in the 1.5-1.6 μm spectral region. This paves the way for compact and versatile dual-band WDM communication systems.

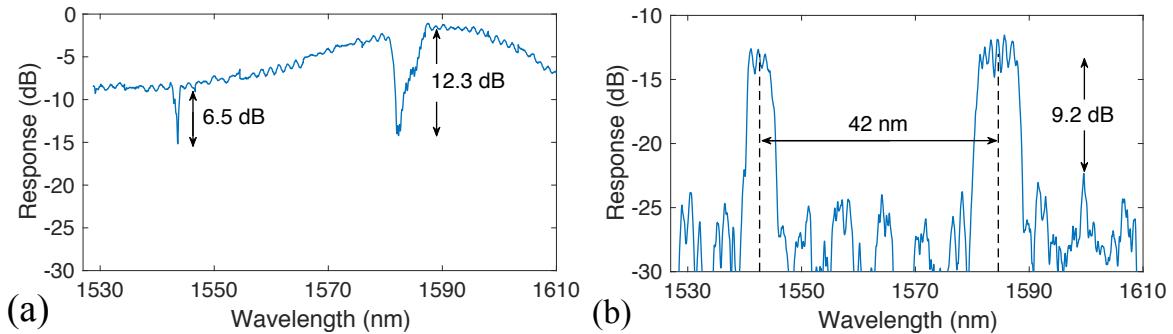


Figure 4.3 – Measured (a) thru-port and (b) drop-port responses.

Chapter 5

Thermally-Chirped Contra-Directional Coupplers for Residue-Less, Bandwidth-Tunable Bragg Filters with Fabrication Error Compensation

Résumé

Les filtres accordables en bande passante sont essentiels dans les réseaux optiques élastiques pour l'allocation dynamique de bande passante. Les solutions existantes en photonique sur silicium sont confrontées à des défis pour répondre aux exigences des applications du monde réel en raison des compromis de conception et des erreurs de fabrication. Nous proposons et démontrons expérimentalement un filtre d'ajout accordable qui consiste en un coupleur contra-directionnel apodisé selon une fonction tangente hyperbolique avec un micro-élément chauffant segmenté. Cette configuration permet de créer un profil de température arbitraire le long du dispositif pour accorder la bande passante des spectres de passage et de chute. Nous montrons que le fonctionnement algorithmique du dispositif peut compenser efficacement la non-uniformité de fabrication locale et améliorer le rapport de suppression hors bande de 69%. L'application de niveaux et de pentes de température appropriés permet d'ajuster en continu la longueur d'onde centrale du filtre sur 8 nm et sa bande passante à 3 dB entre 14,0 et 22,4 nm.

Abstract

Bandwidth-tunable filters are essential in elastic optical networks for dynamic bandwidth allocation. Existing solutions in silicon photonics face challenges meeting requirements in real-world applications due to design trade-offs and fabrication errors. We propose and exper-

imentally demonstrate a silicon photonic tunable add-drop filter in a single-stage, hyperbolic-tangent-apodized contra-directional coupler with a segmented micro-heater. It allows to create an arbitrary temperature profile along the device for bandwidth tuning in both through and drop responses. We show that the algorithmic operation of the device can effectively compensate local fabrication non-uniformity and improve the out-of-band suppression ratio by 69%. Applying proper temperature offsets and slopes allows to continuously tune the filter's center wavelength over 8 nm and its drop-port 3-dB bandwidth between 14.0 and 22.4 nm.

5.1 Introduction

Bandwidth limitation is an ever-increasing barrier to faster and more energy-efficient optical networks. This has led to a desire to use tunable optical filters in order to dynamically allocate bandwidth based on real-time demand [32]. In this context, much work has been put into the development of silicon-based optical filters. Several integrated approaches to the problem have been explored on the silicon-on-insulator (SOI) platform. Microdisk [55] and microring [56] resonator-based systems have both been demonstrated to offer dynamic bandwidth allocation. These resonator-based systems have narrow bandwidths and allow for dense wavelength-division multiplexing (WDM). However, they are limited by relatively small free spectral ranges. By contrast, grating-assisted, contra-directional couplers (CDCs) offer FSR-free, broadband, flat-top filtering [57] and have been demonstrated to be suitable for WDM applications [58, 9].

Micro-heating elements provide versatile on-chip Joule heating that can be used to shift the spectral response of passive silicon components by applying the proper current [8]. State-of-the-art CDC tunability has been demonstrated by cascading two devices, and using such heaters to both tune the central wavelength and the bandwidth by shifting the CDCs' drop responses overlap [4], as shown in Fig. 5.1a. Although the spectral overlap of the devices allows continuous bandwidth tunability, the non-overlapping spectral region, namely the band residue, is lost and cannot be easily integrated back into the main bus. Although it might seem possible to re-connect the second thru-port waveguide back into the first device's add port, the response quality is considerably hindered by the extra delay and phase mismatch it creates.

In this letter, we propose a tunable broadband filter based on a single-stage CDC and a segmented micro-heater (Fig. 5.1b) to create arbitrary temperature profiles along the device. While allowing for central wavelength and bandwidth tunability without a band residue loss, this configuration also allows compensation for local fabrication errors and variations.

Chirp is a design strategy that has found use in fiber Bragg gratings (FBG) to increase the filter bandwidth [27]. Since FBGs have low index contrasts, their bandgaps are typically very small. By linearly chirping the grating pitch along the propagation direction, chirp allows to scan a

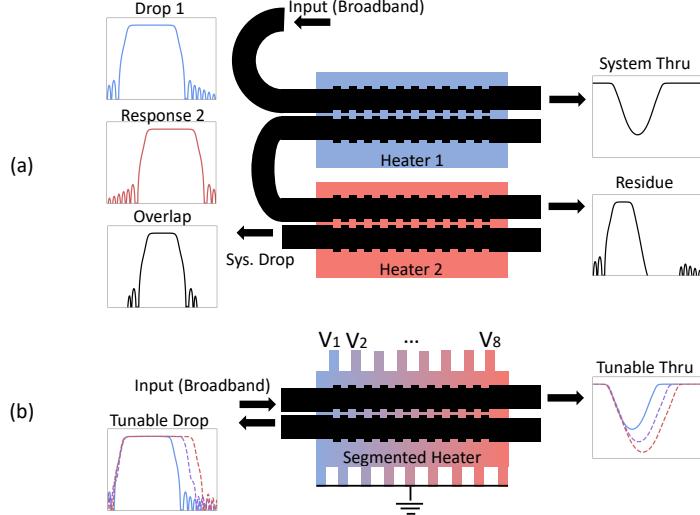


Figure 5.1 – Operational scheme of fully-tunable CDC-based filters. (a) Cascaded filters as proposed in [4]. (b) The segmented micro-heater solution proposed in this work. The heater dimensions are not at scale.

range of phase matching conditions and effectively broaden the reflection band. It is also used for tunable delay lines in waveguide Bragg gratings [59]. FBGs have also been demonstrated with proper temperature profiles allowing for bandwidth tuning [60]. On the SOI platform, thermal chirp has been demonstrated on Bragg gratings by using a heater running diagonally with respect to the grating, hence creating a linear temperature profile [61]. However, it only provides a single degree of freedom in the tunability and no offset or fine tuning can be done.

5.2 Design

Our passive design is based on apodized, corrugation-enabled contradirectional couplers [4], with slight modifications to comply with chirped operation. The two waveguides forming the CDC have widths of 440 and 560 nm, and are separated by an average gap of 100 nm. They comprise 1100 grating periods, a grating pitch of 324 nm and maximum corrugation widths of 32 and 38 nm, respectively. Bragg device apodization is used to avoid undesirable out-of-band side-lobes in the spectral response. Since these side-lobes are due to abrupt changes in the coupling power, using a progressive spatial coupling profile drastically reduces this effect. Moreover, the discrete spatial division of the device into uniform segments for apodization has an effect on this side-lobe suppression. Namely, the use of a sufficient number of discrete segments is required to avoid parasitic reflection bands [62]. While a Gaussian apodization with a proper Gaussian constant generally suppresses CDC side-lobes sufficiently, this work proposes to comply with chirped operation by using a hyperbolic-tangent- (tanh-) shaped

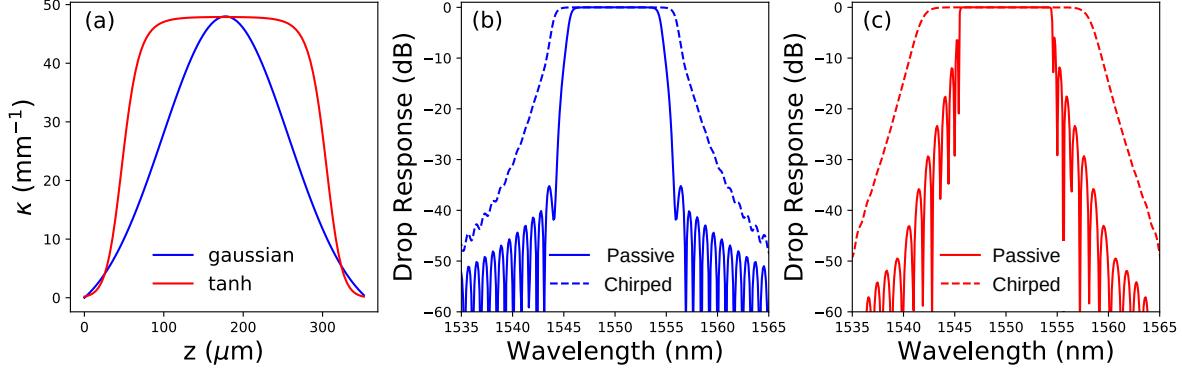


Figure 5.2 – (a) Difference between gaussian and tanh apodization profiles. Effect of a 300 K/mm chirp on a (b) Gaussian-apodized and a (c) tanh-apodized device.

apodization profile. We hence used a corrugation width profile given by

$$\Delta W(n) = \frac{\Delta W_0}{2} \left[1 + \tanh \left(\beta \left(1 - 2 \left| \frac{2n - N}{N} \right|^{\alpha} \right) \right) \right], \quad (5.1)$$

where $\Delta W(n)$ is the n^{th} period's corrugation width, ΔW_0 is the maximum corrugation width of either waveguide, N is the total number of grating periods of the device, and α and β are the tanh parameters seen in Eq. 5.1. $\alpha = 2$ and $\beta = 3$ were chosen to obtain a broad flat top while providing a smooth transition from no coupling to maximum coupling.

Figure 5.2 shows the difference between a Gaussian-shaped and a tanh-shaped apodization. When the device is linearly chirped, the phase-matching condition is swept linearly along the propagation direction. Having a flat-top apodization profile means that when chirped, all phase-matching conditions swept will experience the same coupling power, hence providing a broader spectral response [27]. Figure 5.2 also shows that a consequence of using a tanh apodization profile is that when the device is not chirped, the spectral response shows higher side-lobes. This is because the center section of the apodization profile acts like a uniform CDC, which inevitably brings about side-lobes. However, once a linear temperature chirp is applied on the device, the tanh side-lobes decrease considerably as the resonances are significantly suppressed due to the swept phase-matching condition. In addition, the tanh-apodized device's drop bandwidth becomes broader than a Gaussian-apodized device. In this work, a tanh apodization profile is hence used to provide more bandwidth tunability, at the expense of the passive response quality.

