Electrically reconfigurable integrated photonic switches for emerging computing architectures

Håvard Hem Toftevaag

Linacre College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy

Michaelmas 2025

Abstract

This thesis explores the design, simulation, and experimental demonstration of integrated photonic switch technologies for emerging computing architectures. Emphasis is placed on the interplay between the switch's functionality and the application at hand; that is, ensuring that we benefit from the advantages of the specific switch technology without being impacted by its disadvantages. One example of a technology in this work is phase-change-material-based switches, which are non-volatile (i.e., have no static power consumption) but exhibit limited cyclability. For a switch based on phase-change material, infrequent switching is therefore beneficial.

One relevant application is the use of phase-change materials in photonic tensor cores for more efficient computation. These devices can improve the latency and energy consumption of machine learning accelerators by orders of magnitude, enabling powerful edge devices and the implementation of larger models even in applications where size, weight, and power are crucial. Here, I demonstrate that an analogue computer based on photonic tensor cores with phase-change material and analogue electronic peripherals can solve a model predictive control problem with an update time on the order of tens of nanoseconds, which is significantly faster than the sub-megahertz update frequency of the all-electronic alternative. Moreover, the update frequency is entirely limited by the electronic circuitry, not by the relatively slow phase-change material, as the latter is static during operation.

Next, I present the first prototype of a non-volatile field-programmable coupler array—a reconfigurable photonic platform—with the low-loss phase-change material Sb_2Se_3 as the active element. The use of low-loss phase-change material enables more compact, lower-power systems that competing technologies cannot realise.

Finally, I showcase the first photonic device with a non-linear transfer function that allows for inherent non-uniform weight quantisation in neural networks. The device, a ring-assisted Mach–Zehnder interferometer, is designed to produce a tailored output that follows the weight distribution in the layers of the neural network. This behaviour enables more power- and memory-efficient models with quantised weights while maintaining model accuracy.

Collectively, the devices and architectures developed in this work establish a foundation for advanced photonic computing and signal processing, integrating the strengths of photonics and computational engineering. The demonstrated components highlight pathways to low-power, high-performance photonic hardware for next-generation computing and signal processing.

Electrically reconfigurable integrated photonic switches for emerging computing architectures



Håvard Hem Toftevaag

Linacre College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas 2025

This thesis is dedicated to my parents,
who always believed,
but never pushed

Acknowledgements

I want to begin by expressing my deepest gratitude to my professor, Harish, for his invaluable guidance, encouragement, and support. Although unorthodox in style, your extraordinary supervision enables world-class researchers to reach their full potential. I will carry the many lessons I have learned from you well into my future professional life. I am equally grateful to Nick, who stepped in as my *de facto* supervisor during Harish's sabbatical. Your steady guidance and constructive advice ensured that my work continued to progress (some would say exponentially), and I benefited enormously from your perspective and expertise. I am dreading having to work in photonics without your wisdom so easily accessible. Also, a big thank you to my co-supervisors at DSTL, Jordi and Matt, for providing me with interesting insights into your world and the opportunity to work closely with researchers in a different but equally exciting field. Hopefully, I have taught you enough about photonics over the past four years that you will enjoy reading even the more technical parts of this thesis.

I am indebted to my senior colleagues, in particular Ángel, Bowei, Mengyun, Mingde, Jason, Samarth, Utku, and Yuhan, whose knowledge, generosity, and willingness to support—whether scientifically or personally—have made an immeasurable difference along the way. I am also incredibly grateful to my contemporaries, Yi and Kairan, for making this journey so enjoyable. Sharing the ups and downs with you has been completely frictionless, and I am truly grateful for the camaraderie and understanding that we share. To my junior colleagues, thank you for the energy and joy you bring to the workplace—it has been a privilege to watch you grow and share in your enthusiasm. Especially, I would like to thank Maxine, who has been part of this experience the longest and whose tea breaks I have especially appreciated, and Jock, who not only introduced me to new ways of thinking about photonics but also offered input on everything from job applications to life itself.

To my collaborators—Ángel, Bowei, Jock, Mohamadreza, and Nick—thank you for holding me accountable and pushing me beyond *good enough* towards something far better. Thanks to our combined efforts, the work presented in this thesis is far richer and more compelling than anything I could have produced on my own. I already look forward to future collaborations and cannot wait to see what you are working on next.

I am truly indebted to OUARFC, the team that never lets you down, for welcoming a Norwegian into their ranks and creating an environment where I found so many close friends. My time in Oxford would not have been the same without you. Although the nickname *Fanning* never stuck, the memories did.

Thanks also to the Oxford college system, and in particular to Linacre College, for treating PhD students as exactly what we are—students. Through the Linacre community, I have had countless experiences on the river, the football field, and in the college bar; I have met people from a remarkable range of backgrounds and disciplines, and I have gained a few close friends along the way, too. I could ask for little more.

I would also like to extend my appreciation to the Norwegian community in Oxford and to Aker Scholarship. Ironically, I would not be on this journey were it not for you.

I have long believed that one of my greatest strengths is the people I have in my life. To all my friends: thank you for giving life far more meaning than research alone ever could. An extra thanks goes to my fellow Linacrite, flat- and teammate, committee member, gym partner, amateur psychologist, and good friend, Guy, for being part of this journey. The dog days are finally over. I also owe my close confidant, Martin, a special thanks for being an especially eager ear (a rather large one, in fact), even if he acted as the incident photon to my stimulated emission more often than not. Some friends not only provided invaluable support, but even helped me understand my own work—I therefore owe my old friend William a special thanks for succeeding where others failed. I hope the rest of you are aware of the importance you have had and still have on my life.

Alla dessa vänner som kom och gick,

inte visste jag att det var livet

Finally, I owe the deepest gratitude to my family, to whom this thesis is dedicated.

- Til pappa for å vise oss at dedikasjon og litt nysgjerrighet kan ta deg langt.
- Til mamma for å alltid gjøre alt i din makt for å muliggjøre dette, uten å noen gang ha ment at jeg skulle nå så langt.
- Og til slutt til Jørgen for å gjøre det harde arbeidet med å oppdra meg.

Håvard Hem Toftevaag
Oxford, 5th June 2026

Abstract

This thesis explores the design, simulation, and experimental demonstration of integrated photonic switch technologies for emerging computing architectures. Emphasis is placed on the interplay between the switch's functionality and the application at hand; that is, ensuring that we benefit from the advantages of the specific switch technology without being impacted by its disadvantages. One example of a technology in this work is phase-change-material-based switches, which are non-volatile (i.e., have no static power consumption) but exhibit limited cyclability. For a switch based on phase-change material, infrequent switching is therefore beneficial.

One relevant application is the use of phase-change materials in photonic tensor cores for more efficient computation. These devices can improve the latency and energy consumption of machine learning accelerators by orders of magnitude, enabling powerful edge devices and the implementation of larger models even in applications where size, weight, and power are crucial. Here, I demonstrate that an analogue computer based on photonic tensor cores with phase-change material and analogue electronic peripherals can solve a model predictive control problem with an update time on the order of tens of nanoseconds, which is significantly faster than the sub-megahertz update frequency of the all-electronic alternative. Moreover, the update frequency is entirely limited by the electronic circuitry, not by the relatively slow phase-change material, as the latter is static during operation.

Next, I present the first prototype of a non-volatile field-programmable coupler array—a reconfigurable photonic platform—with the low-loss phase-change material Sb_2Se_3 as the active element. The use of low-loss phase-change material enables more compact, lower-power systems that competing technologies cannot realise.

Finally, I showcase the first photonic device with a non-linear transfer function that allows for inherent non-uniform weight quantisation in neural networks. The device, a ring-assisted Mach–Zehnder interferometer, is designed to produce a tailored output that follows the weight distribution in the layers of the neural network. This behaviour enables more power- and memory-efficient models with quantised weights while maintaining model accuracy.

Collectively, the devices and architectures developed in this work establish a foundation for advanced photonic computing and signal processing, integrating the strengths of photonics and computational engineering. The demonstrated components highlight pathways to low-power, high-performance photonic hardware for next-generation computing and signal processing.

Defence-relevant abstract

This thesis was made possible through the support of and collaboration with researchers at the Defence Science and Technology Laboratory (Dstl). Because it is the potential for real-world impact, not necessarily the most exciting science, that is most interesting to Dstl, I will try to balance my focus between lab-technical and domain-specific intricacies on the one hand and the pathways to practical implementation on the other.

The three chapters referenced in the main abstract are all relevant to both the defence and civilian sectors, with some of the technology being more immediately applicable to the former. First, the photonic tensor core for model predictive control can enable high-speed control problems, such as target interception and plasma control in aircraft actuators, that are unattainable with electronics-only approaches. Fundamentally, the technology is beneficial for any high-speed application that can be reduced to a quadratic programming problem that the photonic tensor core can solve. Secondly, the field-programmable coupler array is more of an enabling technology than the solution to a specific defence-relevant problem. The immediate benefit of this technology will most likely be seen in the acceleration of prototyping and fabrication cycles for photonic integrated circuits. Nevertheless, the platform enables reconfigurability—and thus redundancy—for safety-critical systems, and can function as a reconfigurable optical signal processor for radar. Both technologies mentioned above benefit from the combination of photonic integrated circuits, which promise immunity to electromagnetic interference, and phase-change materials, which function as radiation-hard memory cells. Finally, the non-uniformly quantised photonic neural network weight is helpful for any model where reduced memory, low power, and high accuracy are critical.

For researchers at Dstl who may be reading this work, it is worth noting that, as a research group, we are limited in our scaling capabilities, and our photonic chips require a complex laboratory setup that is neither mobile nor field-deployable. As a result, we are currently unable to conduct large-scale, defence-relevant experiments outside the lab. Nevertheless, I believe that the concepts covered in this thesis are not far from widespread adoption in the coming years.

This DPhil was partly funded by the Defence Science and Technology Laboratory (Dstl) through the Future Sensing and Situational Awareness programme under an Industrial Cooperative Award in Science and Technology (iCASE), agreement number DSTLX1000160238R.

Contents

Acknowledgements **v**

Abstract **vii**

List of Figures **xv**

List of Tables **xvii**

List of Terms **xix**

1 Introduction **1**
1.1 Motivation . . . 3
1.2 Objective . . . 5
1.3 Thesis outline . . . 5
Chapter 2: Literature review . . . 6
Chapter 3: Theory . . . 6
Chapter 4: Device simulation, fabrication, and measurements . . . 6
Chapter 5: Phase-change photonic computing for defence . . . 7
Chapter 6: Non-volatile field-programmable coupler array . . . 7
Chapter 7: Non-uniformly quantised neural network weights . . . 8
Chapter 8: Concluding remarks . . . 8
1.4 Key findings and new science . . . 8
1.5 Statement of originality . . . 9

2 Photonic switch networks for computing technologies: A review **11**
2.1 Photonic integrated switch structures . . . 13
2.1.1 Optical modulators . . . 13
2.1.2 Mach–Zehnder switches . . . 13
2.1.3 Resonator switches . . . 14
2.1.4 MEMS switches . . . 15
2.1.5 Other switch designs . . . 16
2.2 Switching mechanisms . . . 19
2.2.1 Thermo-optic effect . . . 21

2.2.2 Electro-optic effects . . . 22
2.2.3 MEMS actuation . . . 24
2.2.4 Optical phase-change materials . . . 25
2.2.5 Other switching mechanisms . . . 26
2.3 Photonic switches with optical phase-change materials . . . 29
2.3.1 Fundamental properties of phase-change materials . . . 29
2.3.2 Low-loss phase-change materials . . . 31
2.3.3 Integration of phase-change switches . . . 35
2.3.4 Phase-change photonics in extreme environments . . . 35
2.3.5 Application areas of phase-change switches . . . 36
2.4 Photonic computing and signal processing . . . 37
2.4.1 Integrated optical signal processors . . . 38
2.4.2 Photonic mesh networks . . . 40
2.4.3 Photonic recirculating mesh networks . . . 42
2.4.4 Photonic neural networks . . . 46
2.5 Conclusion . . . 52

3 Theory **53**
3.1 Light propagation . . . 54
3.1.1 Effective refractive index . . . 55
3.2 Performance metrics of photonic switches . . . 57
3.3 Directional coupler theory . . . 59
3.4 Ring resonator theory . . . 60
3.4.1 Dynamic behaviour of ring resonators . . . 62
3.5 Mach–Zehnder interferometer theory . . . 63
3.5.1 Governing equations of Mach–Zehnder interferometers . . . 64
3.5.2 Ring-assisted Mach–Zehnder interferometer theory . . . 67
3.6 Phase-change material theory . . . 69
3.7 Neural network theory . . . 71
3.7.1 Weight quantisation in neural networks . . . 72
3.7.2 Working principle of photonic crossbar array . . . 74
3.8 Sensor management and control theory . . . 74
3.8.1 Kalman filter implementation . . . 76
3.8.2 Quadratic programming for model predictive control . . . 76
3.8.3 Primal–dual gradient (Langrangian) method . . . 79

4 Device simulation, fabrication, and measurement 81
4.1 General device fabrication steps . . . 82
4.1.1 Basics . . . 82
4.1.2 Alignment . . . 83
4.1.3 Photonic integrated circuit elements . . . 83
4.2 Fabrication steps for FPCAs . . . 84
4.2.1 Ion implantation . . . 84
4.2.2 Electrodes and ohmic contacting . . . 86
4.2.3 Phase-change material deposition . . . 87
4.3 Fabrication steps for neural network weights . . . 88
4.3.1 Separation layer . . . 89
4.3.2 Heaters and electrodes . . . 90
4.4 Phase-change materials: Experimental basics . . . 92
4.4.1 Phase-change material measurements . . . 92
4.4.2 Phase-change material simulation . . . 96
4.4.3 Phase-change material considerations . . . 100
4.5 Experimental measurement setup . . . 101
4.5.1 Optical measurement setup . . . 102
4.5.2 Electrical measurement setup . . . 104
4.5.3 Optical loss measurements . . . 106
4.5.4 Measurements of directional couplers . . . 107

5 Phase-change photonic computing for defence-relevant tasks 109
5.1 Background . . . 110
5.2 Photonic sensor fusion and its issues . . . 111
5.2.1 Analogue solutions to Kalman filtering . . . 113
5.2.2 A final comment on sensor fusion . . . 113
5.3 Photonic model predictive control . . . 114
5.3.1 Practical example: target tracking . . . 115
5.3.2 Photonic implementation . . . 116
5.4 Practical aspects and performance metrics . . . 120
5.4.1 Performance of the electronic counterparts . . . 121
5.4.2 Speed calculations for photonic hardware . . . 122
5.4.3 Power calculations for photonic hardware . . . 125
5.4.4 Other practical aspects . . . 126
5.5 Future work . . . 127
5.6 Chapter summary . . . 128

6 Non-volatile, small-footprint field-programmable coupler array 129
6.1 Introduction . . . 130
6.1.1 Conceptual overview . . . 131
6.2 Phase shifter with Sb_2Se_3 . . . 131
6.2.1 Design considerations . . . 132
6.2.2 Integrated microheaters . . . 133
6.2.3 Switching energy and multi-level tunability . . . 137
6.2.4 Endurance . . . 139
6.3 Building block: MZI tunable coupler . . . 140
6.3.1 Optical measurements of MZI tunable coupler . . . 142
6.4 Non-volatile FPCA: Optical measurements . . . 143
6.4.1 nv-FPCA as a router . . . 143
6.4.2 nv-FPCA as a ring resonator . . . 144
6.4.3 nv-FPCA as a delay line . . . 146
6.4.4 Loss estimates . . . 148
6.5 Synthesised temporal differentiator . . . 150
6.6 Discussion . . . 150
6.6.1 Optimising the FPCA loss . . . 152
6.6.2 Cyclability of Sb_2Se_3 . . . 153
6.6.3 Technology analysis . . . 155
6.7 Future work . . . 158
6.7.1 Foundry design . . . 158
6.7.2 Future design . . . 159
6.8 Chapter summary . . . 162

7 Non-uniformly quantised analogue neural network weight 165
7.1 Introduction . . . 166
7.1.1 Uniform and non-uniform weight quantisation . . . 167
7.2 Device design . . . 167
7.2.1 Determining the device design through simulations . . . 168
7.2.2 Control of the device . . . 175
7.3 Measurements of ring-assisted MZI weights . . . 178
7.3.1 General behaviour . . . 178
7.3.2 Wavelength dependence . . . 180
7.3.3 Varying the shape of the curve . . . 182
7.3.4 Neural network weight quantisation and accuracy . . . 185
7.3.5 Discussion . . . 190
7.4 Future work . . . 191
7.4.1 Electrical tuning . . . 191

7.4.2 System integration 192
7.4.3 Alternative photonic structures 193
7.4.4 Energy calculations 194
7.5 Chapter summary 196

8 Conclusion and outlook **199**
8.1 Future research and broader impact 202

A Appendices **207**
A.1 Model predictive control 207
A.1.1 Python code 207
A.2 Field-programmable coupler array 210
A.2.1 Overview of switching parameters 210
A.2.2 List of synthesised circuits with four unit cells 210
A.3 Non-uniform neural network weight 212
A.3.1 More INTERCONNECT simulations of RAMZIs 212
A.3.2 More RAMZI transmission spectra 213
A.3.3 Effect of reflection in the light path 213

References **217**

List of Figures

2.1 Mach–Zehnder interferometer structures 14
2.2 Microring resonator structures 15
2.3 MEMS switches 17
2.4 Directional coupler switches 18
2.5 Multi-mode interferometer switches 19
2.6 Photonic crystal switches 20
2.7 Refractive index comparison of phase-change materials 29
2.8 Initial demonstrations of field-programmable coupler arrays 43
2.9 Recent demonstrations of field-programmable coupler arrays 44
2.10 Demonstrations of field-programmable ring/disk arrays 46
2.11 Demonstrations of photonic crossbar arrays 48

3.1 Transmission response and phase delay of ring resonators 61
3.2 Analytical transmission of a ring-assisted MZI 69
3.3 Crystal structures of phase-change materials 71
3.4 Neural network basics and weight distribution 72

4.1 Fabrication procedure for field-programmable coupler arrays 85
4.2 Optical loss measurements of boron-implanted silicon 87
4.3 Fabrication procedure for neural network weights 89
4.4 Different heater designs for neural network weights 91
4.5 Ellipsometry measurements of Sb_2Se_3 93
4.6 Effective refractive index measurements of Sb_2Se_3 95
4.7 Optical loss measurements of Sb_2Se_3 97
4.8 MODE simulations of silicon-on-insulator rib waveguide with Sb_2Se_3 98
4.9 Experimental setups 102
4.10 Experimental validation of 50:50 splitting in directional couplers . . 108

5.1 Model predictive control concept 116
5.2 Model predictive control: simple tracking example 116
5.3 Interconnected photonic crossbar arrays for model predictive control 118
5.4 Model predictive control: advanced tracking example 119
5.5 Design of photonic crossbar arrays for model predictive control . . . 120

6.1 Conceptual overview of field-programmable coupler arrays 132
6.2 Sb_2Se_3 phase shifter with microheater and its structural parameters 134
6.3 Simulations of Sb_2Se_3 phase shifter with integrated microheater . . 136
6.4 Current–voltage sweeps of microheaters 137
6.5 Switching of Sb_2Se_3 phase shifters with integrated microheaters . . 138
6.6 Cyclability of Sb_2Se_3 phase shifters with integrated microheaters . . 140
6.7 Images of field-programmable coupler array components 141
6.8 MZI coupler behaviour 142
6.9 The nv-FPCA used for routing of light 144
6.10 The nv-FPCA synthesised as a ring resonator and delay line 145
6.11 Lorentzian fits to synthesised ring resonator spectra 146
6.12 Synthesised delay line simulations 147
6.13 Simulation of synthesised temporal differentiator 151
6.14 FPCA technology comparison: power and active length 156
6.15 Foundry design of scaled non-volatile FPCA 159
6.16 Future design of tunable coupler for FPCA 160

7.1 Optical simulations of various RAMZI devices 169
7.2 Impact of phase shifts on RAMZI behaviour 171
7.3 Impact of ring coupling on RAMZI behaviour 172
7.4 Unbalanced RAMZI behaviour 173
7.5 RAMZI with imperfect directional couplers 174
7.6 RAMZI weight transmission and curve fit 179
7.7 Wavelength dependence on RAMZI curve shape 181
7.8 Coupling-length dependence on RAMZI curve shape 182
7.9 Effect of coupling heater (design 1) on RAMZI transmission 184
7.10 Effect of coupling heater (design 3) on RAMZI transmission 185
7.11 Non-uniform weight quantisation in photonic devices 187
7.12 Multiplication accuracy of neural network weights 189
7.13 Photonic crossbar array with RAMZI weights 192
7.14 All-pass ring resonators showing Fano resonances 194
7.15 Add-drop ring resonators showing Fano resonances 195

A.1 Examples of synthesised circuits in a larger FPCA 212
A.2 More INTERCONNECT simulations of RAMZIs 213
A.3 Reflection affecting the spectrum of a RAMZI 214
A.4 Reflection affecting the spectrum of a RAMZI 215
A.5 Reflection affecting the non-linear curves of a RAMZI 216

List of Tables

2.1 Comparison of hybrid waveguides with low-loss PCM in literature . 34

4.1 Overview of the refractive indices of Sb_2Se_3 at different wavelengths 93
4.2 Variations in simulation results of Sb_2Se_3 99
4.3 Comparison of the effective refractive index and loss of Sb_2Se_3 . . . 100

6.1 Overview of material parameters used for microheater simulations . 135
6.2 Performance comparison of demonstrated FPCA technologies 164

A.1 Switching parameters for amorphisation (lowest total energy) 211
A.2 Switching parameters for amorphisation (multilevel) 211
A.3 Switching parameters for crystallisation 212

List of Terms, Acronyms, and Abbreviations

1×1 one input and one output

1×2 one input and two outputs

2×2 two inputs and two outputs

$m \times n$ m inputs and n outputs

1-D one-dimensional

2-D two-dimensional

3-D three-dimensional

ADC analogue-to-digital converter

AFM atomic force microscopy

AI artificial intelligence

AIMC analogue in-memory computing

AIST (Ag,In)-doped Sb_2Te

Al aluminium

ALD atomic layer deposition

AlN aluminium nitride

APoT additive power-of-two quantisation

ASPIC application-specific photonic integrated circuit

Au gold

BOX buried oxide

Bragg grating a component with periodically modulated refractive index, acting as a mirror for specific wavelengths of light

BTO barium titanate

CDF cumulative distribution function

CMOS complementary metal–oxide–semiconductor

CNN convolutional neural network

Cr chromium

CROW coupled-resonator optical waveguide

CSAR a type of high-contrast electron-beam resist

CT crosstalk

CW continuous wave

DAC digital-to-analogue converter

DAQ data acquisition

DC directional coupler

DI deionised

DONN diffractive optical neural network

Dstl Defence Science and Technology Laboratory

DVD Digital video/versatile disc

EAM electro-absorption modulator

EBL electron-beam lithography

EDFA erbium-doped fibre amplifier

EIT electromagnetically induced transparency

EMI electromagnetic interference

EO electro-optic

EOM electro-optic modulator

ER extinction ratio

FDTD finite-difference time-domain simulation method

FEM finite element method

FOM figure of merit

FOV field of view

FPCA field-programmable coupler array

FPDA field-programmable disk array

FPGA field-programmable gate array

FPRA field-programmable ring array

FSR free spectral range

FWHM full width at half maximum

GPU graphics processing unit

GSST germanium-antimony-selenium-telluride, typically $Ge_2Sb_2Se_4Te_1$

GST germanium-antimony-telluride, typically $Ge_2Sb_2Te_5$

HF acid hydrofluoric acid

I*–*V current–voltage

IBM International Business Machines Corporation, an American multinational technology company

ICT information and communications technology

IL insertion loss

imec Interuniversity Microelectronics Centre, an international research & development organisation

Industry 4.0 the Fourth Industrial Revolution

INTERCONNECT a photonic integrated circuit simulator by Ansys Lumerical

IoT Internet of Things

IPA isopropyl alcohol

ITO indium tin oxide

Kalman filter an algorithm used in control theory that uses a series of measurements to produce more accurate estimates, named after Rudolf E. Kálmán

LCN London Centre for Nanotechnology

LiDAR light detection and ranging

LLM large language model

LN lithium niobate

LP linear programming

MAC multiply–accumulate

MEMS micro-electromechanical systems

MIBK methyl isobutyl ketone

MMI multi-mode interferometer

MO magneto-optic

MODE a solver for optical waveguides and couplers by Ansys Lumerical

MOS metal–oxide–semiconductor

MPC model predictive control

MRM microring modulator

MRR microring resonator

MVM matrix–vector multiplication

MWP microwave photonics

MZI Mach–Zehnder interferometer

MZM Mach–Zehnder modulator

nichrome nickel–chromium

OSP optical signal processor

p–i–n a junction with an intrinsic semiconductor sandwiched between a p-type semiconductor and an n-type semiconductor

PCB printed circuit board

PCM phase-change material

PD photodetector

PDK process design kit

PECVD plasma-enhanced chemical vapor deposition

PhC photonic crystal

PIC photonic integrated circuit

PID proportional–integral–derivative

PMMA polymethyl methacrylate

PoT power-of-two quantisation

PTC photonic tensor core

Q **factor** quality factor

QCSE quantum-confined Stark effect

QP quadratic programming

RAMZI ring-assisted Mach–Zehnder interferometer

ReLU rectified linear unit

RF radio frequency

RIE reactive-ion etching

RNN recurrent neural network

RTA rapid thermal annealing

$\mathbf{Sb_2S_3}$ antimony trisulfide

$\mathbf{Sb_2Se_3}$ antimony triselenide

SC switching contrast

SCISSOR side-coupled integrated spaced sequence of resonators

SEM scanning electron microscope

Si silicon

SiO_x silicon oxide

SiN silicon nitride

SMU source measure unit

SOI silicon on insulator

SRAM static random-access memory

SWaP size, weight, and power

TCO transparent conducting oxide

TIA transimpedance amplifier

TIR total internal reflection

TO thermo-optic

ToF time of flight

TOPS tera operations per second

TPU Tensor Processing Unit

WDM wavelength-division multiplexing

Neque porro quisquam est qui dolorem ipsum quia dolor sit amet, consectetur, adipisci velit...

There is no one who loves pain itself, who seeks after it and wants to have it, simply because it is pain...

— Cicero's *de Finibus Bonorum et Malorum*

1 Introduction

Contents

1.1 Motivation 3
1.2 Objective 5
1.3 Thesis outline 5
Chapter 2: Literature review 6
Chapter 3: Theory 6
Chapter 4: Device simulation, fabrication, and measurements . 6
Chapter 5: Phase-change photonic computing for defence . . . 7
Chapter 6: Non-volatile field-programmable coupler array . . . 7
Chapter 7: Non-uniformly quantised neural network weights . . 8
Chapter 8: Concluding remarks 8
1.4 Key findings and new science 8
1.5 Statement of originality 9

Humanity is currently undergoing the Fourth Industrial Revolution (Industry 4.0), with key themes including massive interconnection and immense amounts of data, artificial intelligence (AI), and the Internet of Things (IoT). The electronic complementary metal–oxide–semiconductor (CMOS) technology that has taken us where we are today is starting to run into trouble, with the transistor size reaching the point where the fundamental physics of solids will not let it scale down much further [1]. Furthermore, the energy consumption and carbon emissions from the information and communications technology (ICT) sector are becoming significant [2].

Computing using light—known as optical or photonic computing—may help overcome these shortcomings. Light (visible and near-infrared electromagnetic radiation) has long been the medium of choice in telecommunications and data centres [3], as wavelength-division multiplexing (WDM), i.e., simultaneous encoding of several signals onto different wavelengths in the same channel, and high temporal resolution through high data modulation and detection speeds enable massively-parallel, large-bandwidth data transfer with low latency [4]. Furthermore, optical signals in photonic waveguides do not suffer from problems associated with electromagnetic interference (EMI) [5], and they experience lower attenuation, thus producing less heat per unit distance than their electronic counterparts [3].

It is only recently that these and other advantages of photonics have begun to challenge digital electronics for information-processing applications, such as image recognition [4], and the prospects look promising. For instance, the next generation of nanophotonic deep learning architectures will, with all likelihood, be both more energy efficient and more compact—by several orders of magnitude—compared with digital electronics ones [6]. Photonic computing might enable currently unreachable applications, such as high-speed, high-bandwidth, and low-energy processing [3], and we are already seeing the first companies making it out of the prototype stage [7, 8].

For photonic circuits to realise applications beyond the reach of conventional technologies, we must not only learn from the CMOS technology that got electronics to where it is today, but also leverage it. First, we must capitalise on the decades of research that academia and industry have invested in the CMOS technology, resulting in, among other things, high-quality wafers and high-density integration, which are also available for photonic components through foundries [5]. Secondly, to find widespread utilisation, it is paramount that photonics systems are integrated and compact; that is, we need photonic integrated circuits (PICs) that can scale to large numbers of devices, like their electronic counterparts [6]. Finally, co-integration with CMOS electronics is vital in one way or another for virtually all on-chip photonics systems [3], as photonic systems are, with all likelihood, not going

to eliminate the use of many conventional electronic components [3], and we are reliant on electrical driving for most photonic components.

1.1 Motivation

Research on PICs has seen a rapid increase in activity across several fields in recent years, driven by the promise of capabilities beyond electronics-only approaches [9]. Integrated photonics combines ultra-high bandwidth, low latency, and energy efficiency in a way that conventional electronics cannot match, offering a pathway towards systems with minimal size, weight, and power (SWaP). Furthermore, light signals in optical waveguides are immune to EMI, making photonics inherently robust for high-performance information processing. These combined characteristics make photonics naturally suited to computing and signal-processing applications, as well as to high-performance defence scenarios such as target recognition, situational awareness, radar systems, and adaptive control. Several spin-out companies, including iPronics, LightMatter, and Salience Labs, are now pursuing large-scale integration of photonic networks to address such tasks.

Although this work cannot compete with the scaling and integration capabilities of the companies mentioned above, it aims instead to identify better-suited applications for existing device technologies and to develop improved technologies or concepts for existing applications. Photonic devices vary widely in terms of speed, power consumption, footprint, and optical loss, making it crucial to determine the most suitable technology for each specific functional domain. Achieving widespread implementation of photonic circuits, therefore, requires not only the development of components suitable for diverse functions but also a clear understanding of their respective application spaces. As the benefits of photonics depend heavily on architecture, materials platform, electrical interfacing, and other factors, we need to design systems accordingly, because photonics is of no use if we cannot fully exploit its benefits. These guiding principles underpin the investigations presented in this thesis.

Among the many components that constitute a PIC, the photonic switch plays a particularly central role. Photonic switches are modulators and couplers, and serve as core building blocks in integrated systems for optical communications, optical signal-processing platforms, and user-programmable reconfigurable photonics [10], as well as in optical neural networks and quantum information systems [11].

A notable subset of photonic switches is based on phase-change materials (PCMs), a class of materials that enables compact, non-volatile, and reconfigurable devices. Initial work on PCM-based photonic switches utilised either external heaters, free-space lasers, or integrated optical pump lasers to set the state of the switch, none of which are particularly compatible with scaled-up systems. At the time this thesis commenced, a handful of research groups had begun demonstrating integrated heaters in various forms, as such devices provide a valuable and scalable approach to the large-scale integration of future systems. We also wanted to explore this route further. Furthermore, the low-loss PCMs antimony trisulfide (Sb_2S_3) and antimony triselenide (Sb_2Se_3) had only recently been demonstrated in photonic devices [12]. These materials, which modulate only the optical phase rather than amplitude, opened an entirely new class of applications for PCM-based integrated photonics. A significant portion of the early work of this PhD focused on identifying suitable applications for these low-loss materials and developing devices that exploit their unique properties.

Finally, considerable effort was invested in aligning architectural design with application-level requirements. One of the initial goals of this thesis was to find ways to solve computationally intensive defence-related tasks using photonic hardware, for instance, by running neural networks on PCM-based photonic tensor cores (PTCs). However, in its current form, the PTC architecture has several limitations that need to be addressed before the technology can be a one-size-fits-all solution. One such drawback is the limited cyclability and update frequency of the PCM-based photonic switches in the PTC, which means PCM-based PTCs work best for applications where the PCM matrix changes infrequently. Furthermore, the current PTC architecture has a fundamental summation loss and, when used as a machine

learning accelerator, does not inherently allow for non-uniform quantisation of the weights encoded in the PCM elements. By delving deeper into the interplay between architecture and application and reevaluating the problem at hand, we sought to address the current limitations.

1.2 Objective

This thesis aims to demonstrate an alternative approach to the current state of the art in architectures based on photonic switches, with applications in user-programmable reconfigurable photonics and optical neural networks, emphasising the need to solve the right problem with the right platform. In this thesis, I present the following concepts:

- **Defence-relevant computing task on a photonic platform:** A recurring version of the PTC can be used to significantly speed up model predictive control (MPC) for target interception (Chapter 5).

- **Small-footprint and low-power reconfigurable optical platform:** An alternative to existing reconfigurable optical platform technology, based on the low-loss PCM Sb_2Se_3 for non-volatility (Chapter 6).

- **Non-uniformly quantised photonic neural network weight:** A conceptually new type of amplitude modulator for photonic neural networks (Chapter 7).

Through these projects, I hope to bridge the gap between innovation and application.

1.3 Thesis outline

The chapters of this thesis are organised in the following way:

Chapter 2: Literature review

Chapter 2 aims to provide the reader with a comprehensive overview of the field of photonic switches to date. The chapter first introduces the fundamental operating principles, structural configurations, and underlying physical mechanisms of photonic switches. Special attention is given to devices based on PCM. Finally, the state of the art of photonic switch networks used for computing and signal processing is outlined, with a focus on architectures such as photonic neural networks and recirculating mesh networks, which exemplify the potential of programmable photonics for scalable information processing.

Chapter 3: Theory

This chapter introduces the theoretical background and establishes the foundation for the remainder of this thesis. Fundamental concepts of light propagation and modulation are presented, along with the governing equations for simple photonic devices such as Mach–Zehnder interferometers (MZIs), microring resonators (MRRs), and directional couplers (DCs). The characteristics of PCMs and the operating principles of PTCs based on them are then described in detail. Finally, the theoretical framework underlying sensor fusion and MPC is introduced, along with the algorithms commonly used to solve the latter.

Chapter 4: Device fabrication and measurements

In Chapter 4, the general and device-specific fabrication procedures for the PICs developed in this work are described in detail. The chapter outlines the complete process flow, from substrate preparation and patterning to material deposition and device measurement, highlighting the techniques and instruments employed to ensure repeatability and performance consistency. Particular emphasis is placed on the experimental characterisation of the low-loss PCM Sb_2Se_3, including initial optical measurements and supporting simulations used to validate its suitability for integrated photonic applications. Finally, the optical and electrical measurement setups used to evaluate the fabricated devices are presented and discussed.

Chapter 5: Phase-change photonic computing for defence

Chapter 5 focuses on the application of PCM-based computing cores for demanding defence-related scenarios, in particular sensor fusion and MPC. Emphasis is placed on the implementation of photonic architectures capable of solving the quadratic programming problems central to MPC, demonstrating their potential for real-time decision-making in mission- and safety-critical environments. The limited feasibility of employing PTCs for sensor fusion is also examined, illustrating that the developed photonic cores are inherently better suited to some computational tasks than to others. Finally, estimates of latency and power consumption are presented to quantify the performance and to underpin the practical relevance of these PCM-based computing cores for next-generation defence and autonomous systems.

Chapter 6: Non-volatile field-programmable coupler array

The development and characterisation of a small-footprint, non-volatile field-programmable coupler array (FPCA) based on the low-loss PCM Sb_2Se_3 are presented in Chapter 6. The chapter begins with an overview of the FPCA concept and its role as a reconfigurable photonic platform for integrated circuit applications. The design, simulation, fabrication, and characterisation of Sb_2Se_3-based phase shifters integrated with microheaters are then described, with particular attention to device geometry, switching dynamics, and cyclability. These phase shifters are incorporated into MZI couplers to form the fundamental units of the FPCA. Optical measurements demonstrate the platform's reconfigurable functionality through various synthesised structures, including interconnects, ring resonators, and delay lines. Simulations of a synthesised temporal differentiator further illustrate the versatility of the architecture. The chapter concludes with a discussion of the current technological limitations and an outlook on opportunities to enhance performance, scalability, and functionality in future field-programmable photonic systems.

Chapter 7: Non-uniformly quantised neural network weights

Chapter 7 explores the concept and realisation of a non-uniformly quantised weight for analogue neural networks, implemented through an electrically tunable integrated photonic device, a ring-assisted Mach–Zehnder interferometer (RAMZI). The chapter is motivated by the limitations of uniform quantisation in neural networks, where equal spacing between discrete weight levels can lead to suboptimal performance and low-accuracy models. Here, a device architecture is introduced that inherently produces a non-linear electrical-to-optical transfer function, enabling non-uniform quantisation directly in the optical domain. The design process, simulations, and optical measurements of the device are described, and practical aspects of using the RAMZI weight element in a real system are examined in detail. Finally, the potential for system-level integration and energy-efficient operation is discussed, supported by energy calculations and considerations of future design optimisations.

Chapter 8: Concluding remarks

In the final chapter of this thesis, I revisit its objectives, summarise the main achievements and experimental findings, and outline potential directions for future research arising from this work.

