UNIVERSITY OF SOUTHAMPTON

FACULTY OF ENGINEERING AND PHYSICAL SCIENCES

Zepler Institute for Photonics and Nanoelectronics

Optoelectronics Research Centre

All-Optical Characterisation and Wavefront Manipulation of Photonic Integrated Circuits

by

Nicholas Joseph Dinsdale

Thesis for the degree of Doctor of Philosophy

June 2021

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF ENGINEERING AND PHYSICAL SCIENCES
Zepler Institute for Photonics and Nanoelectronics
Optoelectronics Research Centre

Doctor of Philosophy

ALL-OPTICAL CHARACTERISATION AND WAVEFRONT MANIPULATION OF PHOTONIC INTEGRATED CIRCUITS

by Nicholas Joseph Dinsdale

Advanced photonic probing techniques are of great importance for the development of non-contact wafer-scale testing of Photonic Integrated Circuits (PICs). The most advanced characterisation of a PIC is given by the electromagnetic field distribution of the light propagating within it. Obtaining such a complete description typically requires complex photonic probing techniques which utilise a scanning perturbation placed in the near-field to infer information on the fields within. However, such approaches are limited with most real-world devices due to high-cost, low-throughput testing and the use of protective cladding layers that block access to the near-field.

Here, we demonstrate Ultrafast Photomodulation Spectroscopy (UPMS) as a far-field technique for the non-destructive characterisation of individual PIC elements. A scanning optical pump creates a highly localised perturbation in the refractive index profile of a silicon waveguide via free-carrier excitation, which can be raster-scanned over the device to map the internal electric field distributions without requiring direct access to the near-field. Sub-micrometer resolution is obtained, and direct comparison to a rapid and rigorous analytical model enables the quantitative comparison to an ideal design structure. In addition to constituting a promising route for the testing and diagnostics of PICs, the pump-probe system can also be used for the real-time trimming of photonic devices with active feedback. The optical annealing of ion-implanted racetrack resonators is shown here to be capable of locating the critical coupling position with an accuracy of $\pm 200\,\mathrm{nm}$. Multiple refractive index perturbations can be used for intricate wavefront manipulation of photonic structures with complex functionalities, such as optical routers, modulators, (de)multiplexers and mode converters. In this work, perturbation patterns are initially designed using brute-force iterative techniques, leading to the development of a deep-Artificial Neural Network (ANN) that is capable of rapidly ($\ll 1\,\mathrm{ms}$) and accurately ($<6\,\%$ mean transmittance error) generating new perturbation patterns with arbitrary complex transmission matrices for universal optical components.

Contents

List of Publications and Presentations

Publications:

I. A. Ajia, JY. Ou, N. J. Dinsdale, H. J. Singh, T. Chen-Sverre, T. Liu, N. I. Zheludev and O. L. Muskens, "Gigahertz nano-optomechanical resonances in a dielectric SiC-membrane metasurface array," *Nano Letters*, vol. 21, no. 11, pp. 4563–4569, 2021.

N. J. Dinsdale, P. R. Wiecha, M. Delaney, J. D. Reynolds, M. Ebert, I. Zeimpekis, D. J. Thomson, G. T. Reed, P. Lalanne, K. Vynck and O. L. Muskens, "Deep learning enabled design of complex transmission matrices for universal optical components," *ACS Photonics*, vol. 8, no. 1, pp. 283–295, 2021.

M. M. Milosevic, X. Chen, X. Yu, N. J. Dinsdale, O. Aktas, S. Z. Oo, A. Z. Khokhar, D. J. Thomson, O. L. Muskens, H. M. H. Chong, A. C. Peacock, S. Saito, and G T. Reed, "Ion implantation of germanium into silicon for critical coupling control of racetrack resonators," *Journal of Lightwave Technology*, vol. 38, no. 7, pp. 1865–1873, 2020.

K. Vynck*, N. J. Dinsdale*, B. Chen, R. Bruck, A. Z. Khokhar, S. A. Reynolds, L. Crudgington, D. J. Thomson, G. T. Reed, P. Lalanne, and O. L. Muskens, "Ultrafast perturbation maps as a quantitative tool for testing of multi-port photonic devices," *Nature Communications*, vol. 9, no. 1, p. 2246, 2018.

*Equal authorship contribution.

Conference proceedings:

O. L. Muskens, M. Delaney, N. Dinsdale, I. Zeimpekis, D. Hewak, G. Reed and D. Lawson "Reconfigurable nanophotonics using ultrafast photomodulation and phase change materials," Oral presentation at *SPIE Nanoscience + Engineering, Active Photonic Platforms XII*, in San Diego, USA, 2020.

V. Mittal, R. Petra, A. Tarazona, N. J. Dinsdale, A. Khokar, M. Banakar, D. Thomson, O. L. Musken, G. T. Reed, and H. M. H. Chong, "Vertical directional coupler network using a-Si slope waveguides," Oral presentation at *AOS Australian Conference on Optical Fibre Technology (ACOFT) and Australian Conference on Optics, Lasers, and Spectroscopy (ACOLS) 2019*, in Melbourne, Australia, 2019.

N. J. Dinsdale, K. Vynck, M. Milosevic, P. Lalanne, G. T. Reed, and O. L. Muskens, "Ultrafast perturbation mapping for characterization of photonic integrated circuits," Oral presentation at *CLEO Europe, Photonic Integration*, in Munich, Germany, 2019.

K. Vynck, N. J. Dinsdale, R. Bruck, B. Chen, G. T. Reed, P. Lalanne, and O. L. Muskens, "Remote optical testing and reconfiguration of silicon photonic circuits with ultrafast photomodulation spectroscopy," Oral presentation at *PIERS*, in Rome, Italy, 2019.

O. L. Muskens, B. Chen, N. Dinsdale, R. Bruck, D. J. Thomson, G. Mashanovich, G. T. Reed, K. Vynck, and P. Lalanne, "New all-optical and plasmonic strategies for controlling light on a silicon chip," Oral presentation at *CLEO, QELS Fundamental Science*, in San Jose, USA, 2018.

B. Chen, R. Bruck, N. J. Dinsdale, D. J. Thomson, G. Z. Mashanovich, G. T. Reed, and O. L. Muskens, "Coherent absorption with plasmonics nanoantennas on SOI racetrack resonator," Oral presentation at *SPIE Europe, Silicon Photonics: from Fundamental Research to Manufacturing*, in Strasbourg, France, 2018.

O. L. Muskens, R. Bruck, N. Dinsdale, K. Vynck, P. Lalanne, G. Z. Mashanovich, and G. T. Reed, "Reconfigurable silicon photonics: shaping light on a chip," Oral presentation at *SPIE Nanoscience + Engineering, Active Photonic Platforms IX*, in San Diego, USA, 2017.

Underline indicates presenting author.

Declaration of Authorship

I, Nicholas Joseph Dinsdale , declare that the thesis entitled *All-Optical Characterisation and Wavefront Manipulation of Photonic Integrated Circuits* and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- parts of this work have been published as: Dinsdale *et al.*. A full list can be found in the "List of Publications and Presentations" section on the previous page.

Signed:.....

Date:.....

Acknowledgements

I have been privileged to receive a great deal of support and work with some of the most fantastic people in the world over my four years long PhD journey. It has certainly not been easy at times, but things truly worth doing seldom are.

Firstly, I would like to thank my supervisors Otto Muskens and Graham Reed, for giving me this opportunity. Your guidance and support have been invaluable, and I am indebted to you both for the skills and knowledge I have acquired throughout the years.

There is a long list of people who have directly contributed to this work or my personal development during my PhD. In particular, I would like to thank Roman Bruck, Peter Wiecha, Kevin Vynck, Bigeng Chen, Idris Ajia and Ben Mills, without whom this thesis simply would not have been possible.

Next, I would like to thank all the Integrated Nanophotonics Group members, past and present. Various discussions over the years have been indispensable, and group socials have been a highlight of my time at Southampton. I am particularly indebted to Matthew Delaney, with whom I bounced endless ideas off of and always advocated the best course of action — even if it was not the most appealing option at the time.

To all the inhabitants of room 3073 over the years, I wish to express my deepest gratitude for keeping me sane. When times were hard or equipment in the lab failed, I could always go to the office for some respite. A special mention to Oliver Trojak, a fountain of knowledge and mastermind behind the office's ticker-tape printer that checked all incoming e-mails for free food alerts. We never went hungry on your watch.

Special thanks to BBAS for services rendered.

I will be eternally grateful for the unconditional love and encouragement I have received from my family throughout my life. I am reassured that no matter where life takes me, I will always have a home to return to and people who care for me.

Finally, to my best friend and partner Jana, your support, patience and love have been endless, for which I owe you everything under the sun.

Nicholas J. Dinsdale
January 2021
Southampton, UK

> *"Do you find coming to terms with the mindless tedium of it all presents an interesting challenge?"*
>
> —Douglas Adams, *The Hitchhiker's Guide to the Galaxy*

Glossary

a-FMM	aperiodic-Fourier Modal Method
AFM	Atomic Force Microscopy
ANN	Artificial Neural Network
APD	Avalanche Photodetector
AWG	Arrayed Waveguide Grating
BB	Building Block
CCD	Charge-Coupled Device
CMOS	Complementary Metal-Oxide-Semiconductor
CPU	Central Processing Unit
DC	Directional Coupler
DMD	Digital Micromirror Device
EIM	Effective Index Method
EM	Electromagnetic
ER	Extinction Ratio
FDTD	Finite-Difference Time-Domain
FEM	Finite Element Method
FMM	Fourier Modal Method
FPGA	Field-Programmable Gate Array
FTTN	Fibre-To-The-Neighbourhood
FTTP	Fibre-To-The-Premises
FSR	Free-Spectral-Range
FWHM	Full-Width Half-Maximum
GPU	Graphics Processing Unit
HPC	High-Performance Computing
HWP	Half-Wave Plate
IR	Infrared
LP	Longpass
LO	Local Oscillator
MEMS	Micro-Electro-Mechanical System
MFD	Mode Field Diameter
MMI	Multimode Interference

MSE	Mean Squared Error
MZI	Mach-Zehnder Interferometer
MZM	Mach-Zehnder Modulator
NA	Numerical Aperture
ND	Neutral Density
NN	Neural Network
NSOM	Near-field Scanning Optical Microscopy
PCM	Phase Change Material
PLC	Planar Lightwave Circuit
OADM	Optical Add-Drop Multiplexer
OFDR	Optical Frequency Domain Reflectometry
OPO	Optical Parametric Oscillator
OSA	Optical Spectrum Analyser
OTDR	Optical Time Domain Reflectometry
PCM	Phase Change Material
PIC	Photonic Integrated Circuit
RCWA	Rigorous Coupled-Wave Analysis
ReLU	Rectified Linear Unit
RR	Ring Resonator
SEM	Scanning Electron Microscope
SHG	Second-Harmonic Generation
SiP	Silicon Photonics
SMF	Single-Mode Fibre
SNR	Signal-to-Noise Ratio
SOI	Silicon-On-Insulator
SP	Shortpass
SWG	Sub-Wavelength Grating
TE	Transverse Electrical
TIR	Total Internal Reflection
TM	Transverse Magnetic
TMI	Two-Mode Interference
UPMS	Ultrafast Photomodulation Spectroscopy
UV	Ultraviolet
VAE	Variational Autoencoder
varFDTD	variational Finite-Difference Time-Domain

Chapter 1

Introduction

Light plays an essential role in how we as humans communicate, and has done so throughout much of our history, from smoke signals and the lighting of warning beacons to the intricate waving of flags and encoded flashes of lamps. However, all of these methods suffer from the same one significant drawback: they require a direct line of sight, which can easily be blocked by terrain or adverse weather conditions. Even effects such as the curvature of the Earth ultimately limit their ability to transfer information over great distances. For this reason, light must be guided within some medium to be able to form truly global communication networks.

The first description of guiding light via Total Internal Reflection (TIR) was made by the Swiss physics professor Jean-Daniel Colladon in 1842 while attempting to demonstrate turbulence in fluids by releasing multiple jets of water from a tank [1]. In an effort to make the turbulent regions more visible to his class, he decided to focus sunlight from a nearby window into the tank. To his surprise—and great delight—certain incident angles resulted in entire water streams illuminating. This phenomenon is now a common classroom physics experiment known as a 'light fountain', which is typically used with a laser to demonstrate the principle of TIR.

Nowadays, almost all long-haul communications are realised using the same basic concept as Colladon's light fountain, optical signals transported via TIR, with glass fibres acting as the arteries of this information network. We are now seeing what would have traditionally been electrical connections being replaced by optical links, which is occurring at ever shorter length scales. It is now commonplace in the United Kingdom to have Fibre-To-The-Neighbourhood (FTTN), or even Fibre-To-The-Premises (FTTP), and the majority of routing in large-scale data centres is now done optically. Photonic Integrated Circuits (PICs) are required to access even shorter length scales of optical links. Several decades of development in electronics integration has made it possible to shrink computers from the size of a large room to handheld portable devices. A similar trend is now being observed in photonic integration, with bulk optics being integrated

and combined at the chip level to create optical circuits in which photons are analogous to the electrons in electronic circuits. Fibre optics represent an intermediate level of integration and are significantly smaller than their bulk optics equivalent. However, a single-mode telecommunication fibre will have the core diameter of ∼10 µm while a typical single-mode silicon integrated waveguide can be as small as 220 × 400 nm, corresponding to almost three orders of magnitude reduction in cross-sectional area. There are a number of factors driving photonic integration, including cost-saving, size reduction, power reduction, high-volume fabrication and alignment-less production [2].

This integration approach is creating an increased demand for versatile all-optical elements to keep signals in the optical domain, avoiding transformation losses and circumnavigating the electronic speed limit. Ultimately, continued developments in photonic integrated are predicted to lead to new shifts in paradigm with applications in optical Neural Networks (NNs) [3, 4], optical quantum computing [5] or entire lab-on-a-chip [6].

1.1 Silicon photonics

Silicon continues to dominate the electronics industry as the choice of material for fabricating integrated circuits since it was initially conceived in the mid-20th century [7]. Silicon is now also widely considered one of the most feasible platforms for producing PICs for a variety of reasons. Primarily, its potential for cost-effective manufacturing at high-volume production, which draws upon over half a century of fabrication development stemming directly from the electronics industry [8, 9]. Large high-quality mono-crystal Silicon-On-Insulator (SOI) wafers and lithographic systems are commercially available, with 7 nm chips already present in consumer electronics products, such as Central Processing Units (CPUs) and Graphics Processing Units (GPUs), and the intention is to further reduce node size to 3 nm by 2022 [10]. Typically, most photonic chips do not require such small feature sizes, with the smallest features on most common PICs being on the order of ∼100 nm. The complex refractive index of silicon itself provides a few advantageous properties. A high real part of the refractive index, $n \approx 3.5$, at telecommunication wavelengths means that light confinement is very high, which consequentially allows devices to be produced with a very small footprint and tight bend radii without causing excess loss. In addition, the low imaginary part of the refractive index at telecommunication wavelengths results in a negligible material absorption (absorption coefficient of $\sim 10^{-4}$ cm^{-1} [11]) over the path lengths of optical chips. Another advantage of being based on the same material and fabrication techniques as electronics is direct compatibility with Complementary Metal-Oxide-Semiconductor (CMOS) elements. Thereby allowing photonic and electronic components to be combined on to a single die, significantly reducing the distance electrical signals need to travel, and consequently, the overall power consumption of the device [12, 13].

However, silicon also has some significant drawbacks as a material for producing PICs. Its high real part of the refractive index, while resulting in high confinement, also makes coupling light into and out of devices very difficult. Smaller mode sizes mean poor Mode Field Diameter (MFD) matching with fibre optics without the use of spot-size converting devices. Moreover, while the material loss may be very low in silicon, it is sidewall scattering that dominates the propagation loss with a typical roughness of 1 nm resulting in a loss of around $1\,\mathrm{dB\,cm^{-1}}$ [14]; however, overall short device lengths help alleviate this problem somewhat. Another serious issue with silicon is the lack of efficient nonlinear effects, limiting practical modulation methods, and an indirect bandgap means there is no inherent light emission or detection in silicon. Instead, more complicated mechanisms, such as strain engineering, germanium implantation or hybrid integration, must be employed. This hybrid approach can either be through the integration of other material platforms, such as III-V materials, or the direct bonding of external light sources. Finally, silicon's bandgap of 1.11 eV (at 300 K) results in two-photon absorption dominating at high power, limiting permissible optical power to around 25 mW for single-mode thin-SOI waveguides. Silicon as an integrated photonics platform is currently experiencing rapid industry adoption, with predicted market growth to around $4 billion by 2025 (Source: Yole Développement [15]). The predominant driving forces behind this growth are High-Performance Computing (HPC), so-called 'big data' centres and sensing applications. For HPC and big data, the optical domain offers high-speed transmission with minimum attenuation of signals which greatly exceed the capabilities of traditional copper cable connections, and the potential for low power consumption per bit (order of fJ/bit) [16–19]. It is these pluggable optical transceivers which make up the vast majority of the commercial silicon photonics market [15].

While one of the major applications, telecommunications is by no means the only field in which integrated silicon photonics is having an impact. There is also much development in chemical and biological sensing, medical diagnostics and defence applications [20, 21]. The majority of the sensing applications rely on near-field interactions to produce novel, high-sensitivity sensors [21–24]. However, some significant challenges remain to be overcome for silicon as a platform, resulting in many active areas of research. A large proportion of the costs involved in silicon photonics are associated with device testing and packaging, and recent advances in integrated photonics are resulting in increasingly complex circuit designs exacerbating the problem. Wafer-scale testing is a critical component in electronics manufacturing to establish what is a known-good-die. Since packaging represents a large fraction of costs involved, it is often beneficial to expose the bare, unpackaged dies to testing, tuning and burn-in. Integrated photonics is yet to develop as versatile techniques as seen in the electronics industry, typically with only input-output characteristics accessible. Techniques capable of accessing individual elements in a complex PIC, such as erasable coupling ports [25, 26], are important to identify component performance and provide some level of diagnostics capability for the overall circuit; thereby, bypassing the need to package a device prior to testing.

1.2 Thesis outline

The world is currently undergoing an unprecedented era of technological growth, with exponential increases in the demand for data. To meet such a demand, the development of new techniques to achieve more complex functionalities on a smaller footprint will be required. Ideally, integrated photonics will find its equivalent of the Field-Programmable Gate Array (FPGA) in electronics, with which arbitrary functionality can be achieved with one standardised circuit layout. A high degree of reconfigurability is of vital importance to produce the universal operation elements required to achieve the photonic FPGA. The primary goal of the work presented in this thesis is to develop various perturbative approaches in the field of integrated optics for device characterisation, optimisation and arbitrary wavefront manipulation. The organisation of this thesis is outlined below.

Chapter 2 introduces the underlying physics for this thesis, including the fundamentals of light, waveguide theory, common passive PIC components and modulation methods in silicon.

Chapter 3 discusses the methodology, covering the different fabrication, computational and experiments techniques used within this thesis. In particular, this chapter introduces the concept of Ultrafast Photomodulation Spectroscopy (UPMS) as a technique and provides an overview of the system's experimental setup and key performance parameters.

Chapter 4 details the use of UPMS as a characterisation technique and diagnostic tool for a variety of different Multimode Interference (MMI) power splitters. The method allows for the spatial mapping of light within devices, and in combination with an analytical model can be used as a quantitative tool to analyse the quality of fabrication non-destructively at the wafer-scale.

In Chapter 5, the optical real-time trimming of germanium ion-implanted devices is investigated. The ion implantation results in a localised disruption of the silicon crystal lattice, forming amorphous silicon, which has a higher refractive index. Localised pulsed laser annealing then reforms the crystalline silicon and in doing so reverts the refractive index back towards the original value, thereby allowing the active tuning of devices.

Chapter 6 is an investigation of the advanced functionalities that can be achieved through the use of multiple perturbations in the refractive index profile of silicon waveguides. This work amalgamates in the training of a deep-Artificial Neural Network (ANN) that is capable of rapidly and accurately generating perturbation patterns for producing an arbitrary target complex transmission matrix.

Finally, in Chapter 7, the findings and outlook of this thesis will be discussed.

Chapter 2

Background

This chapter outlines the key underlying physics used in this thesis, including; polarisation and propagation modes of light in a waveguide, fundamental slab waveguide theory, the effective index method for approximating the propagation constants in three-dimensional waveguides and multimode interference self-imaging theory. Different common waveguide geometries and components will be discussed. Finally, various modulation techniques for silicon waveguides are outlined and assessed.

2.1 Electromagnetic radiation, polarisation and modes

Light is an Electromagnetic (EM) wave, and as such can be broken down into its component electric and magnetic fields that oscillate sinusoidally with time, shown in Figure 2.1. The behaviour of light in a vacuum is therefore governed by Maxwell's equations: [27]

$$\boldsymbol{\nabla} \cdot \mathbf{E} = \frac{\rho}{\epsilon_0}, \qquad \text{(Gauss's Law)} \tag{2.1}$$

$$\boldsymbol{\nabla} \cdot \mathbf{B} = 0, \qquad \text{(No Magnetic Monopoles)} \tag{2.2}$$

$$\boldsymbol{\nabla} \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \qquad \text{(Faraday's Law)} \tag{2.3}$$

$$\boldsymbol{\nabla} \times \mathbf{B} = \mu_0 \mathbf{J} + \mu_0 \epsilon_0 \frac{\partial \mathbf{E}}{\partial t}, \qquad \text{(Ampère–Maxwell Law)} \tag{2.4}$$

where $\mathbf{E}$ is the electric field, $\mathbf{B}$ is the magnetic field, ρ is the charge density, $\mathbf{J}$ is the current density, t is time and ϵ_0 and μ_0 are the permittivity and permeability of free space, respectively. With these equations, it is possible to prove that the electric and magnetic fields are always perpendicular to one another in a vacuum (see Appendix A.1). However, this is not the case for light propagating in a waveguide, in which the electromagnetic waves can interact directly with the material itself.

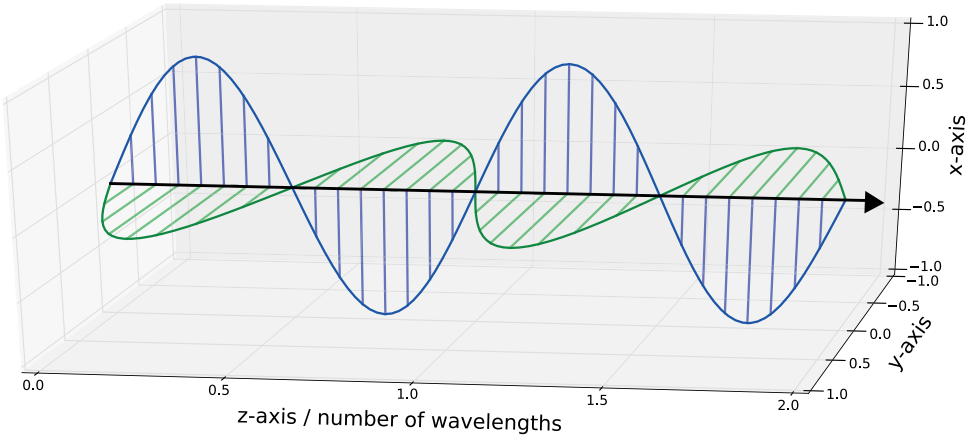


Figure 2.1: Illustration of a linearly polarised EM wave, where blue and green correspond to the electric and magnetic field, respectively, and the light is propagating along the z-axis. By convention, it is the electric field used to describe the polarisation, as such this wave is said to be polarised in the x-axis.

Supported solutions of Maxwell's equations in matter that are constrained by some boundary conditions are referred to as "modes". A slab waveguide has two families of modes; the Transverse Electrical (TE) and Transverse Magnetic (TM) [27]. For the TE mode, the electric field is always perpendicular to the direction of propagation. While in the case of the TM mode, it is the magnetic field vector that is is always perpendicular to the propagation direction. However, this is only the case for infinite slab waveguides that only constrain the electromagnetic fields in one dimension. Waveguide geometry that restricts the electromagnetic fields in two dimensions cannot support pure TE and TM modes. The modes are still orthogonal to each other, but every mode will contain all six field components. For many waveguide geometries, such as ridge or rib, the modes still carry the vast majority of their optical power in the TE or the TM field profiles. Therefore, they are instead referred to as quasi-TE and -TM modes, but it is often convention to drop the "quasi". Figure 2.2 shows the simulated electric fields for both quasi-modes in a 500 nm by 500 nm silicon ridge waveguide. Both of these modes will have additional higher order versions providing the guiding medium is large enough to support multiple integer numbers of electric or magnetic field nodes. This thesis will focus primarily on the fundamental quasi-TE mode, henceforth referred to simply as the TE mode unless directly stated otherwise. Likewise, the fundamental quasi-TM mode will be referred to simply as the TM mode.

In the presence of bulk matter, it is useful to define some of Maxwell's equations in terms of the electric displacement, $\mathbf{D}$, and the magnetic intensity, $\mathbf{H}$:

$$\mathbf{D} = \epsilon \mathbf{E}, \text{ and} \tag{2.5}$$

$$\mathbf{H} = \frac{\mathbf{B}}{\mu}. \tag{2.6}$$

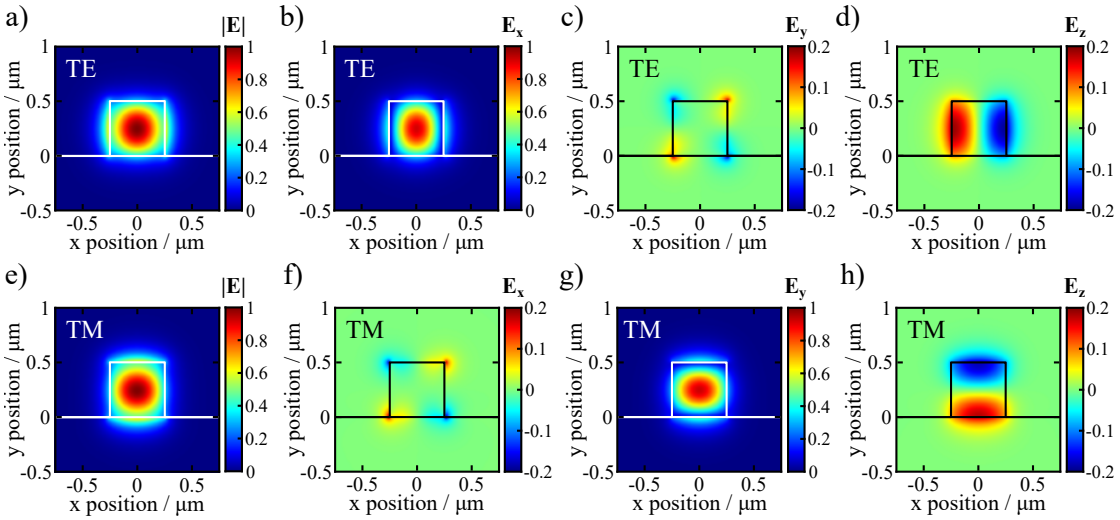


Figure 2.2: The electric field a) magnitude and b-d) axial components for the fundamental quasi-TE mode in a 500 nm by 500 nm ridge waveguide. Similarly, the electric field a) magnitude and b-d) axial components for the fundamental quasi-TM mode. Modes are modelled using Lumerical Finite-Difference Time-Domain (FDTD) solver in 3D at a wavelength of 1550 nm. Refractive indices for the core, substrate and cladding were 3.5, 1.45 and 1.0, respectively.

In which, it is the dielectric permittivity, $\epsilon = \epsilon_0 \cdot \epsilon_r$, and magnetic permeability, $\mu = \mu_0 \cdot \mu_r$, that need to be considered, where ϵ_r and μ_r are the relative permittivity and permeability, respectively [27]. This results in the following adjusted Maxwell's equations in matter:

$$\boldsymbol{\nabla} \cdot \mathbf{D} = \rho_f, \qquad \text{(Gauss's Law)} \tag{2.7}$$

$$\boldsymbol{\nabla} \times \mathbf{H} = \mathbf{J}_f + \frac{\partial \mathbf{D}}{\partial t}, \qquad \text{(Ampère–Maxwell Law)} \tag{2.8}$$

where ρ_f is the free charge density and $\mathbf{J}_f$ is the free current density. By taking the curl of Faraday's law (Equation 2.3) and using the vector identity $\boldsymbol{\nabla} \times \boldsymbol{\nabla} \times \mathbf{A} = \boldsymbol{\nabla}(\boldsymbol{\nabla} \cdot \mathbf{A}) - \boldsymbol{\nabla}^2 \mathbf{A}$ in conjunction with Equation 2.1 and Equation 2.2, we obtain

$$\boldsymbol{\nabla}^2 \mathbf{E} = \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) \mathbf{E} = \epsilon\mu \frac{\partial^2 \mathbf{E}}{\partial t^2}. \tag{2.9}$$

This equation is known as the wave equation, and it gives rise to some very important fundamental definitions [28]. Firstly, the phase velocity of an electromagnetic wave travelling in matter is extracted from the wave equation as

$$v = \sqrt{\frac{1}{\epsilon\mu}}, \tag{2.10}$$

and in free space as

$$c \equiv v_0 = \sqrt{\frac{1}{\epsilon_0 \mu_0}}, \tag{2.11}$$

where c is the speed of light in a vacuum. The bulk refractive index is then defined as the ratio of these two phase velocities

$$n \equiv \frac{c}{v} \geq 1, \tag{2.12}$$

$$= \sqrt{\frac{\epsilon\mu}{\epsilon_0 \mu_0}} = \sqrt{\epsilon_r \mu_r}. \tag{2.13}$$

In most materials, $\mu \approx \mu_0$, which simplifies the index of refraction to just be in terms of the dielectric constant,

$$n \approx \sqrt{\epsilon_r}. \tag{2.14}$$

Solving the macroscopic wave equation for planar monochromatic waves, i.e. $\mathbf{E} = \mathbf{E_0} e^{i(\mathbf{k}\cdot\mathbf{r}-\omega t)}$, yields the dispersion relation given by

$$\omega^2 = v^2 \mathbf{k}^2, \tag{2.15}$$

where $\omega = 2\pi\nu$ and is the temporal angular frequency of the wave and $\mathbf{k} = \frac{2\pi}{\lambda}\hat{\mathbf{k}}$ is the wavevector. Finally, this results in the useful definitions:

$$\mathbf{k} = n\mathbf{k}_0, \tag{2.16}$$

$$\lambda = \frac{2\pi c}{\omega n} = \frac{2\pi}{|\mathbf{k}|}, \tag{2.17}$$

$$\lambda = \frac{\lambda_0}{n}, \tag{2.18}$$

where $\mathbf{k}_0$ and λ_0 are the free-space wavevector and wavelength, respectively. It is possible to confine the propagation of a plane wave in one or more directions by choosing an appropriate geometric structure, thereby enabling the guiding of electromagnetic waves along a predefined path. Such structures are known as waveguides and form the basis of all integrated optics. Their basic operation through confinement by Total Internal Reflection (TIR) is outlined in the following section, along with some of the common wave guiding geometries used.

2.2 Slab waveguide theory

Here, we start with the simple case of a slab waveguide with the goal to form a set of equations that can be solved for the supported TE and TM mode propagation constants discussed in the previous section. A slab waveguide consists of three layers of materials that extend infinitely in the directions parallel to their interfaces. The wavevector, $\mathbf{k} = n_1 k_0 \hat{\mathbf{k}}$, of light propagating through the core of a slab waveguide of refractive index

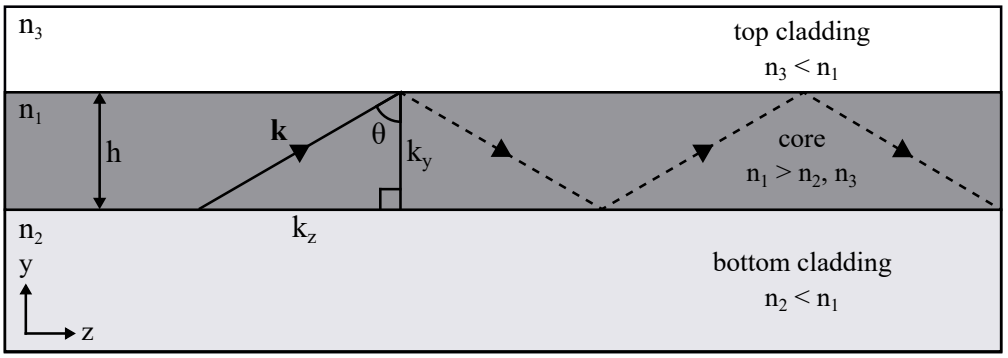


Figure 2.3: Propagation of light in a two-dimensional slab waveguide, with a core of refractive index n_1 and thickness h.

n_1 and thickness h is shown in Figure 2.3. This wavevector can be broken down into its axial components, known as propagation constants, using simple trigonometry [29]. These propagation constants are therefore given by

$$k_z = n_1 k_0 \sin(\theta), \tag{2.19}$$

$$k_y = n_1 k_0 \cos(\theta), \tag{2.20}$$

where θ is the propagation angle in the core and $k_0 = 2\pi/\lambda_0$ is the wavevector in free-space. The phase shift due to round trip propagation in the y-direction, ϕ_y, is given by

$$\phi_y = 2hk_y = 2hn_1 k_0 \cos(\theta). \tag{2.21}$$

Taking into account the phase changes at the boundaries, ϕ_u and ϕ_l for upper and lower boundaries respectively, the total phase change, ϕ_T, can then be written as

$$\phi_T = \phi_y - \phi_u - \phi_l = 2\nu\pi, \tag{2.22}$$

where ν is an integer representing the mode number, with zero being the fundamental mode. By considering these boundary phase shifts, it is possible to calculate the propagation angle, and thus the propagation constants, for the discrete modes of a simple two-dimensional slab waveguides [29, 30]. For symmetric slab waveguide, $n_2 = n_3$, in the TE mode

$$\tan\left(\frac{n_1 h k_0 \cos(\theta) - \nu\pi}{2}\right) = \frac{\sqrt{\sin^2(\theta) - (n_2/n_1)^2}}{\cos(\theta)}, \tag{2.23}$$

and for symmetric slab waveguide in the TM mode

$$\tan\left(\frac{n_1 h k_0 \cos(\theta) - \nu\pi}{2}\right) = \frac{\sqrt{(n_1/n_2)^2 \sin^2(\theta) - 1}}{(n_2/n_1)\cos(\theta)}. \tag{2.24}$$

The maximum number of supported TE modes is therefore given by the floor of Equation 2.23 when the propagation angle is equal to the critical angle. The equation then

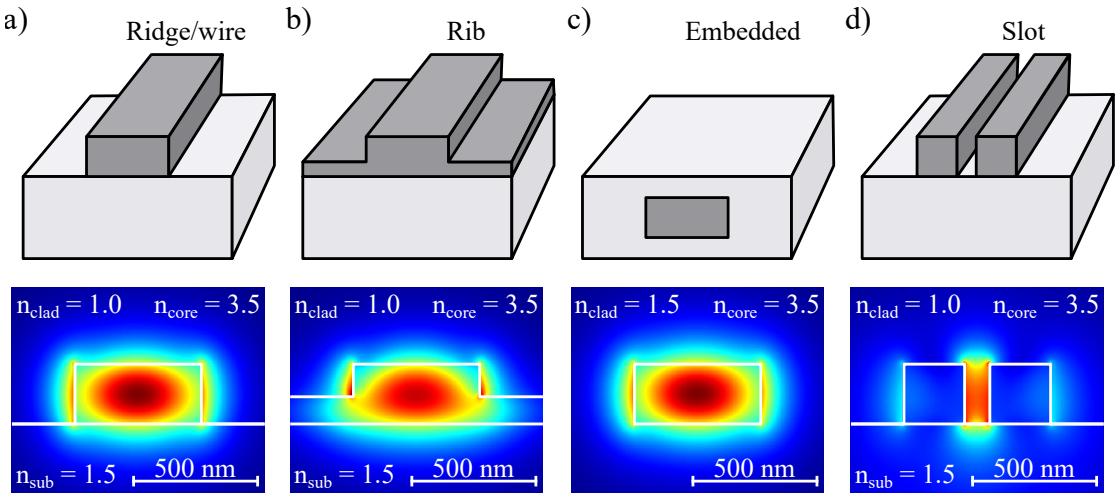


Figure 2.4: Illustration (top) and electric fields (bottom) for a) ridge (also commonly referred to as wire or strip), b) rib, c) embedded and d) slot waveguides. Modelled using Lumerical FDTD solver for the fundamental input mode at a wavelength of 1550 nm.

reduces to

$$\nu_{max} = \frac{n_1 h k_0 \cos(\theta_c)}{\pi}, \tag{2.25}$$

where θ_c is the critical angle and is given by the TIR condition, $\sin(\theta_c) = n_2/n_1$.

For an asymmetric slab waveguide, $n_2 \neq n_3$, the phase changes from the top and bottom boundary will not be equal, so the propagation constant's equations are slightly more complex. For a TE mode, it is described by

$$n_1 h k_0 \cos(\theta) - \nu\pi = \tan^{-1}\left(\frac{\sqrt{\sin^2(\theta) - (n_2/n_1)^2}}{\cos(\theta)}\right) + \tan^{-1}\left(\frac{\sqrt{\sin^2(\theta) - (n_3/n_1)^2}}{\cos(\theta)}\right), \tag{2.26}$$

and a TM mode is adjusted as in Equation 2.24, becoming

$$n_1 h k_0 \cos(\theta) - \nu\pi = \tan^{-1}\left(\frac{\sqrt{(n_1/n_2)^2 \sin^2(\theta) - 1}}{(n_2/n_1)\cos(\theta)}\right) + \tan^{-1}\left(\frac{\sqrt{(n_1/n_3)^2 \sin^2(\theta) - 1}}{(n_3/n_1)\cos(\theta)}\right). \tag{2.27}$$

These solutions for a slab waveguide structure will later be used in conjunction with the Effective Index Method (EIM) to provide approximate solutions for the propagation constants of real-world waveguide structures. Some of the most common geometries are introduced in the next subsection and impose additional constraints on the electric fields compared to the slab scenario.

2.2.1 Waveguide geometries

In the previous section, we only considered a slab waveguide to simplify the mathematical derivation to confinement in only a single dimension. The vast majority of real-world

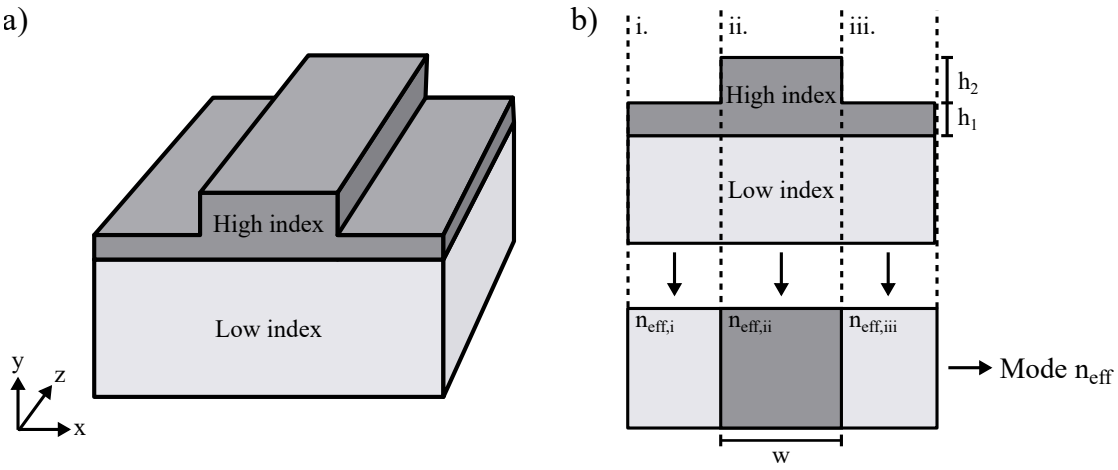


Figure 2.5: An example of reducing a three-dimensional rib waveguide to an effective index slab waveguides using the EIM. a) The rib waveguide with a propagation axis in the z-direction. b) A two-dimensional slice of the rib waveguide with a rib width of w, a rib height of h_2 and a slab height of h_1. Each region (separated by the dashed line and denoted by a Roman numeral) can be seen as its own slab waveguide with its own effective index, which reduces the entire rib structure to another apparent slab waveguide.

applications require confinement in two axes to guide the light along the third, thereby allowing the waveguide to act as the photonic equivalent to an electronic wire. There are many types of waveguide geometries; the most common of which are shown in Figure 2.4 with their electric field profiles. Out of these geometries, it is predominantly ridge or rib that are used to guide light around Photonic Integrated Circuits (PICs). The only difference between the two is the etch depth, with rib leaving a thin high-index slab layer and ridge going all the way down to the substrate. This results in ridge waveguides having higher mode confinement, and consequently, smaller minimum bend radii, but also results in slightly higher propagation losses compared to rib [31, 32]. Slot waveguides have very high confinement of the electric field in the gap between two high index strips, making them ideal for near-field sensing applications [33–35].

2.2.2 The effective index method

The EIM is a useful technique to find approximate solutions for the propagation constants of a simple three-dimensional waveguide [36]. These waveguides can have their geometry broken down into effective two-dimensional slab waveguides, shown in Figure 2.5 for a rib waveguide, from which the effective index of the ν^{th} mode, $n_{\text{eff},\nu}$, can be obtained. Where $n_{\text{eff},\nu}$ is given by

$$n_{\text{eff},\nu} = k_{z,\nu}/k_0, \tag{2.28}$$

which is sometimes written in terms of the phase constant of the waveguide, β_ν, where

$$\beta_\nu = k_{z,\nu} = n_{\text{eff},\nu}\frac{2\pi}{\lambda_0}. \tag{2.29}$$

It is the convention to take the effective index of the fundamental mode in the absence of any ν value. In this method, the propagation constant, and therefore the effective index, is found for one of the apparent slab waveguides using the equations in Section 2.2. Then the second apparent slab waveguide can be solved by using the effective index of the first as the core's new refractive index. This provides an effective index for a particular mode in the entire structure [37]. There are also powerful online mode solvers and commercial packages, such as Lumerical Mode Solutions, capable of rapidly solving the eigenvalue equations for the propagation constants. Most modern solvers rely on slightly adjusted versions of the EIM, such as the so-called 'variational variant' [38]. It is also possible to calculate the chromatic dispersion of the waveguide from frequency-dependent β values.

2.2.3 Waveguide coupling

Coupling light into and out of waveguides remains a very challenging task, especially in the case of very high index and thin guiding layer platforms such as 220 nm Silicon-On-Insulator (SOI) [39]. Other lower index platforms, such as silicon nitride (SiN) and indium phosphide (InP), result in better Mode Field Diameter (MFD) matching but waveguides are still typically significantly smaller than the core of a Single-Mode Fibre (SMF). Their lower refractive indices than silicon also result in smaller Fresnel reflection coefficient, further reducing the coupling loss. Planar Lightwave Circuits (PLCs) are SiO_2 or low-index polymer based waveguides, and as such offer near-perfect refractive index and MFD matching with silica fibres. For this reason, in combination with exceptionally low propagation losses, PLCs are still frequently used for passive power splitters and Arrayed Waveguide Gratings (AWGs) to this day, despite being the oldest integrated photonics material platform.

The techniques for coupling into a integrated waveguide can be broadly divided into three main categories; vertical coupling, lateral coupling and evanescent coupling [40]. Some examples of the most common approaches are shown in Figure 2.6 and outlined in the following sections.

2.2.3.1 Grating couplers

Grating couplers are one of the most common coupling techniques in integrated photonics, primarily due to their efficient coupling of light that can be incident from near perpendicular angles. Their operation stems from a periodic structure, typically one-dimensional, to alter the propagation direction of light from nearly normal to the chip

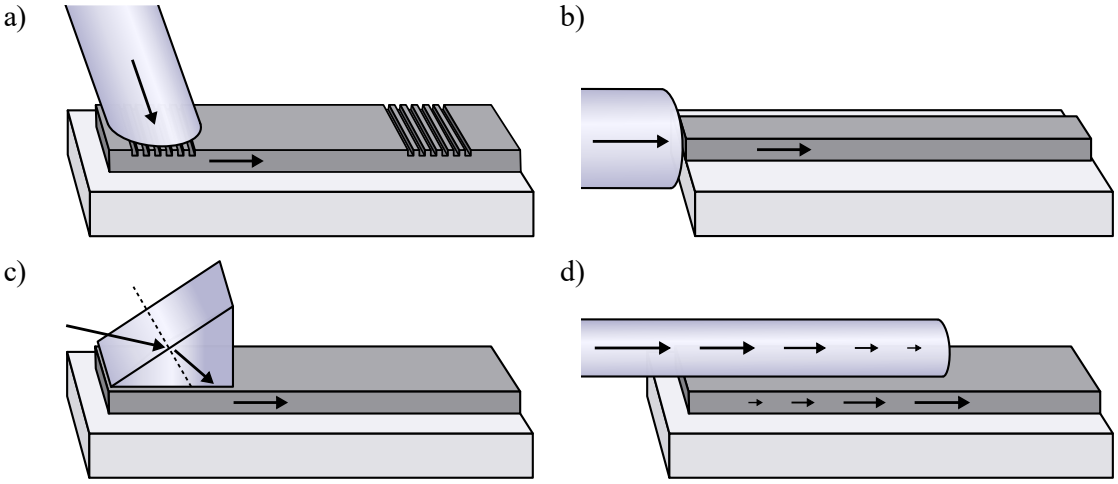


Figure 2.6: Illustrations of a) grating, b) end face, c) prism and d) evanescent coupling. The black arrows indicate the direction of light propagation.

to collinear [14, 41]. They are advantageous compared to lateral coupling as they do not require access to the edge facet of a chip, thereby allowing coupling to devices still at the wafer-scale pre-dicing. Additionally, many separate circuits can be combined on a single chip, and it is relatively fast to couple light into and out of devices under test. A grating coupler's function stems from diffraction from the surface, resulting in a number of diffraction orders that depend on the grating period and effective indices of the grating. From Huygens' wave principle, each period of the grating can be considered a new point-source [42]. These wavelets will spread out in all directions from each point source, but will only constructively interfere for particular directions spaced by the so-called 'grating vector'. This grating vector, K, is defined as

$$K = \frac{2\pi}{\Lambda}, \tag{2.30}$$

where Λ is the period of the grating. Light incident on the grating with wavevector $\mathbf{k}_{\text{inc}}$ at an angle ϕ from the normal to the structure's periodicity will be both reflected and transmitted. To result in efficient coupling between the incident light and a waveguide mode, or vice versa, the difference between the z-component of the wavevector of the beam and waveguide mode must be a multiple of the grating vector. This leads to the coupling condition for grating couplers, given by

$$\mathbf{k}_{\text{inc}} \sin\phi + mK = k_0 n_{\text{eff}}, \tag{2.31}$$

where m is an integer corresponding to the order of diffraction [43]. A small non-zero angle of ϕ is typically used to avoid large unwanted back reflections for incident light perpendicular to a surface. There has been much research into improving the coupling efficiency and bandwidth grating couplers [44–46]. Additionally, more advanced 2D gratings have been demonstrated to have the capability to split TE and TM incident polarisations into separate waveguides [47, 48].

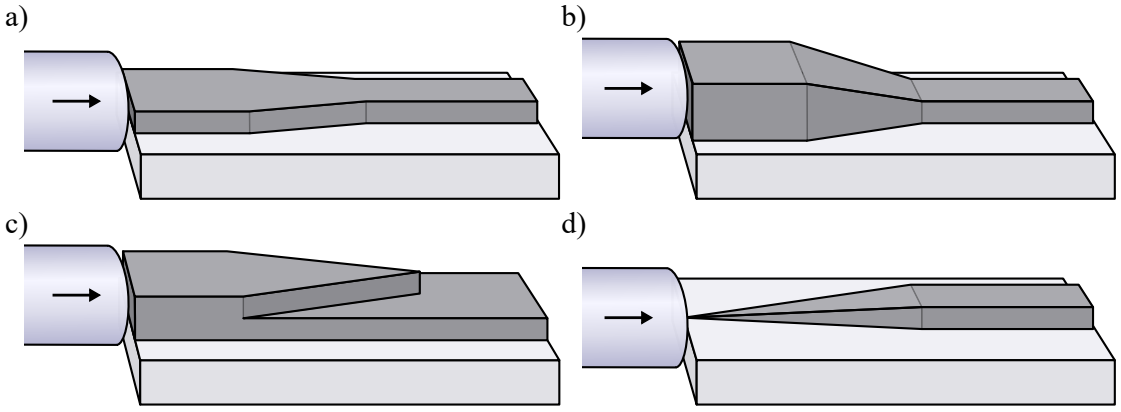


Figure 2.7: Illustrations of a) one-dimensional, b) two-dimensional, c) NVT [49] and d) inverse taper lateral coupling. The black arrows indicate the direction of light propagation.

2.2.3.2 End face coupling

End face coupling, also known as edge coupling, is another standard technique for optical input/output of photonic circuits. It is also frequency used for the hybrid integration of different integrated components, such as lasers, modulators and detectors. Compared to grating coupling, end face coupling has a number of advantages, such as low insertion loss, very broadband and coupling both TE and TM polarisations. However, there are also significant challenges, including MFD mismatch, alignment precision, facet polishing and the need for anti-reflection coatings [14]. In particular, the MFD mismatch can lead to significant losses in silicon due to the orders of magnitude size disparity between an SMF fibre and typical waveguide dimensions. Several spot-size converter geometries have been suggested to help alleviate this coupling loss, some of which are illustrated in Figure 2.7.

2.2.4 Fundamental integrated photonics building blocks

There are many different integrated elements that can be combined to build a photonic circuit. These basic components are sometimes referred to as 'building blocks', and there are integrated equivalents of most bulk optics. However, sometimes physical limitations of the material itself render some integrated components impossible, e.g. optical isolators and circulators in silicon. A complete circuit's performance can be modelled by propagating incident light through the scattering matrix (or S-matrix) for each component building block [14], where the S-matrix describes the complex coupling parameters between all ports, modes and wavelengths.

Some of the most common building blocks that will be frequently referred to within this work are introduced in the following subsections.

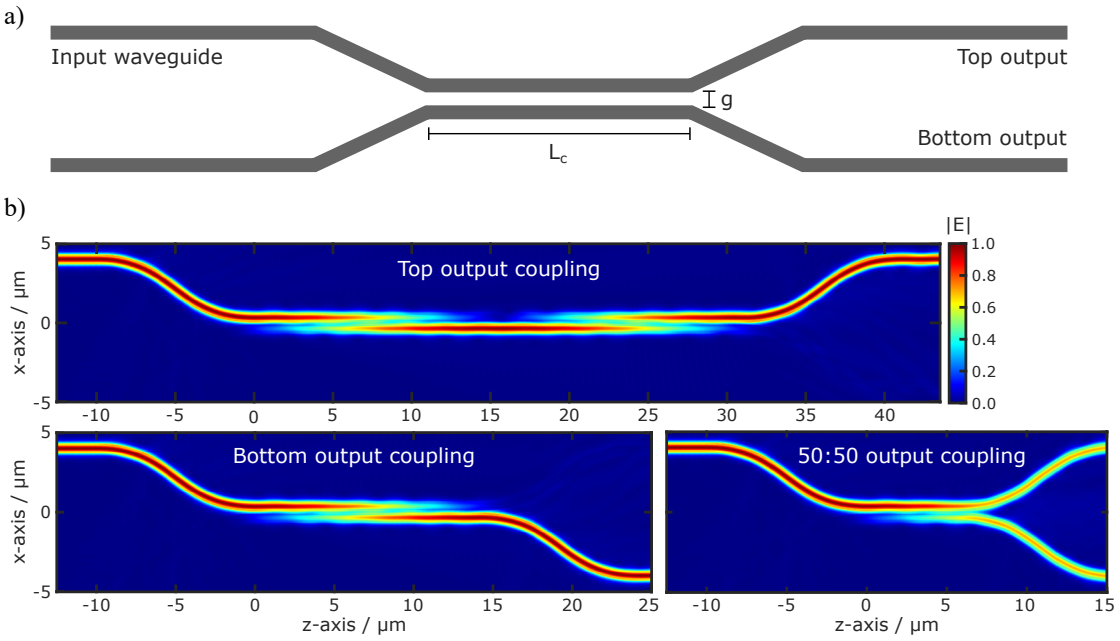


Figure 2.8: a) Schematic for an integrated Directional Coupler (DC). b) Simulated electric field distribution for three values of L_c that correspond to top, bottom and 50:50 output coupling. Simulated using Lumerical FDTD solver for the fundamental TE input mode at a wavelength of 1550 nm.

2.2.4.1 Directional coupler

The evanescent coupling between adjacent waveguides was first demonstrated by Somekh et al. in the 1970s [50]. The structure is known as a Directional Coupler (DC), in which light gradually transfers back and forth between two super-modes corresponding to the top and bottom waveguides. Based on this principle, the DC is frequently used as a simple optical power splitter, where the splitting ratio depends on two key parameters; g, the separation gap and L_c, the coupling length, which are shown in the schematic in Figure 2.8 a).

Light alternates between the two modes of the top and bottom rail waveguides, as shown in Figure 2.8 b), at regular intervals depending on g. The top electric field distribution shows the case where all the light couples to the top output having transferred to the lower rail and back again. L_c can be set to the length where all the light is in the bottom rail, the lower-left case in Figure 2.8 b). The distance after which all the light couples to the bottom waveguide is called the cross-over length, L_{cross}, and is defined as

$$\beta_1 L_{\text{cross}} - \beta_2 L_{cross} = \pi, \tag{2.32}$$

$$L_{\text{cross}} \left[\frac{2\pi n_{\text{eff},1}}{\lambda} - \frac{2\pi n_{\text{eff},2}}{\lambda} \right] = \pi, \tag{2.33}$$

$$L_{\text{cross}} = \frac{\lambda}{2\Delta n_{\text{eff}}}, \tag{2.34}$$

where β_1 and β_2 are the propagation constant of the two super-modes corresponding to the top and bottom waveguides, respectively, where the effective index difference between the two is $\Delta n_{\text{eff}} = n_{\text{eff},1} - n_{\text{eff},2}$ [14]. Most devices require a significant interaction length due to the two propagation constants typically being very similar. The bottom right example in Figure 2.8 b) shows the 50:50 splitting case. However, any desired splitting ratio can be achieved by altering L_c to the appropriate value. The field coupling coefficient, κ, at an arbitrary position, z, along the DC is then given by [14]

$$\kappa = \left[\frac{P_{\text{coupled}}}{P_0}\right]^{1/2} = \left|\sin\left(\frac{\pi \Delta n}{\lambda} \cdot L_c\right)\right|, \tag{2.35}$$

$$= \left|\sin\left(\frac{\pi}{2} \cdot \frac{z}{L_{\text{cross}}}\right)\right|. \tag{2.36}$$

Directional couplers are very narrow bandwidth in function, due to their phase dependence which changes with wavelength. They are a frequent building block component as optical power splitters, and they are often used as a component of other fundamental integrated elements, such as Ring Resonators (RRs). In addition, DCs have been demonstrated as on-chip mode converters [51], and erasable DCs have been utilised as temporary probing points for the testing of complex circuits [26].

2.2.4.2 Multimode interference devices

Multimode Interference (MMI) devices are a common integrated photonic component used in a wide variety of devices. They were initially designed as Two-Mode Interference (TMI) coupler, which are essentially DCs with a zero-gap size [52, 53]. It was found that performance could be improved by increasing the coupling region's width to support multiple modes [54]. These modes are free to interfere with one another and create an electric field distribution that will transmute periodically along the propagation direction [55]. At certain positions, known as self-imaging points, the electric field will converge to N points for the M^{th} time. The distance at which these points occur can be found in terms of the beat length between the fundamental and first-order mode, L_π, which is given by

$$L_\pi = \frac{\pi}{\beta_0 - \beta_1}, \tag{2.37}$$

and can be rewritten as (see Appendix A.2 for derivation)

$$L_\pi \approx \frac{4 n_{\text{eff}} w_{\text{eff}}^2}{3\lambda_0}, \tag{2.38}$$

where β_0 and β_1 are the propagation constants for the fundamental and first-order mode respectively, n_{eff} is the effective refractive index of the waveguide, w_{eff} is the effective width of the MMI region and λ_0 is the wavelength in a vacuum. The effective width takes into account the penetration depth of the mode, and can be approximated to be

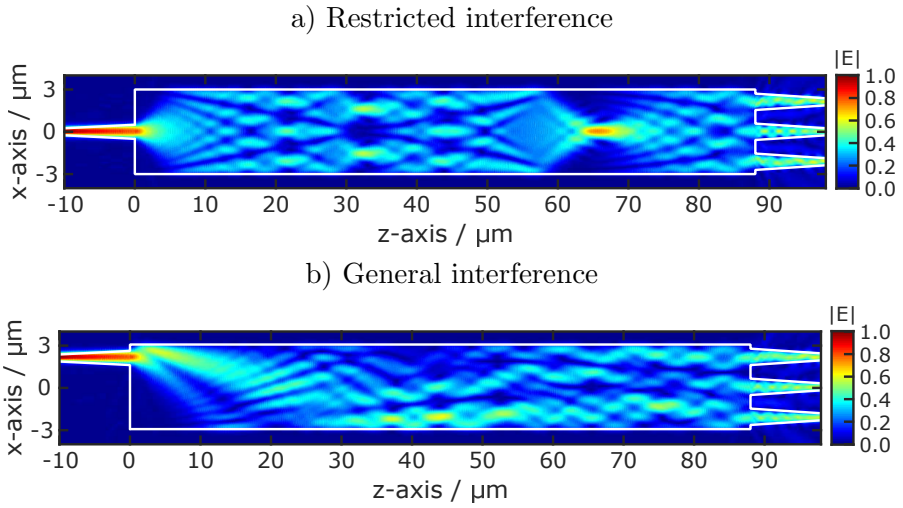


Figure 2.9: Electric field distribution of a symmetrical and asymmetrical MMI device. a) MMI power splitter with one central input, so the field distribution is symmetrical around the x-axis — restricted interference. b) MMI power splitter with an offset input, so the field distribution is asymmetrical around the x-axis — general interference. Both devices are 1×3 MMI splitters with a MMI region of width 6 µm and length 88 µm. Simulated using Lumerical FDTD solver for TE input modes at a wavelength of 1550 nm.

equal to the width of the MMI region, W, for high-contrast waveguide boundaries, such as silicon-air and silicon-silica interfaces [56].