The segmented micro-heating device is designed with a $3.1\text{-}\mu\text{m}$ wide spine running along the whole length of the CDC. Eight electrodes and nine grounds, also $3.1\text{-}\mu\text{m}$ wide, are placed on either side of the heater spine, providing parallel resistance operation as shown in Fig. 5.1b. The heater spine runs $360\text{ }\mu\text{m}$ long and the positive electrodes are placed $40\text{ }\mu\text{m}$ apart. When a current is applied to an electrode, it directly sinks in parallel to the two closest common grounds. The segmented micro-heater is located in the silica cladding, $2\text{ }\mu\text{m}$ above

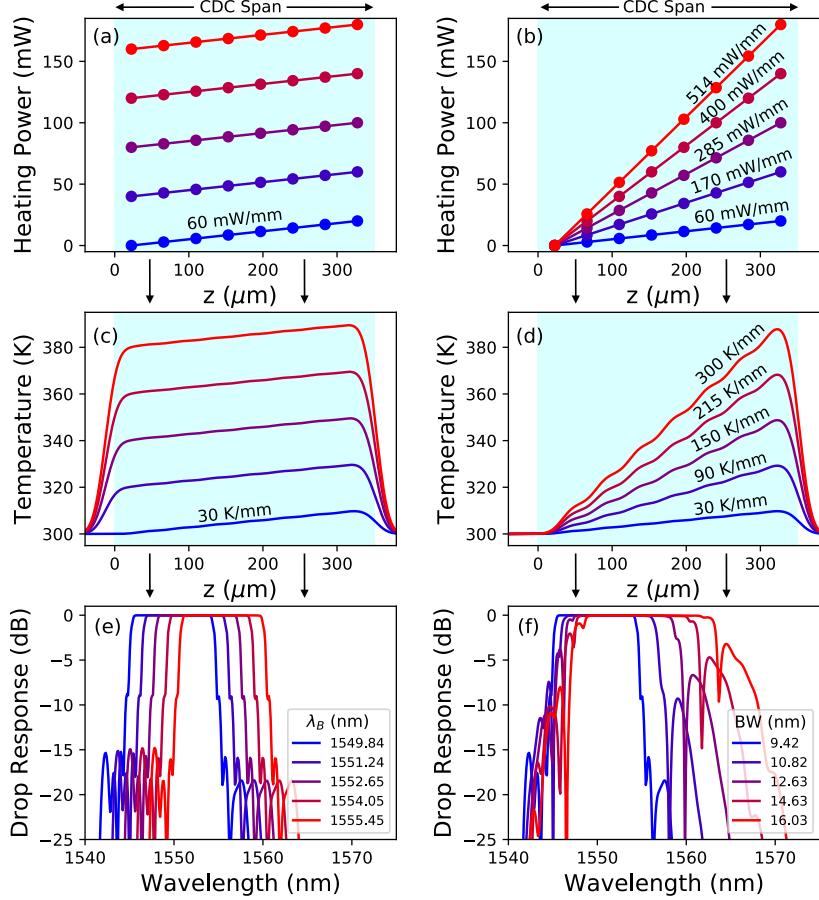


Figure 5.3 – Simulated demonstration of tunability, showing how the application of electrical powers (a-b) creates a temperature profile (c-d), which in turn tunes the response’s central wavelength (e) and bandwidth (f). The light blue region shows the spatial span of the CDC being tuned.

the silicon layer. This allows for a smoother temperature profile at the CDC level because the local temperature averages out between adjacent electrodes. Figure 5.3 shows the simulated effect of applying various electrical power profiles on the heater electrodes. The electro-thermal simulations were performed using Lumerical DEVICE’s finite element method implementation, and the resulting thermal profiles were fed into the CDC transfer-matrix model [57]. For central wavelength tunability, a small chirp of 30 K/mm is always applied to eliminate the side-lobes brought by the tanh apodization. By applying a temperature offset, the response can be shifted in wavelength. By changing the slope of the temperature profile, an arbitrary chirp can be created and the response’s bandwidth can be tuned. Hence, by applying a combination of offset and slope, the center wavelength and bandwidth can be tuned independently. The drop spectra in Fig. 5.3f show considerably higher side-lobes than those in Fig. 5.3e. These side-lobes are attributed to the relatively large granularity of the thermal tuning configuration. Since it is only composed of 8 electrodes, applying a steep chirp slope creates a stair-like

thermal profile, as seen in Fig. 5.3d. This deviation from a smooth, linear chirp profile results in higher side-lobes.

5.3 Results

The device was fabricated using Applied Nanotools's electron-beam lithography fabrication process and features a $3\text{-}\mu\text{m}$ buried oxide layer and a $2\text{-}\mu\text{m}$ oxide cladding. The micro-heater device is made of a 200-nm-thick Ti/W alloy with a sheet resistance of $3.07\ \Omega/\text{sq}$ [63]. The CDC's passive drop response is shown on Fig. 5.4. As expected, the use of the tanh apodization shape creates high side-lobes in the spectral response. Moreover, in-band ripples are present, which considerably limit the response's 3-dB bandwidth. These ripples are caused by the sharp transition in the tanh apodization profile. Since the coupling power goes from null to the maximum value very abruptly at the tanh edges, slight fabrication variations at these locations prove to be very destructive to the response quality.

While making temperature offsets and linear chirps possible, the proposed heater configuration also allows to counteract on-chip fabrication non-uniformities. Since CDCs are rather long integrated devices (typically hundreds of microns), they suffer from chip-scale waveguide width and height variations [64]. Considerable efforts are put toward making CDC-based filters more compact to avoid these variations [65]. Our segmented heater can be used to adapt to the fabrication non-uniformity by compensating for these variations in order to optimize the response quality.

Active operation of the tunable filter is performed using a tuning algorithm to fine-tune the response and meet certain performance criteria. For specified center wavelength and bandwidth figures, a target (ideal) spectral response as shown in Fig. 5.5 is considered.

The target response is a perfect rectangle with a floor at a given level. The choice of the

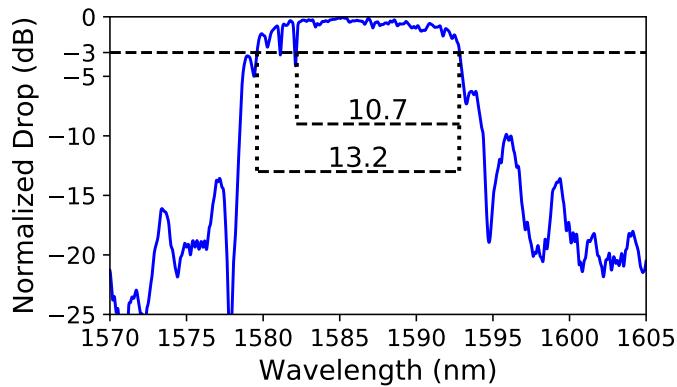


Figure 5.4 – Passive drop response of the CDC device. In-band notches limit the bandwidth to 10.7 nm, while the out-band rejection band is 13.2 nm wide.

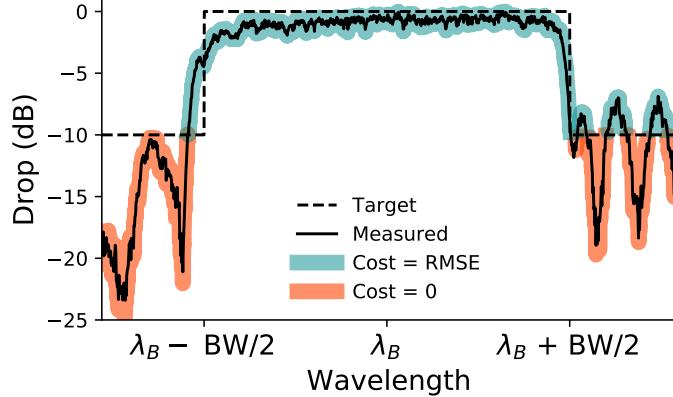


Figure 5.5 – Visual representation of the cost criterion used for algorithmic tuning. The cost is calculated as the root-mean-squared error in the blue-highlighted region. This measurement has a cost of 1.43.

floor value is empirical and based on the filter’s preliminary performance. For an efficient use of the algorithm, we chose a floor value that ensured a portion of the out-of-band response was always above the target. This way, the algorithm has leverage to converge to a solution because the defined cost function has a good dynamic range. A value of -10 dB was chosen for this experiment.

In decision theory, a cost (or loss) function is a mathematical mapping of a given optimization event onto a scalar number indicating the desirability of the event. It can be arbitrarily defined to suit a given problem and highly impacts the convergence of the algorithm. For a given experimental CDC drop response, our cost is defined as the root-mean-square error (RMSE) between the measured response and the target response. Outside of the reflection band, the RMSE is only considered when the measured response is higher than the target (i.e. side-lobes are present). When the out-of-band side-lobes are lower than the floor value of -10 dB, the cost is considered null. Figure 5.5 shows a visual representation of the cost criterion used and shows an example of a measured response and how it compares against the target response.

Initially, I-V curves are measured for each of the 8 heater segments, and their individual resistances are estimated. The applied voltages are then chosen to create a linear electrical power profile along the device. Then, the voltage is tuned by an optimization routine based on the Nelder-Mead algorithm that minimizes the measured cost function [66]. Figure 5.6 shows an example of the optimization process for a targeted reflection band centered at 1587 nm and 14-nm wide. Simply applying voltages based on the I-V regressions is enough to linearly chirp the CDC and eliminate the in-band ripples of the passive response seen in Fig. 5.4. Nevertheless, the response exhibits a poor side-lobe suppression ratio (SLSR) because the applied linear thermal chirp is not perfectly adapted to the as-fabricated CDC. By performing

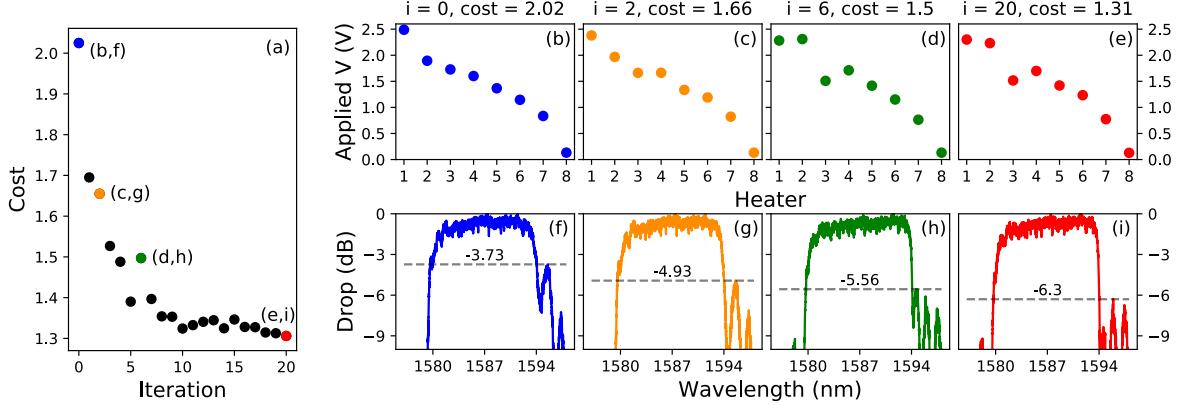


Figure 5.6 – Experimental demonstration of the Nelder-Mead algorithm used to tune the CDC response, with target center wavelength and bandwidth of 1587 and 14 nm respectively. (a) The convergence of the algorithm showing the cost (defined as per Fig. 5.5) at each iteration. Sample states of the filter tuning during the optimization process, showing the 8 voltages applied to the segmented micro-heater’s elements and the measured spectra after 0 (b,f), 2 (c,g), 6 (d,h) and 20 (e,i) iterations.

the Nelder-Mead optimization algorithm, the measured cost decreases exponentially over the iterations. As a result, the SLSR is successfully improved from an initial -3.73 to a final -6.3 dB after 20 iterations, which constitutes a 69% improvement. Meanwhile, the cost criterion is minimized from an initial 2.02 to a final 1.31. The performance of the fine-tuning algorithm is constrained by the limited number of micro-heating elements which creates granularity in the thermal profile. Additionally, the need for an initial I-V-curve-based estimation and the unforeseen presence of local minima in the parameter space are also factors to consider for future improvement.