1.4 Key findings and new science

The key results of this thesis concern the development of photonic switch networks using novel device technologies. Three different concepts were devised, each addressing a separate aspect of computing and signal processing using light. First, I investigated defence-relevant applications where the use of PCM-based PTCs significantly speeds up the process by leveraging the platform's benefits while avoiding its limitations. In particular, MPC appears to be a promising candidate for problems where this low-latency, energy-efficient platform outperforms electronics by orders of magnitude; however, similar optimisation problems can also be solved using this compact analogue computing engine. Secondly, I designed and characterised the

first experimental demonstration of a non-volatile FPCA based on the low-loss PCM Sb_2Se_3 [13]. The device demonstrated field-programmable path and phase control of light via integrated microheaters for PCM switching, enabling reconfigurable circuit functionality suitable for photonic signal routing and processing in a small-footprint platform with no static power consumption from the mesh. Finally, a novel device concept was introduced that implements a non-uniformly quantised weight for analogue neural networks by exploiting the intrinsic non-linear optical transfer function of a carefully designed RAMZI. The device is the first of its kind, offering a new approach to photonic computing and paving the way for few-bit, high-accuracy photonic inference systems. Collectively, these results contribute to the emerging field of integrated photonic computing hardware for high-performance information processing in both civilian and defence-related contexts.

1.5 Statement of originality

This thesis is the original work of the author and reflects research undertaken during the course of their DPhil. Unless explicitly stated otherwise, all results presented are the author's own. Contributions from collaborators have been appropriately acknowledged where relevant.

All intelligent thoughts have already been thought;
what is necessary is only to try to think them again.

— Johann Wolfgang von Goethe

2

Photonic switch networks for optical computing and signal processing: A review

Contents

2.1 Photonic integrated switch structures 13
2.1.1 Optical modulators 13
2.1.2 Mach–Zehnder switches 13
2.1.3 Resonator switches 14
2.1.4 MEMS switches 15
2.1.5 Other switch designs 16
2.2 Switching mechanisms 19
2.2.1 Thermo-optic effect 21
2.2.2 Electro-optic effects 22
2.2.3 MEMS actuation 24
2.2.4 Optical phase-change materials 25
2.2.5 Other switching mechanisms 26
2.3 Photonic switches with optical phase-change materials 29
2.3.1 Fundamental properties of phase-change materials . . . 29
2.3.2 Low-loss phase-change materials 31
2.3.3 Integration of phase-change switches 35
2.3.4 Phase-change photonics in extreme environments 35
2.3.5 Application areas of phase-change switches 36
2.4 Photonic computing and signal processing 37
2.4.1 Integrated optical signal processors 38
2.4.2 Photonic mesh networks 40
2.4.3 Photonic recirculating mesh networks 42
2.4.4 Photonic neural networks 46
2.5 Conclusion 52

Photonic switches are indispensable to PICs. The word *switch* generally refers to an element that changes the transmission of light through the waveguide—also called an *on-off switch* [14] or *modulator* [15]—or it can mean an active splitter, a structure that directs light out of two or more arms based on the state of the switch [16], which is also called an *optical power splitter* [17] or *tunable coupler* [18]. We can therefore separate photonic switches into two different classes: modulators and couplers, where modulators are elements with one input and one output (1×1), while couplers have m inputs and n outputs ($m\times n$), where $m, n \in \mathbb{N}^+$, $m \geq 1$, and $n > 1$.

In this review, I will focus on photonic switches in next-generation PICs for computing and signal processing. As the large-scale deployment of photonic technologies depends on high-density and compact integrated devices, this review is restricted to on-chip optical switches (for an overview of free-space switches, see, for instance, Ref. 19). Finally, only power switches (not mode-, polarisation-, or wavelength-multiplexed switches) are reviewed here. There have been review papers over the years that focus on the field of photonic switches [19–22], but only in the last few years has the growing interest in PCMs been taken into account [23, 24]. PCM-based PICs in general have already been recognised as potential enablers of next-generation systems [3, 4, 25], and there has been much work recently on PCM-based photonic switches. Here, I keep an extra focus on such switches and point out areas where they might offer an advantage.

The remainder of this chapter is organised as follows: In Sections 2.1 and 2.2, respectively, I discuss the different structures and switching mechanisms used in photonic modulators and couplers. Section 2.3 gives an introduction to on-chip photonic switches based on PCM, with emphasis on integration and applications. This chapter concludes with an overview of photonic switch networks for photonic computing and optical signal processing, followed by a summary of the key points.

2.1 Photonic integrated switch structures

There are numerous photonic switch structures in the literature. The simplest form of a switch is a modulator, a single element that changes the phase and/or amplitude of the incoming light. A modulator can be used for coupling by integrating it with other structures, for example, when using a phase modulator in a Mach–Zehnder interferometer (MZI). Some of the designs presented here, such as the MZI, can serve as both modulators and couplers, whereas others function only as couplers. A multitude of designs for photonic modulators and couplers exist, ranging from those based on interferometric effects to those based on light coupling. In this section, I will present the most common of these designs.

2.1.1 Optical modulators

In contrast to the other switch structures in this section, the optical modulator is a type of photonic switch that cannot be used for coupling on its own, but rather for modulation of the intensity or phase of the light.[1] Phase and intensity modulation are achieved by changing the real and imaginary parts of the effective refractive index of the modulator, respectively (see Section 3.1.1 for more on effective refractive index). The mechanisms employed for optical modulators will be covered in greater detail in Section 2.2.

2.1.2 Mach–Zehnder switches

Mach–Zehnder switches are based on the interferometric structure named after the physicists Ludwig Mach and Ludwig Zehnder. In the interferometer, the light is split into two arms, the phases of the light beams are modulated independently, and subsequently the phase-shifted light beams are recombined. An MZI switch can have both one and two input and output ports (see Fig. 2.1) and use virtually any type of phase shifter in its arms, making it a very versatile structure. When the MZI is in a 1×1 configuration, it is commonly referred to as a Mach–Zehnder modulator (MZM). Generally, MZI switches have a wide optical bandwidth [26].

[1] Frequency or polarisation modulation is also possible, but falls outside the scope of this work.

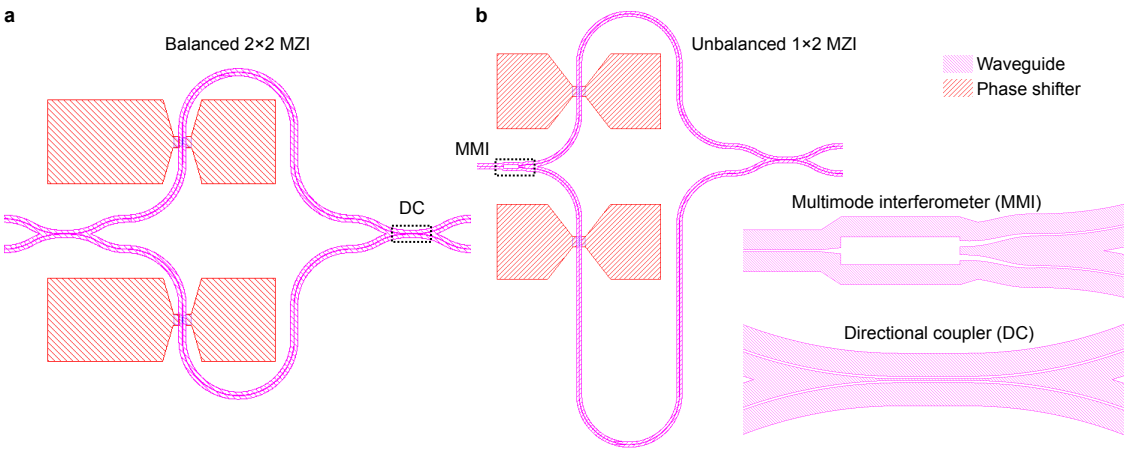

Figure 2.1: Two examples of MZI switches. **a**, A balanced MZI with two inputs and two outputs (2×2) and phase shifters in each arm. **b**, An unbalanced MZI (different path length of the two arms) with one input and two outputs (1×2) and phase shifters in each arm. The insets show a 1×2 passive multi-mode interferometer (MMI) (see Section 2.1.5), used at the input of the device in **b**, and a 50:50 DC, used at the input of the device **a** and at both outputs.

Several alternatives to the basic design have been proposed, for example, with waveguides looping back [27] or spiralling [28] to reduce power consumption and switching time, but at the expense of a larger footprint and a slight increase in loss. For an extensive overview of MZI switch designs, see Fig. 2 of Tu *et al.* [21].

2.1.3 Resonator switches

Ring resonators and disk resonators are resonant structures that are used for photonic switches, filters, temporal integrators and differentiators, and Hilbert transformers [29, 30], to mention some of their many uses. The working principle of ring and disk resonators is based on evanescent coupling and interference between one or more waveguides and a ring or disk resonator. Light is sent through the *input* port and couples to the ring or disk structure. Because light behaves like a wave, we get resonances (covered in Section 3.4), where certain wavelengths experience destructive interference, resulting in minimum transmission through the *through* port, while other wavelengths experience constructive interference and maximum transmission. By changing the resonant wavelengths, which can be done using a phase shifter in the ring, we can effectively modulate the wavelength of interest. In ring resonator couplers, the ring also couples to an output arm, called the *drop* port.

Figure 2.2a shows a microring resonator (MRR) with two inputs and two outputs. In theory, the ring resonator switch can have multiple input and output ports, but it is typically used only in configurations with one or two input and output ports each.

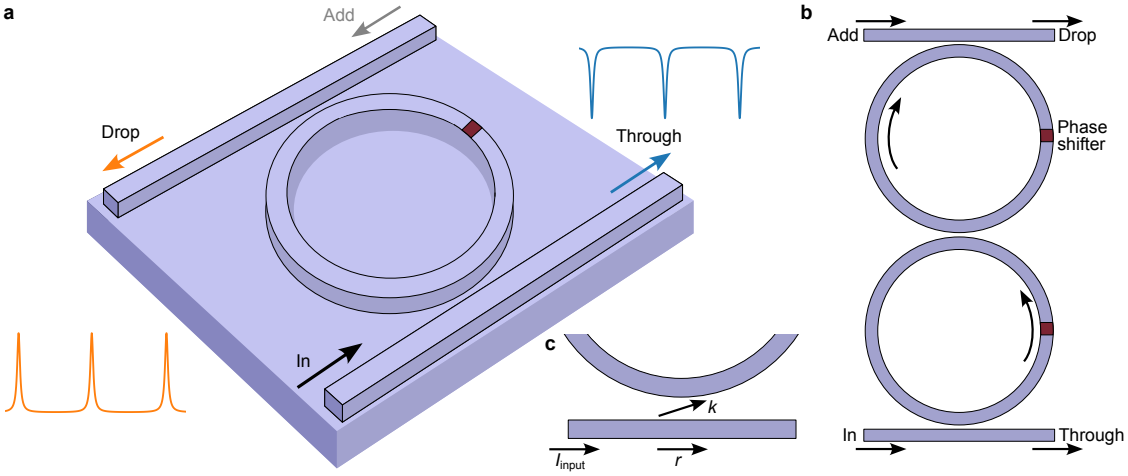

Figure 2.2: Examples of microring resonator switches. **a**, A 2×2 MRR with the characteristic through and drop port transmission in blue and orange, respectively. **b**, Higher-order MRR with two tuned rings, inspired by Fig. 21 in Bogaerts *et al.* [31]. **c**, Zoom-in of the coupling region, where r is the self-coupling coefficient and k is the cross-coupling coefficient.

The inherently narrow Lorentzian-shaped resonances of ring resonators (the blue and orange plots in Fig. 2.2a) give MRRs high modulation strength with comparatively lower power consumption than MZIs. Switches based on ring resonators can be extremely small if the refractive index contrast between the waveguide material and its surroundings is high, with radii down to a few microns in silicon [31, 32]. Higher-order coupled-ring switches have also been shown (Fig. 2.2b), which can provide flatter (non-Lorentzian) passbands and sharper roll-off [20, 31]. These structures allow lower power consumption for the phase shifter and lower crosstalk between the ports, but also result in higher insertion loss at the drop port [20].

2.1.4 MEMS switches

A special class of switches is fundamentally different from the switches discussed in the previous subsections, as their switching is driven by moving parts. In integrated

photonics, these switches are controlled by micro-optoelectromechanical actuators that use electrical signals to mechanically modify components through various physical effects (see Section 2.2.3), thereby affecting the light path. Such devices are generally classified as micro-electromechanical systems (MEMS).

Where other switch designs are generally confined to the lateral plane, MEMS switches can also exploit the vertical dimension [33]; compare Fig. 2.3a,b with Fig. 2.3c. This three-dimensional nature is among the significant benefits of these switches, allowing them to avoid cumulative losses on the chip. In larger systems with many input and output waveguides, such as a photonic switch mesh, the switches cause more optical loss than simple waveguide crossings. MEMS switches can, therefore, allow significantly lower overall on-chip loss as only one switch is used per routing event [33]. As seen in Fig. 2.3d–i, MEMS switches can be used both for coupling and phase modulation of light. More details on the switching mechanism of MEMS switches will be presented in Section 2.2.3.

2.1.5 Other switch designs

In addition to the designs mentioned above, other noteworthy switch designs include directional coupler, multi-mode interferometer, and photonic crystal switches.

Directional coupler switches

The basic principle of a directional coupler (DC) switch is shown in Fig. 2.4 for 1×2 and 2×2 switches, where the switching element changes the effective refractive index, n_{eff}, of one or more of the waveguides. Light couples from the input waveguide to the adjacent waveguide if the effective refractive indices of the two waveguides are similar. In the case of the 2×2 switches in Fig. 2.4b, light is then further coupled from the middle waveguide(s) to the *cross* port (labelled *Drop* in the figure). If Δn_{eff} between the two waveguides is large, it indicates a phase mismatch and the light continues through the initial waveguide to the *bar* port [37]. Alternatively, if the switching element cannot introduce a large difference in Δn_{eff}, it is possible to design the switch in such a way that the light can couple across and back again

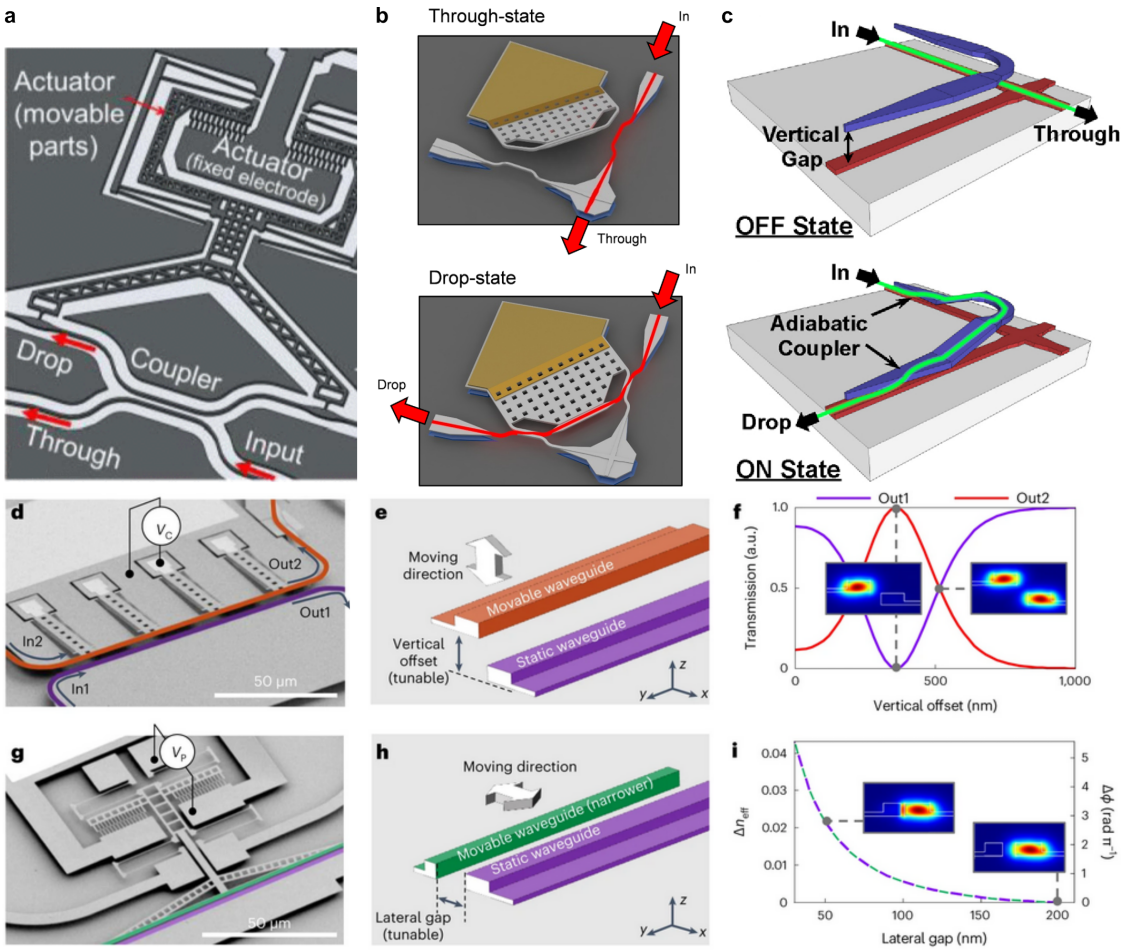

Figure 2.3: Examples of switches with MEMS actuators. **a**, A 2 × 2 MEMS-based DC switch. The actuator brings the drop port towards the input port to couple light. **b**, A 2 × 2 MEMS switch with two coupling sections. **c**, A three-dimensional (3-D) MEMS switch. **d**–**f**, A 2 × 2 capacitive electrostatic MEMS coupler with vertical actuation, **g**–**i**, A capacitive electrostatic MEMS phase modulator with lateral actuation. Figure reproduced with permission from: **a**, Ref. 34, © 2014 IEEE; **b**, Ref. 35, © 2015 Optica Publishing Group; **c**, Ref. 36, © 2016 Optica Publishing Group; **d**–**i**, Ref. 18, © 2023 Springer Nature Ltd.

for the *bar* state and just couple across for the *cross* state [38, 39]. DC switches are narrow in physical width and can have a relatively wide optical bandwidth of tens of nanometres in the telecommunications bands [16].

Multi-mode interferometer switches

Multi-mode interferometer (MMI) switches are made up of m input waveguides going into the central part, a multi-mode waveguide, which is connected to n output waveguides. Figure 2.5a,c shows 1 × 2 MMI switches with the low-loss PCM

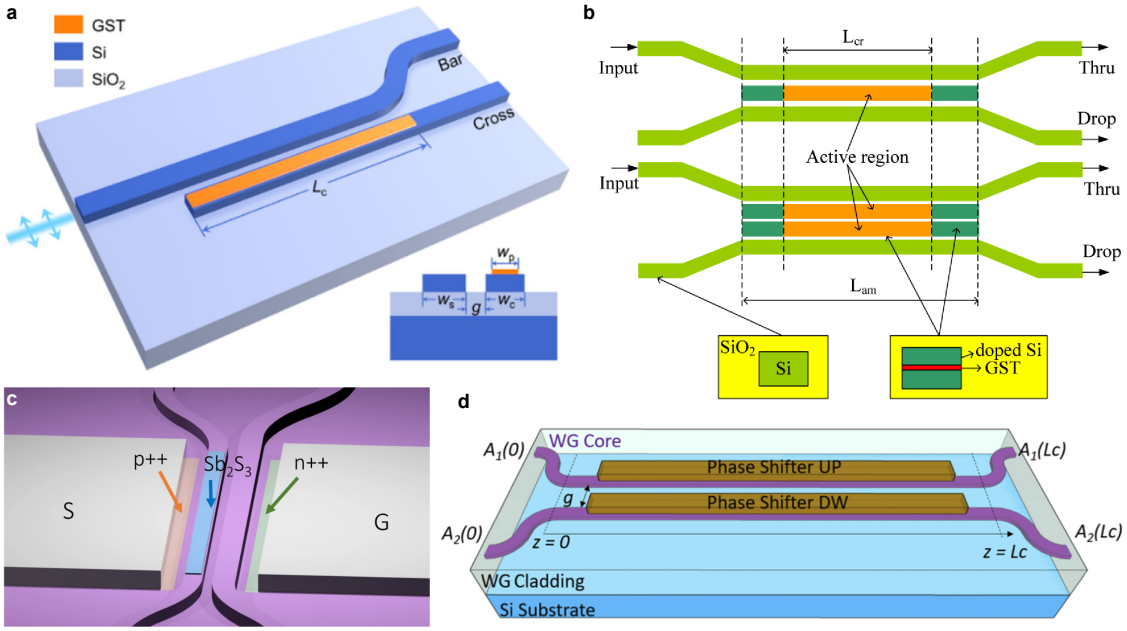

Figure 2.4: Examples of DC switches. **a**, A 1×2 DC switch with the PCM $Ge_2Sb_2Te_5$. **b**, Two 2×2 DC switches with the PCM $Ge_2Sb_2Te_5$, showing both one and two intermediate waveguides for coupling. **c**, A 2×2 DC switch with the low-loss PCM Sb_2Se_3. **d**, A 2×2 DC switch with thermo-optic (TO) heaters on both arms. Figure reproduced with permission from: **a**, Ref. 37, © 2019 American Chemical Society; **b**, Ref. 40, © 2015 Optica Publishing Group; **c**, Ref. 39 under a Creative Commons licence CC BY 4.0; **d**, Ref. 41, © 2019 Optica Publishing Group.

Sb_2Se_3. The coupling of the MMI switches is dictated by the inversely designed PCM pattern (Fig. 2.5a) and the PCM metasurface (Fig. 2.5c). In Fig. 2.5b, two TO heaters change the real part of the refractive index of the multi-mode waveguide, leading to a change of the interference pattern at the output waveguides. Similarly to DC switches, MMI switches have a reasonably narrow physical width and can also have a wide bandwidth [42].

Photonic crystal switches

A photonic crystal (PhC) is an optical nanostructure with periodic changes in the refractive index, which leads to photonic band gaps that prohibit certain wavelengths of light from propagating. One example, a PhC nanobeam resonator, is shown in Fig. 2.6a–c [45]. Because light forms standing waves rather than travelling waves in the structures, the operating bandwidth is intrinsically narrowband. In his review from 2018, R. Soref gives a good overview of PhC-based switches where

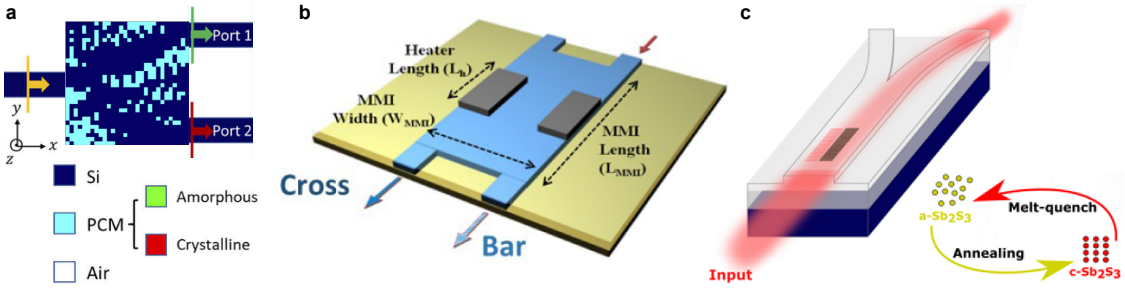

Figure 2.5: Examples of multi-mode interferometer switches. **a**, A 1×2 inversely-designed MMI switch with the low-loss PCM Sb_2Se_3. **b**, A 2×2 MMI switch with TO heaters. **c**, A 1×2 Y-branch switch with a metasurface of the low-loss PCM Sb_2Se_3 controlling coupling. Figure adapted with permission from: **a**, Ref. 43, © 2021 Optica Publishing Group; **b**, Ref. 44, © 2016 Optica Publishing Group; **c**, Ref. 42 under a Creative Commons licence CC BY 4.0.

the PhC is integrated in waveguides in either 2×2 MZI or DC structures [22]. In these structures, the *forbidden band* of the PhC (e.g., a nanobeam with holes of varying size or a Bragg grating; see Fig. 2.6d,e) determines the transmitted and reflected wavelengths. As we can see in Fig. 2.6d–f, the drop and through ports are organised like those of single MRRs rather than those of conventional MZIs or DCs. Two-dimensional (2-D) PhC switches,[2] like the one shown in Fig. 2.6g, has also been demonstrated recently [46], in which the PhC itself serves as both the waveguide and the resonant element. In this design, light can be directed not only out of the through and drop ports, but also out of the add port.

2.2 Switching mechanisms

A wide range of physical effects can be harnessed to realise optical modulation and switching, each offering distinct advantages in terms of speed, energy efficiency, scalability, and fabrication compatibility. The most commonplace switching mechanisms for photonic switch networks—and the ones presented in this thesis—include those based on: the thermo-optic effect, relying on thermal heating; electro-optic effects, where the refractive index is altered directly by an electric field, carrier-based effects, which exploit changes in free-carrier density; MEMS actuators based on

[2]Here, 2-D refers to the dimensionality of the photonic crystal's periodic structure (two spatial dimensions), not to be confused with 2-D materials.

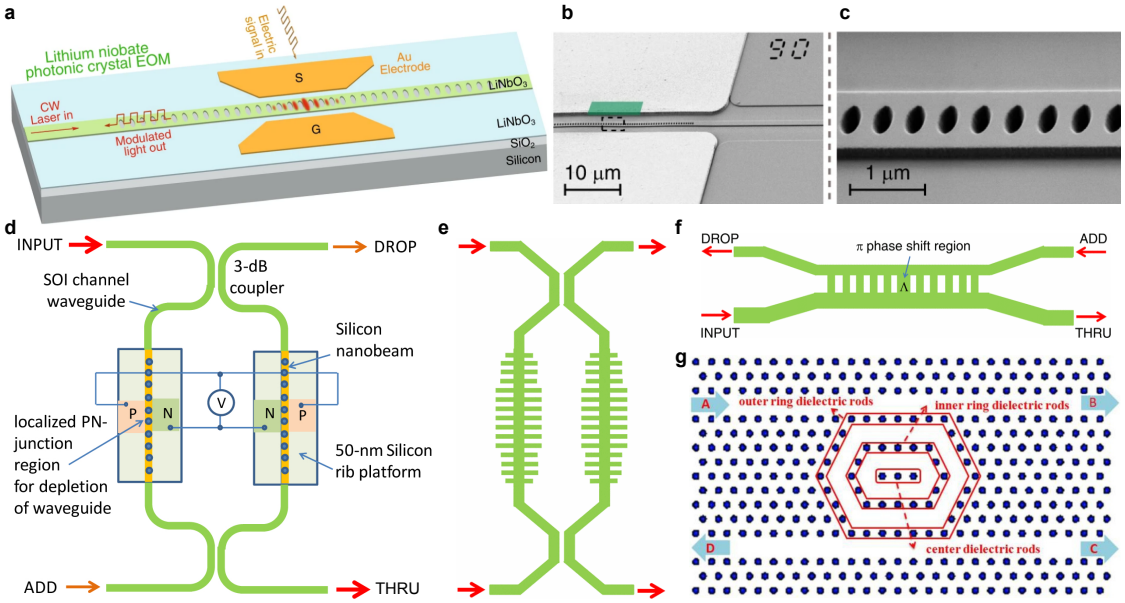

Figure 2.6: Examples of photonic crystal switches. **a**–**c**, One-dimensional (1-D) lithium niobate (LN) PhC nanobeam modulator. **d**, A 2×2 dual-nanobeam PhC MZI switch with p–n junctions for carrier depletion (see Section 2.2.2) of the nanobeam. **e**, Alternative design for the 2×2 PhC MZI switch with a Bragg grating. **f**, A 2×2 PhC DC switch. **g**, 2-D PhC switch with $Ge_2Sb_2Se_4Te_1$ rods making up the circuit. The modulators controlling coupling in **e** and **f** are not shown. Figure reproduced with permission from: **a**–**c**, Ref. 45 under a Creative Commons licence CC BY 4.0; **d**, Ref. 47, © 2015 Optica Publishing Group; **e**–**f**, Ref. 48, © 2017 IEEE; **g**, Ref. 46, © 2020 Elsevier B.V.

electro-mechanical phenomena; and reconfigurable non-volatile materials such as PCMs. In addition, a range of emerging switching mechanisms that remain outside mainstream applications will be briefly addressed.

Before delving into the switching mechanisms, we must first consider the attributes of an ideal photonic switch. The ideal photonic switch does not exist, but is characterised by several metrics that are almost universally favourable. The switch should have a small footprint, low insertion loss, high switching contrast (extinction ratio), and minimal crosstalk between ports. Furthermore, we want it to consume minimal static power and have low dynamic energy consumption. For many applications, especially light modulation in photonic computing, it is also beneficial that the switch offers precise multi-level tunability and stability, a fast switching speed with short reconfiguration times, and high cyclability. Depending on the use, a wide optical bandwidth may also be required. Finally, we want CMOS compatibility, both in terms of the materials (integration in foundries) and the operating powers

and voltages. However, in reality, each technological approach involves characteristic trade-offs among these performance metrics. The following subsections provide an overview of the most widely explored technologies for photonic switches.

2.2.1 Thermo-optic effect

The thermo-optic effect (TO effect, alternatively known as the thermo-optical or thermal–optical effect) is based on a change in the real part of the refractive index of the material with temperature. The change in the phase of the light, $\Delta\varphi$, is proportional to the change in temperature, ΔT, following $\Delta\varphi \propto \partial n/\partial T \Delta T$, where $\partial n/\partial T$ is the TO coefficient (Section 3.1.1). In silicon, the TO coefficient is relatively high, around $1.86 \times 10^{-4}\,\mathrm{K}^{-1}$ at a wavelength of 1550 nm [49], and as a consequence, the TO effect has been frequently used in the silicon on insulator (SOI) platform. On the other hand, in silicon nitride (SiN), $\partial n/\partial T \sim 2.5 \times 10^{-5}\,\mathrm{K}^{-1}$, requiring either a larger temperature difference or a longer interaction length for this platform compared with SOI [50].

TO phase shifters are widely used in silicon due to their simplicity and ease of fabrication. Because TO switches work by changing the real part of the refractive index, they can, in principle, be lossless [51]. The trade-off is that the effect is volatile, meaning the switch is only active when voltage is applied, which leads to constant power consumption. Because we need to heat a relatively large volume, the energy efficiency is usually on the order of tens of milliwatts per π phase shift (i.e., the phase shift required to switch to the other state of an MZI switch) [20]. Furthermore, the TO effect is typically slow, since it involves heating large volumes of material and since some platforms, such as SOI with a thick buried oxide layer, have poor heat-sinking abilities [52]. In silicon, this leads to switching times in the microsecond range [51, 53, 54], depending on the heater material, amongst other things. Therefore, TO switches have typically been considered for low-loss and low-speed applications such as optical routing [55].

2.2.2 Electro-optic effects

Electro-optic (EO) effects refer to modulation methods that arise from alterations in the refractive index induced by an electric field, with the primary effects traditionally useful in semiconductor materials being the Pockels, Kerr, and Franz–Keldysh effects [56]. When an electric field is applied to the material, either the real part of the refractive index, n, changes—known as electro-refraction—or the imaginary part of the refractive index, κ, changes—known as electro-absorption—or a combination of the two happens. Alternatively, the electrical field can introduce a change in the density of free carriers, called the plasma dispersion effect [56], which will be covered at the end of this section.

Electro-refractive effects

Electro-refractive effects include the Pockels and Kerr effects. The Pockels effect is an EO phenomenon in which the refractive index of a material varies linearly with an applied electric field. Because the Pockels effect is weak in pure silicon at telecommunications wavelengths [57], lithium niobate ($LiNbO_3$, LN) has been the dominant platform for decades, offering low-loss, high-quality devices with excellent linearity. More recently, barium titanate ($BaTiO_3$, BTO), lithium tantalate ($LiTaO_3$), aluminium nitride (AlN), and strontium titanate ($SrTiO_3$, STO) have emerged as promising alternatives [58–60]. Additionally, EO polymers can be used, either on their own or as an active cladding for easier integration with silicon [61]. The Kerr effect is usually too small in most practical materials for efficient integrated modulators and has therefore not been reported as much as the Pockels effect [62], although there are some exceptions [63].

Devices based on the Pockels effect are among the fastest phase shifters available, with very low optical insertion losses, making them particularly interesting for high-performance applications like quantum computing [58]. EO modulators typically achieve a half-wave voltage-length product ($V_\pi \cdot L$) of a few V·cm in thin-film LN [29] and down to $\sim 0.5\,\mathrm{V\cdot cm}$ in barium titanate (BTO) thin films [58]. Driving voltages of a few volts are sufficient in many designs, which enables compatibility

with CMOS drivers. The losses can be negligible (e.g., only 53 dB m^{-1} in Ref. 58) and the modulation bandwidth can extend well beyond 50 GHz [29, 64, 65]. The downsides include a typical power consumption of a few milliwatts [29] and lengths of up to a few centimetres [66].

Electro-absorption effects

Electro-absorption modulators (EAMs) are based on changes in the absorption coefficient under an applied electric field, exploiting either the Franz–Keldysh effect in bulk semiconductors or the quantum-confined Stark effect (QCSE) in quantum wells [56]. InP- and Ge-based devices are commonly employed, and their relatively compact footprint makes them attractive for dense integration [61]. Typical device lengths are tens of micrometres [67, 68], with extinction ratios exceeding 10 dB achievable [69]. Since the mechanism is absorptive, the intrinsic insertion loss is greater than that of phase modulators, often exceeding several dB [69]. The required driving voltages are typically in the range of 1 V to 3 V, with modulation depths of less than one dB μm^{-1} and power consumption ranging from a few milliwatts to tens of milliwatts [70, 71]. EAMs are well suited for high-speed applications, with bandwidths beyond 50 GHz demonstrated [72].

Plasma dispersion effect

None of the electro-refractive and electro-absorptive effects in the previous sections are strong enough in silicon for practical applications in the telecommunications window of the electromagnetic spectrum [56]. Consequently, EO switches in silicon are instead based on the plasma dispersion effect, also called the free-carrier effect, in which the concentration of free holes and electrons (through carrier injection, accumulation, or depletion) changes the complex refractive index of the material [56, 57], again changing the phase and intensity of the light. The most common implementations of the plasma dispersion effect include reverse-biased p–n junctions for carrier depletion, p–i–n junctions for carrier injection, and metal–oxide–semiconductor (MOS) capacitors for carrier accumulation, but transparent conducting oxides (TCOs) such as indium tin oxide (ITO) and 2-D

materials have also been used for the latter [61]. The plasma dispersion effect remains attractive due to its CMOS compatibility.

The plasma dispersion effect is volatile and also relatively weak, requiring long path lengths of typically hundreds of micrometres [20, 51]. The major benefit is high speed, exceeding 30 GHz in carrier-depletion devices [73]. Carrier injection devices can provide larger index shifts but suffer from slower response times (a few nanoseconds) due to minority carrier lifetimes [52, 56, 74]. Unlike the other two mechanisms, carrier accumulation devices can be made non-volatile; however, they require high programming voltages and often exhibit limited cyclability [23]. One disadvantage of the plasma dispersion effect in phase modulators is the absorption caused by the free carriers responsible for the phase shift, which results in excess insertion loss. When used in couplers, this phase-dependent loss leads to crosstalk and asymmetric output from the switch [26, 74, 75]. The power consumption of EO switches varies from sub-milliwatt to a few tens of milliwatts, depending on the interaction length and technology [20, 76], with $V_\pi \cdot L$ products around 1 V·cm [73]. For devices that require high concentrations of carriers, the applied power can lead to parasitic self-heating of the phase tuning section that introduces a counteracting TO phase delay [52, 77]. To correct for phase errors in the EO phase shifter of, for instance, an MZI switch, EO and TO phase shifters are sometimes used in conjunction, adding to both the footprint and power consumption of the switch [78].

2.2.3 MEMS actuation

MEMS actuation refers to photonic switches where MEMS technology, i.e., systems with electrostatic, thermal, magnetic, or piezoelectric actuators, is used to modulate or couple light through waveguides [79]. MEMS couplers are based on relatively straightforward concepts, where in-plane [35] or out-of-plane coupling [33, 36], either evanescently [36] or directly through butt-coupling [80], redirects the light from a bus waveguide into another waveguide. Another option is to use the MEMS actuators to allow or disrupt the field coupling between adjacent waveguides [81]. Other designs include rotating MEMS devices, which, in theory, can accommodate

numerous output waveguides [82]. For phase shifters, the effective refractive index of a waveguide system comprising the bus waveguide and an actuated, perturbing waveguide changes with the gap between the waveguides [83, 84] and, therefore, a phase shift can be introduced in the bus waveguide by varying this gap (Fig. 2.3g–i).

MEMS switches have the potential to allow wide optical bandwidth—up to 300 nm was shown for out-of-plane MEMS switches in Fig. 2.3c using adiabatic couplers [36]. Furthermore, optical insertion losses can be very low (<0.5 dB) [18]. Driving voltages are typically relatively high (tens of volts); however, static power consumption and switching energy can be as low as a few femtowatts and tens of picojoules [18, 85], and the technology can even be made non-volatile using latches (although not with multi-level settings) [84, 86, 87]. The speed of MEMS switches is limited by mechanical movement and typically ranges from just below a microsecond [33, 36] to tens of microseconds [18, 34]. Due to the mechanical movement of MEMS switches, these devices often suffer from several problems, including stiction (microstructures getting stuck) and mechanical degradation (fatigue, creep, residual stress, distortion) [88], which limits their long-term use. For more extensive reviews of MEMS switches, see the papers by Chollet [89], Du *et al.* [90], and Errando-Herranz *et al.* [79].