There are two self-imaging mechanisms; restricted interference and general interference [57]. Examples of both of these are shown for a 1×3 MMI power splitter in Figure 2.9. In the case of restricted interference, there is one central input. The device's symmetry about this centralised input results in the electric field distribution also being symmetrical in the x-axis. The various self-imaging points are clearly visible along the length of the device; in particular, the 1-to-1 point can be seen at 66 µm along the z-axis. The distance of the M$^{\text{th}}$ N-fold image, L, is given by

$$L = \frac{M}{N}\left(\frac{3}{4}L_\pi\right) \approx \frac{M}{N}\frac{n_{\text{eff}}W^2}{\lambda_0}. \tag{2.39}$$

While the general interference case describes MMI devices with an off-centre input, such that the electric field is no longer symmetrical around the x-axis. In this case, the length of L is given by

$$L = \frac{M}{N}(3L_\pi) \approx \frac{M}{N}\frac{4n_{\text{eff}}W^2}{\lambda_0}. \tag{2.40}$$

These equations provide a good initial value for the required device dimensions for a particular power splitting ratio. As expected from these equations and shown by simulation in Figure 2.9, the first three-fold imaging point occurs at a quarter of the central input case's length compared to that of the off-centre input. Further optimisation of

device dimensions can then be carried out by parameter sweeps in computer simulations and is discussed further in Section 3.2.

2.2.4.3 Mach-Zehnder interferometer

A Mach-Zehnder Interferometer (MZI) is essentially just two beam splitters configured in such a way that the first splits an input light source into two paths, and the second recombines the two arms. Given that the input light is coherent, these two arms are very sensitive to small optical path length differences of the two arms [14, 28]. When recombined the relative phase difference will dictate whether the light constructively or destructively interferes.

The integrated MZI works in precisely the same manor as its free-space counterpart, but with the beam splitters replaced by their integrated equivalents; either directional, Y-branch or MMI 3 dB couplers. The light is also confined to waveguides rather than propagating in free-space, so no image of the interference pattern can be produced. A schematic for an integrated MZI is shown in Figure 2.10 a). The phase difference, $\Delta\phi$, between the two arms now originates from a different effective path length between the two waveguide arms. This can be caused by either a different physical path length, ΔL, in which case,

$$\Delta\phi = \frac{2\pi}{\lambda_0} n_{\text{eff}} \Delta L = \beta \Delta L, \tag{2.41}$$

or through some change to the effective index, Δn_{eff}, of one arm,

$$\Delta\phi = \frac{2\pi}{\lambda_0} \Delta n_{\text{eff}} d, \tag{2.42}$$

where λ_0 is the wavelength and d is the length of the region with different effective index. An MZI has a theoretical transfer function given by [14, 58]

$$P_{\text{out}} = P_{\text{in}} \cos^2\left(\frac{\Delta\phi}{2}\right). \tag{2.43}$$

Figure 2.10 b) shows the simulated electric field distribution of the output coupler, in this example an MMI device, for both the in-phase and antiphase cases. There is also an additional pair of output waveguides to collect the antiphase component of light. The real part of E_x for the input and output waveguides is shown in Figure 2.10 c) and provides a clear visualisation of the relative phase between the two arms. The transfer function for this simulated device is given in Figure 2.11. Unlike in Equation 2.43, the transmission does not oscillate perfectly between 1 and 0. This is due to the excess loss of the MMI devices used as the couplers. The Extinction Ratio (ER) is also limited by the balance of power between the two arms resulting in imperfect interference, and the variance of phase at the output waveguide. This phase variance can be either from small environmental fluctuations to the phase shift or stem directly from the phase-noise

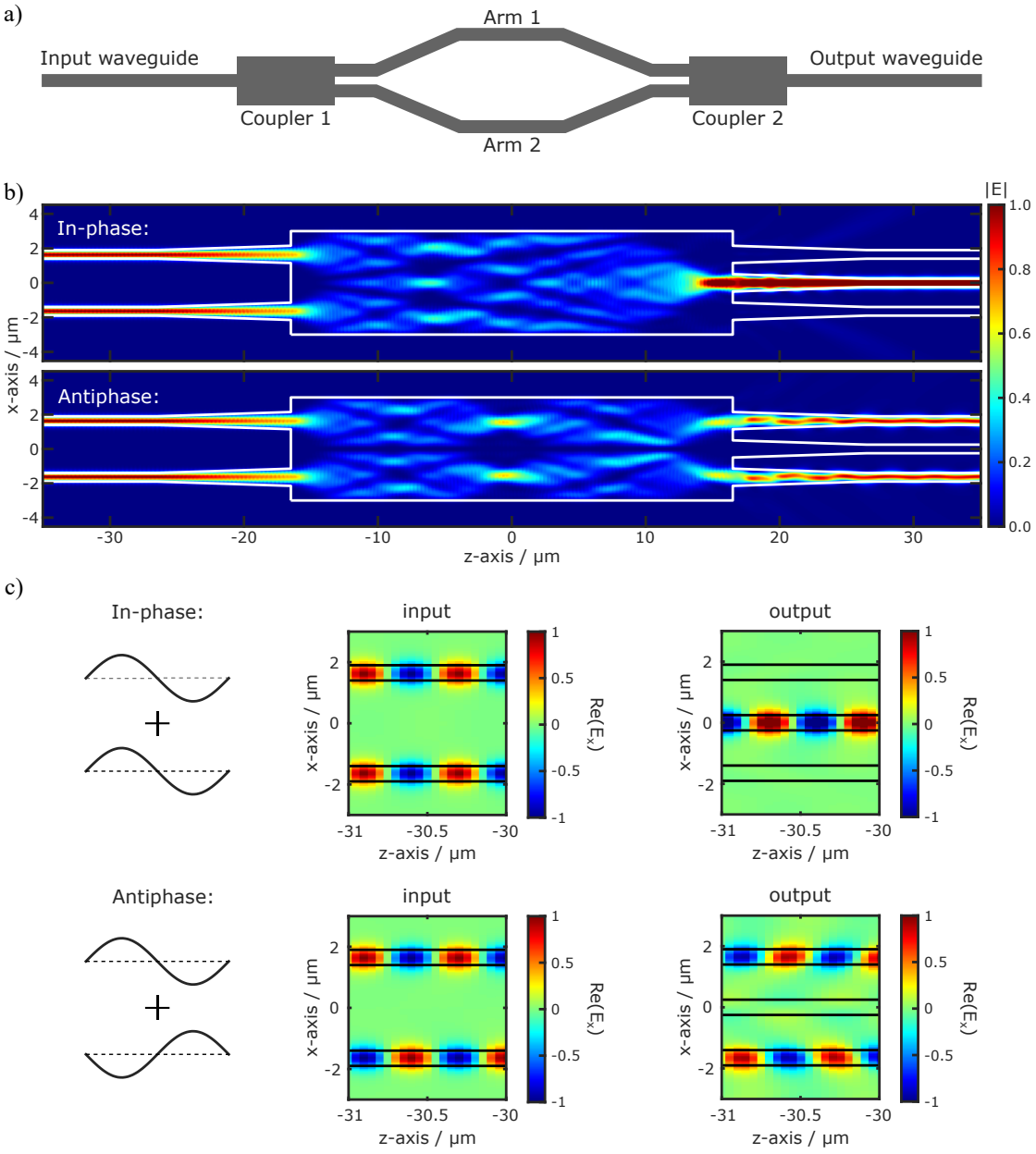


Figure 2.10: a) Schematic for an integrated Mach-Zehnder Interferometer (MZI). b) Simulated electric field distribution of an output MMI coupler for both the in-phase and antiphase cases. c) Corresponding real part of E_x for the input and output waveguides in b). Simulated using Lumerical FDTD solver for the fundamental TE input mode at a wavelength of 1550 nm.

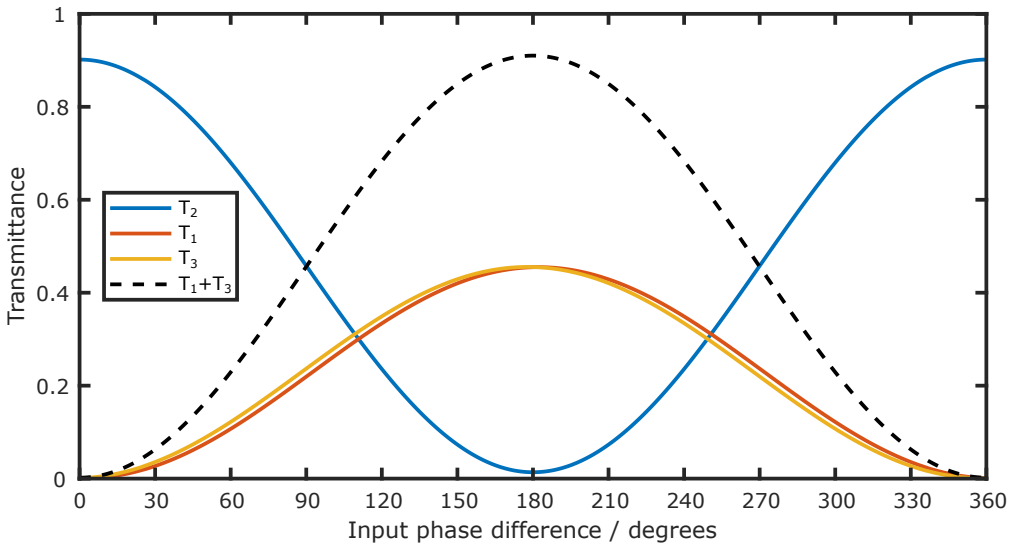


Figure 2.11: The transfer function for the output MMI of the MZI in Figure 2.10. T_1, T_2 and T_3 represent the top, middle and bottom outputs, respectively. Simulated using Lumerical FDTD solver for the fundamental TE input mode at a wavelength of 1550 nm.

or linewidth of the laser source itself. In integrated optics, the Mach-Zehnder is most frequently used as a modulator [59–61] or a switch [62–66], but it also has applications in multiplexing [67–69] and sensing [70–73]. A key performance figure for Mach-Zehnder Modulators (MZMs) is V_π, which is the voltage required to move between a maximum and null position on the transfer function.

2.3 Modulation mechanisms in silicon

Modulation in photonic waveguides is typically achieved in one of two ways; direct increases to the optical absorption coefficient, α, or through phase differences caused by an alteration to the real part of the refractive index that can then be exploited to modulate the intensity of light. These changes can stem from a number of different physical mechanisms, such as the linear electro-optic (Pockels) effect, various non-linear effects, the thermo-optic effect and the plasma dispersion effect. The most common modulation methods in other waveguide platforms are the Pockels, Kerr and Franz-Keldysh effects; however, these are either not possible or very weak in silicon [74].

Crystalline silicon is centrosymmetric in structure, and as such does not exhibit the linear electro-optic effect or any other even-order nonlinear optical effects. However, there have been several studies into inducing these effects in silicon through methods such as strain engineering [75–77]. The Kerr effect is the intensity-dependent refractive index, $n = n_0 + n_2 |E|^2$, third-order non-linearity that is present in silicon; however, studies have shown the Kerr effect to be very small in silicon, $n_2 = 4 \times 10^{-14}\,\mathrm{cm^2\,W^{-1}}$ [78, 79]. While this is around 100 times larger than silica at 1550 nm, this still only amounts

to a $\Delta n = 10^{-4}$ for and applied electric field of $E = 10^6\,\mathrm{V\,cm^{-1}}$ [80]. The Franz-Keldysh effect describes the change in optical absorption of a semiconductor through the distortion of the energy bands when an electric field is applied [29]. The electric field shifts the bandgap energy, changing the absorption properties of the crystal, especially for photons near the bandgap energy. Once again, previous studies have found its effect to be very small in silicon, particularly at telecommunication wavelengths [80].

This leaves the thermo-optic and plasma dispersion effects. The reasonable large thermo-optic effect, $dn/dT = 1.86 \times 10^{-4} K^{-1}$, in silicon does indeed allow for significant modulation. However, Equation 2.42 shows that a $\pi/2$ phase change still results in heater path lengths of $\sim$200 µm for typical thermally-induced index shifts on the order of 2×10^{-3}. Additionally, thermal recovery times are typically too slow for the vast majority of applications and thermal loading of the chip leads to high energy consumption [81]. In fact, temperature dependence is actually a frequent issue in silicon devices and requires careful active control to maintain device performance. Therefore, the thermo-optic effect is typically only exploited for the trimming or tuning of devices through integrated resistive heaters, which can be set to compensate for manufacturing variation and environmental changes [82]. This means that it is the plasma dispersion effect that is most commonly exploited for modulation in silicon photonic devices, which is outlined in more detail in the following subsection.

2.3.1 Plasma dispersion effect

The plasma dispersion effect describes changes to the complex refractive index, $\underline{n} = n + i\kappa$, where κ is the extinction coefficient, in a semiconductor due to free-carrier concentrations. The Drude-Lorenz equations are used to calculate the magnitude of change to the real part of the refractive index and absorption coefficient, which is related to κ by $\alpha = 4\pi\kappa/\lambda_0$, and are given by [80]

$$\Delta n = \Delta n_e + \Delta n_h = -\frac{e^2\lambda_0^2}{8\pi^2 c^2 \epsilon_0 n}\left(\frac{\Delta N_e}{m_{ce}^*} + \frac{\Delta N_h}{m_{ch}^*}\right), \tag{2.44}$$

and

$$\Delta\alpha = \Delta\alpha_e + \Delta\alpha_h = \frac{e^3\lambda_0^2}{4\pi^2 c^3 \epsilon_0 n}\left(\frac{\Delta N_e}{\mu_e \left(m_{ce}^*\right)^2} + \frac{\Delta N_h}{\mu_h \left(m_{ch}^*\right)^2}\right) \quad [\mathrm{cm^{-1}}], \tag{2.45}$$

where Δn_e and Δn_h are the changes to the real part of the refractive index attributed to electrons and holes respectively, similarly $\Delta\alpha_e$ and $\Delta\alpha_h$ are the changes to the absorption coefficient, e is the electron charge, ϵ_0 is the permittivity of free space, ΔN_e and ΔN_h are the electron and hole density per $\mathrm{cm^3}$, m_{ce}^* and m_{ch}^* are the conductivity effective mass of electrons and holes, and μ_e and μ_h are the electron and hole mobility.

For 1550 nm light, these equations have been shown to simplify to [14, 74, 83]

$$\Delta n = -\left[8.8 \times 10^{-22} \Delta N_e + 8.5 \times 10^{-18} \left(\Delta N_h\right)^{0.8}\right], \tag{2.46}$$

and

$$\Delta \alpha = 8.5 \times 10^{-18} \Delta N_e + 6.0 \times 10^{-18} \Delta N_h \quad [\mathrm{cm}^{-1}]. \tag{2.47}$$

It is clear from these equations that there is a trade-off to be made when increasing the number of free-carriers between decreasing electro-refraction and increasing the electro-absorption. The electrical manipulation of the charge density is typically achieved by PN junctions through mechanisms such as carrier injection, carrier accumulation or carrier depletion [74, 84]. In carrier accumulation, a thin oxide barrier is used to separate two halves of a waveguide to form a capacitor, while carrier injection utilises highly doped p- and n-regions separated by an intrinsic region, which contains the waveguide. Forward-biasing this PIN diode injects free electrons and holes into the intrinsic waveguide region. Finally, in carrier depletion, reverse biasing is applied over lightly doped p- and n-doped regions that make up the waveguide structure and form a PN diode. The carrier depletion approach has become the predominate modulation scheme in silicon devices due to its relatively straightforward fabrication [16, 85, 86].

Typical free-carrier concentrations that can be achieved by carrier injection and depletion are on the order of $10^{18}\,\mathrm{cm}^{-3}$ [84], which corresponds to an index shift of $\Delta n = -3 \times 10^{-3}$ at a wavelength of 1550 nm. Such index shifts result in similar path lengths as with the thermal effect but at a much higher modulation speed in the gigahertz regime. However, as previously stated, this electro-refraction change coincides with an undesirable increase in absorption due to the free-carriers. It is therefore fairly commonplace to integrate a more suitable platform, e.g. InP or $LiNbO_3$, in a hybrid manner for devices that require many active components [87].

2.4 Summary

In this chapter we have introduced the key underlying physics used in this thesis, starting with an introduction to fundamental waveguide theory and some of the common assumptions and approximations employed. These concepts provide a solid basis to work from, and many of them will be visited repeatedly throughout this thesis. Next, various input-output coupling schemes were discussed, and many of the fundamental building blocks used to build up integrated photonic circuits are introduced. Finally, different modulation mechanisms for silicon are outlined, particularly free-carrier dispersion, the effect exploited by Ultrafast Photomodulation Spectroscopy (UPMS) which will be introduced in the following chapter.

Chapter 3

Methodology: Fabrication, Simulation and Measurement Techniques

This chapter will cover the different approaches for the fabrication of samples, as well as the main numerical and experimental methods. The goal is to outline the typical approaches used within this thesis and provide a solid foundation to work from going forward. Additionally, the key parameters of the Ultrafast Photomodulation Spectroscopy (UPMS) setup are also characterised in this chapter and will be referred back to multiple times throughout this work.

3.1 Fabrication

The vast majority of samples investigated in this thesis were fabricated in-house by the Silicon Photonics Group, and where samples fabricated by collaborators are used it will be clearly stated. Typically, the in-house samples were fabricated on commercial Silicon-On-Insulator (SOI) wafers, which consists of a thin 220 nm mono-crystal top silicon layer on top of a 3 µm buried oxide layer with a base silicon layer. Two main lithographic techniques were used to directly expose devices into the silicon: Electron-beam lithography (commonly referred to as e-beam lithography) and deep–Ultraviolet (UV) projection lithography. Each has its advantages and disadvantages, which essentially comes down to e-beam lithography being a fantastic research tool while deep–UV lithography is better suited for industry.

E-beam lithography allows for very small feature sizes, on the order of 10 nm, to be written without the requirement of a mask, making it one of the most versatile techniques

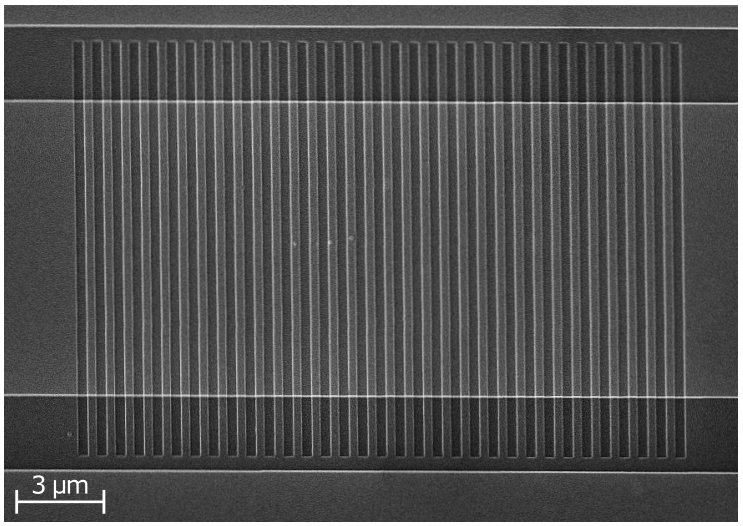


Figure 3.1: SEM image of a grating coupler as fabricated on a 220 nm thick SOI rib waveguide with a 3 µm wide trench and 120 nm etch depth. The waveguide is 10 µm wide to provide good mode matching with the MFD of a standard 1550 nm SMF fibre. The grating itself is made with a 70 nm etch depth, 620 nm pitch and a duty cycle of 0.5. Image is at four thousand times magnification.

for research and prototyping. However, the write times are very long, which limits the number of chips produced and consequently the commercial viability of the approach.

The other approach is deep–UV projection lithography with a Nikon NSR-S204B scanner. This system utilises a KrF excimer laser at 248 nm with a Numerical Aperture (NA) of 0.68 to achieve a minimum feature size of ∼150 nm. The most significant advantage of a scanner system is the speed at which devices can be written. A slit of UV light is scanned over the mask directly projecting the chip design onto the wafer. This approach is the one most widely adopted by industry.

Chips were either fabricated as ridge waveguides with a full 220 nm etch depth or rib waveguides which use a 120 nm etch depth with a 3 µm wide trench region. Rib waveguides have slightly poorer confinement but experience less sidewall scattering [14]. A second etch depth of 70 nm is used to write grating couplers to couple light into and out of the chips. The gratings were made with a pitch of 620 nm and a duty cycle of 0.5, which was confirmed by Scanning Electron Microscope (SEM) as shown in Figure 3.1. The waveguide at the point of the grating is 10 µm wide to provide good mode matching with the Mode Field Diameter (MFD) of a standard telecommunications Single-Mode Fibre (SMF), and then adiabatically taper down to single-mode 500 nm wide waveguides.

Straight test waveguides were included on all chips to be used as references so that the coupling and propagation loss can be measured and used to calculate individual device insertion loss using broadband swept transmission measurements, as discussed in Section 3.3.2.

3.2 Numerical methods

Computer simulation is a vital tool for both the design of new Photonic Integrated Circuits (PICs) and the confirmation of experimental results. Individual components can be modelled and optimised via various parameter sweeps. The scattering matrix, or S-matrix, describes the full complex coupling between every port, wavelength and supported mode of a photonic device. Once the component designs are finalised, their resulting S-matrix parameters can be combined to enable complete circuit simulations, where the final function of the PIC can be assessed [14]. The chip's physical layout can then be finalised as a lithographic mask and verified against various fabrication rules and restrictions. The results of this verification enable circuit simulations to predict the system response due to effects from the environment or fabrication, such as lithography effects, wafer non-uniformity and device temperature. Post-fabrication tests then provide an opportunity to feed the real-world chip performance back into the simulation for further optimisation until convergence is achieved and a design is finalised.

A variety of modelling techniques were used to simulate devices and their physical responses in this work. Primarily Finite-Difference Time-Domain (FDTD) and aperiodic-Fourier Modal Method (a-FMM), which are outlined in the following subsections.

3.2.1 Finite-difference time-domain

The FDTD method, also known as Yee's method after the mathematician Kane Yee who developed the approach in the 1960s [88], is a numerical analysis technique used for modelling electrodynamics by solving Maxwell's equations in discretised volumes. Iteratively solving for the electric and magnetic fields in time steps simulates the propagation of light through the system, and is repeated until the maximum simulation time or the desired field behaviour has fully converged [89]. The spatial mesh does not need to be uniform, allowing for a denser mesh for smaller feature size regions or areas of interest. Being time-domain based gives FDTD a couple of advantages compared to other numerical solvers. Firstly, it is possible to cover a wide range of frequencies with a single simulation run. And secondly, the manner in which simulations are carried out is very intuitive, allowing temporal visualisation of the propagation of light.

For PICs, in which light typically only propagates in-plane, this volumetric approach can be simplified to two-dimensions using the effective index method (see Section 2.2.2) to approximate the waveguide structure [36]. This represents a significant improvement of simulation speed as 3D simulations scale with mesh size to the fourth power, while 2D scales with the third power [90, 91]. A comparison between 2D and 3D FDTD simulations for 1×2 Multimode Interference (MMI) is shown in Figure 3.2, for the same fixed mesh size of 20 nm in all axes. The agreement is very good, with around a 1% transmittance discrepancy in each output, which is explained by the additional out-of-plane

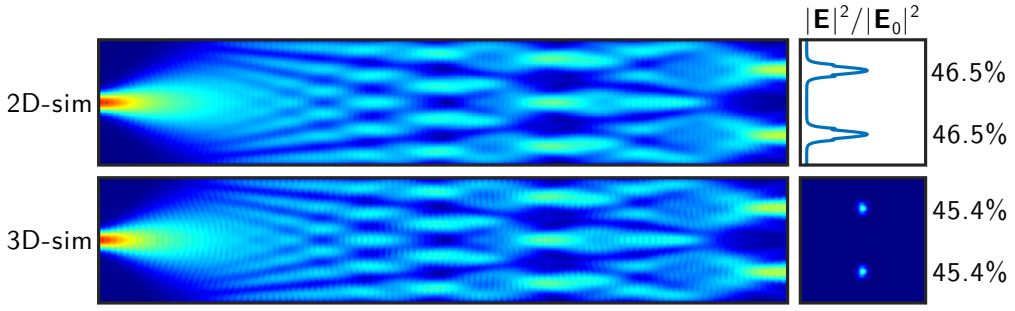


Figure 3.2: Comparison of 2D and 3D FDTD simulations for a 1×2 MMI. Transmission and output profiles are displayed to the right-hand side of the electric field distributions with the respective port transmittance values. Simulations were carried out with Lumerical FDTD for incident light in the fundamental Transverse Electrical (TE) mode at a wavelength of 1550 nm. The top image corresponds to 2D effective index simulations and the bottom image to a slice through the middle of a full 3D simulation.

scattering in 3D that is not considered in the 2D approximation. For the vast majority of simulations in this work, the 2D approximation will be used due to the significantly faster computational simulation times. When relevant additional comparisons will be made in full 3D.

3.2.2 Aperiodic-Fourier modal method

The Fourier Modal Method (FMM), also known as Rigorous Coupled-Wave Analysis (RCWA), is an effective numerical method for solving Maxwell's equations in structures with some spatial periodicity. Initially, it was used to simulate the diffraction of a plane wave from a 1D grating [92]. The approach was then modified to allow the calculation of the eigenmodes of a variety of different 1D and 2D structures [93].

A significant development of FMM was the aperiodic-Fourier Modal Method (a-FMM), in which perfectly matched layers are applied to the edges of the computational cell to isolate adjacent periods [94–96]. This enables the technique to be applied to integrated optics problems by calculating all the input and output modes of the system in a Fourier basis. The input modes are normalised and through the S-matrix formalism, the complex coupling between the propagating modes for each section of the waveguide is calculated, e.g. input waveguides to MMI region to output waveguides. This approach has been demonstrated to be capable of accurately modelling various PICs [97, 98]. Once again, the effective index method is used to reduce the system to a two-dimensional one.

The a-FMM simulations were carried out using a modified version of the open-access "RETICOLO" RCWA software package, that was provided by the Light in Complex Nanostructures Group from the Laboratory for Photonics, Numerics and Nanosciences (LP2N), Bordeaux, France [99].

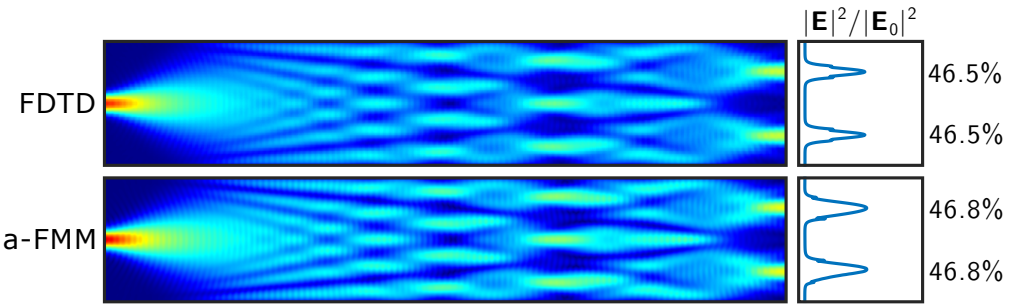


Figure 3.3: Comparison of FDTD and a-FMM simulations for a 1×2 MMI. Transmission and output profiles are displayed to the right-hand side of the electric field distributions with the respective port transmittance values. Simulations were carried out in 2D using the effective index method with the same device parameters. Incident light is in the fundamental TE mode at a wavelength of 1550 nm. The output waveguides for the a-FMM model are 1 µm wide, opposed to 0.5 µm for FDTD, due to a-FMM's inability to efficiently model tapered waveguides.

3.2.3 Comparison of simulation techniques

Since FDTD is a time-domain method, it has the advantage that solutions inherently include the frequency response with a single simulation run up to the Nyquist–Shannon sampling limit. a-FMM is single-frequency, so multiple simulations must be carried out to build up a wavelength response. Meanwhile, a-FMM is significantly faster at extracting the S-matrix parameters of a simple system compared to FDTD. As previously mentioned, FDTD is also better suited at modelling complex geometries due to its physical meshing of the 3D structure, while a-FMM must solve for the modes supported in each unique cross-section of the structure, resulting in a substantial speed penalty. In this thesis, both techniques will be used depending on which is better suited to the problem at hand. In all cases, the technique used will be clearly stated.

Figure 3.3 shows a comparison between the FDTD and a-FMM simulation techniques for a 1×2 MMI device. Both simulations were carried out in 2D using the effective index method with the same device parameters, except for the output waveguides which are 1 µm wide for the a-FMM model, opposed to 0.5 µm for FDTD, due to a-FMM's inability to efficiently model tapered waveguides. The agreement between the two techniques is excellent, with less than 0.5% difference in transmittance to a given output port for this particular device.

3.3 Measurement techniques

This section will outline some of the more basic measurement techniques that will be frequently used throughout this work. Followed by a more in-depth look at Ultrafast Photomodulation Spectroscopy (UPMS) and the characterisation of its key parameters.

3.3.1 Scanning electron microscopy

SEM is a technique that is capable of much higher magnification and resolution than conventional optical microscopy systems. This is due to the de Broglie wavelength of a high energy electron being significantly shorter than the wavelength of visible light, and consequently, the diffraction limit is also reduced. An SEM system consists of an electron beam fired at a sample under vacuum, and the scattered electrons are then collected. The incident beam can then be raster-scanned over the sample to build up a two-dimensional image of the sample. Non-conducting materials need first to be coated in a metal, typically gold, so as to prevent charge build-up on the sample which alters the deflection of the electrons. As a semiconductor, silicon does not require a gold coating, but care must be taken that the build-up of charge does not distort the image. Once calibrated, SEM images are an excellent way of measuring devices on the nanoscale and are frequently used to confirm the correct fabrication of PICs. Figure 3.1 shows an example SEM image of an SOI grating coupler.

3.3.2 Broadband swept transmission measurements

The S-matrix describes the full complex coupling between every port of a photonic device and can be characterised using a network analyser; however, in many instances, only the transmitted intensity is of interest. In these cases, broadband transmission measurements are typically used to assess device performance.

For the broadband swept transmission measurements carried out in this work, an Agilent 8163B lightwave multimeter was used in conjunction with the 81940A tunable laser and 81634B power sensor modules. These modules are directly coupled to Single-Mode Fibres (SMFs), and a three-paddle polarisation controller provides input polarisation control to minimise rippling effects from grating couplers. An overhead camera is used to position two angle-adjustable fibre arms to couple into and out of the device under test. The tunable laser can be swept over 1505 nm to 1630 nm in steps as low as 0.1 pm with a narrow linewidth of 100 kHz, and the InGaAs power sensor is capable of recording powers between 10 dBm to −110 dBm. This enables the broadband responses of devices to be measured along with any very narrow resonances with high extinction ratios that may be present. The device's transmission can then be normalised to the transmission of a straight reference waveguide to obtain the component insertion loss, as any common propagation and coupling losses will be removed.

3.3.3 Raman spectroscopy

Raman spectroscopy is a technique based on the inelastic scattering of light to identify specific vibrational modes of bonds [28]. A large proportion of light will scatter elastically, that is to say at the same wavelength as the incident light, but there will also be a small fraction of the light will lose some energy through the excitation of vibrational modes in the material resulting in a reduction in energy of the scattered photon. This process is referred to as Stokes scattering. The opposite can also occur, called anti-Stokes, where the material is already in a vibrational state, and the scattered photon gains energy from the material.

For the measurements in this thesis, a Renishaw inVia laser Raman spectrometer was used, in which a $\times 50$ objective focuses a Nd:YAG laser with wavelength 532 nm onto the sample under test. A spectrometer then collects the scattered light with variable exposure time and number of accumulations to collect sufficient light to maximise the Signal-to-Noise Ratio (SNR).

3.3.4 Pump-probe spectroscopy

Pump-probe measurements are commonly used to investigate ultrafast phenomena where direct temporal measurements are not possible [100–103]. This is typically due to the response being significantly faster than even high-speed electronics can observe. Instead, the time delay between the probe and pump pulses are controlled spatially via a variable delay stage, where small changes in optical path length corresponding to minuscule changes in time. The pump pulse is used to generate some excitation in the sample, resulting in a subsequent change in the transmission or reflection of the probe pulse. By recording this change as a function of the delay time between the two pulses, it is possible to build up a temporal response at the resolution of the variable time delay stage's step size. An example of such a response is shown in Figure 3.4. The technique also has been demonstrated in waveguide geometries, where the pump and probe pulse are collinearly coupled into the device to investigate non-linear properties [104–106]. The technique used here differs from the traditional collinear pump-probe setup in the fact that the pump light arrives externally onto the surface of the device, perpendicular to the plane in which the probe light propagates. Therefore, the time delay also dictates the pulses' spatial overlap position along the propagation axis of the device under test, due to the travel time of the probe pulse. An illustration of a pump-probe experimental setup for a waveguide device is provided in Figure 3.5.

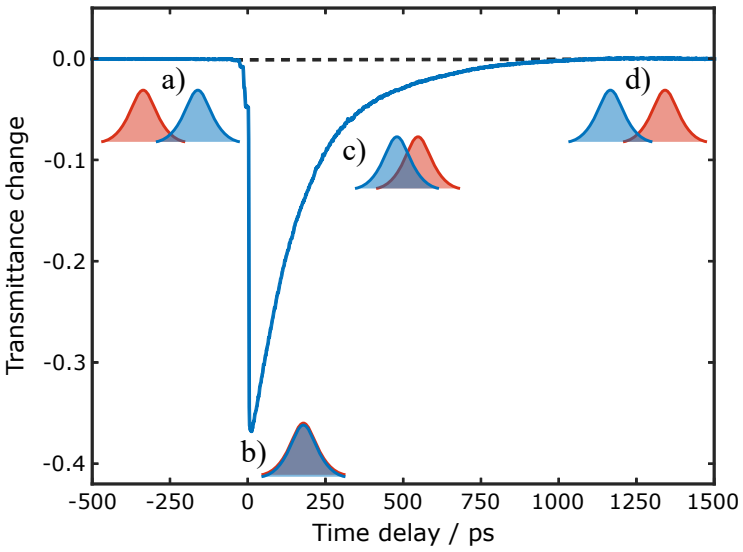


Figure 3.4: An example of a typical pump-probe response taken from externally pumping a silicon waveguide and measuring transmittance. a) The probe pulse arrives before the pump, seeing an unperturbed device, and therefore the transmittance is unchanged. b) The pump and probe pulse arrive simultaneously at the same spatial position in the device, which results in the greatest change in transmittance. c) The pump pulse arrives slightly before the probe. Transmittance depends on how long the device has had to recover. d) The pump pulse arrives sufficiently before the probe, such that the device has fully recovered and transmittance returns to the unperturbed state.

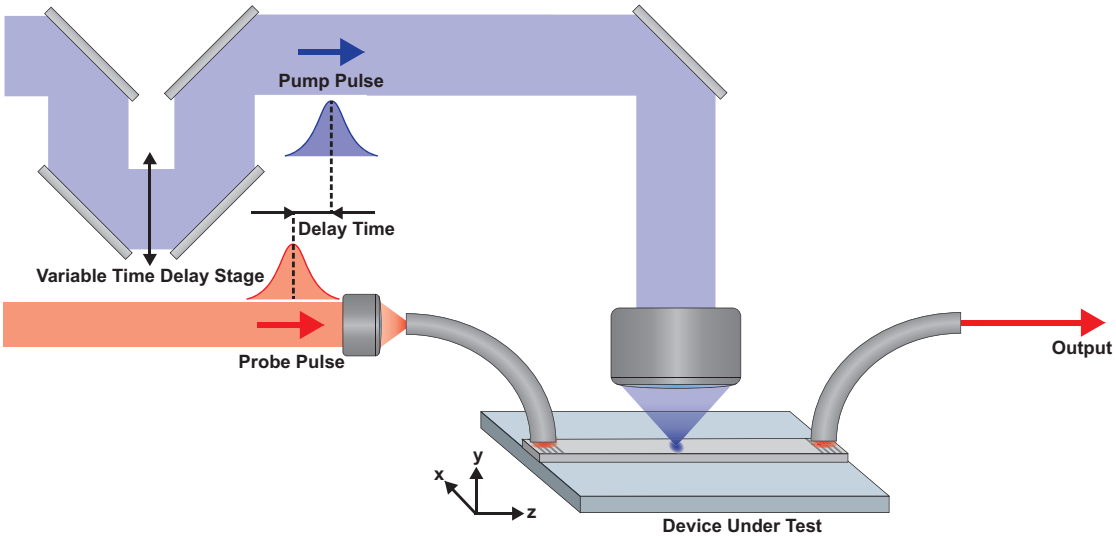


Figure 3.5: An illustration of the pump-probe technique. The probe pulse is coupled into the device via an optical fibre, while the pump pulse is externally focused onto the surface of the device under test. The temporal separation of the two pulses is finely controlled through adjustments to a multi-pass variable time delay stage. Device recovery is then investigated by recording the change in transmission as a function of the delay time.

3.4 Ultrafast photomodulation spectroscopy

In its essence, UPMS is a form of pump-probe measurement with a few key differences. Firstly, rather than the two pulses being coincident, the pump light arrives externally onto the device's surface — perpendicular to the plane in which the probe light propagates. The modulation of the probe light is achieved through highly localised alterations of the refractive index profile of the sample via the plasma dispersion effect (Section 2.3.1). Secondly, the focusing objective is mounted to a three-dimensional nanopositioner, allowing the position of the perturbation to be moved over the surface of the device. This provides the system with a selection of different parameters that can be varied to characterise a device under test, such as the delay time between pulses, the probe wavelength of light, all three spatial dimensions of the pump spot, and the pump power. Varying the time delay allows time-domain responses to be investigated, such as free-carrier lifetimes, direct time of flight measurements, group index of modes and quality factors (also known as Q factor) of cavities. The probe light's central wavelength can be altered and its broadband nature means it can be used in conjunction with a spectrometer to obtain spectral information on the device under test. Two of the three spatial dimensions allow the change in transmission to be recorded as a function of perturbation position, resulting in a spatial transmission sensitivity map (also referred to as photomodulation maps). While the third spatial dimension can be used to change the focus of the pump, which allows the perturbation size, and consequentially magnitude, to be controlled and can also be used to investigate three-dimensional structures. Finally, changes to the pump pulse's power can be used to directly control the strength of modulation in the device.

A schematic of the experimental setup is provided in Figure 3.6. A Coherent Chameleon Ultra II Ti:Sapphire laser is used in conjunction with an Optical Parametric Oscillator (OPO) to provide the variable wavelength probe pulses, while the Second-Harmonic Generation (SHG) of the depleted seed produces the pump pulses. Both of these pulses stem from the same initial seed pulse and therefore have the same pulse duration of 200 fs with a repetition rate of 80 MHz, unless the pulse select is used which halves the pump's repetition rate to 40 MHz. The optical setup then consists of two main beam paths; one for the probe and one for the pump.

The probe pulses are set to a central wavelength of choice by the OPO within the range of 1000 nm to 1600 nm. A Half-Wave Plate (HWP) then rotates the polarisation of the beam, which for grating couplers allows for optimal coupling to the waveguide and in the case of end coupling is also used to select between the TE and Transverse Magnetic (TM) modes. The probe is then free-space coupled to an optical fibre using a fibre collimator. The fibre enables the probe to be easily end-coupled, butt-coupled or grating coupled into the device under test depending on the chips design and various lengths of fibre can be used to compensate for different pulse delay times. A second

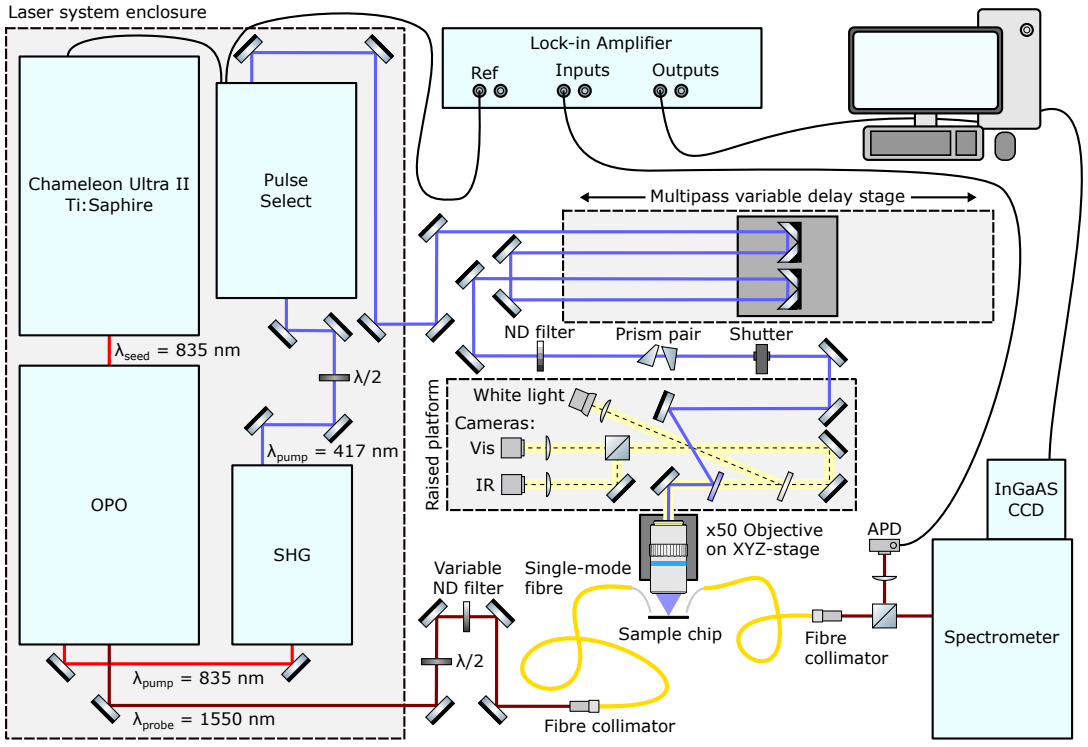


Figure 3.6: Schematic of the ultrafast photomodulation spectroscopy setup using the pulse select. In the case where the pulse select isn't used, it is bypassed and optical choppers provide modulation for both the probe and pump beam paths.

fibre is used to collect the light from the desired output and transmission is recorded by either an Avalanche Photodetector (APD) connected to a lock-in amplifier or a grating spectrometer if wavelength dependence is desired.

Meanwhile, a HWP is also used to rotate the pump pulses' polarisation to optimise the efficiency of the pulse select, which can be bypassed if not in use. The beam is then passed through a four-pass variable delay stage, allowing fine control of the time delay between the probe and pump pulses on the order of sub-picosecond. A variable Neutral Density (ND) filter is used to limit the pump power and set the desired effective index modulation and a 2× magnification anamorphic prism pair is used to correct the ellipticity of the beam. The pump pulses are then focused onto the surface of the device under test using a 50× objective with a NA of 0.55. The beam slightly overfills the focusing objective which is mounted to a nanopositioner, providing accurate repeatable positioning (on the order of <100 nm) and focusing of the pump spot with minimum fluence deviation.

The pump's focusing objective is also used to image the device with either an infrared (IR) or visible light camera. The visible light camera is used for the approximate alignment of fibres and positioning of the pump spot, and the Infrared (IR) camera is used to look at any probe light scattered out of the device.

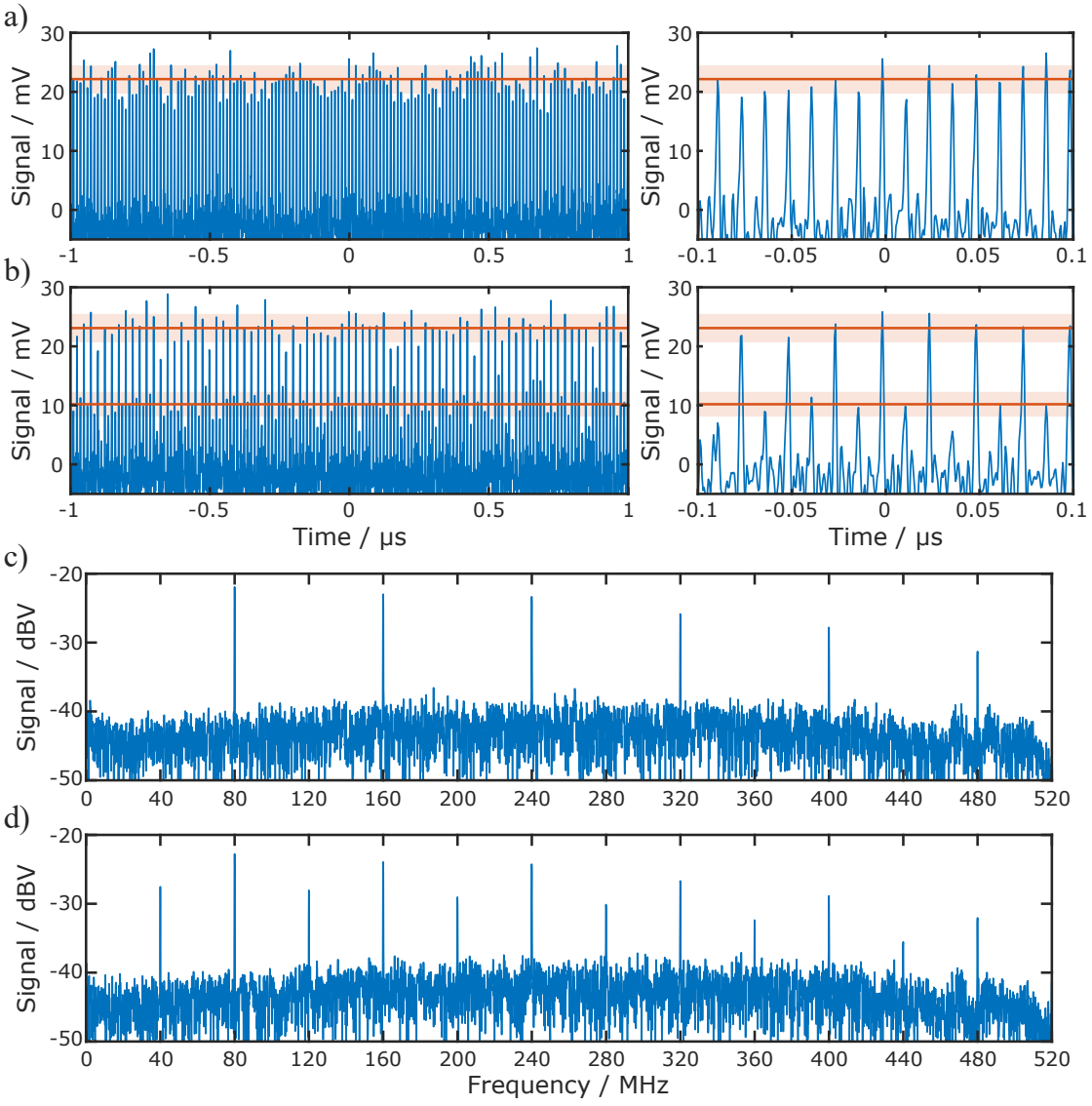


Figure 3.7: Ultrafast photomodulation spectroscopy pump-probe response using the pulse select. a) Temporal signal from the avalanche photodiode with an 80 MHz probe pulse train for a 2 µs (left) and 0.2 µs (right) time windows. b) Same as previous, but with the addition of the 40 MHz pump pulse train. The red lines and shaded regions indicate the mean and standard deviation of the 80 MHz and 40 MHz signals from the full 9 µs time window. c) and d) are the Fourier transforms for a) and b) respectively; clearly indicating the 40 MHz and 80 MHz signals with their harmonics. Measurements were taken with a single-mode 220 nm thick and 500 nm wide SOI waveguide with the pump positioned at the middle of the waveguide.

3.4.1 Experimental method

The femtosecond optical pump pulses locally reduce the silicon's refractive index via the plasma dispersion effect [80]. Due to the ultrafast nature of the pump light, free-carrier concentrations far exceeding those achievable by electrical effects can be obtained [107]. This local perturbation in the refractive index then in turn modulates the transmission of the device. By precisely recording the change in transmission as a function of the perturbation position it is possible to build up a 2D spatial map of the photomodulation response, which provides direct visualisation of the propagation of light through the device. The resolution of such maps is ultimately limited by the size of the perturbation and the minimum step size of the nanopositioner that the objective is mounted to.

To obtain the normalised transmission change, $\Delta T/T$, both beams are modulated either by optical choppers set to different frequencies or directly by their repetition rates if the pulse select is used. These frequencies can then be used as references for the dual-channel lock-in amplifier. In the case of pulse select being used, the signal with reference to the probe beam's frequency records the overall transmission, $T_{probe} = T + \Delta T$, where T is the unperturbed transmission and ΔT is the change in transmission. The other signal at the pump beam's frequency records only the change in transmission, $T_{pump} = \Delta T$. The 400 MHz APD is sufficiently quick to resolve individual pulses, and the signal can then be demodulated at 40 MHz and the second harmonic at 80 MHz to provide T_{pump} and T_{probe}, respectively, as shown in Figure 3.7. The probe and pump signals have SNRs on the order of 20 dB and 15 dB, respectively, for the device and laser parameters used here.

3.4.2 Effect of pump fluence on effective index

Ultrafast optical pump pulses have been shown in previous work to reduce the effective refractive index, n_{eff}, of an SOI waveguide by as much as 0.5, without causing visible damage to the device. The magnitude of effective index shifts was experimentally verified through observing the shifting of fringes in the transmission spectra of an unbalanced Mach-Zehnder Interferometer (MZI) for varying pump fluence on one of the arms [108]. This change in refractive index is attributed to free-carrier excitation in the silicon (see Section 2.3.1), with electron-hole densities in the region of $10^{20}\,\text{cm}^{-3}$ to $10^{21}\,\text{cm}^{-3}$ [107]. Appendix B.1 further investigates the scattering and absorption effects on the perturbation in a single-mode silicon waveguide via simulation.

3.4.3 Characterising the pump spot size

To accurately determine the pump fluence, and thus the effective index shift, it is first important to characterise the pump spot size. This was done by raster scanning over a

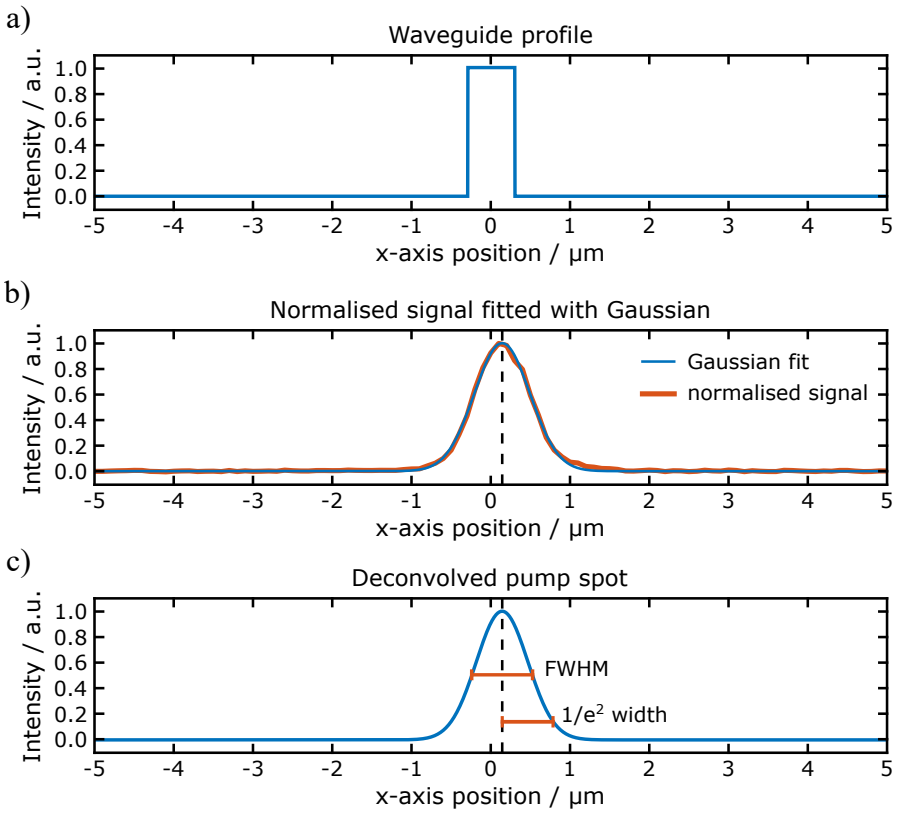


Figure 3.8: Example of the deconvolution method used to calculate the pump spot size. a) 500 nm wide ridge waveguide profile. b) The normalised $\Delta T/T$ signal fitted with a Gaussian. c) The recovered perturbation profile from the deconvolution of the Gaussian fit with the waveguide profile.

100 µm long section of 500 nm wide straight waveguide. To recover a reasonable estimate of the original pump spot size, a Gaussian is convolved with the waveguide profile and fitted to each slice of the experimental photomodulation signal normal to the propagation axis, as demonstrated in Figure 3.8.

The fit with the minimum Mean Squared Error (MSE) is then taken to be the spot's width for that position along the propagation axis. Figure 3.9 shows the Full-Width Half-Maximum (FWHM) of the Gaussian pump spot profile carried out in 1 µm intervals over the 100 µm length of the straight waveguide. The mean spot size at FWHM was found to be 740 nm, which corresponds to a $1/e^2$ radius of $w = 630$ nm. There is good focus stability over this range, as indicated by the standard deviation of 40 nm, and sample tilt is minimal as observed over the device's length. The theoretical diffraction-limited spot size, w_0, can be calculated by

$$w_0 = M^2 \frac{\lambda}{\pi\theta} = M^2 \frac{\lambda}{\pi \sin^{-1}\left(\frac{\mathrm{NA}}{n}\right)}, \tag{3.1}$$

where M^2 is the beam quality factor, λ is the wavelength of light and θ is the half-angle beam divergence, which is related to the NA of the focusing objective by $\mathrm{NA} = n \sin\theta$.

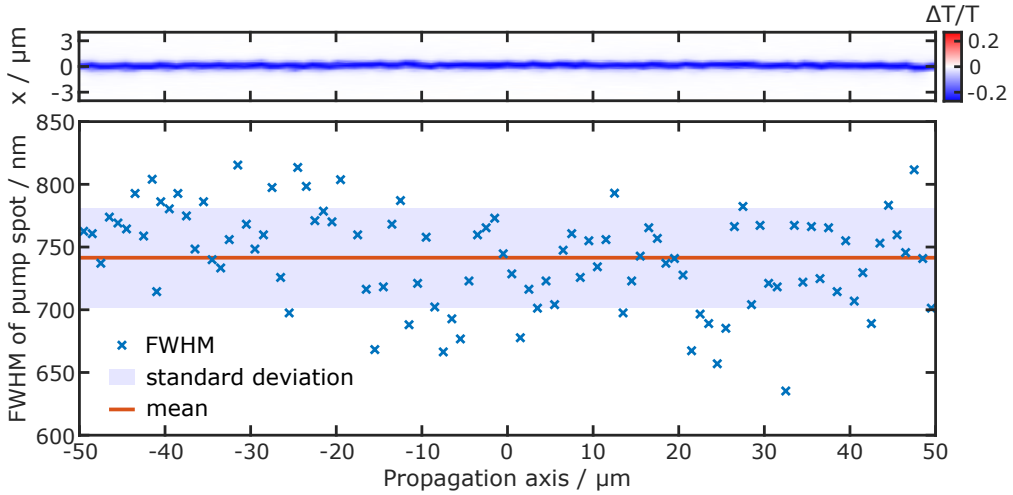


Figure 3.9: The mean pump spot size found using the deconvolution method. (top) Spatial photomodulation map of differential transmission for a 100 µm long section of 500 nm wide straight rib waveguide in 1 µm steps. (bottom) Calculated FWHM of perturbation of each x-axis slice along the propagation axis. A mean perturbation size of 740 nm with a standard deviation of 40 nm was recovered.