To demonstrate the full tunability of the filter, it was operated using the optimization algorithm with various target spectra. Figure 5.7 shows the experimental drop-port responses for various target left and right band edges obtained by applying the same tuning algorithm to a single device. The left edge can be continuously tuned from 1580 to 1586 nm and the right edge can be continuously tuned from 1594 to 1606 nm, while maintaining a bandwidth between 14.0 and 22.4 nm. The drop response mainly remains flat-top. High-chirp micro-heater operation creates more prominent side-lobes because the steep chirps create stair-like thermal profiles. This effect can be seen in Fig. 5.7a and 5.7b, where the rightmost side-lobes become increasingly prominent with increasing chirp, as predicted in Fig. 5.3. This could be avoided by using a greater number of tuning electrodes. The measurements also display in-band wavelength-dependent loss, where the Bragg wavelengths last swept by the chirp suffer from their lengthy propagation inside the grating. It could be avoided by cascading two devices with inverted chirp slopes. This would in turn average out the in-band ripple effects and yield an even more flat-top response.

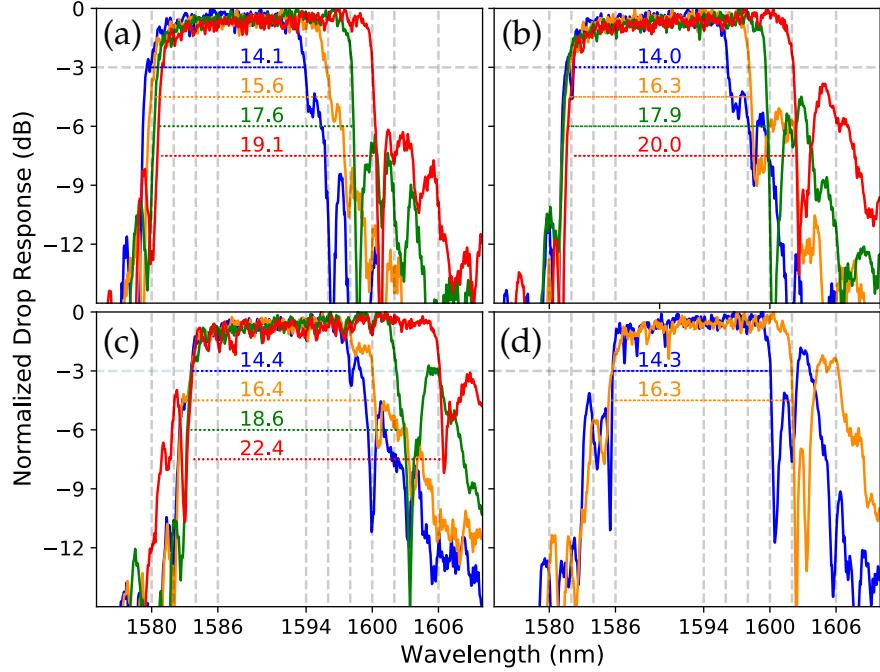


Figure 5.7 – Experimental demonstration of the left- and right-edge band tunability of the device, shown on the drop port response. Various right edge locations are shown for left edges located around (a) 1580 nm, (b) 1582 nm, (c) 1584 nm, and (d) 1586 nm. All responses were measured from the same device.

5.4 Conclusion

In conclusion, we have proposed a novel tunable SOI filter concept based on a single-stage tanh-apodized contradirectional coupler in combination with a segmented micro-heater to achieve thermal chirp and tuning. We have demonstrated that its algorithmic operation considerably improves the in-band ripples and side-lobe suppression ratio by counter-acting the on-chip fabrication non-uniformity. While side-lobe suppression ratios of 69% were achieved, designing a heating device with a larger number of segments with a more localized heat distribution can substantially improve this figure. Moreover, we have demonstrated that the chirped operation allows to continuously tune the center wavelength over 8 nm and the bandwidth over 8.4 nm. Tunability could be further increased by designing a passively chirped CDC in combination with dynamic tuning through the segmented micro-heater. Faster tuning is possible via free-carrier injection [67] at the cost of a higher insertion loss. The device can also be cascaded to further improve the performance such as in-band ripples and out-of-band suppression ratio. It holds great potential for applications such as elastic optical networks that require high-performance filters capable of dynamic bandwidth allocation.

Chapter 6

BraggNet: Complex Photonic Integrated Circuit Reconstruction Using Deep Learning

Résumé

Nous proposons l'utilisation d'un modèle d'apprentissage profond pour reconstruire la conception physique de systèmes complexes à cavités couplées à partir de leur réponse spectrale. Le modèle est démontré à l'aide de coupleurs contra-directionnels constitués de milliers de périodes de Bragg. Les coupleurs contra-directionnels sont modélisés comme des cavités optiques couplées, pour lesquelles un modèle par matrices de transfert est utilisé pour générer un jeu de données synthétique comprenant un espace de paramètres stratégique. Une approche modulaire est adoptée pour construire un réseau de neurones profond composé de 6 sous-modèles pré-entraînés pour faire des prédictions *a priori* et *a posteriori* de haute précision sur des quantités physiques spécifiques. Le modèle de forme libre et indépendant de l'architecture permet à l'utilisateur d'inclure toutes les géométries dans l'espace des paramètres de conception simplement en les incluant dans le jeu de données d'apprentissage. Après un entraînement approprié, le modèle atteint une erreur absolue moyenne de 1,4 % sur la reconstruction et donc se révèle adéquat pour les applications de rétro-conception. Pour montrer le potentiel de l'approche, un deuxième jeu de données est généré pour émuler les conditions post-fabrication d'un dispositif nominale entravé par des imperfections de fabrication. Le modèle est ensuite entraîné sur ce jeu de données et est utilisé pour reconstruire des dispositifs fabriqués à partir de mesures expérimentales, ce qui s'avère prometteur pour des applications de diagnostic post-fabrication.

Abstract

We propose a deep learning model to reconstruct physical designs of complex coupled photonic systems, such as waveguide Bragg gratings, from their spectral responses for inverse design and fabrication diagnosis. Traditional reconstructing algorithms demand considerable computing resources at every query. Conversely, machine learning algorithms use most of the computing resources during the training process and provide effortless and orders-of-magnitude faster analysis in response to queries. This approach is demonstrated using silicon photonic grating-assisted, contra-directional couplers consisting of thousands of Bragg periods. The contra-directional couplers are modeled as coupled cavities, for which a transfer matrix model is used to generate a synthetic dataset comprising a strategic design parameter space. The free-form, architecture-independent model allows to include any geometries to the design parameter space. Upon proper training, the model achieves 1.4% mean absolute percentage error on device reconstruction and thus proves suitable for inverse design applications. To further show its potential for assessment of fabricated devices, another dataset is generated to emulate the fabrication conditions of a nominal design hindered by fabrication imperfections. The model is shown to reconstruct devices from experimental measurements with greater than 600-fold improvement in speed compared to the classical layer-peeling algorithm. This proves promising for data-driven processes required by Industry 4.0.

6.1 Introduction

Silicon photonics offers scalability and low-cost-per-unit manufacturing of photonic integrated circuits (PICs) for a wide range of applications such as telecommunications [68], biosensors [69], and quantum computing [70]. This technology has attracted massive research attention in recent years, leading to significant progress in integrated photonic devices and systems. Most modern silicon PICs are developed based on sub-micrometer-thick waveguides with features comparable or even smaller than operating wavelengths. As a result, their performance is sensitive to wafer non-uniformity and fabrication errors [64], which is one of the main issues limiting their yield. Consequently, considerable effort is required on post-fabrication quality assessment and design for manufacturing, which involves massive data analysis. To address this challenge, we need more efficient modeling techniques for inverse design and reconstruction of physical parameters without using time-consuming, expensive metrology tools such as scanning electron microscope.

While design of single components often relies on rigorous electromagnetic simulations, circuit-level design scales rapidly in complexity. In the realm of linear optics, simulation paradigms such as transfer matrix models based on scattering matrices of optical components have been proposed and used as a powerful frequency-domain simulation tool [8]. Allowing for device encapsulation and versatility, such models rely on linear algebra operations that are com-

putationally efficient and can be accelerated by modern computing hardware like graphics processing units (GPUs). One of their fundamental limitations lies in the fact that linear operations performed in the frequency domain do not allow for direct inverse operations. In other words, the scattering matrix of a complete system encapsulates the system in a black box that does not allow easy retrieval of the individual sub-components.

Therefore, reconstructing a coupled photonic system from its frequency response usually requires much more complex algorithms. In particular, the current state of the art for reconstruction of complex Bragg gratings is the layer peeling (LP) algorithm based on time-domain feature extraction [71, 72]. While this model is based on analytical equations and hence physically accurate, it is sensitive to noise in the input data and becomes unstable when used with strongly-reflecting gratings. In addition, depending on the grating complexity, it can be computationally expensive because one has to run through the complex algorithm at every single query. For those reasons, it is hardly scalable to support large-scale analysis in a data-driven context such as qualification of fabrication imperfections or inverse design for manufacturing. This is where machine learning may play an important role.

Photonics has been proven to be greatly compatible with artificial intelligence (AI). At least three areas of compatibility have emerged, where AI is used to support photonics applications or *vice versa*. First, silicon photonics has been demonstrated a suitable platform to create neuromorphic networks that can be used as an AI accelerator hardware foundation [7]. Many computing functions such as matrix multiplication and non-linear operations can be emulated by photonic components. Implementation of such networks on a photonic chip can bring substantial performance improvement over their electronic counterparts. Second, machine learning techniques can be used to optimize fiber-optic networks. Notable applications include modulation format identification [73] and anomaly detection in optical networks [74]. Last but not the least, inverse design of photonic components can be greatly accelerated by cutting-edge machine learning algorithms. Photonic design optimization often involves a large parameter space which makes it challenging to converge to optimal solutions in a reasonable time. Machine learning algorithms such as stochastic gradient descent are very efficient at solving high-dimensionality problems. Such AI approaches have been demonstrated for simulation and/or inverse design of photonic components such as nanoslit arrays [42] and Bragg gratings [46]. These approaches often assume a fixed function (e.g., a skewed Gaussian function for the Bragg grating response [46]), whose parameters are to be predicted by the model. This approach is not suitable for fabrication diagnosis because fabricated devices do not necessarily follow the predefined functions due to wafer non-uniformity and fabrication errors.

In this work, we propose the utilization of machine learning to reconstruct physical parameters of complex coupled photonic devices such as Bragg gratings with complex phase and amplitude modulation profiles. For the proof of concept, we demonstrate a deep learning model named BraggNet to reconstruct as-fabricated non-uniform contra-directional coupler (CDC) filters

from their frequency responses. Their profiles are treated as pixels for which synthetic datasets are used to limit the model to a desired parameter space. The proposed strategy is to use a modular neural network (NN) model which can be trained on any synthetic dataset obtained from a transition matrix model. Unlike the approach suggested in [46], our free-form approach trains the model without using a predefined grating-profile function, thus allowing arbitrary designs to be learned from strategically-created datasets. It is this distinction that makes our model suitable for fabrication diagnosis, where a given as-fabricated device does not necessarily follow any particular shape.

The rest of this paper is organized as follows. First, Section 6.2 defines the problem and explains how our synthetic datasets are created. Then, in Section 6.3, we introduce the deep learning model’s architecture using modular prior and posterior prediction paradigm. In Section 6.4, we start by showing that the model is suitable for inverse design. From this baseline, we then assess the model performance in the presence of noise in the input data. Then, we show that it is suitable for fabrication diagnosis by training the model on a fabrication-emulating dataset and then feeding it with measured results of fabricated devices to successfully extract their as-fabricated characteristics. This mode of operation is particularly promising for Industry 4.0 applications, characterized by high-level automation, live data and machine learning. Section V concludes the paper.