2.2.4 Optical phase-change materials

Optical PCMs are materials that undergo a reversible phase transition in response to external stimuli [61] (not to be confused with another group of phase-change materials which release or absorb energy at phase transition). Of the optical PCMs, there are two different types: transition metal oxides (e.g., vanadium dioxide, VO_2) that experience a transition from one crystalline phase to another and chalcogenide-based alloys (e.g., germanium-antimony-telluride, typically $Ge_2Sb_2Te_5$ (GST)) that experience a structural transition between an amorphous (disordered) and crystalline (ordered) state. In this work, when writing PCM, it is this latter group that I refer to.

When undergoing a crystalline-to-amorphous phase transition, the optical, electrical, and mechanical characteristics of the PCM change considerably [91], with

large contrast in their refractive index [92], Young's modulus [93], and resistivity [94]. Due to the large refractive index contrast, PCMs have been used for decades in optical storage technology (e.g., rewritable DVDs) [95] and more recently for reflective display [96] and smart glazing technologies [97], as well as for analogue in-memory computing (AIMC) [98, 99].

For photonic switches, the refractive index change of the PCM can be used for amplitude and/or phase modulation, or to allow light to couple evanescently between waveguides in, for example, a DC switch [16]. The strong modulation per unit length allows for ultra-compact devices, often only a few micrometres long. Furthermore, the modulation is non-volatile, meaning that no power is required to maintain the state of the PCM and, by extension, the switch. Because we usually heat the material to induce phase change, the dynamic power can be relatively large. Some of the main challenges stopping the widespread integration of PCMs include material endurance and integration into foundries. I will dive deeper into PCMs in Section 2.3.1 later in this chapter.

2.2.5 Other switching mechanisms

Beyond the switching mechanisms discussed above, a wide range of alternative approaches to optical switching has been explored. A recent review of integrated optical memristors by Youngblood *et al.* highlights several mechanisms with potential for non-volatile optical switching, including the magneto-optic (MO) effect, ferroelectric domain switching in materials such as BTO, electrochromic modulation via ion intercalation, and optically-addressable electronic memristors based on filamentation [23]. Additionally, acousto-optic effects, liquid crystals, 2-D materials such as graphene, and all-optical switching are interesting volatile alternatives.

The emerging material BTO is of particular interest because of its strong Pockels effect (see Section 2.2.2) that enables low-loss, high-efficiency phase modulation [58]. However, BTO is also a ferroelectric material, meaning we can store optical information by switching ferroelectric domains in the material [23]. Recent demonstrations of this non-volatile platform include a phase shifter with $L_\pi = 1\,\mathrm{mm}$ with only $0.5\,\mathrm{dB}$

insertion loss [100]. As early-stage integration challenges are rapidly overcome [58], these devices are emerging as a promising route to relatively dense, non-volatile switch fabrics with attractive modulation efficiencies.

Magneto-optic (MO) modulation represents an alternative approach to non-volatile, phase-only tuning, in which alloys such as cerium-substituted yttrium iron garnet (Ce:YIG) exhibit refractive-index changes under an externally applied magnetic field [101, 102]. The MO effect is relatively weak (and therefore mainly used in microring modulators) and MO materials tend to be lossy; the insertion loss of the silicon/Ce:YIG rings was 2.5 dB in Ref. 101. Nevertheless, MO devices have recently demonstrated energy-efficient GHz-speed modulation with 2.4 billion programming cycles [103] and have been shown to work at cryogenic temperatures [101], attracting interest in this approach for amplitude modulation.

Another active research direction involves incorporating 2-D materials onto silicon photonics. Graphene, in particular, provides extremely strong electro-absorption and broadband electro-refractive modulation [104]. Graphene-based EAMs have been demonstrated with high efficiency, e.g., a 34 µm long modulator with 6 dB modulation strength and only 0.2 dB insertion loss [105]. Other 2-D materials, such as transition metal dichalcogenides, exhibit strong excitonic responses that can be used for electro-refraction [106], though their integration and long-term stability remain active areas of investigation. For more on photonic switches with 2-D materials, see, for instance, Ref. 107.

Mechanical waves can also be exploited for modulation through the acousto-optic effect, in which surface acoustic waves induce strain and thus modulate the refractive index via the photoelastic effect [108]. Interdigital transducers on piezoelectric substrates generate the acoustic waves that couple to adjacent optical waveguides. Bandwidths up to a few GHz have been reported, though the need for piezoelectric layers and relatively long devices makes dense integration challenging.

Another route to refractive index tuning is based on liquid-crystal and opto-fluidic approaches. Liquid crystals exhibit large birefringence and can be reoriented with voltages of only a few volts, providing substantial modulation depth over short

lengths [109]. However, their switching frequencies are limited to the kilohertz regime by molecular reorientation dynamics, which restricts their use in high-speed networks [109]. Recently, Dutta *et al.* showed a liquid crystal ring resonator with an exceptionally low $V_\pi \cdot L$ of only $0.014\,\mathrm{V}\cdot\mathrm{cm}$ [109]. Liquid crystals remain promising candidates for tunable filters, slow-light structures, and low-power, reconfigurable optics, provided that packaging challenges can be addressed.

Electrochromic mechanisms based on ion intercalation into metal oxides (e.g., WO_3, MoO_3, TiO_2) can yield strong and persistent absorptive changes [23]. These effects can be non-volatile, but their switching speeds are typically limited by ion transport, leading to time constants on the order of milliseconds to seconds, and integration with silicon photonics is hindered by CMOS incompatibility and material stability issues [110]. Related concepts, such as optically addressable electronic memristors and filamentary devices, have also been explored for optical modulation [111]. However, the achievable index contrast is generally small, and the optical loss can be high [23].

Finally, all-optical switching based on nonlinear effects (e.g., Kerr nonlinearities and two-photon absorption) has been demonstrated across a variety of platforms [112]. These mechanisms can provide ultrafast response times [113], but generally require high optical intensities, incur significant non-linear losses, are challenging to scale, and lack non-volatility, limiting their relevance for large-scale, low-power integrated switch networks.

Overall, while these alternative mechanisms offer unique functionalities, they presently fall short of the balanced performance characteristics required for large-scale photonic switch fabrics, particularly in terms of speed, loss, footprint, and CMOS compatibility. Nonetheless, they remain promising for specialised applications, hybrid photonic–electronic systems, or future material platforms that may complement the more established mechanisms described earlier in this section.

2.3 Photonic switches with phase-change material

In this section, I will give a summary of PCM-based photonic switches and their application areas. I begin by examining the fundamental properties of PCMs, followed by a discussion of the characteristics of the novel low-loss materials Sb_2S_3 and Sb_2Se_3. I will also address the integration of PCMs into large-scale PICs, their use in extreme environments, and the emerging applications for which they are particularly promising.

2.3.1 Fundamental properties of phase-change materials

Chalcogenide-based PCMs can be reversibly switched between two stable solid-state phases—the crystalline and amorphous states—with distinct optical properties (see Fig. 2.7a). Notable examples of PCMs include GST (typically $Ge_2Sb_2Te_5$, but also includes other stochiometric compositions), (Ag,In)-doped Sb_2Te (AIST), germanium-antimony-selenium-telluride, typically $Ge_2Sb_2Se_4Te_1$ (GSST), Sb_2Se_3, and Sb_2S_3.

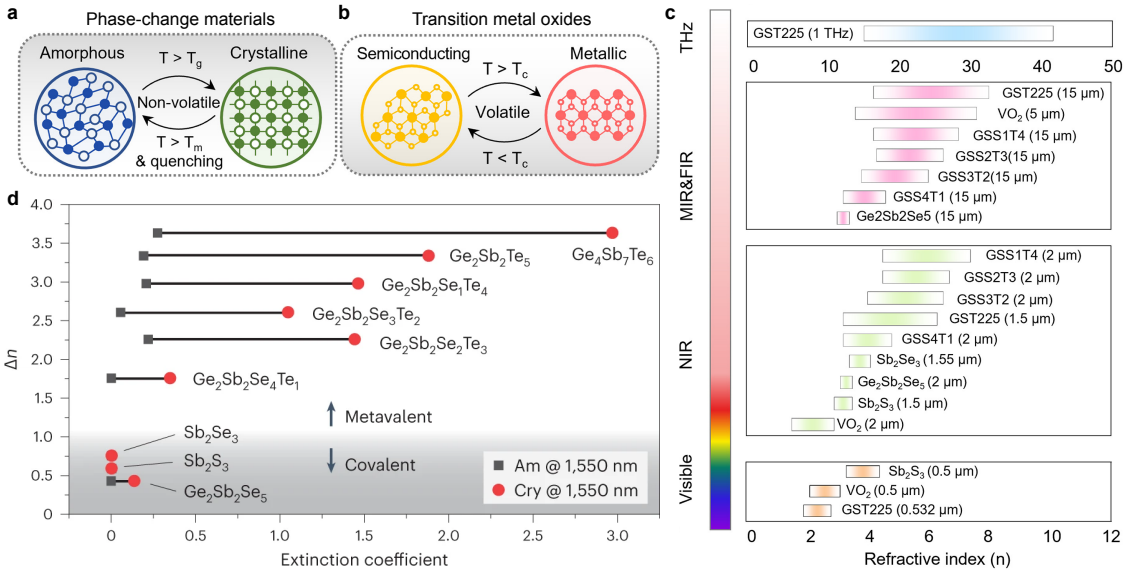

Figure 2.7: Refractive index comparison of phase-change materials. **a**,**b**, Switching behaviour of chalcogenide PCMs (a) and transition metal oxides, e.g., VO_2 (b). **c**, Overview over the refractive index contrast of different PCMs for different optical wavelengths. **d**, Overview of the refractive index change between the crystalline and amorphous states vs. the extinction coefficient of the two states for different PCMs. Figure adapted with permission from: **a**–**c**, Ref. 114, modified under a Creative Commons licence CC BY-NC-ND 4.0; **d**, Ref. 23, © 2023 Springer Nature Ltd.

One of the main benefits of PCMs is the small footprint stemming from the large refractive index contrast between the two phases of the PCM, which is caused by the distinct bonding nature and lattice structure in the material [115]. In the crystalline state, both the real and imaginary parts of the refractive index (often called the *refractive index* and the *extinction coefficient*, respectively; see Section 3.1.1) are larger than those of the amorphous state. In some PCMs, like the most studied one, GST, the change in absorption between the two phases at visible and near-infrared wavelengths is so significant that it is predominantly used as an absorptive element in photonic switches. In others, like Sb_2Se_3, the absorption in both states is almost negligible at near-infrared wavelengths, meaning the PCM can be used as a phase-shifting element (note that *phase shift* refers to the phase of the light passing through the PCM, not the phase of the PCM itself). Importantly, these *low-loss* or *transparent* PCMs have small extinction coefficients for both material states in the wavelength range of interest; in this work, low-loss PCMs refers to Sb_2Se_3 and Sb_2S_3 and similar materials with low absorption for both states in the telecommunications C-band (optical wavelengths around 1550 nm) [12]. The change in absorption and phase is exceptionally prominent in PCMs—it is, in fact, the optical material platform with the largest observed refractive index contrast [92], and thus allows for ultra-compact footprints and interaction lengths. For an overview of the refractive index of different PCMs, see Fig. 2.7.

The second main benefit of PCMs is non-volatility, which means that the material retains its state once it has been set [115]. In other words, you can *set and forget* the state of the system, which exists in this state without standby power consumption and electrical control [115]. This characteristic makes PCMs extremely interesting for long-term memory and other applications with infrequent switching. To switch between the states of the PCM, we normally heat the material [115], although other methods (e.g., threshold switching) also exist [116, 117]. Heating is normally introduced using integrated heaters (via *Joule heating*), or through optical pulses generated by external or integrated laser sources. For more on the switching mechanisms of PCM, see Section 3.6.

PCMs allow for relatively fast (up to $\sim$1 GHz electrically) [118] and reversible switching with high cyclability [119], although these metrics depend on the material and switching mechanism. The switching time can vary from below a nanosecond in GST [118] to milliseconds in Sb_2Se_3 [120] (depending on heater design). The cyclability, i.e., the number of switching cycles before degradation, depends on many factors. In optimised cases, researchers have experimentally demonstrated electrical switching of GST over 10^{12} cycles [121]—with an extrapolated estimate of 6.5×10^{15} cycles [122], and optical switching of Sb_2Se_3 over 10^6 cycles [123]. Although impressive, a finite number of cycles can be detrimental to device performance: as Burr *et al.* pointed out, even the best-case endurance of 10^{12} cycles, achieved for the lowest-energy pulses, is still six orders of magnitude from the target specification for static random-access memory (SRAM) memory [124].

2.3.2 Low-loss phase-change materials

A major limitation of conventional PCMs is the high optical absorption of the crystalline state in the near-infrared region, which results in significant insertion losses and prohibits their use as pure phase shifters. The more recently discovered[3] antimony chalcogenides, antimony trisulfide (Sb_2S_3) and antimony triselenide (Sb_2Se_3), have emerged as promising candidates for low-loss optical phase-change applications. Although the refractive index contrasts between the amorphous and crystalline states of Sb_2S_3 and Sb_2Se_3 are smaller than those of conventional PCMs, such as GST (see Fig. 2.7), they still exhibit relatively large values compared with other technologies, while maintaining extinction coefficients close to zero. This combination enables high modulation efficiency with significantly reduced propagation loss, opening up many new phase-based applications.

The transparent PCMs Sb_2S_3 and Sb_2Se_3 are obvious candidates for ultra-compact phase-shifting elements, because their wide band gaps shift the absorption edge of the extinction coefficient spectrum away from the near-infrared and towards visible wavelengths [12]. However, different research groups have argued for the

[3]By *recently discovered*, I refer to the successful realisation and integration of the optical PCMs in PIC, rather than to the discovery of the material *per se*.

use of one or the other based on different characteristics. First, Sb_2S_3 has a wider bandgap than Sb_2Se_3, with 1.7 eV to 2.0 eV [125] vs. $\sim$ 1.2 eV [126, 127]. The wider bandgap and consequently lower extinction coefficient of Sb_2S_3 mean that it should provide a lower loss when integrated in a hybrid waveguide. However, as Delaney *et al.* points out, the losses for Sb_2S_3 can surprisingly be higher than those for Sb_2Se_3, due to other effects such as grain boundaries or microscopic voids due to sulphur migration [12]. Another related effect is the stronger absorption of Sb_2Se_3 in the visible range, which is beneficial for optical switching with visible lasers [12]. However, the absorption of both states at visible wavelengths in Sb_2Se_3 limits its use for certain switching applications in the visible range (e.g., in DC switches for quantum applications [128]), whereas Sb_2S_3, with $\kappa_{\text{amo}} \approx 0$ down to around 600 nm, is better suited [128].

The refractive index contrast between the two phases of the PCM—the *strength* of the material [129]—is also an important parameter. The refractive index contrast of Sb_2Se_3 at 1550 nm is often quoted to be $\Delta n = 0.77$ [12, 120], while that of Sb_2S_3 is considerably lower, from 0.54 [130] to 0.6 [12]. This difference of around one third means that Sb_2Se_3 is more potent than Sb_2S_3. Furthermore, the real part of the refractive index of both phases of Sb_2Se_3 more closely matches that of silicon than do those of Sb_2S_3 (see Section 4.4.1) [12, 131], which leads to better mode matching in hybrid silicon-Sb_2Se_3 waveguides [12] that again reduces scattering loss.

An aspect that does not directly affect device performance but is equally important is how safely it is deposited. One advantage of using Sb_2S_3 is that it is less likely to contaminate the vacuum chamber during the deposition process, due to the absence of selenium [39, 132]. Furthermore, for the human conducting the deposition process, a low toxicity of the material is essential; according to Chen *et al.*, Sb_2Se_3 has low/no toxicity [133], while other works specifically on selenium toxicity disagree [134].

In Table 2.1, we present some of the relevant parameters for devices fabricated with Sb_2S_3 and Sb_2Se_3. Note that for the figure of merit (FOM) for transparent

PCMs, instead of the conventional FOM for PCMs, given as $\mathrm{FOM_{conv}} = \frac{\Delta n_{\mathrm{eff}}}{\kappa_{\mathrm{eff}}(\mathrm{cry})}$ (used, for instance, in Ref. 135), I use the definition

$$\mathrm{FOM_{transparent}} = \alpha L_\pi, \tag{2.1}$$

where α is the propagation loss in $\mathrm{dB\,\mu m^{-1}}$ for the most lossy state of the hybrid waveguide, typically the crystalline state, and L_π is the length required for a π phase shift. In contrast to the conventional FOM, this definition incorporates all causes of propagation loss and the strength of the PCM in one unit. Although my FOM is similar to the FOM proposed by Delaney *et al.* [12], which was given as

$$\mathrm{FOM_{Delaney}} = \frac{2\pi}{\lambda}\frac{\Delta n_{\mathrm{eff}}}{\alpha} = \frac{\pi}{\alpha L_\pi}, \tag{2.2}$$

I believe mine is more useful, as it represents the loss introduced by a π phase shift rather than a semi-arbitrary number.[4] In Table 2.1, L_π and α and are given for $\lambda = 1550\,\mathrm{nm}$ unless specified otherwise. As shown in the table, Sb_2Se_3 exhibits lower loss per π shift than Sb_2S_3 across most implementations. Furthermore, the highest number of cycles has also been achieved for Sb_2Se_3. However, as discussed in Section 6.6, most of these numbers are obtained for limited volumes of PCMs; therefore, it might be difficult to make a fair comparison between the number of switching cycles for the low-loss PCMs.

Even though several works over the last five years have explored low-loss PCMs, several challenges remain regarding device endurance under repeated cycling and the integration of materials with CMOS-compatible processes. I will discuss these issues in greater detail in Chapter 6. Continued work in material optimisation, commercial integration, and device design is expected to establish Sb_2S_3 and Sb_2Se_3 as central platforms for next-generation small-footprint, non-volatile photonic switches and modulators [114]. For more information on integrated photonic devices using Sb_2Se_3 specifically, a good summary is presented by Chen and co-authors in their review in Ref. 136.

[4] Although, I must admit, a good figure of merit is one that increases with improved device performance, which mine does not. Instead of using the FOM, I encourage the reader to instead use the π-loss, αL_π, on its own.

Table 2.1: Comparison of different hybrid waveguides with transparent phase-change material reported in literature. Dep. = deposition technique, Cov. = coverage of waveguide, α_{cry} = propagation loss for the crystalline state of the PCM, Sp: Sputtering, TE: Thermal evaporation. If no comment on the deposition coverage is made, the PCM is on top of both the waveguide and the slab; S: assumed also on the sidewalls; T: PCM only on top of the waveguide.

	t (nm)	Dep.	Cov.	L_π (µm)	α_{cry} (dB/µm)	FOM (dB)	Cycles	Ref. (year)
Sb_2S_3	22	Sp	—	21.8	0.028	0.61*	∼1000†	[12] (2020)
	20	Sp	T	38.8	0.023	0.89	—	[137] (2021)
	60	Sp	S	7.8	0.322	2.5	—	[135] (2021)
	20	Sp	T	64.6	0.16	10	Failed	[130] (2021)
	20	Sp	T	30.7‡	0.024	0.74	1600	[39] (2023)
	70	Sp	—	25	0.01	0.25	750	[132] (2025)
Sb_2Se_3	22	Sp	—	10.8	0.01	0.11*	4000†	[12] (2020)
	20	Sp	T	15.5	0.06§	0.93	—	[137] (2021)
	30	TE	—	11	0.035¶	0.39#	125	[120] (2022)
	30	Sp	—	14.6	0.019	0.28	—	[138] (2022)
	30	TE	—	9.3	0.036	0.33	—	[138] (2022)
	30	Sp	T	704.5	0.006	4.2	1000	[139] (2022)
	30	Sp	T	14.6	0.02	0.29	3000**	[131] (2023)
	30	Sp	—	16.1	0.021	0.33	—	This work
	60	Sp	—	7.6	0.105	0.80	—	This work

* In the paper by Delaney *et al.*, Δn_{eff} is not shown experimentally, so I assume it is estimated from simulations. Furthermore, Δn_{eff} assumes 25 nm Sb_2Se_3, but the loss measurements were done with 22 nm Sb_2Se_3 [12].

† Switching cycles reported for thin-film reflection measurement (not waveguide).

‡ The Δn_{eff} values are reported for $\lambda = 1.34\,\mu\text{m}$.

§ In the paper, the authors claim 0.07, which I believe is artificially high due to an outlier. A more correct value is likely 0.06.

¶ The authors report $\frac{0.21\,\text{dB}}{6\,\mu\text{m}} = 0.035\,\text{dB}\,\mu\text{m}^{-1}$ in the Supplementary Material, but subtract scattering losses to get $0.018\,\text{dB}\,\mu\text{m}^{-1}$ in the main text.

The authors claim that for longer cells, the change in n_{eff} was slightly smaller than the 0.07 reported (estimated to be 0.0626), giving $L_\pi = 12.5\,\mu\text{m}$ and FOM = 0.44.

** The authors also show 10^4 switching cycles for an array of Sb_2Se_3 patches, but not the overall loss from the array.

2.3.3 Integration of phase-change switches

For the large-scale integration of PCM-based components into PICs, it is crucial that we can switch the PCM efficiently (the theory behind the switching mechanism of PCMs will be covered in Section 3.6). Previous implementations have mainly used external (free-space) laser switching, integrated optical switching, or integrated electro-thermal heating. The free-space version is interesting for lab-based devices (in particular the application in Section 2.4.3), but is not feasible for miniaturised and large-scale integration. Integrated optical switching has been used for the switching of memory elements [4, 99], but lacks potential for scaling and switching of large volumes of PCM (needed, for instance, for most low-loss PCMs). Furthermore, both for free-space and integrated optical switching of transparent PCMs, the wavelength of the switching laser must be different from the wavelength of the probe laser (i.e., be in the absorptive region of the spectrum). Therefore, electrical switching with integrated heaters (made from, e.g., doped silicon, metals, TCOs like ITO, or graphene [140]) provides a more attractive solution that overcomes the problems of optical switching [139]. Nevertheless, electro-thermal switching with integrated heaters typically requires higher energy per switching event than optical switching, with increased insertion loss and switching voltages routinely exceeding the CMOS-compatible 5 V constraint.

2.3.4 Phase-change photonics in extreme environments

PCM-based photonic computing holds promise for use in extreme environments. Photonic devices are inherently immune to many of the EMI-related issues that their electronic counterparts are prone to [141], meaning optical signals in waveguides are normally not impacted by external electromagnetic pulses. The same is true for PCM memories. By substituting exposed electronic components with PICs, for example, replacing the processor core with a PTC, we can significantly reduce the total area of electronic components on our microchips. Even though the electronic peripherals needed for the PICs are still exposed, the reduced number of electronic components means that bit flips and other faults occur at a lower rate. Furthermore,

we can make sure these electrical components are more immune to EMI, but because they are in smaller numbers, the total cost of the chip will be more tolerable.

Furthermore, there is a growing body of research on the effects of high-energy particles on PCMs and PICs [142–145]. Because PCM memories store data without relying on electrons, they are not expected to exhibit the same radiation sensitivity as floating-gate flash memories [146], and are therefore promising candidates for memories in space. However, radiation effects can change the complex refractive index of the photonic circuits [145], thereby affecting performance. Nevertheless, in the review by Terrasanta *et al.*, the authors discuss techniques to mitigate such effects, for instance, by employing a thick SiO_2 cladding to prevent a change in n_{eff} and using thick, high-confinement Si waveguides to reduce losses from radiation-induced defects in the waveguide. Moreover, in cases where the effective refractive index of a waveguide in, for instance, an MZI changes with radiation, we can counteract this change with the phase shifters, as long as we have time to calibrate the MZI occasionally.

Finally, if the extreme environments include high temperatures, data stored in PCM memories might be lost. The main PCM, GST, has retention times over 10 years at 85 °C, but will completely crystallise (and therefore lose its information) in a few minutes as the temperature reaches the glass-transition temperature. However, as shown in Ref. 147, with doping of the GST, the memory is retained even at temperatures as high as 180 °C [147].

2.3.5 Application areas of phase-change switches

As previously noted in this chapter, photonic switches are integral components in almost all PICs. PCM-based switches typically have compact designs and are therefore especially useful for applications with strict footprint requirements. Furthermore, the combination of non-volatility with limited cyclability and relatively slow, energy-intensive switching makes PCMs more applicable to certain applications, in particular those with infrequent switching. Examples include interconnects, photonic recirculating mesh networks, memory devices, and specific computing

hardware. There are also numerous other interesting applications where PCMs excel, such as metasurfaces and reflective displays [96, 97], but these fall outside the scope of this work. I will cover more about computing applications with PCMs in Section 2.4.4.

Among the areas where PCM-based integrated switches might surpass the state of the art, we propose those where slower, multi-level tuning over extended time periods and/or with reduced footprint is paramount. PCMs in integrated systems require relatively large energies per switching event and long switching times, especially when going to the crystalline state. Ultrafast (e.g., GHz) switching applications are therefore out of the question for the currently available PCMs in integrated optoelectronic systems due to their slow response.

If the switching events occur at slower rates, e.g., kHz rates or less, then the non-volatility of PCMs can offer considerable energy savings compared with volatile switching mechanisms, despite the large switching energy [120]. For even more infrequent switching events (e.g., once per second or less), PCM-based switches can enable applications where minimal power consumption is imperative, such as in edge and IoT devices. Due to the long retention time of PCMs (at least for the most studied ones, such as GST [124]), PCM-based switches can keep their output state for extended periods of time at no energy cost. Finally, PCM-based devices have been exploited for their multi-level tunability [148], which is also an asset for many switching applications.

I will discuss more about specific application areas of PCM-based switches in the following sections.

2.4 Photonic computing and signal processing

Photonic computing and signal processing involve using photons (i.e., light signals) instead of electrons (i.e., electrical signals) to process and transmit information in a system. Photonic platforms in general hold promise for high-speed, broadband signal transfer and processing, offering faster, wider-bandwidth, and lower-energy

systems—this is why essentially all long-haul communications now depend on fibre-optic cables. Not only do photons in a waveguide travel faster than electrons in a metal wire, but they are also not susceptible to the capacitance effects found in metal wires, which lead to loss and reduce the quality of electrical signals; as a result, light signals can propagate through the circuit and be modulated significantly faster than electrical signals [149]. Due to the various multiplexing capabilities (wavelength, mode, polarisation, etc.), light signals can carry vast amounts of data through the same circuit, providing wide bandwidth. Moreover, multiplexing offers inherent parallelism, which is vital for many systems, such as those used for machine learning and processing large data sets.

This section outlines key application areas of photonic switches reported in recent years. These include photonic neural networks, integrated optical signal processors, switch fabrics, and quantum computers. For some applications—such as switch fabrics—we simply need to couple light between ports. For others, we may require *tunable* modulators or couplers. And for yet another class of applications, including forward-only meshes for matrix–vector multiplication (MVM) or recirculating meshes for FPCAs, we need tunable 2×2 unitary gates capable of modulating both the coupling and the output phase.

2.4.1 Integrated optical signal processors

As the need for wider bandwidth and reconfigurability increases in communications, integrated photonics offers a promising platform not only for signal transfer but also for signal processing. Optical signal processors (OSPs) exploit the intrinsic characteristics of photonic systems to manipulate information directly in the optical domain. The inherently low loss and SWaP of integrated photonics, in addition to the wide bandwidth tunability, make the use of OSPs much more favourable for next-generation systems compared with their electronic counterparts [150]. Especially for systems where the signals are already in or have been converted to the optical domain for signal transfer, processing in OSPs reduces the need for energy-intensive EO conversions. These capabilities make optical signal processing a compelling

approach for high-capacity communication networks and real-time data analysis. OSP chips can perform operations such as filtering, wavelength conversion, and logic functions at speeds beyond those achievable with conventional electronics. Furthermore, they cover a wide range of application areas [151], ranging from optical transceivers in tele- and datacomms [152] to reconfigurable optical add-drop multiplexers [153], optical phased arrays for light detection and ranging (LiDAR) in autonomous systems [154] to quantum information processing circuits [75]. Photonic switches will be crucial for the proliferation of integrated OSPs, not only for routing and tunability, but even more so in reconfigurable OSPs (Section 2.4.3) [151].

Integrated OSPs are increasingly relevant in sensing and defence-oriented applications. Optical radar systems, for example, benefit from the high bandwidth, low latency, low-loss transmission, EMI immunity, and precise phase control offered by photonic integration. Integrated optical front-ends enable rapid beam steering, high-resolution ranging, and robust detection in compact, low-power platforms with resolution and speed potentially unattainable with electronics [141]. The ability to process return signals directly in the optical domain—such as through photonic temporal filtering, pulse shaping, matched filtering, or chirp generation—reduces electrical bottlenecks and improves dynamic range, making photonic processors well-suited for next-generation optical radar architectures [141].

Similarly, microwave and radio frequency (RF) signal processing stands to benefit considerably from photonic implementations [155]. Integrated photonic processors can handle signals with bandwidths far exceeding the limits of electronic RF front-ends, enabling functions such as frequency-to-time mapping, beam-forming, channelisation, and convolution at rates of tens of GHz [29]. These capabilities are attractive for spectrum sensing, electronic warfare, and advanced radar systems, where rapid, reconfigurable, and wideband processing is critical [156]. The combination of wide optical bandwidth, low-loss delay elements, and programmable linear transformations positions integrated photonics as a strong candidate for future defence-relevant signal-processing hardware.

2.4.2 Photonic mesh networks

Meshes of photonic switches find use in many areas, ranging from switch fabrics to FPCAs and photonic machine learning accelerators. In this section, I will look at some of these technologies and the requirements of the individual switches for these applications. Because of the major focus on FPCAs and photonic machine learning accelerators in this thesis, I will cover these technologies separately in the following chapters.

Switch fabrics and interconnects

In data centres and high-performance computing networks, switches constitute one of the main building blocks [19]. The demand for signal bandwidth in data centres is growing, and therefore the most important metric for optical switches in data centres is optical bandwidth [19]. Although one could think that switching speed would be another factor of paramount importance, Cheng *et al.* argues that reducing the switching time much below 100 ns just introduces new problems; the optical links in data centres are extremely sensitive to phase variations at this short pulse duration, so a link training session of maybe ten to hundred nanoseconds or more would be required to understand the new link conditions [19]. Switching times below the microsecond mark should therefore be sufficient for data centre and supercomputer applications [19].

Switch fabrics for data centre applications require highly uniform and reliable switches [157]. Among other requirements, low-loss and high-extinction-ratio operation is desired, with low power consumption and cost. Fabrics with, for instance, 32 ports or more are necessary [157], opening the door to switches with relatively high per-component loss but low overall loss, such as in some MEMS-based systems. Since the switch size is not critical for this application, as long as the footprint is reasonably small, it is not worth compromising other metrics to reduce it further. Finally, multi-level tunability is unnecessary, as the switches are used only for routing. This relaxed requirement makes it feasible to employ bistable MEMS switches, which are non-volatile in just two states, leading to reduced power

consumption [84]. Different switch fabric topologies and notable large-scale research demonstrations based on 2×2 switches have been presented in other reviews—see Refs. 20 and 21. Finally, it is worth noting that for certain routing configurations, such as splitter trees, switches with only one input port are sufficient.

In addition to routing, quantum applications (see the next section), and feed-forward neural networks (see Section 2.4.4), switch fabrics also have other applications. One example is the on-chip unscrambling of optical modes [158], where a mesh of phase shifters and MZI switches with feedback control from integrated probes was used to separate different optical modes.

Quantum applications

Photonic switch fabrics are also the workhorses behind some of the most advanced quantum systems [58, 159]. Using 2×2 unitary gates composed of MZI switches and phase shifters in tandem, Carolan *et al.* demonstrated some of the earliest fully reconfigurable quantum optical circuits on an integrated platform, implementing a variety of programmable quantum gates for single-photon systems [160, 161]. Their work established the feasibility of large-scale unitary transformations on-chip, paving the way for more complex photonic quantum processors [162]. Building on these foundations, both PsiQuantum and Xanadu have recently reported the development of scalable photonic quantum-computing architectures implemented on integrated photonics platforms [58, 159]. These efforts highlight the growing maturity of photonic integration for quantum information processing and point toward practical fault-tolerant systems based on silicon-compatible technologies.

In parallel, the emerging field of quantum neuromorphic photonics has demonstrated device concepts inspired by classical memristors. Notably, a quantum analogue of the memristor has been realised using an MZI-based architecture with a feedback loop to its phase shifter, enabling history-dependent transformations on single-photon states [163]. Photonic switches can also be used for temporal-to-spatial demultiplexing of single-photon sources [164, 165]. The switching speed required in this case depends on the repetition rate of the single-photon source.

Such developments illustrate the versatility of reconfigurable interferometric meshes, not only for universal linear optics but also for unconventional quantum information paradigms. In all quantum photonics applications, low-loss switches are crucial, and multi-level tunability is required, for instance, to set quantum gates or create superpositions of single photons.

2.4.3 Photonic recirculating mesh networks

Photonic recirculating mesh networks are a type of programmable PICs, i.e., software-controllable reconfigurable circuits. Unlike conventional application-specific photonic integrated circuits (ASPICs), which can offer some tunability but are designed for a specific purpose [75], photonic recirculating mesh networks are truly general-purpose circuits with no specific application. Analogous to the field-programmable gate array (FPGA) in electronics, a generalised photonic version of a photonic recirculating mesh network can be called a field-programmable coupler array (FPCA). Other names that have been used in the literature include programmable photonic (signal) processors [166–168] or arrays [18], multi-/general-purpose programmable photonic processor [169, 170], and field-programmable photonic (gate) array [171, 172]. Due to their universal nature, FPCAs can enable rapid prototyping (lowering development costs and time-to-market) and allow for upgradability of employed devices and improved device lifetime (through reconfigurable back-up redundancy).

Although reconfigurable OSPs exist [173], FPCAs are even more universal platforms. An FPCA has no specific application, but can be used to implement photonic components such as ring resonators, MZIs, delay lines, etc., which again can make up more complex structures (demultiplexers, differentiators, Hilbert transformers, etc.) that find use in optical signal processing applications [75, 174]. The core of the FPCA is a mesh of 2×2 switches arranged in a recirculating manner (meaning that light can be recirculated back to a port it has already seen, unlike forward-only meshes). Each switch is a unitary gate, i.e., it can independently control the coupling and phase of the light, allowing for both phase and amplitude modulation, path selection, beam splitting, and tunable coupling [166].

It is this granular control of the light path that allows us to synthesise entire PICs on this modular platform.

In the original demonstration of the FPCA in 2015, Zhuang *et al.* used long MZI switches as the building blocks that formed a 2-D square mesh (Fig. 2.8a–d) [166]. Pérez *et al.* built on this work [169], demonstrating millimetre-long MZIs in a larger hexagonal mesh (Fig. 2.8e–h), while arguing that the hexagonal configuration is superior to the triangular and square versions [167]. The same authors investigated TO DC switches on an SiN platform a few years later (Fig. 2.9d), showing a lower loss per building block [41]. They also investigated the use of folded heaters on an SiN platform for reduced footprint and power consumption [168]. Finally, in 2024, the authors demonstrated a system with an improved version of their mesh consisting of 72 tunable couplers on an SOI platform [170], built by their spin-out company, iPronics (see Fig. 2.9c,e). The length and loss per tunable coupler are 811 µm and 0.48 dB, respectively, with an average power consumption per π phase shift of 1.3 mW.

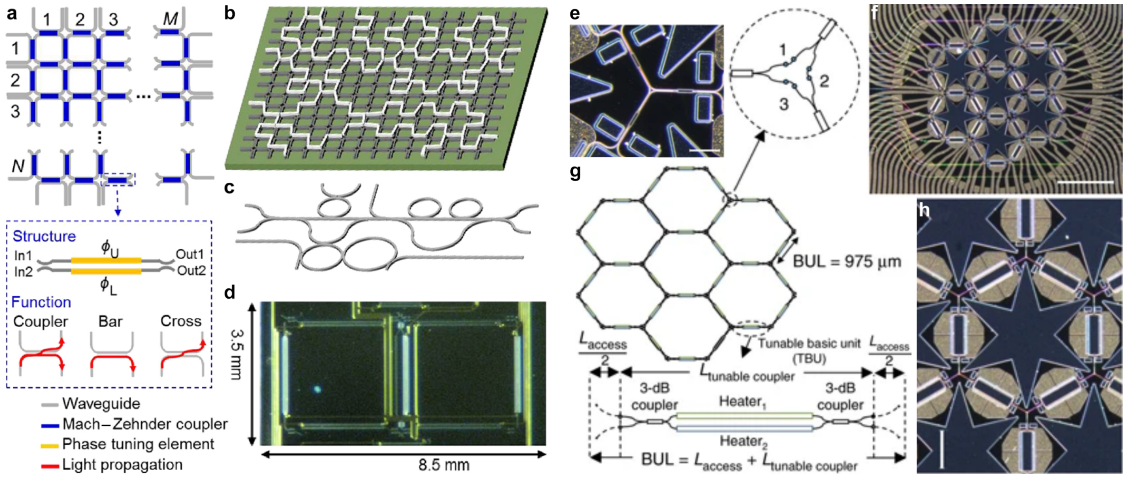

Figure 2.8: Initial demonstrations of field-programmable coupler arrays. **a**, Schematic of a square FPCA mesh consisting of 2×2 MZI couplers. **b**,**c**, Schematics of a synthesised PIC in an FPCA (b) and the corresponding ASPIC (c). **d**, Optical micrograph of two square unit cells with TO MZI couplers. **e**, Optical micrograph and a schematic of an interconnection node of the hexagonal mesh in **f**. **f**, Optical micrograph of a hexagonal FPCA with seven unit cells consisting of MZIs with TO phase shifters. **g**, Schematic of the structure in **f**, with the inset showing a schematic of a tunable coupler. **h**, Optical micrograph of the centre hexagon of the FPCA in **f**, rotated 90°. Scale bars are **e**: 100 µm; **f**: 2 mm; **h**: 500 µm. Figure reproduced with permission from: **a**–**d**, Ref. 166, © 2015 Optica; **e**–**h**, Ref. 169, modified under a Creative Commons licence CC BY 4.0.