The objective has a NA of 0.55, which at a pump wavelength of 415 nm with an M^2 of 1.1 (taken from laser's specification sheet) corresponds to a diffraction limit of $w_0 = 274\,\text{nm}$ in air. This theoretical minimum spot size is significantly smaller than the measured value. This is attributed to the refractive index perturbation being slightly larger than the pump spot size, as the free-carriers that are created can diffuse short distances inside the material depending on the pump-probe delay time. These ballistic free-carriers have a velocity on the order of $1 \times 10^5\,\text{m}\,\text{s}^{-1}$ to $1 \times 10^6\,\text{m}\,\text{s}^{-1}$, corresponding to a few 100 nm drift per picosecond [109]. Also, some beam quality deterioration, increasing the M^2 value, will be caused by the number of optics in the pump's path and the inherently imperfect nature of optics [110, 111]. The incident pump fluence, F, can now be calculated by

$$F = \frac{P_{ave}}{f_{rep}} \cdot \frac{1}{A} = \frac{P_{ave}}{f_{rep}} \cdot \frac{1}{\pi w^2}, \tag{3.2}$$

where P_{ave} is the average pump power after the focusing objective, f_{rep} is the laser repetition rate and A is the cross-sectional area of the beam.

In this work, the pump power was typically set such that $\Delta n_{\text{eff}} = -0.25$, by using the variable ND filter to limit the pump fluence to $\sim 60\,\text{pJ}\,\mu\text{m}^{-2}$. This free-carrier concentration corresponds to an increase of the extinction coefficient of $\Delta\kappa = 2.7 \times 10^{-4}$ as calculated from Equation 2.46 and Equation 2.47, which is shown to be negligible over the <1 µm perturbation scales used here in Appendix B.1.

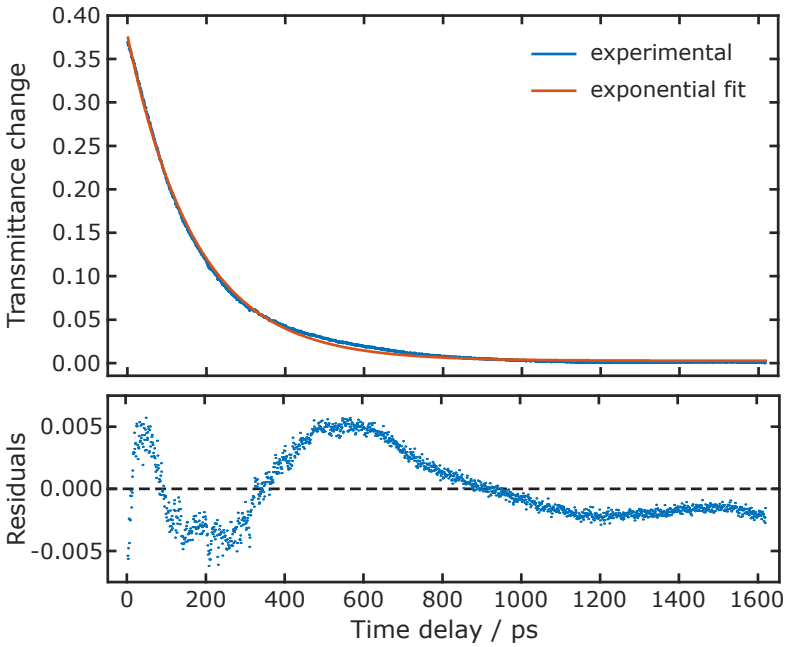


Figure 3.10: The ultrafast recovery response of a straight 500 nm rib waveguide with an exponential decay fit for a pump fluence of 75 pJ µm^{-2}. The bottom graph shows the residuals of the fit and returns a time constant of $\tau = 173 \pm 1$ ps.

3.4.4 Silicon free-carrier lifetime

Once again, a 500 nm wide straight waveguide was investigated to obtain the pump-probe response and thus the recovery time of the silicon. The pump spot was positioned at the midpoint of the waveguide, and the focus adjusted such that the change in transmission was maximised. The transmission change was then recorded as a function of delay time for the full range of the variable delay stage, as described in Section 3.3.4. The fitting of an exponential decay, $|\Delta T/T| = A \cdot e^{-t/\tau}$, to the pump-probe response was used to calculate a recovery time constant of $\tau = 173 \pm 1$ ps (the error is taken from the fit), as shown in Figure 3.10. The residuals from the fit clearly show there are some higher-order components; however, the exponential fit is sufficient for extracting the time constant of the recovery. The value obtained is also in excellent agreement with the value of $\tau = 173.5 \pm 0.7$ ps reported in previous work for SOI waveguide structures [112].

Therefore, the silicon's full recovery is sub-nanosecond, which is substantially faster than the temporal separation of the pulses (12.5 ns for 80 MHz repetition rate) and, as such, each pulse can be considered fully isolated from one another for free-carrier effects. The vast majority of experiments carried out in this work was with the delay stage set to a 3 ps time delay, thereby allowing some positional leeway of the pump spot, as its position along the device will also have a slight effect on the delay time between the pulses related to the group index of the light. The full range of the delay stage can be used to measure devices' temporal response over a range of ∼4 ns with a resolution of ∼30 fs. As this approach simply relies on the constant speed of light, it follows that

longer scan time ranges can be achieved by increasing the number of passes on the delay stage or increasing the stage's travel length. Similarly, the temporal resolution could be reduced using fewer passes of the delay path or a stage with a smaller minimum step size. In both cases, the change in time, Δt, is given simply by

$$\Delta t = \frac{N \cdot \Delta L}{c}, \tag{3.3}$$

where N is the number of delay stage passes, ΔL is the single-pass change in distance of the delay stage and c is the speed of light.

3.4.5 Summary

This chapter has covered the general fabrication, modelling and measurement techniques that will be used throughout the work of this thesis. In addition, this chapter introduced the experimental technique of UPMS and characterised its key performance parameters, such as refractive index modulation, free-carrier lifetime and spatial resolution. Effective index modulations of up to $\Delta n_{\text{eff}} = -0.5$ were found to be obtained over regions of as small 740 nm (at FWHM), without any visible damage to the silicon itself. These perturbations in the refractive index profile of the silicon are not only highly localised spatially but also temporally, with a measured free-carrier lifetime of sub-nanosecond before the structure returns to its unperturbed state. The work in this chapter will be referred back to multiple times throughout the following chapters and forms the foundation of this thesis as a whole.

Chapter 4

Ultrafast Photomodulation Spectroscopy for Characterisation of Photonic Integrated Circuits

The industrialisation of Photonic Integrated Circuits (PICs) as a high-volume, low-cost technology requires new optical testing approaches and diagnostics at the wafer-scale [113, 114]. Compared to the maturity of nanoelectronics' testing, available techniques for testing optical chips at the commercial level are very limited. Standard optical transmission measurements can probe between all input and output ports and can be automated using wafer-scale testing station [115–121]. However, such an approach on its own does not allow access to the performance of individual elements in a complex circuit of many cascaded components. Furthermore, transmission measurements of an entire circuit provide very limited diagnostic capabilities.

One proposed solution is the introduction of erasable ion-implanted elements that allow the transmission of selected components to be measured and/or tuned before the system is fixed into its final state [25]. The temporary coupling point can be achieved via direct or indirect means. The direct approach involves the writing of a grating into the silicon waveguide itself and requires waveguides to taper up to a fibre Mode Field Diameter (MFD) ($\sim$10 µm) which introduces some additional excess loss to the system even after annealing [122, 123]. Another more recent approach utilises the empty space on a chip that arises from their typically linear design and minimal use of bends [26]. This enables the chip to have permanent grating coupling positions which are then themselves coupled to the component of interest by erasable Directional Couplers (DCs). While this approach enables very low loss post-annealing, it does have some drawbacks for rapid high-volume testing which render it perhaps more suitable for the post-fabrication tuning of complex circuits. The primary drawbacks are it is time-consuming and requires extra fabrication steps to implant and anneal. It also still only provides access to transmission

information and not any detail on the electric fields within, which limits its diagnostic capabilities.

Optical Time Domain Reflectometry (OTDR) and Optical Frequency Domain Reflectometry (OFDR) are standard techniques employed in distributed vibration sensing that rely on the very weak Rayleigh backscatter to convert an entire optical fibre into a very sensitive high-resolution sensor [124]. Coherent OFDR has also been demonstrated in PICs, in which the inherent backscatter signal is used to probe the device under test [125]. The spatial resolution demonstrated by Glombitza et al. was only on the order of 50 µm. However, the system is theoretically only limited by the interrogator, and therefore smaller spatial resolution should be achievable by using narrower linewidth sources swept over a larger frequency range or advanced pulse compression techniques.

The most complete characterisation of a PIC is given by the electromagnetic field distribution of the light propagating within it, which is typically obtained by techniques that rely on the use of scanning perturbations to infer information on the electromagnetic field within [126]. Typically, these techniques require direct access to the near-field, such as in Near-field Scanning Optical Microscopy (NSOM) or scanning Atomic Force Microscopy (AFM) [127–131], which is not possible with the majority of real-world devices due to protective cladding layers. In addition to this, the complexity of the nanoprobes involved and the scanning times makes both NSOM and AFM less desirable in an industrial environment. Fluorescence imaging of the light propagating inside a silicon nitride Multimode Interference (MMI) has been demonstrated for visible wavelengths via the application of water containing fluorescent dyes [132]. This approach still requires access to the near-field, making it incompatible with clad devices, and resolution is limited to the imaging system. Other techniques have been demonstrated that utilise a far-field optically induced perturbation, such as in a study by Lian et al. in which a spatial light modulator was used to scan a thermo-optic perturbation over a photonic crystal waveguide to observe spectral shifts in their resonances and reconstructed the mode profile [133]. However, the slow response of thermal effects results in long dwell times and thermal diffusion limiting the perturbation spot size, which are both undesirable for component testing. The study was also limited to resonant structures and is unlikely to work as well for non-resonant devices. Finally, adaptive 3D multimodal microscopy has been demonstrated for far-field optical component-wise testing of the beam-splitting ratio of coupled-mode integrated beam splitters [134].

Optical pump-probe techniques have been demonstrated as an effective method for the spatial and temporal visualisation of light transport in disordered media and photonic crystals [135–139]. Previous work has shown Ultrafast Photomodulation Spectroscopy (UPMS) to be a viable far-field technique for all-optical non-destructive characterisation of the flow of light in PIC for a range of devices [97, 108, 140]. In particular interest here, Bruck et al. utilised spatial ultrafast photomodulation mapping to investigate the mode structure of a 1×2 MMI power splitter [108].

In this chapter, we demonstrate photomodulation mapping with increased spatial resolution maps, higher Signal-to-Noise Ratio (SNR) and significantly faster device scan times. Furthermore, we develop the technique from a purely qualitative one to a quantitative one through the use of an analytical model developed by a collaborating research group. We start by introducing the model that can be used to predict the effect of a refractive index perturbation for an arbitrary integrated photonic device. This technique is then used to quantitatively characterise the performance of increasingly complex multiple-input multiple-output devices as fabricated. Finally, we show that under the right circumstances it is possible to infer information on the light intensity fields within PICs to the spatial resolution of the perturbation size — even in the presence of protective cladding layers.

4.1 Formulation of the analytical model

To calculate the theoretical perturbation maps an analytical method that exploits Lorentz reciprocity was developed by a collaborating group, the Light in Complex Nanostructures Group from the Laboratory for Photonics, Numerics and Nanosciences (LP2N), Bordeaux, France [98]. The goal of forming an analytical model is to develop a general and rigorous approach that allows the prediction of transmittance perturbation maps for arbitrary linear multiport photonic devices with great accuracy and minimal computational cost. This is achieved by exploiting the Lorentz reciprocity theorem applied to normalised waveguide modes, and allows the calculation of spatial perturbation maps by computing the response of the unperturbed system for only two independently excited ports; one forward and one backward excitation. In contrast, numerical brute-force approaches require the fields to be calculated for the number of perturbation positions considered plus the unperturbed case, which results in quadratic scaling with the resolution of the map leading to rapidly untenable computational times.

An illustration of the problem is shown in Figure 4.1 for a device with M input and N output modes. The values of M and N are the products of the number of ports and supported modes for the input and output, respectively. A specific mode is denoted by $\tilde{\mathbf{\Phi}}_i^{\pm}$, where the sign of the superscript indicates if the mode is incoming or outgoing from the system and i is the m-th input port or n-th output port mode. The unperturbed transmission coefficient for the coupling between $\tilde{\mathbf{\Phi}}_m^{+}$ and $\tilde{\mathbf{\Phi}}_n^{-}$ is written as t_{mn}^0.

In the presence of a perturbation, the relative permittivity, $\boldsymbol{\epsilon}(\mathbf{r})$, can be represented as

$$\boldsymbol{\epsilon}(\mathbf{r}) = \boldsymbol{\epsilon}_\mathrm{b} + \Delta\boldsymbol{\epsilon}(\mathbf{r}), \tag{4.1}$$

where $\boldsymbol{\epsilon}_\mathrm{b}$ is the background permittivity, $\Delta\boldsymbol{\epsilon}(\boldsymbol{r})$ is the permittivity change due to the perturbation and is related to the refractive index by $\epsilon = \epsilon_0 n^2$. This results in a

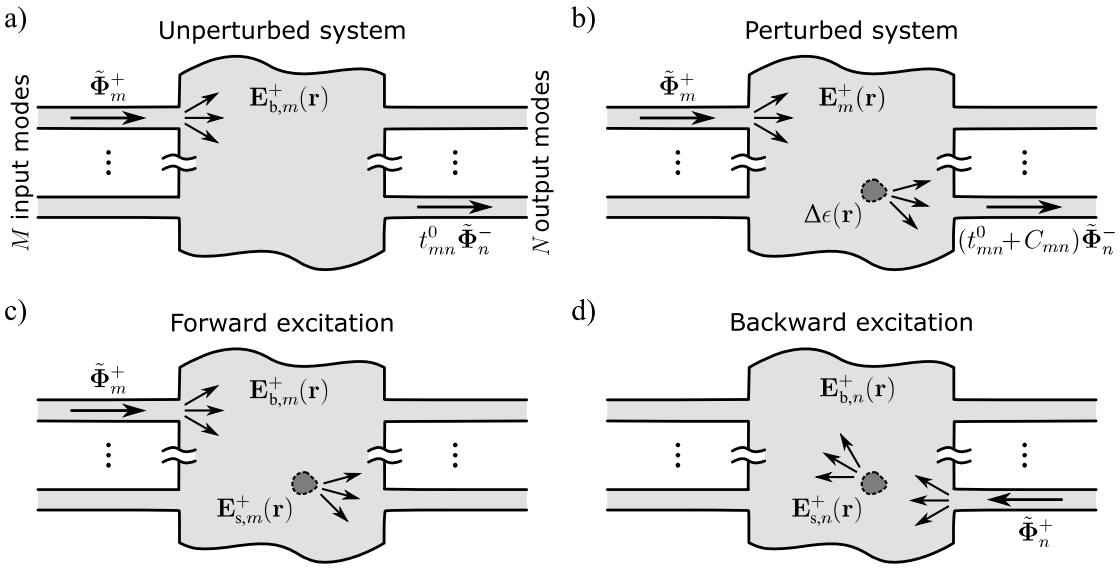


Figure 4.1: Illustration of the analytical method for an arbitrary multiport photonic device. a) The unperturbed system with M inputs and N outputs. b) The perturbed system for an arbitrary perturbation with permittivity change $\Delta\epsilon$. c) and d) the forward and backward excitations of the system, respectively.

perturbed transmission coefficient of

$$t_{mn} = t^0_{mn} + C_{mn}, \tag{4.2}$$

where C_{mn} is the coupling coefficient between the m-th incoming mode and the n-th outgoing mode caused by the scatterer $\Delta\boldsymbol{\epsilon}(\boldsymbol{r})$. The normalised transmission change, which can be experimentally measured, can be written in terms of transmission coefficients as

$$\frac{\Delta T}{T} = \frac{|t_{mn}|^2}{|t^0_{mn}|^2} - 1 = \frac{|C_{mn}|^2}{|t^0_{mn}|^2} + 2\mathrm{Re}\left[\frac{C_{mn}}{t^0_{mn}}\right]. \tag{4.3}$$

It is therefore C_{mn} that must be represented analytically at an arbitrary position in terms of the background and scattered electric fields, $\mathbf{E}^+_{b,i}$ and $\mathbf{E}^+_{s,i}$ respectively, for both the forward ($i = m$) and backward ($i = n$) excitations, as shown in Figure 4.1 c) and d).

The perturbed system's total electric field is simply the linear sum of the background and scattered fields, $\mathbf{E}^+_i = \mathbf{E}^+_{b,i} + \mathbf{E}^+_{s,i}$, and can be described by conventional $\exp(-i\omega t)$ notation, where ω is the angular frequency and t is time. By solving Maxwell's equations (Section 2.1) for the electric fields, $\mathbf{E}$, in terms of the current density, $\mathbf{J}$, we obtain the vector wave equation, which is given by

$$\nabla \times \nabla \times \mathbf{E}^+_i(\mathbf{r}) - k_0^2 \boldsymbol{\epsilon}(\mathbf{r}) \mathbf{E}^+_i(\mathbf{r}) = i\omega\mu_0 \mathbf{J}^+_i(\mathbf{r}), \tag{4.4}$$

where $\mathbf{J}^+_i$ is the current density source produced by $\tilde{\boldsymbol{\Phi}}^+_i$, μ_0 is the permeability of free-space and k_0 is the wave-vector of the incident light, which is related to the angular

frequency and speed of light by, $k_0 = \omega/c$. By breaking the total electric field into its background and scattering field components and subbing Equation 4.1 into Equation 4.4, we can form the scattered field relation [141]. This can be written as

$$\nabla \times \nabla \times \mathbf{E}^{+}_{\mathrm{s},i}(\mathbf{r}) - k_0^2 \boldsymbol{\epsilon}_{\mathrm{b}} \mathbf{E}^{+}_{\mathrm{s},i}(\mathbf{r}) = k_0^2 \Delta\boldsymbol{\epsilon}(\mathbf{r}) \mathbf{E}^{+}_{i}(\mathbf{r}). \tag{4.5}$$

Using the relation $c = \left(\sqrt{\mu_0 \epsilon_0}\right)^{-1}$, the perturbation $\boldsymbol{\epsilon}(\mathbf{r})$ can be considered as a source for the scattered field in the unperturbed system with a current density of

$$\mathbf{J}^{+}_{\mathrm{s},i}(\mathbf{r}) = -i\omega\epsilon_0 \Delta\boldsymbol{\epsilon}(\mathbf{r}) \mathbf{E}^{+}_{i}(\mathbf{r}). \tag{4.6}$$

This equation describes how the excitation of the permittivity change from an incoming field gives rise to the source of the scattered field, as shown by Figure 4.1 c) and d). We must now calculate the coupling efficiency between the scattered field and an outgoing mode $\tilde{\boldsymbol{\Phi}}^{-}_{i}$, which we do through the exploitation of the Lorentz reciprocity theorem in waveguide geometry. Providing the materials that form the waveguide are reciprocal, a relation between the current density sources and the fields produced in an arbitrary system can be established [142]. On the condition that the modes of the input and output waveguides $\tilde{\boldsymbol{\Phi}}$ form a complete set, that is to say, an arbitrary propagating field in the waveguides can be written as a linear combination of both guided and radiating waveguide modes, one finds that the excitation amplitude of an outgoing mode by a current source is given by [143]

$$\begin{aligned} C_{mn} &= -\frac{1}{4} \int \mathbf{J}^{+}_{\mathrm{s},m}(\mathbf{r}) \cdot \mathbf{E}^{+}_{n}(\mathbf{r}) d\mathbf{r}, \\ &= \frac{i\omega\epsilon_0}{4} \int \Delta\boldsymbol{\epsilon}(\mathbf{r}) \mathbf{E}^{+}_{m}(\mathbf{r}) \cdot \mathbf{E}^{+}_{n}(\mathbf{r}) d\mathbf{r}, \end{aligned} \tag{4.7}$$

where the 1/4 factor results from mode normalisation with unitary power flux and Equation 4.6 is substituted in for $\mathbf{J}^{+}_{\mathrm{s},m}(\mathbf{r})$ with input mode $i = m$. In this equation, C_{mn} is described by the overlap integral between the total forward and backwards propagating fields over the perturbation.

Equation 4.7 as it stands is exact and applies to any arbitrary multiport photonic device and any arbitrary refractive index perturbation (as shown in Figure 4.1), providing the materials involved are reciprocal. However, it is currently represented in terms of the total electric fields that are still dependant on the position of the perturbation and, therefore, it would be more beneficial to describe it in terms of the background fields, which are independent of the perturbation, with a fixed perturbation variation. For small ($\ll \lambda$) perturbations, it is possible to model the scatterers as small cylinders where the local field correction can be calculated from the dipole moment of the perturbation itself. This dipolar approximation has been shown to be in excellent quantitative agreement for $\ll \lambda$ scatterers when compared to exact numerical calculations [98].

4.1.1 Wavelength-scale perturbations

For the case of UPMS, the perturbations are on the same scale as the probing wavelength and their magnitude is dictated by the incident pump power. Consequently, these perturbations do not behave as electric dipoles (higher-order multipoles will also be excited), as such the dipolar approximation is no longer valid. Instead, another approach to model the effect of the perturbations must be used to calculate the total electric fields from the background fields, as required by Equation 4.7 to obtain a spatially invariant formula for C_{mn}.

A simple 1D Fabry-Peŕot was decided upon to model the wavelength-scale perturbations, which is based upon two approximations; the propagation is paraxial and the perturbations can be considered as squares with a fixed index shift. The scatters in Figure 4.1 are now considered as these 1D Fabry-Peŕot cavities with a length equal to perturbation experimentally induced by the optical pump (see Figure 3.9). To the first order, the transmission and reflection amplitude coefficients of the interface between the background medium and the cavity, t_{12} and r_{21}, respectively, are given by the Fresnel equations at normal incidence [28]

$$\mathrm{t}_{12} = 2n_\mathrm{b}/\left(2n_\mathrm{b} + \Delta n\right), \tag{4.8}$$

$$\mathrm{r}_{21} = \Delta n/\left(2n_\mathrm{b} + \Delta n\right), \tag{4.9}$$

where n_b is the background refractive index and Δn is the index change caused by the perturbation. Considering a 1D cavity of length d, the change to the forwards and backwards propagating electric fields can be obtained from the equations for layered media with normal incident light [28]. This allows the total electric fields to be represented exclusively in terms of the perturbation itself and the background electric fields

$$\mathbf{E}_m^+(\mathbf{r}) = \frac{\mathrm{t}_{12}\exp\left[i\omega\Delta n(z-(z_0-d/2))/c\right]}{1-\mathrm{r}_{21}^2\exp[i\omega(n_b+\Delta n)d/c]}\mathbf{E}_{\mathrm{b},m}^+(\mathbf{r}), \tag{4.10}$$

and

$$\mathbf{E}_n^+(\mathbf{r}) = \frac{\mathrm{t}_{12}\exp\left[-i\omega\Delta n(z-(z_0+d/2))/c\right]}{1-\mathrm{r}_{21}^2\exp[i\omega(n_b+\Delta n)d/c]}\mathbf{E}_{\mathrm{b},n}^+(\mathbf{r}), \tag{4.11}$$

where z is the axial position and z_0 is the central position of the perturbation. These two equations can then be substituted into Equation 4.7 to obtain

$$C_{mn} = \frac{i\omega\epsilon_0}{4}\frac{\mathrm{t}_{12}^2\exp\left[i\omega\Delta nd/c\right]}{\left(1-\mathrm{r}_{21}^2\exp[i\omega(n_b+\Delta n)d/c]\right)^2}\Delta\epsilon\int_{S_\mathrm{p}}\mathbf{E}_{\mathrm{b},m}^+(\mathbf{r})\cdot\mathbf{E}_{\mathrm{b},n}^+(\mathbf{r})d\mathbf{r}. \tag{4.12}$$

The equation for C_{mn} is no longer dependant on the total electric fields but still does require an integration of the background fields over the surface of the perturbation, S_p. In the case of low index contrast perturbations, $\Delta n/n_\mathrm{b} \ll 1$, it is the dephasing term,

$\exp[i\omega\Delta nd/c]$, which is dominant and Equation 4.12 simplifies to

$$C_{mn} = \frac{i\omega\epsilon_0}{4}\exp\left[i\frac{\omega}{c}\Delta nd\right]\Delta\epsilon\int_{S_\mathrm{p}}\mathbf{E}^+_{\mathrm{b},m}(\mathbf{r})\cdot\mathbf{E}^+_{\mathrm{b},n}(\mathbf{r})d\mathbf{r}. \tag{4.13}$$

The variation of the total transmission can now be calculated from Equation 4.3 and plotted as a function of perturbation's central position to give theoretical perturbation maps which can be directly compared to the experimentally obtained photomodulation maps.

4.1.2 Field intensity recovery

Intuitively, the magnitude of the transmission change will depend on the intensity of light at any given position, so one would expect the perturbation map to be related to the electric field distributions. Equation 4.7 and Equation 4.13 show that transmittance perturbation maps contain information on the background electric fields. Indeed, in the limits of small perturbations to the transmission, $|C_{mn}| \ll |t^0_{mn}|$, and near-unity transmission, $|t^0_{mn}| \sim 1$, it is possible to show that $\Delta T/T$ is directly proportional to the forward propagating background light intensity field when integrated over the perturbation surface area. Under the first assumption, Equation 4.3 to the first-order reduces to

$$\frac{\Delta T}{T} \approx 2\mathrm{Re}\left[\frac{C_{mn}}{t^0_{mn}}\right], \tag{4.14}$$

and from here we can substitute in C_{mn} for large low-index contrast perturbations from Equation 4.13, to obtain

$$\frac{\Delta T}{T} \approx 2\mathrm{Re}\left[\frac{1}{t^0_{mn}}\frac{i\omega\epsilon_0}{4}\exp\left[i\frac{\omega}{c}\Delta nd\right]\Delta\epsilon\int_{S_\mathrm{p}}\mathbf{E}^+_{\mathrm{b},m}(\mathbf{r})\cdot\mathbf{E}^+_{\mathrm{b},n}(\mathbf{r})d\mathbf{r}\right]. \tag{4.15}$$

Under the second assumption, of near unity unperturbed transmission, Lorentz reciprocity states that the inverse propagating field is just the complex conjugate of the forward propagating electric field, i.e. $\mathbf{E}^+_{\mathrm{b},n}(\mathbf{r}) = \left[\mathbf{E}^+_{\mathrm{b},m}(\mathbf{r})\right]^\star$. This can be understood from the conservation of energy in the time-reversed case. Equation 4.15 then becomes

$$\frac{\Delta T}{T} \approx \frac{\omega\epsilon_0}{2}\mathrm{Re}\left[\frac{i\exp[i\omega\Delta nd/c]\,\Delta\epsilon}{\exp[i\phi_t]}\right]\int_{S_\mathrm{p}}|\mathbf{E}^+_{\mathrm{b},m}(\mathbf{r})|^2 d\mathbf{r}, \tag{4.16}$$

where $\phi_t = \arg(t^0_{mn})$. From this equation, it can be seen that under that $\Delta T/T$ is directly proportional to the light field intensity integrated over the perturbation surface area S_p under the previously mentioned assumptions.

4.2 Experimental setup

A photograph and schematic of the experimental setup used is provided in Figure 4.2 and is described in more detail in Section 3.4. A Coherent Chameleon Ultra II Ti:Sapphire laser is used in conjunction with an Optical Parametric Oscillator (OPO) to provide variable wavelength probe pulses, while the second harmonic generation of the seed produces the pump pulses. Both of these pulses stem from the same initial seed pulse and therefore have the same pulse duration of 200 fs with a repetition rate of 80 MHz. The experiments carried out in this chapter were before the addition of the pulse select and, as such, optical choppers were instead used to modulate the two beams at different frequencies.

The probe pulses are set to a central wavelength of 1550 nm with a bandwidth of 20 nm, and free-space coupled into an optical fibre; allowing the probe to then be grating coupled into the device under test. A second fibre is then coupled to the desired output port, also via a grating coupler, and collects the transmitted light. The transmission is measured using an InGaAs APD connected to a lock-in amplifier demodulated at the frequencies of the optical choppers.

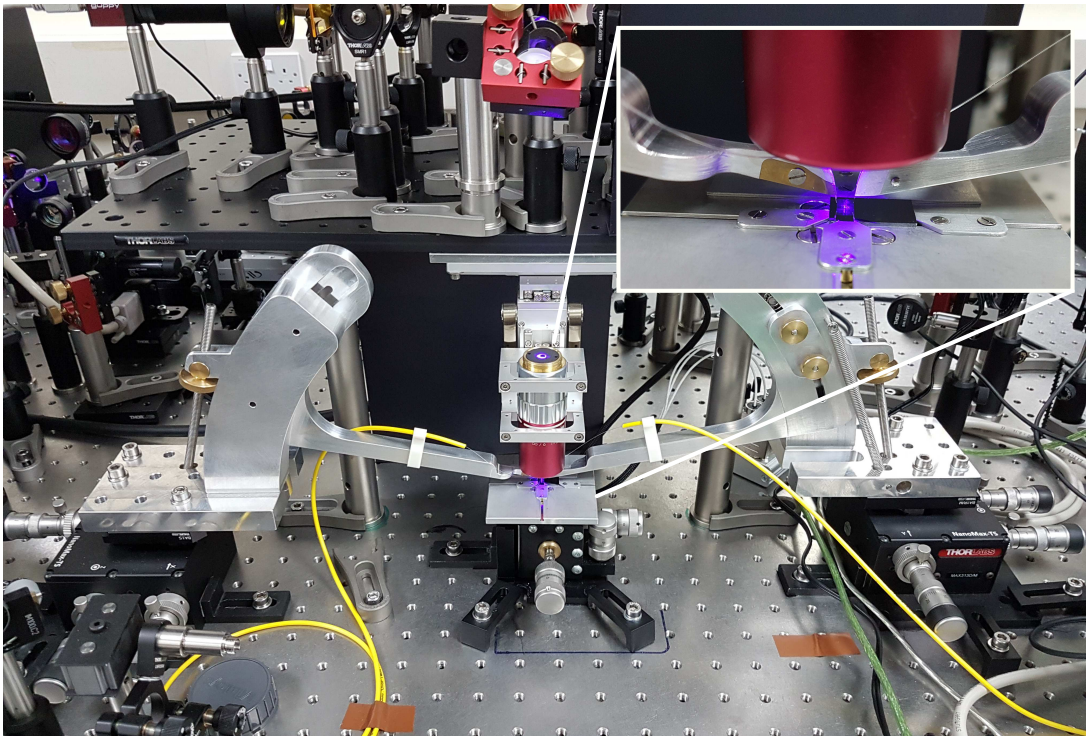
Meanwhile, the pump pulses are collimated and passed through a four-pass variable delay stage, allowing fine control of the time delay between the two pulses. A variable neutral density filter is used to limit the pump power and set the desired effective index modulation. The pulses are then focused onto the device's surface using a 100× objective with a Numerical Aperture (NA) of 0.5. The beam slightly overfills the focusing objective which is mounted to a 3D nanopositioner, providing accurate positioning and focusing of the pump spot with minimum fluence deviation.

The pump's focusing objective is also used to image the device with either an IR or a visible light camera. The visible light camera is used for the approximate alignment of fibres and positioning of the pump spot, and the IR camera is used to image any out-of-plane scattered probe light.

4.2.1 Method

To method used to obtain the normalised transmission change, $\Delta T/T$, is outlined in Section 3.4.1. The data in this section was taken before the addition of the pulse select system. Instead, both beams were modulated by optical choppers set to different frequencies. The signal that is chopped with reference to the probe beam therefore records the overall transmission, $T_{probe} = (T + \Delta T)/2$, where T is the unperturbed transmission, ΔT is the change in transmission and the factor of 2 originates from half of the light being blocked by the optical chopper's blades. The other signal that is chopped with reference to the pump beam records only the change in transmission,

a) Photograph of the experimental setup



b) Schematic of the experimental setup

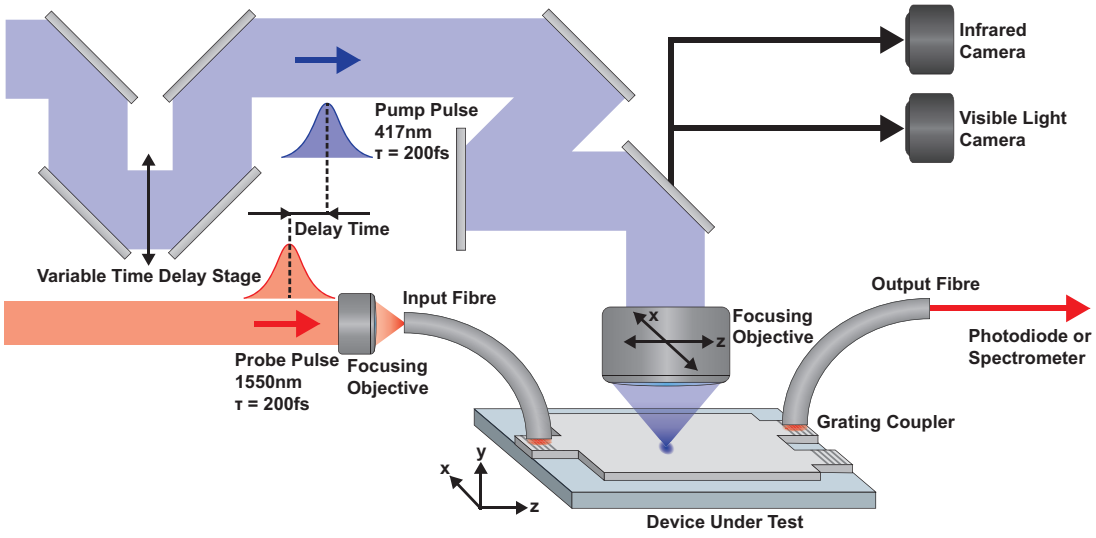


Figure 4.2: a) Photograph and b) schematic of the UPMS experimental setup. Infrared probe pulses are grating-coupled into the waveguide, travel through the device and then out-coupled by another grating. The transmitted light is recorded by either an InGaAs Avalanche Photodetector (APD) connected to a lock-in amplifier or a spectrometer. The pump pulses are incident perpendicular to the device's surface, and the delay time between the two pulses is controlled via a variable time delay stage. The pump focusing objective is mounted to a 3D nanopositioner, allowing the pump spot to be focused and scanned over the device. The focusing objective is also used for imaging in both visible and near-Infrared (IR) light.

$T_{pump} = (\Delta T)/2$. A spatial transmission map is built up be recording these values as a function of optical pump position.

Once the experimental map is obtained, it can be compared to the model by the two-dimensional correlation coefficient, Γ, which is given by

$$\Gamma = \frac{\sum_m \sum_n \left(\mathbf{A}_{mn} - \bar{\mathbf{A}}\right)\left(\mathbf{B}_{mn} - \bar{\mathbf{B}}\right)}{\sqrt{\left(\sum_m \sum_n \left(\mathbf{A}_{mn} - \bar{\mathbf{A}}\right)\right)^2 \left(\sum_m \sum_n \left(\mathbf{B}_{mn} - \bar{\mathbf{B}}\right)\right)^2}}, \tag{4.17}$$

where $\mathbf{A}$ and $\mathbf{B}$ are the $\Delta T/T$ matrices for the experimental and modelled photomodulation maps respectively, and $\bar{\mathbf{A}}$ and $\bar{\mathbf{B}}$ are the corresponding matrices filled with their mean values. This limits the correlation to only spatial and relative intensity variations.

This function returns a value between -1 and 1; with 1 representing two identical matrices, -1 two totally opposite matrices and 0 for no correlation. Any required rotation correction is applied to the experimental data, and an area of equal dimensions of the modelled MMI region is scanned over the experimental map. The degree of correlation is calculated for each possible position, thus accounting for any error in the initial zero position during experimental alignment. The maximum correlation is then recorded and should occur for the best overlap between the experiment and computational model.

4.2.2 Devices under test

The devices investigated in this chapter are Silicon-On-Insulator (SOI) MMI power splitters, which are a standard silicon photonics component that has been shown to be capable of achieving insertion losses of significantly less than 1 dB per device with very accurate splitting ratios [144]. They are frequently used as power splitters in more complex circuits due to the high localisation of light at the self-imaging points (see Section 2.2.4.2). Initially, a simple 50-50 (typically written as 1×2) device was investigated, then we gradually increased the complexity to multiple-input multiple-output devices. Simulations of the devices were calculated using the Lumerical Finite-Difference Time-Domain (FDTD) solver for the fundamental Transverse Electrical (TE) input mode at a wavelength of 1550 nm for varying MMI region width and length parameter space to find dimensions for the optimal splitting ratios. Figure 4.3 shows the |E| field distribution and port transmission for each unique input-output port combination for the 1×2, 1×4 and 3×3 MMI devices as designed, and the transmittance value for varying MMI region width and length.

The waveguides were then fabricated with a rib structure (120 nm rib thickness) from an SOI wafer, consisting of a 220 nm thick layer of silicon on top of a 2 µm thick isolation layer of silica. Scanning Electron Microscope (SEM) images of the devices as-fabricated are shown in Figure 4.4. The input and output waveguides are single-mode, 500 nm wide

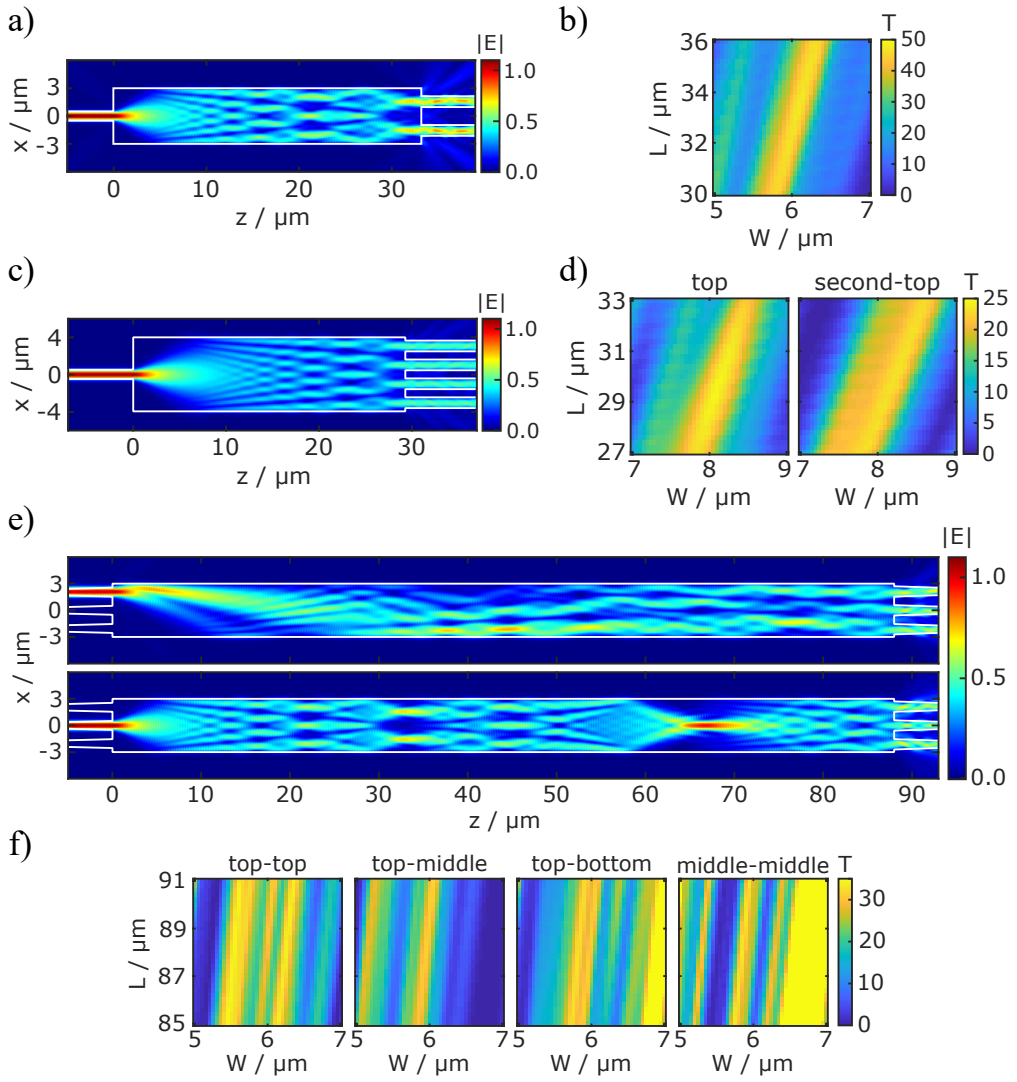


Figure 4.3: a) 1×2, c) 1×4 and e) 3×3 MMI forward propagating |E| fields for the fundamental TE input mode at $\lambda = 1550$ nm. b,d,f) show their corresponding percentage transmission for each unique output port in MMI region length, L, and width, W, parameter space. Simulated using Lumerical FDTD solver for TE input modes at a wavelength of 1550 nm.

and tapered to 1 µm wide at the MMI region boundary — reducing the loss of the device. The MMI region itself is typically several micrometres wide with varying lengths depending on the device's function. The 1×2 device was designed with a width and length of $W = 6.0$ µm and $L = 33.2$ µm, while the 1×4 and 3×3 had design dimensions of $W = 8.0$ µm, $L = 29.4$ µm and $W = 6.0$ µm, $L = 88.0$ µm respectively. Identical chips with either air or silica top cladding were produced. 500 nm wide straight waveguides were also added to the chip to provide information on coupling and propagation loss, as well as to be used to characterise the perturbation. Broadband swept transmission measurements were then carried out to assess the performance of the devices as fabricated, as outlined in Section 3.3.2, and insertion losses were confirmed to be within 1 dB of the design's target with minimal imbalance values of <0.5 dB.

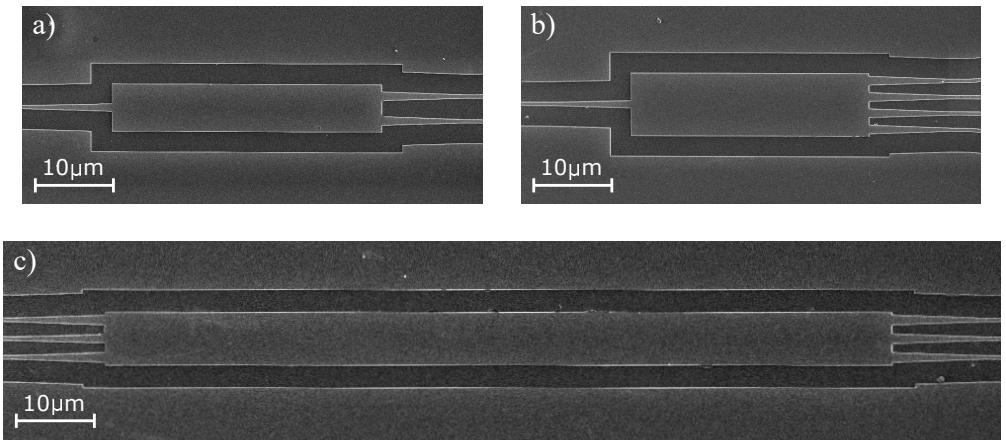


Figure 4.4: SEM images of the MMI devices under test. a) 1×2 MMI power splitter with measured dimensions of $W = 6.0\,\mu m$, $L = 33.8\,\mu m$. b) 1×4 MMI power splitter with measured dimensions of $W = 8.0\,\mu m$, $L = 30.2\,\mu m$. c) 3×3 MMI power splitter with measured dimensions of $W = 5.9\,\mu m$, $L = 88.0\,\mu m$.

4.3 Photomodulation mapping of multimode interference devices

Experimental perturbation maps were obtained using the technique outlined in Section 4.2.1. These maps reflect the spatial contributions of the device to the transmission of the device for a particular output. Where negative $\Delta T/T$ (blue) indicates regions of reduced transmission, with light being diverted away from the coupled output; directly revealing the path of light through the device to the coupled output. In contrast, positive $\Delta T/T$ (red) positions are areas of increased transmission, which correspond to redirecting light back towards the coupled output. This light could have either been coupled initially to a different output or lost through scattering out of the device. A greater magnitude of change in transmission coincides with a higher intensity of light at that location, as any perturbation will have a more significant overall effect on the transmission of the output. To test the capability of UPMS, a simple 1×2 MMI device was initially investigated and compared to previously published results [108]. The device complexity was then increased to a greater number of outputs, and ultimately multiple-input multiple-output devices (see Section 4.2.2). For each device, experimental photomodulation maps are directly compared to the modelled maps at the design dimensions for each input-output combination. The 2D correlation coefficients are then obtained against a library of theoretical perturbation maps in MMI region width and length parameter space centred around the design dimensions.

The 1×2 MMI device is designed to produce an equal split of intensity coupled to the two output ports, and therefore an ideal device should exhibit perfectly symmetrical maps about the x-axis due to device symmetry. Figure 4.5 a) shows the resulting perturbation maps for the 1×2 MMI device, with design dimensions of $L = 33.2$ µm and

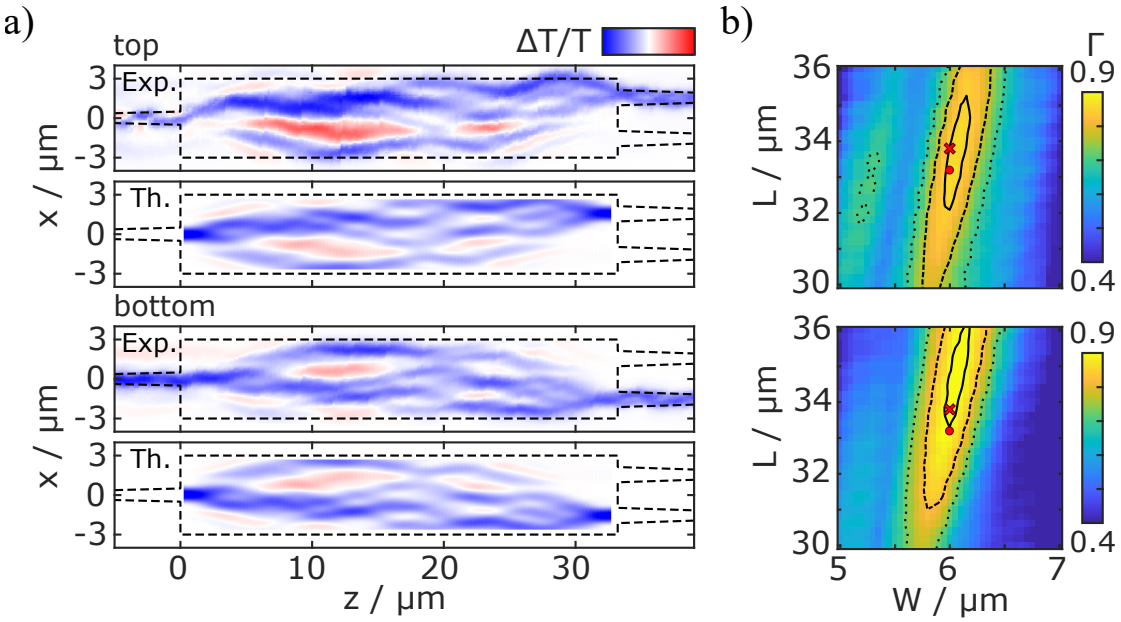


Figure 4.5: a) Experimental (top) and modelled (bottom) perturbation maps for both the top and bottom outputs of the 1×2 MMI device. The design parameters are $L = 33.2$ µm and $W = 6.0$ µm, with fabricated length $L = 33.8$ µm and width $W = 6.0$ µm as measured by SEM. The experimental scans were carried out for 1550 nm probe pulses and 417 nm pump pulses with a delay time of 3 ps. The pump was set to a fluence of 60 pJ µm^{-2} at the focus. The computed maps were obtained for a 740 nm perturbation, with $\Delta n_{\text{eff}} = -0.25$, for the design dimensions of the device. Colour bar range is between $\Delta T/T = \pm 0.25$ for experiment and $\Delta T/T = \pm 0.35$ for simulation. b) 2D correlation maps for varying theoretical MMI region dimensions L and W, for top and bottom outputs. The red dot and cross indicate the MMI design and SEM measured dimensions, respectively, and the contour lines indicate a 2% (solid), 10% (dash) and 20% (dot) decrease of the correlation with respect to the maximum correlation.

$W = 6.0$ µm. There is a very good positional and relative amplitude agreement between the experimental and calculated maps for the design dimension, with correlation coefficients, Γ, of 86% and 89% for the top and bottom outputs. The prominence of blue regions compared to red indicates that, despite being a 50-50 power splitter, a reduction in transmission is significantly more likely than an increase. This means that it is much easier to scatter light away from the coupled output than successfully redirect it towards another output. Asymmetry in the device is most visible in the first 10 µm of the MMI region, with the top output port coupling deviating slightly from the modelled response. It appears to be caused by the non-adiabatic expansion of the fundamental mode resulting in a zigzag-like pattern in the input taper. Interestingly, this is not observed for the bottom output coupling even though the device coupling remained unchanged between the experiments, which suggests the device fabrication itself causes the deviation. The modulation's magnitude is slightly weaker in the experimental maps than in the simulated ones. This is potentially due to an incorrect pump fluence value or

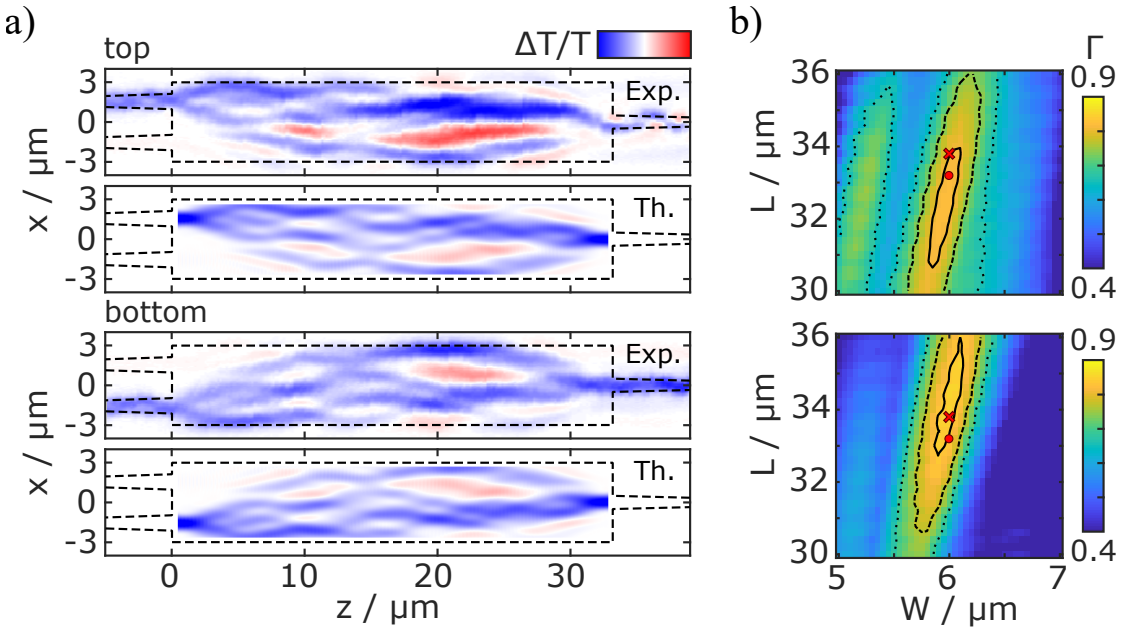


Figure 4.6: Experimental verification of reciprocity on a 1×2 MMI device. a) Experimental (top) and modelled (bottom) perturbation maps for the device as in Figure 4.5, but with the chip rotated by 180° such that light is incident from one of the two ports. The experimental scans were carried out for 1550 nm probe pulses and 417 nm pump pulses with a delay time of 3 ps. The pump was set to a fluence of 60 pJ μm^{-2} at the focus. The computed maps were obtained for a 740 nm perturbation, with $\Delta n_{\text{eff}} = -0.25$, for the design dimensions of the device. Colour bar range is between $\Delta T/T = \pm 0.25$ for experiment and $\Delta T/T = \pm 0.35$ for simulation. (b) 2D correlation maps for varying theoretical MMI region dimensions L and W, for top and bottom outputs. The red dot and cross indicate the MMI design and SEM measured dimensions, respectively, and the contour lines indicate a 2% (solid), 10% (dash) and 20% (dot) decrease of the correlation with respect to the maximum correlation.

the approximate values made in the model to reproduce the experiments. Figure 4.5 b) shows the correlation study in MMI width and length parameter space. The device is significantly more sensitive the width changes than length changes, as predicted from MMI self-imaging theory (see Section 2.2.4.2) where the required length for a particular splitting is proportional to the square of the width [56]. There is also a diagonal series of maxima corresponding to device dimensions with the same self-imaging points, i.e. shorter and thinner or longer and wider devices. The asymmetry seen in the experimental maps is also reflected in the difference in correlation strength for the two outputs, with the top coupling in slightly poorer agreement with the computed map. Optimal correlation values were found to be 87% and 91% for the top and bottom output couplings respectively, and the region of highest correlation is in good agreement with the design and SEM measured dimensions.

According to Equation 4.7, perturbation maps are symmetric upon reversing the direction of propagation of light through the device — by switching $\mathbf{E}_m^+(\mathbf{r})$ and $\mathbf{E}_n^+(\mathbf{r})$.

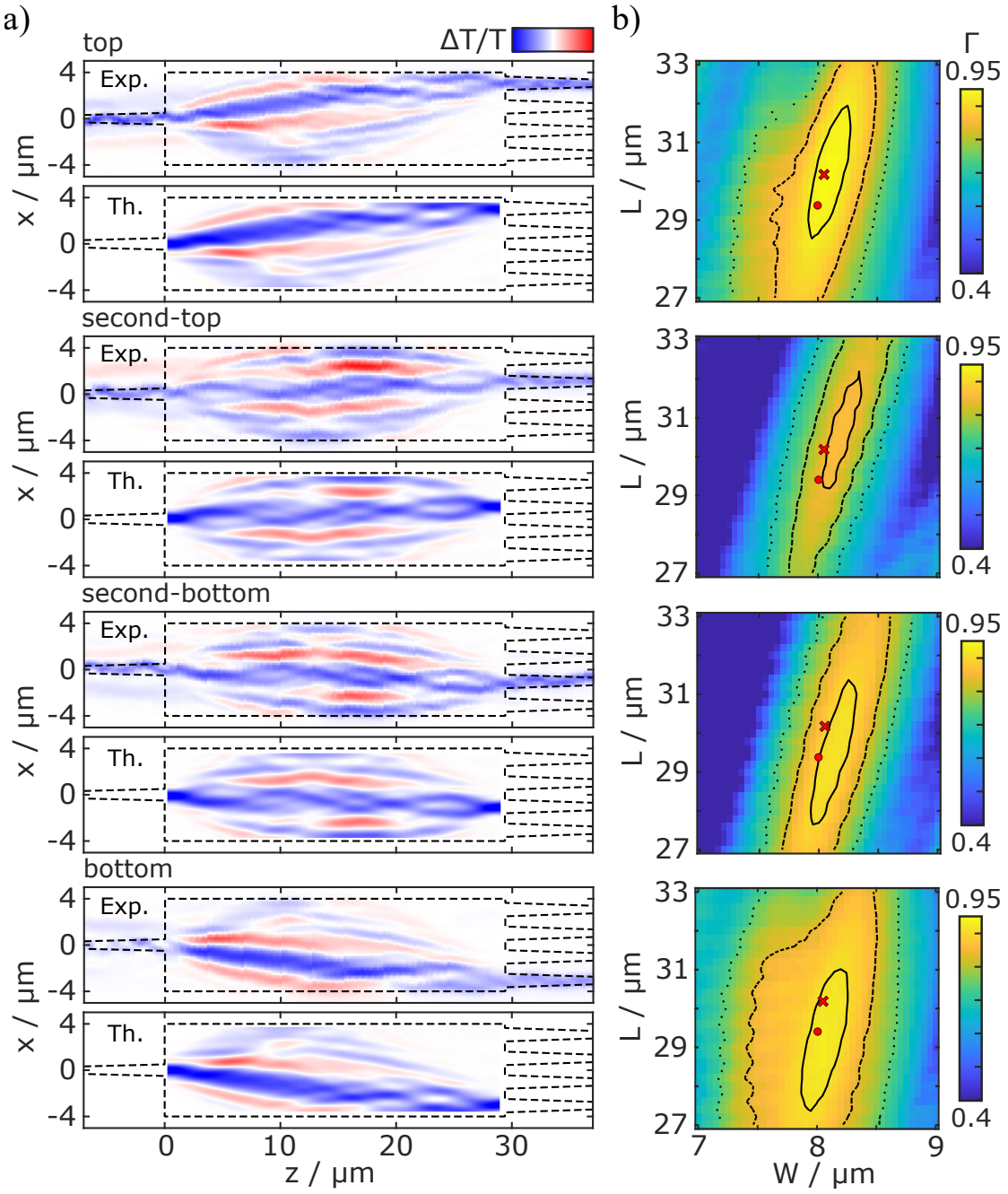


Figure 4.7: a) Experimental (top) and computed (bottom) perturbation maps for all four outputs of the 1×4 MMI device. The design parameters are $L = 29.4$ µm and $W = 8.0$ µm, with fabricated length $L = 30.2$ µm and width $W = 8.05$ µm as measured by SEM. The experimental scans were carried out for 1550 nm probe pulses and 417 nm pump pulses with a delay time of 3 ps. The pump was set to a fluence of 60 pJ µm^{-2} at the focus. The computed maps were obtained for a 740 nm perturbation, with $\Delta n_{\text{eff}} = -0.25$, for the design dimensions of the device. Colour bar range is between $\Delta T/T = \pm 0.25$ for experiment and $\Delta T/T = \pm 0.35$ for simulation. b) 2D correlation maps for varying theoretical MMI region dimensions L and W, for top and bottom outputs. The red dot and cross indicate the MMI design and SEM measured dimensions, respectively, and the contour lines indicate a 2% (solid), 10% (dash) and 20% (dot) decrease of the correlation with respect to the maximum correlation.