6.2 Problem Definition

The proposed demonstration of complex system reconstruction uses contra-directional couplers, a Bragg filtering device that finds use in telecommunications [4] and quantum computing [6]. The basic operation principle of the CDC consists in designing an asymmetric waveguide coupler (i.e., a pair of optical waveguides with different propagation constants) whilst creating Bragg gratings on the waveguides to achieve effective coupling between two confined lateral modes propagating in opposite directions. It is essentially a four-port device where selected pass-band wavelengths are dropped while the remaining spectral content propagates through the input waveguide [31]. While uniform gratings are their simplest form, CDCs often incorporate non-uniform spatial profiles such as apodization and chirp in order to obtain specific performances [4]. In that case, the device is spatially segmented in sub-devices that exhibit their own spectral characteristics. A transfer matrix model (TMM) then can be used to cascade the coupled response of each individual segment along the grating to compute the overall complex electric field spectral response. A non-uniform CDC device is split into N_{seg} uniform grating segments. Fig. 6.1 shows the geometrical and optical characteristics of a single segment as well as the complete system. The mode coupling coefficients κ_{mn} are given by overlap integrals dictated by coupled-mode theory [14],

$$\kappa_{mn} = \frac{\omega}{4} \iint \mathbf{E}_m^*(x,y) \Delta\epsilon(x,y) \mathbf{E}_n(x,y) dy dx, \quad (6.1)$$

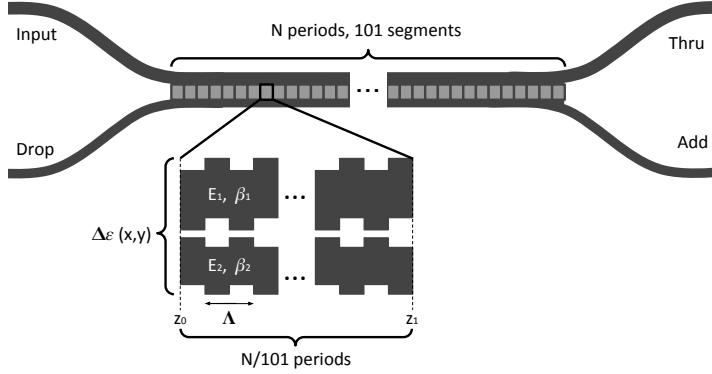


Figure 6.1 – Schematic of a non-uniform CDC device and one of its segments.

where \mathbf{E}_m and \mathbf{E}_n are the complex electric fields of the optical modes guided in waveguides m and n, respectively. The phase-matching condition is $\beta_B = \pi/\Lambda$. Hence, the waveguides' supported modes are detuned from the phase matching condition by

$$\begin{aligned}\Delta\beta_1 &= \beta_1 - \pi/\Lambda - j\alpha \\ \Delta\beta_2 &= \beta_2 - \pi/\Lambda - j\alpha,\end{aligned}\tag{6.2}$$

where α is the propagation loss parameter. For each uniform segment, interface matrix S_1 and propagation matrix S_2 are obtained as follows:

$$\begin{aligned}S_1 &= \begin{bmatrix} j\Delta\beta_1 & 0 & 0 & 0 \\ 0 & j\Delta\beta_2 & 0 & 0 \\ 0 & 0 & -j\Delta\beta_1 & 0 \\ 0 & 0 & 0 & -j\Delta\beta_2 \end{bmatrix}, \\ S_2 &= \begin{bmatrix} -j\Delta\beta_1 & 0 & & \\ 0 & -j\Delta\beta_2 & & \\ j\kappa_{11}^* e^{-2j\Delta\beta_1 z_1} & j\kappa_{12}^* e^{-j(\Delta\beta_1 + \Delta\beta_2)z_1} & & \\ j\kappa_{12}^* e^{-j(\Delta\beta_1 + \Delta\beta_2)z_1} & j\kappa_{22}^* e^{-2j\Delta\beta_2 z_1} & & \\ -j\kappa_{11} e^{2j\Delta\beta_1 z_1} & -j\kappa_{12} e^{j(\Delta\beta_1 + \Delta\beta_2)z_1} & & \\ -j\kappa_{12} e^{j(\Delta\beta_1 + \Delta\beta_2)z_1} & -j\kappa_{22} e^{2j\Delta\beta_2 z_1} & & \\ j\Delta\beta_1 & 0 & & \\ 0 & j\Delta\beta_2 & & \end{bmatrix}.\end{aligned}\tag{6.3}$$

The transfer matrix of the n^{th} segment is then given by

$$M_n = e^{(S_{1,n} + S_{2,n})(z_{1,n} - z_{0,n})},\tag{6.4}$$

where the matrix exponential is obtained by a Padé approximation of order 13 [75]. Provided that the CDC is separated in N_{seg} segments, the total transfer matrix of the system is obtained

by backward multiplication of the individual segments' transfer matrices:

$$M = \prod_{n=1}^{N_{seg}} M_{(N_{seg}-n+1)} \cdot \quad (6.5)$$

The drop and thru electric fields can then be obtained by basic operations on M [31]. Equation 6.5 shows that the system spectral response is given by N_{seg} cascaded 4 X 4 matrix multiplications, which is why it is difficult to extract individual segment matrices from the total transfer matrix. It is worth noting that the propagation loss parameter α in Eq. 6.2 is always positive (i.e., no gain is present). This ensures a reflectivity no greater than 0 dB.

For this demonstration, a dataset was created by using devices with 101 grating segments ($N_{seg} = 101$). A given segment is completely defined by three physical quantities: its number of grating periods, its coupling power κ and its phase-matching condition λ_B . Therefore, a complete CDC device is completely defined by its total number of grating periods N , its coupling power profile along the propagation direction $\kappa(z)$ (101 values total), and its phase-matching condition profile along the propagation direction $\lambda_B(z)$ (101 values total). With the proposed TMM model, the knowledge of these 203 numerical values allows to fully define a given CDC and simulate its spectral response. For the proof of concept, pseudo-random physical parameters were employed to simulate CDCs and create a dataset. It was decided that N would vary between 300 and 3000 periods, κ would have a Gaussian shape comprising maximum coupling values between 10 and 50 mm⁻¹ and Gaussian indices between 0 and 10, hence creating a variety of apodization profile shapes. The effective indices were set to fixed values 2.4 and 2.6, while the grating pitch was set to be between 305 and 315 nm, resulting in phase-matching conditions λ_B between 1525 and 1575 nm. All $\lambda_B(z)$ profiles were set to be linear and centered at 1550 nm, with random slopes. Technically speaking, the two pseudo-randomly generated grating profiles used floats, which provides a virtually infinite number of design combinations (limited by computer float precision).

A total of 18,262 such pseudo-random devices were generated and simulated, creating a synthetic dataset. The spectral response was simulated for 1001 wavelength values between 1500 and 1600 nm. For the reconstruction task, the 203 values defining the physical grating are the targets, and the drop amplitude and group delay spectra are the inputs. The dataset hence exhibits size reduction 2002 → 203. Since the data was randomly generated, a given sample is not necessarily a good design. It is just a random one generated from the parameter space defined above. In order to emulate experimental conditions, noise was added to the dataset spectra. Noisy transformations on the dataset were applied to add a Gaussian noise to the spectra. Standard deviations of 5 dB and 1 ps were applied on the drop amplitude and group delay spectra, respectively. Those figures were obtained from a commonly-used optical spectrum analyzer's specification sheet [76]. A noise floor of -50 dB was added to the data to increase realism. Fig. 6.2 shows a few random samples of the dataset.

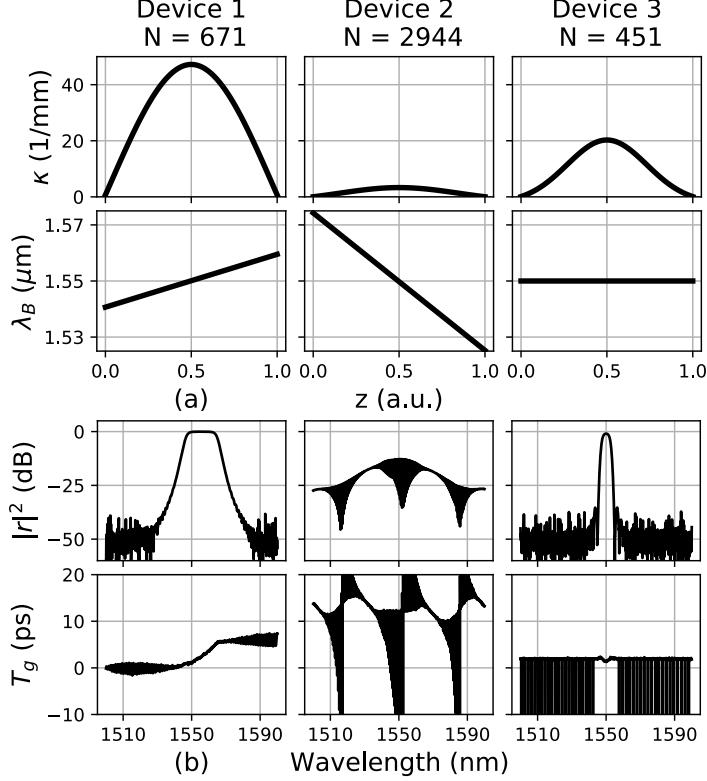


Figure 6.2 – Three samples from the created dataset. (a) Pseudo-randomly-generated grating physical parameters including the number of grating periods (N), the Gaussian apodization profile along the grating ($\lambda_B(z)$), and the linear coupling power profile along the grating ($\kappa(z)$), where z is the normalized propagation axis of light traveling inside the device. (b) The simulated spectral responses of the CDC devices with added noise, including the drop response ($|r|^2$) and the drop group delay (T_g).

6.3 Deep Learning Model

Inverse modeling of non-uniform CDC devices consists in finding a model f with parameters θ such that

$$N, \kappa(z), \lambda_B(z) = f \left(\frac{|r(\lambda)|^2}{\tau_g(\lambda)}, \theta \right), \quad (6.6)$$

which constitutes a high-dimensionality multivariate regression problem. As photonics designers, we can use prior knowledge of the system to choose a good model architecture and provide knowledge transfer. The CDC task suffers from its high dimensionality because the outputs can show inter-dependence. For instance, two different combinations of design parameters can yield very similar spectral responses. For this reason, the prior knowledge of certain targets can go a long way in predicting other targets of the system. For instance, filter bandwidth depends on both the number of grating periods and the coupling power of the device. Prior knowledge of one of those values can thus be useful in inferring the other one. A naive approach, as displayed in Fig. 6.3a, would consist in decoupling the three output variables and

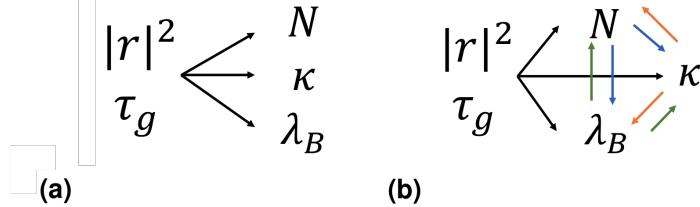


Figure 6.3 – (a) A naive approach to the problem, in which all outputs are decoupled. (b) A more complete approach, in which outputs are considered interdependent. The direction of the arrows indicates the flow of relevant information that can be used to reach an accurate solution.

training a unidirectional model to predict the design from the CDC’s spectra alone. However, in order to increase the coherence between the outputs and acknowledge their coupled nature, a more complete model is proposed, displayed in Fig. 6.3b. In this model, each output is dependent not only upon the model inputs, but also upon the other outputs. Prior knowledge is useful in predicting the inverse design. As highlighted by Bayes’ Theorem and Bayesian network theory [77], estimators widely benefit from the combined use of priors and posteriors because it restricts the parameter space dimensionality.