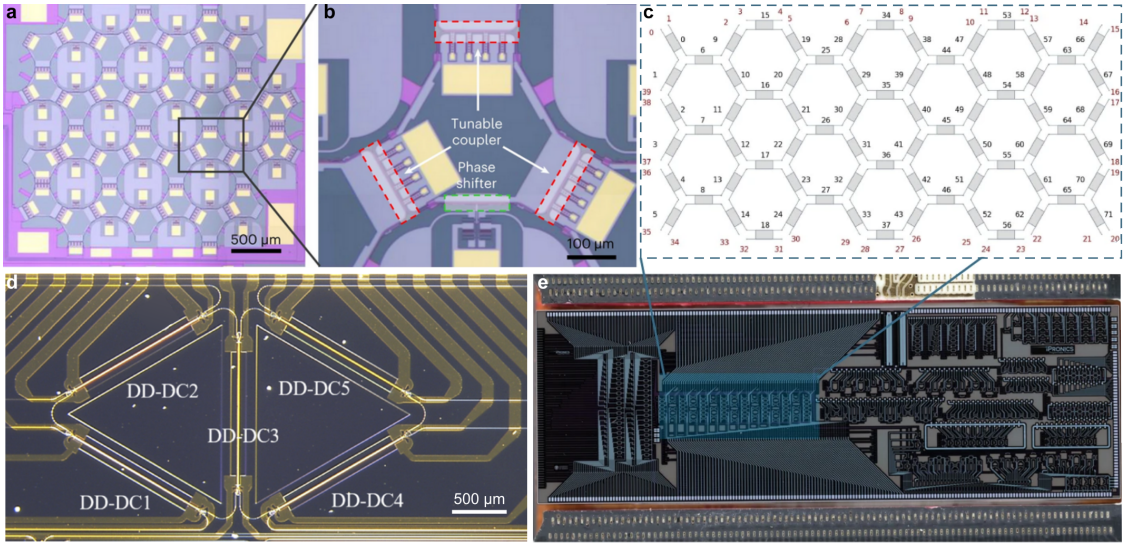

Figure 2.9: Recent demonstrations of field-programmable coupler arrays. **a**, Optical micrograph of a triangular FPCA mesh consisting of 2×2 MEMS couplers and phase shifters. **b**, Optical micrograph of one unit cell from **a**. **c**, Schematic of the hexagonal FPCA mesh demonstrated in **e**. **d**, Optical micrograph of two triangular unit cells with the SiN DC switch shown in Fig. 2.4d. **e**, Picture of a chip with the recirculating mesh region highlighted. Figure reproduced with permission from: **a**,**b**, Ref. 18, © 2023 Springer Nature Ltd.; **c**,**e**, Ref. 170, modified under a Creative Commons licence CC BY 4.0; **d**, Ref. 41, © 2019 Optica.

A promising alternative to TO couplers was demonstrated by Kim *et al.*, shown in Fig. 2.9a,b [18]. Each 2×2 unitary gate consists of the MEMS coupler and phase shifter presented in Fig. 2.3d-i, with compact footprints of 150 µm and 100 µm, respectively. The loss of these components was between 0.12 dB and 0.13 dB for the phase shifter and 0.18 dB and 0.43 dB for the coupler in its extreme states, with switching times between 13.6 µs and 20.3 µs. The extinction ratio can be very large (between 32 dB to 49 dB was demonstrated) because the MEMS coupler physically displaces the two coupling waveguides to prohibit evanescent coupling. Finally, the maximum total switching energy and static power consumption per unitary gate were below 43.5 pJ and 16.6 fW, respectively, due to the nature of the capacitive electrostatic actuators. However, to operate the coupler and phase shifter, a constant voltage of 7.5 V to 10.6 V must be applied at all times.

For synthesised photonic components in FPCA meshes, for example, ring resonators, each switch element of the mesh should ideally be of short length, thereby enabling a large free spectral range (FSR; the frequency spacing between two

resonances) [75]. Furthermore, switches in FPCAs should allow multi-level tunability and low-loss switching, with low overall power consumption [75]. Broadband operation could be an advantage, and, depending on the application, so is high-speed switching.

Field-programmable ring/disk arrays

Alternatives to the relatively large MZI couplers have also been proposed. In 2020, Zhang and Yao showed a field-programmable disk array (FPDA) with a disk radius of 6.4 µm and a unit cell footprint of only 70 µm × 70 µm (see Fig. 2.10a,b,d) [174]. The power consumption was on the order of a few milliwatts, much lower than that in the initial works, due to the lower power required for resonant structures. Sun *et al.* showed a field-programmable ring array (FPRA) with an asymmetric (i.e., unbalanced) MZI-assisted MRR structure (Fig. 2.10c,e,f) [175]. The unit cell area was 85 µm × 42 µm, and the power consumption was again on the order of a few milliwatts. Note that the demonstrated FPRA had phase shifters also on the asymmetric MZI arms (Fig. 2.10). The advantages of using ring- or disk-based structures lie mainly in the reduced SWaP. However, one major limitation is that resonant structures are inherently narrowband. A narrow linewidth can be great for many applications (e.g., filtering), but it makes these systems less widely applicable than FPCAs: for instance, they cannot synthesise certain structures, such as MZIs, which means they are not truly general-purpose platforms.

Freeform rewritable photonic integrated circuits

The coupler arrays presented above have a natural limitation in the fixed length of the couplers—an arbitrarily small path length cannot be achieved. An alternative approach solves this problem by employing a laser-rewritable film of low-loss PCM stacked with a layer of SiN (either above or below the PCM) [176, 177]. When some of the PCM is in the crystalline state, light injected into the SiN-PCM stack from a different laser will propagate only in the crystalline region of the stack, effectively acting as a waveguide [176]. Any passive photonic component can be written using the free-space laser, even fine structures such as grating couplers for light coupling.

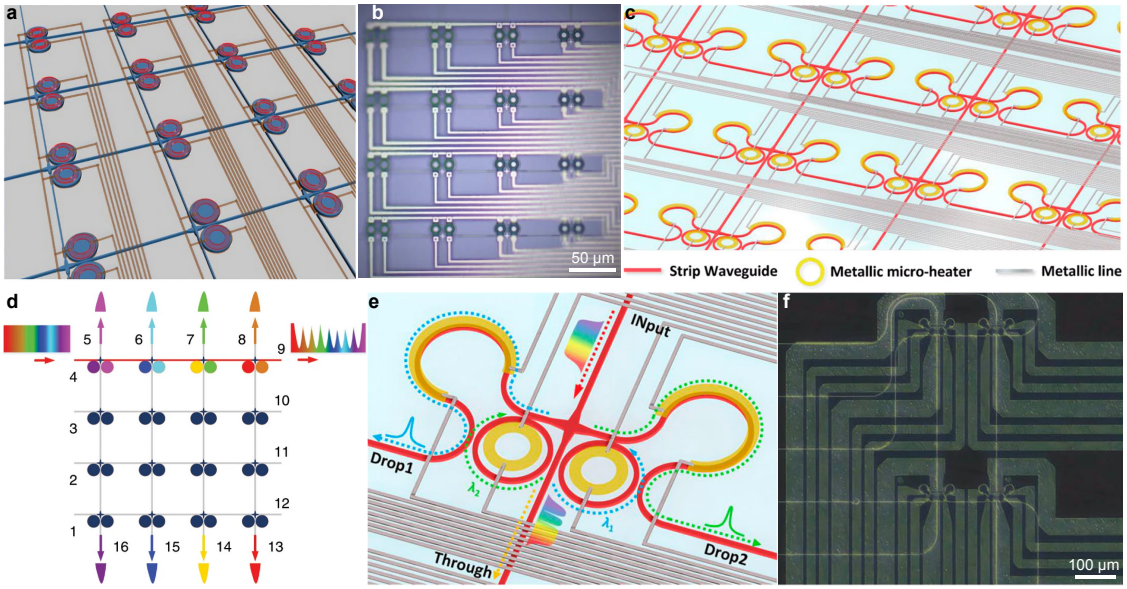

Figure 2.10: Demonstrations of field-programmable ring/disk arrays. **a**, Schematic of an FPDA. **b**, Optical micrograph of the demonstrated FPDA with 32 ring resonators and 16 ports. **c**, Schematic of an asymmetric MZI-assisted FPRA. **d**, Schematic of the working principle behind the FPDA as a wavelength demultiplexer. **e**, Working principle of the asymmetric MZI-assisted FPRA in **c**. **f**, Optical micrograph of the demonstrated FPRA with 8 rings and 8 ports. Figure reproduced with permission from: **a**,**b**,**d**, Ref. 174, modified under a Creative Commons licence CC BY 4.0; **c**,**e**,**f**, Ref. 175, modified under a Creative Commons licence CC BY-NC 4.0.

Furthermore, because PCMs are switchable, the circuits can be erased and rewritten. The composed waveguides exhibit a very low propagation loss, despite the fact that the light is propagating partially in the crystalline state of the PCM, making it a very promising and suitable tool for prototyping of PICs. One of the main disadvantages of this technology is its reliance on a free-space laser, which limits its integration and means it is not a reconfigurable photonic *integrated* circuit *per se*.

2.4.4 Photonic neural networks

Photonic neural networks harness the unique advantages of light, such as ultrafast modulation, high bandwidth, and low loss, to perform neural network computations with improved speed and energy efficiency compared with traditional electronic hardware. A variety of architectures have been developed to implement these systems, each exploiting different optical mechanisms to realise core operations. Most notable implementations of photonic computing hardware to date have been

neuromorphic—that is, computing architectures inspired by the brain—which in practice perform computations using artificial neurons realised directly in hardware. Among the most prominent photonic neural network architectures are photonic crossbar arrays [4], MZI meshes [11], MRR weight banks [178], diffractive optical neural networks (DONNs) [179], and MMI networks [180]. In addition to the more 'conventional' neural network models typically implemented with the aforementioned architectures, there have also been photonic implementations of spiking neural networks [181], reservoir computing [182], and Ising machines [183]. Together, these platforms illustrate the diverse strategies by which optics can accelerate and expand the capabilities of artificial intelligence hardware. For more information about photonic implementations of the various architectures and computing frameworks, see Ref. 184.

Photonic crossbar array

In this thesis, the emphasis will be on photonic crossbar arrays, with examples shown in Fig. 2.11. The photonic crossbar array, whose name and layout are inspired by the electronic crossbar array [185], is a type of photonic tensor core (PTC) that performs MVM. In the original demonstration of the photonic crossbar array, the vector is encoded as light signals entering rows of waveguides that pass through a matrix of non-volatile amplitude modulators (patches of GST) encoding the matrix weights (Fig. 2.11a). After the light has been weighted, it combines in a bus waveguide via DCs in each column [4]. One of the main benefits of the original design is that we can employ WDM schemes to process multiple vectors simultaneously. For more on the working principle of the photonic crossbar array, see Section 3.7.2. The photonic crossbar array is often used as the hardware accelerator for the convolutional layer of a convolutional neural network (CNN). However, more generally, it can be used in any system where MVMs are needed [4].

Since the seminal work, different alterations to the original architecture have been proposed. Aggarwal *et al.* introduced photonic crossbar arrays with reduced rank of operation (Fig. 2.11b) for improved scalability and reduced complexity of

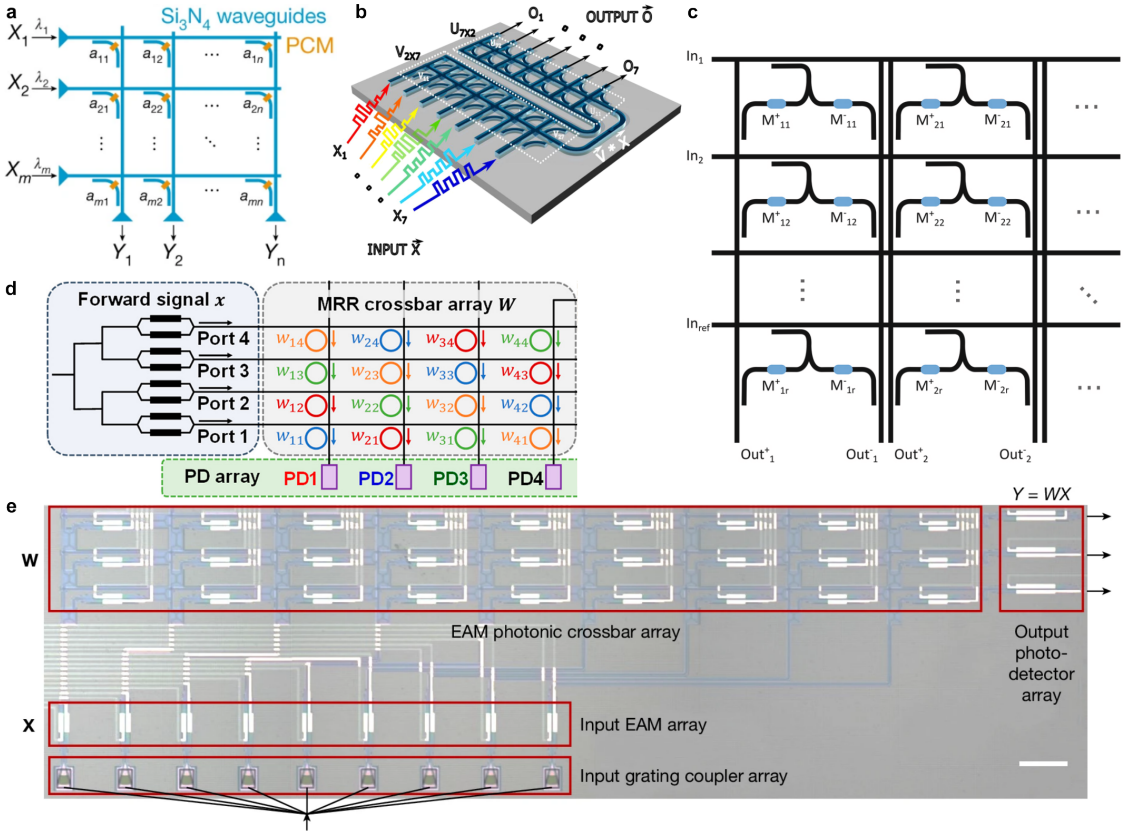

Figure 2.11: Demonstrations of photonic crossbar arrays. **a**, Schematic of the initial demonstration by Feldmann *et al.* [4]. **b**, Reduced rank photonic crossbar array. **c**, Photonic crossbar array with reference arms and balanced photodetection for encoding of negative values. **d**, Implementation with rings for coupling and modulation. **e**, Implementation with EAMs for amplitude modulation. Figure reproduced with permission from: **a**, Ref. 4, © 2021 Springer Nature Ltd.; **b**,**c**,**e**, Refs. 186, 187, and 188, respectively, modified under a Creative Commons licence CC BY 4.0; **d**, Ref. 189, modified under a Creative Commons licence CC BY-NC-ND 4.0.

the system, in which two smaller matrices replace the original matrix and some weight information is omitted without significant sacrifice of accuracy [186]. For some matrices, especially those used as kernels in convolution layers, there can even be no loss of information and therefore no loss of accuracy with these reduced-rank architectures. Another drawback of the original design is that we need two computations to encode negative weight values (as discussed in Section 3.7.2). Brückerhoff-Plückelmann *et al.* solved this issue by including a reference arm in each matrix unit cell and used balanced photodetectors at the end of each column to enable read-out of both positive and negative values (Fig. 2.11c) [187]. Dong *et al.* showed that we can add an additional dimension to the data (alongside

spatial memories and wavelength multiplexing) with time-domain multiplexing [188], opening the door to even more powerful systems. The same authors also showed that we can enable larger-size PTCs by employing partially coherent light sources at the input, and introduced ultrafast EAMs as weights (Fig. 2.11e) [190]. Finally, the use of microrings as combined amplitude modulators and couplers to the bus waveguide has also been shown (Fig. 2.11d) [189, 191]. For the photonic crossbar arrays in Chapters 5 and 7, we use two different amplitude modulators as weights; in Chapter 5, we use patches of GST, and in Chapter 7, we use RAMZIs with a characteristic output curve.

Performance metrics of photonic neural networks

Some of the more interesting performance metrics of photonic neural networks include latency, throughput, power efficiency, and compute density. The latency includes only the time of flight (ToF) of the light signals through the optical circuit. Among photonic accelerator architectures, photonic crossbar arrays exhibit the lowest latency compared with MZI meshes and MRR arrays [184]. Throughput is defined as the number of operations performed across the entire matrix per unit time and is commonly expressed in tera operations per second (TOPS). In a photonic crossbar array, each unit cell performs a multiply–accumulate (MAC) operation, with the weighting and summation stages counted as two separate operations. The overall throughput is typically calculated as the product of the number of operations per cycle and the update frequency, which is determined by the optical latency and the bandwidth of the optoelectronic components (e.g., modulators and photodetectors). Finally, power efficiency and compute density are expressed as the number of operations per unit of power and per unit of area, respectively. For these calculations, we should also include the power and area of the peripherals, such as analogue-to-digital converters (ADCs).

The power consumption of photonic neural networks is a debated topic, in part because the energy requirements of photonics are often presented as vanishingly small. As mentioned before, the signal loss is minuscule compared with electronic systems

thanks to the low-loss transmission of photonics, meaning the energy requirements for both signal transmission and heat extraction to prevent chip overheating could be significantly reduced compared with electronic solutions. Furthermore, the energy requirements of some photonic switches, e.g., phase-change memories, can be negligible in the grand scheme of things. In practice, it is the peripheral components, such as lasers and ADCs, that consume considerable power [4, 7], and these components also significantly limit the compute density of photonic neural networks. For the photonics researcher working on ultra-compact, low-power computing cores, it can therefore be tempting to omit the peripherals in calculations [4].

The energy consumption of photonic computing architectures can be broken down into five categories, as seen in Fig. 6 of Farmakidis *et al.*; optical input, transmitter, weighting, receiver, and analogue-to-digital conversion [184]. First, the optical input typically comes from a laser. Although light sources can be made relatively efficient, they still consume a considerable share of the total power. Importantly, a higher component loss in the circuit requires us to compensate by increasing the optical input power. The transmitter, for example, an electro-optic modulator, is the electro-optical device that modulates the optical input, while receivers, typically photodetectors with transimpedance amplifiers (TIAs), convert the light output to an analogue voltage signal that the ADC digitalises. These devices also contribute significantly to the overall power consumption. Nevertheless, it is likely that optical power, rather than the power of the transmitter, receiver, or ADC, becomes the limiting factor for power efficiency as we scale to larger computing architectures [184]. Therefore, in these cases, low-loss components and architectures are crucial for overall efficiency and system scalability.

Additionally, we must include the power required to maintain the weights of a photonic neural network, i.e., the power of weighting. In Fig. 6 of Ref. 184, weighting is below 1×10^{-6} W/TOPS, which assumes non-volatile or ultra-low-power devices. Instead, if the weights are volatile TO phase-shifters [192], the power consumption can be as high as 0.1 W/TOPS, which is an order of magnitude larger than the power consumption of the other components in Fig. 6 of Ref. 184, and

detrimental to the overall power consumption. Zero- or low-power weights, such as PCM elements, are therefore crucial for power-efficient photonic neuromorphic computing architectures. Additionally, when using non-volatile weights, the photonic accelerator is an example of an in-memory computing platform, which offers significant advantages over traditional von Neumann architectures, particularly in reducing data-transfer latency [185].

Another critical performance consideration in neural network accelerators is the precision. The effective bit-depth is a measure of the real dynamic range of devices such as ADCs. It is limited by noise and signal degradation, for instance, due to optical loss through the network and modulators with finite extinction ratios. Even in the most advanced photonic computers to date, the effective number of bits is limited to below 10 bits [7, 8]. However, many artificial intelligence inference algorithms are designed to operate with fixed-point precisions below 8 bits, making them well suited to the constraints of optical computing [193, 194].

Photonic computing with phase-change materials

The goal of this section is to provide insight into areas where photonic neuromorphic computing architectures based on phase-change memories excel. The two main reasons why PCM-based memories and computing architectures are particularly interesting are the non-volatility and small footprint. Non-volatility means that the material and any computing primitives and architectures based on it, e.g., the weights in a neural network, remain in their state without external stimuli. From a power consumption perspective, this inherent behaviour of PCMs means that the energy used depends only on the energy required to change the state of the PCM (i.e., the switching energy between the crystalline and amorphous states), which, in the big picture, becomes diminishingly small if the switching events occur rarely. However, it should be mentioned that the switching energy per event can be relatively high, e.g., 9 nJ for GST in Zhou *et al.* [195] and 38 µJ for Sb_2Se_3 in Ríos *et al.* [120].

The small footprint of the devices arises from the large refractive index contrast between the crystalline and amorphous states of the PCM. Because the change

in absorption and/or phase is so prominent in PCMs, the length of devices based on PCMs is often on the scale of (tens of) micrometres or even less, which for photonics is extremely small. In comparison, modulators in LN platforms can easily be up to a centimetre long [29]. When used as elements in scaled systems, the small footprint of the PCM devices distinguishes it from most other technologies: in Fig. 1b in Ref. 129, PCMs are identified as close to the ideal.

To make PCMs excel, we need to find applications where their limitations are not detrimental. As seen in Fig. 13 in Ref. 196, of all optical technologies employed for weighting in neural networks, PCMs offer the lowest power consumption and the smallest device footprint. Still, their reconfiguration time is not the fastest. Current PCM devices also have a limited bit resolution. However, for some applications, such as weight kernels in CNNs, the performance characteristics of PCM elements, with slow, energy-intensive switching and limited bit resolution, are acceptable [4].

2.5 Conclusion

In this review, I have summarised the progress in the field of integrated photonic switches over the past decades. Different types of photonic switch structures and switching mechanisms have been presented, along with notable application areas. Based on the metrics presented in this chapter, I believe that PCM-based photonic switches will have a significant impact on large-scale systems with stringent SWaP requirements and the need for multi-level switching, while meeting relaxed speed and endurance requirements. We have yet to see any large-scale implementations of PCM-based photonic switches. Still, I expect they will perform well in areas where non-volatility and tunability are particularly important.

Experience without theory is blind, but theory without experience is mere intellectual play.

— Immanuel Kant

3 Theory

Contents

3.1 Light propagation 54
3.1.1 Effective refractive index 55
3.2 Performance metrics of photonic switches 57
3.3 Directional coupler theory 59
3.4 Ring resonator theory 60
3.4.1 Dynamic behaviour of ring resonators 62
3.5 Mach–Zehnder interferometer theory 63
3.5.1 Governing equations of Mach–Zehnder interferometers . 64
3.5.2 Ring-assisted Mach–Zehnder interferometer theory . . . 67
3.6 Phase-change material theory 69
3.7 Neural network theory 71
3.7.1 Weight quantisation in neural networks 72
3.7.2 Working principle of photonic crossbar array 74
3.8 Sensor management and control theory 74
3.8.1 Kalman filter implementation 76
3.8.2 Quadratic programming for model predictive control . . 76
3.8.3 Primal–dual gradient (Langrangian) method 79

This chapter establishes the theoretical framework used throughout this thesis. I begin with an overview of light propagation in guided media in Section 3.1, introducing key concepts such as the effective refractive index and the theory behind some of the mechanisms by which it can be modulated. In Section 3.2, I then define the performance metrics of photonic switches—switching speed, footprint,

insertion loss (IL), extinction ratio (ER), crosstalk (CT), power consumption, and optical bandwidth—and clarify how these quantities trade off in practice. Next, I review evanescent coupling and the resonant and interferometric phenomena underlying our devices, including DCs, ring resonators, MZIs, and RAMZIs. In Section 3.6, I summarise the materials physics relevant to non-volatile switching with PCMs, distinguishing between amorphisation and crystallisation and discussing their practical implications for device operation. Finally, in Section 3.7, I connect these photonic primitives to computation and control by outlining neural network concepts and photonic crossbar operation, followed by the foundations of sensor management, Kalman filtering, and MPC in Section 3.8.

3.1 Light propagation

From Maxwell's equations, we have the plane wave equation

$$E = E_0 e^{i(\omega t - \beta z)}, \tag{3.1}$$

where ω is the angular frequency of the electromagnetic wave, $\beta = \frac{2\pi}{\lambda_0}\bar{n}$ is the (generally complex) propagation constant, λ_0 is the wavelength of the electromagnetic wave, and $\bar{n}$ is the refractive index of the homogeneous propagation medium. The refractive index of a material is generally complex-valued, $\bar{n} = n + i\kappa$, where n is the real part (often just called the *refractive index*) that governs phase propagation and κ is the imaginary part (called the *extinction coefficient*) that accounts for absorption losses in the medium. Electromagnetic waves can be confined to propagate within a medium (*the core*) when the refractive index of the propagation medium, n, exceeds that of the surrounding medium (*the cladding*), n_0, provided the core geometry supports at least one guided mode. This confinement occurs through total internal reflection (TIR) at the core-cladding interface when light propagates at sufficiently shallow angles with respect to the interface. This confining structure is called a *waveguide*, which in integrated photonics is typically realised as a (quasi-)rectangular or ridge-shaped strip of high-index material on a substrate, surrounded by a lower-index cladding or air.

The speed of light signals in a waveguide is given by the simple equation

$$v_{\mathrm{g}} = \frac{c}{n_{\mathrm{g}}}, \tag{3.2}$$

where $c = 299\,792\,458\,\mathrm{m/s}$ is the speed of light in vacuum and n_g is the group index of the medium. The group index considers the dispersion in the waveguide [31] and is defined as

$$n_{\mathrm{g}} = n_{\mathrm{eff}} - \lambda_0 \frac{dn_{\mathrm{eff}}}{d\lambda}, \tag{3.3}$$

where λ_0 is the centre wavelength and n_{eff} is the effective refractive index of the waveguide (see Section 3.1.1). It is assumed that the reader is familiar with the basic principles of light propagation in optical waveguides, as well as the basic working principle of grating couplers for coupling light into waveguides from free space, as these topics will not be further explored in this work.

3.1.1 Effective refractive index

In the same way that the propagation of light through a homogeneous medium is determined by $\beta = \frac{2\pi}{\lambda_0}\bar{n}$, the phase constant in a waveguide is given as $\beta = \frac{2\pi}{\lambda_0}\bar{n}_{\mathrm{eff}}$.

Effective refractive index, n_{eff} (real part)

In optics, a *phase shift* refers to the change of the phase of the light wave as it propagates through a medium.[1] A phase shift relative to the phase of the free-space wave occurs when the light propagates through a medium of effective refractive index n_{eff} and physical length L, following the equation

$$\Delta\phi = \beta' L = \frac{2\pi}{\lambda_0} \cdot n_{\mathrm{eff}} L, \tag{3.4}$$

where β' is the real part of the propagation constant and λ_0 is the free-space wavelength of the light. For a phase shifter to achieve a full phase shift of the propagating wave passing through it, the phase component $\phi = \beta' z$ of the electric

[1] Or reflects off a surface; however, here, only the phase shift upon propagation is covered.

field $E(t) = E_0 e^{i(\omega t - \phi)}$ must change by an integer multiple of 2π, that is, $\Delta\phi = 2m\pi$, where $m \in \mathbb{Z}$. The physical length needed to achieve a π phase shift is therefore

$$L_\pi = \frac{\lambda_0}{2\Delta n_{\text{eff}}}, \tag{3.5}$$

where Δn_{eff} is the difference in effective refractive index of the phase shifter. In resonant and interferometric devices, such as MRRs and MZIs, we often refer to the free spectral range (FSR); one FSR is equal to a 2π phase shift.

Effective extinction coefficient, κ_{eff} (imaginary part)

The effective extinction coefficient is the imaginary part of the effective refractive index. It is given as

$$\kappa_{\text{eff}} = \frac{\alpha\lambda_0}{4\pi}, \tag{3.6}$$

where α is the intensity absorption coefficient given in units of inverse length and can be found from the normalised transmission change of a device of length L through the Beer–Lambert law, $P_{\text{out}} = P_{\text{in}} \exp(-\alpha L)$:

$$\alpha = -\frac{1}{L} \ln\left(\frac{P_{\text{out}}}{P_{\text{in}}}\right) \approx \frac{(P_{\text{in}} - P_{\text{out}})/P_{\text{in}}}{L}, \tag{3.7}$$

where P_{in} and P_{out} are the optical input and output powers, and the approximation holds for $P_{\text{out}}/P_{\text{in}} \sim 1$. The easiest way to measure the loss, and therefore the effective extinction coefficient, of a device (e.g., a hybrid waveguide-PCM system in Chapter 6), is to measure the optical transmission of a plain waveguide and then the device, with all other parameters constant. The transfer matrix method is an alternative approach for determining the loss of a phase shifter in an MZI [197, 198].

Thermo-optic and electro-optic effects

The TO effect is based on changing the refractive index of a material upon heating. When the refractive index of the material changes, a phase shift is introduced, given by the formula [44]

$$\Delta\varphi = \frac{2\pi}{\lambda_0}\Delta n L = \frac{2\pi}{\lambda_0}\frac{\partial n}{\partial T}\Delta T L, \tag{3.8}$$

where Δn is the change in refractive index, L is the interaction length (and usually the length of the heater), $\partial n/\partial T$ is the TO coefficient, and ΔT is the change in temperature. As we can see from the equation, the TO effect is, in theory, lossless, as it only affects the real part of the refractive index.

The EO plasma dispersion effect introduces carriers into the light path. The phase shift of the plasma dispersion effect is given by [20]

$$\Delta\varphi = \left(\frac{2\pi}{\lambda_0}\Delta n(N) + \frac{i\Delta\alpha(N)}{2}\right) L, \tag{3.9}$$

where N is the carrier density and $\Delta n(N)$ and $\Delta\alpha(N)$ are the changes in the effective refractive index and free-carrier absorption, respectively.

3.2 Performance metrics of photonic switches

The performance metrics of photonic switches include switching speed, extinction ratio, insertion loss, crosstalk, power consumption, device footprint, and optical bandwidth. Other important features of photonic switches include the stability and (mainly in the case of tunable switches) precision of the output. Several of these metrics are inversely related, creating trade-off situations. Because the ideal switch does not exist, the big question is what trade-offs one is willing to accept for the application at hand.

The switching time is typically given as the slowest of the rise time (10 % to 90 % of maximum transmission) and fall time (90 % to 10 %) of the switch [76]. To define insertion loss (IL), extinction ratio (ER), switching contrast (SC), and crosstalk (CT), we take a 1×2 coupler as an example, with output ports x and y and states X and Y corresponding to the state with maximum output out of port x and y, respectively. The IL for state X is given as

$$\mathrm{IL}_X = 10\log\left(\frac{P^{in}}{P_x^{out}(X)}\right), \tag{3.10}$$

where $P_x^{out}(X)$ is the power out of port x for the coupler in state X. To find the IL of a device with length L and absorption coefficient α, we use

$$\mathrm{IL} = -10\log(e^{-\alpha L}) = \frac{10}{\ln(10)}\alpha L. \tag{3.11}$$

The ER is often defined as the discrimination between the two output ports of a coupler:

$$\mathrm{ER}_X = 10\log\left(\frac{P_x^{out}(X)}{P_y^{out}(X)}\right). \tag{3.12}$$

On the other hand, for a modulator, the ER is typically defined as the on-off ratio of the modulator. In this work, I will instead call this latter definition the SC,[2] which can then be extended also to cover couplers [199]; for output port x, the on and off states are X and Y, respectively, so

$$\mathrm{SC}_x = 10\log\left(\frac{P_x^{out}(X)}{P_x^{out}(Y)}\right). \tag{3.13}$$

We define the optical CT of the switch as [200]

$$\mathrm{CT}_X = -\left(\mathrm{IL}_X + \mathrm{ER}_X\right) = 10\log\left(\frac{P_y^{out}(X)}{P^{in}}\right). \tag{3.14}$$

With these definitions, IL, ER, and SC are typically positive, while CT is negative [135, 199, 201]. Although definitions with opposite signs also exist [200], I adopt a convention in which IL, ER, and SC take positive values, while CT is negative. This choice reflects how we refer to the terms in relation to real devices; we want to *reduce* the insertion loss and have devices with *high* extinction ratio and switching contrast, but *low* crosstalk. All of the equations above are expressed in decibels (dB), but can also be expressed as just the ratio in parentheses.

Comparing power consumption between different switches is not easy, as the number of switching events per time dictates total power consumption, especially for non-volatile switches. Both P_{avg} and P_π, i.e., the average power and the power needed for a π shift, respectively, are used [166, 169]. For optical modulators, the energy required to produce a bit of data is often used as a metric of power consumption [56]. For tunable couplers, a similar metric is not useful, as the switch might stay in a state for a long time.

Finally, the bandwidth of a photonic device can be a confusing term as it is also quoted in different ways. For example, the 3 dB bandwidth is a widely used metric

[2]With these definitions, an optical modulator does not have an extinction ratio, only a switching contrast.

for optical modulators, denoting the highest frequency at which modulation of the light is meaningful. In the context of light sources, the optical bandwidth refers to the width of the optical spectrum of the source (also called *linewidth*), while for photodetectors, the bandwidth denotes the frequency at which fluctuations in optical power can be observed. Lastly, optical bandwidth often refers to wavelength ranges, for instance, the transmission bandwidth of an optical fibre or photonic device. For this definition, a commonly used metric is the full width at half maximum (FWHM; see Section 3.4), but for some applications, such as tunable couplers, it might be more meaningful to instead quote the range of wavelengths for which the other parameters, such as ER or IL, do not vary significantly [16, 202]. To convert between the bandwidth given in terms of wavelength and frequency, we can use the equation

$$\Delta\nu = \frac{c}{\lambda_0^2}\Delta\lambda, \tag{3.15}$$

which we obtain by differentiating $\nu = c/\lambda$ with respect to λ. Note that this relationship is only valid for a small wavelength interval around the centre wavelength.

3.3 Directional coupler theory

The power coupled across a DC of length L_c is given by [203]

$$P_\mathrm{cross} = \frac{\xi^2}{S^2}\sin^2(SL_\mathrm{c}) = \frac{\xi^2}{\left(\xi^2 + \left(\frac{\Delta\beta'}{2}\right)^2\right)}\sin^2\left(\sqrt{\xi^2 + \left(\frac{\Delta\beta'}{2}\right)^2}L_\mathrm{c}\right), \tag{3.16}$$

where ξ is the field coupling coefficient (in units of 1/length), $S = \left(\xi^2 + (\Delta\beta'/2)^2\right)^{0.5}$, and $\Delta\beta' = 2\pi\Delta n_\mathrm{eff}/\lambda_0$ is the difference in the phase constant between the two arms, also called the phase mismatch. If the two arms are identical, the power oscillates sinusoidally as $P_\mathrm{cross} = \sin^2(\xi L_\mathrm{c})$, with complete power transfer after coupling length $L_\mathrm{c} = \frac{\pi}{2\xi}$.

To tune the coupling of a DC, there are three main approaches: changing either the phase mismatch ($\Delta\beta$), the coupling coefficient (ξ), or the interaction length (L_c). The first approach is to change the effective refractive index, for example, by heating one of the arms or by depositing PCM. The other two can be modified by changing

the physical parameters of the coupling region with tunable MEMS couplers; for instance, by varying the coupling gap (changing ξ) or the interaction length (changing L_c). Finally, in evanescent devices such as DCs, the optical field in the gap between the waveguides decays quasi-exponentially, which means that a small error in the coupling region can significantly affect the coupling [204]. Therefore, all fabrication parameters in the coupling region are essential for optimal device performance, such as the etch depth of the rib waveguides, the waveguide width, and the gap.

3.4 Ring resonator theory

Ring resonators were described in Section 2.1.3. They are inherently narrowband because of the resonant condition $\lambda_{\text{res}} = \frac{n_{\text{eff}} L}{m}$, where n_{eff} is the effective refractive index of the ring waveguide, L is the round-trip length of the ring (i.e., circumference), and $m = 1, 2, 3, \ldots$ is the mode number of the ring resonator [31]. In other words, the wavelength of the light resonantly coupling into the ring must satisfy the condition that an integer multiple of the wavelength fits inside the optical round-trip length of the ring. The ring is on resonance when the round-trip phase shift $\phi = \frac{2\pi}{\lambda} n_{\text{eff}} L_{\text{ring}}$ is a multiple of 2π, i.e., $\phi_{\text{res}} = 2m\pi$ [31].

The transmission from the through port of a ring resonator is given as [31]

$$t = \frac{a^2 - 2ra\cos\phi + r^2}{1 - 2ra\cos\phi + (ra)^2}, \tag{3.17}$$

where r is the self-coupling coefficient and a is the round-trip loss coefficient, which is related to the power attenuation coefficient α as $a^2 = \exp(-\alpha L_{\text{ring}})$ [31]. Note that the self-coupling coefficient r and the cross-coupling coefficient k (see Fig. 2.2c) are related by the equation $r^2 + k^2 = 1$ when there are no losses in the coupling section. Furthermore, when the coupled power equals the power lost in the ring, $1 - a^2 = k^2$ or $r = a$, the ring is at critical coupling, and the transmission at the resonance wavelengths is 0.