Indeed, near-identical experimental maps are obtained when light is launched into one of the outputs and collected from the input port, as shown in Figure 4.6. Interestingly, the same port coupling asymmetry can be observed in the now single output taper, with the non-perfect collection of light in the top input case. This is a nice demonstration of the device's reciprocal nature under test and furthers the argument that the asymmetry is due to the fabrication of the device and not an experimental artefact of the imaging system.

Next, the number of outputs was increased to a 1×4 MMI device, designed to produce an equal split of intensity to each of the four output ports. Once again the device is symmetrical about the x-axis, which results in two unique output couplings. The experimental and simulated perturbation maps are shown in Figure 4.7 a) for a device with MMI region design dimensions of $L = 29.4$ µm and $W = 8.0$ µm. Spatial and relative modulation intensity agreement is once again excellent, with correlation values as high as 95% (top output). However, the second-top output coupling correlates slightly worse compared to the others at only 86%. This deviation appears to stem from the larger positive $\Delta T/T$ region towards the top of the MMI region than the model predicts. Figure 4.7 b) shows the correlation strength for another set of simulated MMI widths and lengths around the design dimensions. The correlation region is wider for the top and bottom output couplings, suggesting these outputs are less sensitive to width changes than the central two outputs.

Finally, a multiple-input multiple-output device consisting of three inputs and three outputs (3×3 MMI) was investigated. Such a device has only four unique coupling combinations (e.g. top-top, top-middle, top-bottom and middle-middle) with the others bearing some kind of symmetry to these — as the device now has symmetry in both the x-axis and the z-axis. To function as a 3×3 device, it must split the intensity three ways irrespective of input. To achieve this, the device was designed to have outputs at the first three-fold self-imaging point for one of the off-centre inputs, which is equivalent to the third three-fold self-imaging point for the central input (see Section 2.2.4.2). Figure 4.8 a) and Figure 4.9 b) show the resulting experimental and computed perturbation maps for the unique and symmetrical input-output combinations. The device was designed with an MMI region of $L = 88.0$ µm and $W = 6.0$ µm. Agreement varies drastically more than previous two devices measured, owing to the increased size and complexity, with a maximum and minimum correlation coefficients of 93% (top-top) and 71% (top-middle), respectively. Due to this device's length, it is much more sensitive to changes in device width. Looking at Figure 4.8 b) and Figure 4.9 b) shows that the region of maximum correlation occurs at a slightly thinner device, around $W = 5.9$ µm, which is confirmed by the SEM measurement (Figure 4.4). Averaging the correlation maps for each output provides some indication of the device dimensions, but this approach remains best suited for viewing the asymmetry of a device that might stem from nanoscale fabrication defects or slight thickness variations in the SOI wafer.

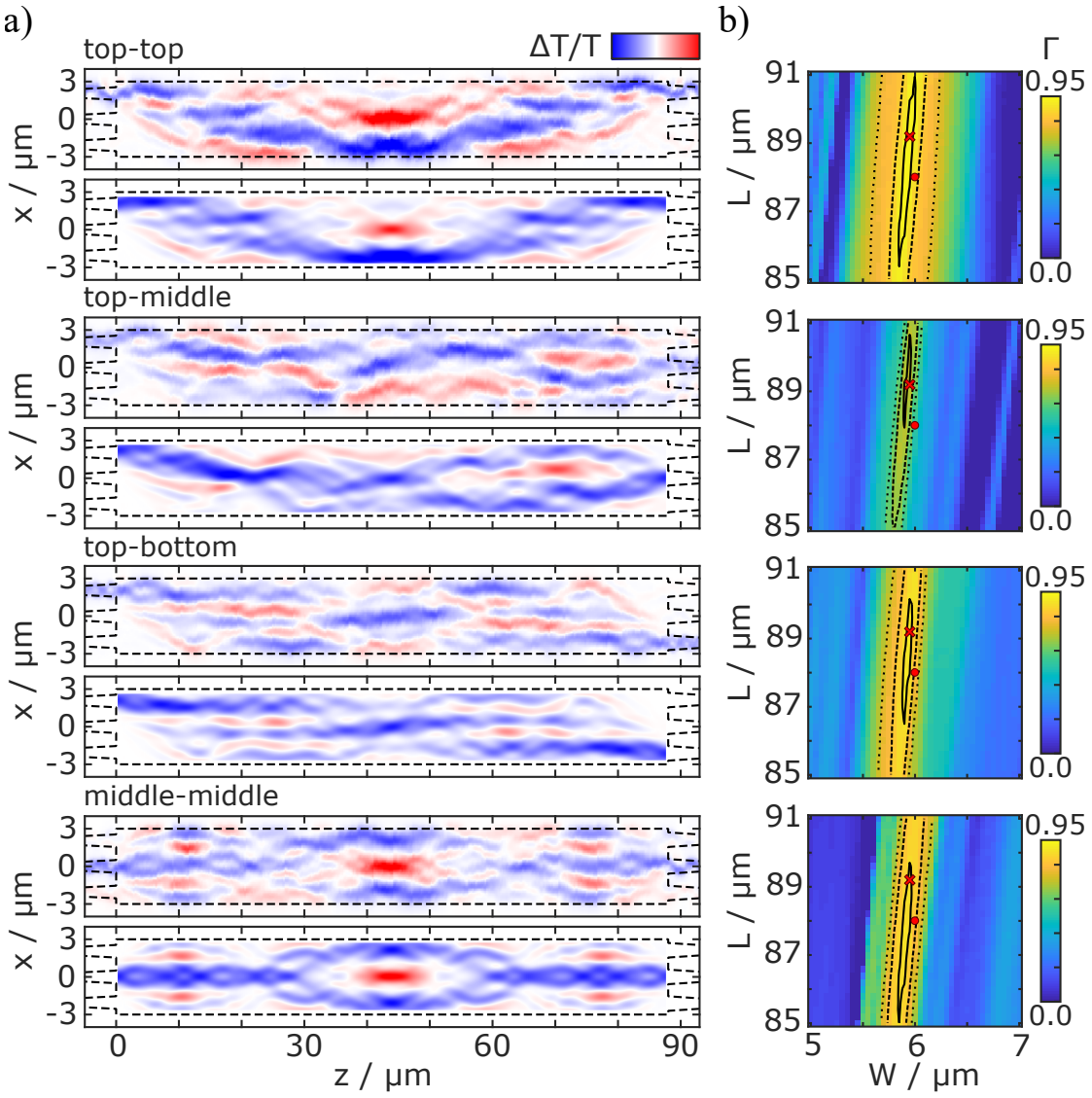


Figure 4.8: a) Experimental (top) and computed (bottom) perturbation maps for each of the unique input-output combination of the 3×3 MMI device. The design parameters are $L = 88.0$ µm and $W = 6.0$ µm, with fabricated length $L = 89.2$ µm and width $W = 5.95$ µm as measured by SEM. The experimental scans were carried out for 1550 nm probe pulses and 417 nm pump pulses with a delay time of 3 ps. The pump was set to a fluence of 60 pJ µm^{-2} at the focus. The computed maps were obtained for a 740 nm perturbation, with $\Delta n_{\text{eff}} = -0.25$, for the design dimensions of the device. Colour bar range is between $\Delta T/T = \pm 0.3$ for experiment and $\Delta T/T = \pm 0.4$ for simulation. (b) 2D correlation maps for varying theoretical MMI region dimensions L and W, for top and bottom outputs. The red dot and cross indicate the MMI design and SEM measured dimensions, respectively, and the contour lines indicate a 2% (solid), 10% (dash) and 20% (dot) decrease of the correlation with respect to the maximum correlation.

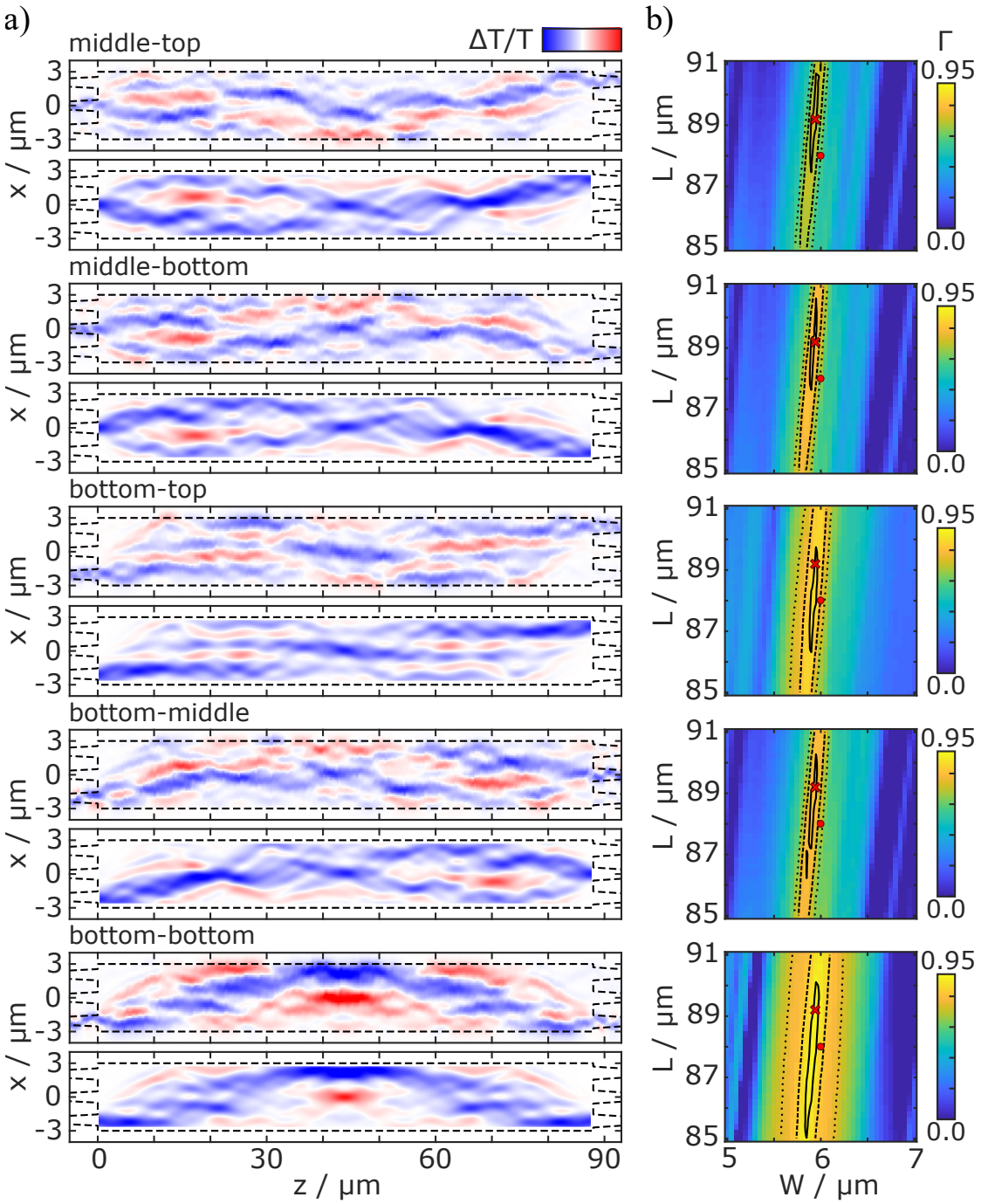


Figure 4.9: a) Experimental (top) and computed (bottom) perturbation maps for the other symmetrical input-output combination of the 3×3 MMI device. The experimental scans were carried out for 1550 nm probe pulses and 417 nm pump pulses with a delay time of 3 ps. The pump was set to a fluence of 60 pJ µm^{-2} at the focus. The computed maps were obtained for a 740 nm perturbation, with $\Delta n_{\text{eff}} = -0.25$, for the design dimensions of the device. Colour bar range is between $\Delta T/T = \pm 0.3$ for experiment and $\Delta T/T = \pm 0.4$ for simulation. (b) 2D correlation maps for varying theoretical MMI region dimensions L and W, for top and bottom outputs. The red dot and cross indicate the MMI design and SEM measured dimensions, respectively, and the contour lines indicate a 2% (solid), 10% (dash) and 20% (dot) decrease of the correlation with respect to the maximum correlation.

4.4 Light intensity field information retrieval

The current state-of-the-art probing techniques capable of revealing the flow of light in photonic systems rely on scanning perturbations to recover information on the electromagnetic field [126]. Typically, these techniques require direct access to the near-field, such as in NSOM and AFM [131]. However, this is not possible in most real-world devices, due to a capping layer's requirement to protect the near-field from any surface interactions.

In this section, we demonstrate photomodulation mapping's capability of retrieving information about the light intensity field with far-field optical pulses. Full recovery of the electric fields is not carried out here, as only the transmitted intensity is recorded with no phase information. In practice, phase information could be easily measured by in-phase (I) and quadrature (Q) demodulation using a 90° hybrid and a Local Oscillator (LO) but was found to lead to no additional useful information for these structures (see Appendix B.2). Additionally, no reflection or out-of-plane scattering is considered here, thereby restricting the study to exclusively transmitted light. Finally, the resolution of the measured photomodulation map is limited by the perturbation size. However, we have shown in Section 4.1.2 that in the case where losses are negligible, perturbation map of the total forward fields can be used to infer the transmitted light intensity field. This approach is based on two assumptions. Firstly, a negligible loss device can only exhibit negative $\Delta T/T$ areas; that is to say, transmission cannot be increased beyond the unperturbed case. Therefore, any regions of positive $\Delta T/T$ must correspond to light that was ultimately scattered out of the unperturbed device. And secondly, a larger change in transmission coincides with a higher intensity of light at that position, as any perturbation will have a greater overall effect on the transmission. This means that the magnitude of the $\Delta T/T$ signal should be proportional to the intensity field strength at that location.

Figure 4.10 shows the recovered light intensity field intensity for a silica-clad 1×2 MMI with length $L = 33.2$ µm and width $W = 6.0$ µm. The use of a second 1×2 MMI spitter, but reversed (referred to as a 2×1 MMI splitter), allows the two ports to be recombined to a single output, essentially forming a balanced Mach-Zehnder Interferometer (MZI). The FDTD simulated $|\mathbf{E}|$-field for forward excitation is shown in Figure 4.10 a) and is compared to the experimental and modelled perturbation maps in part b). Qualitative agreement between the experimental result and simulated field is very good, with up to the 5-fold self-imaging point still clearly visible. However, much of the finer detail structure towards the first 10 µm is not recovered and appears washed out due to the limited resolution from the perturbation's size. Asymmetry in the device is visible as different magnitudes in the device's outputs, which indicates stronger coupling to the bottom output in this case. Also, the input taper does not seem to be providing very good enlargement of the fundamental mode as intended, instead, two distinct beam

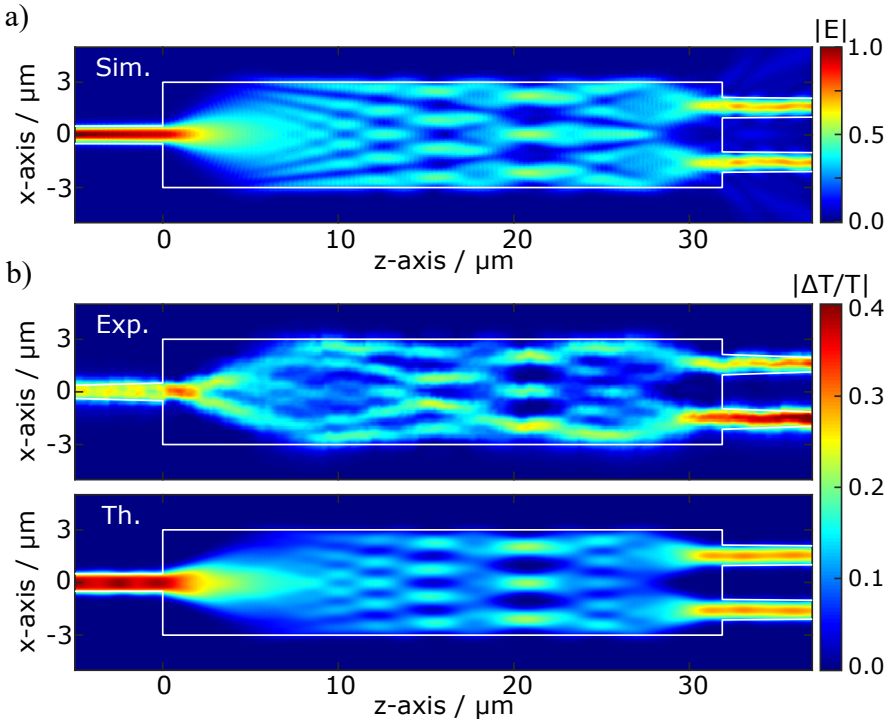


Figure 4.10: Electric field intensity recovery for a silica-clad 1×2 MMI with length $L = 33.2$ µm and width $W = 6.0$ µm. a) FDTD simulated $|\mathbf{E}|$-field forward excitation. b) Experimentally measured and computationally modelled perturbation maps of the device. To experimentally obtain the photomodulation map for both outputs simultaneously a reversed 1×2 MMI (2×1) splitter was used to recombine both outputs to a single grating. The experimental scan was carried out for 1550 nm probe pulses and 417 nm pump pulses with a delay time of 3 ps. The pump was set to a fluence of 60 pJ µm^{-2} at the focus. The computed map was obtained for a scanning 740 nm sided square perturbation.

paths can be seen crossing over one another. It is likely to be this phenomenon that is causing the 'zigzag' pattern previously seen in some of the single output photomodulation maps, where one path couples more strongly to a particular output. Nevertheless, this clearly demonstrates photomodulation mapping as a technique capable of recovering light intensity field distribution information of a PIC, even in the presence of protective cladding layers that restricts near-field approaches.

4.4.1 Summed photomodulation maps

In instances where the outputs cannot be recombined, it is possible to retrospectively sum individual output photomodulation maps, providing that the unperturbed device performs well with equal power splitting between each output port. This approach is not as exact as in the combined case due to the measurements not occurring simultaneously, and therefore slight coupling and environmental variations between scans will deteriorate the recovery. A better alternative for testing real-world devices would be to use a fibre

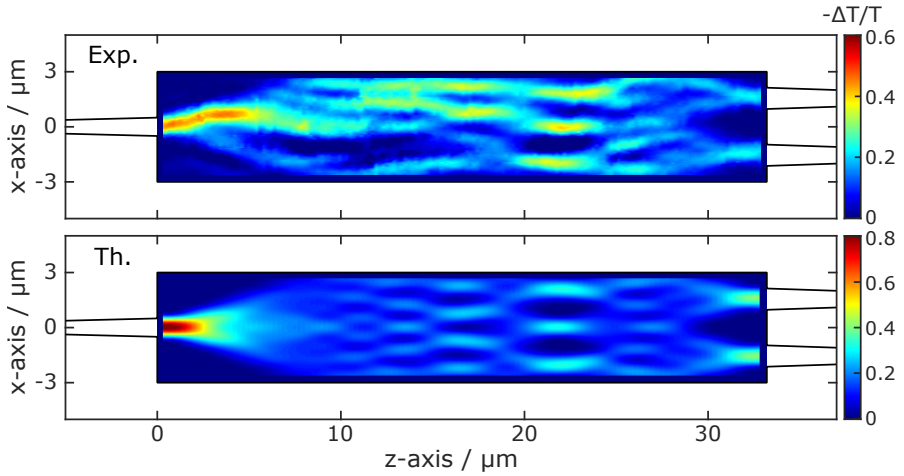


Figure 4.11: Summed experimental (top) and modelled (bottom) perturbation maps for the 1×2 MMI device in Figure 4.5.

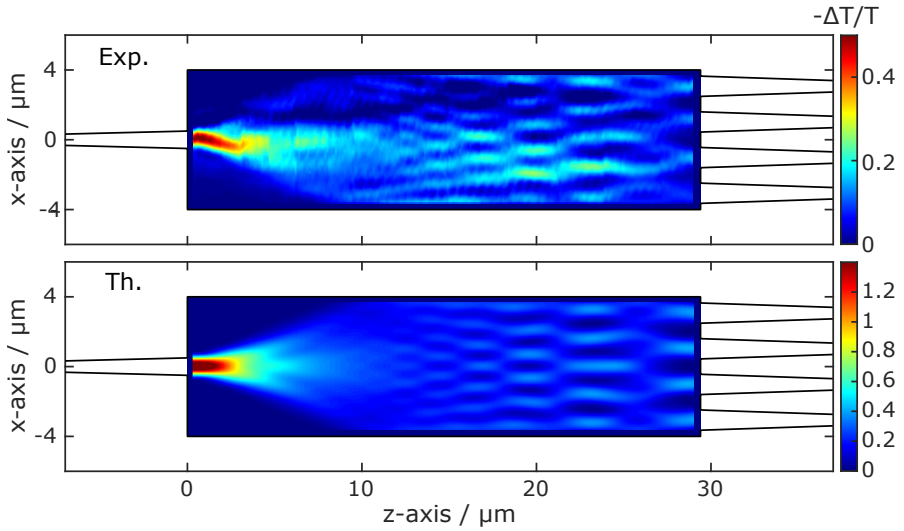


Figure 4.12: Summed experimental (top) and modelled (bottom) perturbation maps for the 1×4 MMI device in Figure 4.7.

array for simultaneous measurement of all outputs' transmission. Individual components of a larger, more complex PIC can be directly measured using erasable grating, where temporary gratings can be written directly into the device under test and erased after taking the measurements [25, 26, 145].

Figure 4.11 shows the summed experimental and modelled perturbation maps for the 1×2 device corresponding to Figure 4.5. The agreement is not as good as the combined case in the previous section; however, the general structure of the MMI region is still clearly visible with the self-imaging points in the correct position. The asymmetry in the individual photomodulation maps is clearly seen in the degradation of the lower first-third of the MMI region. Once, again this is attributed to the non-symmetrical expansion of the input mode.

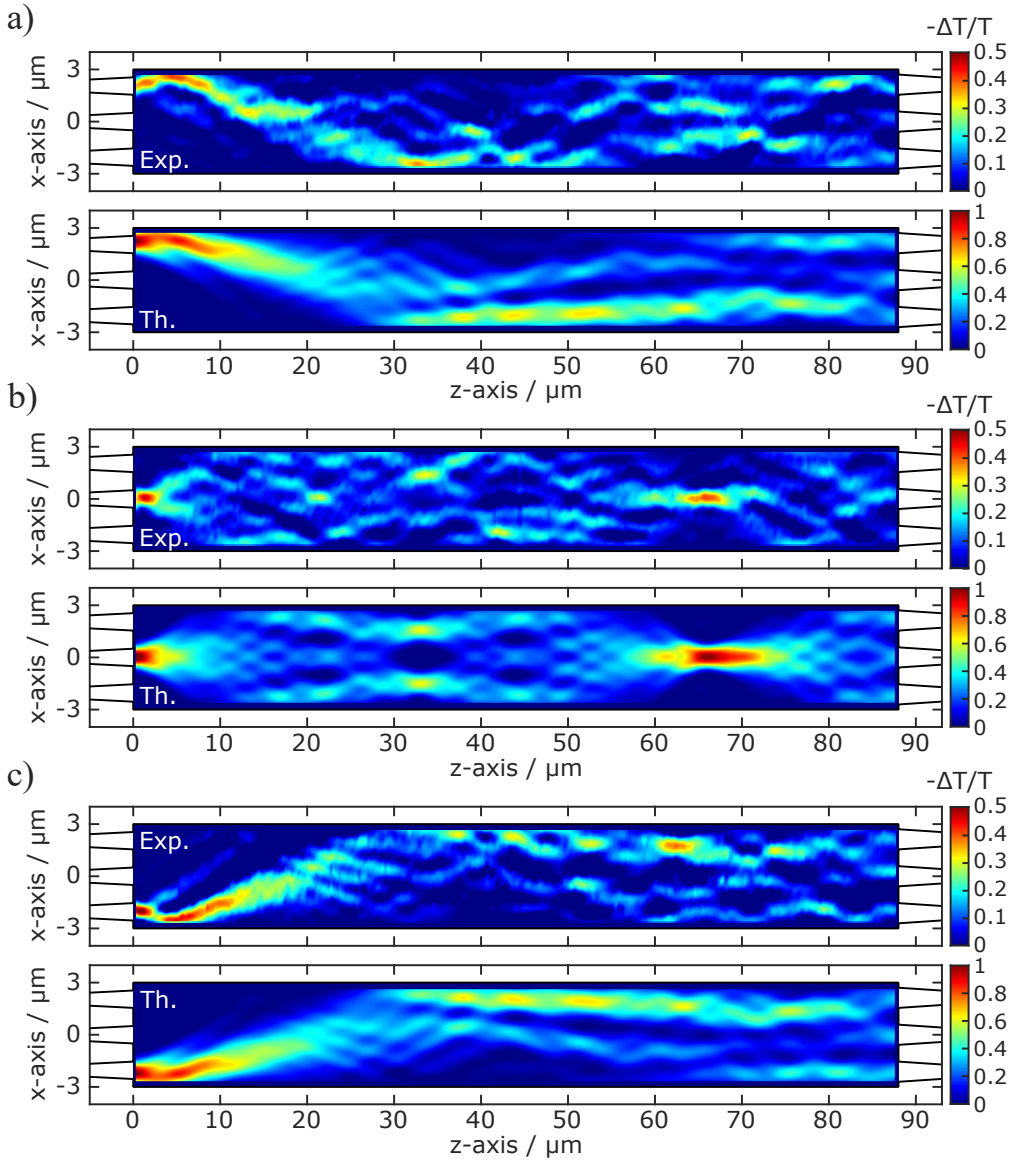


Figure 4.13: a) Top, b) middle and c) bottom summed (top) and modelled (bottom) perturbation maps for each of the inputs for the 3×3 MMI device in Figure 4.8 and Figure 4.9.

Similarly, Figure 4.12 and Figure 4.13 show the summed experimental and modelled perturbation maps for the 1×4 and 3×3 devices, respectively. The simulated electric field distributions of all the structures are provided in Figure 4.3. Once again, the summed photomodulation maps show good qualitative agreement, with the key self-imaging points clearly visible within the pump spot's resolution. The poorest correlating 1×4 port coupling results in the breakdown of the experimentally retrospectively summed maps for the top half of the MMI region compared to the modelled distribution. Meanwhile, the 3×3 device summed maps are not as straightforward as for the previous two devices, with significantly fewer self-imaging points clearly visible. However, the general overall structural agreement is still good, and the deviations are attributed to the more sensitive nature of the 3×3 device due to its longer length. All the summed

maps exhibit almost exclusively negative $\Delta T/T$ values, which is expected for low loss devices where the transmission cannot be improved much beyond that of the unperturbed case.

4.5 Summary

In this chapter, we first introduced a general analytical model to predict the transmittance coupling change between any input and output mode of an arbitrary linear photonic system. As with other adjoint methods, the technique only requires two electromagnetic field simulations to be undertaken; one for both the forward and backward propagating electric fields of the desired input and output modes. This represents a significant saving in computational time compared to brute-force approaches which scale quadratically with device area. For $\ll\lambda$ scale and low magnitude refractive index perturbations this approach is found to be nearly exact. Whilst for wavelength-scale perturbations an additional approximation is required, in which the perturbation is modelled as a 1D Fabry-Pérot cavity. This is valid for systems where the majority of light is travelling along the propagation axis, such as the devices we consider here. The transmission and reflection coefficients of the cavity can then be calculated simply from the perturbation's effective index shift. In the case of low refractive index contrast, it is found that it is the dephasing term which is dominant, as the reflection coefficient will be nearly negligible. Next, we demonstrated that for low-loss devices the total forward propagating field can be used to directly infer information on the light intensity field within a device.

Next, we verified the analytical model via direct comparison to experimentally obtained perturbation maps of multiport photonic devices, using the UPMS technique outlined in Chapter 3. Overall, very good agreement was found between the experimental and predicted maps from our model. This enabled us to harness its fast computational speed to carry out full parametric studies of MMI length and width design space, thereby indicated any fabrication deviations compared to design and allowing quantitative characterisation of fabrication compared to an ideal design structure. The parametric study could in principle be expanded to include other features, such as SOI wafer thickness, cladding and thermal variations. UPMS's non-destructive and non-evasive nature makes it an ideal technique to be applied at the wafer-scale. In principle, individual components can be investigated even contained within a much larger and more complex circuit, as long as there is access to the desired input and output ports. Characterisation speed is one limiting factor, which currently restricts the technique to more diagnostic applications. For typical scan parameters used here, we achieved scan speeds on the order of a few micrometres per second, resulting in full scanning times on the order of minutes depending on the device size. However, this can be significantly improved by either increasing the raster scan step-size or moving to a different scanning system. Mirror

galvanometers can have kilohertz scanning frequencies that can result in device scanning times on the order of seconds if combined with high-frequency optical modulators to reduce dwell times significantly.

Finally, we demonstrated photomodulation mapping's capability of revealing the light intensity field distribution within a device even in the presence of protective cladding layers, which render conventional near-field techniques impossible. The approach requires the total propagating light field to be measured and is theoretically limited to devices with high transmission and perturbations that induce small effective index shifts. Nevertheless, even in non-ideal cases, it is possible to infer some information on the propagation of light within. The field distribution obtained is also limited in resolution by the perturbation size, which washes out some of the fine detail structure present in devices. In the event that each output port cannot be simultaneously measured, we have also shown that it is sufficient to obtain a photomodulation map of each individual output port and retrospectively combine them. However, small shifts between scans, such as environmental changes, were found to deteriorate the summed maps' quality compared to the total throughput's simultaneous measurement. In real-world settings, this could be achieved via the use of fibre arrays, and individual components could be directly probed using erasable couplers or gratings.

Chapter 5

Optical Trimming of Ion Implanted Silicon for Critical Coupling Control of Racetrack Resonators

Silicon photonics continues to be adopted as a mainstream technology. More recent advancements are giving rise to increasingly complex circuitry, with resulting chips having hundreds of densely packed individual components which all need to perform within tightly defined parameters [146–149]. Even with the decades of experience to directly draw upon from the electronics industry, these tight tolerances during the fabrication process for Photonic Integrated Circuits (PICs) may give rise to a variety of problems that negatively impact device performance, particularly for resonant structures. Parameters such as the silicon wafer uniformity itself can result in significant deviations from the design structure's performance and require post-fabrication correction. Even devices fabricated well within tolerances still have to contend with environmental factors, such as temperature, that can result in detuning.

Typically device performance is restored via active means, such as integrated micro-heaters or thermoelectric coolers, which use the thermo-optic effect (see Section 2.3) to control the refractive index of a section of silicon. This approach can be used to offset both the fixed error of fabrication defects and the dynamic error caused by the device's environment. However, heaters increase both the total energy consumption and complexity of a chip, requiring additional space and electrical contact pads. The most common configuration of an integrated heater is a metal resistive element separated from the waveguide by around one micrometre of SiO_2 [150]. Their performance is typically given in terms of tuning efficiency and speed. The tuning efficiency is the amount of power required to shift the wavelength by 1 nanometre (mW/nm), or in the case of

ring resonator devices, it is common to quote the value in terms of the Free-Spectral-Range (FSR) to remain consistent for varying ring size [151]. Moreover, the tuning speed is simply how fast the heater can cause a shift in wavelength and is governed by the material's thermal response time. The SiO_2 layer is needed to protect the optical mode; however, it reduces the heater performance due to the generated heat having to diffuse down to the waveguide below [152].

This chapter investigates the permanent trimming approach of silicon photonic devices via germanium (Ge) ion implantation. This approach cannot alleviate shifts in performance due to environmental reasons but can be used to offset any fabrication errors and tune devices back to their target specification. Therefore, the use of integrated heaters will still be required, but they can be driven with significantly lower energy consumption, which is a significant concern in data centres currently [17, 153, 154]. The refractive index tuning is achieved through the localised disruption of the silicon crystal lattice resulting in amorphous silicon, which has a significantly higher refractive index [155, 156]. Localised laser annealing can then be used to regrow the crystalline lattice with minimal impact on device performance. Ion implantation has been previously used to write temporary grating for component testing [25, 26, 157], enhanced photodetection [158], as well as the trimming of ring resonators and Mach-Zehnder Interferometers (MZIs) [159–162].

5.1 Experimental setup

The experimental setup is broadly similar to that used for Ultrafast Photomodulation Spectroscopy (UPMS) with the grating spectrometer. An illustration of the annealing setup is shown in Figure 5.1. A 200 fs mode-locked Ti:Sapphire laser is used in conjunction with an Optical Parametric Oscillator (OPO) to produce the broadband probe pulse centred around 1550 nm with a 20 nm bandwidth and 80 MHz repetition rate. The second harmonic of the 834 nm seed pulse provided the pump beam at 417 nm, which is externally focused onto the device's surface using a 100× objective with a numerical aperture of 0.5. The resulting focused pump spot has a Full-Width Half-Maximum (FWHM) of 740 nm (see Section 3.4.3). The focusing objective is mounted to a 3D piezo nanopositioner, allowing the pump spot to be scanned over the ion-implanted region. At the same time, the device's output is analysed using a grating spectrometer equipped with an InGaAs 512-pixel Charge-Coupled Device (CCD) array providing 0.08 nm spectral resolution over a 41 nm range which is sufficient to cover the entire bandwidth of the probe pulse. The spectrometer's use also means there is no requirement to apply separate modulation to the two beams. Instead, a shutter is placed in the pump beam to differentiate between the pumped and unpumped output when recording output spectra. A narrow linewidth swept-source laser can also be coupled into the input fibre for

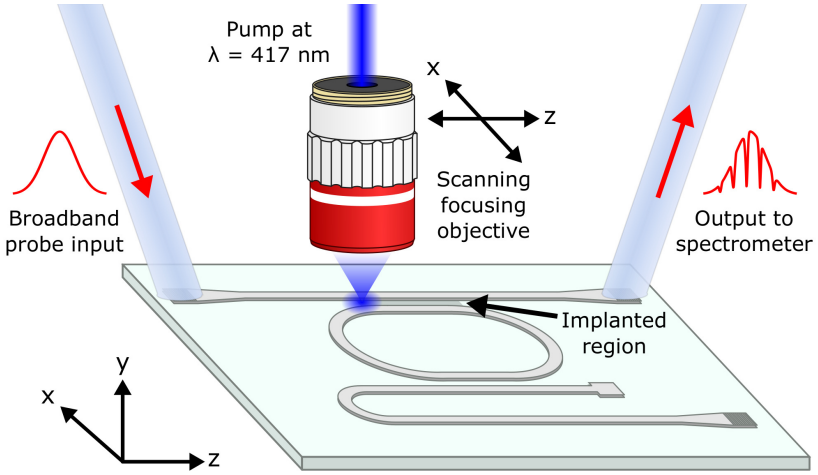


Figure 5.1: Broadband infrared probe pulses are grating-coupled into and out of the device under test. The transmitted spectrum is then analysed using a grating spectrometer with InGaAs CCD array. The annealing pump beam is incident perpendicular to the device's surface through a focusing objective mounted to a 3D piezo nanopositioner, allowing the pump to be focused and scanned over the device.

broadband transmission measurements (see Section 3.3.2) [163–165], and is required for the gradient feedback when tuning to a specific target wavelength [160].

5.1.1 Device design and fabrication

The devices were designed and fabricated by the Silicon Photonics Group using electron beam lithography and inductively coupled plasma etching, and are standard 220 nm thick Silicon-On-Insulator (SOI) rib waveguides with a 100 nm thick slab layer. Figure 5.2 shows an optical microscope image of a fabricated racetrack resonator. 500 nm wide waveguides are used for their single-mode operation, and the racetrack itself is comprised of a 25 µm bend radius with a straight coupling length of 10 µm and a coupling gap of 260 nm edge-to-edge. Grating couplers were used to allow light to be coupled into and out of the device, with access to both the through and drop ports of the racetrack resonator. Un-implanted devices with the same dimensions were also fabricated for reference. Figure 5.3 shows the transmission spectrum for one such reference device using the narrow linewidth swept-source laser, which was found to have an Extinction Ratio (ER) of >25 dB and FSR of 3.43 nm. Simulations of the un-implanted device were carried out using Lumerical Mode Solutions software using the 2.5D variational Finite-Difference Time-Domain (varFDTD) solver. Simulations were run for 25 000 fs to ensure convergence and with a 25 pm spectral resolution. General agreement between the measured fabricated device and simulation is good, with a ∼50 pm difference in FSR. However, the small wavelength shift shown in Figure 5.3 indicates a significant

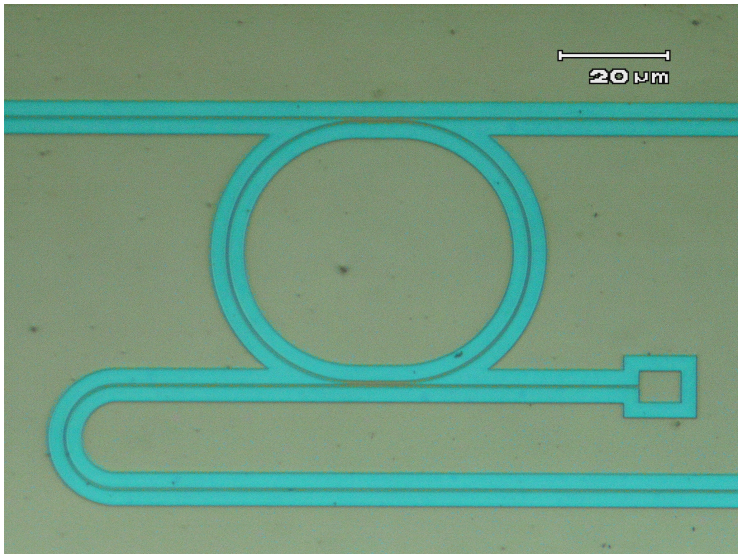


Figure 5.2: An optical microscope image of the reference racetrack resonator.

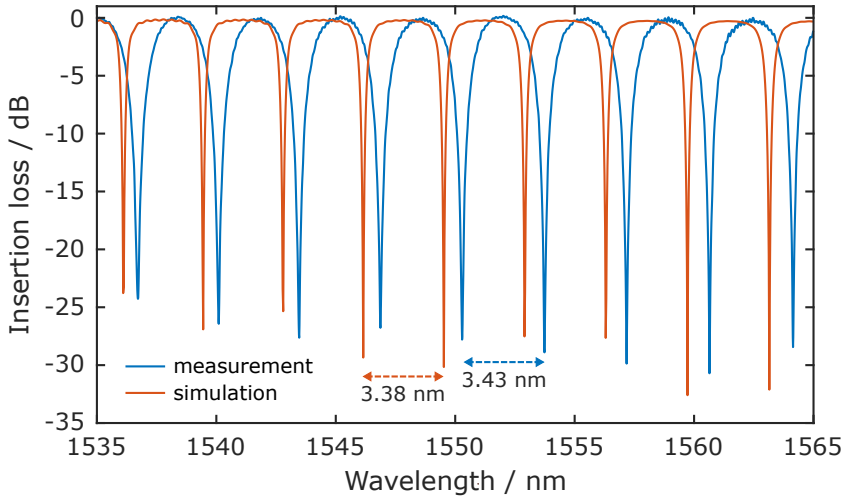


Figure 5.3: Transmission spectra of the through-port for an un-implanted racetrack resonator as simulated and measured, with FSRs of 3.38 nm and 3.43 nm, respectively. The experimental measurement is normalised to a straight reference waveguide, and Lumerical Mode Solutions software was used to simulate the response for the fundamental Transverse Electrical (TE) input mode.

loss of performance for wavelength critical applications – highlighting the importance of post-fabrication trimming of devices.

Germanium ion implantation was used to induce lattice damage to the silicon slab layer between the bus waveguide and the racetrack (see Figure 5.1). The ion energy of 130 keV and a fluence 1×10^{15} ions/cm^2 were used, which have previously been demonstrated as the optimal parameters for the complete amorphisation of a 100 nm thick silicon layer [123]. The refractive index of the implanted region has a refractive index change of $\Delta n = 0.48$ compared to crystalline silicon, with an increase to the imaginary part of 8.5×10^{-4} [122, 166].

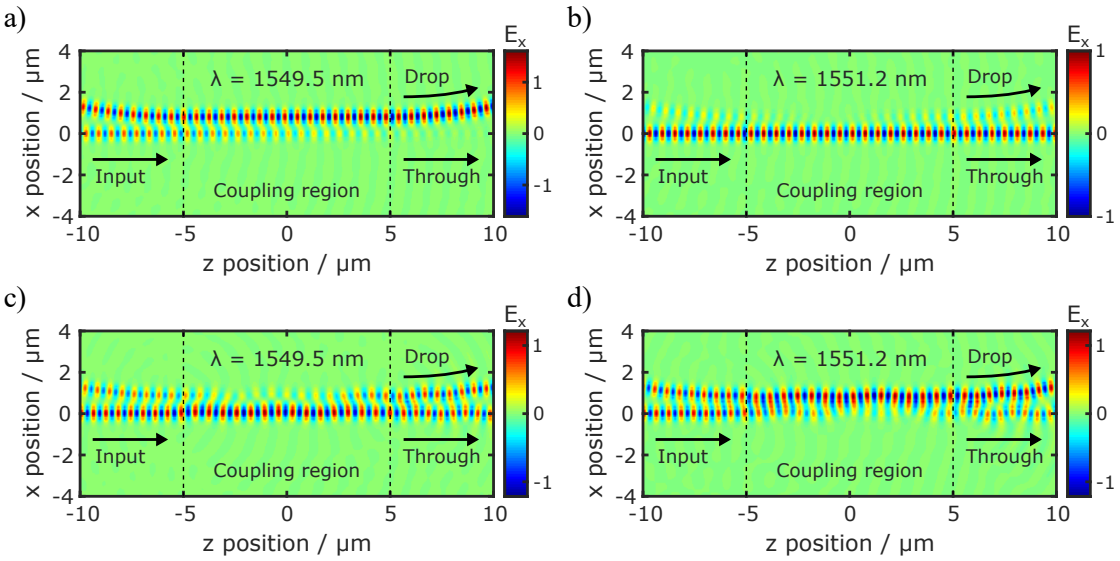


Figure 5.4: Electric field profile (x-component) of the coupling region between the bus waveguide and racetrack resonator for the a) on-resonance and b) off-resonance wavelengths of 1551.2 nm and 1549.5 nm respectively. c) and d) show the electric field profiles for the same device with Ge ions implanted in the 10 µm long gap between the ring and bus waveguide. Simulated using Lumerical Mode solver for the fundamental TE input mode.

Raman spectroscopy was carried out to confirm the ion-implanted silicon's complete amorphisation. The measurements were taken with a ×50 objective lens using a 0.2 mW Nd:YAG laser at a wavelength of 532 nm with an exposure time of 10 s. Data was accumulated over three exposures to improve Signal-to-Noise Ratio (SNR). Figure 5.5 shows the complete amorphisation of the ion-implanted silicon, as indicated by the broad peaks at 150 cm^{-1} and 465 cm^{-1} combined with the complete suppression of the 521 cm^{-1} crystalline silicon peak, and is consistent with published values [167–169]. There is also a very shallow peak that can be observed at 300 cm^{-1}, which corresponds to the very low dosage of implanted germanium ions [170].

5.1.2 Method

The spatial monitoring of the annealing was carried out by raster scanning the pump spot over the ion implanted region plus an additional 2.5 µm before and after the implanted region to provide both pre- and post-annealing data. A scan width of 1 µm was chosen to ensure equal exposure to the focused laser spot over the entire implanted region. The 3D nanopositioner was set to the optimal focus, and 100 nm steps were used in both scanning directions. The shutter is opened at every position, exposing the device to the pump for 500 ms and a spectrum is recorded before the shutter is closed again. After the shutter is closed, the unpumped measurement is delayed by another 500 ms to allow sufficient time for both the free-carrier and thermal contributions from the pump to recover. The nanopositioner then moves to the next spot, and the process is repeated.

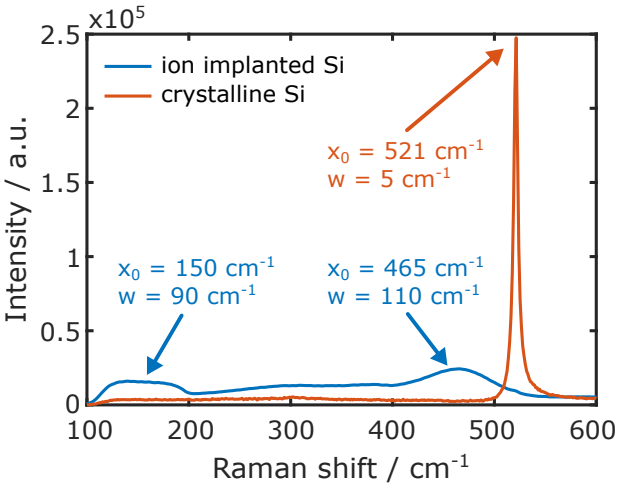


Figure 5.5: Raman spectra of Ge ion-implanted and crystalline silicon. Arrows indicate the key Raman peaks with their central position, x_0, and FWHM, w.

The combination of limited spectral resolution and background scattering with a 16-bit CCD detector array resulted in an available dynamic measurement range of only ∼16 dB. This was confirmed by testing the same un-implanted reference device with both a swept-source and the grating spectrometer and observing the respective extinction ratios.

5.2 Results

The results section is split into two parts; the simulated complete annealing response and the experimental localised pulsed laser annealing response. In the paper [161], the continuous wave annealing was also investigated using a 488 nm argon-ion laser, in which the complete crystalline regrowth of the ion-implanted silicon was shown. A spot size of width 2.5 µm was obtained with a 20× focusing objective. The same setup has previously been used to demonstrate the annealing of other Ge ion-implanted devices [159] and the crystallisation of amorphous silicon core fibres [171].

5.2.1 Simulated annealing response

The transmission response during annealing was modelled by varying the ion implanted region's length, as shown in Figure 5.6 a). A simulated annealing step size of 500 nm was chosen, due to the long simulation times caused by the large device footprint and high energy storage of the ring. The direction of the annealing was carried out along the bus waveguide's propagation axis. Figure 5.6 b) shows the pre-annealing and optimal annealing insertion loss spectra, from which a clear reduction to the envelope loss of ∼1 dB can be seen as well as a large increase to the extinction ratio and slight blue-shifting of resonance peaks. The evolution of one peak around 1550 nm is shown in Figure 5.6 c), where the darker shade of blue represents a longer annealed length. This also clearly

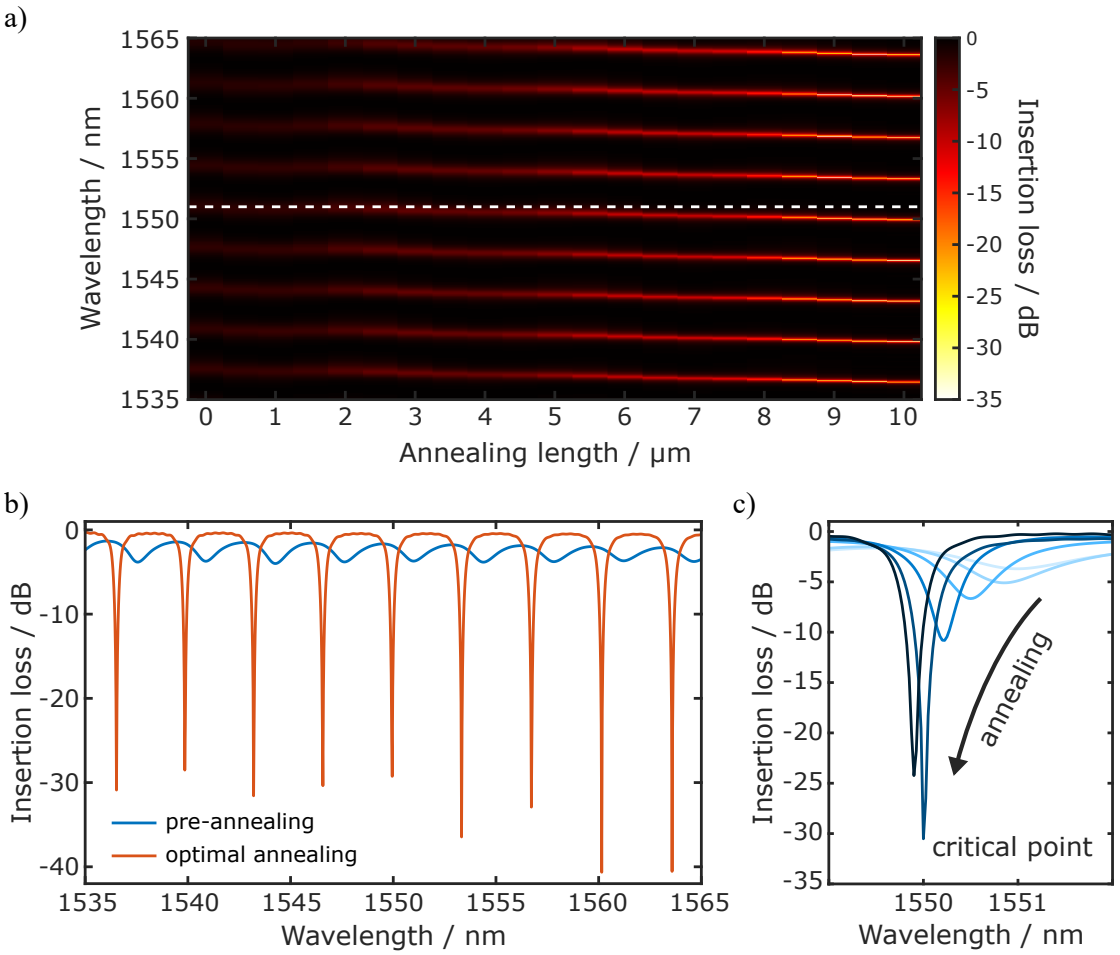


Figure 5.6: Simulated annealing of an ion-implanted racetrack resonator. a) Insertion loss as a function of annealed length in 500 nm steps. b) Transmission spectra for the pre-annealed and optimally annealed device. c) Evolution of one resonance peak around 1550 nm as indicated by the white dashed line in a). Darker shades of blue correspond to longer annealed length; 0.0 µm, 2.5 µm, 5.0 µm, 7.5 µm, 9.0 µm and 10.0 µm, respectively, and the critical coupling point is indicated by the largest extinction ratio. Simulated using Lumerical Mode solver for the fundamental TE input mode.

shows the passing of the critical coupling position at ∼9.0 µm as the extinction ratio starts to decrease once again, while the resonance position continues to blue-shift.

The mean resonance extinction ratio and shift are displayed as a function of the annealed length in Figure 5.7 a). The extinction ratio follows a non-linear trend, as expected from changing the coupling length of a Directional Coupler (DC) [14]. The critical coupling position was found to be in the region 9.0 ± 0.5 µm, with a mean extinction ratio of 35.7 dB and a standard deviation of 6.0 dB over the nine resonance peaks within the wavelength range. The critical coupling position accuracy is limited by the number of implanted region lengths simulated. Meanwhile, the resonance shift follows a quasi-linear response of 0.033 ± 0.001 FSR/µm (error is from linear fit) as the ring's effective optical path length is slightly reduced during annealing. Figure 5.7 b) shows the increasing Q

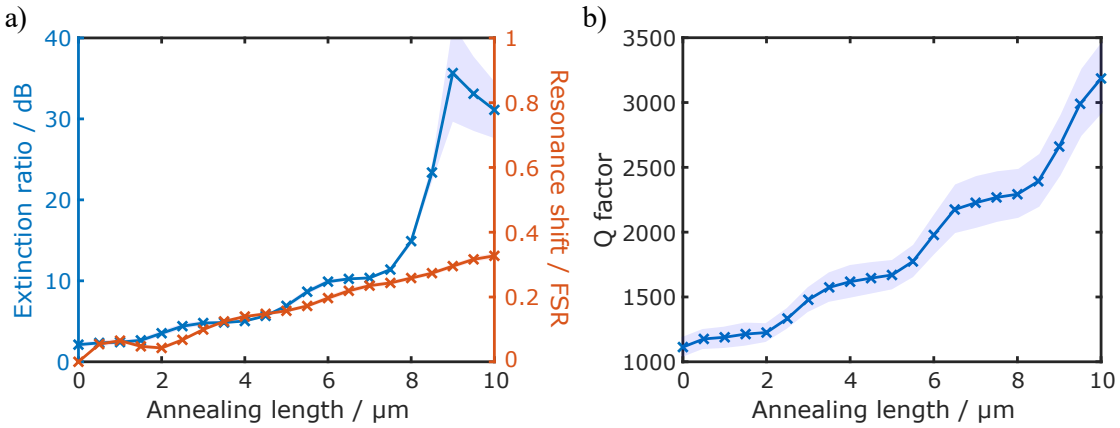


Figure 5.7: a) Mean extinction ratio and resonance shift of peaks between 1535 nm and 1565 nm during simulated annealing of an ion-implanted racetrack resonator in 500 nm steps. Optimal annealing length was found at 9.0 µm with a resonance peak shift of approximately a third of an FSR. b) Mean Q factor of resonance peaks as a function of annealing position. Shaded areas represent the nine resonance peaks' standard deviation within the wavelength range investigated. Simulated using Lumerical Mode solver for the fundamental TE input mode.

factor of the ring as the implanted region is reduced, with the initial implanted device having a mean $Q = 1110 \pm 80$ and the fully annealed device returning to $Q = 3190 \pm 270$, where errors are the standard deviation of the fitted FWHMs for the nine investigated peaks.

5.2.2 Localised pulsed laser annealing

Initially, pulse power tests were run on bulk ion-implanted regions of silicon to find the optimal pump power for annealing, without causing damage. Raman spectroscopy was then performed to access the quality of the silicon recrystallisation. A sufficiently large area needs to be annealed to obtain a reliable Raman spectroscopy reading; therefore, 5 µm squares were chosen to be annealed in 100 nm steps with varying pump powers and dwell times. The best resulting Raman spectrum is shown in Figure 5.8 in which a clear silicon peak can be seen at 518 cm^{-1} with a reduction to the broad, amorphous silicon bands centred around 150 cm^{-1} and 465 cm^{-1}. However, it is clear from the Raman spectrum that the silicon is not fully crystallised. The silicon peak is located at a smaller Raman shift than for crystalline silicon, and a shoulder is readily apparent to the left of the peak, reducing the peak symmetry. This, coupled with a significantly larger FWHM of 13 cm^{-1}, indicates that the resulting state is polycrystalline. There is also a small peak that can just about be seen at 300 cm^{-1}, which is indicative of germanium. This peak's very low intensity is due to the low dosage used in the ion implantation process.

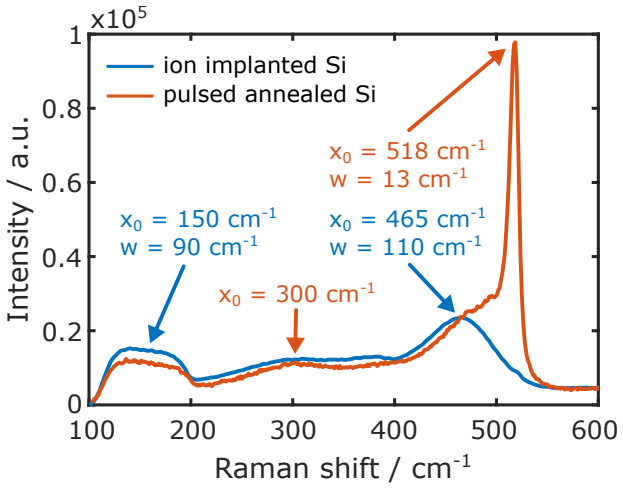


Figure 5.8: Raman spectra of Ge ion-implanted and pulsed laser annealed silicon. Arrows indicate the key Raman peaks with their central position, x_0, and FWHM width, w.

However, in all cases it was found that pulsed laser annealing was insufficient to provide full regrowth of the silicon crystalline lattice, as shown by the shoulder at 480 cm^{-1}. Further increases to the pump power result in visible damage and rapid loss of transmission in the devices. The average pump power of 5 mW was chosen for pulsed laser annealing as it resulted in the clearest silicon peak without any visible damage and the same power previously reported to work well for the same laser system [160].