In order to obtain a complete dynamic model as displayed in Fig. 6.3b, as well as to provide encapsulation and scalability, the problem is separated into six sub-problems to which six sub-models are sought. For each output, one naive (prior) model and one complete (posterior, depending on other outputs) model are defined. These six sub-models are defined by functions f_1 through f_6 , given in Eq. 6.7. Upon obtaining such functions, the six models can interact with each other dynamically in order to create a complete neural network (NN) model.

$$\begin{array}{ll}
 \textit{Prior} & \textit{Posterior} \\
 \tilde{N}_{\text{naive}} = f_1 \left(\frac{|r|^2}{\tau_g} \right), & \tilde{N}_{\text{complete}} = f_2 \left(\frac{|r|^2}{\tau_g} \middle| \kappa, \lambda_B \right) \\
 \tilde{\kappa}_{\text{naive}} = f_3 \left(\frac{|r|^2}{\tau_g} \right), & \tilde{\kappa}_{\text{complete}} = f_4 \left(\frac{|r|^2}{\tau_g} \middle| N, \lambda_B \right) \\
 \tilde{\lambda}_B_{\text{naive}} = f_5 \left(\frac{|r|^2}{\tau_g} \right), & \tilde{\lambda}_B_{\text{complete}} = f_6 \left(\frac{|r|^2}{\tau_g} \middle| N, \kappa \right)
 \end{array} \tag{6.7}$$

6.4 Results

6.4.1 Inverse Design

The dataset was split in a 80-20 ratio between training and test data. The test data was set aside and never trained upon, in order to provide accurate performance assessment and avoid overfitting. Each individual sub-model is trained at performing its own simple task. This makes the learning process easier because the solution space is limited and models can converge relatively easily. This approach is commonly referred as transfer learning [78], since it uses the input as well as previously-acquired knowledge to generate a final output. Each

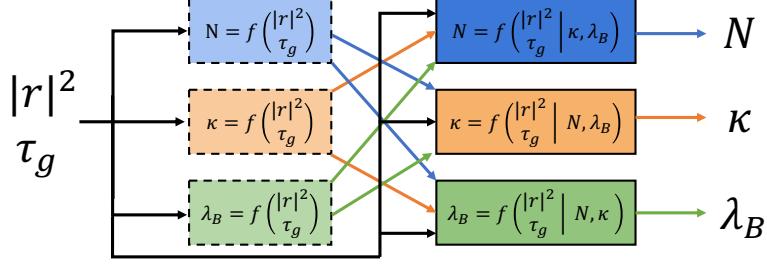


Figure 6.4 – General architecture of the BraggNet model. Six sub-models are trained to achieve either high prior or posterior performance. The drop amplitude and group delay spectra first enter the naive (prior) models in order to obtain zeroth-order estimations. Those estimations are then fed into the complete (posterior) models to obtain higher-accuracy estimations.

sub-model consists in a 3-layer fully-connected NN (one hidden layer) and various layers sizes. During the training process, techniques such as the Adam optimizing algorithm [35] and batch size increase [79] were used to minimize the loss criterion on the test dataset. Poutyne, a convenient training framework for PyTorch, was used to conduct most training routines [80]. A custom loss criterion was defined as the mean absolute percentage error (MAPE), averaged on the three physical quantities of interest (N , $\kappa(z)$ and $\lambda_B(z)$). Since the three quantities have different orders of magnitude, this allows to optimize the regression model uniformly.

Upon proper training of the six sub-models, they were cascaded as shown in Figure 6.4. Since all sub-models were pre-trained to perform their specific task, further training of the whole BraggNet allowed to reach the best performances. Like in a sport team, each player is good at their own position. However, team trainings are necessary in order to optimize the team performance.

Table 6.1 contains a summary of all trained models in terms of number of parameters and performance on the test data, while Fig. 6.5 shows its performance of three sample predictions. All sub-models show MAPEs below 3% on their predictions. Once the six pre-trained models are put together to form BraggNet and trained further, the complete trained model shows a

Table 6.1 – Summary of all trained models' number of parameters and performance on the test dataset.

| Model | Num. Params (10^3) | Mean Abs. % Error |
|------------------------|------------------------|-------------------|
| N - naive | 20.3 | 1.57 |
| N - complete | 23.9 | 0.70 |
| κ - naive | 200.8 | 2.78 |
| κ - complete | 351.9 | 2.02 |
| λ_B - naive | 200.8 | 0.71 |
| λ_B - complete | 262.0 | 0.48 |
| BraggNet | 1069.8 | 1.40 |

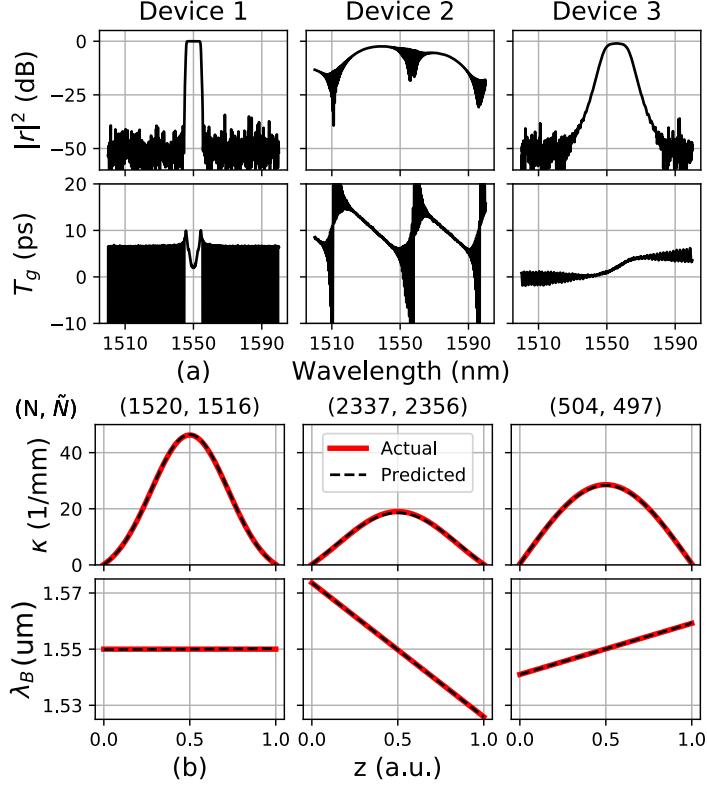


Figure 6.5 – Random examples of BraggNet’s reconstructions (predictions) of test data and comparison with the actual design values. N and \tilde{N} are the design number of periods and reconstructed number of periods, respectively.

MAPE of 1.40% on its device reconstructions. It is worth noting the effect of the use of a cascaded model as opposed to a naive one. The naive approach would consist in using the three naive models for predictions, which would result in a MAPE of 1.69% on all physical quantities. By using a complete, posterior approach, the performances are slightly higher. It is also worth noting that the average of the complete sub-models’ performances (1.07%) is significantly lower than BraggNet’s performance. This is because the complete sub-models, trained using actual target values, are used in BraggNet with estimations provided by the naive models. Therefore, it is reasonable to consider that the accuracy of the predicted priors used have an impact on the accuracy of BraggNet.

6.4.2 Noise Analysis

To use BraggNet on experimental data for fabrication diagnosis, a proper noise analysis is needed to ensure the usability of the model and assess what performance to expect. Since it features average pooling layers at its inputs, the model is mostly insensitive to the Gaussian noise applied on the transmission and group delay spectra because the averaging kernel operations act as filters. However, relative the noise floor present (relative to the maximum reflection) in a given drop spectrum plays a major role in the device reconstruction task.

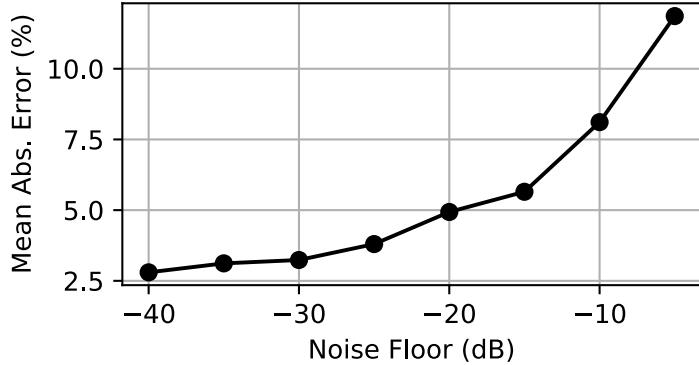


Figure 6.6 – Effect of the measuring instrument’s noise floor on the reconstruction performance of BraggNet. The dashed line shows the best-obtained performance without any noise applied to the dataset.

Side-lobe suppression ratio (SLSR) is mainly dependent on the apodization and chirp of the CDC device. Therefore, the noise floor level inevitably influences the accuracy of the possible reconstruction. In order to quantify this effect, a standard BraggNet training procedure was defined: 100 epochs at a learning rate of 0.01 followed by 50 epochs at a learning rate of 0.001. For different noise floor values, BraggNet was trained from scratch using the standard procedure to determine what reconstruction performance it could achieve with the given noise floor value. Figure 6.6 shows the MAPEs on the test data obtained during this experimentation. For high noise floor values, the error on reconstruction is very important and goes beyond 10% at the noise level of -8 dB. Therefore, the model is ineffective in presence of a high noise level due to the lack of information obtainable from the side lobes. The noise floor is referred to is relative, For experimental measurements of CDC devices, the relative noise floor is typically around -25 dB beneath the maximum reflection, which coincides with typical single-stage device SLSR values [81]. At this floor level, the training procedure yields a MAPE of 3.8% on test data.

6.4.3 Fabrication Diagnosis

To demonstrate the potential of the deep learning model for practical applications, it was further used to reconstruct fabricated CDC devices from their measured responses. In the case of inverse design, a wide parameter space was covered in the dataset in order to enable flexible and versatile device design. For fabrication diagnosis, a much more narrowed-down dataset was created to emulate on-chip fabrication variation and non-uniformity. Namely, a nominal design was chosen and stochastic variations were applied to the design to emulate fabrication imperfection. The chosen nominal design is a broadband apodized CDC with no chirp as shown in [4].

The modifications to the nominal design were all made randomly, following a normal distribu-

tion centered at the nominal value and with a determined standard deviation. The maximum and minimum coupling powers κ_{max} and κ_{min} of the apodization profile were offset from the nominal values. For the chirp profile, five points were randomly generated along the grating, with variations from the nominal values following a normal distribution. From those five points, a third-degree polynomial fit was obtained and from this fit, the chirp profile of the sample was determined. By choosing five points semi-randomly separated from each other, we ensured that the dataset comprised different frequency contents in the chirp variation (i.e. some samples have very low-frequency chirp while some other ones have higher-frequency chirp). These chirp profiles hence emulate on-chip index variations along the grating [82]. The distorted profiles were fed into the transfer matrix model that yielded the corresponding filter responses. A noise floor value of -25 dB on the drop spectrum and a Gaussian noise with an amplitude of 0.1 ps were applied to once again emulate typical experimental conditions observed with our instruments [76].