The transmission response of one resonance of a ring resonator takes the shape of the Cauchy–Lorentz distribution (often just called a Lorentzian function)

$$y = \frac{1}{\pi} \frac{\gamma}{(x - x_0)^2 + \gamma^2}, \tag{3.18}$$

where x_0 is the location and γ is the scale parameter, given as $\gamma = \text{FWHM}/2$, i.e., half of the width of the peak at half of the maximum transmission. The FWHM of the resonance can be calculated from the equation [31]

$$\text{FWHM} = \frac{(1 - ra)\,\lambda_{\text{res}}^2}{\pi n_{\text{g}} L_{\text{ring}} \sqrt{ra}}. \tag{3.19}$$

The quality factor (Q factor) is a parameter commonly used when referring to resonators, which describes the sharpness of the resonance, and is given as

$$Q = \frac{\lambda_{\text{res}}}{\text{FWHM}}. \tag{3.20}$$

The resonance is repeated once every FSR, which, within a first-order approximation of the dispersion, is given as [31]

$$\text{FSR} = \frac{\lambda^2}{n_{\text{g}} L_{\text{ring}}}. \tag{3.21}$$

In the frequency domain, $\text{FSR}_\nu = \frac{c}{n_{\text{g}} L_{\text{ring}}}$. Finally, the ER of a ring resonator is often defined as the ratio between a peak and its neighbouring trough. Some of the definitions presented in this section are shown in Fig. 3.1.

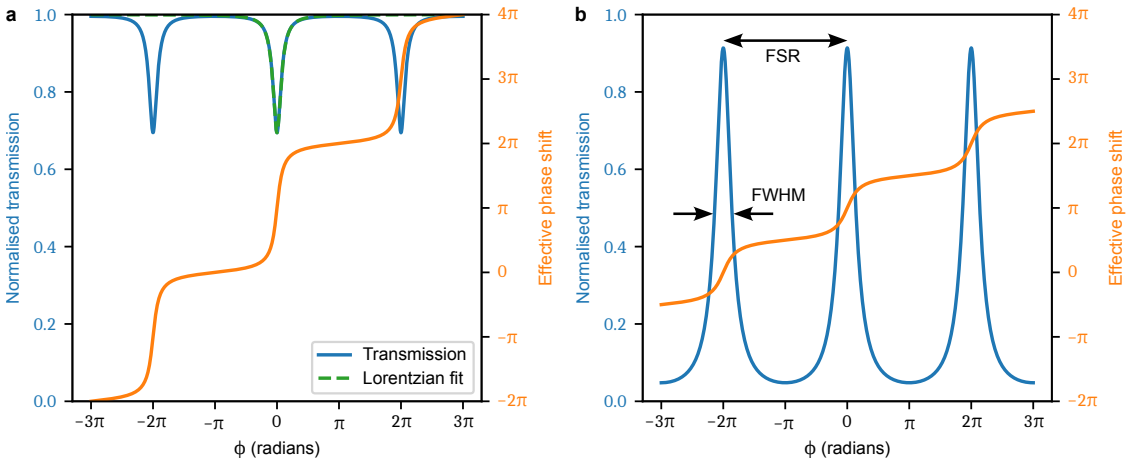

Figure 3.1: Transmission response and phase delay of ring resonators. **a**, For an all-pass ring resonator, with a Lorentzian fit to the transmission peak. **b**, For the drop port of an add-drop ring resonator. The FSR and full width at half maximum (FWHM) are indicated by arrows. For both rings, $a = 0.98$ and $r = 0.8$.

The effective phase shift of an all-pass ring resonator, φ, is given as [31]

$$\varphi = \pi + \phi + \arctan\frac{r\sin\phi}{a - r\cos\phi} + \arctan\frac{ra\sin\phi}{1 - ra\cos\phi}. \tag{3.22}$$

As seen in Fig. 3.1, the effective phase shift of the light travelling through an all-pass ring resonator is $\Delta\varphi = 2\pi$ across one FSR (detuning of $\phi = 2\pi$), whereas for an add-drop resonator, $\Delta\varphi = \pi$. Furthermore, the phase accumulates across several resonances. Practically, this accumulation is not usually relevant, because a 2π phase shift is equivalent to a 0 phase shift. However, for the RAMZI presented in Chapter 7, the accumulated phase shift across resonances serves a critical function. The detuning in Fig. 3.1 can be implemented practically by adding a phase shifter to the ring.

3.4.1 Dynamic behaviour of ring resonators

Ring resonators can act as high-speed temporal differentiators [29, 30, 205], which are found in microwave photonics (MWP) systems and are used, for instance, in edge detection for image recognition [29]. Edge detection means that if we send in an optical pulse, we only detect its edges: for a rectangular-like pulse, only the rising and falling edges are detected, which manifests as an optical spike when using temporal differentiators. An intuitive way of thinking about the classical working principle of the temporal differentiator is as follows: when light is sent into the input port of the ring resonator, k^2 of the light couples into the ring while r^2 continues through to the through port. When enough time has passed that the light in the ring has traversed one round-trip, we will start to see interference in the coupling region. At steady state, for a ring near critical coupling, some wavelengths do not propagate out of the through port due to destructive interference. If we set the laser to one of these wavelengths and measure the temporal behaviour, we will see that initially some light is allowed through, but after the interference starts happening and the interference condition reaches a steady state, no light is allowed through. In other words, we can detect the rising edge by seeing a spike. For the falling edge, there is suddenly no light entering the input port, so the light in the ring can exit the through port without experiencing the previous steady-state interference. As a result, we detect the falling edge by seeing another spike.

The same phenomenon is seen in unbalanced MZIs (see next section). In these structures, the light is split into two arms, one with a longer path length, before the light interferes with itself at the combiner. Because the longer arm effectively works as a delay line, the same working principle as above is also true for the unbalanced MZI: for the rising edge, the light in the shortest arm traverses the MZI quicker than the light in the long arm, and therefore there is no interference for a short period, while for the falling edge, the final light in the longest arm exits after the light in the shortest arm has passed, again without interfering.

3.5 Mach–Zehnder interferometer theory

An MZI is a fundamental photonic device that operates on the principle of optical interference (see Section 2.1.2 for examples). Similar to the ring resonator, MZIs find use in optical filters and temporal differentiators [29, 155], but more importantly, the MZI is the go-to structure for photonic modulators and couplers. The device consists of a splitter at the input and a combiner at the output connected by two optical paths (arms). The operation of an MZI can be understood by considering the phase relationship between the two arms. When an optical field E_{in} enters the device, it is split into two components at the input splitter. These components propagate through the upper and lower arms, accumulating different phase shifts φ_1 and φ_2, respectively, because of different path lengths and/or different effective refractive indices in the two arms. The relative phase difference between the two arms is given by

$$\Delta\varphi = \varphi_2 - \varphi_1 = \beta_2' L_2 - \beta_1' L_1 = \frac{2\pi}{\lambda_0}(n_{\text{eff},\,2}\, L_2 - n_{\text{eff},\,1}\, L_1) \tag{3.23}$$

where $n_{\text{eff},\,1}$ and $n_{\text{eff},\,2}$ are the effective refractive indices and L_1 and L_2 are the physical lengths of the upper and lower arms, respectively. The FSR of an MZI is more or less equivalent to that of a ring resonator (Eq. (3.21)), but instead of the ring circumference L, we use the path length difference $\Delta L = L_2 - L_1$.

3.5.1 Governing equations of MZIs

The general equation governing a 2×2 MZI is given as [206]

$$\begin{bmatrix} E_{\text{out},1} \\ E_{\text{out},2} \end{bmatrix} = \Big[\text{Splitter}\Big] \Big[\text{Arms}\Big] \Big[\text{Combiner}\Big] \begin{bmatrix} E_{\text{in},1} \\ E_{\text{in},2} \end{bmatrix}, \tag{3.24}$$

where $E_{\text{out},1}$ and $E_{\text{out},2}$ are the E-fields out of the port on the same and opposite side as $E_{\text{in},1}$, respectively, $\Big[\text{Splitter}\Big]$ and $\Big[\text{Combiner}\Big]$ are the transfer matrices representing the splitter and coupler of the MZI, and the transfer matrix $\Big[\text{Arms}\Big]$ represents the effect of the two arms. For a 1×2 MZI, we simply set $E_{\text{in},2} = 0$.

Transfer matrices of splitters and combiners

The splitters and combiners used in this thesis are either 1×2 Y-branches or 2×2 DCs. A 2×2 DC is characterised by its coupling coefficient k and transmission coefficient r, which describe the fraction of power coupled between waveguides (as given by Eq. (3.16)) and transmitted through the same waveguide, respectively. For a lossless coupler, these coefficients satisfy the relation $k^2 + r^2 = 1$. For a 2×2 DC, the transfer matrix equation is

$$\begin{bmatrix} E_{\text{out},1} \\ E_{\text{out},2} \end{bmatrix} = \begin{bmatrix} r & ik \\ ik & r \end{bmatrix} \begin{bmatrix} E_{\text{in},1} \\ E_{\text{in},2} \end{bmatrix}. \tag{3.25}$$

When light couples across the DC, it accumulates a $\pi/2$ phase shift, which is why we introduce the imaginary unit i. For a coupler with a 50:50 splitting ratio, the coupling coefficients become $r = k = 1/\sqrt{2}$, and the transfer matrix equation simplifies to

$$\begin{bmatrix} E_{\text{out},1} \\ E_{\text{out},2} \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & i \\ i & 1 \end{bmatrix} \begin{bmatrix} E_{\text{in},1} \\ E_{\text{in},2} \end{bmatrix}. \tag{3.26}$$

The same equations hold for when MMIs are used. For a Y-splitter, the transfer matrix equation is given as

$$\begin{bmatrix} E_{\text{out},1} \\ E_{\text{out},2} \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} E_{\text{in}}, \tag{3.27}$$

and for a Y-combiner, it is given as

$$E_{\text{out}} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} E_{\text{in},1} \\ E_{\text{in},2} \end{bmatrix}. \tag{3.28}$$

Note, in real systems, the coupling of a DC is wavelength dependent, so assuming $r = k = 1/\sqrt{2}$ is typically not valid for all wavelengths in the range.

Transfer matrix of Mach–Zehnder interferometer arms

For normal MZIs, the propagation matrix $\Big[\text{Arms}\Big]$ accounts for both phase accumulation and amplitude attenuation in both arms, and is given by

$$\Big[\text{Arms}\Big] = \begin{bmatrix} a_1 e^{i\varphi_1} & 0 \\ 0 & a_2 e^{i\varphi_2} \end{bmatrix}, \tag{3.29}$$

where $a_j = e^{-\alpha_j L_j}$ is the amplitude transmission coefficient accounting for propagation loss, α_j is the power loss coefficient (in units of inverse length), L_j is the arm length, and $\varphi_j = \frac{2\pi}{\lambda_0} n_{\text{eff},\,j}\, L_j$ is the phase accumulated in arm j.

The transfer matrix of a 2×2 MZI with 2×2 DCs at the input and output is

$$\mathbf{T}_{\text{MZI}} = \begin{bmatrix} r & ik \\ ik & r \end{bmatrix} \begin{bmatrix} a_1 e^{i\varphi_1} & 0 \\ 0 & a_2 e^{i\varphi_2} \end{bmatrix} \begin{bmatrix} r & ik \\ ik & r \end{bmatrix}. \tag{3.30}$$

If we set the splitter and combiner to 50:50 splitting ratios, then $r = k = 1/\sqrt{2}$ and the transfer matrix of the MZI becomes

$$\mathbf{T}_{\text{MZI}} = \frac{1}{2} \begin{bmatrix} a_1 e^{i\varphi_1} - a_2 e^{i\varphi_2} & i(a_1 e^{i\varphi_1} + a_2 e^{i\varphi_2}) \\ i(a_1 e^{i\varphi_1} + a_2 e^{i\varphi_2}) & -a_1 e^{i\varphi_1} + a_2 e^{i\varphi_2} \end{bmatrix} \tag{3.31}$$

For an input field E_{in} at port 1 (with port 2 having zero input), the output fields are:

$$E_{\text{out},\,1} = T_{11} E_{\text{in},\,1} = \frac{1}{2}(a_1 e^{i\varphi_1} - a_2 e^{i\varphi_2}) E_{\text{in},\,1}, \tag{3.32}$$

$$E_{\text{out},\,2} = T_{21} E_{\text{in},\,1} = \frac{i}{2}(a_1 e^{i\varphi_1} + a_2 e^{i\varphi_2}) E_{\text{in},\,1}. \tag{3.33}$$

The corresponding output powers from the MZI are given as

$$P_{\text{out},\,1} = |E_{\text{out},\,1}|^2 = \frac{|E_{\text{in},\,1}|^2}{4}(a_1^2 + a_2^2 - 2a_1 a_2 \cos\Delta\varphi), \tag{3.34}$$

$$P_{\text{out},\,2} = |E_{\text{out},\,2}|^2 = \frac{|E_{\text{in},\,1}|^2}{4}(a_1^2 + a_2^2 + 2a_1 a_2 \cos\Delta\varphi). \tag{3.35}$$

where $\Delta\varphi$ is the phase difference between the two arms given in Eq. (3.23).

In the lossless case where $a_1 = a_2 = 1$, these expressions reduce to:

$$P_{\text{out},\,1} = |E_{\text{in},\,1}|^2 \sin^2\left(\frac{\Delta\varphi}{2}\right), \tag{3.36}$$

$$P_{\text{out},\,2} = |E_{\text{in},\,1}|^2 \cos^2\left(\frac{\Delta\varphi}{2}\right). \tag{3.37}$$

These equations reveal the interferometric nature of the device. When $\Delta\varphi = 2m\pi$, destructive interference occurs at the bar port (out 1) while constructive interference occurs at the cross port (out 2). Conversely, when $\Delta\varphi = (2m+1)\pi$, constructive interference occurs at the bar port. The two output ports are complementary, satisfying $P_{\text{out},1} + P_{\text{out},2} = |E_{\text{in},1}|^2$ at all times, ensuring energy conservation. From Eqs. (3.36) and (3.37), we see the effect of introducing an arm length difference in an unbalanced MZI, that is, where $\Delta\varphi = 2\pi n_{\text{eff}}/\lambda_0 \times (L_2 - L_1)$. The alternative is to introduce phase shifters that alter the effective refractive index. In most practical systems with active control, one or both arms will have a phase shifter, and the phase difference is therefore $\Delta\varphi = 2\pi L/\lambda_0 \times (n_{\text{eff},2} - n_{\text{eff},1})$ for a balanced MZI.

In Chapter 7, I present an MZI with a Y-branch splitter and a DC combiner. Substituting the DC at the input with a Y-branch gives

$$\mathbf{T}'_{\text{MZI}} = \frac{1}{2}\begin{bmatrix} a_1 e^{i\varphi_1} + i a_2 e^{i\varphi_2} \\ i a_1 e^{i\varphi_1} + a_2 e^{i\varphi_2} \end{bmatrix} \tag{3.38}$$

and the power

$$P'_{\text{out},1} = |E'_{\text{out},1}|^2 = \frac{|E_{\text{in},1}|^2}{4}(a_1^2 + a_2^2 - 2a_1 a_2 \sin\Delta\varphi), \tag{3.39}$$

$$P'_{\text{out},2} = |E'_{\text{out},2}|^2 = \frac{|E_{\text{in},1}|^2}{4}(a_1^2 + a_2^2 + 2a_1 a_2 \sin\Delta\varphi). \tag{3.40}$$

In the lossless case, we get

$$P'_{\text{out},1} = |E_{\text{in},1}|^2 \sin^2\left(\frac{\Delta\varphi - \pi/2}{2}\right), \tag{3.41}$$

$$P'_{\text{out},2} = |E_{\text{in},1}|^2 \cos^2\left(\frac{\Delta\varphi - \pi/2}{2}\right). \tag{3.42}$$

As we can see, substituting one of the DCs with a Y-branch is equivalent to adding a constant phase shift of $\pi/2$ to the upper arm. For $\Delta\varphi = 2\pi m$, both outputs give 50 % power out, which is called the *quadrature point*.

If there is a phase error or power imbalance between the two arms of the MZI, the CT between the output ports for couplers increases, and the SC for modulators decreases. A power imbalance is usually caused by splitter and combiner imperfections or loss imbalance in the arms [20]. When loss is present and unequal

between the arms ($a_1 \neq a_2$), the MZI no longer exhibits complete destructive interference. The maximum and minimum power out of the ports become

$$P_{\text{extrema}} = \frac{|E_{\text{in},1}|^2}{4}(a_1^2 + a_2^2 \pm 2a_1a_2), \tag{3.43}$$

and the combined output power $P_{\text{out},1} + P_{\text{out},2} = \frac{|E_{\text{in},1}|^2}{2}(a_1^2 + a_2^2)$ is less than the input power, reflecting the optical loss in the system.

3.5.2 Ring-assisted Mach–Zehnder interferometer

If we add a ring resonator to one of the arms of an MZI, the result is a ring-assisted Mach–Zehnder interferometer (RAMZI). Various versions exist, with single or multiple all-pass or add-drop rings [206, 207], but in this work, we will focus on MZIs with a single all-pass ring. The transfer matrix of the MZI arms in a RAMZI with an all-pass ring in the upper arm (arm 1) is given as

$$\mathbf{T}_{\text{RAMZI}} = \begin{bmatrix} a_1 t \exp\left(i\left(\varphi_{\text{ring}} + \varphi_1\right)\right) & 0 \\ 0 & a_2 \exp(i\varphi_2) \end{bmatrix}, \tag{3.44}$$

where t is the transmission response of the ring resonator, φ_{ring} is the effective phase shift of the transmitted light at the through port of the ring in the upper arm, and φ_1 (φ_2) is the phase shift in the upper (lower) arm excluding the contribution from the ring resonator. If the arms of the MZI have no loss, the RAMZI transmission response consists of two overlapping responses: the phase-dependent interference condition between the arms of the MZI and the phase and transmission introduced by the ring. For a RAMZI with a Y-splitter at the input, a DC at the output, $a_1 = a_2 = 1$, and $\varphi_1 = \varphi_2$, we get

$$P_{\text{out},1} = |E_{\text{out},1}|^2 = \frac{|E_{\text{in}}|^2}{4}(1 + t^2 - 2t\sin\varphi_{\text{ring}}), \tag{3.45}$$

$$P_{\text{out},2} = |E_{\text{out},2}|^2 = \frac{|E_{\text{in}}|^2}{4}(1 + t^2 + 2t\sin\varphi_{\text{ring}}). \tag{3.46}$$

This equation is quite complex, but I can still make a few general comments. First, at resonance, the drop in the ring transmission suppresses the maximum

transmission available to the RAMZI. However, outside of the resonances, $t \sim 1$ and $t^2 \sim 1$ when the ring loss is low ($a \sim 1$) and the equation simplifies to

$$P_{\text{out},1} \approx \frac{|E_{\text{in}}|^2}{2}(1 - \sin\varphi_{\text{ring}}) = |E_{\text{in}}|^2 \sin^2\left(\frac{\varphi_{\text{ring}} - \pi/2}{2}\right). \tag{3.47}$$

I have shown before that the transmission spectrum of a ring follows a Lorentzian function (Eq. (3.18)), but this fitting only really holds inside the resonance when the loss of the ring is low. In a similar vein, we can fit the transmission spectrum of the RAMZI described here with the inverse of the cumulative distribution function (CDF) of a Lorentzian, which is the tangent function. In contrast to the Lorentzian fit of the ring resonator spectrum, this fit is only valid between the resonances. In this work, this equation will be given by

$$T = \gamma' \tan\left(\frac{\pi\,(x - x_0)}{d}\right) + T_0, \tag{3.48}$$

where γ' is the scale parameter, x_0 and T_0 are the locations of the midpoint on the x-axis and y-axis, respectively, and d is related to the FSR. In this work, x, x_0, and d will be given in both wavelength and phase shift. Note that the CDF is a probability distribution and that the inverse of the CDF (called the *quantile function*) would normally be defined as $Q : [0, 1] \mapsto \mathbb{R}$, which means $d = 1$ and $x_0 = 1/2$.

Equation (3.45) is presented in Fig. 3.2a for different a and r, together with the corresponding ring transmission and phase delay. We see that unless $a = 1$, the ring transmission near resonance will lead to a transmission drop in the RAMZI. In Fig. 3.2b, I present the curve fit to Eq. (3.48). Interestingly, we see that the d parameter is necessary, as $d = 2\pi$ does not give a good fit. Only in the case of $a \sim 1$ and $r \sim 1$ is $d \approx 2\pi$.

In Chapter 7, we want to find the derivative of the inverse of the RAMZI transmission. The inverse of Eq. (3.48) is

$$x = \frac{d}{\pi}\arctan\left(\frac{T - T_0}{\gamma'}\right) + x_0, \tag{3.49}$$

and its derivative is then given as

$$\frac{dx}{dT} = \frac{d}{\pi}\frac{1}{1 + \left(\frac{T-T_0}{\gamma'}\right)^2}\frac{1}{\gamma'} = d \times \frac{1}{\pi}\frac{\gamma'}{(T - T_0)^2 + \gamma'^2}, \tag{3.50}$$

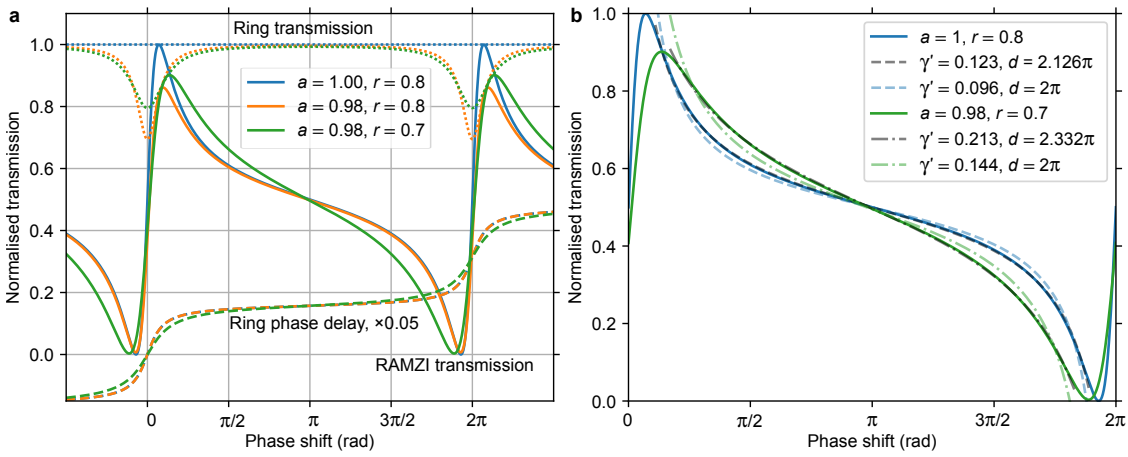

Figure 3.2: Analytical solution to the optical power output from RAMZIs. **a**, Ring transmission, ring phase delay, and RAMZI transmission for devices with $a = 1.0$ and $r = 0.8$ (blue), $a = 0.98$ and $r = 0.8$ (orange), and $a = 0.98$ and $r = 0.7$ (green). The phase delay is multiplied by 0.05 to fit in the figure. **b**, Curve fits to the blue and green plots in **a**. The coloured curve fits have a fixed $d = 2\pi$.

which is d times a Lorentzian function with scale γ'. Note that the scale parameter γ' is not the same as the γ of the Lorentzian function of the ring.

Finally, if the ring is lossless, it becomes a pure all-pass phase shifter, and we can get virtually any RAMZI curve without loss of transmission close to the resonances. In reality, however, we are limited by propagation losses in the ring and losses in the coupling region. Nevertheless, values for the loss coefficient of a ring resonator have repeatedly reached close to unity [208]. For example, Han *et al.* were able to demonstrate propagation losses as low as $0.24\,\mathrm{dB\,cm^{-1}} = \left(1 - 10^{-\frac{0.24}{10}}\right) = 0.054\,\mathrm{cm^{-1}}$ for a ring with effective radius $R_{\mathrm{eff}} = 20\,\mu\mathrm{m}$ [209]. Ignoring coupling losses, the single-pass amplitude transmission of this ring is $a = \exp\left(-\frac{0.054\times 2\pi\cdot 20\times 10^{-4}}{2}\right) = 0.9997$. Practically speaking, for $a = 0.999$, the maximum RAMZI transmission is above 0.9 for all $r < 0.984$, which allows for a wide range of shapes of the RAMZI transmission curve determined by γ'.

3.6 Phase-change material theory

In PCMs used for integrated photonics, the optical response is typically characterised by a change in the refractive index or extinction coefficient. For phase shifters, we are interested in the change in the refractive index between the crystalline

and amorphous states, $\Delta n_{\text{eff}} = n_{\text{eff,cry}} - n_{\text{eff,amo}}$ [120], that gives us L_π through Eq. (3.5). For absorptive elements, we are interested in the switching contrast (%) or modulation depth (dB) [195].

Although the operational principles of PCMs were introduced in Section 2.3, their behaviour is ultimately governed by the underlying materials physics of the phase transition. Switching from the crystalline to amorphous state of the PCM (amorphisation) requires the temperature to be increased above the melting temperature, T_m, and subsequently quenched (i.e., rapidly cooled) at rates higher than $10^9\,\text{K}\,\text{s}^{-1}$ [95]. This process results in the formation of a covalently bonded, solid-state matrix that lacks long-range atomic order, with atoms arranged similarly to those of the liquid phase [95]. In contrast, to crystallise the material (crystallisation), the temperature is increased above the crystallisation temperature (also known as the glass-transition temperature) but below the melting temperature, $T_\text{c} < T < T_\text{m}$. At these temperatures, the atoms are more mobile and gain sufficient energy to settle into the energetically favourable ordered state [95]. Because this reorganisation is a kinetic process, the crystallisation pulse must last long enough for the full atomic rearrangement to complete; crystallisation is therefore generally slower than amorphisation.

The kinetics of crystallisation depend strongly on whether the material is nucleation-dominated or growth-dominated [95]. In nucleation-dominated materials (e.g., GST), many small crystalline nuclei form throughout the amorphous matrix. These nuclei then grow until they impinge on one another, eventually resulting in complete crystallisation. In growth-dominated materials (e.g., AIST), only a few nuclei form, and crystallisation proceeds primarily through the lateral expansion of these nuclei. Because the number of nuclei is small, the time required for a single nucleus to grow and cover the full material volume is relatively long, making growth-dominated PCMs slower to crystallise. Sb-rich PCMs tend towards growth-dominated crystallization [123, 210], meaning the low-loss PCMs Sb_2S_3 and Sb_2Se_3 are expected to be slower than GST. According to Müller *et al.*, bond stiffening due to an increased number of shared electrons between atoms is the leading cause

of the slow switching of Sb_2Se_3 [210]. However, it is worth noting that, from an operational standpoint, the switching mechanisms of the low-loss materials remain similar to those of GST, so we can leverage the practical know-how from GST switching for these materials as well.

The crystal structures of the PCMs presented in this thesis are shown in Fig. 3.3. GST crystallises in the trigonal $\bar{P}3m1$ space group, as seen in Fig. 3.3a. The structure is two-dimensional and consists of a GST sheet oriented in the (001) direction. In the structure, Ge^{2+} is bonded to six Te^{2-} atoms, forming $GeTe_6$ octahedra that share corners with three equivalent $SbTe_6$ octahedra. Furthermore, the $GeTe_6$ octahedra shares edges with three equivalent $SbTe_6$ octahedra and six equivalent $GeTe_6$ octahedra. Both Sb_2Se_3 and Sb_2S_3 adopt stibnite-type structures and crystallise in the orthorhombic *Pnma* space group, as seen in Fig. 3.3b,c. The structures are described as layered, comprising two Sb_2Se_3 (or Sb_2S_3) sheets per unit cell stacked along the (001) direction.

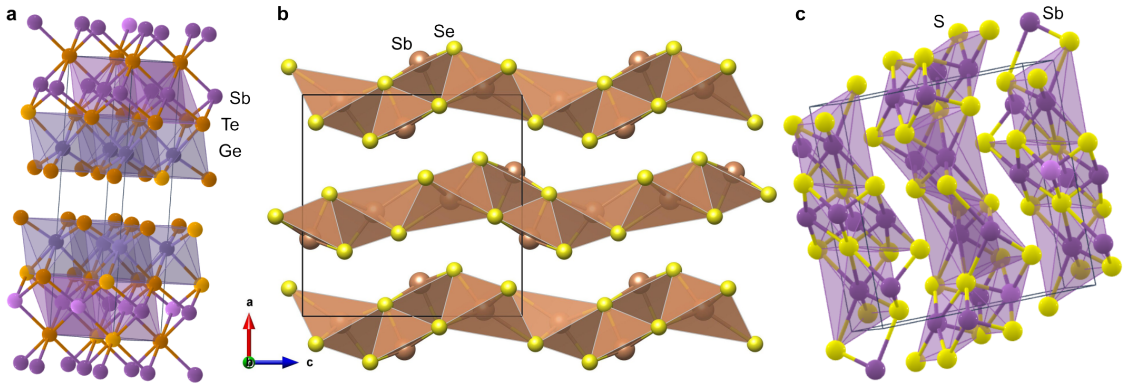

Figure 3.3: Crystal structures of the PCMs in this thesis. **a**, $Ge_2Sb_2Te_5$. **b**, Sb_2Se_3 with the unit cell indicated by the square. **c**, Sb_2S_3. Figures **a** and **c** are from The Materials Project webpage (`https://next-gen.materialsproject.org`; codes mp-3534 and mp-2809, respectively) under a Creative Commons licence CC BY 4.0. Figure **b** by Xmineralx, Wikimedia Commons, file: `Structure of Sb2Se3.jpg`, modified under a Creative Commons licence CC BY-SA 4.0.

3.7 Neural network theory

In neural networks, data flows from the neurons in the input layer, through the hidden layers and to the output layer (Fig. 3.4a). Every two neurons, *a* and *b*, are

connected across layers by a weight, w, representing the connection from a to b. The weight has a numerical value that represents the excitatory or suppressive effect (positive or negative sign) and magnitude (higher or lower value) of this connection. After the signals have been weighted, they are summed and passed through a non-linear activation function, for instance, a rectified linear unit (ReLU) or sigmoid function. To give the model more flexibility, a bias term that shifts the activation function's output is often introduced. The output from the activation function is the input to neurons in the next layer. The process explained here is shown in Fig. 3.4b.

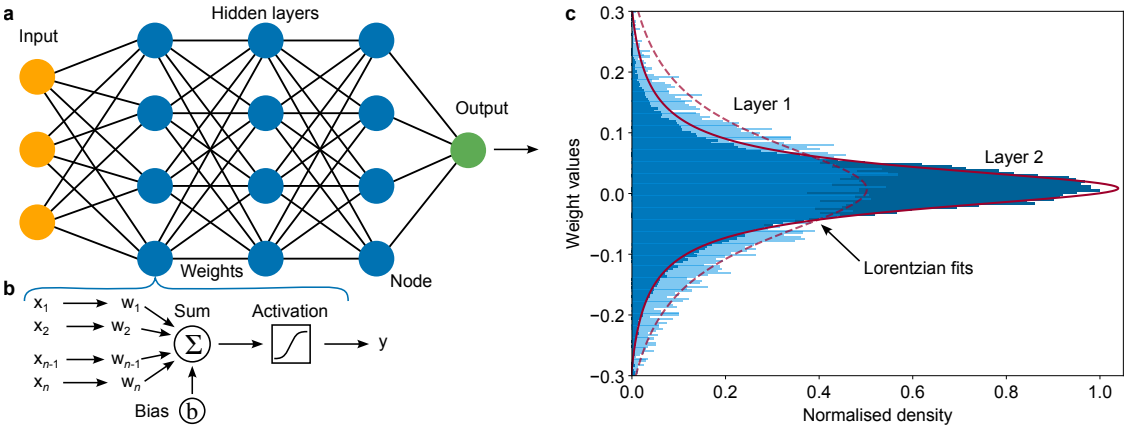

Figure 3.4: **a**, Schematic of a neural network with three hidden layers, with weights represented by lines and nodes by circles. **b**, Schematic of the computation performed at each node. **c**, Weight distribution from two of the layers of a fully-connected feedforward neural network for image recognition.

3.7.1 Weight quantisation in neural networks

In real systems, each stored number (for weights and activations) has only a limited number of obtainable values, given by the number of bits used to represent the value. The number of bits is often so large that we can almost treat the obtainable values as continuous. However, if the values are discrete, we refer to the number as quantised.

Weight quantisation in neural networks involves reducing the precision of the network's weight parameters. This process can significantly reduce the computational and memory requirements of neural networks, making them feasible for deployment on resource-constrained devices. By representing weights with lower bit-widths, such as using 8-bit integers instead of 32-bit floating-point numbers, quantisation reduces

the model size and increases inference speed. Furthermore, quantisation is often a natural choice for some hardware accelerators, such as those in edge computing devices, which are often optimised for lower-precision operations. However, gains in computational efficiency usually come at the cost of reduced model accuracy.

Generally speaking, the approach for quantisation is as follows: the weight values are normalised to the continuous range $[-1, 1]$, and the obtainable values are discretised to 2^N values, where N is the number of bits. Quantisation can be done in a uniform or non-uniform way. In uniform quantisation, the discretised values are evenly spaced on the number line, while in non-uniform quantisation for neural networks, the discretised values are normally more closely spaced around 0. Some works even perform extreme quantisation with only three (ternary; -1, 0, 1) and two (binary; 0, 1) obtainable values [211, 212].

It is well known that the distribution of weights in many neural networks, e.g., CNNs for image recognition, follows a bell-shaped curve centred around a mean value with fewer weights at extreme values (Fig. 3.4c) [213, 214], and the same is true also for large language models (LLMs) [215, 216]. This phenomenon occurs because most neural connections in big networks are not influential—if they were, the model would be overfitting. Owing to this behaviour of neural networks, it is logical to employ non-uniform quantisation with greater emphasis (i.e., finer spacing) on values closer to the mean. I will investigate this further in Chapter 7.

Weight setting in analogue neural networks

In analogue systems, setting the weight requires a digital-to-analogue converter (DAC) that converts the digital value to a discrete analogue voltage. Naturally, the analogue weight values are discretised into 2^N levels (due to the working principle of the DAC), meaning we could argue that analogue neural networks are inherently quantised. However, if N is high, we can treat the weight value as quasi-continuous, and we do not have to worry about errors in the expected vs. measured output, meaning these systems are considered not quantised. Because the power consumption of DACs is usually said to grow exponentially with the number of bits [217, 218], reducing the

bit-width of the DACs can provide us with massive energy savings.[3] Furthermore, in many analogue systems, such as photonic crossbar arrays, there is often already limited precision in setting the weights; consequently, moderate weight quantisation enables substantial energy savings without significantly degrading accuracy.

3.7.2 Working principle of photonic crossbar array

The photonic crossbar array is one of the photonic computing architectures introduced in Section 2.4.4 and the one used in this work. The working principle of the photonic crossbar array is as follows: the vector is encoded as light amplitudes (with values from 0 to 1) entering the rows of the matrix, where the matrix elements are encoded as the transmission of absorptive components (for example, PCMs, with values from 0 to 1). The value 0 (1) is defined as the minimum (maximum) intensity or transmission of light. If we want negative values for the matrix weights, we can simply define a reference point (typically the middle value) as 0, the maximum transmission as 1, and the minimum as -1; however, this definition requires a reference measurement per computation. Alternatively, complex-valued operations can be implemented in coherent photonic crossbar arrays by using a phase shifter in conjunction with the absorptive component, which allows for negative numbers [220]. The interaction between individual light inputs and individual absorptive components can be considered a multiplication, the output light intensity out of each column of the matrix can be considered a dot product between two vectors (i.e., the input light vector times the column of absorptive elements), and therefore the light outputs of all columns contain the matrix–vector product [4].

3.8 Sensor management and control theory

Sensor management is the process of managing and coordinating multiple sensors in a resource-constrained environment to accomplish specific tasks, with sensor fusion (i.e., combining data from two or more sensors to reduce the uncertainty of the

[3] Although the device architecture, the number of channels, and the sample rate impact the total power consumption of DACs and ADCs more than the number of bits alone [219].

resulting information) constituting one of its key components. Sensor management is a challenging problem for many reasons; for example, there are often several trade-offs in terms of competing—and sometimes conflicting—goals, and the number of potential actions grows exponentially with the number of sensors and objectives. Potential improvements in more efficient sensor management lie in the realms of software-based solutions, such as fast optimisation methods, and hardware-based solutions, such as parallel computing and in-memory computing to counter the limitation of data movement in conventional computers [221–223]. Software solutions fall outside the scope of this thesis; instead, the focus of this chapter and Chapter 5 will be on photonic in-memory computing with inherent parallelism, which holds promise of immense performance gains for sensor management.

To understand how photonic neuromorphic computing can deliver performance gains for defence-relevant sensor management tasks (Chapter 5), we first need to understand the equations that govern sensor management and its building blocks, with an emphasis on the constituents of simple, lab-deployable models. For sensor fusion, the different algorithms can be classified into fusion based on probabilistic models (e.g., evidence theory), fusion based on least squares methods (e.g., Kalman filtering), or intelligent fusion (e.g., neural networks) [224]. The equations governing sensor management and other defence-relevant tasks relevant to this work are presented here. In the following subsections, I describe the Kalman filter and how an MPC problem can be solved using an iterative approach.