The spectral response during the annealing process is shown in Figure 5.9 a) for a 5 mW average pump power. Once the pump laser spot reaches the implanted region, the resonance positions start to blue-shift during annealing, as expected from the refractive index reduction of going from amorphous to crystalline silicon. Figure 5.9 b) shows the pre-annealed and optimal annealing length for the racetrack resonator. The resonance shift here is around half an FSR per 10 µm annealed, which is significantly smaller than the whole FSR per 10 µm annealed that has been previously reported [160]. This is attributed to the ion implantation here being in the gap between the ring and the bus waveguide, opposed to directly along a segment of the ring itself. Therefore, the effective length change of the ring is significantly smaller in this work. It is, however, slightly larger than one would expect from the simulations (see Figure 5.7), which is likely caused by fabrication tolerances during the ion implantation process. If the ring itself experiences some ion dosage, then the ring's effective length will be significantly increased compared to the simulations. An enhancement of ∼8 dB to the ER was found at the optimal annealed length. The evolution of one resonance peak about 1550 nm is shown in Figure 5.9, which indicates the resonance passes through the critical coupling position with a ∼1 dB reduction to the ER in the last ∼1 µm of the implanted region.

Figure 5.10 a) shows the mean extinction ratio and resonance spectral shift extracted from the spectra in Figure 5.9 as a function of annealing position. Annealing can be

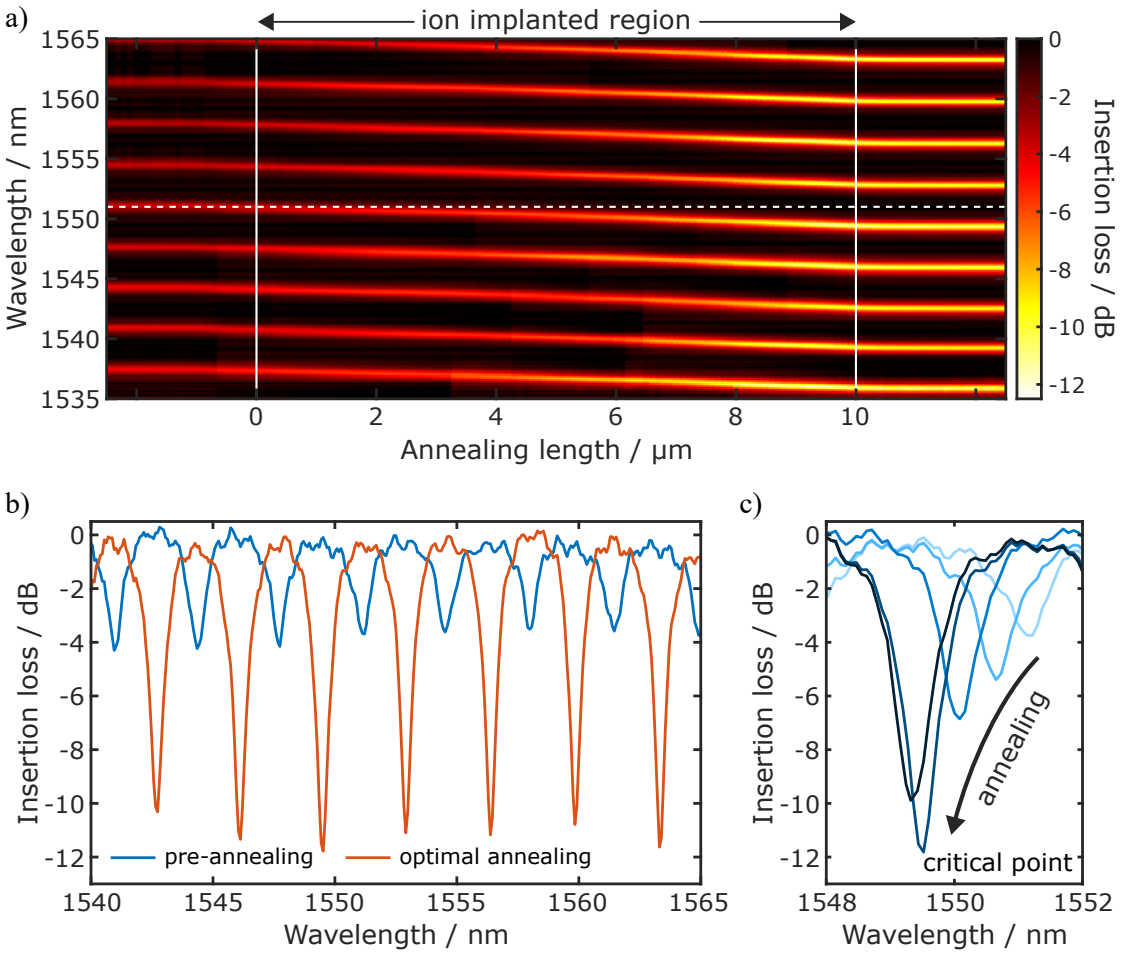


Figure 5.9: Experimental pulsed laser annealing of ion-implanted racetrack resonator. a) Insertion loss, normalised to probe pulse envelope, as a function of annealed length in 100 nm steps for 5 mW average pump power. Solid white lines indicate the boundaries of the implanted region. b) Transmission spectra for the pre-annealed and optimal annealed device. c) Evolution of one resonance peak around 1550 nm as indicated by the white dashed line in a). Darker shades of blue correspond to longer annealed length, and the critical coupling point is indicated by the largest extinction ratio.

seen for 400 nm before and after the implanted region, which is attributed to the pump laser spot size at the focus of 740 nm at FWHM. The extinction ratio can be seen to increase non-linearly towards the critical coupling position at 9.3 ± 0.2 µm annealing length, where the mean extinction ratio is 11.0 dB with a standard deviation of 0.7 dB. After the critical coupling position, the ER begins to degrade by ∼1 dB. Meanwhile, the resonance shift is found to follow a linear relationship with the length of annealed implanted region, yielding a gradient of 0.05 ± 0.02 FSR/µm. The resonance shift continues to increase past the optimal coupling position as the ring's effective length is still being reduced as more of the amorphous silicon is annealed. The quality factor of the ring was not found to significantly improve during annealing, as shown by Figure 5.10 b). The initial implanted device had $Q = 1360 \pm 250$, as opposed to after annealing where

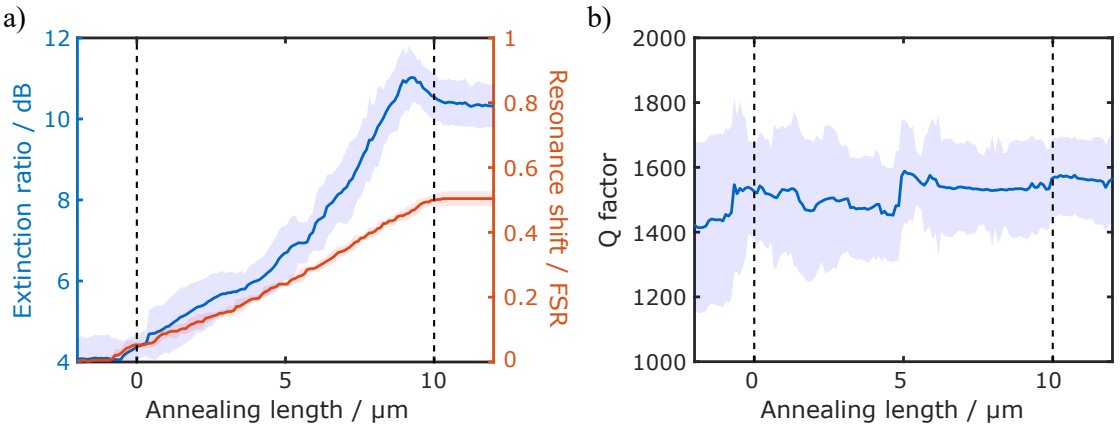


Figure 5.10: a) Mean extinction ratio and resonance shift of peaks between 1535 nm and 1565 nm during pulsed laser annealing for 5 mW average pump power. Optimal annealing length was found at 9.3 ± 0.2 µm with a resonance peak shift of approximately half an FSR. b) Mean Q factor of resonance peaks as a function of annealing position. Shaded areas represent the nine resonance peaks' standard deviation within the wavelength range investigated.

the quality factor was found to increase slightly to $Q = 1580 \pm 120$. This relatively low increase compared to the simulations is attributed to the incomplete recrystallisation of the silicon with pulsed laser annealing.

After the experiment, the device was left to cool down fully and then a final spectrum was taken to check for any heating effects that may be present from the average power of the pump pulses. No shift was found within the resolution of the spectrometer, indicating that the 500 ms dwell time after the pump shutter closes provides sufficient time for the small amount of energy deposited by the pump to dissipate entirely.

5.3 Summary

Post-fabrication component trimming is essential to ensure good performance of photonic circuits, especially as integrated photonics continues to mature and chip design is becoming ever more complicated. Traditionally, such control is carried out via integrated heaters, which can correct fabrication defects and offset any environmental shifts; however, the inclusion of heaters drastically increases the power consumption of PIC. Data centres already account for a significant proportion (∼1–3%) of the world's total energy consumption [154], which is only expected to increase over the coming decade. The situation is further exasperated by the development of interferometric meshes for new emerging fields, such as quantum and neuromorphic computing, which can have hundreds of individual MZIs that all require tuning and control.

In this chapter, we demonstrated the use of ion implantation and localised laser annealing for the trimming of racetrack resonators. The approach once carried out, offers permanent offsetting of fabrication defects and wafer thickness variations without the need for continuous power draw. Heaters are still required for dynamic control and environmental change offsetting. Additionally, the ion implantation process is Complementary Metal-Oxide-Semiconductor (CMOS) compatible and can be carried out with only one additional alignment step during fabrication. We initially simulated the racetrack resonator's annealing response, which shows a significant increase in the extinction ratio with annealed length (>35 dB), up until the critical coupling position of ∼9.0 µm. Past this point, the extinction ratio starts to fall once again, as some light transferred to the ring couples back to the bus waveguide. The process of annealing is also found to reduce the ring's effective length slightly, thereby blue-shifting the resonances, despite the ion implantation only changing the refractive index of the gap between the ring and bus waveguide. This resonance shift is quasi-linear with a gradient of 0.033 ± 0.001 FSR/µm and continues past the critical coupling position.

Next, we experimentally test localised laser annealing performance using a pulsed laser source. Raman spectroscopy of some test structures reveal that the recrystallisation is incomplete, and instead, the silicon is found to be in a polycrystalline form. There is a clear peak around 520 cm^{-1}, however, the relative intensity is very low, and the peak width is too large for completely crystalline silicon. Nevertheless, the pulse laser annealing was found to increase the extinction ratio by >11 dB at its critical coupling length of 9.3 ± 0.2 µm. Resonance shift followed the same quasi-linear trend, but at a greater gradient compared to simulations of 0.050 ± 0.002 FSR/µm. This is attributed to some ion implantation occurring within the ring itself, where it will have a significantly greater impact on the ring's effective length. Continuous-wave laser annealing resulted in full silicon recrystallisation and the recovery of an extinction ratio of greater than 35 dB for an annealing length of $\sim 9.0 \pm 0.5$ µm [161]. The ion implantation of the 10 µm silicon slab layer region only resulted in a small increase of additional loss of <0.2 dB compared to a reference un-implanted racetrack resonator.

Despite its more inferior material quality after silicon recrystallisation, localised annealing using the pulsed laser source provides more precise information about the required annealing length than the continuous wave laser annealing. Additionally, pulsed laser annealing can be used for gradient feedback real-time resonant wavelength trimming of optical components [160], with the full tuning of devices carried out in <1 min per device.

Chapter 6

Optimisation of Refractive Index Perturbation Patterns for Advanced Photonic Functionalities

Programmable Photonic Integrated Circuits (PICs) will likely be required for emerging new fields of photonics-based quantum, neuromorphic and analogue computing [147, 172–178]. Recent years have seen numerous breakthrough studies in the field; for example, a programmable nanophotonic processor for quantum transport simulations by Harris et al. [5], spoken vowel recognition in an optical Neural Network (NN) photonic circuit by Shen et al. [3] and an optical NN realised using non-volatile Phase Change Material (PCM) capable of detecting letters by Feldmann et al. [4]. Such reconfigurable photonic networks are typically achieved via meshes of many individual Mach-Zehnder Interferometers (MZIs) used to control phase and amplitude information; however, this geometry results in very large device footprints and cascading losses [148, 149]. Micro-Electro-Mechanical System (MEMS) is another very promising platform for reconfigurable circuits, but also has footprint constraints and cascading loss limitations [179].

More advanced photonic functionality can be packed into a denser region through 'inverse design', the process of designing a device for a specific pre-defined functionality. Inverse design problems typically cannot be solved analytically, and usually require approaches such as brute-force, topological optimisation or evolutionary algorithms [180–182]. More recently, machine learning has proven to be a potent tool for data processing, image recognition and solving other problems of enormous complexity [183, 184]. Various applications have been reported in the field of nanophotonics, such as phase recovery in optical microscopy [185], optical characterisation and classification of nanometer-scale specimens [186–188] and the real-time interpretation of light scattering through complex

media [189–191]. Deep-Artificial Neural Networks (ANNs) have also been demonstrated to be capable of being trained on the solutions of inverse problems, thereby learning to rapidly design new complex nanophotonic devices [192–198].

The optimisation of PICs with complex functionalities is being made widely available, with many highly advanced open-access software packages being added to online repositories; such as MetaNet, a collection of various photonic design algorithms [199]. The table below outlines some of the currently available packages on the MetaNet repository.

Software	Group	Description	Ref.
GLobal Optimization NETwork (GLOnet)	Jonathan Fan, Stanford University	Generative neural network for optimising ensembles of highly efficient topology-optimised metasurfaces	[200]
Metagrating Topology Optimization	Jonathan Fan, Stanford University	Adjoint-based gradient descent topology optimisation for optimisation of periodic metasurfaces	[201]
Angler	Shanhui Fan, Stanford University	Adjoint nonlinear gradient descent method for simulating linear and nonlinear devices in the frequency domain, and adjoint-based inverse design	[202]
Wavetorch	Shanhui Fan, Stanford University	Recurrent Neural Network (RNN) time-domain scalar wave equation solver for inverse design and optimisation of photonic devices	[203]
LumOpt	Eli Yablonovitch, University of California, Berkeley	Adjoint-based optimisation for inverse design of photonic devices using Finite-Difference Time-Domain (FDTD) simulations	[204]
SPINS-B	Jelena Vucković, Stanford University	Adjoint-based gradient descent optimisation for inverse design of photonic devices	[205]

Wavefront shaping has been demonstrated as an effective method for characterising and modulating light in complex media [206–210]. Previous sections of this work have already shown that even a single small-magnitude refractive index perturbation can significantly impact the light within PICs. One can imagine that significantly more advanced functionalities can be achieved through the application of multiple scatters within such systems. The principle is simple, given fine enough control and sufficient

design space, full control over the light's wavefront can be attained. This approach was demonstrated in a recent paper from Bruck et al. for an all-optical spatial light modulator [112], in which Digital Micromirror Device (DMD) was used to project an optical perturbation pattern onto a 1×2 Multimode Interference (MMI) device with >97.5% of light transmitted to a single output port. The multiple perturbations alter the mode distribution and with the correct geometry can shape the wavefront of the light, effectively guiding the light to a particular output. Storing optimal patterns for each output allows for rapid switching, that is theoretically only limited by the device's recovery time. In Section 3.4.4, it was shown that the full recovery of a silicon waveguide is sub-nanosecond, which corresponds to a switching speed in the gigahertz regime.

In this chapter, we will discuss various optimisation approaches for refractive index perturbation patterns for different applications. It starts with simple brute-force iterative optimisation of single function devices, and ultimately leads to the development of an ANNs capable of rapidly generating patterns for arbitrary complex electric field outputs for a given MMI region. The capability of such patterns is then demonstrated experimentally through shallow etched perturbations in a 1×2 MMI device, easily fabricatable using standard industrial lithographic techniques.

6.1 Iterative optimisation of perturbation patterns

The iterative approach is a brute-force optimisation technique used to calculate perturbation patterns. A flow chart diagram of the process used is shown in Figure 6.1. The model is initialised with the structure's physical parameters, and the unperturbed device is simulated to obtain the initial coupling. This coupling can be optimised for any parameter, e.g. intensity, wavelength or mode, depending on the desired functionality. The first perturbation is applied to the model and then simulated to obtain the new coupling parameter. If the coupling is improved, the perturbation is retained; otherwise, it is removed once again. The next perturbation is then applied in the same manner, and the process is repeated until all perturbation positions have been tested or some stop condition is met. Finally, the optimal perturbation pattern is simulated for both the forward and reciprocal cases.

The order in which the perturbations are applied can either be ordered or randomised. For the ordered case, the top left perturbation is chosen to be applied first, and positions are tested column-wise along the device's propagation axis, where only perturbations that improve the desired coupling are kept. This approach yields repeatable patterns but is prone to becoming trapped in local minima, i.e. a pixel that initially improves coupling may later actually be deteriorating it. Even just reversing the starting position from the top to the bottom of the device can significantly affect the final pattern's performance. To help alleviate this either multiple passes can be used, where previously

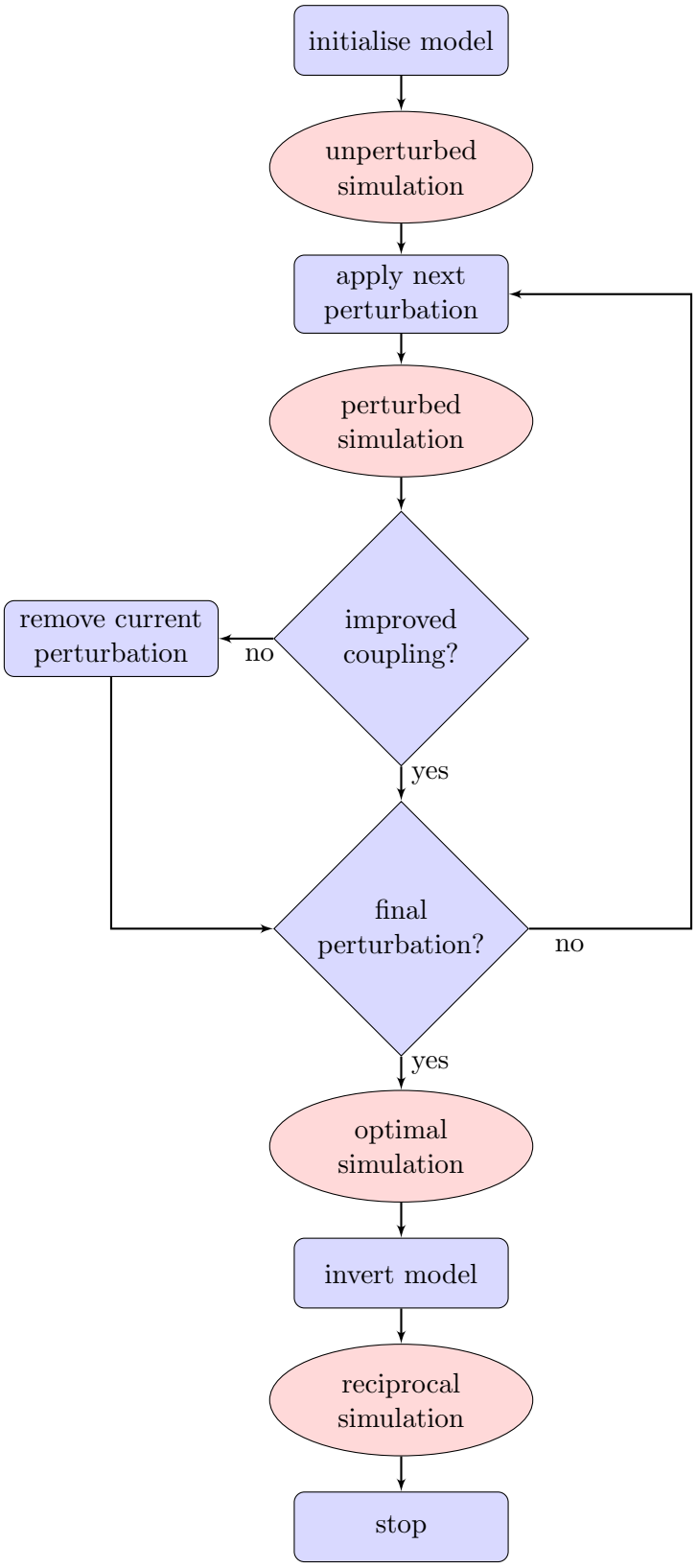


Figure 6.1: Flow chart of the iterative refractive index perturbation pattern optimisation process. Blue rectangles indicate actions, blue diamonds indicate decisions and red ellipses indicate simulation runs.

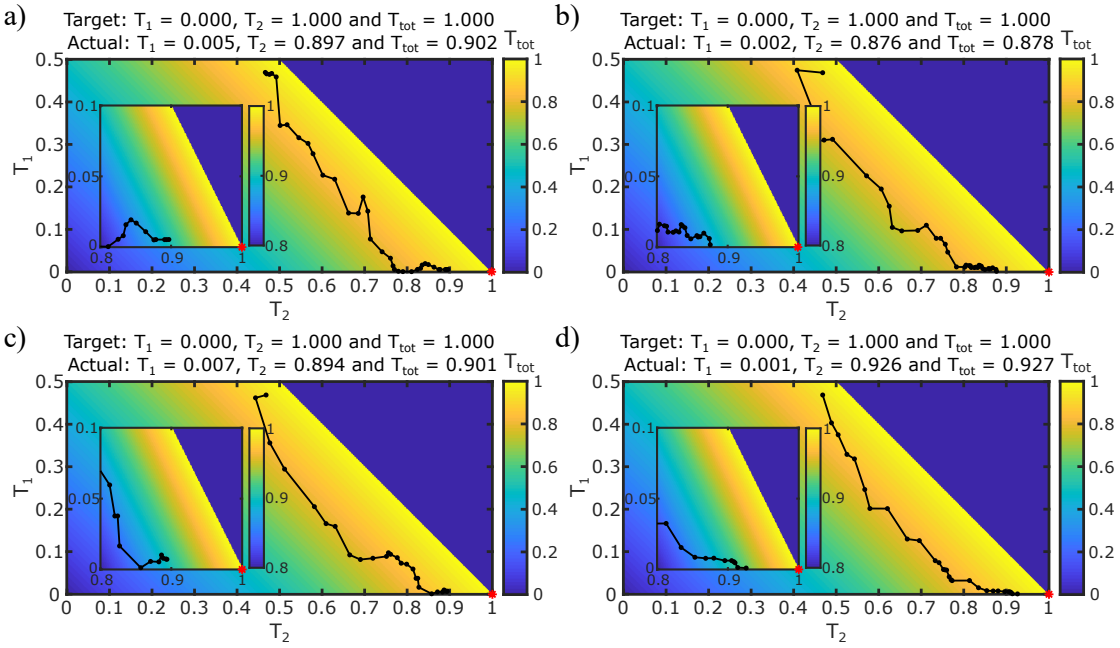


Figure 6.2: Iterative perturbation pattern optimisation process for single output port. a) Ordered and b–d) randomised optimisation of the T_2 output port of a 1×2 MMI device for 750 nm sided square non-overlapping perturbations with $\Delta n_{\text{eff}} = -0.25$. Colour maps show the value of the total transmittance, T_{tot}, at each position, the red asterisks mark the target port transmittance of $T_2 = 1$ and the black dots indicate a perturbation being accepted during optimisation. Inserts show a magnified region around the target port transmittance. Optimised using the in-house aperiodic-Fourier Modal Method (a-FMM) solver for the fundamental Transverse Electrical (TE) input mode at a wavelength of 1550 nm.

accepted perturbations can be removed, or some degree of randomness can be introduced to the optimisation process. The level of randomness can itself be varied; for example, the entire perturbation position order can be fully randomised or randomised only row- or column-wise. However, computational time scales linearly with the number of passes and random runs. The more randomness introduced into the model inevitably results in more variance on the device performance between subsequent runs.

Figure 6.2 illustrates a selection of optimisation runs for maximising the transmission to a single output port (T_2) of a 1×2 MMI device for both the ordered and randomised positioning of scatterers. In general, the final device's performance is very similar to the target transmittance values, typically within a few percent. However, it is worth noting that the ordered case does not correspond to the highest coupling. The optimisation routes are broadly similar for all the examples, with the total transmission of the device, indicated by the colour map, remaining relatively constant and optical power being transferred from one output to the other. A variety of different photonic devices are investigated using this iterative approach in the following subsections.

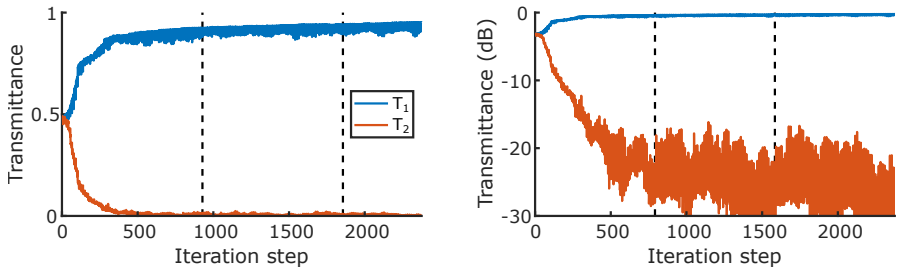


Figure 6.3: The optimisation of mode transmittance for T_1 (top output) of a 1×2 MMI optical router after three iterative passes (indicated by the black dashed lines). Perturbations are 500 nm sided squares with $\Delta n_{\text{eff}} = -0.25$. Simulated using Lumerical FDTD solver for the fundamental TE input mode at a wavelength of 1550 nm.

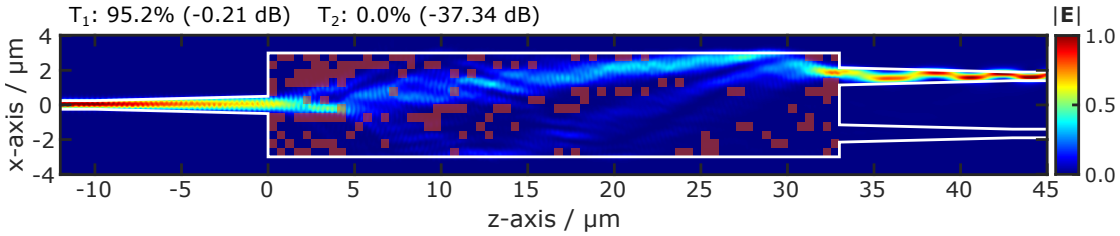


Figure 6.4: The electric field distribution for the optimised 1×2 MMI optical router perturbation pattern for T_1 (top output). Perturbations are 500 nm sided squares with $\Delta n_{\text{eff}} = -0.25$. Simulated using Lumerical FDTD solver for the fundamental TE input mode at a wavelength of 1550 nm.

6.1.1 Optical router

The first device investigated is a basic optical router that guides light from an input port to a particular output. The light can be dynamically switched between outputs by changing the applied perturbation patterns.

Figure 6.3 shows the simulated optimisation of an 1×2 optical router based on a MMI device with dimensions of 6 µm by 33 µm, using the ordered-positioning iterative approach. The perturbation size was chosen to be 500 nm with $\Delta n_{\text{eff}} = -0.25$, as this is approximately the diffraction-limited spot size for blue light with an index shift that has been experimentally demonstrated. However, the perturbations could be modelled with whatever dimensions, shape or magnitude one desires. During this optimisation, only the transmittance to the top output, T_1, is investigated, as the device symmetry allows a simple flip of the perturbation pattern about the propagation axis to recover the same response in the other port. Three passes of the device were carried out, providing the opportunity to remove perturbations on subsequent passes; however, the vast majority of the transmittance optimisation occurs during the first pass. The final optimised pattern and its corresponding electric field distribution are shown in Figure 6.4. The simulated coupling to the top output port is found to be 95.2 % (−0.21 dB) with a less than −35 dB

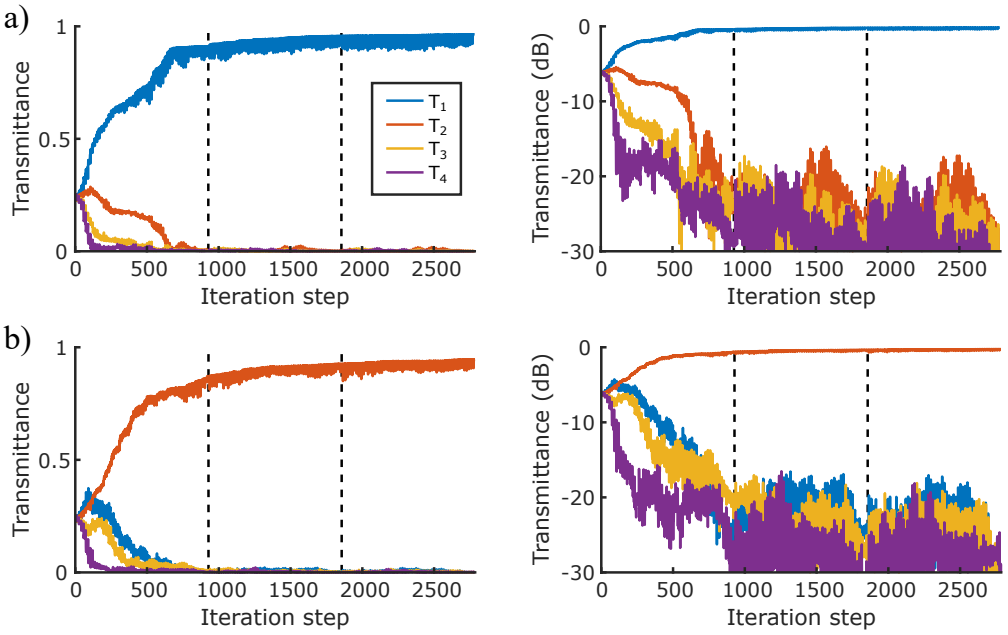


Figure 6.5: The optimisation of mode transmittance for a) T_1 (top output) and b) T_2 (second-top output) of a 1×4 MMI optical router after three iterative passes (indicated by the black dashed lines). Perturbations are 500 nm sided squares with $\Delta n_{\text{eff}} = -0.25$. Simulated using Lumerical FDTD solver for the fundamental TE input mode at a wavelength of 1550 nm.

crosstalk to the other output. It was found that increasing the perturbation size to 1 µm the coupling was reduced to 85.2 % (−0.69 dB) with crosstalk less than −25 dB, but the simulation is four times faster to complete.

The same iterative process was then applied to a 1×4 MMI device with dimensions of 8 µm by 29 µm. The transmittance optimisation to the top, T_1, and second-top, T_2, is shown in Figure 6.5. The same response for the other two ports can be obtained via the exploitation of the device's symmetry. Once again, three passes of the device were carried out, and the majority of optimisation was found to occur in the first pass, with only a few percent of coupling improvements from subsequent passes. The optimised patterns and their electric field distributions, shown in Figure 6.6, were found to have 96.5 % (−0.16 dB) and 94.7 % (−0.24 dB) for the top and second-top output couplings, respectively. Crosstalk in less than 30 dB in all cases. Increasing the perturbation size to 1 µm again results in reduced coupling for the top output of 89.7 % (−0.47 dB) and increased crosstalk of around 20 dB.

6.1.2 Mode converters

Up until this point, we have focused primarily on the transmission of a fundamental input mode to a fundamental output mode; however, refractive index patterns have also been shown to be capable of producing (de)multiplexers [211] and spatial mode

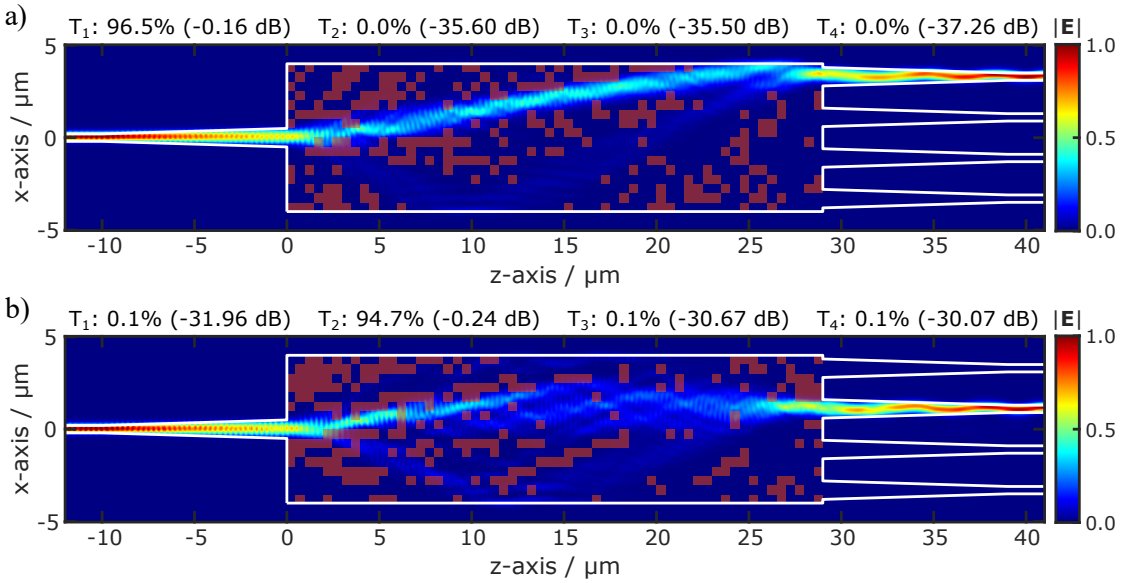


Figure 6.6: The electric field distribution for optimised 1×4 MMI optical router perturbation patterns for a) T_1 (top output) and b) T_2 (second-top output). Perturbations are 500 nm sided squares with $\Delta n_{\text{eff}} = -0.25$. Simulated using Lumerical FDTD solver for the fundamental TE input mode at a wavelength of 1550 nm.

converters [212]. Mode converters are a critical class of component frequently used in integrated photonic circuits that operate in polarisation diverse settings, i.e. both TE and Transverse Magnetic (TM) modes are present, such as coherent transceivers [213, 214]. Due to the birefringent nature of most PICs, it is usually much simpler to split and rotate one polarisation state than to design an entire circuit that is well loss- and phase-matched. Higher-order spatial modes have also been suggested to encode more information within the same optical bandwidth [68, 215, 216]. Several different design geometries have been investigated as mode converters, the most common of which are based on Directional Couplers (DCs) or MMIs [217].

The mode converter designed here is based on a 1×1 MMI device with dimensions of 6 µm by 66 µm, with 500 nm perturbations. The input and output waveguides are 1.1 µm wide, which is wide enough to support the first three TE modes. Figure 6.7 shows the optimisation of different TE mode conversions using the iterative approach with three passes. The resulting optimised patterns are simulated for both the forward and reciprocal cases, as shown by the electric field distributions in Figure 6.8. Simulated performance is very good with < -1 dB coupling losses and < -30 dB crosstalk in all cases. Due to the forward propagation's high coupling operation, it was found that the patterns were also very efficient in their reciprocal cases with similar coupling values, with only slightly increased average crosstalk. Perturbations with dimensions of 1×1 µm resulted in only around a -2 dB coupling efficiency, with crosstalk values of approximately -20 dB. The various modes can be separated out of the device through the use of a mode splitting taper [218].

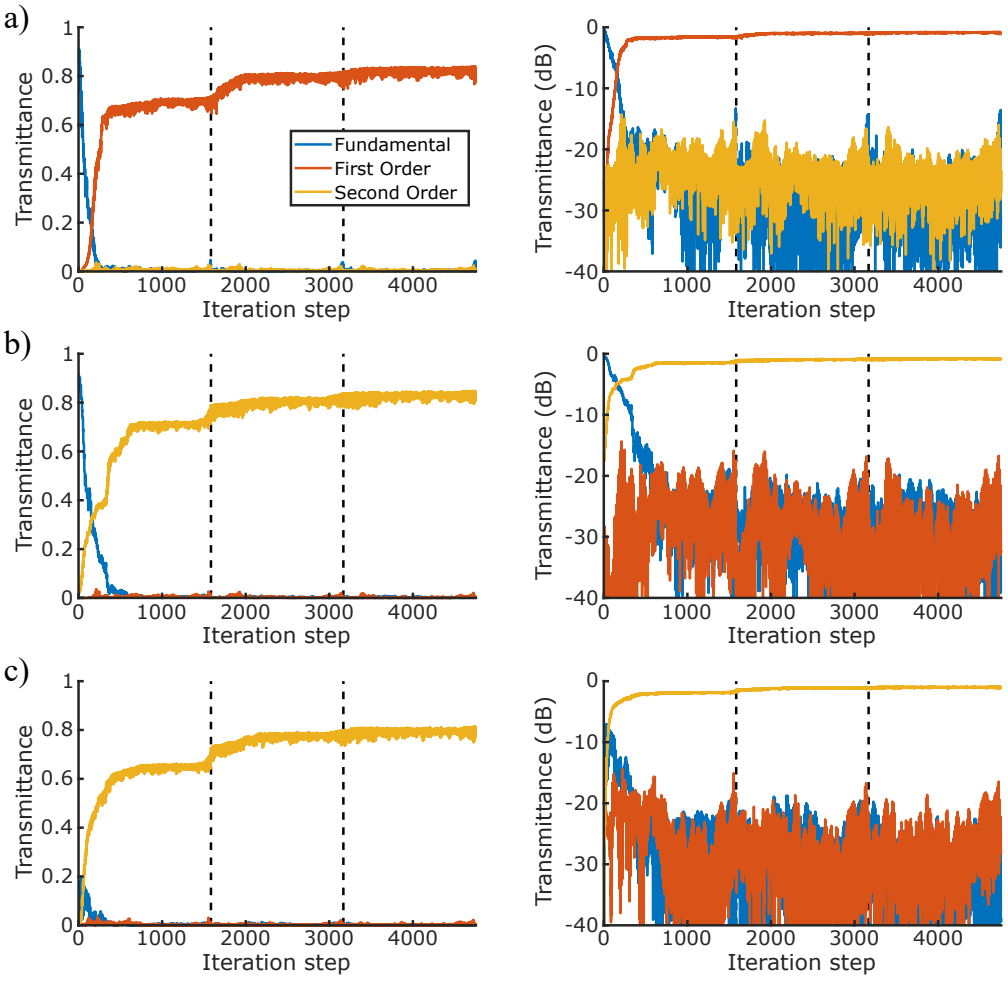


Figure 6.7: The optimisation of mode transmittance for a) TE_0-to-TE_1, b) TE_0-to-TE_2 and c) TE_1-to-TE_2 for three iterative passes (indicated by the black dashed lines). Perturbations are 500 nm sided squares with $\Delta n_{\text{eff}} = -0.25$. Simulated using Lumerical FDTD solver for TE input modes at a wavelength of 1550 nm.

6.1.3 Multi-objective optimisation

The optimisation function, f, used to generate arbitrary splitting ratio perturbation patterns for a 1×2 MMI device is shown below

$$f = 1 - \sqrt{\left(\frac{T_1^{\text{tar}}}{T_2^{\text{tar}}} - \frac{T_1^{\text{act}}}{T_2^{\text{act}}}\right)^2 + (T_1^{\text{tar}} - T_1^{\text{act}} + T_2^{\text{tar}} - T_2^{\text{act}})^2}, \tag{6.1}$$

where T is the transmission for the port given by the subscript, and the superscript denotes if it is the target or actual value. This equation optimises both the splitting ratio, $R = \dfrac{T_1}{T_2}$, and the total transmission, $T_{\text{tot}} = T_1 + T_2$, with equal weighting. The acceptance condition for a perturbation is an increase in f, with an additional extra threshold amount (typically 0.1%) to prevent numerical noise accepting perturbations.

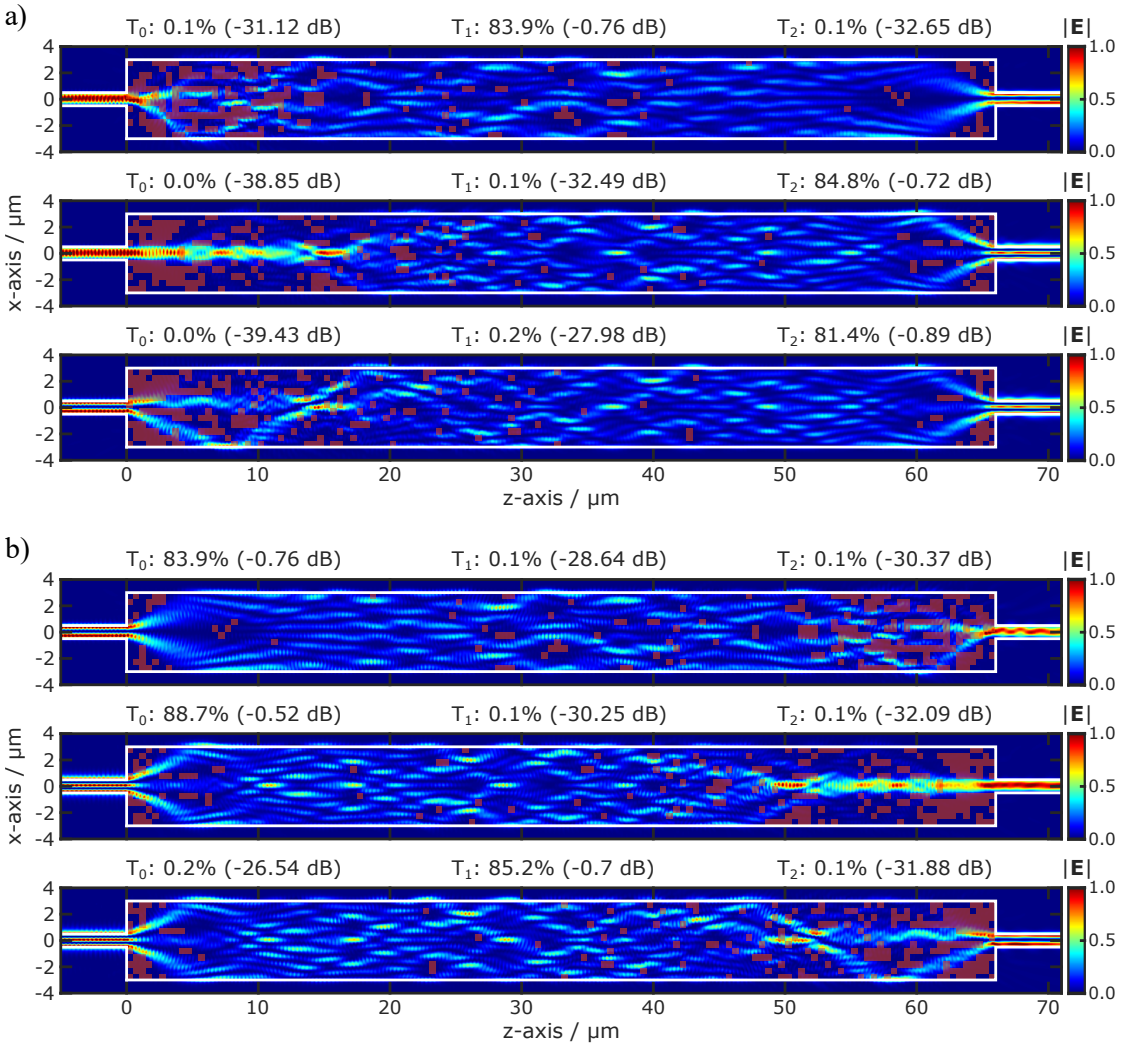


Figure 6.8: a) The electric field distribution for optimised mode converter perturbation patterns for TE_0-to-TE_1 (top), TE_0-to-TE_2 (middle) and TE_1-to-TE_2 (bottom) after three iterative passes. Perturbations are 500 nm sided squares with $\Delta n_{\text{eff}} = -0.25$. b) The corresponding reciprocal mode conversion electric field distributions from a). Simulated using Lumerical FDTD solver for TE input modes at a wavelength of 1550 nm.

Figure 6.9 shows the optimisation of a selection of randomised splitting ratios with total target transmittance values of 100%, 90% and 80%. The colour map corresponds to the value of f for any given position in transmission parameter space. In general, the multi-objective optimisation works well with target transmission values and splitting ratio being achieved for all but the 100% total target transmittance cases, where the optimisation comes up against the physical limits of the system. Even for the 50:50 target, Figure 6.9 g), the total transmittance is only increased slightly ($\sim 2\%$) compared to the unperturbed device.

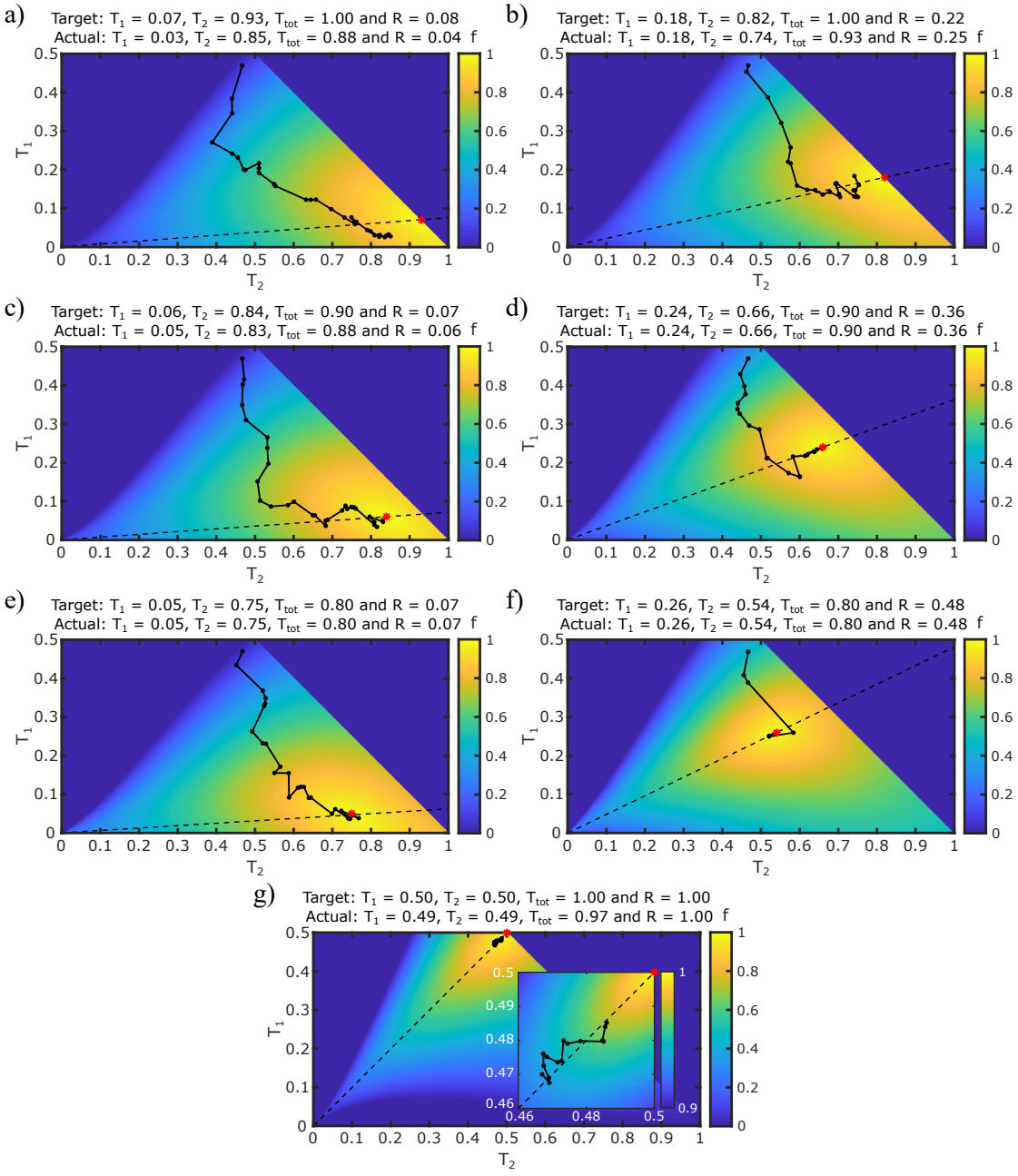


Figure 6.9: Iterative perturbation pattern optimisation process for variable splitting ratios of a 1×2 MMI device for 750 nm sided square non-overlapping perturbations with $\Delta n_{\text{eff}} = -0.25$. Randomised perturbation order and target port transmittance values were used for a–b,g) no loss, c–d) 10% loss and e–f) 20% loss. Colour maps show the optimisation function's value, f, at each position and the black dots indicate a perturbation being accepted during optimisation. All port transmittance values with the target splitting ratio, R, are shown by the black dashed line, and the target port transmittance values are indicated by the red asterisk. Insert in g) is a magnified region around the target port transmittance values. Optimised using the in-house a-FMM solver for the fundamental TE input mode at a wavelength of 1550 nm.

6.2 Deep artificial neural network optimisation

ANNs are a relatively recent approach to solving problems which are traditionally very difficult for computers to solve numerically due to their large optimisation parameter space [183, 184]. A neural network can be trained on only a relatively small subset of solutions to form a set of weight parameters capable of describing a physical system. These weight parameters are applied to the inputs of nodes, often referred to as neurons, the output of which is ultimately decided by some user-defined activation function. Layers of these neurons are then built up to produce the so-called 'deep-ANN' that is capable of predicting the response of a system for a given input. Of particular interest here, deep-ANNs with many convolutional layers have been used for image recognition [219, 220]. Networks based on the same principle have been demonstrated to be capable of predicting both the response of complex nanophotonic devices, as well as their inverse design for a specific target function [200, 203]. Inverse design problems are generally not analytically solvable, and usually require approaches, such as brute-force, topological optimization, evolutionary algorithms or other computationally expensive techniques [201, 202, 205, 221].

While we have previously shown in this chapter that it is relatively simple to optimise a single coupling parameter using iterative brute-force approaches, this becomes increasingly complex as the problem becomes multi-objective. Additionally, poor computational scaling with the number perturbations considered rapidly render such approaches untenable with increasing design space, and impossible for any real-time applications. Even for the most simple case of a binary perturbation pattern, the problem contains 2^P possible solutions, where P is the number of perturbation positions. Considering a $6 \times 33\,\mu m$ MMI region with 750 nm sided square non-overlapping perturbations, as we do here, this represents 2^{352} or 10^{106} possible configurations — a number that renders a systematic search for specific MMI designs impossible.

The development of the networks used in this chapter was carried out in conjunction with Dr Peter Wiecha [198]. A combination of two separate networks was decided upon in a predictor-generator configuration. The first is the forward network which acts as a 'physical predictor', taking a 2D perturbation pattern as input and returning the predicted 1D output intensity profile. While the second is the generator network and essentially acts as the inverse of the forward network, in which the desired 1D intensity output profile is given as input, and it produces a perturbation pattern for the given target profile. The networks are based on a 2D-to-1D encoder-decoder residual neural networks, referred to as "ResNets", originally designed for image recognition [222, 223]. There has been much development of these kinds of networks in recent year and they can be directly applied to the problem posed here, where the 'image' is the perturbation pattern and the output is the light intensity field profile.

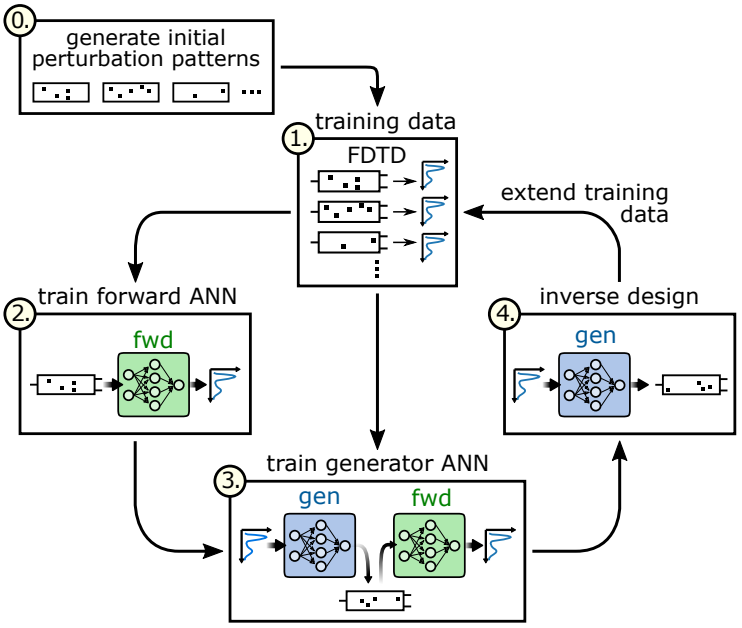


Figure 6.10: Artificial Neural Network (ANN) training scheme.

The complete training process is then broken down into four main steps; the simulation of training data, the forward network's training, the training of the generator network and the testing of the inverse design against simulation, as illustrated in Figure 6.10. This training scheme can be iterated through multiple times to optimise the network's performance and converge on an accurate model of the physical system by specifically targeting previous iteration's weaknesses. Two MMI devices were chosen to be investigated using this approach; a 1×2 with shallow etched perturbations and a 3×3 with weak optical perturbations. Both share the same size for the perturbation of 750 nm and MMI region dimensions of 6 µm by 33 µm. The centre-to-centre waveguide spacing is 3 µm and 2.15 µm for the 1×2 and 3×3 devices, respectively. This breaks the MMI region down to an 8×44 grid corresponding to all the possible non-overlapping positions. The input and output waveguides are single-mode with a width of 500 nm that taper up to 1 µm wide at the MMI region boundary over a length of 10 µm, thereby allowing the fundamental TE mode's adiabatic expansion and reducing the transmission losses of the device. It is important to note that the unperturbed 1×2 MMI performs well as a 50:50 power splitter; however, the 3×3 device does not perform as a natural three-way power splitter. This renders the design-problem even more challenging but maintains the compact device footprint.

6.2.1 1×2 MMI device with shallow etched perturbations

The first device chosen to be investigated was a 1×2 MMI device with etched 750 nm sided square perturbations. The perturbations were modelled as a shallow etch of 120 nm. This is conveniently the same etch depth typically used when fabricating rib

structures from 220 nm Silicon-On-Insulator (SOI), thereby simplifying the fabrication of such devices and eliminating the need for additional alignment steps. This results in an effective index change of $\Delta n_{\text{eff}} = -0.71$. The unperturbed device has a simulated transmittance of 46.5% for each of the output ports.

Initial training dataset patterns were produced using the in-house aperiodic-Fourier Modal Method [95] (see Section 3.2.2), as shown in step 0 of Figure 6.10. The a-FMM technique relies on a scattering-matrix formalism, in which the transmittance coupling between a set of input and output modes for an MMI device can be calculated accurately and rapidly [98]. Optimisation of the patterns was carried out using the randomised iterative approach outlined previously in Section 6.1, in which the fitness function (Equation 6.1) is maximised. The target transmittance values were either chosen to optimise coupling to a single output port or for an arbitrary splitting ratio, as shown in Figure 6.11. Perturbations were then applied in a randomised order of positions for the grid. If the cost function is reduced for a given perturbation, it is accepted and its position stored; otherwise, it is removed. The next perturbation is then applied to the following randomised position, and the process is repeated for every location in the grid. To reduce the number of duplicate patterns during a particular optimisation (referred to as a 'family') the pattern is only saved for each accepted perturbation after the initial one. A total of 2500 families were produced for the initial training dataset, with up to a maximum of 30 accepted perturbations in each family.

Next, in step 1 of Figure 6.10, each accepted pattern was simulated in 2D using the Lumerical FDTD solver to obtain the complex electric field profiles at both the MMI-output interface and across the single-mode output waveguides at a distance of 20 µm from the MMI output interface. FDTD was chosen instead of a-FMM because it allows closer comparison the experimental device, a-FMM cannot efficiently model the input and output tapers, and faster computational time to calculate the complex electric field output profiles. The simulations were carried out with the same parameters as for the initial dataset generation, with the incident light at a wavelength of 1550 nm and in the fundamental TE mode. Appendix B.3 demonstrates the validity of the 2D simulation approach, with a slight decrease in performance in 3D being explained by the inclusion of out-of-plane scattering. The simulated dataset was then randomly split into two separate groups; training and validation, with approximately a 9:1 ratio. Specific families of patterns from the same optimisation were kept together to avoid excessively similar patterns occurring in both the training and validation datasets. Finally, both datasets were doubled in size by exploiting the symmetry of the device about the propagation axis, and it also helps to teach the networks the symmetry of the system explicitly. This was done by flipping both the pattern and output profile in the x-axis, taking into account the complex electric fields' sign.