Additionally, other considerations were taken to make the dataset compliant with experimental data. Firstly, instead of defining a phase-matching conditions centered at 1550 nm on all samples, the phase matching condition is normalized in terms of relative detuning from λ_B . Since the phase matching condition is linear in terms of wavelength ($\lambda_B = 2\bar{n}_{eff}\Lambda$), this allows to apply the model on a measurement with any center wavelength. Secondly, instruments measuring group delays do not eliminate the effects of other components of the optical circuit such as routing waveguides and grating couplers. For this reason, there is an arbitrary offset between measured and simulated group delays. To normalize this situation, we set the group delay value to be zero at the phase-matching wavelength on all samples. Thirdly, experimental group delay is often very noisy outside the reflection band. Since there is negligible optical power reflected at these wavelengths, the value of the simulated/measured group delay has no physical meaning. By weighting the group delay with respect to drop-port optical power (in linear scale), we drastically suppress the noisy, meaningless group delay values. The weighted group delay is given by:

$$\tau_{g,w} = \tau_g 10^{\frac{drop \text{ [dB]}}{10}}. \quad (6.8)$$

An experimental-emulating dataset comprising 50,000 data samples was generated this way. Fig. 6.7 shows three samples of the dataset. It can be seen that the chirp profile has a significant impact on the quality of the filter's response, which can be distorted by wafer non-uniformity and fabrication errors.

BraggNet itself was slightly simplified to comply with experimental operation. While fabrication errors affect the yielded apodization and chirp profiles through smoothing effects and silicon height variations, the device's number of periods remains unhindered by fabrication and thus can be deemed exact. Consequently, there is no need for N-models to make predictions on N. Instead, a simple bypass is used on the model, as displayed in Fig. 6.8. The known value of N (1000) is fed directly to the posterior sub-models. This also contributes to increase

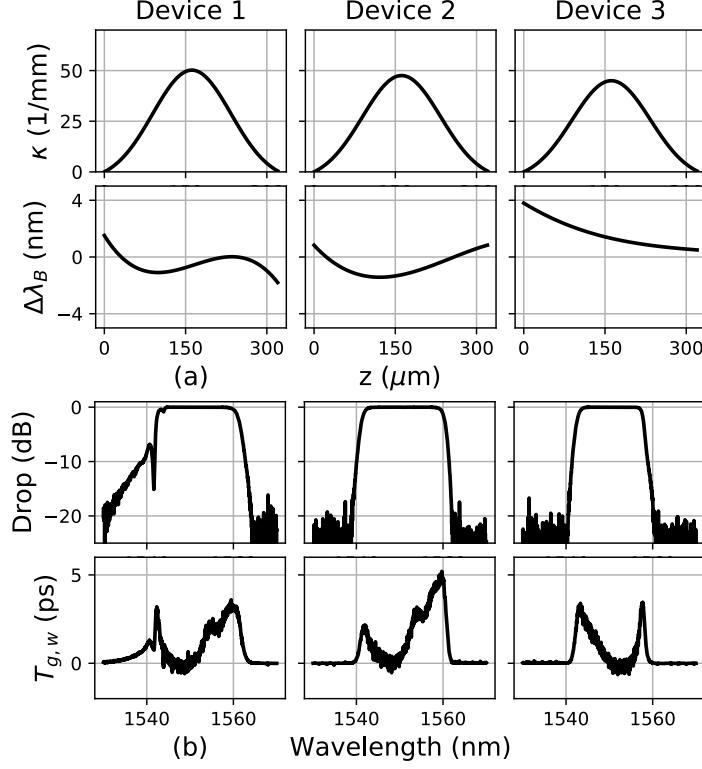


Figure 6.7 – Samples of the experimental-oriented dataset showing (a) the coupling and phase-matching condition profiles of the device along the propagation direction, and (b) the amplitude and group delay spectral responses of the fabrication-distorted devices.

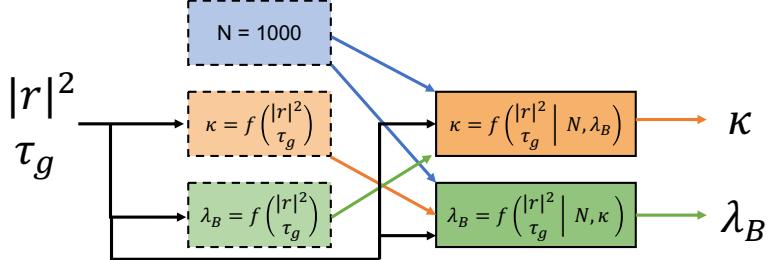


Figure 6.8 – Adaptation of BraggNet for applications on experimental data. The number of grating period remains unhindered by fabrication, which makes it known from the design. N-models can hence be replaced by the known value of 1000.

accuracy because posterior models will use an exact N value as opposed to the usual estimated N value. BraggNet was re-trained on this new dataset, following the same training routine as before. Performances similar to those shown in Fig. 6.6 were reached (with an MAE of 3.2%).

To demonstrate the fabrication diagnosis capabilities of the experimental-oriented model, some on-chip non-uniformities were manually created along the fabricated device. To do so, the device was fabricated with a segmented micro-heater $2 \mu\text{m}$ above it, allowing to create virtually-arbitrary temperature profiles that change the local effective index and therefore the local

phase-matching condition. This technique is used in fiber Bragg grating to generate chirp for delay compensation applications [60]. The device was then measured using different thermal chirps. The drop and group delay response were fed into BraggNet, as shown on Fig. 6.9. The grating profile predicted by BraggNet were then fed into the transfer matrix model for comparison with the experimental measurements.

Fig. 6.9 shows three measurement samples displaying different levels of chirp. A passive (not chirped), a lightly-chirp and a heavily-chirped device are each shown. In all three cases, it can be seen that the predicted coupling coefficient κ is smaller than the designed value. This is expected since lithography smoothing typically leads to bigger-than-designed inter-waveguide gaps and hence weaker coupling. Interestingly, the predicted κ_{max} on all three devices are similar despite the remarkable difference in bandwidth and filter shape. The model hence understood that this spectral widening comes from chirp instead of stronger coupling. Looking at the chirp reconstruction on Fig. 6.9b, it is evident that the model caught the chirping effect. As smoothed-out temperature profiles were applied along the device, BraggNet detected that the maximum phase mismatch increases with the bandwidth and shifts towards longer wavelength. Looking at the left-most chirp profile on Fig. 6.9, namely the passive-response chirp, it is seen that the reconstructed and predicted chirps are similar but not quite the same. This deviation is attributed to the on-chip fabrication non-uniformity, that typically cause effective index variations of $\Delta n \approx 1 \times 10^{-3}$ along the silicon waveguides [64].

Fig. 6.9c shows the output of the transfer matrix model, where the BraggNet-predicted grating profiles are fed and noise is applied. It can be seen that the reconstruction is very accurate at the drop and group delay level. That is, BraggNet picked up on the features underlying the chirp and apodization. Comparing the BraggNet-reconstructed responses to the experimental results, we can see that the overall behavior shows excellent agreement, while the small features (e.g., location of ripples) are slightly off. The weighted group delays also match well in the profile with mismatch in fine features.

This leads to understanding the limitations of the model. Since it is based on 101 grating segments, it treats each segment as a lumped entity. This brings granularity when used with experimental measurements. In our case, the devices have 1000 grating periods, which means that each grating segment comprises roughly 10 periods. Any fabrication variations occurring within the 10-period scale cannot be resolved by the model in training or in application. Therefore, the current model is more suitable for extraction of low-frequency features of CDC devices. In this regard, the performance may be greatly improved by increasing the resolution (i.e., the grating periods per segment) of the dataset. This, however, indicates processing a larger number of inputs that can be better handled by convolutional neural networks (CNNs) whose efficacy has been proven in image processing applications [83]. Note that although two of the three measured spectra in Fig. 6.9 show a stronger positive chirp because of random

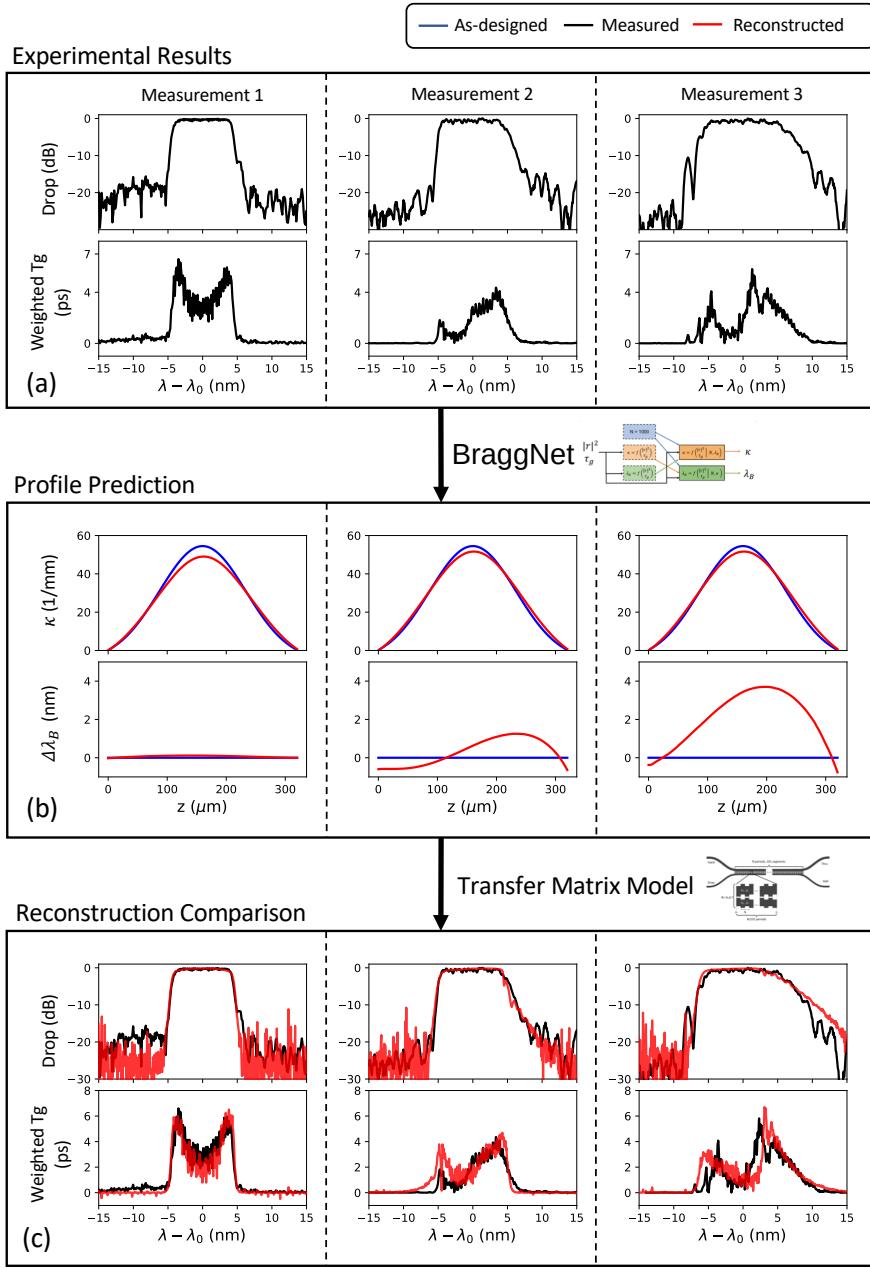


Figure 6.9 – Demonstration of BraggNet applied on experimental data. (a) three conditioned measurement of a CDC device under different thermal chirp conditions, including the measured drop and weighted group delay spectra. (b) BraggNet prediction of the apodization and chirp profile of the experimental data. (c) Comparison between the original measurements and the output of the transfer matrix model fed with BraggNet’s profile prediction for the three measurements, including the drop and weighted group delay spectra. The central wavelength was offset manually to match the experimental and reconstructed spectra.

fabrication errors, the model should perform equivalently well on negatively chirped devices, since the training dataset was generated symmetrically (e.g., Device 3 in Fig. 6.7).