3.8.1 Kalman filter implementation

One of the most well-known algorithms in control theory, the Kalman filter, takes the following recursive form, as described in more detail in Ref. 225:

$$\begin{cases} P_{0,0} = \mathrm{Var}(\mathbf{x}_0) \\ P_{k,k-1} = A_{k-1} P_{k-1,k-1} A_{k-1}^\top + \Gamma_{k-1} Q_{k-1} \Gamma_{k-1}^\top \\ G_k = P_{k,k-1} C_k^\top \Big(C_k P_{k,k-1} C_k^\top + R_k \Big)^{-1} \\ P_{k,k} = \Big(I - G_k C_k \Big) P_{k,k-1} \\ \hat{\mathbf{x}}_{0|0} = E(\mathbf{x}_0) \\ \hat{\mathbf{x}}_{k|k-1} = A_{k-1} \hat{\mathbf{x}}_{k-1|k-1} \\ \hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + G_k \Big(\mathbf{v}_k - C_k \hat{\mathbf{x}}_{k|k-1} \Big) \\ k = 1, 2, \dots \end{cases} \tag{3.51}$$

I do not intend to dive into the meaning of these equations (for more information, see Chapter 2 of Ref. 225), but it is clear from them that the Kalman filter contains several matrices (e.g., the state transition matrix A_k, the observation matrix C_k, the Kalman gain matrix G_k, etc.), some of which must be computed recursively, and many repeated MVMs and matrix multiplications. Because photonic neuromorphic computers can perform MVM with low latency and high parallelism, the Kalman filter appears to be an ideal candidate for photonic computing.

3.8.2 Quadratic programming for model predictive control

The MPC algorithm allows for optimising an action at the current time step while accounting for future time steps and system constraints, and is used in power electronics, automated driving, drone guidance, and related areas. The algorithm considers the effects of current and future actions on future time steps, thereby outperforming traditional proportional–integral–derivative (PID) control methods [226].

It is possible to frame the MPC problem as a quadratic programming (QP)

problem, which is what I will show here. First, the state matrices are given by

$$\begin{aligned} x_{k+1} &= Ax_k + Bu_k, \\ y_k &= Cx_k, \end{aligned} \tag{3.52}$$

where $x_k \in \mathbb{R}^{n_x}$ is the state at time step k, $u_k \in \mathbb{R}^{n_u}$ is the control input, and $y_k \in \mathbb{R}^{n_y}$ is the output. We want to predict $y_{k+1}, y_{k+2}, \ldots, y_{k+N}$ over a finite prediction horizon of N steps. By expanding recursively, we get

$$y_{k+1} = CAx_k + CBu_k, \tag{3.53}$$

for the first iteration,

$$\begin{aligned} x_{k+2} = Ax_{k+1} + Bu_{k+1} &= A(Ax_k + Bu_k) + Bu_{k+1} \\ &= A^2x_k + ABu_k + Bu_{k+1}, \\ y_{k+2} = Cx_{k+2} &= CA^2x_k + CABu_k + CBu_{k+1}. \end{aligned} \tag{3.54}$$

for the second time step, and in the general case

$$y_{k+i} = CA^i x_k + \sum_{j=0}^{i-1} CA^{i-1-j} B\, u_{k+j}, \quad \text{for } i = 1, \ldots, N. \tag{3.55}$$

In compact matrix form, we can express the prediction in the following way:

$$\mathbf{y} = \begin{bmatrix} y_{k+1} \\ y_{k+2} \\ \vdots \\ y_{k+N} \end{bmatrix} = \begin{bmatrix} CA \\ CA^2 \\ \vdots \\ CA^N \end{bmatrix} x_k + \begin{bmatrix} CB & 0 & \cdots & 0 \\ CAB & CB & \cdots & 0 \\ \vdots & & \ddots & 0 \\ CA^{N-1}B & CA^{N-2}B & \cdots & CB \end{bmatrix} \mathbf{u} = Fx_k + H\mathbf{u} \tag{3.56}$$

with $\mathbf{u} = [u_k, u_{k+1}, \cdots, u_{k+N-1}]^\top$.

We define a quadratic objective function to minimise the output error (the first term) and control effort (the second term):

$$J(\mathbf{u}) = \frac{1}{2} \|\mathbf{y} - \mathbf{r}\|^2 + \frac{1}{2} \|\mathbf{u}\|^2, \tag{3.57}$$

where $\mathbf{r}$ is the desired reference trajectory. Substituting $\mathbf{y} = Fx_k + H\mathbf{u}$, we get:

$$J(\mathbf{u}) = \frac{1}{2}\mathbf{u}^\top (H^\top H + I)\mathbf{u} + \mathbf{u}^\top H^\top (Fx_k - \mathbf{r}) + \text{const.} \tag{3.58}$$

So, the quadratic programming problem is:

$$\min_{\mathbf{u}} \ \frac{1}{2}\mathbf{u}^\top P\mathbf{u} + q^\top \mathbf{u}, \tag{3.59}$$

with the Hessian matrix and gradient vector given by, respectively,

$$P = H^\top H + I \tag{3.60}$$

and

$$q = H^\top (F x_k - \mathbf{r}). \tag{3.61}$$

Our system will, in most cases, have constraints, and the ability to respect constraints is one of the advantages of using MPC compared with other methods. We want to enforce the acceleration $a_{\min} \leq u_k \leq a_{\max}$ and position constraints $\theta_{\text{low}} \leq y_k \leq \theta_{\text{high}}$. These translate to linear inequalities on $\mathbf{u}$ in the following manner:

$$G\mathbf{u} \leq h, \tag{3.62}$$

with G built row-by-row from standard basis vectors $\pm e_i^\top$ (e.g., $e_y^\top = [0\ 1\ 0]$ in three dimensions) for the input (control) constraints, with the corresponding rows in h being $a_{\max}$ and $-a_{\min}$, and rows $\pm H_i$ for the output constraints $\theta_{\text{high}} - F_i x_k$ and $-\theta_{\text{low}} + F_i x_k$ in h, such that

$$G\mathbf{u} = \begin{bmatrix} e_0^\top \\ -e_0^\top \\ H_0 \\ -H_0 \\ \vdots \\ e_{N-1}^\top \\ -e_{N-1}^\top \\ H_{N-1} \\ -H_{N-1} \end{bmatrix} \mathbf{u} \leq \begin{bmatrix} a_{\max} \\ -a_{\min} \\ \theta_{\text{high}} - F_0 x_k \\ -\theta_{\text{low}} + F_0 x_k \\ \vdots \\ a_{\max} \\ -a_{\min} \\ \theta_{\text{high}} - F_{N-1} x_k \\ -\theta_{\text{low}} + F_{N-1} x_k \end{bmatrix} = h. \tag{3.63}$$

Here, H_i (F_i) refers to the i-th row of H (F). The whole QP problem now looks like

$$\begin{aligned} &\min_{\mathbf{u}} \ \frac{1}{2}\mathbf{u}^\top P \mathbf{u} + q^\top \mathbf{u} \\ &\text{subject to} \ \ G\mathbf{u} \leq h. \end{aligned} \tag{3.64}$$

In this work, I will not delve into the many computational methods for solving MPC problems. For more information about MPC algorithms, the reader is directed to Ref. 227.

3.8.3 Primal–dual gradient (Langrangian) method

One way to solve the QP problem is to use the primal–dual gradient method, which finds the optimal point by iteratively solving the associated primal and dual updates. Forming the Lagrangian of the QP problem (Eq. (3.59)) by adding the inequality constraints (Eq. (3.62)) multiplied by their Lagrange multipliers gives

$$\mathcal{L}(u, \lambda) = \frac{1}{2} u^\top P u + q^\top u + \lambda^\top (Gu - h), \tag{3.65}$$

where $\lambda \geq 0$ is the vector of Lagrange multipliers associated with the inequality constraints. Note: in this section, λ refers to the Lagrange multiplier, not the wavelength.

The gradients of this equation are

$$\nabla_u \mathcal{L}(u, \lambda) = Pu + q + G^\top \lambda \tag{3.66}$$

and

$$\nabla_\lambda \mathcal{L}(u, \lambda) = Gu - h. \tag{3.67}$$

Due to the nature of the saddle-point structure of the Lagrangian, we want to subtract $\nabla_u \mathcal{L}(u, \lambda)$ from u and add $\nabla_\lambda \mathcal{L}(u, \lambda)$ to λ to find the optimum point, i.e.,

$$u_{k+1} = u_k - \eta \nabla_u \mathcal{L} = u_k - \eta P u_k - \eta q - \eta G^\top \lambda_k = (I - \eta P) u_k - \eta q - \eta G^\top \lambda_k \tag{3.68}$$

and

$$\lambda_{k+1} = \max(0, \lambda_k + \eta \nabla_\lambda \mathcal{L}) = \max(0, \lambda_k + \eta G u_k - \eta h), \tag{3.69}$$

where η is the step size. The value of u_k at steady state is used as the control input for that time step in the MPC algorithm.

Here, we identify the following components: the vectors u and λ, which are iteratively updated until a steady-state solution is found for each time step; the vectors q and h, which are updated after each action has been implemented (once every time step); and three matrices, $(I - \eta P)$, ηG, and $\eta G^\top$, which are all constant under the same conditions and constraints.

In theory, there is no difference between theory and practice. But in practice, there is.

— Benjamin Brewster

4

Device simulation, fabrication, and measurement

Contents

4.1 General device fabrication steps 82
4.1.1 Basics 82
4.1.2 Alignment 83
4.1.3 Photonic integrated circuit elements 83
4.2 Fabrication steps for FPCAs 84
4.2.1 Ion implantation 84
4.2.2 Electrodes and ohmic contacting 86
4.2.3 Phase-change material deposition 87
4.3 Fabrication steps for neural network weights 88
4.3.1 Separation layer 89
4.3.2 Heaters and electrodes 90
4.4 Phase-change materials: Experimental basics 92
4.4.1 Phase-change material measurements 92
4.4.2 Phase-change material simulation 96
4.4.3 Phase-change material considerations 100
4.5 Experimental measurement setup 101
4.5.1 Optical measurement setup 102
4.5.2 Electrical measurement setup 104
4.5.3 Optical loss measurements 106
4.5.4 Measurements of directional couplers 107

This chapter outlines the methods used to simulate, fabricate, and measure the nanophotonic elements and PICs in this thesis. First, I describe the general

fabrication procedure for the devices demonstrated in this work (Section 4.1), followed by fabrication steps specific to individual projects (Sections 4.2 and 4.3). Then, I discuss the basic experimental measurements and simulations of the low-loss PCM Sb_2Se_3 (Section 4.4). Finally, I present the optical and electrical measurement setup, along with supporting measurements of DCs, as these are standard structures used throughout this thesis (Section 4.5).

4.1 General device fabrication steps

In this section, I present the general process flow for fabricating PICs in our cleanroom. First, I introduce the standard cleanroom procedures performed on all chips before presenting the general process flow, which includes patterning of alignment marks and etching of optical circuits (i.e., grating couplers and waveguides). In subsequent sections, I will cover the deposition of active elements and aid structures, such as PCM-elements and metal pads.

4.1.1 Basics

All devices in this work were manufactured on SOI wafers from University Wafer, with 220 nm silicon (Si) on a 2 µm buried oxide (BOX) layer and a total wafer thickness of 675 µm. The orientation of the lightly-boron-doped Si was $\langle 100 \rangle$ and the resistivity was between 10 Ω cm and 20 Ω cm. The wafers were diced into individual 10 mm × 10 mm chips for simpler handling.

Before each step, all samples were immersed in a beaker with acetone under ultrasonic agitation for 5 min, followed by 5 min in isopropyl alcohol (IPA) under ultrasonic agitation. The samples were then dried, first with a nitrogen blow-dry and then on a hotplate at >150 °C for >5 min.

For all layers, we use conventional cleanroom techniques with electron-beam lithography (EBL) and e-beam resist for patterning. Our EBL system (JEOL JBX-5500FS) has a 50 kV electron beam and two different modes, mode 2 and mode 4, with maximum field sizes of 1000 µm × 1000 µm and 100 µm × 100 µm, respectively. mode 2 has a minimum scanning step (or shot pitch) of 5 nm, 10 times larger than

that of mode 4. Therefore, for waveguides and other fine features <1 µm, we use mode 4, but for larger structures, such as metal pads, we generally use mode 2 due to its larger field size and faster writing speed. EBL designs were produced in Python using the `gdshelpers` library created and distributed by the Responsive Nanosystems lab at the University of Münster (led by Professor Wolfram Pernice), unless explicitly stated otherwise. In the designs, the waveguide width was 500 nm.

4.1.2 Alignment

The first step in all fabrication processes is marker patterning, which we use to align subsequent layers. A layer of AR-P 6200 (CSAR 62; a positive e-beam resist) was spin-coated at 4000 RPM for 1 min and subsequently soft-baked at 150 °C for 3 min. With the EBL system in mode 2, we used a beam current of 10 nA to expose the resist with a dose of 180 µC/cm^2. After lithography, the samples were developed in AR 600-546 developer, methyl isobutyl ketone (MIBK), and IPA for 30 s, 15 s and 15 s, respectively. The sample is now covered by a layer of resist with openings for alignment marks. Next, 5 nm of chromium (Cr) and 75 nm of gold (Au) was deposited onto the sample by thermal evaporation in an Edwards 306 thermal evaporator system. The Cr acts as an adhesion layer for the Au, which gives high contrast in the scanning electron microscope (SEM) images used for subsequent alignment in the EBL system. Finally, the samples were immersed in REM-700 remover overnight at 50 °C, which removes the resist and metal layer covering the resist in a lift-off process, leaving only the alignment marks. All samples in this work have global alignment marks for rough alignment, while some samples also have local alignment marks near each design for improved alignment accuracy. For samples where we used local alignment marks, the writing was performed in a separate step from the global alignment marks to ensure higher accuracy.

4.1.3 Photonic integrated circuit elements

The second layer of the fabrication procedure is the patterning of the PIC with waveguides and grating couplers. Again, I used the positive resist CSAR 62 for

patterning. However, for the EBL step, I used mode 4, a beam current of 100 pA, and a dose of 160 µC/cm^2. For this step, using negative resist would significantly reduce write time, since we need to write buffer zones on the sides of the waveguide with positive resist rather than simply writing the actual waveguide with negative resist. However, previous work in our group has shown that negative resist is more prone to charging effects that limit alignment precision and give rounded waveguide edges [228]. After development, the wafers were dry-etched in a reactive-ion etching (RIE) system (Oxford Instruments PlasmaPro 80) with a mixture of 15 SCCM SF_6 and 40 SCCM CHF_3 gases at an RF power of 100 W. After about 1 min of etching,[1] 110 nm to 120 nm of the Si was etched, leaving what is called a rib-waveguide with a 110 nm to 100 nm thick slab. For some of my devices, the etching process was immediately followed by oxygen plasma at 100 SCCM and 200 W for 5 min to remove the resist. In others, the sample was immersed in REM-700 remover at 50 °C for a couple of hours until the resist was completely removed.

4.2 Fabrication steps for FPCAs

For the FPCA in Chapter 6, I used Sb_2Se_3 phase shifters integrated with microheaters to switch the PCM. The heaters are formed by ion implantation, contacted with aluminium (Al) pads, and partially covered with PCM. In this section, I explain these process steps in more detail, which were performed sequentially after PIC patterning. Most steps are shown in Fig. 4.1. For every patterning step, I performed cleaning, spin-coating, soft-baking, EBL, and development, and subsequently resist removal. For the resist removal after waveguide patterning, I used oxygen plasma cleaning for 5 min, as described in Section 4.1.3.

4.2.1 Ion implantation

The windows for ion implantation were opened with EBL in a 3 µm thick triple-layer of polymethyl methacrylate (PMMA) positive resist (one layer of 495 PMMA A8, two of 950 PMMA A8) to protect the rest of the sample from the high-energy

[1] The time varied throughout the duration of my PhD and had to be calibrated before each run.

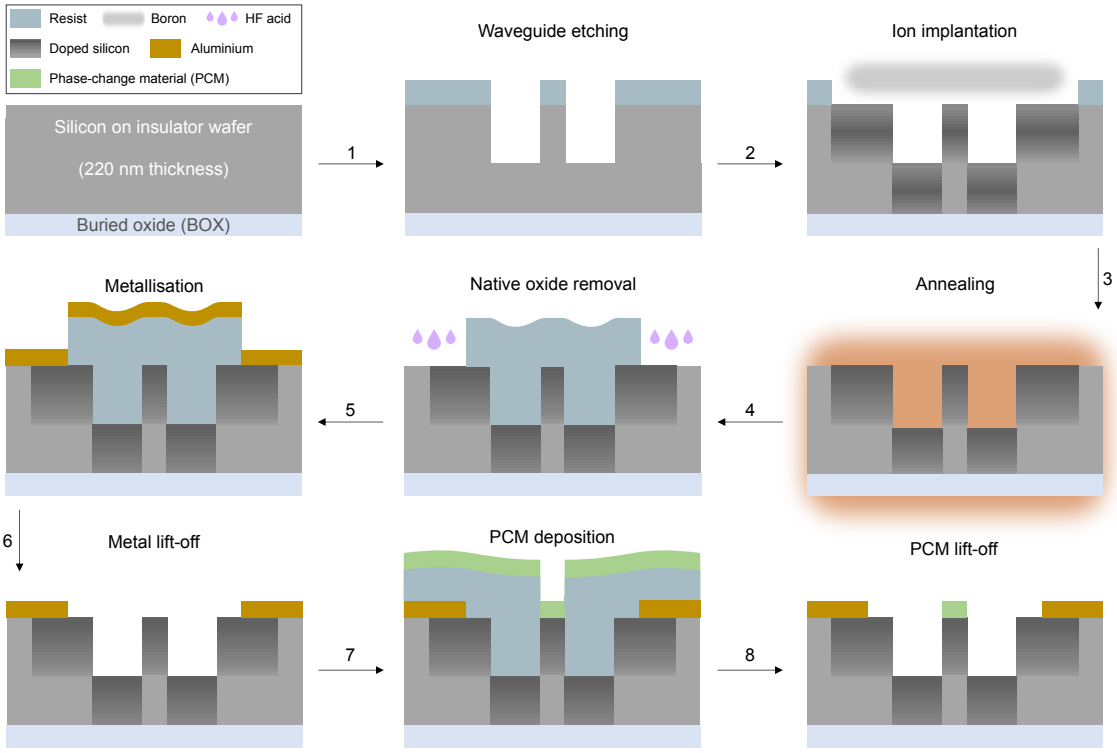

Figure 4.1: Schematic of the fabrication procedure for the FPCA samples. The SOI sample undergoes the following steps: (1) waveguide patterning with EBL and etching with RIE, (2) opening of windows with EBL and ion implantation with boron, (3) tube-furnace annealing for dopant activation, (4) opening of windows with EBL for subsequent metal deposition and etching of native oxide with hydrofluoric acid (HF acid), (5) e-beam evaporation of Al, (6) lift-off in acetone, (7) patterning with EBL and sputtering of PCM (Sb_2Se_3) and capping layer ($ZnS–SiO_2$, not shown) on the doped waveguides, and (8) lift-off in acetone. The spin-coating before steps 1, 4, and 7, the resist removal and spin-coating before step 2, and the resist removal and piranha clean before step 3 are not shown.

particles. Individual resist layers were spin-coated at 1500 RPM and 3000 RPM for 30 s each and soft-baked at 150 °C for 5 min. To write the windows, we used mode 4, a beam current of 3 nA, and a dose of 650 $\mu C/cm^2$. We resorted to mode 4 due to the significant thickness of the resist layer, which made alignment in mode 2 difficult. Development was performed in a 3:1 mixture of IPA and MIBK for 1 min, immediately followed by nitrogen blow-dry.

Ion implantation was performed by Dr Luke Antwis at the Surrey Ion Beam Centre at the University of Surrey. Boron ions were implanted with a dose of 3×10^{15} ions/cm^2 at an energy of 10 keV. These are the same parameters as in Refs. 195 and 229 and result in very heavily p-doped (p^{++}-doped) Si. Following

doping, the sample was immersed in acetone overnight at 67 °C to remove the PMMA resist layer. Because there could still be resist or other organic residue on the sample, the sample was immersed for 5 min in a piranha solution made from a 3:1 mixture of H_2SO_4 and 30 % H_2O_2. Following the piranha clean, the sample was immersed in three different beakers of deionised (DI) water.

Annealing

After ion implantation, we need to activate the boron dopants through annealing. The elevated heat from the process also mitigates some of the damage to the Si caused by ion implantation. In Refs. 195 and 229, annealing of the SOI samples was performed in a rapid thermal annealing (RTA) system. Unfortunately, the RTA system previously used by our group was miscalibrated, resulting in damage to some of the samples. Because of limited and expensive samples, a safer approach was to use a tube furnace for annealing. The samples in Chapter 6 were annealed at 920 °C for 5 min in a Lenton Thermal Design LTF 12 tube furnace under argon flow, before turning off the furnace and letting the samples reach room temperature inside the tube.

Heater loss

The doped Si introduces a significant loss to the optical waveguide. Figure 4.2 shows measurement data of the effect of the doping, with an estimated loss of $0.108\,\mathrm{dB\,\mu m^{-1}}$. For more information on this measurement and setup, see Section 4.5.3.

4.2.2 Electrodes and ohmic contacting

We deposited metal electrodes to contact the microheaters in a lift-off process. A 2 µm thick double-layer of positive resist (495 PMMA A8 and 950 PMMA A8) was spin-coated, and windows were opened in it using EBL, with all parameters kept the same as for the writing of doping windows in Section 4.2.1. Before metal deposition, the native oxide on top of the Si was removed in the openings by immersing the sample in 5 % HF acid for 5 s, to ensure ohmic contact between the metal pads and

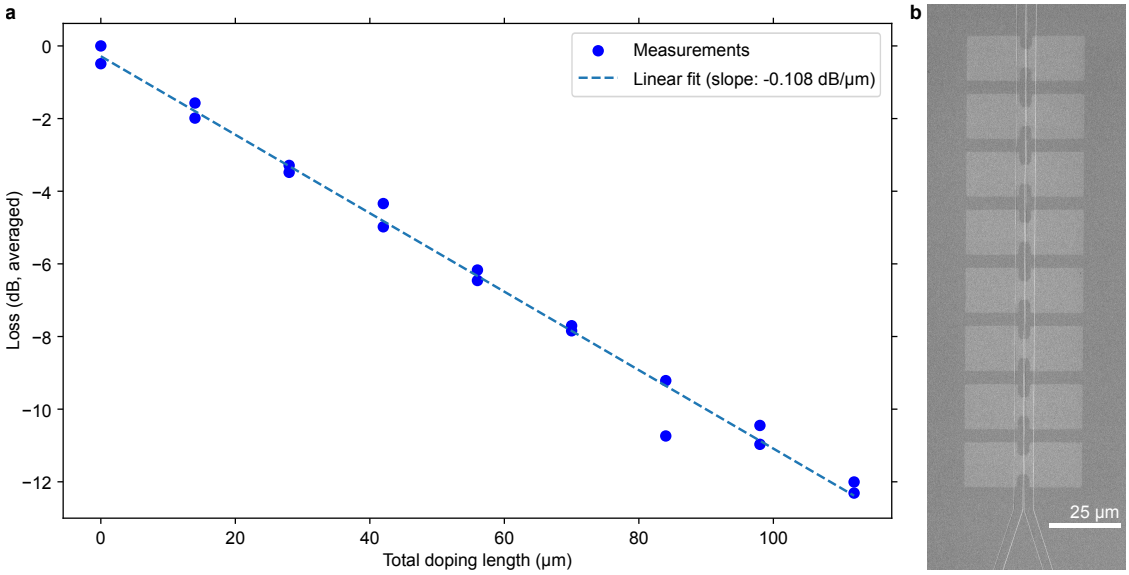

Figure 4.2: **a**, Optical loss measurements of multiple 7 µm patches of boron-implanted Si. **b**, Scanning electron micrograph of a waveguide with eight patches of doped Si. Each patch is 7 µm long in the waveguide region and larger on the sides to contact the metal pads.

the doped silicon. Subsequently, 300 nm of Al was deposited by e-beam evaporation at 130 mA in a Plassys MEB550S e-beam deposition system. Aluminium was used because of its cost-effectiveness and because it diffuses into p-type Si at elevated temperatures, which enhances the ohmic contacting. The sample was immersed in acetone overnight at 67 °C to remove the excess metal.

4.2.3 Phase-change material deposition

The final step of the fabrication procedure is deposition of PCM in a lift-off process. A 400 nm thick double-layer of positive resist (495 PMMA A6 and 950 PMMA A3) was used. Each layer was individually spin-coated at 6000 RPM for 1 min and soft-baked on a hotplate at 180 °C for 10 min. Windows were opened using EBL in mode 4, with 100 pA and 620 µC/cm^2, and subsequently developed in 3:1 IPA:MIBK for 1 min, followed by a nitrogen blow-dry.

A 30 nm or 60 nm thick layer of Sb_2Se_3 was deposited by RF-magnetron sputtering in an AJA International ATC ORION 5 system by striking a two-inch Sb_2Se_3 target with a composition of 40 atomic% Sb and 60 atomic% Se and purity of 99.995 % from Pioneer Materials. Different layer thicknesses were used in preliminary

measurements, while 60 nm was used for the devices in Chapter 6. The sputtering rate was ∼5.3 nm/min, giving a total sputtering time of 11 min 20 s for the 60 nm layer and half of that for the 30 nm layer. The target was sputtered with a power of 30 W and an argon flow of 3 SCCM in the chamber at a pressure of 3×10^{-3} torr (the base pressure of the chamber before deposition was 3×10^{-8} torr). The estimated size of the sputtering region was ∼30 cm^2 [230]. The sample holder was rotating to achieve uniform deposition and was kept at room temperature during sputtering. A 20 nm thick capping layer of $ZnS-SiO_2$ was deposited by sputtering with a power of 50 W immediately after Sb_2Se_3 deposition within the same vacuum, to impede oxidation and delamination of the PCM. The sputtering rate of the $ZnS-SiO_2$ layer was ∼1.4 nm/min, giving a total sputtering time of 14 min 17 s. The capping layer was primarily chosen out of convenience—its sputtering rate is higher than that of pure SiO_x, and it exhibits lower loss than ITO in the wavelength range of interest (and is not conductive). The sample was immersed in acetone overnight at 67 °C to lift off the excess material.

4.3 Fabrication steps for neural network weights

The phase shifters used in Chapter 7 are not based on PCMs but are TO metal heaters. Because metal heaters lead to significant optical loss if deposited immediately on top of the waveguides, we must separate the waveguide and heater by a transparent separation layer, here, silicon oxide (SiO_x) deposited by plasma-enhanced chemical vapor deposition (PECVD). The process steps described in this section were performed sequentially after the patterning of the PIC. Figure 4.3 shows schematics of the fabrication steps (for a top-view of the layers, see Fig. 4.4). For every patterning step, I performed cleaning, spin-coating, soft-baking, EBL, and development, and subsequently resist removal. After waveguide etching, the sample was immersed in REM-700 remover at 50 °C until the resist was completely removed.

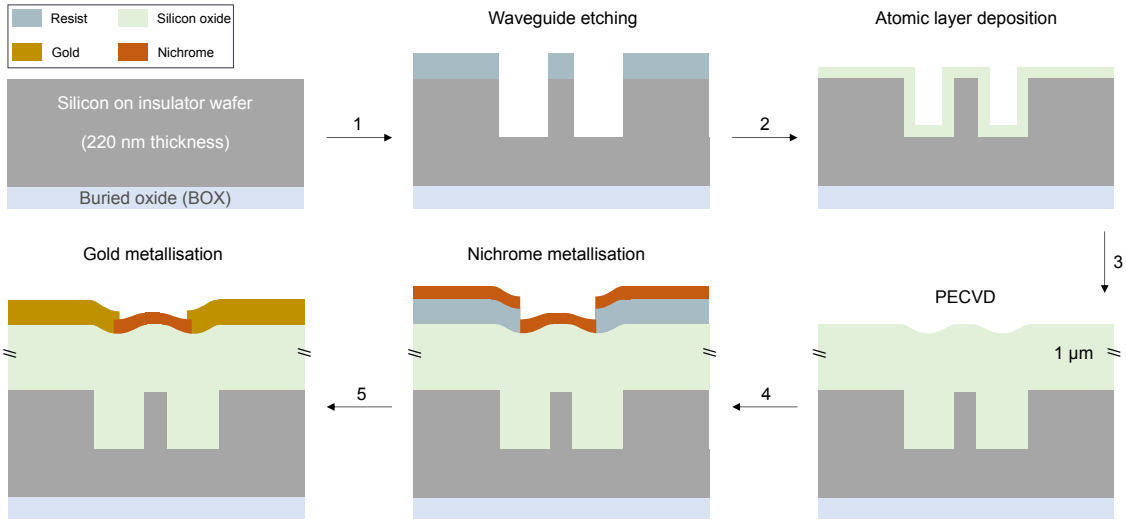

Figure 4.3: Schematics of the fabrication procedure for the neural network weights. The SOI sample goes through the following steps: (1) waveguide patterning with EBL and etching with RIE, (2) atomic layer deposition (ALD) of 20 nm SiO_x, (3) PECVD of $\sim$1 µm SiO_x, (4) opening of windows with EBL and sputtering of nickel–chromium (nichrome), (5) opening of windows with EBL, thermal evaporation of Au, and lift-off. Not shown in the figure are the spin-coating before steps 1, 4, and 5, resist removal and plasma cleaning before step 2, and lift-off after step 4.

4.3.1 Separation layer

For SOI devices, SiO_x is a natural choice for the separation layer. To properly separate the metal heaters from the waveguides, we deposited approximately 1 µm of SiO_x in a PECVD process. However, because the PECVD process requires elevated chamber temperatures exceeding 300 °C, a thin SiO_x layer was first deposited by ALD to prevent contamination of the PECVD chamber, in particular from the Au alignment markers. Since ALD operates at significantly lower temperatures, the ALD chamber is inherently less susceptible to such contamination. Nevertheless, to further protect the ALD system, the samples were subjected to 1 min of oxygen plasma cleaning in a Henniker Plasma HPT-200 plasma treatment system (40 kHz, 200 W, 50 SCCM), before the deposition of 20 nm SiO_x in a Veeco Savannah S200 Atomic Layer Deposition System at 150 °C.

For the PECVD, we used two different systems in this thesis; an Oxford Instruments PlasmaPro 80 PECVD (in our cleanroom) for samples with design 1 in Section 4.3.2 and Chapter 7 and an STS PECVD system (at the London Centre for

Nanotechnology (LCN)) for samples with design 2 and 3. The former system was operated at a power of 20 W, a frequency of 13.56 MHz, and a temperature of 350 °C, with gases N_2O (710 SCCM) and 5 % SiH_4/95 % Ar_2 (170 SCCM), and a chamber pressure of 1000 mtorr. For the samples deposited in this system, we did not measure the final thickness; instead, we deposited for 20 min at an assumed deposition rate of 50 nm min^{-1} (based on prior experience), yielding approximately 1 µm. The second PECVD system was operated at a power of 60 W, a frequency of 380 kHz, and a temperature of 300 °C, with gases N_2 (165 SCCM), N_2O (1420 SCCM), and 5 % SiH_4/95 % Ar_2 (240 SCCM), and a chamber pressure of 550 mtorr. A deposition time of 9.5 min resulted in a film thickness of around 0.97 µm, as measured by ellipsometry, giving a rate of 102 nm min^{-1}.

4.3.2 Heaters and electrodes

The TO heater is made of a nichrome alloy with Au pads for contacting. First, we deposited nichrome in a lift-off process. A 2 µm thick double-layer of positive resist (495 PMMA A8 and 950 PMMA A8) was spin-coated with the same parameters as in Section 4.2.2. Next, windows were opened using the EBL system in mode 4, with a beam current of 100 pA and a dose of 650 µC/cm^2. Because of the smaller features of design 2, we instead used a double-layer of 495 PMMA A6 and 950 PMMA A3 with parameters for spin-coating and EBL as mentioned in Section 4.2.3 for both designs 2 and 3. Approximately 100 nm nichrome was deposited by RF-magnetron sputtering in an AJA International ATC ORION 5 system. The target was sputtered for 48 min at an assumed rate of 2.1 nm s^{-1}, with a power of 50 W and an argon flow of 3 SCCM in the chamber at a base pressure of 3×10^{-8} torr.

Finally, Au pads for contacting were deposited in a lift-off process. Windows were opened in a double-layer of 495 PMMA A8 and 950 PMMA A8, with all parameters of the spin-coating kept the same as in Section 4.2.2 and the EBL set to mode 4, 3 nA, and 650 µC/cm^2. 5 nm Cr and 300 nm Au was deposited onto the sample by thermal evaporation in an Edwards 306 thermal evaporator system. The excess metal was lifted off in acetone overnight at 67 °C.

Heater designs

In the above paragraphs, I refer to three designs used for the TO metal heaters presented in Chapter 7, which are labelled designs 1–3. GDS layouts of these designs are shown in Fig. 4.4a–c, respectively. The main difference between the designs is that in Fig. 4.4a, the heater only heats the coupling region, while in Fig. 4.4b,c, the heater also heats the other MZI arm. The aim of this design alteration is to offset the heating of the MZI upper arm—that leads to a phase-dependent transmission change of the RAMZI—by also heating the lower arm, with the aim that we will only see a change in the coupling into the ring. The difference between Fig. 4.4b,c is just the width of the nichrome heaters. For design 1, the heater in the coupling region was placed with its axis overlapping with the propagation axis of the upper MZI arm (Fig. 4.4d). In designs 2 and 3, we instead offset this heater by a distance $d = 0.6\,\mu\text{m}$ below MZI arm (Fig. 4.4e,f), as done in Refs. 231 and 232. The width of the heaters in design 2, $w_2 = 0.9\,\mu\text{m}$, is the same as in Refs. 231 and 232, while $w = 3\,\mu\text{m}$ in designs 1 and 3.

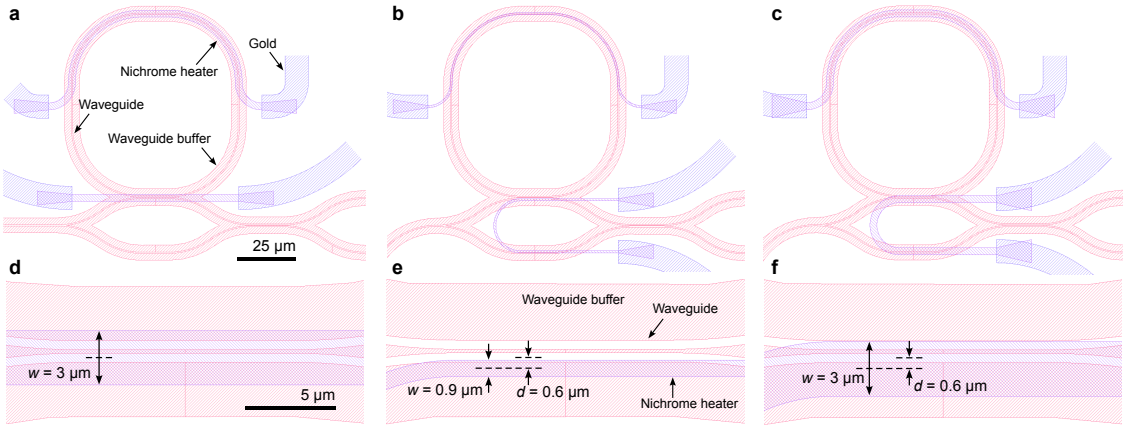

Figure 4.4: Different heater designs for the neural network weights. **a**–**c**, GDS designs of the three different RAMZI designs presented in Chapter 7. In **a**, the heater only heats the coupling region, while in **b** and **c** the heater also heats the other MZI arm. The difference between **b** and **c** is the width of the nichrome heaters. **d**–**f**, Zoom-in of the coupling region into the ring for design 1 (d), design 2 (e), and design 3 (f). The parameters w and d refer to the width of the nichrome heater and the offset from the centre of the MZI arm (lower waveguide in the coupling region), respectively. The Au metal connections are shown in blue, the nichrome heater in purple, and the waveguide buffer in pink.

4.4 Phase-change materials: Experimental basics

In this work, I use the recently discovered low-loss PCM Sb_2Se_3 as a phase shifter. Although previous works have both used and characterised Sb_2Se_3, there is no one-size-fits-all principle when it comes to the use of PCMs—therefore, it is crucial to understand how *our* material behaves. In this section, I present preliminary experimental measurements and simulations of Sb_2Se_3 on an SOI platform.

4.4.1 Phase-change material measurements

In this section, I show ellipsometry data on Sb_2Se_3 and experimentally determine changes in the effective refractive index and loss between the amorphous and crystalline states of Sb_2Se_3 on a waveguide.

Ellipsometry of Sb_2Se_3

The first experimental procedure we typically perform when using a new PCM is ellipsometry, to obtain the complex refractive index of the material. All measurements, analysis, and data presented in this subsection were performed together with Dr Guoce Yang. Ellipsometry measurements of a thin-film of Sb_2Se_3 were conducted using a J.A. Woollam RC2 ellipsometer and are presented in Fig. 4.5. The thickness was 180 nm, measured with a contact profilometer (Veeco DekTak 6M Stylus Profiler). To fit the data, we used two methods. The first used only the Cody-Lorentz model, while the second used two oscillators in the fitting model: the Cody-Lorentz and the Gaussian models. I will refer to these methods as the *basic* and *advanced* models, respectively. With the additional Gaussian term, the fitting gives a more reasonable κ value. We only fitted the curve for the wavelength range 400 nm to 1690 nm, because the model may not work well for wavelengths below 400 nm.