The forward network was trained on the training dataset (step 2 of Figure 6.10) with the validation dataset being reserved to benchmark each epoch's true loss. The validation

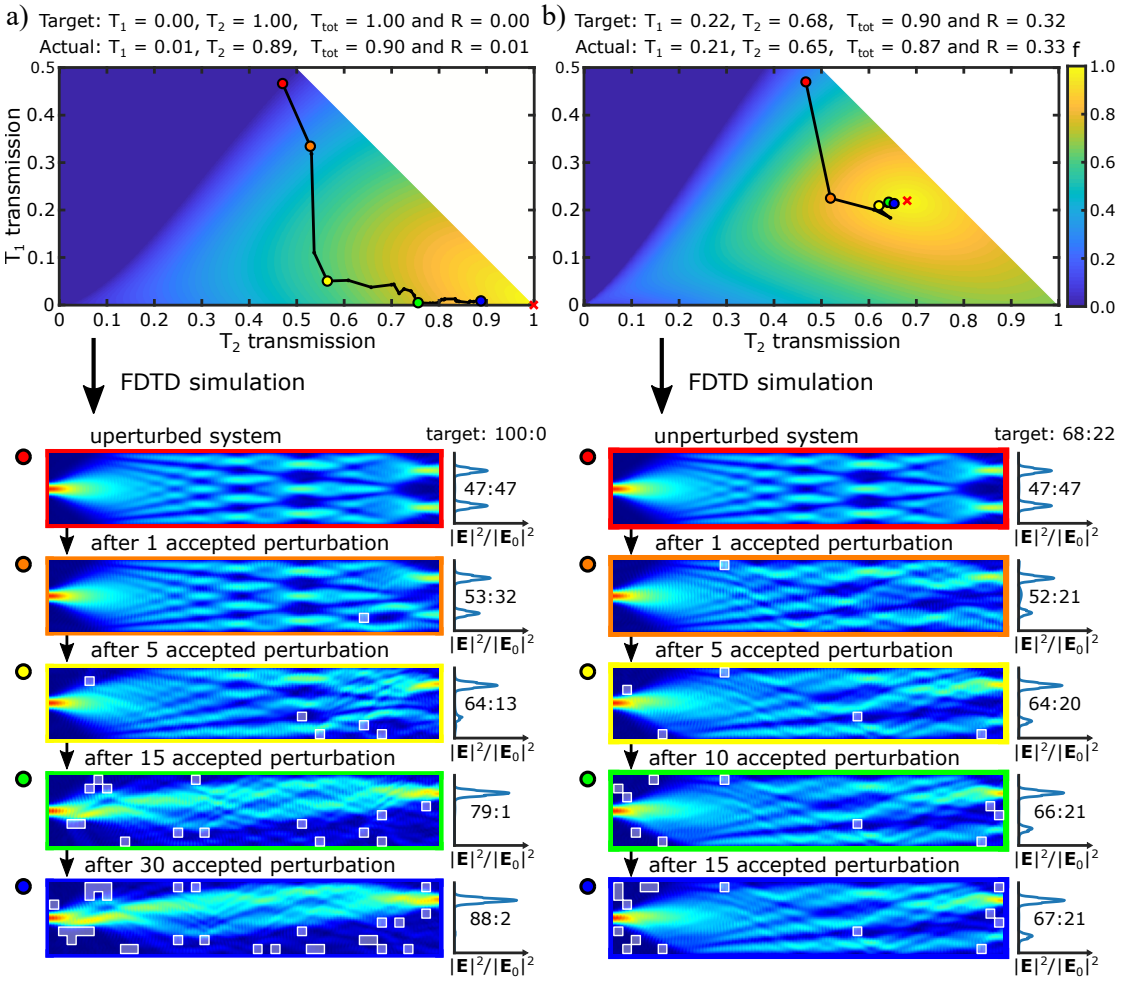


Figure 6.11: Example of initial dataset generation for optimising coupling to a) a single output port and b) a specific target splitting ratio, R. The top section shows the iterative perturbation pattern optimisation process using the in-house a-FMM model. The coloured circles indicate the FDTD simulated patterns below, which show the perturbation pattern and its corresponding electric field distribution and output intensity profile.

dataset is kept unseen by the network during training and, therefore, it provides the true performance value and ensures the network is not merely overfitting to the initial data. The forward predictor network's structure is outlined in Figure 6.12 a). Each network is comprised of residual down- and up-sampling convolutional blocks and allows efficient training convergence to be maintained even for very 'deep' network layouts. The encoder takes the 8×44 perturbation pattern, or 'image', and uses 2D convolutional downsampling to extract the MMI features. This intermediate fully connected section acts as a dense interconnect and transitions between 2D and 1D data. The decoder then uses 1D convolutional upsampling to obtain an output profile. All the blocks here use the leaky Rectified Linear Unit (ReLU) activation function, a simple $y = x$ function for positive values of x with some significantly reduced gradient for negative values, with the exception of the network's output which uses a Sigmoid activation function. This

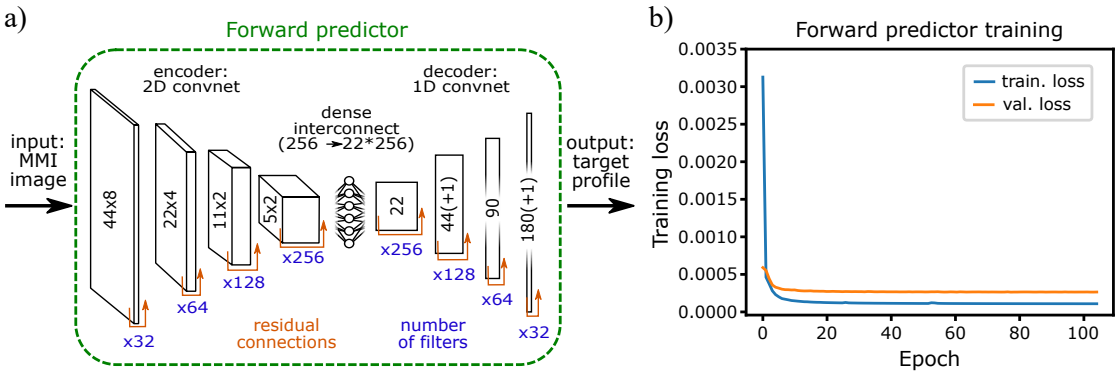


Figure 6.12: a) Forward predictor network structure. b) Example of forward network training and validation Mean Squared Error (MSE) losses.

constrains values between 0 and 1, as physically required by the $|\mathbf{E}|^2$–profile. The "(+1)" in the decoder blocks represent zero-padding, which is necessary to reach the correct dimensionality when scaling up from the dense interconnect to the output. As only the intensity profile for the output waveguides is considered here, the network returns a single output profile. Multiple network channels can be used to allow the prediction of the full complex output profile, where each channel represents one component of the complex electric field, as will be demonstrated later. However, training on the complete complex fields was found to slightly reduce the network's performance here, which is attributed to the limited design space and the prioritisation of phase information even for low-intensity outputs.

The network is then trained for 120 epochs using increasing batch size, which has been shown to lead to significantly faster convergence for ResNets than traditional decaying learning rate methods with little to no loss of accuracy [224]. Batch size is initially set to 32 and doubled every 30 epochs with a fixed learning rate using the 'ADAM' solver [225]. The MSE network training loss is calculated after each epoch and characterises the difference between the network prediction and simulation results. The training and validation losses during training are shown in Figure 6.12 b). After the forward network is fully trained, its performance statistics are calculated using the validation dataset comparing the network prediction error against the number of perturbations. Using a consumer NVIDIA GeForce GTX 1080 Ti GPU, the average prediction time was found to be <0.2 ms per pattern.

Step 3 of Figure 6.10 is the generator ANN training on the same dataset. The network's structure is nearly identical to the forward predictor but reversed, as shown in Figure 6.13 a). Now, the encoder takes a 1D target intensity profile as the input, and the decoder outputs an 8×44 image of the perturbation pattern. The generated designs are compared to the training patterns to yield an MSE loss value; however, it has been shown that this pattern loss alone is not sufficient for a "many-to-one" problem [192, 226]. This is because significantly different patterns can result in nearly identical output intensity profiles, resulting in instability during network training. To help alleviate this problem,

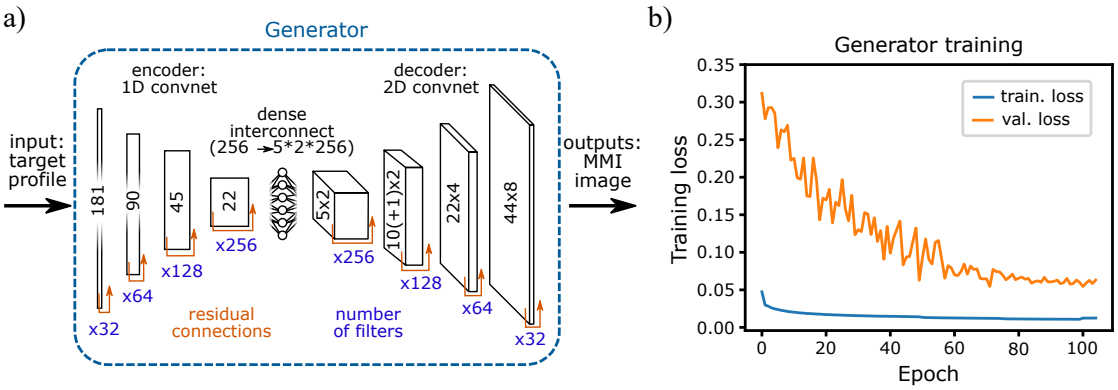


Figure 6.13: a) Generator network structure. b) Example of generator network training and validation MSE losses.

the predicted pattern is fed back into the forward network and a second MSE loss term of the predicted output profile is considered (as shown by step 3 in Figure 6.10). Such tandem network architectures have recently been demonstrated successfully in the inverse design of nanophotonic systems [193, 227, 228]. The total loss was taken as the sum of the two individual MSE loss values and used to benchmark the generator network loss for each epoch against the validation dataset. It was found that a weighting of a factor of 50 for the intensity profile loss compared the pattern image loss yielded the best results. Once again, the network is trained for 120 epochs using increasing batch size method and minimising the MSE total loss value. The total training and validation losses during during the generator's training are shown in Figure 6.13 b).

Once both networks were fully trained, the generator network can be used to inverse design new perturbation patterns (step 4 in Figure 6.10) for a desired target profile. However, the predicted perturbation pattern image contains continuous output values of between 0 and 1 for each perturbation due to the sigmoid activation layer, as shown by the example in Figure 6.14 a). Therefore, some threshold level needs to be applied to convert the greyscale image to a binary map. Various levels of perturbation could be achieved experimentally via different etch heights; however, this would increase fabrication complexity and require additional alignment steps. Due to the rapid prediction times of <0.3 ms per pattern for the predictor network, on a commercial NVIDIA GeForce GTX 1080 Ti GPU, it was decided to test threshold values between 0 and 1 in steps of 0.01. The value with the minimal MSE between the target and predicted profiles is then selected as the binary threshold value. An example of this process is shown in Figure 6.14 b). A more detailed statistical investigation of 150 inverse designed patterns is provided in Appendix B.4. The entire inverse design time per perturbation pattern, including threshold value analysis, is <50 ms and <150 ms on a commercially available Graphics Processing Unit (GPU) and Central Processing Unit (CPU), respectively.

It was found that the initial set of trained networks did not generalise particularly well with mean absolute transmittance errors, $|T_{\mathrm{err}}|$, of ~4% and ~8% between prediction

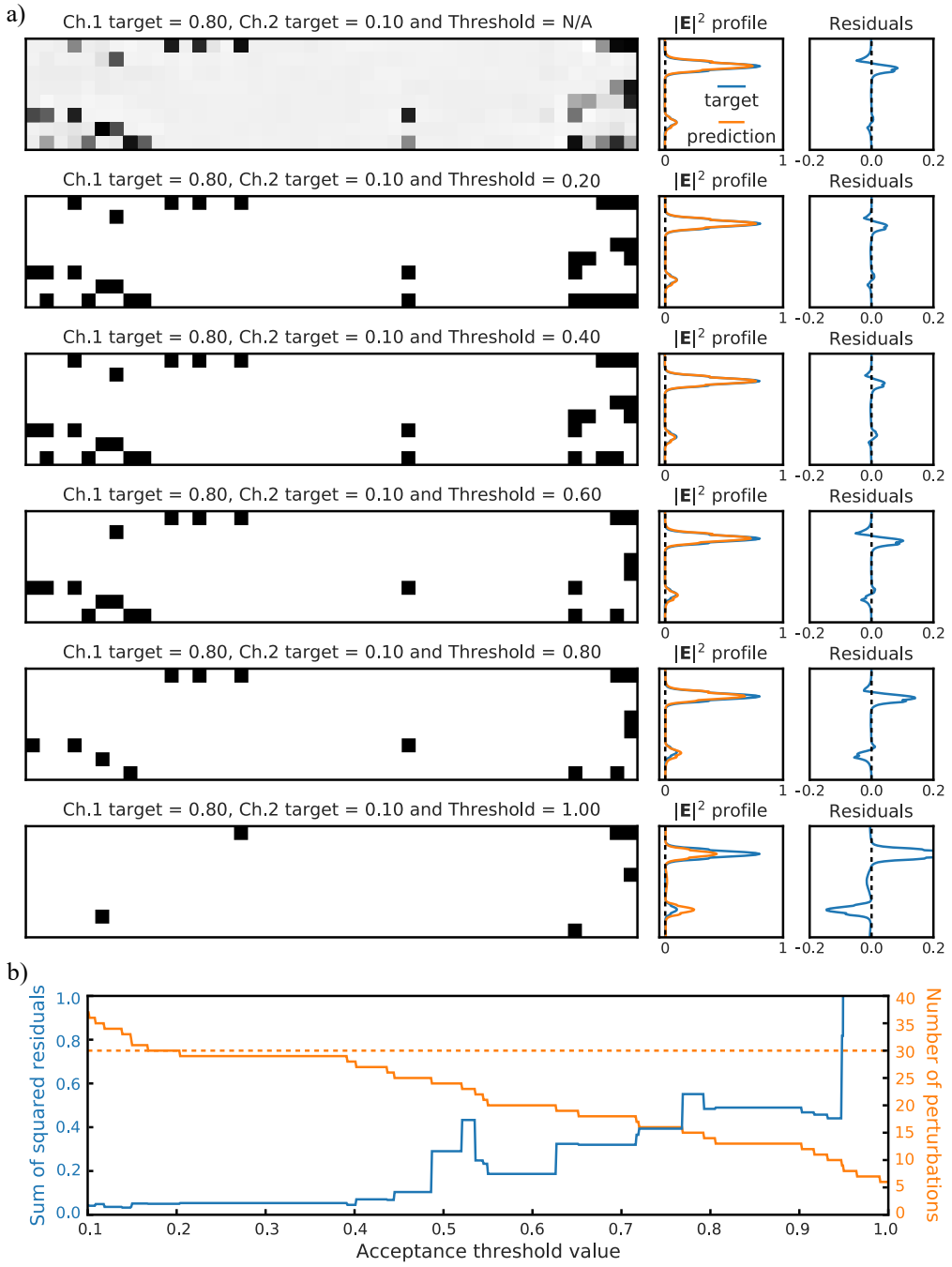


Figure 6.14: Example of the effect of perturbation acceptance threshold on the inverse design pattern for a linear activation layer. a) The predicted perturbation patterns (left column), the forward network predicted intensity profiles (central column) and the residuals between the prediction and target (right column) for a selection of different perturbation acceptance threshold values. The target intensity splitting was set to 80% and 10% in channels 1 and 2. b) The sum of squared residuals between the prediction and target for acceptance threshold values between 0.10 and 1.00 in steps of 0.01 and the corresponding number of predicted perturbations. The orange dashed line represents the maximum number of perturbations considered in the initial training dataset.

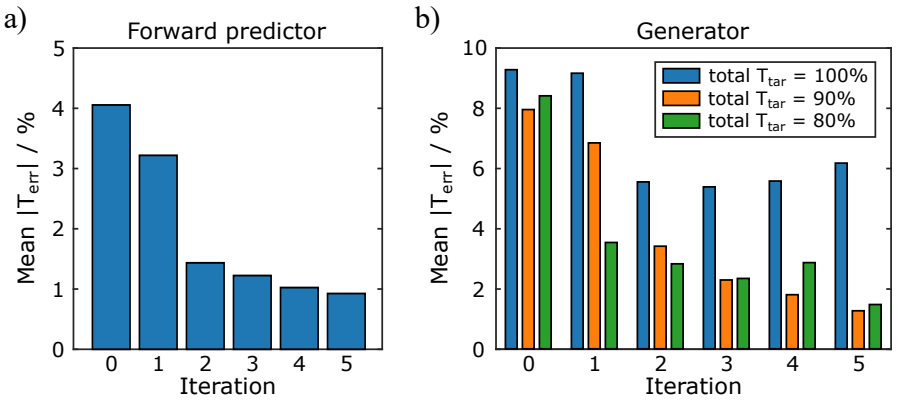


Figure 6.15: 1×2 MMI mean absolute transmittance error between the ANNs and simulations for the consecutive data-generation iterations, for a fixed test dataset. a) Forward network error from prediction against simulation. b) Generator network error of design-targets against simulations of inverse-designed patterns.

and simulation for the validation dataset of the forward predictor and generator networks, respectively. To improve the accuracy of both ANNs, the generator was used to rapidly inverse design a new dataset of 10000 perturbation patterns with an arbitrary splitting ratio and total transmittance target. This was done by generating two random numbers as the targets for each output waveguide and normalising them to the total target transmittance between 80–100%. The newly generated patterns are then be taken back to step 0 in Figure 6.10 and their actual responses simulated using FDTD, thus closing the loop of the optimisation scheme. This new dataset can then be combined with the initial dataset and used to train a new generation of predictor and generator networks. Any features of the system that were previously poorly modelled are essentially fed directly back into the networks, thereby improving its performance beyond just increasing the initial dataset size. This process was repeated a total of five times to end up with six iterations labelled from 0 through to 5, where 0 denotes the initial network.

Finally, the performance for each of the iterations is evaluated by calculating the average transmittance error for a fixed dataset consisting of 300 different perturbation patterns. The target profiles were input into the generator for each of the iterations, and the inverse designed patterns were simulated using the FDTD solver. Figure 6.15 shows the mean absolute transmittance error between both of the networks and simulation for each of the six iterations. There is a clear improvement in subsequent iterations, with less than a quarter of the average error in the final iteration compared to the first for 80% and 90% total transmittance targets. The 100% transmittance target case shows only slight improvement with later network iterations, which due to the physical limit of maximum transmittance that the etched perturbations are able to achieve. In Appendix B.5, the scattering properties are investigated for full 3D FDTD simulations of both a single perturbation in a single-mode waveguide and an example

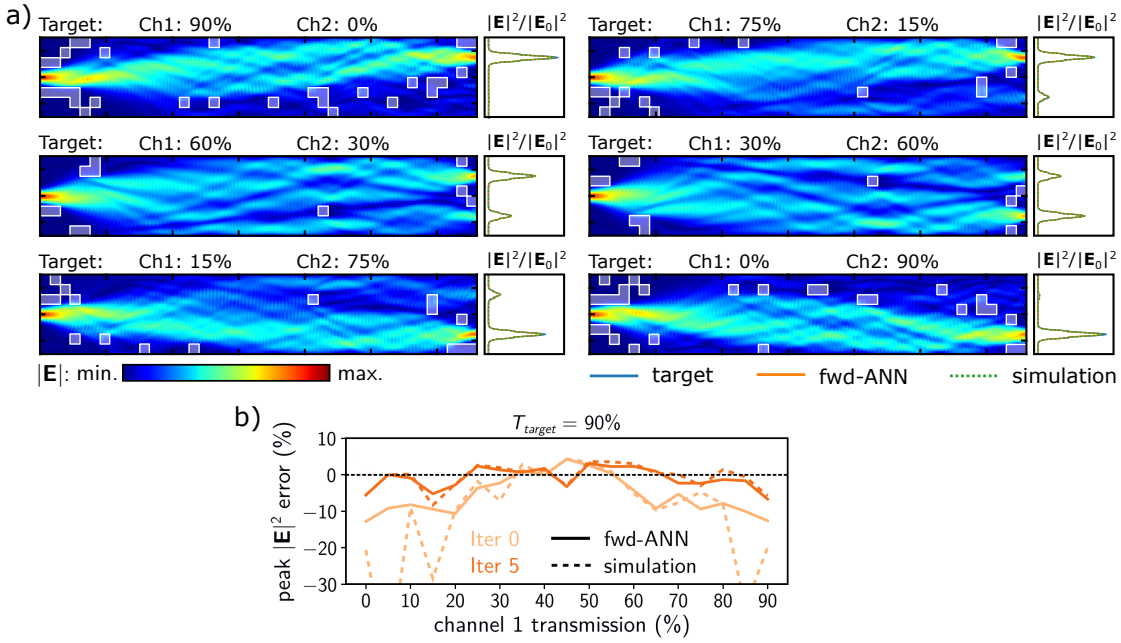


Figure 6.16: a) Examples of inverse designed 1×2 MMI perturbation patterns with a 90% total transmittance target. Generated from the iteration 5 networks. The target transmittance values are listed about the electric field distribution, with different splitting ratios for switching the outputs between the channel 1 and 2. b) Comparison between ANN prediction (solid line) and simulation (dashed line) versus target for the full set of inverse designed splitting ratios in 5% steps for a T_{target} of 90%.

multiple perturbation patterned MMIs. No pattern was found to be able to increase transmittance above the unperturbed value of 94% — suggesting that high transmittance targets are limited by both the MMI's physical design parameters and the scattering properties of the etched perturbations.

Figure 6.16 a) shows an example of switching power between the 1×2 MMI device's two outputs with a 90% total transmittance target for the final iteration networks. Agreement with the target for both the forward predictor and simulation is excellent in all cases. The MMI regions' electric field distributions show that the perturbations are essentially acting to block light, diverting it away, due to the large index contrast and lack of supported mode in the etched region. A comparison between ANN prediction and simulation inverse design MMI peak intensity error is shown in Figure 6.16 b). The initial generation of ANNs clearly shows a decrease in performance towards the high transmittance limits of the two output ports, as well as a relatively poor agreement between the forward network prediction and simulated response. These both are found to improve drastically with subsequent network iterations, with the final generation showing a reasonably flat error for different splitting ratios and significantly better agreement between prediction and simulation, as expected from the statistics shown in Figure 6.15.

One remaining crucial question is if the network is actually generating new, previously

unseen perturbation patterns, or if it is simply acting as an advanced lookup table. The table below lists the average number of mismatched perturbations between new inverse designed patterns and the initial training dataset, and the average total number of perturbations in any given pattern.

ANN	avg. mismatch count	avg. perturbation count
1×2 MMI iteration 0	10.1 ± 10.2	19.1 ± 15.0
1×2 MMI iteration 5	1.9 ± 1.5	16.1 ± 5.1

By the final iteration of networks, almost any inverse designed pattern is contained within the initial training data, with the average number of non-identical perturbations being just ∼2. This is attributed to the relatively low complexity of the 1×2 problem. We move next to a 3×3 device to confirm the technique's ability to fully generalise a system too complicated solve in an exclusively brute-force manner.

6.2.2 3×3 MMI device with weak optical perturbations

The 1×2 MMI discussed so far represents only the relatively simple problem of a variable power splitter, but it was nevertheless an important first step in demonstrating capability the of ANNs to model such systems. Next, a more complex coupling geometry of a 3×3 MMI is investigated with the goal of generating arbitrary intensity transmission matrices. Of particular interest are the cases where light is routed between each input port to each output port and their permutations.

The device was also chosen to have a 6 µm by 33 µm MMI region. These dimensions result in very poor unperturbed splitting performance; however, the natural length of a self-imaging device with the same width is four times longer for off-centre inputs (see Section 2.2.4.2). This would result in significantly longer simulation and optimisation times for the initial dataset generation, as well as increasing the overall device footprint. Both the input and output waveguides are centralised about the MMI region with a separation of 2.15 µm. Another change from the 1×2 case is that the perturbations were modelled with a reduction to the effective index of $\Delta n_{\text{eff}} = -0.25$, which have been previously experimentally demonstrated in an all-optical spatial modulator for reconfigurable silicon photonic circuits [112].

Initially, the same shallow etched perturbations were also investigated for the 3×3 device; however, the effective index shift of the etched perturbations ($\Delta n_{\text{eff}} = -0.71$) was found to be too strong for the fine control required to efficiently route light within this particular device and its limited design space. This is attributed to the unperturbed device's poor splitting performance and the perturbation's larger magnitude.

The initial dataset was generated in the same manner as before with the slightly modified fitness function of

$$f = 1 - \left[\sum_{m=1}^{M} \sum_{n=1}^{N} \left(T_{m,n}^{\mathrm{tar}} - T_{m,n} \right)^2 \right]^{0.5}, \tag{6.2}$$

where $T_{m,n}^{\mathrm{tar}}$ and $T_{m,n}$ are the target and modelled transmittance for the coupling between the m^{th} input and n^{th} output ports of an M-by-N MMI device. This tries to minimise the distance between all port-to-port coupling values and their respective targets. A total of 2500 training pattern families were produced for the initial training dataset for the 3×3 weak optical perturbation case, with a maximum of 50 accepted perturbations in any given pattern. Note that the maximum number of perturbations accepted was increased due to their weaker magnitude and the problem's increased complexity. Each pattern is simulated using Lumerical FDTD solver for all three of the input ports, and the complex electric field profiles at both the MMI-output interface and across the single-mode output waveguides at a distance of 20 µm from the MMI are recorded. As before, the simulated dataset was doubled through symmetry exploitation, and then split with an approximately 9:1 ratio for training and validation datasets. Specific families were once again kept together so as to avoid near-identical patterns occurring in both datasets.

The forward predictor and generator network structures are the same as shown in Figure 6.12 and Figure 6.13, but now with three separate channels — one for each of the input ports. The networks were trained on the intensity profiles at the single-mode waveguide output using the increasing batch size approach for 120 epochs. As with the 1×2 device, this process was iterated a total of six times, mixing in an additional 10,000 inverse designed patterns from each of the previous generator networks. The targets for the new patterns were produced by generating a 3×3 array of random numbers, in which the rows represent the input ports and the columns correspond to the output ports. Each coupling value was then given a one-in-three chance of being replaced by a null target, thus reducing the homogenisation of target values. Total target transmittance for a given input cannot sum to greater than 1, which was prevented by normalising each row to a total target transmittance of between 70–100%. Similarly, columns cannot sum to greater than 1 due to reciprocity, and were therefore also normalised to the total target transmittance.

Performance of each generation of the ANNs was evaluated by mean absolute transmittance error for a fixed dataset of 1000 targets, as shown in Figure 6.17. The testing dataset was generated in the same way as during step 4 of the iterative training scheme, with a quarter of the targets normalised to a total transmittance of 70%, 80%, 90% and 100%, respectively. The test dataset was also evaluated for a version of the iteration 5 forward network that was trained exclusively on the transmission values opposed to the entire intensity profile. The mean absolute transmittance of the predicted response was found to increase slightly from 2.5% to 3.0%. While this is not a significant difference,

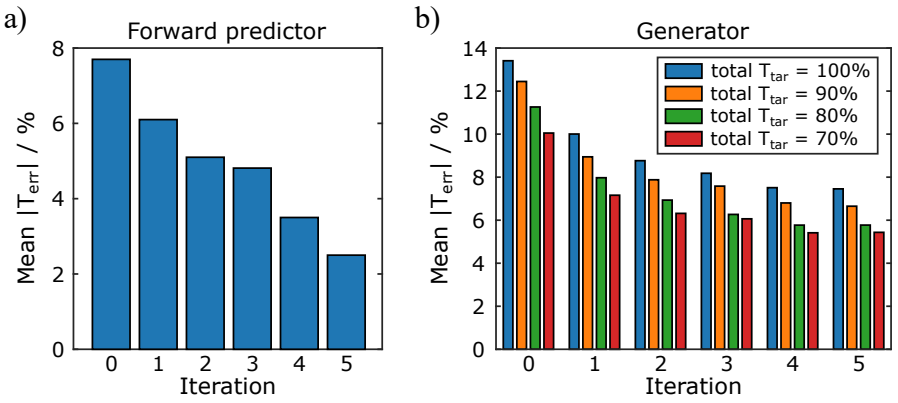


Figure 6.17: 3×3 MMI mean absolute transmittance error between the ANNs and simulations for the consecutive data-generation iterations, for a fixed test dataset. a) Forward network error form prediction against simulation. b) Generator network error of design-targets against simulations of inverse-designed patterns.

we note that the inverse design's performance is highly dependant on the forward prediction, as both the generator network and threshold value utilise the forward ANN. Both the prediction and inverse design are found to improve with subsequent iterations. After the complete iterative training, both the forward and inverse ANNs have experienced significant performance improvements of a factor of 3 and 2, respectively, compared to the initial dataset of iteration 0.

Figure 6.18 shows the example case of the output coupling between the second and third inputs being swapped while leaving the remaining input port coupling unaffected. Absolute transmittance coupling for the top input coupling varies by less than 10% for all cases. These cases' good performance indicates the robustness of the networks in archiving arbitrary splitting ratios, providing it is within the physical limits of the system.

Next, the full set of port-to-port cases and their permutations are shown in Figure 6.19. Once again the agreement is very good, demonstrating the system's ability to dynamically route light between any given input and output without adversely affecting the other couplings. As previously discussed, interferometric meshes are traditionally used to achieve controllably routing of light on-chip. Here, similar functionality can be achieved in a fraction of the overall device footprint.

The transmission matrices of these cases (denoted by the Roman numerals in Figure 6.19), plus the additional instances where one port coupling is unitary and the other two are split 50:50 between the remaining outputs, are shown in Figure 6.20. In this colour-array representation, the row and column numbers correspond to input and output ports, respectively. Excellent agreement is found between the target and simulated transmission for all inverse designed patterns, with a mean absolute transmittance error of less than 5%.

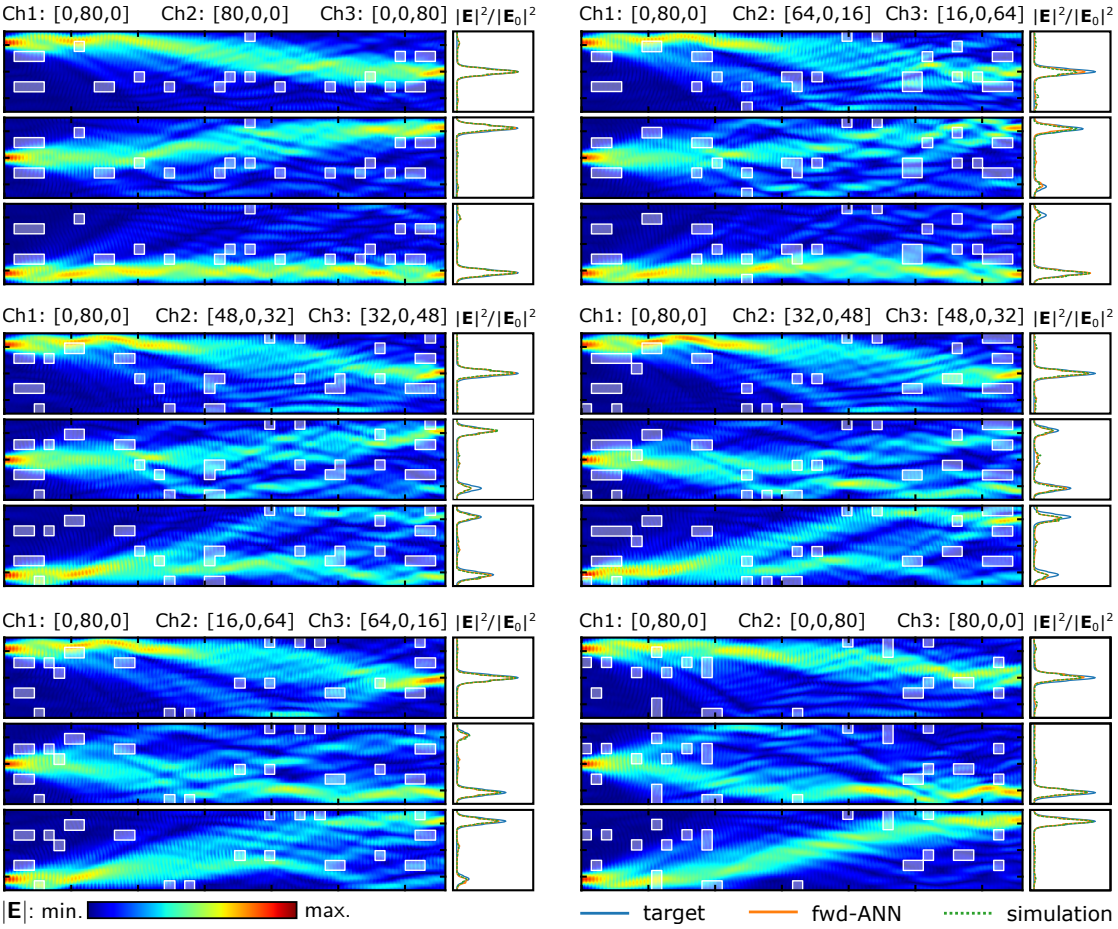


Figure 6.18: Examples of inverse designed 3×3 MMI perturbation patterns transferring output coupling between the top and bottom ports with an 80% total transmittance target for each input. Generated from the iteration 5 networks. The target percentage transmittance values are listed above the electric field distribution for each channel input.

Once again, it is important to ask if the networks have actually managed to generalise the physical system or if they are simply acting as an advanced lookup table. The average number of mismatched perturbations and the average total number of perturbations are given in the table below for the initial and final network generations compared to their respective training datasets. The error is the standard deviation of values.

ANN	Perturbation mismatch count	Total perturbation count
3×3 MMI iteration 0	21.4 ± 12.5	45.6 ± 20.9
3×3 MMI iteration 5	17.1 ± 12.2	36.0 ± 13.6

Inverse designed patterns for the 3×3 case are significantly different from the patterns provided to the networks during training, as can be seen by the high average perturbation mismatch. This number does not decrease substantially for subsequent generations of networks, suggesting that the ANNs have generalised the problem well and can produce

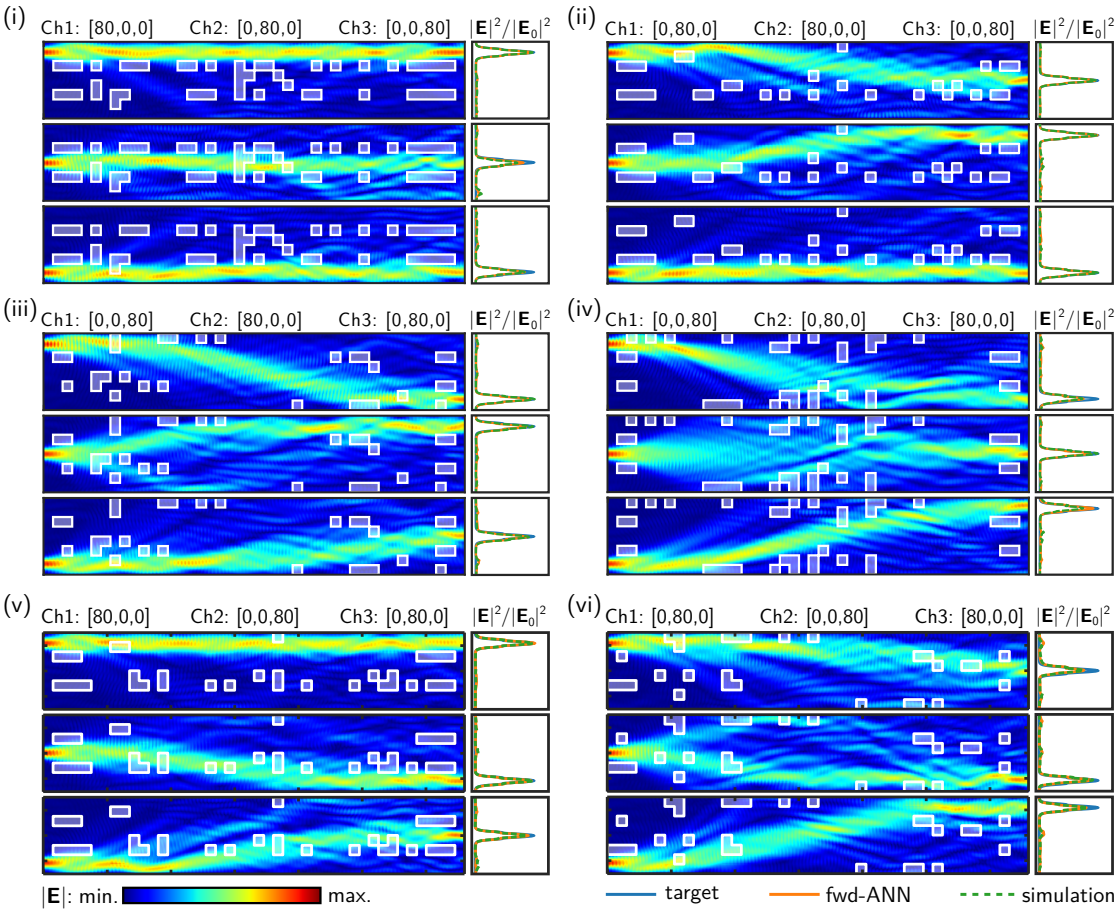


Figure 6.19: Examples of inverse designed 3×3 MMI perturbation patterns for port coupling permutations with an 80% total transmittance target for each input. Generated from the iteration 5 networks. The target percentage transmittance values are listed above the electric field distribution for each channel input.

new, original solutions. Also, note the decreasing average total number of perturbations, indicating that later iterations are more capable of placing perturbations in optimal locations as well as having higher overall performance.

6.2.2.1 Variational autoencoder for latent space operations

Another type of deep learning model that has gained lots of interest in recent years is Variational Autoencoders (VAEs) [229, 230]. A VAE is defined as an autoencoder whose training is regularised such as to avoid overfitting and ensure that the compressed dimensional, or 'latent', space has good properties that enable the generative process. In particular, β-VAE networks have been demonstrated as a novel approach to generating a smooth set of network outputs, thereby allowing the gradual morphing between designs and even the combining and subtracting of specific features [230–232]. The dense interconnect has an additional regularisation, the Kullback-Leibler divergence term, that

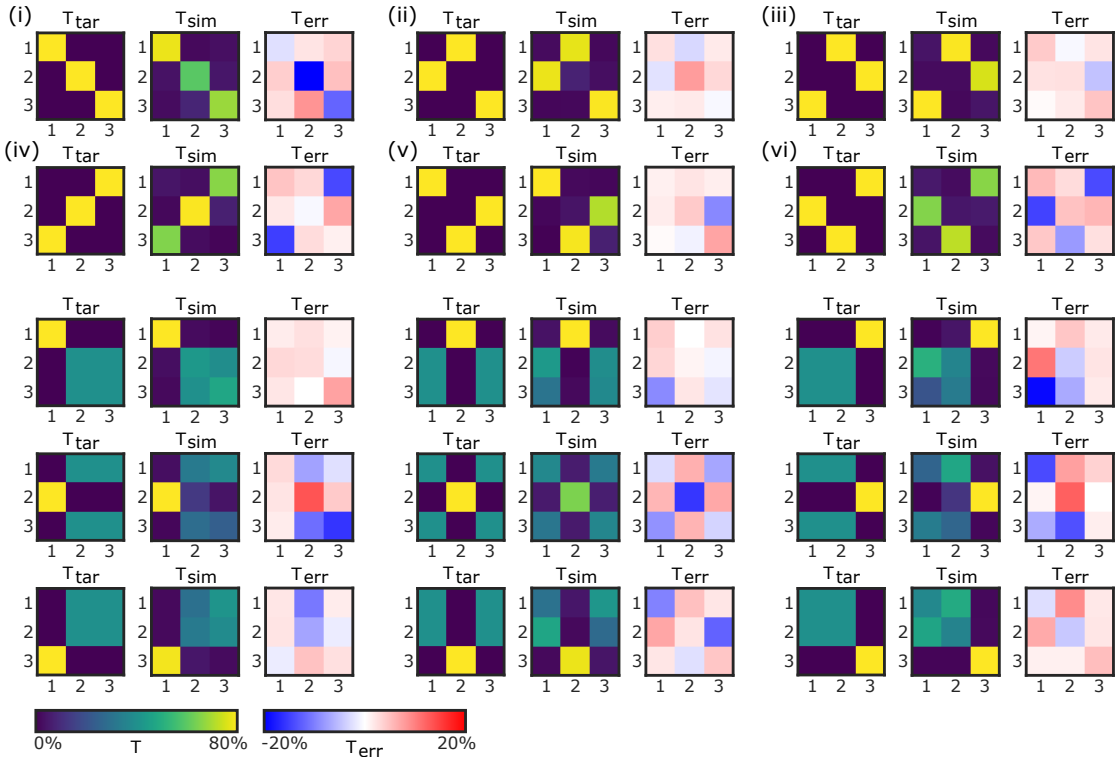


Figure 6.20: Transmission matrices for all six input-output permutation cases (Roman numerals correspond to the examples given in Figure 6.19), as well as for the additional cases where one port coupling is unitary and the other two are split 50:50 between the remaining outputs. The patterns are generated from the final generation of networks (iteration 5), and total transmittance target is set to 80%.

tends to regularise the latent space's organisation by making the distributions returned by the encoder close to a standard normal distribution [233]. This Kullback-Leibler divergence term acts as a third loss parameter during training, with a weighting of 0.2 that was found to work well with the previous 1:50 weighting between MMI image and output profile. It essentially ensures that the network learns a smooth and continuous representation of the physical features in latent space. Thereby allowing the smooth transition between designs by moving directly between two positions in the vectorial latent space. Similarly, features can be added or subtracted from a design through simple vector arithmetic.

Figure 6.21 shows examples of different latent space operations for the 3×3 MMI VAE generator's final iteration. A clear, smooth transition between designs is observed during interpolation, as shown by Figure 6.21 a), with the mid-point states that one would logically expect. This process is equivalent to the manual interpolation carried out previously in Figure 6.18, but only requires two inverse designs to obtain the full transition of states. Examples of latent space addition and subtraction are demonstrated between pairs of transmission matrices in Figure 6.21 b). Features, in this instance port-to-port coupling values, are seen to be added or subtracted from the latent space output.

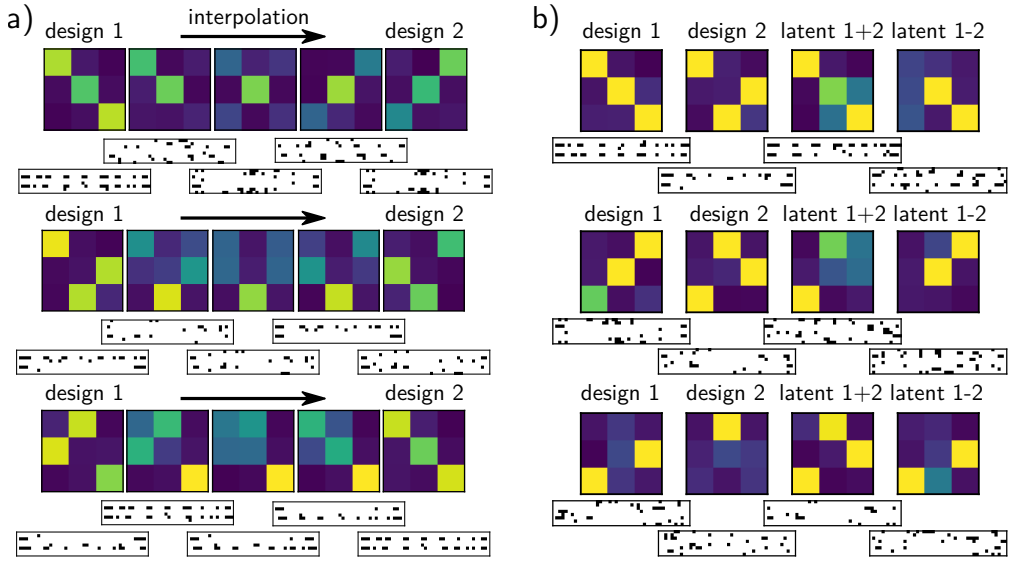


Figure 6.21: Examples of VAE generator latent space a) interpolation and b) arithmetic operations. The colour-array transmission matrices are for the simulated response of the inverse designs shown below each matrix, and the colour bar limit is set to the total transmittance target of 80% for all designs.

However, this process is still clearly limited by the system's physical limits; namely, transmission cannot be greater than 100% or less than 0%. Note that new patterns are being generated during the latent space vectorial operations that do not merely correspond to the combination of, or interpolation between, the design MMI patterns themselves. The capability to carry out vector arithmetic has potential applications in analogue photonic computing [234–238].

6.2.2.2 Design of full complex transmission matrices

Finally, we demonstrate the deep learning enabled design of the complete complex transmission matrix for the development of a real-time universal optical component. Such reconfigurable photonic networks are traditionally achieved via meshes of many individual MZIs to control phase and intensity information [172, 239–241], with applications in coherent datacommunication, quantum optics, and analogue photonic computing [3, 5, 242–245]. However, this geometry results in very large device footprints and cascading losses that depend on the number of interferometers in a given path [148].

The phase-aware version of our networks are based on the same 3×3 MMI geometry as covered previously, but the final iteration of the networks were retrained on the full complex electric field information rather than just intensity. This results in four output channels for each input channel; the real and imaginary parts of E_x and E_z, where the z-axis is the propagation axis and the x-axis is the perpendicular in-plane axis. There is no E_y (out-of-plane) component as the simulations were carried out in 2D using the effective index method for the fundamental TE mode. The forward ANN's accuracy

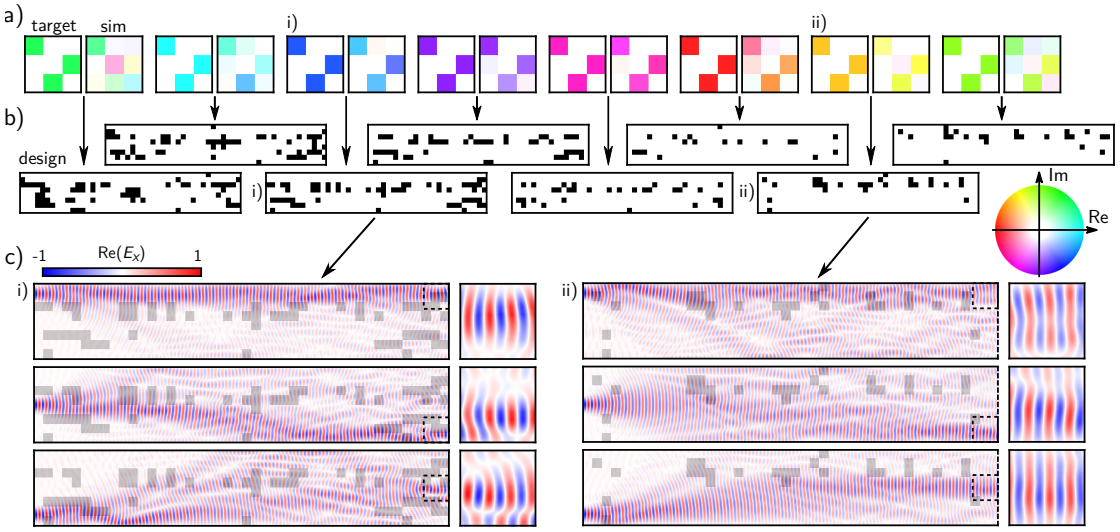


Figure 6.22: a) Examples of complex transmission matrices design for a fixed intensity target with phase varying over 2π for all outputs. b) The ANN inverse designed perturbation patterns. c) Examples of the real part of the E_x field distribution simulated using FDTD for the targets with a π phase difference, corresponding to the targets denoted by the Roman numerals in part a). The square subplots on the right-hand side show a zoomed $2 \times 2\,\mu m$ region at the target output for each input channel.

was found to be slightly reduced compared to the intensity-only version of the network, owing to the increased complexity of the problem and the same design space. The phase in the outputs is approximated by $\theta = \arctan\left(\mathrm{Re}\left(E_x\right)/\mathrm{Im}\left(E_x\right)\right)$, under the assumption that $E_z \ll E_x$, and target values are restricted to the complex values at the centre of each output channel.

An example of a fixed intensity target with varying phase over 2π for all outputs is shown in Figure 6.22. A hue and brightness based colour coding is used for the complex transmission matrices, where hue corresponds to the phase and brightness to the intensity. Intensities below 5% are clipped to white to allow clear visualisation of the complex transmission matrix. There is a slight degradation of performance compared to the purely intensity case. This is attributed to two main causes. Firstly, the phase-aware networks were only trained on the iteration 5 datasets, and the networks have not therefore undergone the feedback effect of the full iterative optimisation with phase information. Secondly, we are asking more from the same design space, which was already approaching its limits in the intensity case. However, despite this, the networks display good qualitative agreement, as shown by Figure 6.22, with a constant phase offset cycling through almost a complete 2π. It is primarily at the limits of this phase range where the inverse design performance deteriorates. It is also worth noting that to achieve these large shifts in output phase the generator network has learnt it has to place perturbations directly in the light's path, a feature it typically tried to avoid for purely intensity-based targets.

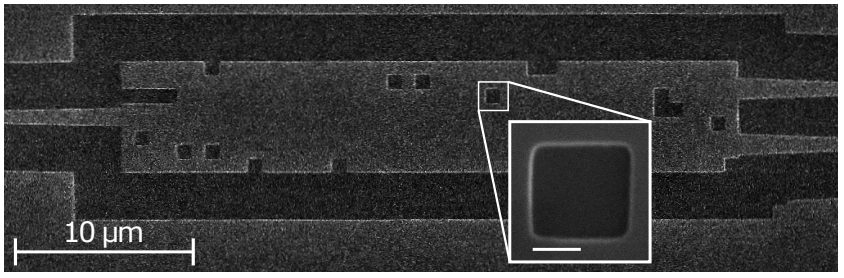


Figure 6.23: Scanning Electron Microscope (SEM) image of the etched 1×2 MMI device. The MMI region was fabricated with dimensions of $W = 6.0\,\mu\text{m}$, $L = 33.0\,\mu\text{m}$ and square etched holes of average side length $L_{\text{pert}} = 760 \pm 10\,\text{nm}$. Insert scale bar represents 400 nm.

6.3 Etched multiple perturbation patterns

Inverse design in nanophotonics is a field with rapidly expanding interest over the last decade, enabling complex photonic functionality on a small-scale footprint [217, 246]. Significantly sub-wavelength etched devices have been demonstrated as broadband (de)multiplexers [211], arbitrary power splitters [247], optical routers [248] and mode converters [249]. Dynamic control has been proposed using thermo-optic and electro-optic beam steering [250] and experimentally demonstrated using all-optical wavefront shaping [112]. Phase Change Materials (PCMs) have also been experimentally demonstrated for non-volatile reconfigurable beam steering from metasurfaces [251], and have recently shown promising results in low-loss reconfigurable PICs [252, 253].

Here, shallow etched perturbations in a 1×2 MMI were decided upon to experimentally confirm the capabilities of multiple perturbations in this work, due to their relative ease to fabricate. However, the weak optical perturbations used in the case of the 3×3 MMI networks have also been previously experimentally demonstrated [112]. The fabricated pattern's design was produced using the basic iterative approach to maximise the transmission of the bottom output and compared to the final iteration of the 1×2 forward predictor ANN covered in Section 6.2.1.

A 220 nm SOI wafer with a 120 nm etch depth and 3 µm trench width was used to produce the structures. A 120 nm etch depth was also chosen for the perturbations to keep the fabrication a single etch process and remove any alignment errors from a second alignment process. Single-mode input and output waveguides were chosen, with a width of 500 nm that is adiabatically tapered up to 1 µm over a 10 µm length at the MMI region boundaries, which has dimensions of $6 \times 33\,\mu\text{m}$ itself. The output waveguides are separated by 3 µm centre-to-centre symmetrically about the middle of the MMI region, and the perturbations were chosen to be 750 nm sided squares. The SEM image of the device as fabricated are shown in Figure 6.23.

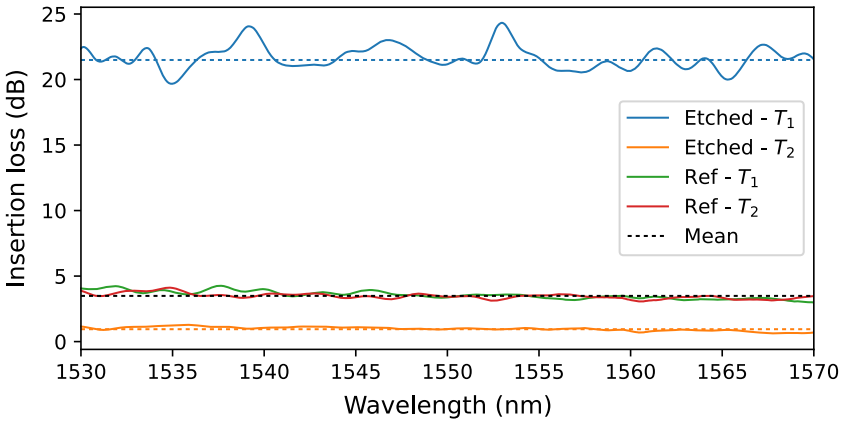


Figure 6.24: Experimentally measured insertion loss for the etched MMI device (blue and orange lines), normalised to the average insertion loss of six straight waveguides. The green and red lines show the insertion loss for an unperturbed reference 1×2 MMI device with identical dimensions.

Measurements taken from the SEM images confirm the positioning of etched holes all to be within 50 nm of the design, which is of the limit the image's resolution at this magnification. The insert in Figure 6.23 is at ×50k magnification and allows for more precise measurement of the perturbation. Each perturbation's dimensions were measured, and it was found that they were fabricated slightly larger than designed, with an average perturbation length of $L_{\text{pert}} = 760 \pm 10$ nm. The corners were also found to be slightly rounded with an average radius of $r_{\text{pert}} = 30 \pm 10$ nm. However, these slight discrepancies are to be expected from the etch process. The larger perturbation size is more resistant to fabrication defects than the significantly sub-wavelength perturbations commonly used in literature, where exact dimensions and positions can be critical to device performance [254]. Additionally, such $\ll\lambda$ devices cannot be fabricated using industry-standard deep-Ultraviolet (UV) lithography and have to rely on electron beam lithography instead.

Figure 6.24 shows the broadband swept-source transmission (outlined in Section 3.3.2) for both outputs of the fabricated device, as well as for an unperturbed reference 1×2 MMI device with identical dimensions. The insertion loss values are normalised to the average loss of six straight waveguides fabricated on the same chip, thereby removing any common coupling and propagation losses. Insertion loss was measured to be −21.1 dB and −1.0 dB (at $\lambda = 1550$ nm) for the top and bottom outputs, respectively. This corresponds to an increase of excess loss of approximately 0.2 dB compared to an unperturbed device.

The individual output port photomodulation maps are shown in Figure 6.25 a), and show the majority of light is coupled to the bottom output with minimal regions of increased coupling (positive $\Delta T/T$). On the other hand, the top output shows many regions of increasing transmission due to the very low initial coupling to that port. It is relatively easy to increase the transmission by 50% if the unperturbed coupling has an insertion

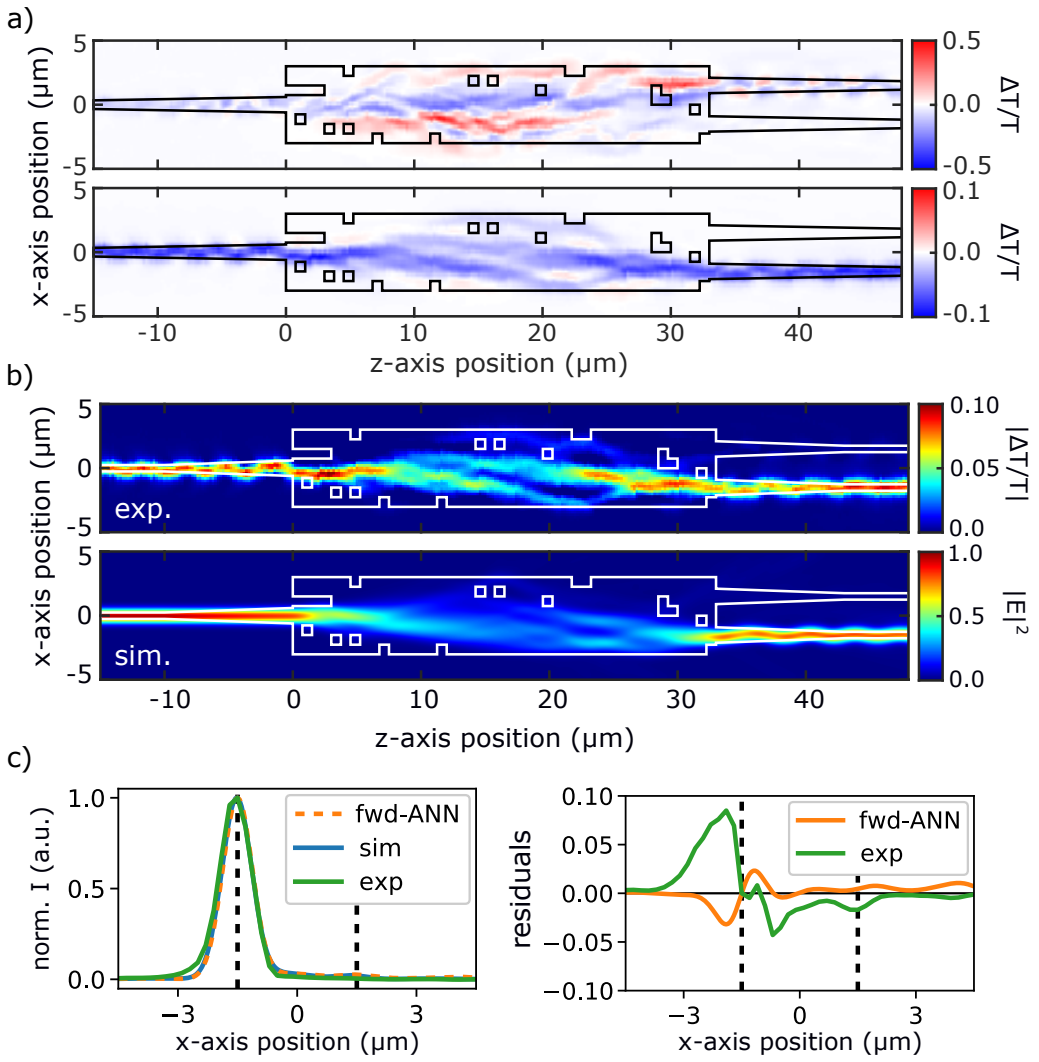


Figure 6.25: a) Experimental photomodulation map for individual top and bottom outputs. Both outputs were measured simultaneously via a dual fibre array, and the pump optical fluence was set to $20\,\mathrm{pJ\,\mu m^{-2}}$, resulting in an effective index shift of $\Delta n_{\mathrm{eff}} = -0.1$ in the silicon. b) Combined experimental photomodulation map (top) and FDTD simulated $|\mathbf{E}|^2$ distribution (bottom) of the etched 1×2 MMI device. The simulated light intensity response was convolved with a Gaussian of width 740 nm, corresponding to the pump laser spot size. c) Normalised intensity profile across the single-mode output waveguides for the experiment (green), simulation (blue) and forward prediction of the ANN (orange). The plot on the right-hand side shows the residuals between the simulation and ANN prediction as well as the simulation and experimental perturbation map. Simulations were carried out using Lumerical FDTD solver for the fundamental TE input mode at a wavelength of 1550 nm.