We studied the speed of the approach and determined that the AI model running on a local laptop took approximately 1.5 ms to make a prediction, no matter the complexity of the device. On the same laptop, our classically-used layer-peeling algorithm takes an average of 900 ms for similar designs, and becomes slower for more complex and longer designs. Such a 600-fold speed improvement proves very useful in a mass-production context. BraggNet contains 1,069,800 parameters and only takes up 4 MB of storage, which makes it suitable for easy deployment in an industrial setting. Further investigation will be put towards model simplification to make it even faster. This could be once again achieved by employing CNNs.

6.5 Conclusion

We have demonstrated that a free-form deep learning approach can be used for efficient reconstruction (inverse design) of a non-uniform contra-directional coupler comprising a large number of grating segments, which is essentially a complex coupled-cavity optical system. With such a free-form approach, the model can learn any design parameter space included in the training dataset. The approach is modular because it comprises sub-models trained specifically for particular tasks. These sub-models are further trained together to form a comprehensive model, namely BraggNet. Our model has shown an efficient and accurate inverse-design capability in the realm of Gaussianly-apodized, linearly-chirped contra-directional couplers, reaching a mean absolute percentage error of 1.4% on validation data. A noise analysis was performed considering a standard training procedure. This showed that BraggNet is able to offer sub-5% mean absolute percentage error on validation data as long as the amplitude noise floor was below 20 dB.

To show the potential and versatility of BraggNet, a second dataset emulating fabrication variations and comprising realistic noise figures was created. Upon proper re-training and simple architecture simplifications, BraggNet achieved a 3.2% mean absolute percentage error on validation data. Not only does BraggNet offer a 600-fold speed improvement over the layer peeling algorithm, it is also more robust to noise. Due to its granular pixelization approach, BraggNet’s computation time independent of the device’s design, as opposed to the layer peeling algorithm whose computing time increases for gratings with a large number of periods. It was then used to reconstruct as-fabricated CDC devices with manually-induced thermal chirp profiles. Once fed back into the transfer-matrix model, these reconstructions proved to be accurate and capable of picking up on most low-frequency features such as bandwidth, band-edge angle and group delay profile. Additionally, querying the model takes as little as 1.5 ms on a conventional laptop, which is a 600-fold improvement over the classical layer-peeling algorithm. Despite its great potential, the model is currently limited in extraction of higher-

frequency variations along the grating and can be improved in future work by increasing the granularity of grating segments. We made all our source code available and open source, including the models, datasets, and dataset creation scripts [84].

Based on BraggNet’s performances for inverse design and comprehensive fabrication diagnosis, we expect to achieve eventually robust data-driven fabrication assessment and diagnosis. These techniques should be incorporated in automated and reliable fabrication assessment tools, playing an important role in the transition to photonic Industry 4.0.

Conclusion

Ultimately, this work has laid the foundation of silicon photonic Bragg-based devices and machine learning. It has then presented my original contributions, pertaining both to the hardware design and demonstration of novel devices, as well as to software for computationally-effective simulation, inverse design, fabrication-error compensation and fabrication quality assessment. In this sense, this work truly constitutes a valuable contribution in the hardware and software advances of Bragg-based integrated devices. While the demonstrations and results contained in this work prove very promising for eventual applications in the industry, major improvement work needs to be put toward making the proposed solutions more robust and performant.

At the hardware level, two main aspects of thermally-chirped contra-direction couplers need further optimization. First, the contra-directional coupler's passive response shows less-than ideal quality. I believe the future of the approach lies not in adopting cascaded devices, but rather to investigate and build a mathematical model for parallel-stacked contra-directional-coupling devices [85]. Second, the segmented heater device can be greatly improved by increasing the number of electrodes, thereby decreasing granularity in the temperature profile. Taken a step further, new shapes and architectures could be explored to create arbitrary temperature profiles with more power efficiency. My contra-directional coupler numerical simulation software will always remain open-source on Github [25].

At the software level, major improvements can be made to BraggNet, which will remain available open-source on Github as well [86]. First, it was considered that the input and output resolution limits the reconstruction precision of the model. Hence, employing grating profiles of significantly more than 101 grating segments would prove useful for enhanced performance. Second, fully-connected neural networks are the most basic types of neural networks. Convolutional neural networks (CNN) have demonstrated record-breaking performance in the field of image and speech recognition, which somewhat resembles the problem BraggNet is looking to tackle. In this sense, an updated BraggNet implementation using CNNs seems highly promising [87]. With proper training, such a deep learning model has the potential to propel silicon photonics design into the Industry 4.0 era and allow for data-driven mass production.

Bibliography

- [1] Jonathan Cauchon and Wei Shi. Dual-band optical filters using integrated multimode Bragg gratings. In *Optical Fiber Communication Conference (OFC) 2020*, page W2A.7. Optical Society of America, 2020.
- [2] Jonathan Cauchon, Jonathan St-Yves, and Wei Shi. Thermally chirped contra-directional couplers for residueless, bandwidth-tunable Bragg filters with fabrication error compensation. *Optics Letters*, 46(3):532–535, 2021.
- [3] IEEE Photonics Society. Call for papers. <https://www.photonicsociety.org/publications/journal-of-selected-topics-in-quantum-electronics/call-for-papers>. Online; accessed 5 April 2021.
- [4] Jonathan St-Yves, Hadi Bahrami, Philippe Jean, Sophie LaRochelle, and Wei Shi. Widely bandwidth-tunable silicon filter with an unlimited free-spectral range. *Optics Letters*, 40(23):5471–5474, 2015.
- [5] Lukas Chrostowski, Samantha Grist, Jonas Flueckiger, Wei Shi, Xu Wang, Eric Ouellet, Han Yun, Mitch Webb, Ben Nie, Zhen Liang, et al. Silicon photonic resonator sensors and devices. In *Laser Resonators, Microresonators, and Beam Control XIV*, volume 8236, page 823620. International Society for Optics and Photonics, 2012.
- [6] Abdelrahman Afifi, Lukas Chrostowski, Mustafa Hammood, Nicolas AF Jaeger, Sudip Shekhar, and Jeff F Young. Contra-directional couplers as pump rejection and recycling filters for on-chip photon-pair sources. In *2019 IEEE 16th International Conference on Group IV Photonics (GFP)*, volume 1949, pages 1–2. IEEE, 2019.
- [7] Alexander N Tait, Thomas Ferreira De Lima, Ellen Zhou, Allie X Wu, Mitchell A Nahmias, Bhavin J Shastri, and Paul R Prucnal. Neuromorphic photonic networks using silicon photonic weight banks. *Scientific Reports*, 7(1):1–10, 2017.
- [8] Lukas Chrostowski, Hossam Shoman, Mustafa Hammood, Han Yun, Jaspreet Jhoja, Enxiao Luan, Stephen Lin, Ajay Mistry, Donald Witt, Nicolas AF Jaeger, et al. Silicon photonic circuit design using rapid prototyping foundry process design kits. *IEEE Journal of Selected Topics in Quantum Electronics*, 25(5):1–26, 2019.
- [9] Mustafa Hammood, Ajay Mistry, Han Yun, Minglei Ma, Lukas Chrostowski, and Nicolas AF Jaeger. Four-channel, silicon photonic, wavelength multiplexer-demultiplexer with high channel isolations. In *2020 Optical Fiber Communications Conference and Exhibition (OFC)*. IEEE, 2020.
- [10] Charith Perera, Chi Harold Liu, Srimal Jayawardena, and Min Chen. A survey on internet of things from industrial market perspective. *IEEE Access*, 2:1660–1679, 2014.

- [11] Lynn Conway. Reminiscences of the VLSI revolution: how a series of failures triggered a paradigm shift in digital design. *IEEE Solid-State Circuits Magazine*, 4(4):8–31, 2012.
- [12] Lukas Chrostowski and Michael Hochberg. *Silicon photonics design: from devices to systems*. Cambridge University Press, 2015.
- [13] Amnon Yariv. Coupled-mode theory for guided-wave optics. *IEEE Journal of Quantum Electronics*, 9(9):919–933, 1973.
- [14] Amnon Yariv and Pochi Yeh. *Photonics: optical electronics in modern communications*. Oxford University, 2006.
- [15] Joé Habel, Tommy Boilard, Jean-Simon Frenière, François Trépanier, and Martin Bernier. Femtosecond fbg written through the coating for sensing applications. *Sensors*, 17(11):2519, 2017.
- [16] Jens Buus, Markus-Christian Amann, and Daniel J Blumenthal. *Tunable laser diodes and related optical sources*. John Wiley & Sons, 2005.
- [17] Pochi Yeh and HF Taylor. Contradirectional frequency-selective couplers for guided-wave optics. *Applied Optics*, 19(16):2848–2855, 1980.
- [18] Yun Wang, Xu Wang, Jonas Flueckiger, Han Yun, Wei Shi, Richard Bojko, Nicolas AF Jaeger, and Lukas Chrostowski. Focusing sub-wavelength grating couplers with low back reflections for rapid prototyping of silicon photonic circuits. *Optics Express*, 22(17):20652–20662, 2014.
- [19] Christos Riziotis and Mikhail N Zervas. Design considerations in optical add/drop multiplexers based on grating-assisted null couplers. *Journal of Lightwave Technology*, 19(1):92, 2001.
- [20] J-P Weber. Spectral characteristics of coupled-waveguide Bragg-reflection tunable optical filter. *IEE Proceedings J (Optoelectronics)*, 140(5):275–284, 1993.
- [21] Lumerical inc. Comprehensive Optical Waveguide Design Environment. <https://www.lumerical.com/products/mode/>. Online; accessed 5 April 2021.
- [22] Guido Van Rossum and Fred L. Drake. *Python 3 Reference Manual*. CreateSpace, Scotts Valley, CA, 2009.
- [23] Charles R. Harris, K. Jarrod Millman, St'efan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernandez del Rio, Mark Wiebe, Pearu Peterson, Pierre G'erard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi,

- Christoph Gohlke, and Travis E. Oliphant. Array programming with NumPy. *Nature*, 585(7825):357–362, September 2020.
- [24] Bradley J Frey, Douglas B Leviton, and Timothy J Madison. Temperature-dependent refractive index of silicon and germanium. In *Optomechanical technologies for Astronomy*, volume 6273, page 62732J. International Society for Optics and Photonics, 2006.
- [25] Jonathan Cauchon. Contra-Directional Coupler Model Written in Python.
<https://github.com/JonathanCauchon/Contra-DC>. Online; accessed 5 April 2021.
- [26] Jonathan St-Yves. Contra-directional couplers as optical filters on the silicon on insulator platform. Master’s Thesis, Université Laval, 2017.
- [27] İsa Navruz and Ahmet Altuncu. Design of a chirped fiber Bragg grating for use in wideband dispersion compensation. In *New Trends In Computer Networks*, pages 114–123. World Scientific, 2005.
- [28] Raman Kashyap. *Fiber Bragg gratings*. Academic Press, 2009.
- [29] Han Yun, Mustafa Hammod, Stephen Lin, Lukas Chrostowski, and Nicolas AF Jaeger. Broadband flat-top SOI add-drop filters using apodized sub-wavelength grating contradirectional couplers. *Optics Letters*, 44(20):4929–4932, 2019.
- [30] Han Yun, Mustafa Hammod, Lukas Chrostowski, and Nicolas AF Jaeger. Optical add-drop filters using cladding-modulated sub-wavelength grating contra-directional couplers for silicon-on-insulator platforms. In *2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, pages 0926–0931. IEEE, 2019.
- [31] Wei Shi, Xu Wang, Charlie Lin, Han Yun, Yang Liu, Tom Baehr-Jones, Michael Hochberg, Nicolas AF Jaeger, and Lukas Chrostowski. Silicon photonic grating-assisted, contra-directional couplers. *Optics Express*, 21(3):3633–3650, 2013.
- [32] Masahiko Jinno, Hidehiko Takara, Bartłomiej Kozicki, Yukio Tsukishima, Yoshiaki Sone, and Shinji Matsuoka. Spectrum-efficient and scalable elastic optical path network: architecture, benefits, and enabling technologies. *IEEE Communications Magazine*, 47(11), 2009.
- [33] Ethem Alpaydin. *Introduction to machine learning*. MIT press, 2020.
- [34] Kobus Barnard, Pinar Duygulu, David Forsyth, Nando de Freitas, David M Blei, and Michael I Jordan. Matching words and pictures. *Journal of Machine Learning Research*, 3(Feb):1107–1135, 2003.