As can be seen in Fig. 4.5, the refractive index difference between the crystalline and amorphous states of our Sb_2Se_3 is $\Delta n = n_{\mathrm{cry}} - n_{\mathrm{amo}} \approx 0.65$ for the advanced model, with a wavelength dependency on Δn of only 0.014 between 1500 nm and 1600 nm. For the basic model, the same values are 0.75 and 0.012. We present some of the measurement results in Table 4.1. For the basic model, the extinction

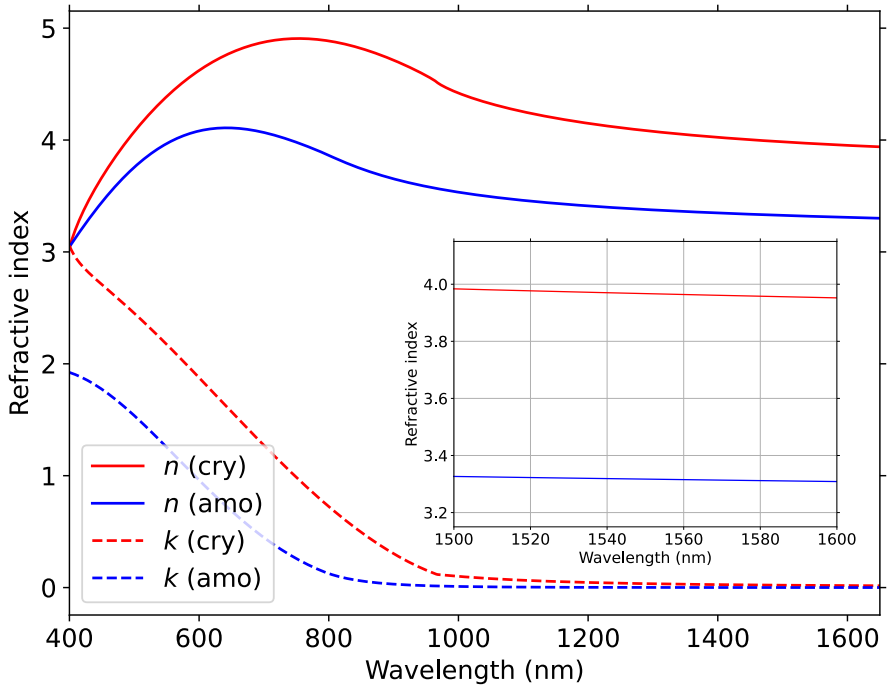

Figure 4.5: Ellipsometry measurements of a 180 nm thick layer of the PCM Sb_2Se_3, fitted with the advanced model. The inset shows that the real refractive index between 1500 nm and 1600 nm is almost constant.

coefficients of both states of the PCM go to 0 for all wavelengths above 1085 nm, while the advanced model gives $\kappa_{\text{cry}} \leq 0.021$ and $\kappa_{\text{amo}} \leq 0.0004$ for wavelengths above 1500 nm. We see that the refractive index of Sb_2Se_3 is closely matched to that of Si, $n_{\text{Si}} = 3.48$ at 1550 nm. As noted in Section 2.3.2, this is closer to Si than Sb_2S_3 with $n_{\text{amo}} = 2.712$ and $n_{\text{cry}} = 3.308$ [12].

Table 4.1: Overview of the refractive indices of Sb_2Se_3 at different wavelengths, for two different fitting models.

	Basic model: Cody-Lorentz				Advanced model: Cody-Lorentz and Gaussian				
Wavelength	n_{cry}	n_{amo}	Δn	κ	n_{cry}	n_{amo}	Δn	κ_{cry}	κ_{amo}
1500 nm	4.037	3.282	0.755	0	3.984	3.327	0.657	0.0212	0.0004
1550 nm	4.024	3.275	0.749	0	3.967	3.317	0.650	0.0193	0.0003
1600 nm	4.012	3.269	0.743	0	3.952	3.309	0.644	0.0176	0.0002

Effective refractive index of Sb_2Se_3

It is important to determine the complex refractive index of new materials; however, this information does not automatically reveal the exact characteristics of devices

based on those materials. To learn more about the device response, we can either run simulations (as in Section 4.4.2) or measure the loss and the effective refractive index experimentally (as done here).

To find the effective refractive index of the waveguide with PCM on top, $n_{\mathrm{eff,\,PCM}}$, we designed unbalanced MZIs with PCM in one arm. From Section 3.5.1 and Eq. (3.23), we have that the output varies sinusoidally with the change in phase. In an unbalanced MZI with a PCM patch of length L_{PCM} and effective refractive index $n_{\mathrm{eff,\,PCM}}$ in the lower arm and arm length difference $\Delta L = L_2 - L_1$, the term $(n_{\mathrm{eff},2}L_2 - n_{\mathrm{eff},1}L_1)$ in Eq. (3.23) becomes

$$(n_{\mathrm{eff,\,PCM}}L_{\mathrm{PCM}} + n_{\mathrm{eff},2}(L_2 - L_{\mathrm{PCM}})) - n_{\mathrm{eff},1}L_1 = n_{\mathrm{eff,\,PCM}}L_{\mathrm{PCM}} + n_{\mathrm{eff}}(\Delta L - L_{\mathrm{PCM}}), \tag{4.1}$$

where $n_{\mathrm{PCM}} = n_{\mathrm{eff,\,PCM}}$ to save space and we have assumed that the effective refractive indices of the arms are identical outside of the PCM section, $n_{\mathrm{eff},1} = n_{\mathrm{eff},2} = n_{\mathrm{eff}}$. In this equation, the second term on the right-hand side is constant, so the change in the optical phase is only given by the state of the PCM, that is,

$$\Delta\varphi = \frac{2\pi}{\lambda_0}\Delta n_{\mathrm{eff,\,PCM}}L_{\mathrm{PCM}}. \tag{4.2}$$

Because a phase shift introduces a spectral shift $\Delta\lambda = \Delta\varphi\frac{\mathrm{FSR}}{2\pi}$ in the output of the MZI [197], we obtain

$$\Delta n_{\mathrm{eff,\,PCM}} = \frac{\Delta\lambda}{\mathrm{FSR}}\frac{\lambda_0}{L_{\mathrm{PCM}}}, \tag{4.3}$$

which we use to determine the change in the effective refractive index of Sb_2Se_3.

Figure 4.6 shows measurements of unbalanced MZIs with Sb_2Se_3. In Fig. 4.6a, I show an atomic force micrograph of a waveguide with Sb_2Se_3, with the scans in Fig. 4.6b indicating that the height of the stack of Sb_2Se_3 and $ZnS{-}SiO_2$ was 52 nm, in good agreement with the expected $30 + 20$ nm. Optical transmission spectra of a device with and without 5 µm long PCM in the amorphous and crystalline states are shown in Fig. 4.6c, with the device shown in the inset. To crystallise the Sb_2Se_3, we put the sample on a hotplate at 250 °C for 5 min. By measuring the blueshift between the different states, we can find $\Delta\lambda$ and thus $\Delta n_{\mathrm{eff,\,PCM}}$.

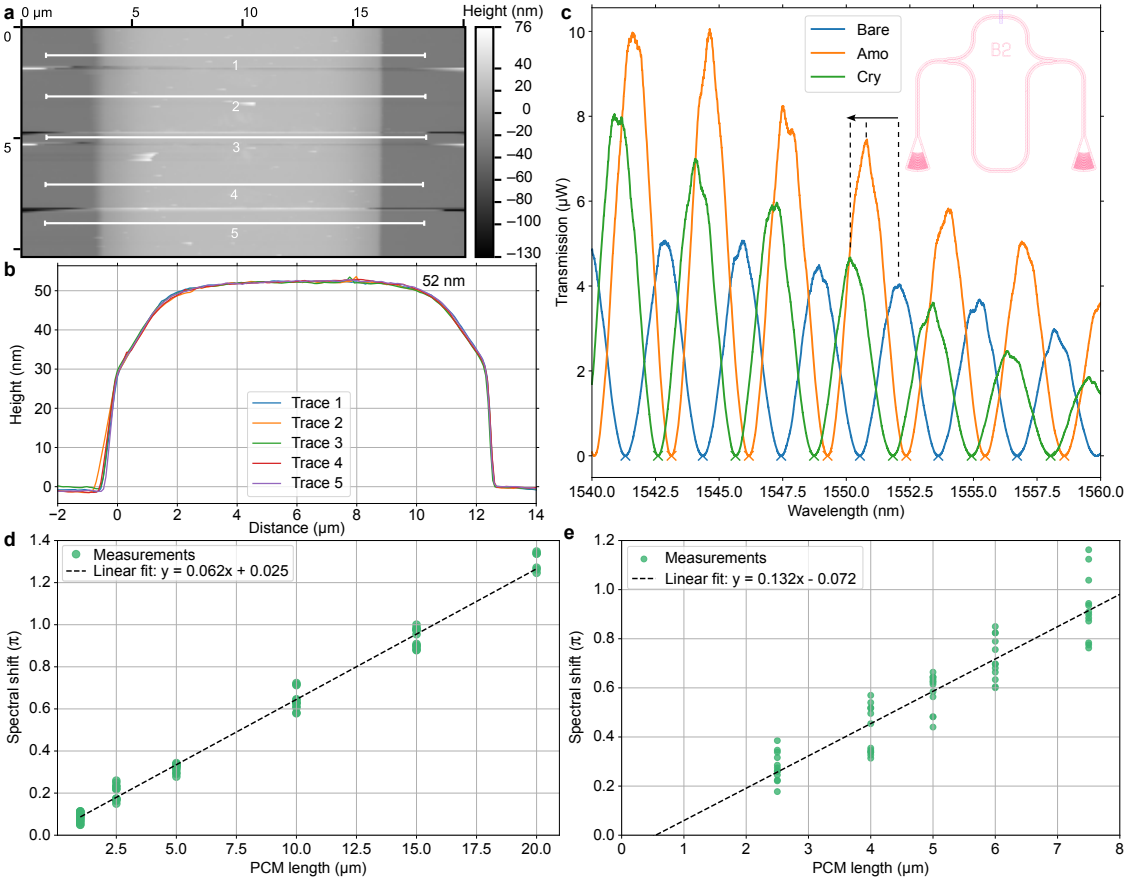

Figure 4.6: Effective refractive index measurements of 30 nm and 60 nm thick Sb_2Se_3. **a**, Atomic force micrograph of a 12 µm long patch of Sb_2Se_3. The waveguide and buffer regions are visible as horizontal lines. Because the scans are parallel to the waveguide edges, the height measurements in the horizontal edge regions are incorrect. **b**, Scans of the Sb_2Se_3 patch indicated by the numbers in **a**. The height of the patch is 52 nm. Note: The device with 12 µm Sb_2Se_3 was from a dummy device on the same chip and was not used in the rest of this figure. **c**, Transmission spectra of an unbalanced MZI without PCM (Bare), with amorphous Sb_2Se_3 (Amo), and crystalline Sb_2Se_3 (Cry), with PCM length 5 µm. The transmission change is most likely caused by variations in the setup between measurements, such as the distance or rotation between the fibre array and the grating couplers. The inset shows the GDS design of the MZIs used. **d**,**e**, Spectral shift in units of π vs. PCM length for Sb_2Se_3 of thickness 30 nm (d) and 60 nm (e).

In Fig. 4.6d and e, I show $\Delta\lambda$ given in units of π for 30 nm and 60 nm thick Sb_2Se_3, respectively. From the plots, we get $\Delta n_{\text{eff, PCM}} = 0.0481$ for 30 nm and $\Delta n_{\text{eff, PCM}} = 0.102$ for 60 nm, which corresponds to $L_\pi = \frac{\lambda_0}{2\Delta n_{\text{PCM}}} = 16.1$ µm and 7.6 µm, respectively, at $\lambda_0 = 1550$ nm. Here, $\Delta n_{\text{eff, PCM}} = n_{\text{eff, cry}} - n_{\text{eff, amo}}$. For the measurements with 30 nm Sb_2Se_3, the height of the waveguide was measured by profilometry to be in the range 118 nm to 119 nm. For 60 nm Sb_2Se_3, the height

of the waveguide was measured to be in the range 112 nm to 115 nm. Ideally, the etch depths and PCM thicknesses should have been consistent between the loss and effective refractive index measurements.

The effect of depositing PCM on a bare waveguide can be found if we insert $\Delta n = n_{\mathrm{eff,\,cry}} - n_{\mathrm{eff,\,Si}}$ and $\Delta n = n_{\mathrm{eff,\,amo}} - n_{\mathrm{eff,\,Si}}$ into the equation instead, which gives us $L_\pi = 5.8\,\mu\mathrm{m}$ and 4.3 µm for 30 nm thick a- and c-Sb_2Se_3, respectively (not shown in the figure). Such a short patch of PCM can be used, for instance, to introduce a static phase shift with minimal footprint to a photonic device [135].

Loss of Sb_2Se_3

For loss measurements, we deposited Sb_2Se_3 on a plain waveguide (see Sections 4.1.3 and 4.2.3 for fabrication steps). In Fig. 4.7a, an atomic force microscopy (AFM) image of a 30 nm thick PCM with a 20 nm thick capping layer is shown. Part of the waveguide buffer region is shown horizontally at the top of the micrograph. Note that because we scan along the direction of the waveguide, the height between horizontal lines in the waveguide area might be wrong. Figure 4.7b shows three AFM scans across the PCM patch, all showing similar heights from 55 nm to 57 nm, which is 10 % to 14 % higher than the expected height of 30 + 20 nm. The width of the PCM patch is also close to the design specification of 20 µm, although the sidewalls are tilted.

In Fig. 4.7c,d, I present the optical loss measurements from waveguides with 30 nm and 60 nm Sb_2Se_3, respectively, and a capping layer of 20 nm $ZnS–SiO_2$. The results show that the loss of crystalline Sb_2Se_3 is 0.021 dB/µm and 0.105 dB/µm for 30 nm and 60 nm thick Sb_2Se_3, respectively. The loss of amorphous Sb_2Se_3 was indistinguishable from that of a plain waveguide in our setup. For more information on the measurement setup, see Section 4.5.3.

4.4.2 Phase-change material simulation

The values obtained from the ellipsometry measurements were used in optical simulations in Ansys Lumerical MODE's Finite Difference Eigenmode solver. When PCM is deposited on top of an optical waveguide, we get a change in the effective

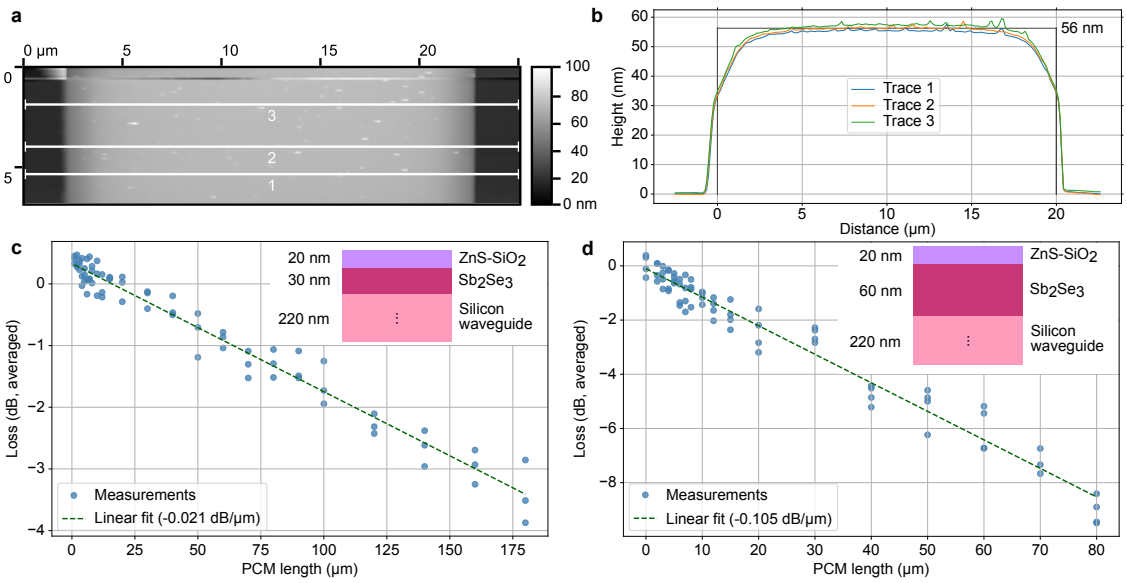

Figure 4.7: Optical loss measurements of Sb_2Se_3 on a waveguide. **a**, Atomic force micrograph of a 30 nm thick Sb_2Se_3 patch with 20 nm $ZnS-SiO_2$ on top. The waveguide buffer is visible horizontally at the top. **b**, Individual scans, as indicated by the lines in the micrograph in **a**, showing consistent height and width of ∼56 nm and 20 µm, respectively. **c**,**d**, Optical loss measurements of crystalline Sb_2Se_3 with a 20 nm thick capping layer of $ZnS-SiO_2$. The Sb_2Se_3 height was 30 nm and 60 nm in **c** and **d**, respectively.

refractive index, n_{eff}, of the hybrid waveguide-PCM system, which we can vary by switching the PCM. Figure 4.8 shows cross-sectional mode profiles for hybrid waveguides with different thicknesses of PCM in the amorphous and crystalline states, respectively. In the simulations shown in the figure, the waveguide is a 500 nm wide rib waveguide with a 100 nm thick slab, and the PCM is capped with 20 nm of $ZnS-SiO_2$. The electric field in the y direction, E_y, is shown in the figure, with the z direction upwards and the x direction pointing out of the plane of the page.

Table 4.2 summarises the simulation results for a waveguide with an Sb_2Se_3 layer of thickness 30 nm or 60 nm and a 20 nm thick capping layer of $ZnS-SiO_2$ covering the top of the waveguide. The table shows the effective refractive index of the two states of the PCM ($n_{\text{eff, amo}}$ and $n_{\text{eff, cry}}$), the effective refractive index difference between the two states (Δn_{eff}), and the length required for a full π phase shift (L_π). As shown in the table, increasing the thickness of the PCM can significantly reduce the device length. Furthermore, the waveguide parameters also affect L_π: a width of 500 nm instead of 460 nm reduces L_π, and the same is true for an etch depth of

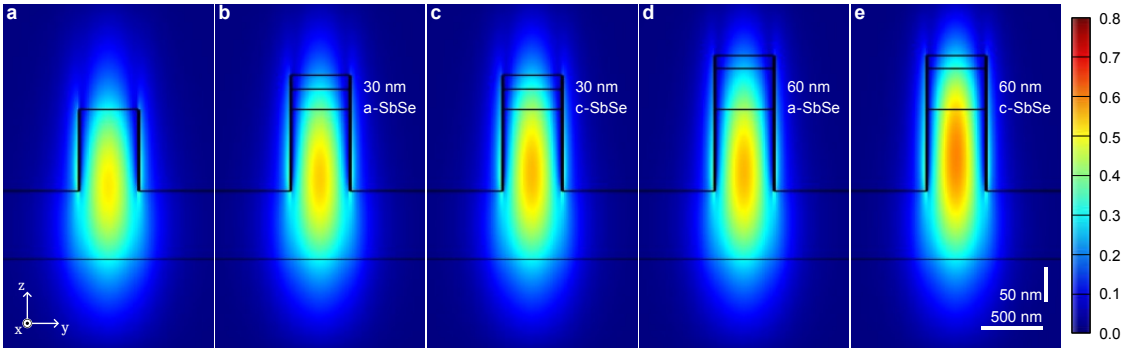

Figure 4.8: Simulations of an SOI rib waveguide with Sb_2Se_3 and a capping layer of ZnS–SiO_2 on top. **a**, A plain rib waveguide with dimensions $w = 500\,\text{nm}$, $h_{\text{tot}} = 220\,\text{nm}$, $h_{\text{wg}} = 120\,\text{nm}$, and $h_{\text{slab}} = 100\,\text{nm}$. **b**–**e**, The waveguide in **a** with 30 nm amorphous Sb_2Se_3 (a-Sb_2Se_3) and 20 nm ZnS–SiO_2 (b), 30 nm crystalline Sb_2Se_3 (c-Sb_2Se_3) and 20 nm ZnS–SiO_2 (c), 60 nm a-Sb_2Se_3 and 20 nm ZnS–SiO_2 (d), and 60 nm c-Sb_2Se_3 and 20 nm ZnS–SiO_2 (e). Note that the y- and z-directions are not 1:1 and have different scale bars.

120 nm instead of 110 nm (equivalent of $h_{\text{slab}} = 100\,\text{nm}$ and 110 nm, respectively). Finally, it is interesting to see the difference between the basic and advanced models; in the most extreme case, the simulated L_π differs by almost 3 µm!

In Ríos *et al.*, the authors state that due to the relatively flat refractive index in the range 1500 nm to 1600 nm ($\Delta n(\lambda) \sim C$, where C is a constant), the phase modulation of Sb_2Se_3 (given in units of π/µm from $\frac{1}{L_\pi} = \frac{2\Delta n}{\lambda}$) is mainly dependent on λ directly [120]. If we assume $\Delta n = 0.05$ is constant with wavelength, we get $L_\pi = \frac{\lambda}{2\Delta n} = 15\,\mu\text{m}$ and 16 µm for $\lambda = 1500$ and 1600 nm, respectively, with phase modulation of $0.0667\,\pi\,\mu\text{m}^{-1}$ and $0.0625\,\pi\,\mu\text{m}^{-1}$, respectively. In Table 4.2, for the advanced model with 30 nm Sb_2Se_3, $w = w_2$, and $h = h_2$, we have $\Delta n(\lambda) = 0.0471$ and 0.0462 at 1500 nm and 1600 nm, respectively; $\Delta n(\lambda)$ therefore has a small wavelength dependence. Consequently, the phase modulation lies within $0.058\,\pi$/µm–$0.063\,\pi$/µm for 30 nm Sb_2Se_3 and $0.122\,\pi$/µm–$0.135\,\pi$/µm for 60 nm.

We also obtain loss estimates from the MODE simulations with the advanced fitting model of the ellipsometry data. The hybrid waveguide with $w = w_2$, $h = h_2$, and 30 nm (60 nm) Sb_2Se_3 has a loss of $6.0 \times 10^{-4}\,\text{dB}\,\mu\text{m}^{-1}$ ($1.2 \times 10^{-3}\,\text{dB}\,\mu\text{m}^{-1}$) for the amorphous state and $0.058\,\text{dB}\,\mu\text{m}^{-1}$ ($0.133\,\text{dB}\,\mu\text{m}^{-1}$) for the crystalline state at 1550 nm.

Table 4.2: Variations in simulation results of Sb_2Se_3 depending on wavelength and waveguide parameters for 30 nm and 60 nm thick Sb_2Se_3. The effective refractive index of the hybrid waveguide with amorphous and crystalline Sb_2Se_3 (stylised here as n_{state} instead of $n_{\text{eff, state}}$ to save space, where state $\in$ [cry, amo]) is shown together with Δn_{eff} and L_π (given in µm), for the following combinations of parameters: the basic and advanced models; Sb_2Se_3 thicknesses $t = 30$ nm and 60 nm; wavelengths $\lambda = 1500$, 1550, and 1600 nm; waveguide widths $w_1 = 460$ nm and $w_2 = 500$ nm; and etch depths $h_1 = 110$ nm and $h_2 = 120$ nm. The shortest and longest L_π values are set in boldface.

				Basic model: Cody-Lorentz				Advanced model: Cody-Lorentz and Gaussian			
t	λ	w	h	n_{amo}	n_{cry}	Δn_{eff}	L_π	n_{amo}	n_{cry}	Δn_{eff}	L_π
30 nm SbSe	1500	w_1	h_1	2.6108	2.6600	0.0493	15.2	2.6131	2.6558	0.0427	17.6
			h_2	2.5891	2.6407	0.0516	14.5	2.5915	2.6362	0.0447	16.8
		w_2	h_1	2.6491	2.7013	0.0523	14.3	2.6515	2.6968	0.0453	16.6
			h_2	2.6318	2.6861	0.0543	**13.8**	2.6344	2.6815	0.0471	**15.9**
	1550	w_1	h_1	2.5576	2.6062	0.0486	15.9	2.5598	2.6017	0.0419	18.5
			h_2	2.5335	2.5846	0.0511	15.2	2.5359	2.5799	0.0441	17.6
		w_2	h_1	2.5979	2.6498	0.0519	14.9	2.6003	2.6450	0.0447	17.3
			h_2	2.5788	2.6328	0.0540	14.4	2.5813	2.6279	0.0466	16.6
	1600	w_1	h_1	2.5014	2.5495	0.0481	**16.6**	2.5035	2.5449	0.0413	**19.4**
			h_2	2.4749	2.5256	0.0507	15.8	2.4771	2.5207	0.0435	18.4
		w_2	h_1	2.5439	2.5954	0.0515	15.5	2.5462	2.5904	0.0443	18.1
			h_2	2.5228	2.5766	0.0538	14.9	2.5251	2.5714	0.0462	17.3
60 nm SbSe	1500	w_1	h_1	2.6534	2.7611	0.1077	6.96	2.6580	2.7509	0.0929	8.07
			h_2	2.6339	2.7462	0.1123	6.68	2.6388	2.7356	0.0969	7.74
		w_2	h_1	2.6944	2.8080	0.1136	6.60	2.6993	2.7973	0.0981	7.65
			h_2	2.6791	2.7965	0.1175	**6.38**	2.6842	2.7856	0.1014	**7.40**
	1550	w_1	h_1	2.6008	2.7060	0.1052	7.37	2.6052	2.6955	0.0903	8.58
			h_2	2.5792	2.6891	0.1100	7.05	2.5837	2.6782	0.0944	8.21
		w_2	h_1	2.6441	2.7556	0.1115	6.95	2.6487	2.7445	0.0958	8.09
			h_2	2.6271	2.7426	0.1156	6.71	2.6319	2.7312	0.0993	7.81
	1600	w_1	h_1	2.5453	2.6483	0.1030	**7.76**	2.5494	2.6375	0.0881	**9.08**
			h_2	2.5213	2.6293	0.1080	7.41	2.5256	2.6180	0.0924	8.66
		w_2	h_1	2.5909	2.7007	0.1097	7.29	2.5953	2.6892	0.0939	8.52
			h_2	2.5720	2.6860	0.1140	7.02	2.5766	2.6742	0.0975	8.20

Comparison between simulation and optical measurements

Table 4.3 summarises the simulated and measured effective refractive index and loss values for 30 nm and 60 nm Sb_2Se_3. The values for L_π are in good agreement but differ slightly—this discrepancy could be caused by fabrication variations, for instance, in etch depth, waveguide width, and Sb_2Se_3 thickness. Furthermore, in

the simulations, I have assumed that the Sb_2Se_3 is only on top of the waveguide. However, in the fabricated devices, I opened windows for Sb_2Se_3 deposition that were wider than the waveguide, meaning there is Sb_2Se_3 on the slab on both sides of the waveguide and potentially on the waveguide sidewalls as well. Based on previous experience in the research group, our waveguides are typically narrower than 500 nm, which should lead to a longer L_π than in simulations. However, the extra PCM on the slabs might counteract this effect in the measured devices. Finally, in Table 4.3, there is a considerable difference in the loss, especially for 30 nm. This discrepancy arises from the difficulty of accurately measuring small extinction coefficient values with ellipsometry.

Table 4.3: Comparison of the effective refractive index change and loss of Sb_2Se_3 at $\lambda = 1550$ nm. The advanced model was used for the simulations, with $w = 500$ nm and etch depth 120 nm ($h_{\text{slab}} = 100$ nm). The first column contain the thickness of the Sb_2Se_3 patch, and the units for L_π, α_{cry}, and $\alpha_{\text{cry}}L_\pi$ are µm, dB/µm, and dB, respectively.

Thickness	Type	Δn_{eff}	L_π	α_{cry}	$\alpha_{\text{cry}}L_\pi$
30 nm	Measurement	0.0481	16.1	0.021	0.33
	Simulation	0.0466	16.6	0.058	0.96
	Difference (%)	-3.1	3.1	176.2	190.9
60 nm	Measurement	0.1020	7.60	0.105	0.80
	Simulation	0.0993	7.81	0.133	1.04
	Difference (%)	-2.6	2.8	26.7	30

4.4.3 Phase-change material considerations

When using PCMs, several parameters should be considered, including PCM thickness, waveguide width, and the deposition technique used for the material. Furthermore, for each of these parameters, several different considerations must be taken into account. For instance, the deposition technique, typically thermal evaporation [120] or RF magnetron sputtering, significantly affects the real and imaginary refractive indices of the material (see Table 2.1). Also, as discussed in the previous section, the waveguide shape affects the light's overlap with the PCM and, consequently, the performance.

The thickness of the PCM is a particularly important parameter, as increasing the PCM thickness enhances the achievable phase shift per unit length while

simultaneously increasing optical loss, thereby introducing a trade-off between compactness and loss. In Fig. S3 in Ref. 132, the authors present the total insertion loss required to achieve a π phase shift (αL_π) as a function of Sb_2S_3 thickness, and select the global minimum of this curve to determine the PCM thickness used in their device. A similar approach could have been used in this work. Note, however, that the authors in Ref. 132 fail to include the non-negligible loss of their integrated microheaters, which should be incorporated in the calculation to find the lowest total loss.

Although a thicker PCM patch gives a more compact device footprint, working with very thick PCMs can be challenging. First, a thick PCM can make uniform vertical switching difficult, and we might not obtain a complete phase transition throughout the volume [39]. Secondly, thick layers of Sb_2Se_3 might experience unintentional reamorphisation [39]. Thirdly, a thicker film requires us to supply more heat to the PCM–waveguide interfacial region to reach sufficient temperatures at the top of the PCM. This extra heat at the interface can lead to ablation, i.e., deformation and/or removal of material [39]. Yang *et al.* solved this problem effectively by depositing multiple microstrips of PCM instead of a single patch [131]; however, due to the air gaps between the microstrips, this solution also results in a longer effective L_π. Although all of the above considerations are generally true, Pang *et al.* recently demonstrated over 1500 switching cycles of a 70 nm thick patch of Sb_2S_3 [132], proving the feasibility of using thicker films.

In summary, there will inevitably be many trade-offs between optical loss, footprint, switching voltage, energy consumption, cyclability, and so on, with several optimal points that can be found through extensive simulations and measurements.

4.5 Experimental measurement setup

In this section, I describe the optical and electrical measurement setups used in this thesis, along with some preliminary measurements. Figure 4.9 provides an overview of the setups and measurement configurations used in Chapters 6 and 7, for the FPCA and the non-uniform photonic neural network weight, respectively. Any

deviations from the setups described here will be explicitly noted where relevant. In the rest of this section, I will introduce the various components that make up the optical and electrical setups.

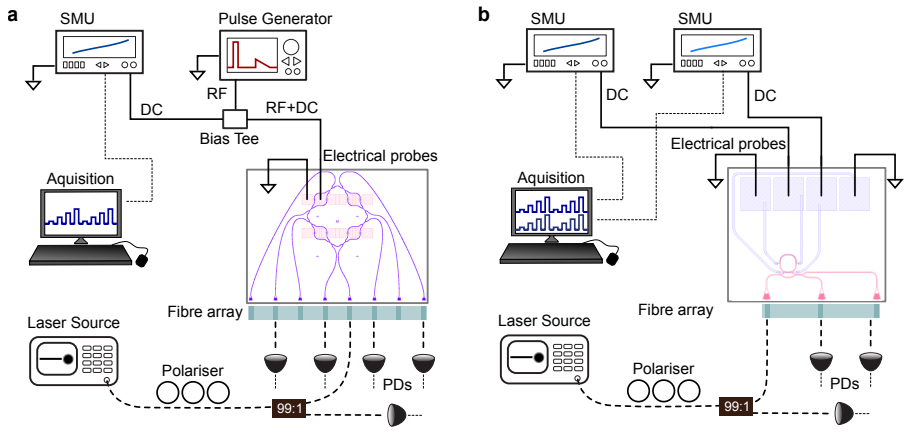

Figure 4.9: a,**b**, Experimental setups used in this thesis, for the FPCA measurements in Chapter 6 (a) and the non-uniform neural network weight in Chapter 7 (b). The photodetectors (PDs) are connected to a data acquisition (DAQ) unit again connected to the computer (not shown here). SMU: source measure unit; DC (here): direct current.

Although the optical and electrical setups are not separated during experiments (as seen in Fig. 4.9), the different connections between components might be more apparent to the reader if we separate the setup into two groups. I will therefore cover the optical and electrical measurement setups separately.

4.5.1 Optical measurement setup

Our most basic optical measurement setup consists of a laser source, a polariser, a fibre array for coupling to the chip, and a photodetector. In our integrated setup, every optical component is fibre-coupled (using SMF-28 single-mode fibres), with the sole exception of the fibre array that couples light into and out of the grating couplers on chip. To align the fibre array and grating couplers, we used lateral stages to control the sample's position in the plane and a vertical stage for the fibre array.

Optical components

Below is a summary of the optical components used in this thesis.

Laser source A tunable laser from Santec (model TSL-570) was used to generate light in and around the telecoms C–L bands (from 1500 nm to 1630 nm). The laser offers continuous wave (CW) lasing at an output power of up to 13 dBm, a low signal-to-noise ratio of 90 dB/0.1 nm, and highly accurate tunability with an error of less than ±20 pm. A key benefit of the Santec laser is its sweep capability, which enables the rapid acquisition of spectral data from resonant or interferometric devices. Unless otherwise noted, the laser power was set to 1 mW in the experiments.

Polariser Because the on-chip grating couplers in this work are polarisation-selective, we need to ensure the laser light is polarised for maximum coupling to our devices. The Thorlabs FPC032 Fiber Polarization Controller is a manual polariser that uses stress-induced birefringence in a single-mode fibre to manipulate the polarisation of the light. These in-line polarisers are devices made up of a quarter-wave plate, a half-wave plate, and a second quarter-wave plate arranged in series, allowing the generation of any desired output polarisation from an arbitrary input. The quarter-wave plate transforms the input polarisation state into a linear polarisation state, the half-wave plate rotates the orientation of the linearly polarised light, and the second quarter-wave plate converts the linear state into an arbitrary polarisation state. Thus, by tuning each of the three paddles (called fibre retarders), one can fully control the resulting polarisation state at the output.

Fibre-optic tap couplers In some of the experiments, we use a 99:1 fibre-optic tap coupler (or other ratios, including 90:10 and 50:50) to measure the light's intensity directly after the polariser. This additional measurement can be used to verify the laser power entering the chip. In this work, we used a Thorlabs PN1550R1A1 single-mode polarisation-maintaining fibre optic coupler with centre wavelength 1550 nm and a bandwidth of ±15 nm. The insertion loss is below 0.5 dB and 24.4 dB for the two ports. Note that we did not use this component in experiments where the operating wavelengths were outside the bandwidth range.

Fibre array The fibre array consists of evenly spaced SMF-28 fibres with 127 µm or 250 µm spacing between each fibre. Depending on the design, light coupled via the fibre array enters and exits at an angle relative to the normal of the PIC plane.

Photodetectors In this work, we mainly used 200 kHz InGaAs Fiber-Optic Receivers from Newport (model 2011-FC) for off-chip detection of continuous optical signals. The output voltages from the low-frequency PDs are recorded using a 16-channel DAQ unit from National Instruments (model PCI-6259) at a sampling rate of 1 kHz. For measuring fast dynamics, we used a 12 GHz InGaAs Fiber-Optic Receiver from Newport (model 1544-B). The output from this PD was captured by an oscilloscope (see Section 4.5.2) to circumvent the sampling limitations of the DAQ system.

Electro-optic modulator In some of the experiments, the amplitude of the optical input was modulated using an electro-optic modulator (EOM). In this work, a 10 Gbps EOM from Lucent (model 2623NA) with a $V_\pi = 3.8\,\mathrm{V}$ was used. The operating principle of the EOM is based on a phase shift produced by the Pockels effect in an LN MZM. When an electrical voltage is applied, this phase shift is converted into amplitude modulation through the interferometric behaviour of the MZM.

Erbium-doped fibre amplifier For the same experiments mentioned in the previous paragraph, we also attempted to amplify the light using a high-peak-power erbium-doped fibre amplifier (EDFA) from Pritel (model HPPPMFA-21-10). For an input power of 3 dBm, the EDFA delivers up to 20 dB gain. By adjusting the EDFA pump current, high-power pulses can be produced with average powers up to 189 mW and peak powers exceeding 10 kW, while preserving the pulse shape.

4.5.2 Electrical measurement setup

The electrical measurement setup, which complements the optical setup, consists of microwave probes and various instruments, including arbitrary waveform generators

and source measure units (SMUs). The probes provide electrical access to TO nichrome heaters and to microheaters used for switching the PCM. For the microheater experiments, a bias tee was used to enable both direct-current measurements and the application of electrical pulses without reconfiguring the electrical wiring.

Electrical components

The following is a summary of the electrical components used in this thesis.

Electrical probes An array of tungsten probes from GGB Industries (model PICOPROBE® 40A) was used to establish electrical contact with on-chip metal pads, allowing electrical switching and characterisation of devices with signal frequencies up to 40 GHz. The array featured probes with a fixed tip-to-tip pitch of 150 µm and was mounted on a three-axis micromanipulator stage for easy access to all devices on the same sample. Instead of using electrical probes, we could have wirebonded the heaters to a printed circuit board (PCB), to prevent the probes from moving the sample and thereby affecting the optical measurements. However, no PCBs were developed for the devices in this work; instead, we used double-sided tape to keep the samples in place.