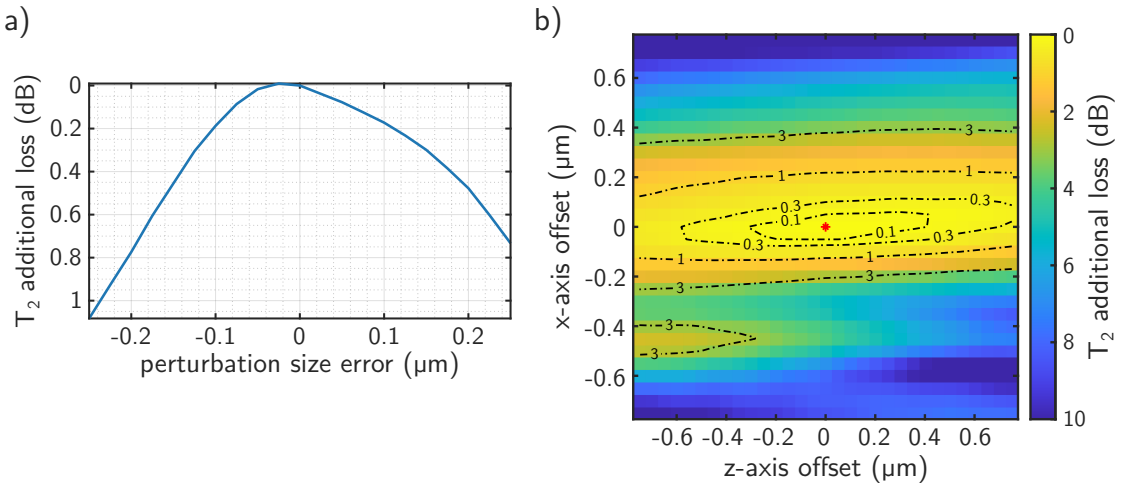


Figure 6.26: Simulated additional loss of optimised output port (T_2) against (a) perturbation size error and (b) x- and z-axis offset for the pattern of the experimental shallow etched device. 0 µm size error corresponds to 750 × 750 nm perturbations, and the red star in b) indicates the zero-offset position.

loss of −20 dB. A comparison between the simulated $|\mathbf{E}|^2$ distribution and the etched device's combined photomodulation maps is shown in Figure 6.25 b). The combined case compares the change in transmission in a given output to the total unperturbed transmission, i.e. $\Delta T_i/(T_1 + T_2)$, where $i = 1$ or 2 and corresponds to the output port number. While these measurements are not directly equivalent, we have previously demonstrated in Chapter 4 they are proportional in high transmission cases, such as here. The qualitative comparison between the two is reasonable, showing the majority of light coupling to the bottom output, as confirmed by the experimental transmission measurement. The greatest discrepancy appears in the input taper, where it is clear the incident mode is not being symmetrically expanded as designed, but can instead be seen to 'zigzag' back and forth. Figure 6.25 c) compares the intensity profile across the single-mode output waveguides for the experiment and simulation, in addition to the forward prediction of the ANN developed in the previous section. The residuals show there is some slight misalignment in the x-axis, caused by mode in the output waveguide oscillating back and forth, but in general, the agreement is very good. There also appears to be a very small negative peak in the top output for the experimental data, indicating that there is slightly less crosstalk in the fabricated device than simulations suggest. It is also possible that the perturbation is resulting in a small reflection back into the MMI region, causing a slight increase in coupling to the bottom output.

An additional advantage of large-scale perturbations compared to $\ll\lambda$ perturbation sizes is that they are more resistant to fabrication tolerances [254]. The experimentally fabricated shallow etched device's simulated tolerance against perturbation size error is shown in Figure 6.25 a). Errors of up to ±100 nm result in less than 0.2 dB additional loss in the optimised output port. Another factor to consider is alignment accuracy of

different mask layers during the fabrication process. For the experimental device presented in this work, the alignment problem can be bypassed via the use of a single etch process, that is to say, the perturbation etch depth is the same as that of the rib structure. However, typical alignment overlay accuracy of ±20 nm for deep-UV lithography results in <0.1 dB additional loss for this device, as demonstrated by the simulations in Figure 6.25 b), thereby demonstrating the robustness of the technique.

6.4 Summary

Developing more advanced functionalities from individual integrated photonic elements is essential as circuit complexity continues to increase. Currently, most circuits are designed for a specific application; however, there is a growing demand for programmable photonic components that can be reconfigured with applications in datacommunications, photonics-based quantum, neuromorphic and analogue computing. Such reconfigurable photonic chips are typically achieved via meshes of many individual interferometers to control phase and amplitude information, but this geometry results in very large device footprints and path-dependent cascading losses. Ultra-compact universal optical elements are needed to bring similar control to the component-level.

In this chapter, we first demonstrated the capability of using multiple scatterers to create compact integrated components, such as arbitrary power splitters, multiple-port optical routers and mode converters. The optimisation was initially carried out using brute-force iterative placement of 500×500 nm square perturbations, with $\Delta n_{\text{eff}} = -0.25$, into a multimode region, and was shown to result in high-performance devices with excess loss values typically <1 dB. However, the computational time required is extremely long, scaling quadratically with the device area and linearly with the number of desired outputs. It is possible to recover some additional outputs without additional optimisation runs by exploiting device symmetry and reciprocity, providing the transmission is high. Larger 1 µm-sided perturbations were also investigated and found to typically reduce the optimised device's performance, but were ×4 faster to run.

The simple iterative optimisation approach was used to produce datasets of perturbation patterns that were used to train a deep-ANN capable of generating new, high-performance patterns for specific output targets. Two separate systems were investigated; the 1×2 power splitter and the generation of full transmission matrices for a 3×3 device. Both are based upon a compact 33×6 µm multimode waveguide region, representing a significantly smaller footprint than interferometric meshes typically employed in the literature. An iterative training scheme was used to optimise further the networks' accuracy, in which new generated patterns were fed back into the training of the next generation of networks, thereby improving their performance beyond just

increasing the initial dataset size. Ultimately, the mean port transmittance error compared to numerical simulations was found to be $\sim 2\%$ and $\sim 6\%$ for the 1×2 and 3×3 networks, respectively. Furthermore, we have shown that independent phase control can be readily implemented via this approach; however, we found our system's design space limited performance. This could be alleviated by increasing the MMI region size or changing the perturbation's size or magnitude.

Generation of the initial dataset is by far the most time-consuming part of developing the ANNs, taking on the order of weeks to fully optimise and simulate. However, while also relatively computationally demanding to train the networks ($\sim$5 hours), their operation is exceptionally fast ($\ll$1 ms) in the generation of new designs and, hence, could be used in real-time applications. The vast majority of the full inverse design time is actually dedicated to the binary threshold value search, which depends on the number of threshold values considered. Here, we chose 100 steps, which was found to be a good compromise between inverse design performance and speed. In principle, it is possible that a network could be trained on exclusively the complex scattering matrix to decrease the time taken to generate the initial dataset drastically. This would also take into account any reflections in our system. It was found that using transmission values exclusively for the final iteration of the 1×2 networks only increased the mean transmittance error from 2.5% to 3.0%. However, the generation of new patterns is highly dependant on the accuracy of the physical predictor network, so even a small increase in the error may result in significantly poorer performing inverse design.

Finally, we experimentally verified the performance of multiple perturbation patterns via the shallow etching of a 1×2 MMI device. Such a device's fabrication is easily achievable with standard industry deep-UV lithographic techniques. Insertion loss was measured to be around -1 dB for the optimised port over a 40 nm bandwidth, with less than -20 dB crosstalk. Simulations show that these large-scale perturbations are more resistant to fabrication tolerances compared to $\ll\lambda$ perturbations, with size errors of ±100 nm resulting in less than 0.2 dB additional loss in the optimised output port. Additionally, using the same etch depth for the perturbations as the rib structure, as is the case here, results in no alignment errors. If a different etch depth is chosen, then a typical alignment overlay accuracy of ±20 nm was found through simulation to result in $<$0.1 dB additional loss for this device.

Chapter 7

Conclusion and Outlook

In this thesis, we have demonstrated the power of perturbative techniques to characterise, tune and manipulate Photonic Integrated Circuits (PICs). We have continued to develop Ultrafast Photomodulation Spectroscopy (UPMS) as an all-optical non-destructive technique for the characterisation of individual photonic elements at the wafer-scale. Optical pump pulses are used to perturb the flow of light in a photonic device through slight alterations to the silicon's refractive index profile. Mapping this perturbation's effect on transmission as a function of position results in a spatial photomodulation map, allowing direct visualisation of the flow of light inside a device at sub-micron resolution. Similarly, the strategic positioning of multiple perturbations can be used to shape the propagating wavefront within photonic circuits arbitrarily.

In Chapter 3, we introduced an improved UPMS experimental setup, which enables higher resolution mapping, better signal-to-noise ratio and faster device scan times. The system yields a perturbation size of 740 nm at Full-Width Half-Maximum (FWHM). The theoretical diffraction-limited pump spot size is calculated to be smaller than the measured value; however, this does not take into account free-carrier transport within the silicon or beam quality deterioration. Temporary effective refractive index shifts of a Silicon-On-Insulator (SOI) waveguide by as much as $\Delta n_{\text{eff}} = -0.5$ were achieved, without causing any damage to the device. The strength of the effective index shift induced by the perturbations can be controlled through variations of the pump fluence. Free-carrier recovery time was found to be sub-nanosecond, as such each perturbation can be considered completely isolated from one another. Longer lifetime thermal effects were confirmed to be minimal.

Next, in Chapter 4, we demonstrated ultrafast perturbation mapping as a tool for the testing of photonic circuits capable of revealing the flow of light within a device, even in the presence of protecting cladding layers that render conventional near-field techniques impossible. The development of a new general and rigorous analytical model based on Lorentz reciprocity allows the prediction of transmittance perturbation maps

for arbitrary linear photonic systems with great accuracy and minimal computational cost. This approach was found to be nearly exact for low magnitude and significantly sub-wavelength perturbations, while for wavelength-scale perturbations an additional approximation is required, in which the perturbation is modelled as a 1D Fabry-Pérot cavity. Excellent agreement was found between experimental photomodulation maps and the model, with correlation values typically in the 80–95% range, demonstrating our capability to predict and measure the perturbation maps of photonic devices. We have also shown that through a parametric study this technique can be used to predict the fabricated Multimode Interference (MMI) dimensions with good accuracy, around ±1.5 µm in length and ±0.1 µm in width, depending on the device under test. Any asymmetry in a device is visible directly from the perturbation maps and reflected in the resulting correlation map. Currently, we have only considered MMI region width and length parameter space; however, this technique can easily be expanded to also account for wafer thickness variations, cladding layer defects, input-output taper positions/sizes and other potential fabrication defects or environmental factors. Finally, we demonstrated UPMS as a potential technique for directly visualising the light intensity distribution in high-transmittance devices, without requiring direct access to the near-field. Good agreement between the simulated electric field distribution and perturbation map is achieved, with up to the 5-fold self-imaging point clearly visible. In addition to constituting a promising route for non-destructive wafer-scale testing in photonics manufacturing, ultrafast perturbation mapping may have applications in the design optimisation of photonic structures with reconfigurable functionalities.

Chapter 5 covers the trimming of germanium ion-implanted SOI devices. The ion implantation results in a localised disruption of the silicon crystal lattice, forming amorphous silicon, which has a higher refractive index. Localised annealing then recrystallises the silicon, and in doing so, reverts the refractive index back towards the original value, thereby allowing the tuning of devices depending on the annealed length. Unlike trimming with resistive heaters, ion implantation does not require continuous power draw to maintain once tuned; however, the approach cannot be used to offset dynamic environmental features, such as temperature. We demonstrate the approach for the critical coupling of a racetrack resonator using pulsed laser annealing. Recrystallisation was found to be incomplete for the pulsed laser. Nevertheless, a significant enhancement of the extinction ratio (>11 dB) was observed, and the critical coupling location was found to within 200 nm. It is suspected that the sub-picosecond pulse duration does not allow for sufficient annealing, and further optimisation of the pulse parameters may result in better recrystallisation.

Finally, in Chapter 6, we investigate some of the advanced photonic functionalities that can be achieved for small-footprint devices with multiple sub-wavelength refractive index perturbations. The optimisation of designs is initially carried out using a brute-force

iterative approach with basic optimisation functions. We demonstrate various high-performance devices through this approach, including optical routers, mode converters, and power splitters with arbitrary ratios. The work amalgamates into the training of a deep-Artificial Neural Network (ANN) capable of rapidly and accurately generating perturbation patterns to achieve an arbitrary target complex transmission matrix of a 3×3 device. A feedback training scheme was used, improving the networks' accuracy beyond simply expanding the initial train dataset. The robustness of both the approach and the ANNs was then experimentally confirmed through shallow etched perturbations in a 1×2 MMI device, resulting in <1 dB insertion loss and less than −20 dB crosstalk. Such perturbations are easily fabricated using industry-standard lithographic techniques, without the requirement of additional alignment steps.

7.1 Outlook

The field of integrated photonics is currently undergoing a period of rapid development, with ever-expanding device functionalities and consequently increasingly complex circuit design and overall device footprint. The next decade will see a shift of cascading many relatively simple components to fewer, more intricate and reconfigurable elements. The work undertaken in this thesis lays the groundwork for a number of future studies into the application of perturbations for device characterisation and advanced functionalities through wavefront manipulation.

Photomodulation mapping could be developed into a true wafer-scale testing technique for mapping internal light fields by incorporating rapid scanning perturbations using high-speed mirror galvanometers. This would reduce device characterisation times from the order of minutes to sub-seconds. In addition, we have currently only investigated the effect of the perturbation on transmission. The pump objective could capture changes in out-of-plane scattering, and reflections could be separated from incident light using an optical circulator. The experimental setup and a proof of concept example for a 1×2 MMI device using the out-of-plane scattering perturbative approach is presented in Appendix B.6. These approaches have the advantage of only requiring a single input fibre coupling to test the whole device, without any out-coupling required.

There is significant optimisation of the pump pulse parameters required to achieve complete recrystallisation of the silicon for the germanium ion-implanted devices. Longer pulse durations may result in better annealing, but not produce high enough free-carrier densities to allow the differential transmission to be used for gradient feedback tuning [160]. Potentially, a continuous wave laser could be used in conjunction with the ultrafast pump to provide complete recrystallisation of the silicon; however, this would require very accurate collinear alignment and focusing of both beams.

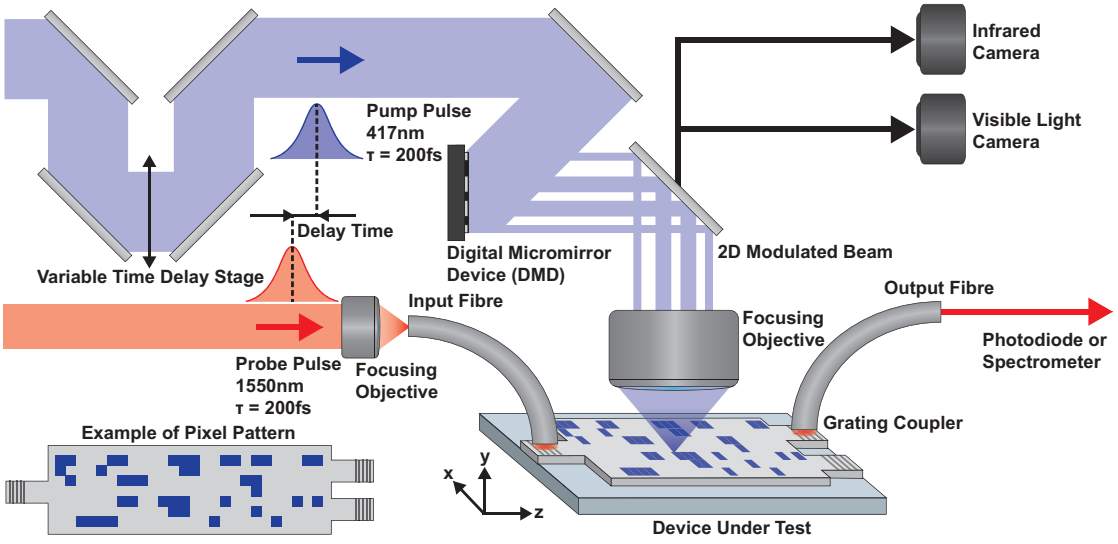


Figure 7.1: Illustration of the UPMS multiple perturbation setup. Two-dimensional modulation of the pump beam is achieved through the use of a DMD, and transmission is measured by fibre coupling probe pulses into and out of the device under test using grating couplers.

The work in this thesis has shown that even a single perturbation was able to achieve changes in transmission as high as 40%. Previous work has shown that even greater modulation can be achieved through multiple pump spots, with >97.5% of light transmitted into a single output of a 1×2 MMI device [112]. The approach is a general one, given fine enough control over the refractive index profile one can directly shape the propagating wavefront within a device. The multiple perturbations can be created by fixing the focusing objective in place and using a Digital Micromirror Device (DMD) to provide two-dimensional modulation of the pump beam itself, as illustrated in Figure 7.1. This results in an array of pump light pixels, which, when focused onto device results in localised shifts in the refractive index profile. These multiple perturbations alter the mode distribution and with the correct pattern can shape the wavefront of the light, effectively guiding the light to a particular output. Storing the optimal patterns for each output allows for rapid switching between them, that is theoretically only limited by the free-carrier recovery time and corresponds to a switching speed in the gigahertz regime.

Combining this reconfigurable approach with the ANNs developed in this thesis would allow for network training on real experimental data. The same iterative feedback effect used to optimise the networks can be carried out on experimental data. Once sufficient data has been collected to allow an accurate physical prediction of the system, the ANN should allow full real-time control of the system for arbitrary targets. Additionally, we have so far limited ourselves to binary patterns in Chapter 6; however, multiple levels of effective index shift could be achieved via modulation of individual perturbations. In

the work by Bruck et al. [112], a 10×10 array of pixels were used to project the perturbation pattern, which theoretically corresponds to 101 possible levels of effective index shift magnitudes. Additionally, in Chapter 6 we demonstrated that multiple scatters can be used for a variety of different device functionalities, which could be achieved experimentally by this approach.

Ultimately, the most significant drawback of the all-optical approach is the requirement of ultrafast lasers with pulse widths on the femtosecond time-scale. Such systems are typically very large, cumbersome and expensive, as well as requiring careful alignment of optics and drawing a large amount of power. Therefore, it is not feasible to use this method for real-world devices; however, the approaches outlined in Chapter 6 are very general and can be applied for a number of different perturbation types, as shown by the shallow etched perturbation pattern device. Dynamically reconfigurable devices in the real world have been achieved using Phase Change Materials (PCMs), which offer a stable transition between states [252, 255]. Stable PCM-based approaches have the additional advantage of being non-volatile, thereby requiring no power draw to maintain a particular state, only when the device needs to be reconfigured. For devices which require reconfiguring infrequently, this represents significant power savings compared to actively controlled and maintained elements. In recent years, many important studies have been carried out combining PCMs with photonic circuits [256–259], in particular, a paper by Delaney et al. demonstrates multiple perturbation patterns, as considered in this work, in a thin PCM layer to dynamically route light in an MMI device [253]. PCMs have been shown to be electronically switchable [260], an essential characteristic for their adoption by industry.

In conclusion, I hope this work well portrayed the potential of perturbative approaches in the field of integrated optics. If this unprecedented era of technological growth is to continue, then the integration of increasing advanced systems is essential for their wide-scale adoption. Exponential increases in the demand for data must be matched technological development to meet bandwidth targets and drive down total power consumption of devices. Ultimately, many problems remain to be overcome; however, the future of all-optical computing and complete lab-on-a-chip already does not seem as far away as it once did — even compared to a few years ago since I first undertook my small part in this future.

Appendix A

Background appendix

A.1 Proof electric and magnetic field of EM wave are perpendicular in a vacuum

Maxwell's equation adjusted for a vacuum are

$$\begin{aligned}
\nabla \cdot \mathbf{E} &= 0, && \text{(Gauss's Law)} \\
\nabla \cdot \mathbf{B} &= 0, && \text{(No Magnetic Monopoles)} \\
\nabla \times \mathbf{E} &= -\frac{\partial \mathbf{B}}{\partial t}, && \text{(Faraday's Law)} \\
\nabla \times \mathbf{B} &= \mu_0 \varepsilon_0 \frac{\partial \mathbf{E}}{\partial t}. && \text{(Ampere's Law)}
\end{aligned}$$

The $\mathbf{E}$ and $\mathbf{B}$ fields can be written as

$$\begin{aligned}
\mathbf{E} &= \hat{\mathbf{E}} \cos(\omega t - \mathbf{k} \cdot \mathbf{r}), \\
\mathbf{B} &= \hat{\mathbf{B}} \cos(\omega t - \mathbf{k} \cdot \mathbf{r}).
\end{aligned}$$

From Faraday's law

$$\begin{aligned}
\nabla \times \mathbf{E} &= \left(\frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} \right) \hat{\mathbf{x}} + \left(\frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x} \right) \hat{\mathbf{y}} + \left(\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} \right) \hat{\mathbf{z}}, \\
&= \left[\left(\hat{E}_z k_y - \hat{E}_y k_z \right) \hat{\mathbf{x}} + \left(\hat{E}_x k_z - \hat{E}_z k_x \right) \hat{\mathbf{y}} + \left(\hat{E}_y k_x - \hat{E}_x k_y \right) \hat{\mathbf{z}} \right] \sin(\omega t - \mathbf{k} \cdot \mathbf{r}), \\
-\frac{\partial \mathbf{B}}{\partial t} &= \omega \hat{\mathbf{B}} \sin(\omega t - \mathbf{k} \cdot \mathbf{r}), \\
\therefore \hat{\mathbf{B}} &= 1/\omega \left[\left(\hat{E}_z k_y - \hat{E}_y k_z \right) \hat{\mathbf{x}} + \left(\hat{E}_x k_z - \hat{E}_z k_x \right) \hat{\mathbf{y}} + \left(\hat{E}_y k_x - \hat{E}_x k_y \right) \hat{\mathbf{z}} \right].
\end{aligned}$$

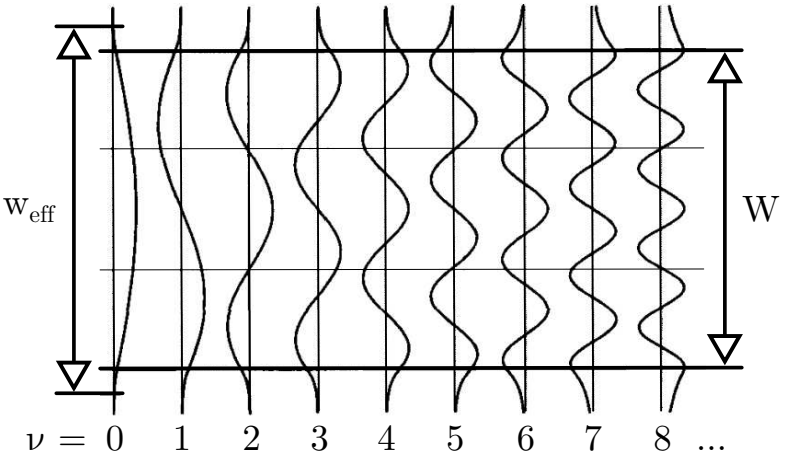


Figure A.1: The lateral field profiles of the first nine guided modes for a step-index multimode waveguide, where w_{eff} is the effective width of the mode and W is the Multimode Interference (MMI) region width. Figure adapted from Soldano and Pennings [56].

Taking the dot product of the $\hat{\mathbf{E}}$ and $\hat{\mathbf{B}}$ vectors

$$\begin{aligned}\omega\hat{\mathbf{E}}\cdot\hat{\mathbf{B}} &= \left[E_x\hat{\mathbf{x}}+E_y\hat{\mathbf{y}}+E_z\hat{\mathbf{z}}\right]\cdot\left[\left(\hat{E}_zk_y-\hat{E}_yk_z\right)\hat{\mathbf{x}}+\left(\hat{E}_xk_z-\hat{E}_zk_x\right)\hat{\mathbf{y}}+\left(\hat{E}_yk_x-\hat{E}_xk_y\right)\hat{\mathbf{z}}\right],\\ &= \hat{E}_x\hat{E}_zk_y-\hat{E}_x\hat{E}_yk_z+\hat{E}_y\hat{E}_xk_z-\hat{E}_y\hat{E}_zk_x+\hat{E}_z\hat{E}_yk_x-\hat{E}_z\hat{E}_xk_y=0.\end{aligned}$$

As the dot product is equal to zero, the $\mathbf{E}$ and $\mathbf{B}$ fields must be orthogonal.

A.2 Derivation for the approximate form for the beat length

Derivation for Equation 2.38 in the main text [56]. Showing that

$$L_\pi = \frac{\pi}{\beta_0-\beta_1}, \tag{A.1}$$

can be approximated by

$$L_\pi \approx \frac{4n_{\text{eff}}w_{\text{eff}}^2}{3\lambda_0}. \tag{A.2}$$

Figure 2.3 shows that k can be broken down into its axial components k_y and k_z

$$k_{y,\nu}^2+k_{z,\nu}^2=k_\nu^2, \tag{A.3}$$

where ν is the mode number with $\nu=0$ representing the fundamental mode.

From Equation 2.28, k_ν can be rewritten as $k_\nu = n_{\text{eff},\nu}k_0$, so Equation A.3 becomes

$$k_{y,\nu}^2 + k_{z,\nu}^2 = n_{\text{eff},\nu}^2 k_0^2, \tag{A.4}$$

where $k_0 = 2\pi/\lambda_0$ and $k_{y,\nu} = (\nu+1)\pi/w_{\text{eff},\nu}$. The effective width, $w_{\text{eff},\nu}$, takes into account the penetration depth of the mode (shown in Figure A.1), and can be approximated to be equal to the width of the MMI region, W, for high contrast waveguide boundaries, such as silicon-air and silicon-silica interfaces. For non-high refractive index contrast boundaries, it is sufficient to approximate $w_{\text{eff},\nu}$ as the effective width for the fundamental mode, $w_{\text{eff},0}$. It is common when referring to the fundamental mode to simply forgo the ν, i.e. $w_{\text{eff},0} = w_{\text{eff}}$.

The propagation constant, $k_{z,\nu}$, can then be found by doing a binomial expansion

$$(x+y)^{1/2} = x^{1/2} + \frac{1}{2}x^{-1/2}y + ... \tag{A.5}$$

in the case of $k_{y,\nu}^2 \ll n_{\text{eff},\nu}^2 k_0^2$, giving

$$\beta_\nu = k_{z,\nu} \simeq \frac{2\pi n_{\text{eff},\nu}}{\lambda_0} - \frac{(\nu+1)^2\pi\lambda_0}{4n_{\text{eff},\nu}w_{\text{eff}}^2}. \tag{A.6}$$

Inserting this into Equation A.1 for the case of $\nu = 0$ and $\nu = 1$ gives Equation A.2. The difference in the effective index for the first two modes is minimal, allowing $n_{\text{eff},\nu}$ to be simply approximated as n_{eff}.

Appendix B

Appendix

B.1 Effect of perturbation size and magnitude

The effect of perturbation length on the transmission of a straight waveguide was modelled using the aperiodic-Fourier Modal Method (a-FMM) (outlined in Section 3.2.2) is shown in Figure B.1. The waveguide was set to be single-mode for incident light of wavelength 1550 nm with a fixed width of 450 nm and a height of 220 nm.

Figure B.1 compares different complex effect refractive index values for perturbations of varying lengths, with the blue line corresponding to the real index shift typically used in this work's perturbation mapping experiments. The red line has the same real refractive index shift with an additional absorption coefficient as calculated from the Drude-Lorentz free-carrier equations, Equation 2.46 and Equation 2.47, on the assumption that the number of free-electrons equals the number of holes [80]. The divergence of these two lines is found to be minimal for perturbations of length less than 1 µm—as is the case for the perturbations in this work—and, as such, allows for the absorptive effects to be neglected. The solid yellow and dashed purple lines show the effects of more extreme changes to the real and imaginary part of the refractive index on transmission, respectively. A greater reduction in transmission can be seen in both these cases resulting from the perturbation. The rippling observed with changing perturbation length arises from the perturbation acting as a simple one-dimensional Fabry-Peŕot cavity [28].

B.2 Complex perturbation mapping

The phase variations at different output ports can be readily measured using coherent techniques, such as homodyne and heterodyne mixing with a Local Oscillator (LO) in a 90° hybrid. This approach was not experimentally verified, but the phase coupling effect can easily be assessed using the analytical model developed in Chapter 4. In Figure B.2,

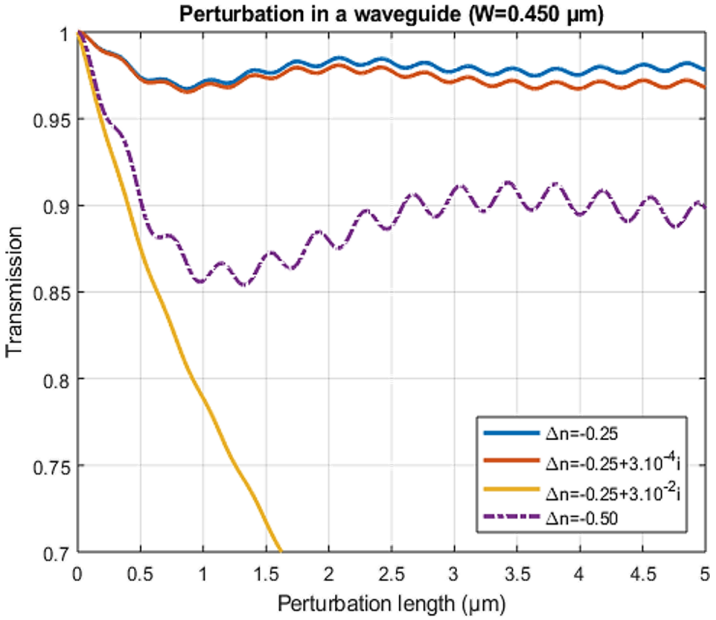


Figure B.1: Effect of the perturbation complex effective index change on transmission for varying perturbation lengths. Simulated using the in-house a-FMM scripts for a 450 nm wide waveguide with incident light in the fundamental Transverse Electrical (TE) mode at a wavelength of 1550 nm.

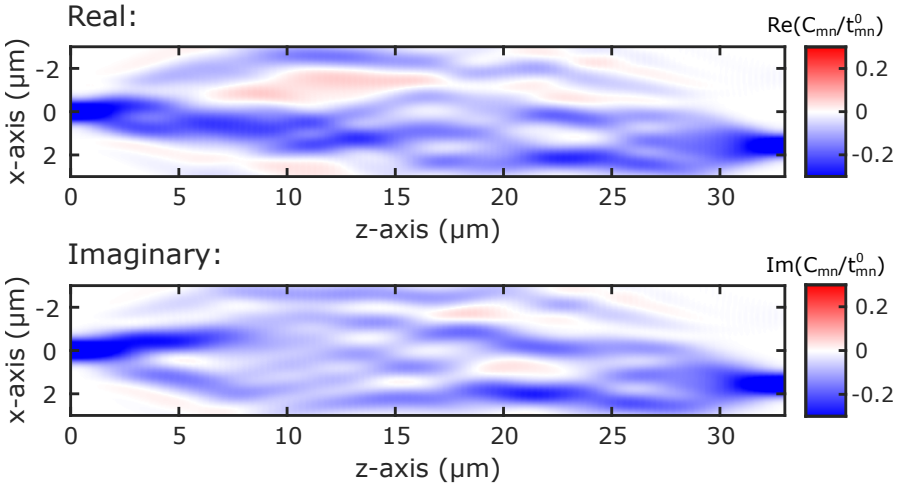


Figure B.2: Comparison of real (top) and imaginary (bottom) components of perturbation mapping for a 1×2 MMI device. Simulations were carried out using a-FMM for incident light in the fundamental TE mode at λ =1550 nm.

we show the real and imaginary parts of C_{mn}/t^0_{mn} for a scanning 750 nm perturbation with $\Delta n_{\text{eff}} = -0.25$ in a 1×2 Multimode Interference (MMI) device.

The real part is very similar to the full $\Delta T/T$ map, as this is the dominating component in Equation 4.3 of the main text when $C_{mn} \ll t^0_{mn}$. The imaginary component yields a map that is overall very similar to the real part but with some slight differences in the fine structure. It is possible that the complex term may become more critical in resonant structures, such as Mach-Zehnder Interferometers (MZIs) or Ring Resonators (RRs).

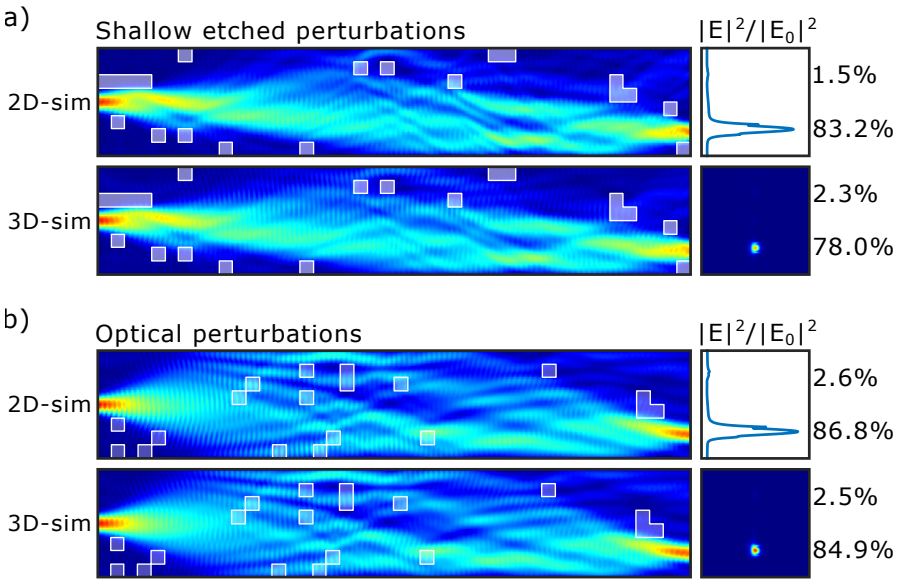


Figure B.3: Comparison of 2D and 3D FDTD simulations for a 1×2 MMI with a) shallow etched (pattern experimentally fabricated and discussed in Section 6.3) and b) weak optical perturbations. Transmission and output profiles are displayed to the right-hand side of the electric field distributions. Simulations were carried out with Lumerical FDTD for incident light in the fundamental TE mode at a wavelength of 1550 nm. The top image corresponds to 2D effective index simulations and the bottom image to full 3D simulation.

B.3 2D vs 3D multiple perturbations FDTD simulations

In order to confirm the validity of the 2D Effective Index Method (EIM) Finite-Difference Time-Domain (FDTD) simulations, comparisons were made with full 3D FDTD simulations. Figure B.3 shows examples of a 1×2 MMI with shallow etched (pattern experimentally fabricated and discussed in Section 6.3) and weak optical perturbations. The 2D effective index approximation is found to be in very good agreement with the full 3D models, typically within a few percent, and we note the significantly faster run time of the order of tens of seconds in 2D opposed to tens of minutes in 3D. Generating datasets of a suitable size for Artificial Neural Network (ANN) training in 3D is possible, but would require significantly more computational time, or power, than we had available. The transmission discrepancy is attributed to the out-of-plane scattering, which is not described by the 2D simulations (discussed further in Appendix B.5). The good agreement, in particular for the case of weak optical perturbations, justifies the use of the 2D effective index approximation for the ANN training data generation.

B.4 Ideal MMI pattern binary threshold search

Deep learning is an optimisation process relying on the gradients of the ANN output with respect to internal weighting parameters, consequently, the output of an ANN is

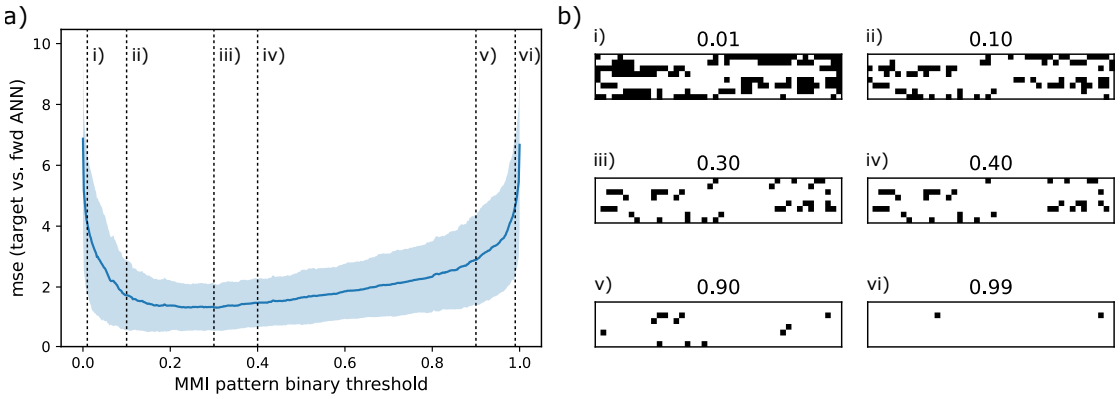


Figure B.4: a) Statistics of ideal binary threshold value search of 150 inverse designed patterns from the final iteration of the 1×2 MMI ANNs. The solid blue line is the MSE between target and forward prediction, and the blue shaded region is the standard deviation of values. b) Example perturbation patterns of one target with different binary threshold values correspond to the numbers above and Roman numerals in a).

inherently continuous with values depending on the chosen activation function. For our problem, a threshold value must therefore be applied to convert the greyscale 'image' of the MMI region into the binary map of perturbations. Different levels of perturbation could be experimentally achieved via varying the etch height; however, this would require multiple fabrication alignment steps and will result in different out-of-plane scattering losses, as shown in Appendix B.5. Regarding the optical perturbations considered in the 3×3 networks, some intensity modulation could be applied to each projected pump spot to achieve similar control over the effective index shift.

Figure B.4 a) shows the Mean Squared Error (MSE) between target and forward network prediction of ideal binary threshold value search of 150 inverse designed patterns from the final iteration of the 1×2 MMI ANNs. The error is highest for threshold values around 0 and 1, which corresponds to patterns with very few and many perturbations, respectively. The binary map examples clearly show this for different threshold values of an inverse design target in Figure B.4 b). Patterns around the optimal value of ∼0.3 are found to be very similar, with the majority of differences corresponding to pixels around the outside of the design that have minimal impact on the device's final performance. On a commercially available Graphics Processing Unit (GPU) the generation of a single pattern is <0.3 ms, resulting in a total inverse design time of <50 ms for testing values between 0 and 1 in 0.01 steps as in this work. If even faster design speeds are required, the range considered can be reduced, or step size increased.

B.5 MMI out-of-plane scattering losses

We have already shown in Figure 3.2 of the main text and Appendix B.1 that 2D simulations are a very effect way of modelling slab waveguide structures that can be described by their effective indices. The main discrepancy between 2D effective index simulations compared to full 3D models is the total lack of any out-of-plane scattering. Here, we perform 3D FDTD simulations to model the out-of-plane scattering for devices with a variety of different perturbation types.

In the Table B.1, we consider out-of-plane scattering for the case of a 500 nm wide rib waveguide with a single perturbation of length 750 nm placed in the middle. This situation corresponds to the reasonable worst-case scenario, in which all the light confined within the waveguide is forced to interact directly with the perturbation. The unperturbed device has negligible out-of-plane scattering over the simulated region. Shallow etched perturbations, with the same etch depth as the rib structure, are found to result in notable loss of transmission due to scattering. Light is primarily lost through the bottom of the device due to the asymmetrical cladding and rib structure. Moving to a full etch depth increased the out-of-plane scattering significantly to nearly 30%. Finally, the optical perturbations are found to induce very little excess scattering, due to their relatively weak magnitude. However, in reality optical perturbations would not be perfect cuboids, instead, there will likely be some distribution of optical power which attenuates through the silicon's depth.

Table B.1: Single perturbation out-of-plane scattering.

out-of-plane scattering	unperturbed	shallow etch	full etch	optical perturbation
top	0.0%	0.6%	9.0%	0.0%
bottom	0.0%	5.6%	19.8%	0.1%

Next, we evaluate the out-of-plane scattering for a 1×2 MMI device with the experimentally fabricated shallow etched perturbation pattern from Section 6.3 and an example optical perturbation pattern. The out-of-plane scattering results for the 3D FDTD simulation are presented in Table B.2. Even the bare, unperturbed device was found to have some loss due to scattering, which is attributed to the imperfect collection of light at the MMI output tapers. The shallow etched device was found to have increased out-of-plane scattering, but only on the order of a few percent. And once again, the optical perturbations showed no significant increase to loss compared to the bare device.

Table B.2: Multiple perturbation pattern out-of-plane scattering.

out-of-plane scattering	unperturbed	shallow etch	optical perturbation
top	0.1%	0.5%	0.2%
bottom	1.1%	2.9%	0.9%

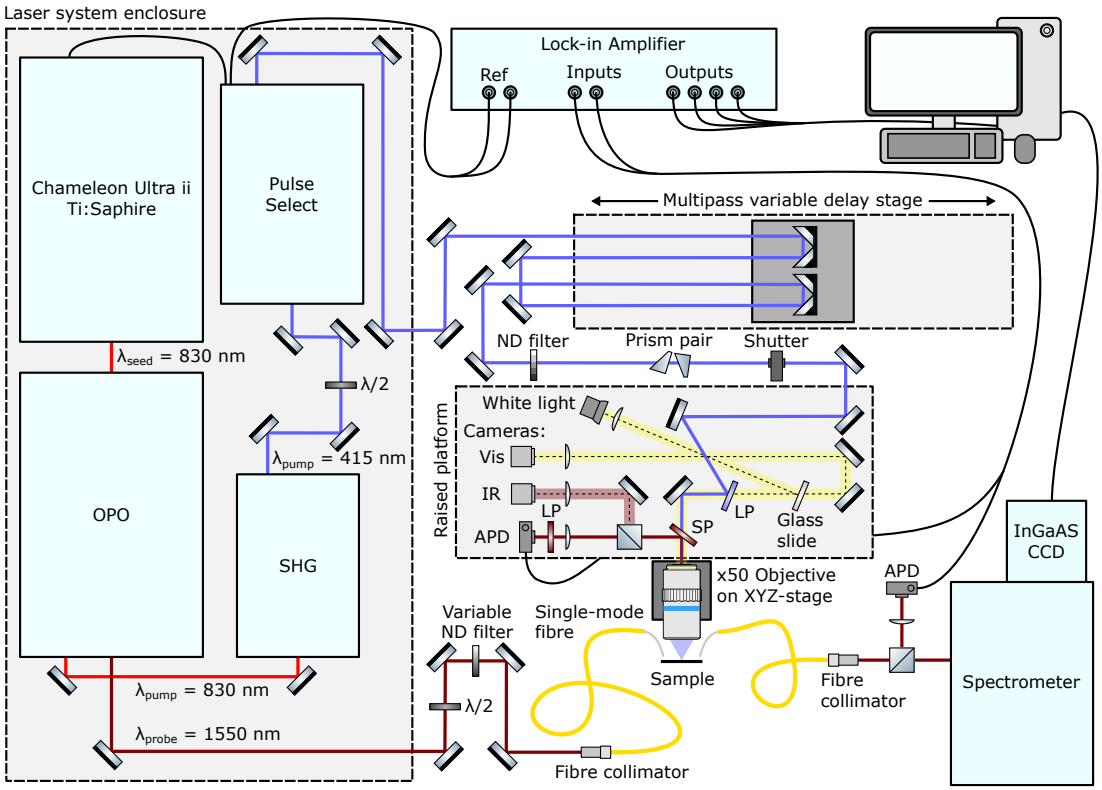


Figure B.5: Schematic of the simultaneous out-of-plane scattering and transmission Ultrafast Photomodulation Spectroscopy (UPMS) setup.

B.6 Out-of-plane scattering photomodulation mapping

In Chapter 4, we demonstrated how UPMS could be used to infer information on the transmission light intensity field, either through the simultaneous or post-measurement combination of all output ports. This was analytically proven for the case of high transmission devices in Section 4.1.2. However, to function most effectively as a diagnostic tool for Photonic Integrated Circuits (PICs), which may be fabricated with poor performance, it is also important to consider the out-of-plan scattering.

The experimental setup was modified, as shown in Figure B.5, to include a second InGaAs Avalanche Photodetector (APD) to record the out-of-plane scattered light signal, S, that is collected by the focusing objective. A 650 nm Shortpass (SP) dichroic mirror was used to separate the scattered probe light from the pump beam, and a 1100 nm Longpass (LP) filter in front of the photodetector removed any residual reflected illumination white light. The output of the APD is connected to the lock-in amplifier, and the signal is demodulated at the same probe and pump frequencies as the transmission signal resulting in S_{probe} and S_{pump}, respectively. The use of a dual-reference and dual-input lock-in amplifier allows for simultaneous transmission and scattering mapping during the experiment. The normalised scattered light change is given by $\Delta S/S$. The value of ΔS does not actually represent the total change of scattered light, as only

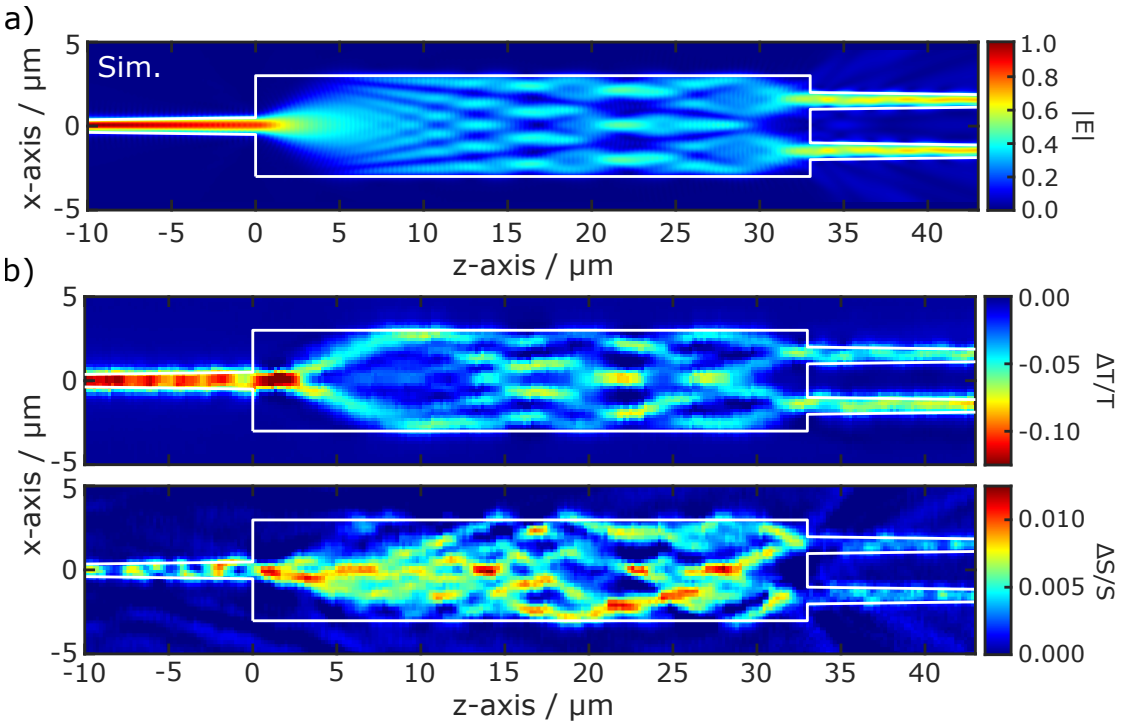


Figure B.6: a) Electric field distribution simulation of a 1×2 MMI device, with length 33 µm and width 6 µm. Simulated using Lumerical FDTD solver for TE input mode at a wavelength of 1550 nm. b) Experimental transmission (top) and out-of-plane (bottom) photomodulation maps of the fabricated device.

the component of light scattered within the Numerical Aperture (NA) of the objective will be collected. For the objective used here, which has an NA of 0.55 this corresponds to cone with an angle of 33.4° from the surface's normal. However, the measured change should be directly proportional to the total scattered light change and therefore should be corrected for when considering the normalised scattered light change.

Figure B.6 shows an example of an out-of-plane scatter photomodulation map compared to the FDTD simulated electric field distribution. Additionally shown is the transmission photomodulation map for the combined ports, which was obtained using a dual fibre array to enable both outputs' simultaneous measurement. It is immediately clear that the optical pump has a much greater effect on transmission than on the out-of-plane scattering. It is also clear that the electric field distribution information is much better retained in the transmission map; nevertheless, the scattering map allows all the light within a system to be assessed, not just the transmitted light, and it does not require any output waveguide coupling.

References

[1] J.-D. Colladon, "Sur les réflexions d'un rayon de lumière à l'intérieur d'une veine liquide parabolique. (French) [On the reflections of a ray of light inside a parabolic liquid vein]," *Comptes Rendus*, vol. 15, pp. 800–802, 1842.

[2] B. Jalali and S. Fathpour, "Silicon photonics," *Journal of Lightwave Technology*, vol. 24, no. 12, pp. 4600–4615, 2006.

[3] Y. Shen, N. C. Harris, S. Skirlo, M. Prabhu, T. Baehr-Jones, M. Hochberg, X. Sun, S. Zhao, H. Larochelle, D. Englund, *et al.*, "Deep learning with coherent nanophotonic circuits," *Nature Photonics*, vol. 11, no. 7, p. 441, 2017.

[4] J. Feldmann, N. Youngblood, C. D. Wright, H. Bhaskaran, and W. Pernice, "All-optical spiking neurosynaptic networks with self-learning capabilities," *Nature*, vol. 569, no. 7755, pp. 208–214, 2019.

[5] N. C. Harris, G. R. Steinbrecher, M. Prabhu, Y. Lahini, J. Mower, D. Bunandar, C. Chen, F. N. Wong, T. Baehr-Jones, M. Hochberg, *et al.*, "Quantum transport simulations in a programmable nanophotonic processor," *Nature Photonics*, vol. 11, no. 7, p. 447, 2017.

[6] C. García-Meca, S. Lechago, A. Brimont, A. Griol, S. Mas, L. Sánchez, L. Bellieres, N. S. Losilla, and J. Martí, "On-chip wireless silicon photonics: from reconfigurable interconnects to lab-on-chip devices," *Light: Science & Applications*, vol. 6, no. 9, pp. e17053–e17053, 2017.

[7] R. Noyce, "Semiconductor device-and-lead structure," Apr. 25 1961. US Patent 2,981,877.

[8] M. Streshinsky, R. Ding, Y. Liu, A. Novack, C. Galland, A.-J. Lim, P. G.-Q. Lo, T. Baehr-Jones, and M. Hochberg, "The road to affordable, large-scale silicon photonics," *Optics and Photonics News*, vol. 24, no. 9, pp. 32–39, 2013.

[9] Y. A. Vlasov, "Silicon CMOS-integrated nano-photonics for computer and data communications beyond 100G," *IEEE Communications Magazine*, vol. 50, no. 2, 2012.

[10] G. Bae, D.-I. Bae, M. Kang, S. Hwang, S. Kim, B. Seo, T. Kwon, T. Lee, C. Moon, Y. Choi, *et al.*, "3nm GAA technology featuring multi-bridge-channel FET for low power and high performance applications," in *2018 IEEE International Electron Devices Meeting (IEDM)*, pp. 28–7, IEEE, 2018.

[11] J. Steinlechner, C. Krüger, N. Lastzka, S. Steinlechner, A. Khalaidovski, and R. Schnabel, "Optical absorption measurements on crystalline silicon test masses at 1550 nm," *Classical and Quantum Gravity*, vol. 30, no. 9, p. 095007, 2013.

[12] C. Gunn, "CMOS photonics for high-speed interconnects," *IEEE micro*, vol. 26, no. 2, pp. 58–66, 2006.

[13] C. Batten, A. Joshi, J. Orcutt, A. Khilo, B. Moss, C. W. Holzwarth, M. A. Popovic, H. Li, H. I. Smith, J. L. Hoyt, *et al.*, "Building many-core processor-to-dram networks with monolithic CMOS silicon photonics," *IEEE Micro*, vol. 29, no. 4, 2009.

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

[15] J.-L. M. Alexis Debray and E. Mounier, "Silicon Photonics: Market and Technology Report 2020," tech. rep., Yole Développement, April 2020.

[16] P. Dong, S. Liao, D. Feng, H. Liang, D. Zheng, R. Shafiiha, C.-C. Kung, W. Qian, G. Li, X. Zheng, *et al.*, "Low Vpp, ultralow-energy, compact, high-speed silicon electro-optic modulator," *Optics Express*, vol. 17, no. 25, pp. 22484–22490, 2009.

[17] D. A. B. Miller, "Energy consumption in optical modulators for interconnects," *Optics Express*, vol. 20, no. 102, pp. A293–A308, 2012.

[18] E. Timurdogan, C. M. Sorace-Agaskar, J. Sun, E. S. Hosseini, A. Biberman, and M. R. Watts, "An ultralow power athermal silicon modulator," *Nature Communications*, vol. 5, p. 4008, 2014.

[19] D. Thomson, A. Zilkie, J. E. Bowers, T. Komljenovic, G. T. Reed, L. Vivien, D. Marris-Morini, E. Cassan, L. Virot, J.-M. Fédéli, *et al.*, "Roadmap on silicon photonics," *Journal of Optics*, vol. 18, no. 7, p. 073003, 2016.

[20] R. Soref, "Silicon photonics: a review of recent literature," *Silicon*, vol. 2, no. 1, pp. 1–6, 2010.

[21] V. Passaro, C. d. Tullio, B. Troia, M. L. Notte, G. Giannoccaro, and F. D. Leonardis, "Recent advances in integrated photonic sensors," *Sensors*, vol. 12, no. 11, pp. 15558–15598, 2012.

[22] A. Densmore, D.-X. Xu, P. Waldron, S. Janz, P. Cheben, J. Lapointe, A. Delge, B. Lamontagne, J. Schmid, and E. Post, "A silicon-on-insulator photonic wire

based evanescent field sensor," *IEEE Photonics Technology Letters*, vol. 18, no. 23, pp. 2520–2522, 2006.

[23] M. Iqbal, M. A. Gleeson, B. Spaugh, F. Tybor, W. G. Gunn, M. Hochberg, T. Baehr-Jones, R. C. Bailey, and L. C. Gunn, "Label-free biosensor arrays based on silicon ring resonators and high-speed optical scanning instrumentation," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 16, no. 3, pp. 654–661, 2010.

[24] Y. Chen, H. Lin, J. Hu, and M. Li, "Heterogeneously integrated silicon photonics for the mid-infrared and spectroscopic sensing," *ACS Nano*, vol. 8, no. 7, pp. 6955–6961, 2014.

[25] R. Loiacono, G. T. Reed, G. Z. Mashanovich, R. Gwilliam, S. J. Henley, Y. Hu, R. Feldesh, and R. Jones, "Laser erasable implanted gratings for integrated silicon photonics," *Optics Express*, vol. 19, no. 11, pp. 10728–10734, 2011.

[26] X. Chen, M. M. Milosevic, A. F. Runge, X. Yu, A. Z. Khokhar, S. Mailis, D. J. Thomson, A. C. Peacock, S. Saito, and G. T. Reed, "Silicon erasable waveguides and directional couplers by germanium ion implantation for configurable photonic circuits," *Optics Express*, vol. 28, no. 12, pp. 17630–17642, 2020.

[27] R. P. Feynman, R. B. Leighton, and M. Sands, *The Feynman lectures on physics, Vol. 2: mainly electromagnetism and matter*. New York: Basic Books, 2011.

[28] E. Hecht, *Optics*. Pearson education, Addison Wesley, 2002.

[29] G. T. Reed and A. P. Knights, *Silicon photonics: an introduction*. John Wiley & Sons, 2004.