- [35] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [36] Mario Miscuglio and Volker J Sorger. Photonic tensor cores for machine learning. *Applied Physics Reviews*, 7(3):031404, 2020.
- [37] Philipp-Immanuel Schneider, Xavier Garcia Santiago, Victor Soltwisch, Martin Hammerschmidt, Sven Burger, and Carsten Rockstuhl. Benchmarking five global optimization approaches for nano-optical shape optimization and parameter reconstruction. *ACS Photonics*, 6(11):2726–2733, 2019.
- [38] Kan Yao, Rohit Unni, and Yuebing Zheng. Intelligent nanophotonics: merging photonics and artificial intelligence at the nanoscale. *Nanophotonics*, 8(3):339–366, 2019.
- [39] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4401–4410, 2019.
- [40] Jiaqi Jiang, David Sell, Stephan Hoyer, Jason Hickey, Jianji Yang, and Jonathan A Fan. Free-form diffractive metagrating design based on generative adversarial networks. *ACS nano*, 13(8):8872–8878, 2019.
- [41] Zhaocheng Liu, Dayu Zhu, Sean P Rodrigues, Kyu-Tae Lee, and Wenshan Cai. Generative model for the inverse design of metasurfaces. *Nano Letters*, 18(10):6570–6576, 2018.
- [42] André-Pierre Blanchard-Dionne and Olivier JF Martin. Teaching optics to a machine learning network. *Optics Letters*, 45(10):2922–2925, 2020.
- [43] Sunae So, Jungho Mun, and Junsuk Rho. Simultaneous inverse design of materials and structures via deep learning: Demonstration of dipole resonance engineering using core–shell nanoparticles. *ACS Applied Materials & Interfaces*, 11(27):24264–24268, 2019.
- [44] Alex Zunger. Inverse design in search of materials with target functionalities. *Nature Reviews Chemistry*, 2(4):1–16, 2018.
- [45] Laura Pilozzi, Francis A Farrelly, Giulia Marcucci, and Claudio Conti. Machine learning inverse problem for topological photonics. *Communications Physics*, 1(1):1–7, 2018.
- [46] Alec M Hammond and Ryan M Camacho. Designing integrated photonic devices using artificial neural networks. *Optics Express*, 27(21):29620–29638, 2019.
- [47] Itzik Malkiel, Michael Mrejen, Achiya Nagler, Uri Arieli, Lior Wolf, and Haim Suchowski. Plasmonic nanostructure design and characterization via deep learning. *Light: Science & Applications*, 7(1):1–8, 2018.

- [48] Wei Ma, Feng Cheng, and Yongmin Liu. Deep-learning-enabled on-demand design of chiral metamaterials. *ACS nano*, 12(6):6326–6334, 2018.
- [49] Mohammad H Tahersima, Keisuke Kojima, Toshiaki Koike-Akino, Devesh Jha, Bingnan Wang, Chungwei Lin, and Kieran Parsons. Deep neural network inverse design of integrated nanophotonic devices. *arXiv preprint arXiv:1809.03555*, 2018.
- [50] Iridan. Dual-band CWDM. <https://www.iridian.ca/product-category/telecom-filters/cwdm-dual-band/>. Online; Accessed 20 December 2019.
- [51] Qingzhong Huang, Xinliang Zhang, Jinsong Xia, and Jinzhong Yu. Dual-band optical filter based on a single microdisk resonator. *Optics Letters*, 36(23):4494–4496, 2011.
- [52] Huiye Qiu, Jianfei Jiang, Ting Hu, Ping Yu, Jianyi Yang, Xiaoqing Jiang, and Hui Yu. Silicon add-drop filter based on multimode Bragg sidewall gratings and adiabatic couplers. *Journal of Lightwave Technology*, 35(9):1705–1709, 2017.
- [53] Toru Mizunami, Tzvetanka V Djambova, Tsutomu Niiho, and Sanjay Gupta. Bragg gratings in multimode and few-mode optical fibers. *Journal of Lightwave Technology*, 18(2):230–235, 2000.
- [54] Dominique Charron and Wei Shi. O-band add-drop filter in Bragg-grating-assisted mach-zehnder interferometers for cwidm. In *CLEO: Applications and Technology*, pages JTh2A–44. Optical Society of America, 2019.
- [55] Bo Zhang, David Leuenberger, Ming-Chang M Lee, Alan E Willner, and Ming C Wu. Experimental demonstration of dynamic bandwidth allocation using a mems-actuated bandwidth-tunable microdisk resonator filter. *IEEE Photonics Technology Letters*, 19(19):1508–1510, 2007.
- [56] Jun Rong Ong, Ranjeet Kumar, and Shayan Mookherjea. Ultra-high-contrast and tunable-bandwidth filter using cascaded high-order silicon microring filters. *IEEE Photonics Technology Letters*, 25(16):1543–1546, 2013.
- [57] Wei Shi, Xu Wang, Charlie Lin, Han Yun, Yang Liu, Tom Baehr-Jones, Michael Hochberg, Nicolas AF Jaeger, and Lukas Chrostowski. Silicon photonic grating-assisted, contra-directional couplers. *Optics Express*, 21(3):3633–3650, 2013.
- [58] Wei Shi, Han Yun, Charlie Lin, Xu Wang, Jonas Flueckiger, Nicolas Jaeger, and Lukas Chrostowski. Silicon CWDM demultiplexers using contra-directional couplers. In *CLEO: Science and Innovations*, pages CTu3F–5. Optical Society of America, 2013.
- [59] Hao Sun, Yue Wang, and Lawrence R Chen. Integrated discretely tunable optical delay line based on step-chirped subwavelength grating waveguide Bragg gratings. *Journal of Lightwave Technology*, 38(19):5551–5560, 2020.

- [60] Yves Painchaud, Carl Paquet, and Martin Guy. Optical tunable dispersion compensators based on thermally tuned fiber Bragg gratings. *Optics and Photonics News*, 18(9):48–53, 2007.
- [61] Charalambos Klitis, Marc Sorel, and Michael J Strain. Active on-chip dispersion control using a tunable silicon Bragg grating. *Micromachines*, 10(9):569, 2019.
- [62] Hajime Sakata. Sidelobe suppression in grating-assisted wavelength-selective couplers. *Optics Letters*, 17(7):463–465, 1992.
- [63] Applied Nanotools inc. Process specifications. <https://www.appliednt.com/nanosoi/sys/resources/specs/>. Online; Accessed 19 October 2020.
- [64] Zeqin Lu, Jaspreet Jhoja, Jackson Klein, Xu Wang, Amy Liu, Jonas Flueckiger, James Pond, and Lukas Chrostowski. Performance prediction for silicon photonics integrated circuits with layout-dependent correlated manufacturing variability. *Optics Express*, 25(9):9712–9733, 2017.
- [65] Minglei Ma, Zhitian Chen, Han Yun, Yun Wang, Xu Wang, Nicolas AF Jaeger, and Lukas Chrostowski. Apodized spiral Bragg grating waveguides in silicon-on-insulator. *IEEE Photonics Technology Letters*, 30(1):111–114, 2017.
- [66] John A Nelder and Roger Mead. A simplex method for function minimization. *The Computer Journal*, 7(4):308–313, 1965.
- [67] Weifeng Zhang and Jianping Yao. A fully reconfigurable waveguide Bragg grating for programmable photonic signal processing. *Nature communications*, 9(1):1–9, 2018.
- [68] Wei Shi, Ye Tian, and Antoine Gervais. Scaling capacity of fiber-optic transmission systems via silicon photonics. *Nanophotonics*, 9(16):4629–4633, 2020.
- [69] Maria Soler, Olalla Calvo-Lozano, M-Carmen Estevez, and Laura M Lechuga. Nanophotonic biosensors: Driving personalized medicine. *Optics and Photonics News*, 31(4):24–31, 2020.
- [70] Xiaogang Qiang, Xiaoqi Zhou, Jianwei Wang, Callum M Wilkes, Thomas Loke, Sean O’Gara, Laurent Kling, Graham D Marshall, Raffaele Santagati, Timothy C Ralph, et al. Large-scale silicon quantum photonics implementing arbitrary two-qubit processing. *Nature Photonics*, 12(9):534–539, 2018.
- [71] Ricardo Feced, Michalis N Zervas, and Miguel A Muriel. An efficient inverse scattering algorithm for the design of nonuniform fiber Bragg gratings. *IEEE Journal of Quantum Electronics*, 35(8):1105–1115, 1999.

- [72] A Buryak, J Bland-Hawthorn, and V Steblina. Comparison of inverse scattering algorithms for designing ultrabroadband fibre Bragg gratings. *Optics Express*, 17(3):1995–2004, 2009.
- [73] Faisal Nadeem Khan, Kangping Zhong, Waled Hussein Al-Arashi, Changyuan Yu, Chao Lu, and Alan Pak Tao Lau. Modulation format identification in coherent receivers using deep machine learning. *IEEE Photonics Technology Letters*, 28(17):1886–1889, 2016.
- [74] David Côté. Using machine learning in communication networks. *Journal of Optical Communications and Networking*, 10(10):D100–D109, 2018.
- [75] Awad H Al-Mohy and Nicholas J Higham. A new scaling and squaring algorithm for the matrix exponential. *SIAM Journal on Matrix Analysis and Applications*, 31(3):970–989, 2010.
- [76] Luna Innovations. *OVA 5000 Optical Vector Analyzer*, 7 2020.
- [77] Finn V Jensen et al. *An introduction to Bayesian networks*, volume 210. UCL Press London, 1996.
- [78] Lisa Torrey and Jude Shavlik. Transfer learning. In *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*, pages 242–264. IGI global, 2010.
- [79] Samuel L Smith, Pieter-Jan Kindermans, Chris Ying, and Quoc V Le. Don’t decay the learning rate, increase the batch size. *arXiv preprint arXiv:1711.00489*, 2017.
- [80] Frédéric Paradis. Poutyne: Keras-like framework for py-torch, 2018.
- [81] Mustafa Hammod, Ajay Mistry, Minglei Ma, Han Yun, Lukas Chrostowski, and Nicolas AF Jaeger. Compact, silicon-on-insulator, series-cascaded, contradirectional-coupling-based filters with > 50 dB adjacent channel isolation. *Optics Letters*, 44(2):439–442, 2019.
- [82] Alexandre D Simard, Guillaume Beaudin, Vincent Aimez, Yves Painchaud, and Sophie LaRochelle. Characterization and reduction of spectral distortions in silicon-on-insulator integrated Bragg gratings. *Optics Express*, 21(20):23145–23159, 2013.
- [83] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 25:1097–1105, 2012.
- [84] Jonathan Cauchon. Braggnnet, 2021. <https://github.com/JonathanCauchon/BraggNet>.
- [85] Mustafa Hammod, Ajay Mistry, Minglei Ma, Lukas Chrostowski, and Nicolas AF Jaeger. Compact contra-directional-coupler-based filters for cwdm applications. In *2018 IEEE 15th International Conference on Group IV Photonics (GFP)*, pages 1–2. IEEE, 2018.

- [86] Jonathan Cauchon. BraggNet: Complex Photonic Integrated Circuit Reconstruction Using Deep Learning. <https://github.com/JonathanCauchon/BraggNet>. Online; accessed 5 April 2021.
- [87] Keiron O’Shea and Ryan Nash. An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458*, 2015.