[30] M. J. Adams, *An introduction to optical waveguides*, vol. 14. Wiley, 1981.

[31] K. Okamoto, *Fundamentals of optical waveguides*. Academic press, 2006.

[32] C. Pollock and M. Lipson, *Integrated photonics*, vol. 20. Springer, 2003.

[33] C. A. Barrios, K. B. Gylfason, B. Sánchez, A. Griol, H. Sohlström, M. Holgado, and R. Casquel, "Slot-waveguide biochemical sensor," *Optics Letters*, vol. 32, no. 21, pp. 3080–3082, 2007.

[34] T. Claes, J. G. Molera, K. De Vos, E. Schacht, R. Baets, and P. Bienstman, "Label-free biosensing with a slot-waveguide-based ring resonator in silicon on insulator," *IEEE Photonics journal*, vol. 1, no. 3, pp. 197–204, 2009.

[35] M. Butt, S. Khonina, and N. Kazanskiy, "Silicon on silicon dioxide slot waveguide evanescent field gas absorption sensor," *Journal of Modern optics*, vol. 65, no. 2, pp. 174–178, 2018.

[36] R. G. Hunsperger, *Integrated Optics: Theory and Technology*, vol. 4. Springer, 1995.

[37] K. Chiang, "Dual effective-index method for the analysis of rectangular dielectric waveguides," *Applied Optics*, vol. 25, no. 13, pp. 2169–2174, 1986.

[38] A. Ivanova, R. Stoffer, and M. Hammer, "A variational mode solver for optical waveguides based on quasi-analytical vectorial slab mode expansion," *arXiv preprint arXiv:1307.1315*, 2013.

[39] R. Marchetti, C. Lacava, L. Carroll, K. Gradkowski, and P. Minzioni, "Coupling strategies for silicon photonics integrated chips," *Photonics Research*, vol. 7, no. 2, pp. 201–239, 2019.

[40] H. S. Dutta, A. K. Goyal, V. Srivastava, and S. Pal, "Coupling light in photonic crystal waveguides: A review," *Photonics and Nanostructures-Fundamentals and Applications*, vol. 20, pp. 41–58, 2016.

[41] L. Cheng, S. Mao, Z. Li, Y. Han, and H. Fu, "Grating couplers on silicon photonics: Design principles, emerging trends and practical issues," *Micromachines*, vol. 11, no. 7, p. 666, 2020.

[42] R. P. Feynman, R. B. Leighton, and M. Sands, *The Feynman lectures on physics, Vol. 1: mainly mechanics, radiation, and heat.* New York: Basic Books, 2011.

[43] T. Suhara and H. Nishihara, "Integrated optics components and devices using periodic structures," *IEEE Journal of Quantum Electronics*, vol. 22, no. 6, pp. 845–867, 1986.

[44] P. Cheben, S. Janz, D.-X. Xu, B. Lamontagne, A. Delâge, and S. Tanev, "A broadband waveguide grating coupler with a subwavelength grating mirror," *IEEE photonics technology letters*, vol. 18, no. 1, pp. 13–15, 2005.

[45] C. Li, H. Zhang, M. Yu, and G. Lo, "CMOS-compatible high efficiency double-etched apodized waveguide grating coupler," *Optics Express*, vol. 21, no. 7, pp. 7868–7874, 2013.

[46] R. Marchetti, C. Lacava, A. Khokhar, X. Chen, I. Cristiani, D. J. Richardson, G. T. Reed, P. Petropoulos, and P. Minzioni, "High-efficiency grating-couplers: demonstration of a new design strategy," *Scientific Reports*, vol. 7, no. 1, pp. 1–8, 2017.

[47] D. Taillaert, H. Chong, P. I. Borel, L. H. Frandsen, R. M. De La Rue, and R. Baets, "A compact two-dimensional grating coupler used as a polarization splitter," *IEEE Photonics Technology Letters*, vol. 15, no. 9, pp. 1249–1251, 2003.

[48] Y. Luo, Z. Nong, S. Gao, H. Huang, Y. Zhu, L. Liu, L. Zhou, J. Xu, L. Liu, S. Yu, *et al.*, "Low-loss two-dimensional silicon photonic grating coupler with a backside metal mirror," *Optics Letters*, vol. 43, no. 3, pp. 474–477, 2018.

[49] I. Day, I. Evans, A. Knights, F. Hopper, S. Roberts, J. Johnston, S. Day, J. Luff, H. Tsang, and M. Asghari, "Tapered silicon waveguides for low insertion loss highly-efficient high-speed electronic variable optical attenuators," in *Optical Fiber Communication Conference*, Optical Society of America, 2003.

[50] S. Somekh, E. Garmire, A. Yariv, H. Garvin, and R. Hunsperger, "Channel optical waveguide directional couplers," *Applied physics letters*, vol. 22, no. 1, pp. 46–47, 1973.

[51] Z. Zhang, X. Hu, and J. Wang, "On-chip optical mode exchange using tapered directional coupler," *Scientific Reports*, vol. 5, no. 1, pp. 1–7, 2015.

[52] M. Papuchon, A. Roy, and D. Ostrowsky, "Electrically active optical bifurcation: BOA," *Applied Physics Letters*, vol. 31, no. 4, pp. 266–267, 1977.

[53] F. Rottmann, A. Neyer, W. Mevenkamp, and E. Voges, "Integrated-optic wavelength multiplexers on lithium niobate based on two-mode interference," *Journal of Lightwave Technology*, vol. 6, no. 6, pp. 946–952, 1988.

[54] L. B. Soldano, F. B. Veerman, M. K. Smit, B. H. Verbeek, A. H. Dubost, and E. C. Pennings, "Planar monomode optical couplers based on multimode interference effects," *Journal of Lightwave Technology*, vol. 10, no. 12, pp. 1843–1850, 1992.

[55] T. Uematsu, Y. Ishizaka, Y. Kawaguchi, K. Saitoh, and M. Koshiba, "Design of a compact two-mode multi/demultiplexer consisting of multimode interference waveguides and a wavelength-insensitive phase shifter for mode-division multiplexing transmission," *Journal of Lightwave Technology*, vol. 30, no. 15, pp. 2421–2426, 2012.

[56] L. B. Soldano and E. C. Pennings, "Optical multi-mode interference devices based on self-imaging: principles and applications," *Journal of Lightwave Technology*, vol. 13, no. 4, pp. 615–627, 1995.

[57] S.-Y. Tseng, C. Fuentes-Hernandez, D. Owens, and B. Kippelen, "Variable splitting ratio 2×2 MMI couplers using multimode waveguide holograms," *Optics Express*, vol. 15, no. 14, pp. 9015–9021, 2007.

[58] A. Martinez, A. Griol, P. Sanchis, and J. Marti, "Mach–Zehnder interferometer employing coupled-resonator optical waveguides," *Optics Letters*, vol. 28, no. 6, pp. 405–407, 2003.

[59] L. Liao, D. Samara-Rubio, M. Morse, A. Liu, D. Hodge, D. Rubin, U. D. Keil, and T. Franck, "High speed silicon Mach-Zehnder modulator," *Optics Express*, vol. 13, no. 8, pp. 3129–3135, 2005.

[60] W. M. Green, M. J. Rooks, L. Sekaric, and Y. A. Vlasov, "Ultra-compact, low RF power, 10 Gb/s silicon Mach-Zehnder modulator," *Optics Express*, vol. 15, no. 25, pp. 17106–17113, 2007.

[61] D. Pérez, C. Baudot, T. Hirtzlin, S. Messaoudène, N. Vulliet, P. Crozat, F. Boeuf, L. Vivien, and D. Marris-Morini, "Low voltage 25Gbps silicon Mach–Zehnder modulator in the O-band," *Optics Express*, vol. 25, no. 10, pp. 11217–11222, 2017.

[62] I. Rendón-Salgado, E. Ramírez-Cruz, and R. Gutiérrez-Castrejón, "640 Gb/s all-optical AND gate and wavelength converter using bulk SOA turbo–switched Mach–Zehnder interferometer with improved differential scheme," *Optics & Laser Technology*, vol. 109, pp. 671–681, 2019.

[63] S. Armaghani, S. Khani, and M. Danaie, "Design of all-optical graphene switches based on a Mach-Zehnder interferometer employing optical Kerr effect," *Superlattices and Microstructures*, vol. 135, p. 106244, 2019.

[64] S. Zhao, L. Lu, L. Zhou, D. Li, Z. Guo, and J. Chen, "16×16 silicon Mach–Zehnder interferometer switch actuated with waveguide microheaters," *Photonics Research*, vol. 4, no. 5, pp. 202–207, 2016.

[65] W. M. Green, R. K. Lee, G. A. DeRose, A. Scherer, and A. Yariv, "Hybrid InGaAsP-InP Mach-Zehnder racetrack resonator for thermooptic switching and coupling control," *Optics Express*, vol. 13, no. 5, pp. 1651–1659, 2005.

[66] Y. Baek, R. Schiek, and G. Stegeman, "All-optical switching in a hybrid Mach–Zehnder interferometer as a result of cascaded second-order nonlinearity," *Optics Letters*, vol. 20, no. 21, pp. 2168–2170, 1995.

[67] Q. Huang, W. Jin, and K. S. Chiang, "Broadband mode switch based on a three-dimensional waveguide Mach–Zehnder interferometer," *Optics Letters*, vol. 42, no. 23, pp. 4877–4880, 2017.

[68] C. Sun, Y. Yu, G. Chen, and X. Zhang, "Integrated switchable mode exchange for reconfigurable mode-multiplexing optical networks," *Optics Letters*, vol. 41, no. 14, pp. 3257–3260, 2016.

[69] A. Tervonen, P. Poyhonen, S. Honkanen, and M. Tahkokorpi, "A guided-wave Mach-Zehnder interferometer structure for wavelength multiplexing," *IEEE photonics technology letters*, vol. 3, no. 6, pp. 516–518, 1991.

[70] B. Luff, J. S. Wilkinson, J. Piehler, U. Hollenbach, J. Ingenhoff, and N. Fabricius, "Integrated optical Mach-Zehnder biosensor," *Journal of Lightwave Technology*, vol. 16, no. 4, p. 583, 1998.

[71] X. Guan, X. Wang, and L. H. Frandsen, "Optical temperature sensor with enhanced sensitivity by employing hybrid waveguides in a silicon Mach-Zehnder interferometer," *Optics Express*, vol. 24, no. 15, pp. 16349–16356, 2016.

[72] B. Ouyang, Y. Li, M. Kruidhof, R. Horsten, K. W. van Dongen, and J. Caro, "On-chip silicon Mach–Zehnder interferometer sensor for ultrasound detection," *Optics Letters*, vol. 44, no. 8, pp. 1928–1931, 2019.

[73] R. S. El Shamy, D. Khalil, and M. A. Swillam, "Mid infrared optical gas sensor using plasmonic Mach-Zehnder interferometer," *Scientific Reports*, vol. 10, no. 1, pp. 1–9, 2020.

[74] G. T. Reed, G. Mashanovich, F. Gardes, and D. Thomson, "Silicon optical modulators," *Nature Photonics*, vol. 4, no. 8, p. 518, 2010.

[75] C. Schriever, C. Bohley, and R. B. Wehrspohn, "Strain dependence of second-harmonic generation in silicon," *Optics Letters*, vol. 35, no. 3, pp. 273–275, 2010.

[76] B. Chmielak, M. Waldow, C. Matheisen, C. Ripperda, J. Bolten, T. Wahlbrink, M. Nagel, F. Merget, and H. Kurz, "Pockels effect based fully integrated, strained silicon electro-optic modulator," *Optics Express*, vol. 19, no. 18, pp. 17212–17219, 2011.

[77] P. Damas, X. Le Roux, D. Le Bourdais, E. Cassan, D. Marris-Morini, N. Izard, T. Maroutian, P. Lecoeur, and L. Vivien, "Wavelength dependence of Pockels effect in strained silicon waveguides," *Optics Express*, vol. 22, no. 18, pp. 22095–22100, 2014.

[78] M. Dinu, F. Quochi, and H. Garcia, "Third-order nonlinearities in silicon at telecom wavelengths," *Applied Physics Letters*, vol. 82, no. 18, pp. 2954–2956, 2003.

[79] Q. Lin, O. J. Painter, and G. P. Agrawal, "Nonlinear optical phenomena in silicon waveguides: modeling and applications," *Optics Express*, vol. 15, no. 25, pp. 16604–16644, 2007.

[80] R. Soref and B. Bennett, "Electrooptical effects in silicon," *IEEE journal of quantum electronics*, vol. 23, no. 1, pp. 123–129, 1987.

[81] G. Cocorullo and I. Rendina, "Thermo-optical modulation at 1.5 um in silicon etalon," *Electronics Letters*, vol. 28, no. 1, pp. 83–85, 1992.

[82] W. A. Zortman, D. C. Trotter, and M. R. Watts, "Silicon photonics manufacturing," *Optics Express*, vol. 18, no. 23, pp. 23598–23607, 2010.

[83] A. Liu, R. Jones, L. Liao, D. Samara-Rubio, D. Rubin, O. Cohen, R. Nicolaescu, and M. Paniccia, "A high-speed silicon optical modulator based on a metal–oxide–semiconductor capacitor," *Nature*, vol. 427, no. 6975, p. 615, 2004.

[84] K. Debnath, D. J. Thomson, W. Zhang, A. Z. Khokhar, C. Littlejohns, J. Byers, L. Mastronardi, M. K. Husain, K. Ibukuro, F. Y. Gardes, *et al.*, "All-silicon carrier accumulation modulator based on a lateral metal-oxide-semiconductor capacitor," *Photonics Research*, vol. 6, no. 5, pp. 373–379, 2018.

[85] D. J. Thomson, F. Y. Gardes, J.-M. Fedeli, S. Zlatanovic, Y. Hu, B. P. P. Kuo, E. Myslivets, N. Alic, S. Radic, G. Z. Mashanovich, *et al.*, "50-Gb/s silicon optical modulator," *IEEE Photonics Technology Letters*, vol. 24, no. 4, pp. 234–236, 2011.

[86] D. Thomson, F. Gardes, Y. Hu, G. Mashanovich, M. Fournier, P. Grosse, J. Fedeli, and G. Reed, "High contrast 40Gbit/s optical modulation in silicon," *Optics Express*, vol. 19, no. 12, pp. 11507–11516, 2011.

[87] C. Monat and Y. Su, "Hybrid photonics beyond silicon," *APL Photonics*, vol. 5, p. 020402, 2020.

[88] K. Yee, "Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media," *IEEE Transactions on antennas and propagation*, vol. 14, no. 3, pp. 302–307, 1966.

[89] W. Yu and R. Mittra, "A conformal finite difference time domain technique for modeling curved dielectric surfaces," *IEEE Microwave and Wireless Components Letters*, vol. 11, no. 1, pp. 25–27, 2001.

[90] A. Mohammadi, H. Nadgaran, and M. Agio, "Contour-path effective permittivities for the two-dimensional finite-difference time-domain method," *Optics Express*, vol. 13, no. 25, pp. 10367–10381, 2005.

[91] Y. Zhao and Y. Hao, "Finite-difference time-domain study of guided modes in nano-plasmonic waveguides," *IEEE transactions on antennas and propagation*, vol. 55, no. 11, pp. 3070–3077, 2007.

[92] M. Moharam, E. B. Grann, D. A. Pommet, and T. Gaylord, "Formulation for stable and efficient implementation of the rigorous coupled-wave analysis of binary gratings," *Journal of the Optical Society of America A*, vol. 12, no. 5, pp. 1068–1076, 1995.

[93] D. A. Bykov, E. A. Bezus, and L. L. Doskolovich, "Use of aperiodic Fourier modal method for calculating complex-frequency eigenmodes of long-period photonic crystal slabs," *Optics Express*, vol. 25, no. 22, pp. 27298–27309, 2017.

[94] P. Lalanne and E. Silberstein, "Fourier-modal methods applied to waveguide computational problems," *Optics Letters*, vol. 25, no. 15, pp. 1092–1094, 2000.

[95] E. Silberstein, P. Lalanne, J.-P. Hugonin, and Q. Cao, "Use of grating theories in integrated optics," *Journal of the Optical Society of America A*, vol. 18, no. 11, pp. 2865–2875, 2001.

[96] P. Lalanne, M. Besbes, J. Hugonin, S. Van Haver, O. Janssen, A. Nugrowati, M. Xu, S. Pereira, H. Urbach, A. Van de Nes, *et al.*, "Numerical analysis of a slit-groove diffraction problem," *Journal of the European Optical Society-Rapid Publications*, vol. 2, 2007.

[97] R. Bruck, C. Liu, O. L. Muskens, A. Fratalocchi, and A. Falco, "Ultrafast all-optical order-to-chaos transition in silicon photonic crystal chips," *Laser & Photonics Reviews*, vol. 10, no. 4, pp. 688–695, 2016.

[98] K. Vynck, N. J. Dinsdale, B. Chen, R. Bruck, A. Z. Khokhar, S. A. Reynolds, L. Crudgington, D. J. Thomson, G. T. Reed, P. Lalanne, *et al.*, "Ultrafast perturbation maps as a quantitative tool for testing of multi-port photonic devices," *Nature Communications*, vol. 9, no. 1, pp. 1–10, 2018.

[99] J.-P. Hugonin and P. Lalanne, "Light-in-complex-nanostructures/reticolo: V9," jan 2021.

[100] O. L. Muskens, L. Bergamini, Y. Wang, J. M. Gaskell, N. Zabala, C. De Groot, D. W. Sheel, and J. Aizpurua, "Antenna-assisted picosecond control of nanoscale phase transition in vanadium dioxide," *Light: Science & Applications*, vol. 5, no. 10, pp. e16173–e16173, 2016.

[101] M. C. Fischer, J. W. Wilson, F. E. Robles, and W. S. Warren, "Invited review article: pump-probe microscopy," *Review of Scientific Instruments*, vol. 87, no. 3, p. 031101, 2016.

[102] S. S. Lo, H. Y. Shi, L. Huang, and G. V. Hartland, "Imaging the extent of plasmon excitation in au nanowires using pump-probe microscopy," *Optics Letters*, vol. 38, no. 8, pp. 1265–1267, 2013.

[103] G. Greetham, D. Sole, I. Clark, A. Parker, M. Pollard, and M. Towrie, "Time-resolved multiple probe spectroscopy," *Review of Scientific Instruments*, vol. 83, no. 10, p. 103107, 2012.

[104] M. Kolarczik, C. Ulbrich, P. Geiregat, Y. Zhu, L. K. Sagar, A. Singh, B. Herzog, A. W. Achtstein, X. Li, D. Van Thourhout, *et al.*, "Sideband pump-probe technique resolves nonlinear modulation response of PbS/CdS quantum dots on a silicon nitride waveguide," *APL Photonics*, vol. 3, no. 1, p. 016101, 2018.

[105] R. J. Suess, M. M. Jadidi, K. Kim, and T. E. Murphy, "Characterization of optical nonlinearities in nanoporous silicon waveguides via pump-probe heterodyning technique," *Optics Express*, vol. 22, no. 14, pp. 17466–17477, 2014.

[106] K. L. Hall, G. Lenz, E. P. Ippen, and G. Raybon, "Heterodyne pump–probe technique for time-domain studies of optical nonlinearities in waveguides," *Optics Letters*, vol. 17, no. 12, pp. 874–876, 1992.

[107] K. Sokolowski-Tinten and D. von der Linde, "Generation of dense electron-hole plasmas in silicon," *Physical Review B*, vol. 61, no. 4, pp. 2643–2650, 2000.

[108] R. Bruck, B. Mills, B. Troia, D. J. Thomson, F. Y. Gardes, Y. Hu, G. Z. Mashanovich, V. M. N. Passaro, G. T. Reed, and O. L. Muskens, "Device-level

characterization of the flow of light in integrated photonic circuits using ultrafast photomodulation spectroscopy," *Nat Photon*, vol. 9, pp. 54–60, Nov 2015.

[109] C. Jacoboni, C. Canali, G. Ottaviani, and A. A. Quaranta, "A review of some charge transport properties of silicon," *Solid-State Electronics*, vol. 20, no. 2, pp. 77–89, 1977.

[110] A. E. Siegman, *Lasers*. University Science Books, 1986.

[111] "Lasers and laser-related equipment-test methods for laser beam widths, divergence angles and beam propagation ratios," *ISO Standard*, pp. 11146–2, 2005.

[112] R. Bruck, K. Vynck, P. Lalanne, B. Mills, D. J. Thomson, G. Z. Mashanovich, G. T. Reed, and O. L. Muskens, "All-optical spatial light modulator for reconfigurable silicon photonic circuits," *Optica*, vol. 3, no. 4, pp. 396–402, 2016.

[113] T. Baehr-Jones, T. Pinguet, P. L. Guo-Qiang, S. Danziger, D. Prather, and M. Hochberg, "Myths and rumours of silicon photonics," *Nature Photonics*, vol. 6, no. 4, pp. 206–208, 2012.

[114] A. E.-J. Lim, J. Song, Q. Fang, C. Li, X. Tu, N. Duan, K. K. Chen, R. P.-C. Tern, and T.-Y. Liow, "Review of silicon photonics foundry efforts," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 20, no. 4, pp. 405–416, 2013.

[115] R. Polster, L. Y. Dai, O. A. Jimenez, Q. Cheng, M. Lipson, and K. Bergman, "Wafer-scale high-density edge coupling for high throughput testing of silicon photonics," in *2018 Optical Fiber Communications Conference and Exposition (OFC)*, pp. 1–3, IEEE, 2018.

[116] T. Pinguet, S. Denton, S. Gloeckner, M. Mack, G. Masini, A. Mekis, S. Pang, M. Peterson, S. Sahni, and P. De Dobbelaere, "High-volume manufacturing platform for silicon photonics," *Proceedings of the IEEE*, vol. 106, no. 12, pp. 2281–2290, 2018.

[117] Y. Maeda, T. Miura, H. Fukuda, and S. Matsuo, "Novel fiber alignment method for on-wafer testing of silicon photonic devices with pn junction embedded grating couplers," in *2018 IEEE 15th International Conference on Group IV Photonics (GFP)*, pp. 1–2, IEEE, 2018.

[118] S. Jordan, "Commercial tests and assembly in silicon photonics: Speeding production of single-and multi-channel photonic devices with breakthrough parallel nanoalignment," *Optik & Photonik*, vol. 12, no. 3, pp. 31–33, 2017.

[119] J. De Coster, P. De Heyn, M. Pantouvaki, B. Snyder, H. Chen, E. J. Marinissen, P. Absil, J. Van Campenhout, and B. Bolt, "Test-station for flexible semi-automatic wafer-level silicon photonics testing," in *2016 21th IEEE European Test Symposium (ETS)*, pp. 1–6, IEEE, 2016.

[120] T. Horikawa, D. Shimura, S.-H. Jeong, M. Tokushima, K. Kinoshita, and T. Mogami, "Process control and monitoring in device fabrication for optical interconnection using silicon photonics technology," in *2015 IEEE International Interconnect Technology Conference and 2015 IEEE Materials for Advanced Metallization Conference (IITC/MAM)*, pp. 277–280, IEEE, 2015.

[121] A. Mekis, S. Gloeckner, G. Masini, A. Narasimha, T. Pinguet, S. Sahni, and P. De Dobbelaere, "A grating-coupler-enabled cmos photonics platform," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 17, no. 3, pp. 597–608, 2010.

[122] R. Topley, L. O'Faolain, D. Thomson, F. Gardes, G. Mashanovich, and G. Reed, "Planar surface implanted diffractive grating couplers in SOI," *Optics Express*, vol. 22, no. 1, pp. 1077–1084, 2014.

[123] R. Topley, G. Martinez-Jimenez, L. O'Faolain, N. Healy, S. Mailis, D. J. Thomson, F. Y. Gardes, A. C. Peacock, D. N. R. Payne, G. Z. Mashanovich, and G. T. Reed, "Locally erasable couplers for optical device testing in silicon on insulator," *Journal of Lightwave Technology*, vol. 32, no. 12, pp. 2248–2253, 2014.

[124] A. H. Hartog, *An introduction to distributed optical fibre sensors*. CRC press, 2017.

[125] U. Glombitza and E. Brinkmeyer, "Coherent frequency-domain reflectometry for characterization of single-mode integrated-optical waveguides," *Journal of Lightwave Technology*, vol. 11, no. 8, pp. 1377–1384, 1993.

[126] N. Rotenberg and L. Kuipers, "Mapping nanoscale light fields," *Nature Photonics*, vol. 8, no. 12, p. 919, 2014.

[127] L. Gilburd, X. G. Xu, Y. Bando, D. Golberg, and G. C. Walker, "Near-field infrared pump–probe imaging of surface phonon coupling in boron nitride nanotubes," *The Journal of Physical Chemistry Letters*, vol. 7, no. 2, pp. 289–294, 2016.

[128] R. J. Engelen, Y. Sugimoto, H. Gersen, N. Ikeda, K. Asakawa, and L. K. Kuipers, "Ultrafast evolution of photonic eigenstates in k-space," *Nature Physics*, vol. 3, no. 6, pp. 401–405, 2007.

[129] L. Lalouat, B. Cluzel, P. Velha, E. Picard, D. Peyrade, J.-P. Hugonin, P. Lalanne, E. Hadji, and F. De Fornel, "Near-field interactions between a subwavelength tip and a small-volume photonic-crystal nanocavity," *Physical Review B*, vol. 76, no. 4, p. 041102, 2007.

[130] A. F. Koenderink, M. Kafesaki, B. C. Buchler, and V. Sandoghdar, "Controlling the resonance of a photonic crystal microcavity by a near-field probe," *Physical Review Letters*, vol. 95, no. 15, p. 153904, 2005.

[131] R. C. Dunn, "Near-field scanning optical microscopy," *Chemical reviews*, vol. 99, no. 10, pp. 2891–2928, 1999.

[132] S. Romero-García, F. Merget, F. Zhong, H. Finkelstein, and J. Witzens, "Silicon nitride CMOS-compatible platform for integrated photonics applications at visible wavelengths," *Optics Express*, vol. 21, no. 12, pp. 14036–14046, 2013.

[133] J. Lian, S. Sokolov, E. Yüce, S. Combrié, A. De Rossi, and A. P. Mosk, "Measurement of the profiles of disorder-induced localized resonances in photonic crystal waveguides by local tuning," *Optics Express*, vol. 24, no. 19, pp. 21939–21947, 2016.

[134] J. Guan, A. J. Menssen, X. Liu, J. Wang, and M. J. Booth, "Component-wise testing of laser-written integrated coupled-mode beam splitters," *Optics Letters*, vol. 44, no. 12, pp. 3174–3177, 2019.

[135] L. Pattelli, R. Savo, M. Burresi, and D. S. Wiersma, "Spatio-temporal visualization of light transport in complex photonic structures," *Light: Science & Applications*, vol. 5, no. 5, pp. e16090–e16090, 2016.

[136] P. H. Otsuka, K. Nanri, O. Matsuda, M. Tomoda, D. Profunser, I. Veres, S. Danworaphong, A. Khelif, S. Benchabane, V. Laude, *et al.*, "Broadband evolution of phononic-crystal-waveguide eigenstates in real- and k-spaces," *Scientific Reports*, vol. 3, p. 3351, 2013.

[137] D. M. Profunser, E. Muramoto, O. Matsuda, O. B. Wright, and U. Lang, "Dynamic visualization of surface acoustic waves on a two-dimensional phononic crystal," *Physical Review B*, vol. 80, no. 1, p. 014301, 2009.

[138] D. M. Profunser, O. B. Wright, and O. Matsuda, "Imaging ripples on phononic crystals reveals acoustic band structure and Bloch harmonics," *Physical Review Letters*, vol. 97, no. 5, p. 055502, 2006.

[139] Y. Sugawara, O. Wright, O. Matsuda, M. Takigahira, Y. Tanaka, S. Tamura, and V. Gusev, "Watching ripples on crystals," *Physical Review Letters*, vol. 88, no. 18, p. 185504, 2002.

[140] B. Chen, R. Bruck, D. Traviss, A. Z. Khokhar, S. Reynolds, D. Thomson, G. Mashanovich, G. Reed, and O. L. Muskens, "Hybrid photon-plasmon coupling and ultrafast control of nanoantennas on a silicon photonic chip," *Nano Letters*, 2017.

[141] J. D. Joannopoulos, S. G. Johnson, J. N. Winn, and R. D. Meade, *Photonic Crystals: Molding the Flow of Light (Second Edition)*. Princeton University Press, 2 ed., 2008.

[142] A. W. Snyder and J. D. Love, *Optical waveguide theory*. Chapman and Hall, 1983.

[143] G. Lecamp, J.-P. Hugonin, and P. Lalanne, "Theoretical and computational concepts for periodic optical waveguides," *Optics Express*, vol. 15, no. 18, pp. 11042–11060, 2007.

[144] D. J. Thomson, Y. Hu, G. T. Reed, and J.-M. Fedeli, "Low loss MMI couplers for high performance MZI modulators," *IEEE Photon. Technol. Lett*, vol. 22, pp. 1485–1487, Oct 2010.

[145] G. Z. Mashanovich, F. Y. Gardes, D. J. Thomson, Y. Hu, K. Li, M. Nedeljkovic, J. S. Penades, A. Z. Khokhar, C. J. Mitchell, S. Stankovic, *et al.*, "Silicon photonic waveguides and devices for near- and mid-IR applications," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 21, no. 4, pp. 407–418, 2015.

[146] J. Wang, S. Paesani, Y. Ding, R. Santagati, P. Skrzypczyk, A. Salavrakos, J. Tura, R. Augusiak, L. Mančinska, D. Bacco, D. Bonneau, J. W. Silverstone, Q. Gong, A. Acín, K. Rottwitt, L. K. Oxenløwe, J. L. O'Brien, A. Laing, and M. G. Thompson, "Multidimensional quantum entanglement with large-scale integrated optics," *Science*, vol. 360, no. 6386, pp. 285–291, 2018.

[147] D. Pérez, I. Gasulla, and J. Capmany, "Field-programmable photonic arrays," *Optics Express*, vol. 26, no. 21, pp. 27265–27278, 2018.

[148] I. Zand and W. Bogaerts, "Effects of coupling and phase imperfections in programmable photonic hexagonal waveguide meshes," *Photonics Research*, vol. 8, no. 2, pp. 211–218, 2020.

[149] X. Chen, P. K. Stroobant, M. Pickavet, and W. Bogaerts, "Graph representations for programmable photonic circuits," *Journal of Lightwave Technology*, 2020.

[150] N. Sherwood-Droz, H. Wang, L. Chen, B. G. Lee, A. Biberman, K. Bergman, and M. Lipson, "Optical 4×4 hitless silicon router for optical networks-on-chip (NoC)," *Optics Express*, vol. 16, no. 20, pp. 15915–15922, 2008.

[151] J. E. Cunningham, I. Shubin, X. Zheng, T. Pinguet, A. Mekis, Y. Luo, H. Thacker, G. Li, J. Yao, K. Raj, *et al.*, "Highly-efficient thermally-tuned resonant optical filters," *Optics Express*, vol. 18, no. 18, pp. 19055–19063, 2010.

[152] K. Padmaraju and K. Bergman, "Resolving the thermal challenges for silicon microring resonator devices," *Nanophotonics*, vol. 3, no. 4-5, pp. 269–281, 2014.

[153] N. Jones, "How to stop data centres from gobbling up the world's electricity," *Nature*, vol. 561, no. 7722, pp. 163–167, 2018.

[154] E. Masanet, A. Shehabi, N. Lei, S. Smith, and J. Koomey, "Recalibrating global data center energy-use estimates," *Science*, vol. 367, no. 6481, pp. 984–986, 2020.

[155] K. Volz, J. Lindner, and B. Stritzker, "Ion beam induced amorphization and recrystallization of Si/SiC/Si layer systems," *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, vol. 120, no. 1-4, pp. 133–138, 1996.

[156] L. Pelaz, L. A. Marqués, and J. Barbolla, "Ion-beam-induced amorphization and recrystallization in silicon," *Journal of applied physics*, vol. 96, no. 11, pp. 5947–5976, 2004.

[157] M. Fujimaki, Y. Ohki, J. L. Brebner, and S. Roorda, "Fabrication of long-period optical fiber gratings by use of ion implantation," *Optics Letters*, vol. 25, no. 2, pp. 88–89, 2000.

[158] X. Zhao, M. Moeen, M. Toprak, G. Wang, J. Luo, X. Ke, Z. Li, D. Liu, W. Wang, C. Zhao, *et al.*, "Design impact on the performance of Ge PIN photodetectors," *Journal of Materials Science: Materials in Electronics*, vol. 31, no. 1, pp. 18–25, 2020.

[159] M. M. Milosevic, X. Chen, W. Cao, A. F. Runge, Y. Franz, C. G. Littlejohns, S. Mailis, A. C. Peacock, D. J. Thomson, and G. T. Reed, "Ion implantation in silicon for trimming the operating wavelength of ring resonators," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 24, no. 4, pp. 1–7, 2018.

[160] B. Chen, X. Yu, X. Chen, M. M. Milosevic, D. J. Thomson, A. Z. Khokhar, S. Saito, O. L. Muskens, and G. T. Reed, "Real-time monitoring and gradient feedback enable accurate trimming of ion-implanted silicon photonic devices," *Optics Express*, vol. 26, no. 19, pp. 24953–24963, 2018.

[161] M. M. Milosevic, X. Chen, X. Yu, N. J. Dinsdale, O. Aktas, S. Z. Oo, A. Z. Khokhar, D. J. Thomson, O. L. Muskens, H. M. Chong, *et al.*, "Ion implantation of germanium into silicon for critical coupling control of racetrack resonators," *Journal of Lightwave Technology*, vol. 38, no. 7, pp. 1865–1873, 2020.

[162] V. Biryukova, G. J. Sharp, C. Klitis, and M. Sorel, "Trimming of silicon-on-insulator ring-resonators via localized laser annealing," *Optics Express*, vol. 28, no. 8, pp. 11156–11164, 2020.

[163] X. Chen, M. M. Milosevic, D. J. Thomson, A. Z. Khokhar, Y. Franz, A. F. Runge, S. Mailis, A. C. Peacock, and G. T. Reed, "Post-fabrication phase trimming of Mach–Zehnder interferometers by laser annealing of germanium implanted waveguides," *Photonics Research*, vol. 5, no. 6, pp. 578–582, 2017.

[164] X. Yu, B. Chen, X. Chen, M. Milosevic, S. Saito, G. Reed, and O. Muskens, "Real-time phase trimming of Mach-Zehnder interferometers by femtosecond laser annealing of germanium implanted waveguides," in *2018 IEEE 15th International Conference on Group IV Photonics (GFP)*, pp. 1–2, IEEE, 2018.

[165] X. Chen, M. M. Milosevic, X. Yu, B. Chen, A. F. Runge, A. Z. Khokhar, S. Mailis, D. J. Thomson, A. C. Peacock, S. Saito, *et al.*, "Germanium implanted photonic devices for post-fabrication trimming and programmable circuits," in *Nanophotonics and Micro/Nano Optics IV*, vol. 10823, p. 108230U, International Society for Optics and Photonics, 2018.

[166] X. Chen, M. M. Milosevic, A. F. Runge, X. Yu, A. Z. Khokhar, S. Mailis, D. J. Thomson, A. C. Peacock, S. Saito, and G. T. Reed, "Towards an optical FPGA-programmable silicon photonic circuits," *arXiv preprint arXiv:1807.01656*, 2018.

[167] H. Richter, Z. Wang, and L. Ley, "The one phonon Raman spectrum in microcrystalline silicon," *Solid State Communications*, vol. 39, no. 5, pp. 625–629, 1981.

[168] A. Zwick and R. Carles, "Multiple-order Raman scattering in crystalline and amorphous silicon," *Physical Review B*, vol. 48, no. 9, p. 6024, 1993.

[169] Z. Xu, Z. He, Y. Song, X. Fu, M. Rommel, X. Luo, A. Hartmaier, J. Zhang, and F. Fang, "Topic review: application of Raman spectroscopy characterization in micro/nano-machining," *Micromachines*, vol. 9, no. 7, p. 361, 2018.

[170] J. Parker Jr, D. Feldman, and M. Ashkin, "Raman scattering by silicon and germanium," *Physical Review*, vol. 155, no. 3, p. 712, 1967.

[171] N. Healy, S. Mailis, N. M. Bulgakova, P. J. Sazio, T. D. Day, J. R. Sparks, H. Y. Cheng, J. V. Badding, and A. C. Peacock, "Extreme electronic bandgap modification in laser-crystallized silicon optical fibres," *Nature Materials*, vol. 13, no. 12, pp. 1122–1127, 2014.

[172] W. Bogaerts, D. Pérez, J. Capmany, D. A. Miller, J. Poon, D. Englund, F. Morichetti, and A. Melloni, "Programmable photonic circuits," *Nature*, vol. 586, no. 7828, pp. 207–216, 2020.

[173] D. Melati, Y. Grinberg, M. K. Dezfouli, S. Janz, P. Cheben, J. H. Schmid, A. Sánchez-Postigo, and D.-X. Xu, "Mapping the global design space of nanophotonic components using machine learning pattern recognition," *Nature Communications*, vol. 10, no. 1, pp. 1–9, 2019.

[174] B. J. Shastri, A. N. Tait, T. F. de Lima, W. H. Pernice, H. Bhaskaran, C. D. Wright, and P. R. Prucnal, "Photonics for artificial intelligence and neuromorphic computing," *arXiv preprint arXiv:2011.00111*, 2020.

[175] T. Zhang, Q. Liu, Y. Dan, S. Yu, X. Han, J. Dai, and K. Xu, "Machine learning and evolutionary algorithm studies of graphene metamaterials for optimized plasmon-induced transparency," *Optics Express*, vol. 28, no. 13, pp. 18899–18916, 2020.

[176] D. Pérez, A. López, P. DasMahapatra, and J. Capmany, "Multipurpose self-configuration of programmable photonic circuits," *Nature Communications*, vol. 11, no. 1, pp. 1–11, 2020.

[177] A. López, D. Pérez, P. DasMahapatra, and J. Capmany, "Auto-routing algorithm for field-programmable photonic gate arrays," *Optics Express*, vol. 28, no. 1, pp. 737–752, 2020.

[178] Z. A. Kudyshev, V. M. Shalaev, and A. Boltasseva, "Machine learning for integrated quantum photonics," *ACS Photonics*, vol. 8, no. 1, pp. 34–46, 2021.

[179] W. Bogaerts, H. Sattari, P. Edinger, A. Y. Takabayashi, I. Zand, X. Wang, A. Ribeiro, M. Jezzini, C. Errando-Herranz, G. Talli, *et al.*, "Morphic: programmable photonic circuits enabled by silicon photonic MEMS," in *Silicon Photonics XV*, vol. 11285, p. 1128503, International Society for Optics and Photonics, 2020.

[180] J. S. Jensen and O. Sigmund, "Topology optimization for nano-photonics," *Laser & Photonics Reviews*, vol. 5, no. 2, pp. 308–321, 2011.

[181] T. Feichtner, O. Selig, M. Kiunke, and B. Hecht, "Evolutionary optimization of optical antennas," *Physical Review Letters*, vol. 109, no. 12, p. 127701, 2012.

[182] P. R. Wiecha, A. Arbouet, C. Girard, A. Lecestre, G. Larrieu, and V. Paillard, "Evolutionary multi-objective optimization of colour pixels based on dielectric nanoantennas," *Nature Nanotechnology*, vol. 12, no. 2, pp. 163–169, 2017.

[183] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[184] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[185] Y. Rivenson, Y. Zhang, H. Günaydın, D. Teng, and A. Ozcan, "Phase recovery and holographic image reconstruction using deep learning in neural networks," *Light: Science & Applications*, vol. 7, no. 2, pp. 17141–17141, 2018.

[186] Y. Jo, S. Park, J. Jung, J. Yoon, H. Joo, M.-h. Kim, S.-J. Kang, M. C. Choi, S. Y. Lee, and Y. Park, "Holographic deep learning for rapid optical screening of anthrax spores," *Science advances*, vol. 3, no. 8, p. e1700606, 2017.

[187] P. R. Wiecha, A. Lecestre, N. Mallet, and G. Larrieu, "Pushing the limits of optical information storage using deep learning," *Nature nanotechnology*, vol. 14, no. 3, pp. 237–244, 2019.

[188] B. Han, Y. Lin, Y. Yang, N. Mao, W. Li, H. Wang, V. Fatemi, L. Zhou, J. I.-J. Wang, Q. Ma, Y. Cao, D. Rodan-Legrain, Y.-Q. Bie, E. Navarro-Moratalla, D. Klein, D. MacNeill, S. Wu, W. S. Leong, H. Kitadai, X. Ling, P. Jarillo-Herrero, T. Palacios, J. Yin, and J. Kong, "Deep learning enabled fast optical characterization of two-dimensional materials," *arXiv:1906.11220 [cond-mat, physics:physics]*, 2019.

[189] B. Rahmani, D. Loterie, G. Konstantinou, D. Psaltis, and C. Moser, "Multimode optical fiber transmission with a deep learning network," *Light: Science & Applications*, vol. 7, no. 1, pp. 1–11, 2018.

[190] N. Borhani, E. Kakkava, C. Moser, and D. Psaltis, "Learning to see through multimode fibers," *Optica*, vol. 5, no. 8, pp. 960–966, 2018.

[191] U. Kürüm, P. R. Wiecha, R. French, and O. L. Muskens, "Deep learning enabled real time speckle recognition and hyperspectral imaging using a multimode fiber array," *Optics express*, vol. 27, no. 15, pp. 20965–20979, 2019.

[192] Z. Liu, D. Zhu, S. P. Rodrigues, K.-T. Lee, and W. Cai, "Generative model for the inverse design of metasurfaces," *Nano letters*, vol. 18, no. 10, pp. 6570–6576, 2018.

[193] Z. Liu, L. Raju, D. Zhu, and W. Cai, "A hybrid strategy for the discovery and design of photonic structures," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 10, no. 1, pp. 126–135, 2020.

[194] C. C. Nadell, B. Huang, J. M. Malof, and W. J. Padilla, "Deep learning for accelerated all-dielectric metasurface design," *Optics Express*, vol. 27, no. 20, pp. 27523–27535, 2019.

[195] M. H. Tahersima, K. Kojima, T. Koike-Akino, D. Jha, B. Wang, C. Lin, and K. Parsons, "Deep neural network inverse design of integrated photonic power splitters," *Scientific Reports*, vol. 9, no. 1, pp. 1–9, 2019.

[196] Y. Tang, K. Kojima, T. Koike-Akino, Y. Wang, P. Wu, M. Tahersima, D. Jha, K. Parsons, and M. Qi, "Generative deep learning model for a multi-level nano-optic broadband power splitter," in *2020 Optical Fiber Communications Conference and Exhibition (OFC)*, pp. 1–3, IEEE, 2020.

[197] J. Lenaerts, H. Pinson, and V. Ginis, "Artificial neural networks for inverse design of resonant nanophotonic components with oscillatory loss landscapes," *Nanophotonics*, 2020.

[198] N. J. Dinsdale, P. R. Wiecha, M. Delaney, J. Reynolds, M. Ebert, I. Zeimpekis, D. J. Thomson, G. T. Reed, P. Lalanne, K. Vynck, *et al.*, "Deep learning enabled design of complex transmission matrices for universal optical components," *ACS Photonics*, vol. 8, no. 1, pp. 283–295, 2021.

[199] J. Jiang, R. Lupoiu, E. W. Wang, D. Sell, J. P. Hugonin, P. Lalanne, and J. A. Fan, "Metanet: a new paradigm for data sharing in photonics research," *Optics Express*, vol. 28, no. 9, pp. 13670–13681, 2020.

[200] J. Jiang and J. A. Fan, "Global optimization of dielectric metasurfaces using a physics-driven neural network," *Nano Letters*, vol. 19, no. 8, pp. 5366–5372, 2019.

[201] D. Sell, J. Yang, S. Doshay, R. Yang, and J. A. Fan, "Large-angle, multifunctional metagratings based on freeform multimode geometries," *Nano Letters*, vol. 17, no. 6, pp. 3752–3757, 2017.

[202] T. W. Hughes, M. Minkov, I. A. Williamson, and S. Fan, "Adjoint method and inverse design for nonlinear nanophotonic devices," *ACS Photonics*, vol. 5, no. 12, pp. 4781–4787, 2018.

[203] T. W. Hughes, I. A. Williamson, M. Minkov, and S. Fan, "Wave physics as an analog recurrent neural network," *Science advances*, vol. 5, no. 12, p. eaay6946, 2019.

[204] C. M. Lalau-Keraly, S. Bhargava, O. D. Miller, and E. Yablonovitch, "Adjoint shape optimization applied to electromagnetic design," *Optics Express*, vol. 21, no. 18, pp. 21693–21701, 2013.

[205] L. Su, D. Vercruysse, J. Skarda, N. V. Sapra, J. A. Petykiewicz, and J. Vučković, "Nanophotonic inverse design with spins: Software architecture and practical considerations," *Applied Physics Reviews*, vol. 7, no. 1, p. 011407, 2020.

[206] A. P. Mosk, A. Lagendijk, G. Lerosey, and M. Fink, "Controlling waves in space and time for imaging and focusing in complex media," *Nature Photonics*, vol. 6, no. 5, pp. 283–292, 2012.

[207] T. Strudley, R. Bruck, B. Mills, and O. L. Muskens, "An ultrafast reconfigurable nanophotonic switch using wavefront shaping of light in a nonlinear nanomaterial," *Light: Science & Applications*, vol. 3, no. 9, pp. e207–e207, 2014.

[208] R. Sarma, A. G. Yamilov, S. Petrenko, Y. Bromberg, and H. Cao, "Control of energy density inside a disordered medium by coupling to open or closed channels," *Physical Review Letters*, vol. 117, no. 8, p. 086803, 2016.

[209] R. Sarma, A. Yamilov, and H. Cao, "Enhancing light transmission through a disordered waveguide with inhomogeneous scattering and loss," *Applied Physics Letters*, vol. 110, no. 2, p. 021103, 2017.

[210] R. French, S. Gigan, and O. L. Muskens, "Speckle-based hyperspectral imaging combining multiple scattering and compressive sensing in nanowire mats," *Optics Letters*, vol. 42, no. 9, pp. 1820–1823, 2017.

[211] A. Y. Piggott, J. Lu, K. G. Lagoudakis, J. Petykiewicz, T. M. Babinec, and J. Vučković, "Inverse design and demonstration of a compact and broadband on-chip wavelength demultiplexer," *Nature Photonics*, vol. 9, no. 6, pp. 374–377, 2015.

[212] D. Ohana, B. Desiatov, N. Mazurski, and U. Levy, "Dielectric metasurface as a platform for spatial mode conversion in nanoscale waveguides," *Nano Letters*, vol. 16, no. 12, pp. 7956–7961, 2016.

[213] C. Doerr, L. Chen, D. Vermeulen, T. Nielsen, S. Azemati, S. Stulz, G. McBrien, X.-M. Xu, B. Mikkelsen, M. Givehchi, *et al.*, "Single-chip silicon photonics 100-Gb/s coherent transceiver," in *Optical Fiber Communication Conference*, pp. Th5C–1, Optical Society of America, 2014.

[214] C. Doerr, J. Heanue, L. Chen, R. Aroca, S. Azemati, G. Ali, G. McBrien, L. Chen, B. Guan, H. Zhang, *et al.*, "Silicon photonics coherent transceiver in a ball-grid array package," in *2017 Optical Fiber Communications Conference and Exhibition (OFC)*, pp. 1–3, IEEE, 2017.

[215] L.-W. Luo, N. Ophir, C. P. Chen, L. H. Gabrielli, C. B. Poitras, K. Bergmen, and M. Lipson, "WDM-compatible mode-division multiplexing on a silicon chip," *Nature Communications*, vol. 5, no. 1, pp. 1–7, 2014.

[216] D. Zhou, C. Sun, Y. Lai, Y. Yu, and X. Zhang, "Integrated silicon multifunctional mode-division multiplexing system," *Optics Express*, vol. 27, no. 8, pp. 10798–10805, 2019.

[217] M. Teng, A. Honardoost, Y. Alahmadi, S. S. Polkoo, K. Kojima, H. Wen, C. K. Renshaw, P. LiKamWa, G. Li, S. Fathpour, *et al.*, "Miniaturized silicon photonics devices for integrated optical signal processors," *Journal of Lightwave Technology*, vol. 38, no. 1, pp. 6–17, 2020.

[218] J. Wang, P. Chen, S. Chen, Y. Shi, and D. Dai, "Improved 8-channel silicon mode demultiplexer with grating polarizers," *Optics Express*, vol. 22, no. 11, pp. 12799–12807, 2014.

[219] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[220] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[221] L. F. Frellsen, Y. Ding, O. Sigmund, and L. H. Frandsen, "Topology optimized mode multiplexing in silicon-on-insulator photonic wire waveguides," *Optics Express*, vol. 24, pp. 16866–16873, July 2016.

[222] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[223] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 31, 2017.

[224] S. L. Smith, P.-J. Kindermans, C. Ying, and Q. V. Le, "Don't decay the learning rate, increase the batch size," *arXiv preprint arXiv:1711.00489*, 2017.

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

[226] D. Liu, Y. Tan, E. Khoram, and Z. Yu, "Training deep neural networks for the inverse design of nanophotonic structures," *ACS Photonics*, vol. 5, pp. 1365–1369, Apr. 2018.

[227] L. Xu, M. Rahmani, Y. Ma, D. A. Smirnova, K. Z. Kamali, F. Deng, Y. K. Chiang, L. Huang, H. Zhang, S. Gould, *et al.*, "Enhanced light–matter interactions in dielectric nanostructures via machine-learning approach," *Advanced Photonics*, vol. 2, no. 2, p. 026003, 2020.

[228] W. Ma, F. Cheng, Y. Xu, Q. Wen, and Y. Liu, "Probabilistic representation and inverse design of metamaterials based on a deep generative model with semi-supervised learning strategy," *Advanced Materials*, vol. 31, no. 35, p. 1901111, 2019.

[229] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.

[230] D. P. Kingma and M. Welling, "An introduction to variational autoencoders," *arXiv preprint arXiv:1906.02691*, 2019.

[231] C. P. Burgess, I. Higgins, A. Pal, L. Matthey, N. Watters, G. Desjardins, and A. Lerchner, "Understanding disentangling in β-VAE," *arXiv:1804.03599 [cs, stat]*, Apr. 2018.

[232] I. Higgins, L. Matthey, A. Pal, C. Burgess, X. Glorot, M. Botvinick, S. Mohamed, and A. Lerchner, "β-VAE: Learning basic visual concepts with a constrained variational framework," *ICLR conference*, 2017.

[233] W. Liu, R. Li, M. Zheng, S. Karanam, Z. Wu, B. Bhanu, R. J. Radke, and O. Camps, "Towards visually explaining variational autoencoders," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8642–8651, 2020.

[234] C. D. Wright, Y. Liu, K. I. Kohary, M. M. Aziz, and R. J. Hicken, "Arithmetic and biologically-inspired computing using phase-change materials," *Advanced Materials*, vol. 23, no. 30, pp. 3408–3413, 2011.

[235] D. R. Solli and B. Jalali, "Analog optical computing," *Nature Photonics*, vol. 9, no. 11, pp. 704–706, 2015.

[236] Y. Jiang, P. T. DeVore, and B. Jalali, "Analog optical computing primitives in silicon photonics," *Optics Letters*, vol. 41, no. 6, pp. 1273–1276, 2016.

[237] J. Feldmann, M. Stegmaier, N. Gruhler, C. Ríos, H. Bhaskaran, C. Wright, and W. Pernice, "Calculating with light using a chip-scale all-optical abacus," *Nature Communications*, vol. 8, no. 1, pp. 1–8, 2017.

[238] H. Babashah, Z. Kavehvash, A. Khavasi, and S. Koohi, "Temporal analog optical computing using an on-chip fully reconfigurable photonic signal processor," *Optics & Laser Technology*, vol. 111, pp. 66–74, 2019.

[239] L. Lu, L. Zhou, and J. Chen, "Programmable scow mesh silicon photonic processor for linear unitary operator," *Micromachines*, vol. 10, no. 10, p. 646, 2019.

[240] D. A. B. Miller, "Perfect optics with imperfect components," *Optica*, vol. 2, no. 8, pp. 747–750, 2015.

[241] A. Crespi, R. Osellame, R. Ramponi, D. J. Brod, E. F. Galvao, N. Spagnolo, C. Vitelli, E. Maiorino, P. Mataloni, and F. Sciarrino, "Integrated multimode interferometers with arbitrary designs for photonic boson sampling," *Nature Photonics*, vol. 7, no. 7, pp. 545–549, 2013.

[242] J. Wang, F. Sciarrino, A. Laing, and M. G. Thompson, "Integrated photonic quantum technologies," *Nature Photonics*, vol. 14, no. 5, pp. 273–284, 2020.

[243] A. Aspuru-Guzik and P. Walther, "Photonic quantum simulators," *Nature Physics*, vol. 8, no. 4, pp. 285–291, 2012.

[244] J. L. O'brien, A. Furusawa, and J. Vučković, "Photonic quantum technologies," *Nature Photonics*, vol. 3, no. 12, pp. 687–695, 2009.

[245] N. M. Estakhri, B. Edwards, and N. Engheta, "Inverse-designed metastructures that solve equations," *Science*, vol. 363, no. 6433, pp. 1333–1338, 2019.

[246] S. Molesky, Z. Lin, A. Y. Piggott, W. Jin, J. Vucković, and A. W. Rodriguez, "Inverse design in nanophotonics," *Nature Photonics*, vol. 12, no. 11, pp. 659–670, 2018.

[247] K. Xu, L. Liu, X. Wen, W. Sun, N. Zhang, N. Yi, S. Sun, S. Xiao, and Q. Song, "Integrated photonic power divider with arbitrary power ratios," *Optics Letters*, vol. 42, no. 4, pp. 855–858, 2017.

[248] J. Lu and J. Vučković, "Nanophotonic computational design," *Optics Express*, vol. 21, no. 11, pp. 13351–13367, 2013.

[249] H. Jia, T. Zhou, X. Fu, J. Ding, and L. Yang, "Inverse-design and demonstration of ultracompact silicon meta-structure mode exchange device," *ACS Photonics*, vol. 5, no. 5, pp. 1833–1838, 2018.

[250] M. van Niekerk, J. A. Steidle, G. A. Howland, M. L. Fanto, N. Soures, F. T. Zohora, D. Kudithipudi, and S. F. Preble, "Approximating large scale arbitrary unitaries with integrated multimode interferometers," in *Quantum Information Science, Sensing, and Computation XI*, vol. 10984, p. 109840J, International Society for Optics and Photonics, 2019.

[251] C. R. de Galarreta, A. M. Alexeev, Y.-Y. Au, M. Lopez-Garcia, M. Klemm, M. Cryan, J. Bertolotti, and C. D. Wright, "Nonvolatile reconfigurable phase-change metadevices for beam steering in the near infrared," *Advanced Functional Materials*, vol. 28, no. 10, p. 1704993, 2018.

[252] M. Delaney, I. Zeimpekis, D. Lawson, D. W. Hewak, and O. L. Muskens, "A new family of ultralow loss reversible phase-change materials for photonic integrated circuits: Sb2S3 and Sb2Se3," *Advanced Functional Materials*, vol. 30, no. 36, p. 2002447, 2020.

[253] M. Delaney, I. Zeimpekis, H. Du, X. Yan, M. Banakar, D. J. Thomson, D. W. Hewak, and O. L. Muskens, "Non-volatile programmable silicon photonics using an ultralow loss Sb2Se3 phase change material," *arXiv preprint arXiv:2101.03623*.

[254] A. Y. Piggott, J. Petykiewicz, L. Su, and J. Vučković, "Fabrication-constrained nanophotonic inverse design," *Scientific Reports*, vol. 7, no. 1, pp. 1–7, 2017.

[255] K. Shportko, S. Kremers, M. Woda, D. Lencer, J. Robertson, and M. Wuttig, "Resonant bonding in crystalline phase-change materials," *Nature Materials*, vol. 7, no. 8, pp. 653–658, 2008.

[256] C. Ríos, M. Stegmaier, P. Hosseini, D. Wang, T. Scherer, C. D. Wright, H. Bhaskaran, and W. H. Pernice, "Integrated all-photonic non-volatile multi-level memory," *Nature Photonics*, vol. 9, no. 11, pp. 725–732, 2015.

[257] M. Wuttig, H. Bhaskaran, and T. Taubner, "Phase-change materials for non-volatile photonic applications," *Nature Photonics*, vol. 11, no. 8, pp. 465–476, 2017.

[258] K. J. Miller, R. F. Haglund, and S. M. Weiss, "Optical phase change materials in integrated silicon photonic devices," *Optical Materials Express*, vol. 8, no. 8, pp. 2415–2429, 2018.

[259] J. Faneca, L. Trimby, I. Zeimpekis, M. Delaney, D. W. Hewak, F. Y. Gardes, C. D. Wright, and A. Baldycheva, "On-chip sub-wavelength bragg grating design based on novel low loss phase-change materials," *Optics Express*, vol. 28, no. 11, pp. 16394–16406, 2020.

[260] M. Wuttig and N. Yamada, "Phase-change materials for rewriteable data storage," *Nature Materials*, vol. 6, no. 11, pp. 824–832, 2007.