Source measure unit An SMU combines a stable, reconfigurable source with a sensitive measurement unit, able to source and measure voltage or current simultaneously. In Chapters 6 and 7, we used a Keithley SourceMeter® (model 2618B) for resistance measurements of doped-silicon microheaters and TO heaters, respectively. In Chapter 7, we used the SMU together with an additional Keithley SourceMeter® (model 2400) to apply voltages to the two TO heaters in Fig. 4.4.

Function generator We used a 100 MHz, 1 GS s^{-1} arbitrary function generator from Tektronix (model AFG3102C) for PCM switching. Arbitrary pulse profiles can be produced using a custom-made virtual instrument in LabVIEW. In this work, we used single-step rectangular pulses for amorphisation and single-step

rectangular pulses, double-step rectangular pulses, and trapezoidal pulses (single-step rectangular pulse followed by linear decay to zero) for crystallisation. In our work, the function generator used assumes a 50 Ω impedance at the microheater, but because the actual resistance is >50 Ω, the voltage from the function generator is higher by a factor $F(R)$ [120], given as

$$F(R) = \frac{R}{R_{\text{load}}} \frac{R_{\text{load}} + R_{\text{int}}}{R + R_{\text{int}}}, \tag{4.4}$$

where R is the actual resistance of the microheater, $R_{\text{load}} = 50\,\Omega$ is the assumed resistance of the microheater, and $R_{\text{int}} = 50\,\Omega$ is the internal resistance of the function generator. For a 50 Ω impedance at the microheater, the maximum output from the function generator was 10 V, but for higher resistances, we can get up to $V_{\text{max}} = F(\infty) \times 10\,\text{V} = 20\,\text{V}$ (since $F(\infty) = \frac{R_{\text{load}}+R_{\text{int}}}{R_{\text{load}}} = 2$).

Bias tee A bias tee combines signals from an RF source and a direct-current source. In Chapter 6, we used a bias tee from Mini Circuits (model ZFBT-4R2GW+) for the measurement of current–voltage (I–V) curves and subsequent application of electrical pulses without having to redo the electrical wiring. The RF port of the bias tee has a frequency range from 100 kHz to 4.2 GHz, which means that for RF pulses longer than 10 µs, we will only see edge spikes, not a square pulse.

Oscilloscope For measurements in which we captured optical signal dynamics using fast PDs, a 4 GHz digital phosphor oscilloscope from Tektronix (model TDS7404B) was used for signal acquisition. The oscilloscope has a 0 % to 90 % signal rise time of 100 ps and was interfaced to the computer using a LabVIEW virtual instrument.

4.5.3 Optical loss measurements

All optical measurements conducted in this thesis are described in detail in the sections where the corresponding data are presented. However, for the optical loss measurements of the p^{++}-doped microheaters in Section 4.2.1 and the crystalline Sb_2Se_3 presented in Section 4.4.1, a more detailed description is warranted.

For optical loss measurements of lossy structures, the following approach was used. First, I made several devices and changed only one parameter, for instance, the doping length along the waveguide. All other parameters were identical between devices. Next, I measured the optical transmission spectra of the devices and found the maximum transmission. For the plots, I found the average transmission in a range around the maximum transmission value, i.e.,

$$T_{\mathrm{avg}} = \frac{1}{N}\sum_{j=0}^{N-1} T(\lambda = \lambda_{\mathrm{max}} - \frac{\lambda_{\mathrm{span}}}{2} + \frac{j\lambda_{\mathrm{span}}}{N-1}),$$

where N is the total number of transmission values measured in the range, λ_{max} is the wavelength that satisfies $T(\lambda) = T_{\mathrm{max}}$, and $\lambda_{\mathrm{span}} = 2\,\mathrm{nm}$. For the measurements of the p^{++}-doped Si waveguide, the loss of the devices was compared to that of a reference device without doping. To reduce variability in the loss measurements for the PCM patches, I instead compared each device to itself in the bare state (before deposition) with the fibre array at the same distance from the substrate. Because grating couplers can vary between devices, leading to different transmission in the reference device than in the other devices, the latter approach will reduce the measurement uncertainty.

4.5.4 Measurements of directional couplers

In this thesis, I will often use DCs for splitting and coupling light, for instance, as 50:50 couplers. Measurements carried out by my colleague Sijing Zhong showed that a DC with a 200 nm gap and 30 µm bend radius is governed by the following equation:

$$P_{\mathrm{cross}} = \sin^2\left(0.13L_{\mathrm{c}} + 0.65\right), \tag{4.5}$$

for $\lambda = 1550\,\mathrm{nm}$, as shown in Fig. 4.10a. Notice that the bends into the DC shift the curve, meaning there is coupling even at $L_{\mathrm{c}} = 0$. This is the equation used in Chapter 7, where several coupling coefficients are introduced.

For the work in Chapter 6, I needed an exact 50:50 splitting ratio. Due to fabrication imperfections between devices and variations in the measurement setup, the approach presented above does not guarantee an accurate splitting ratio. A

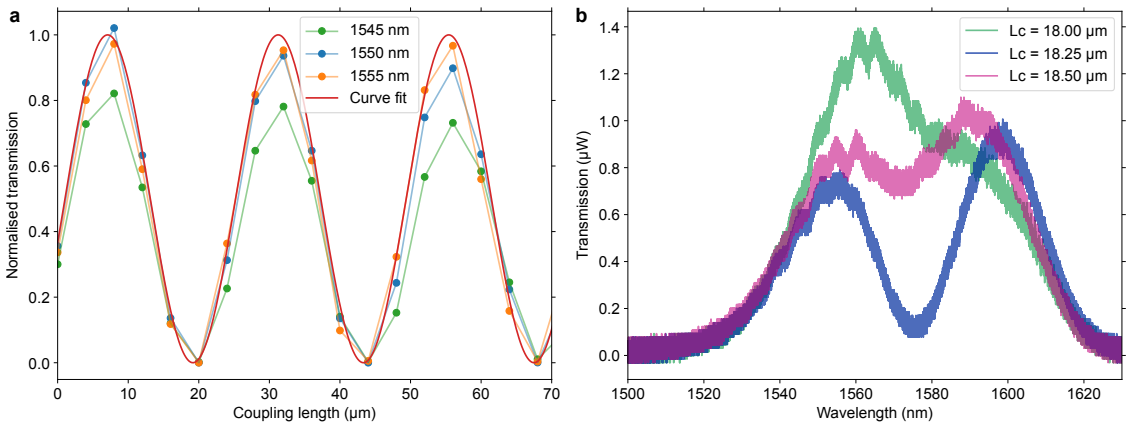

Figure 4.10: Experimental validation of 50:50 splitting in DCs. **a**, Normalised transmission of DCs with different coupling lengths, for three different wavelengths. The DCs have gaps of 200 nm and bend radii of 30 µm. The curve fit follows the equation $\sin^2(0.13x + 0.65)$. **b**, Transmission spectra from the bar port of balanced MZIs with 2×2 DCs at the input and output, for different DC coupling lengths.

more accurate approach includes measuring balanced MZIs. From Eq. (3.36), we know that the bar state output of a balanced MZI with DCs at both the input and the output should be 0 for 50:50 splitting. Therefore, I decided that it was better to measure the output of such MZIs for different DC coupling lengths. Note that for the work in Chapter 6, I used a 250 nm gap instead of the previous 200 nm gap. Initial tests of individual DCs with 250 nm gap showed 100 % coupling for $L_c = 10.5\,\mu m$ and 0 % coupling for $L_c = 26\,\mu m$, so 50 % coupling should occur around $L_c \approx 18.25\,\mu m$ (compared with $L_c = 13.97\,\mu m$ in Eq. (4.5) for DCs with a 200 nm gap). In Fig. 4.10b, I show the transmission spectra from the bar port of three MZIs with DC coupling lengths of 18.00 µm, 18.25 µm and 18.50 µm. The transmission drops to 0 at wavelengths around 1575 nm for the device with a DC coupling length of 18.25 µm, in excellent agreement with the initial tests.

Men rise from one ambition to another: first, they seek to secure themselves against attack, and then they attack others.

— Niccolò Machiavelli

5

Phase-change material-based photonic computing for defence-relevant tasks

Contents

5.1 Background 110
5.2 Photonic sensor fusion and its issues 111
5.2.1 Analogue solutions to Kalman filtering 113
5.2.2 A final comment on sensor fusion 113
5.3 Photonic model predictive control 114
5.3.1 Practical example: target tracking 115
5.3.2 Photonic implementation 116
5.4 Practical aspects and performance metrics 120
5.4.1 Performance of the electronic counterparts 121
5.4.2 Speed calculations for photonic hardware 122
5.4.3 Power calculations for photonic hardware 125
5.4.4 Other practical aspects 126
5.5 Future work 127
5.6 Chapter summary 128

The objective of this chapter—and, in fact, the original objective of this entire thesis—is to identify areas where photonic computing schemes based on PCM outperform other technologies, particularly for defence-relevant tasks like sensor management. Here, my aim is to lay the groundwork for understanding the potential advantages of photonic neuromorphic computing in defence-oriented computing. First, I will cover computationally intensive tasks, such as matrix multiplication,

used, for example, in the Kalman filter. However, due to the nature of our PCM-based photonic computing platform, I shifted my focus to another computationally heavy task, model predictive control (MPC), which benefits from the best qualities of our technology—the non-volatility of the PCM-based matrix elements and the fast modulation and detection of light signals—without being limited by our current experimental and device-level limitations.

For researchers at the Defence Science and Technology Laboratory (Dstl) who might be reading this work, it is worth noting the following. First, as a research group, we are constrained in our scaling capabilities, primarily due to limited measurement equipment and fabrication constraints, which limit us to weight matrices of size $M \times N$, with $M, N \lesssim 10$. In comparison, IBM recently developed a 64-core electronic in-memory compute chip with a crossbar array design similar to our photonic design, but with a magnificent 256×256 weights per matrix core [233]. Theoretically, similar matrix sizes can also be achieved on photonic chips, but doing so requires large teams, advanced foundries, and sufficient funds to cover the high cost per die area. Secondly, our photonic chips require a complex lab setup that is neither movable nor field-deployable. The setup, consisting of lasers, PDs, modulators, etc., can be replaced with on-chip components, but even our foundry-fabricated chips do not contain all components on a single chip. As a result, we are currently unable to use our devices outside of the lab. Thirdly, the most mature technology presented in this thesis—the platform discussed in this chapter—was ironically the only technology that was not fabricated, due to significant delays from the foundry. It becomes evident that although much of the technology is ready for widespread implementation, the ecosystem is not yet fully in place.

5.1 Background

Sensor fusion and sensor management are important to many non-military and defence applications and have therefore been studied intensively over the last decades [234, 235]. The general goal of sensor management is to dynamically choose actions for individual sensors to maximise overall network utility. Because there

are often conflicting goals and a considerable number of sensors and objectives, the problem quickly becomes computationally difficult to solve. Hence, sensor management problems are not trivial to optimise and are often affected by various trade-offs. Potential improvements lie in software- and hardware-based solutions; that is, put simply, better algorithms and better processors, respectively. With these potential improvements in mind, the original objective of the proposed work was to provide key insights into the performance gains that photonic neuromorphic processors based on PCM could provide for sensor fusion.

Where sensor fusion answers the question '*What is going on?*' and sensor management answers '*What should we observe next?*', a similar methodology relevant to both defence and non-military applications alike is MPC, which answers '*What should we do next?*'. The MPC algorithm considers effects of current and future actions on future time steps, which means it outperforms traditional proportional–integral–derivative (PID) control methods [226]; however, this non-myopic characteristic makes the control law complex and relatively slow. As a result, MPC implemented in electronics is limited to the sub-MHz, typically kHz, range [236, 237]. For some applications, slow is not necessarily a problem—for example, an update time of 0.1 s is acceptable for the self-driving car in Ref. 238—while for other applications, such as controlling nuclear fusion power plants or the trajectory of a drone, MHz operation would be required. To reach such speeds, a photonic implementation of MPC might be the only way to overcome the limits of electronics-only approaches.

5.2 Photonic sensor fusion and its issues

In Section 3.8, we introduced sensor management and sensor fusion and covered the mathematics that govern their simple implementations, such as the Kalman filter. Here, I argue that sensor fusion algorithms, including the Kalman filter, might not be the best-suited problems to solve with our PCM-based photonic hardware in its current form.

Primarily, the implementation of the Kalman filter in our PCM-based photonics platform is not straightforward, due to the recursive behaviour of the matrices. If the Kalman filter only contained MVMs in which the matrices were constant, and only the vectors changed, we could have engineered the chip design so that the matrix–vector products were optically cascaded from matrix to matrix. Both the number of matrices and the information transfer between them would be difficult—but possible—to devise. However, the recursiveness of the matrices in the Kalman filter (such as the Kalman gain, G_k) means that they have elements that are potentially updated every time step k. Therefore, these matrices should not be implemented with PCM elements, as updating the PCM elements frequently is not ideal; if the phase-change memory in Zhou *et al.* [195] was updated between the maximum and minimum absorption levels at a frequency of 1 MHz, the power consumption would be $1\,\mathrm{MHz} \times (6.9\,\mathrm{nJ} + 8.8\,\mathrm{nJ})/2 = 7.9\,\mathrm{mW}$, which is comparable to that of TO switches [24]. Furthermore, even if we optimistically assume a cyclability of 1×10^{12} cycles, the device would fail in less than 12 days if operated at this switching frequency. It is evident that photonic tensor cores based on PCM are not suitable for applications that require frequent switching of the matrix elements.

Viable solutions to photonic sensor fusion include using EAMs or other fast modulation techniques to set the matrix weights, but even this approach is not practical in the near term. With the numerous matrix multiplications and the inverse operation in the Kalman gain (Eq. (3.51)), it is likely that the information is converted between the optical and electrical domains so often that the benefit of photonics is lost to some extent. As an illustrative example, we can easily multiply the product of an MVM, $y = Ax$, with another matrix, i.e., $By = BAx$, without converting the optical signal to the electrical domain between matrices—the output vector of the first multiplication is the input vector of the second. On the other hand, if the result of one multiplication is used to set the weights of another matrix (as is the case for the recursive nature of the Kalman filter), the most feasible option is to convert the signal to the electrical domain and then set the weights electrically (with the likely intermediate step of saving the value to memory). Although a photonic

implementation is in all likelihood still faster and more energy efficient than pure electronics implementations [7, 8, 188], the overhead of EO conversion dilutes the expected benefits of using photonics, especially for our simple, lab-based setups.

5.2.1 Analogue solutions to Kalman filtering

At least two papers have investigated alternative electrical solutions to sensor fusion. Chen *et al.* described an analogue circuit that represents a hardware implementation of the continuous-time analogue Kalman filter [239], where the Kalman gain is implemented using a memristor. The authors claim that the memristor only needs to be programmed once (or whenever the environment changes dramatically) because the Kalman gain converges to a constant after several iterations [239]. In both speed and energy efficiency, an enhancement of one order of magnitude was observed, primarily due to the intrinsic properties of analogue processing, as well as the memristor's non-volatile characteristics, resulting in lower energy consumption. A similar result was found for an analogue circuit with a memtransistor-based Kalman gain, where the power consumption was approximately 25 % of that of a digital Kalman filter [240]. These demonstrations show that, at least for relatively simple applications, the equations outlined in Eq. (3.51) can be solved with analogue circuits rather than digital processors, with the advantages of lower latency and lower energy consumption.

Implementing a similar solution in photonics is not as straightforward as in electronics. In both reports mentioned above, the analogue electrical circuits were relatively simple; however, the behaviour of some of these components, such as the op-amp and capacitor, is challenging to realise in photonics. We tried to devise a practical PIC equivalent to those shown in Refs. 239 and 240, but were unable to do so. Therefore, I will not discuss this topic further in this thesis.

5.2.2 A final comment on sensor fusion

In the current times, any problem you can think of, a neural network can probably solve. However, some problems might be better suited due to their mathematical

similarity, and this is true for the Kalman filter; in Freire's tutorial on artificial neural networks [241], the authors underline the interrelation between recurrent neural network (RNN) models and Kalman filter theory. Because the weight matrix does not change at each time step in the RNN, the use of a PCM-based PTC for an RNN could have been a viable approach to solving a sensor-fusion-relevant problem with photonic hardware. Nevertheless, since we were unable to develop a scientifically interesting and feasible scheme for this approach, this topic will not be further covered in this thesis. Ironically, it was only at the end of this project that we realised the similarities between the photonic platform for MPC covered in the next section and photonic RNNs used to solve the Kalman filter.

5.3 Photonic model predictive control

As outlined in the introduction to this chapter, we endeavoured to identify a domain related to sensor management and sensor fusion where PICs would accelerate the process and/or reduce the system's overall power consumption. The problem at hand needs to be sufficiently simple for our lab setup with its current limitations, while at the same time interesting enough to demonstrate that we can perform computing for real-world, defence-relevant tasks. In this section, I discuss how we can use model predictive control (MPC) to solve a simple 1-D target tracking problem for real-time decision-making in mission- and safety-critical environments. I chose this example because of its combined simplicity and defence relevance, and because it also integrates well with sensor management [242]. Although there are only a few parameters and constraints on the system, the example still shows how robust the algorithm is. Furthermore, for more complex problems, the implementation can easily be scaled up.

Photonic neuromorphic computing is expected to achieve ultrafast MPC, with estimated speeds on the order of hundreds of MHz [226]. An implementation of photonic MPC was proposed by de Lima *et al.* [226]. In their system, light passes through a weight matrix, and the output light is acted on by an activation function before it is used as the input light to the weight matrix in the next iteration [226].

The photonic MPC iteratively converges to a steady-state solution that represents the action at the current time step. For the next time step, the matrix remains the same, while the input light is set to a different initial condition, and the system again iteratively converges to a steady-state solution, which is then used as the action for that time step. Because we can solve the MPC problem iteratively, much of the speed-up comes from the fast modulation and detection of light signals, which happens repeatedly at each time step. By designing the analogue system so that we avoid conversion between the analogue and digital domains at each iteration and use ADCs only to implement the action given by the steady-state solution at each time step, we can avoid significant conversion-related delays and considerable power consumption.

5.3.1 Practical example: target tracking

Here, I present a simple example of MPC, with the general MPC scheme shown in Fig. 5.1a. In this example, a camera with a narrow field of view (FOV) is tasked with photographing an object. We assume that the camera receives information from a sensor about the estimated position of the object, meaning the camera itself is not actively tracking the object's position; instead, it is fed the estimated position and a model of the object's movement for the next $N-1$ time steps. In this 1-D example, we control the angle that the camera is pointing in, θ, which is restricted by two constraints: upper limits for the angular acceleration ($u_k < a_{\max}$; we cannot have sudden jumps) and the angle ($\theta < \theta_{\max}$; there is a blind sector). This scenario is shown schematically in Fig. 5.1b. The prediction horizon is $N = 3$ time steps, which gives $3 \times 1 = 3$ control parameters and $3 \times 2 = 6$ constraints for the system. Initially, we assume that the object follows a sinusoidal path, given by $\theta = 40 + 30\cos(t/8)$, and that the camera starts at $\theta = 0$.

Primal–dual method for model predictive control

In the tutorial by de Lima *et al.* [226], the authors suggest using Nengo to map the QP problem in Eq. (3.64) to a photonic neural network. Nengo is a Python

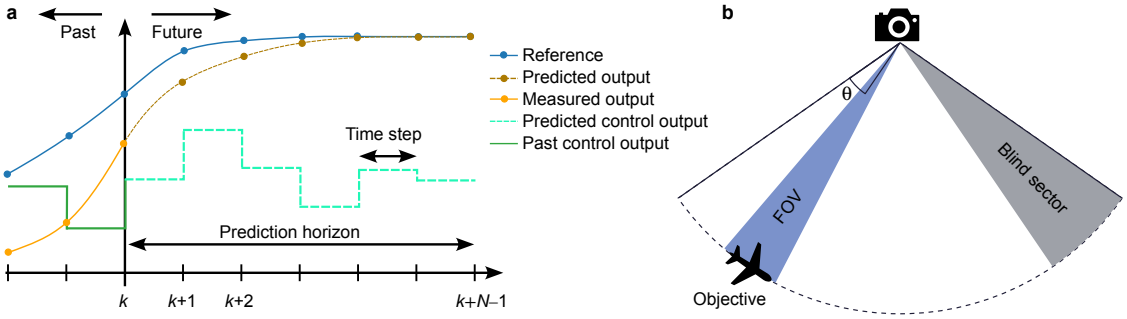

Figure 5.1: Model predictive control concept. **a**, MPC scheme showing the effect of current and future control outputs. **b**, Schematic showing a camera tracking an object travelling on the dotted line. The camera, pointing at an angle θ, has a limited FOV and cannot track in the blind sector. Figure **a** by Martin Behrendt, Wikimedia Commons, file: `MPC_scheme_basic.svg`, modified under a Creative Commons licence CC BY-SA 3.0.

package for simulating large-scale neural systems, built upon the Neural Engineering Framework [243]. However, we were unable to use this approach. Instead, in this work, I will use the primal–dual method outlined in Section 3.8.3. For this example, $P \in \mathbb{R}^{3\times3}$, $G \in \mathbb{R}^{6\times3}$, $u \in \mathbb{R}^3$, $q \in \mathbb{R}^3$, $\lambda \in \mathbb{R}^6$, and $h \in \mathbb{R}^6$. By solving the equations in Sections 3.8.2 and 3.8.3, we get the result shown in Fig. 5.2. We see clearly that the algorithm respects the acceleration constraint in the first few seconds and the angle constraint around 50 s.

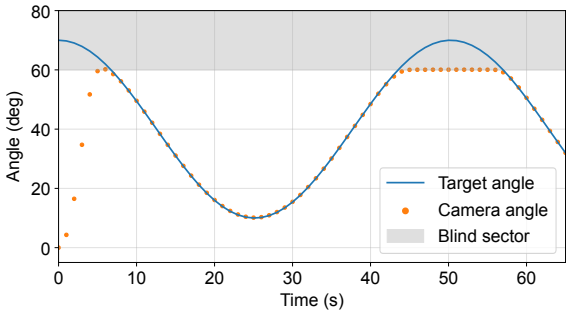

Figure 5.2: Simple tracking example of an object following the path $\theta = 40 + 30\cos(t/8)$, for $N = 3$, $u_k < 5\,\mathrm{deg/s^2}$, and $\theta < 60\,\mathrm{deg}$.

5.3.2 Photonic implementation

Figure 5.3 shows a schematic of interconnected PTCs that can be used to implement the simple algorithm in Sections 3.8.2 and 3.8.3 in photonic hardware. The flow of the computation is as follows: each of the light signals $(u_k(i), \lambda_k(i))$ is split into

two waveguides and enters the rows of the PTCs, the signals are weighted, and the output from each column gets summed up at the PDs. Note that each PD in this design sums up the output signals from two different light paths. The electrical photocurrents from the PDs are fed to EO converters that determine the amplitude of the light signals for the next iteration, i.e., u_{k+1} and λ_{k+1}. To implement the effect of the h and q vectors, we introduce an electrical bias to the EO converters. At steady state, an ADC reading the same voltage signal as the EO converter will convert the analogue signal to a digital signal used to implement the next control action, $u(0)$. As discussed at the beginning of this chapter, the advantage of this approach is that the matrices stay constant as long as the constraints do not change, which is ideal for PCM-based amplitude modulators. Furthermore, the algorithm's iterative nature pairs well with the ultrafast modulation and detection of light signals traversing the matrices.

Several practical considerations warrant note. First, the EO converter (modulator, laser, or similar) that determines $\lambda_{k+1}(i)$ also needs to contain a non-linear activation function (e.g., ReLU) to ensure $\lambda_{k+1}(i) \geq 0$. Practically, this non-linear activation can be implemented using a simple electrical circuit [244, 245]. Alternatively, if we encode the q and h vectors as optical signals to the PDs, the non-linearity could, in theory, happen in the optical domain. Secondly, because the u and λ vectors are encoded as light signals, we cannot use actual negative values for the vector elements. The λ_k vector is automatically positive, while for u_k, we add a global offset to the vector, u_{offset}, to ensure that only positive optical intensities are encoded. After the optical amplitudes are summed in the PD, we then have to account for the offset in u_{k+1} and λ_{k+1} by subtracting $(I - \eta P)u_{\text{offset}}$ and $\eta G u_{\text{offset}}$, respectively. This offset is added to the PDs together with the h and q vectors electrically. Note that each $u_{\text{offset}} = [o_0, o_1, \cdots, o_{N-1}]^\top$ and o_i can be different for each iteration, but I have chosen to keep them the same throughout the calculations for simplicity. Furthermore, by choosing a single global u_{offset}, we do not need to calculate each individual o_i in each iteration, which would have significantly increased the latency of the process.

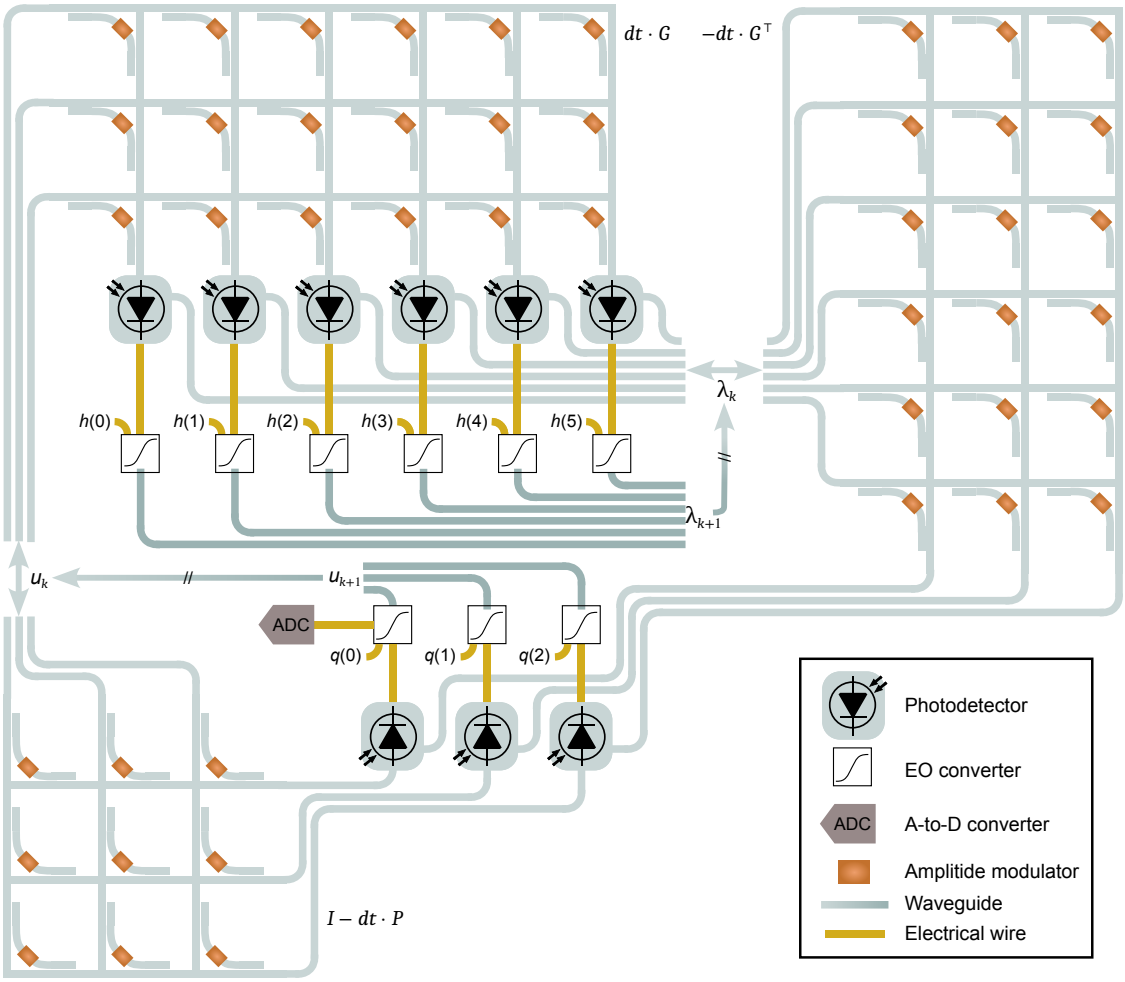

Figure 5.3: Interconnected photonic crossbar arrays for MPC. The design implements the update rule given in Eqs. (3.66) and (3.67). The flow is as follows: light signals $(u_k(i), \lambda_k(i))$ enter the rows of the crossbar arrays, the output gets summed up at the PDs, the electrical photocurrent is fed together with an electrical bias to an electro-optical (EO) converter (also acting as a non-linear activation function in the case of λ_k), which determines the amplitude of the light signals for the next iteration. Note that the waveguides connected to a PD are always leading into it, and that the electrical connections are going out of it, with the sole exception of the wire to the ADC converting the steady-state electrical signal to a control action. Also, note that the EO converter in this figure includes a photocurrent-to-voltage converter (e.g., a TIA) not shown. Note: in this chapter, λ does not refer to wavelength, but the Lagrange multiplier.

Simulation of photonic primal–dual solver

In Fig. 5.4, I show a more advanced example of MPC that could be solved with a photonic primal–dual solver, in which I implement the offset described in the previous paragraph to obtain only positive optical amplitudes. Furthermore, I include periods of false detection, that is, when the model provides the camera with

incorrect reference angles, which could occur if the target's position is difficult to determine (e.g., due to obstructions, partial tracking due to stealth capabilities, or sensor downtime). As a result, the model fed to the camera will suddenly update, leading to discontinuities in the reference trajectory that the MPC algorithm must account for. I assume the model is only updated after the discontinuity has occurred, so we cannot predict it. Here, the model angle is given as

$$\theta_{\text{target}}(t) = 40 + 30\cos\Big(t/8 - d(t)\Big), \qquad d(t) = \begin{cases} \dfrac{\pi}{4}, & 69 < t < 95, \\ 0, & \text{otherwise}, \end{cases} \tag{5.1}$$

for $t < 117$ and $\theta_{\text{target}}(t > 117) = 50$. Additionally, in each iteration, I add an error to u and λ drawn from a Gaussian function with $\sigma = 0.1$. The code for the advanced 1-D example is shown in Appendix A.1.1.

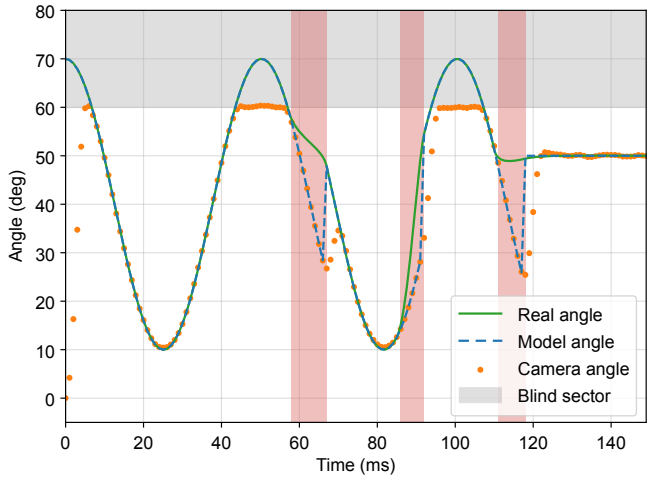

Figure 5.4: Advanced tracking example of an object, where the model follows the path described by Eq. (5.1). The red areas indicate where the angle predicted by the model deviates from the object's actual angle.

Foundry design

To solve the MPC examples covered in this section, I designed a PIC using the photonic design automation framework Luceda IPKISS and the process design kit (PDK) for the imec iSiPP50G silicon photonic platform. Part of the imec design is shown in Fig. 5.5. In the design, we see a 3×6 matrix (left half, with input from the bottom and output on the left) and a 6×3 matrix (right half, with input from

the left and output at the bottom). The optical output is read using integrated photodetectors. Each unit cell contains DCs at the input and output and a PCM cell, in addition to three grating couplers. To set the state of the PCM, we send light into the middle grating coupler and measure the transmission through the PCM using the two grating couplers on the side. A difference between this design and the one presented in Fig. 5.3 is the individual photodetectors at the end of each column, meaning the EO converters receive electrical signals from multiple PDs. Not shown in the figure are the 3×3 matrix and the metal pads used for electrical contacting.

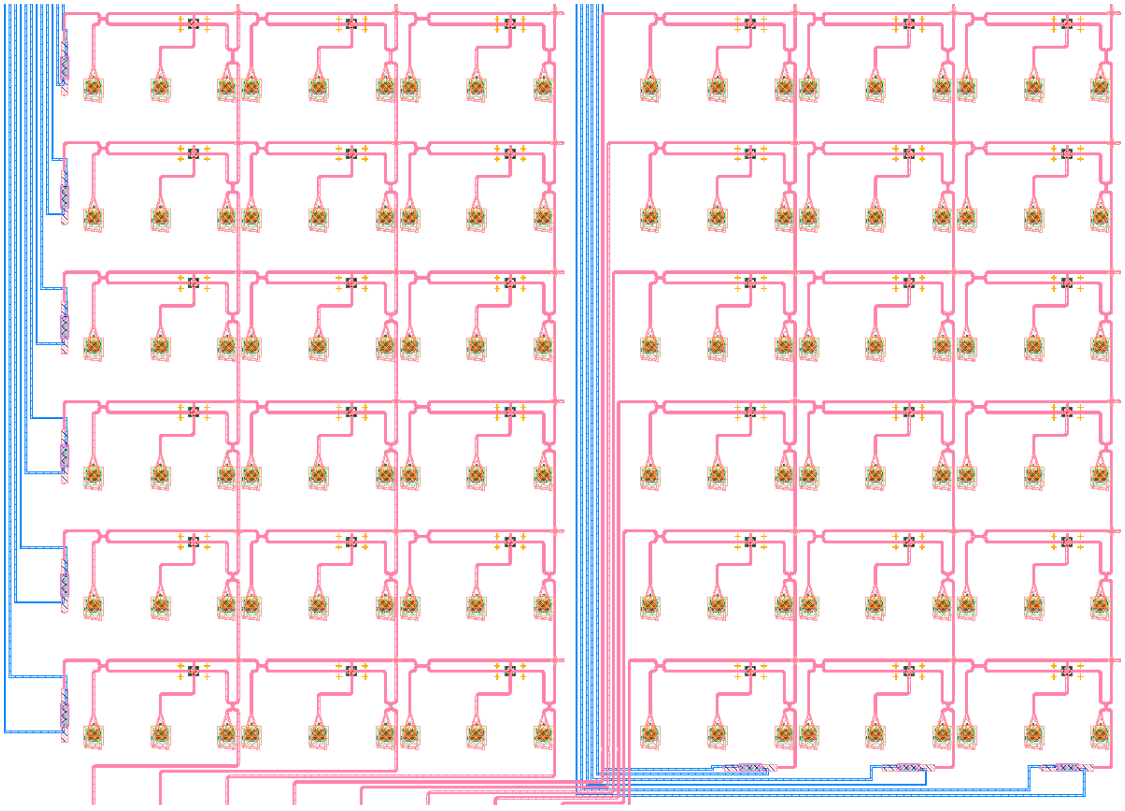

Figure 5.5: GDS design of photonic crossbar arrays for MPC. The two matrices shown here encode the 3×6 and the 6×3 matrices.

5.4 Practical aspects and performance metrics

There are several reasons why MPC problems can benefit massively from a photonic implementation with phase-change memories. First, a photonic implementation can dramatically speed up the calculation. The iterative process to find the solution

can require hundreds of iterations before reaching steady state, which means that the signal's ToF and processing (modulation and detection) must be ultrafast. This requirement is achieved here by using an optical signal carrier. Secondly, the calculation is performed on AIMC hardware, alleviating the von Neumann bottleneck of data shuffling between memory and the processing unit, which is especially beneficial for the frequently updated u and λ vectors. The third reason is that the matrices in this work are photonic crossbar arrays with PCM-based amplitude modulators that encode the matrix elements. Because the matrix stays constant throughout the whole computation, the non-volatile behaviour of PCM is fully taken advantage of. In addition, the potential downsides of PCMs, such as limited cyclability and slow switching, are not relevant metrics to consider for constant matrices. Finally, the photonic crossbar array allows for WDM and therefore parallelism. In theory, we could use the same static matrices with different wavelengths encoding different problems; however, only as long as the problem size (the number of control parameters and constraints) remains the same. Still, in practice, implementing WDM schemes on a chip using the design proposed in Fig. 5.3 would be difficult due to the additional routing to the (de)multiplexers.

In the following subsections, I will present estimates for the speed and power consumption of PCM-based photonic crossbar arrays for MPC and compare them with those of electronic solutions. The speed and power consumption of the photonic hardware can be calculated with both an optimistic, state-of-the-art-inclusive approach or a more sober approach that incorporates only components readily available in foundries. Before I present my calculations for the photonic platform, however, I will introduce its electronic counterparts.

5.4.1 Performance of the electronic counterparts

According to de Lima *et al.* [226], the speed at which electronics can perform MPC (in practice, the underlying QP problems) is limited to the kilohertz range. Although this metric is somewhat dictated by the size of the problem at hand, it is mainly determined by the limitations of electronics. In the paper by Vichik and

Borrelli [246], the authors solve a linear programming (LP) problem—the equivalent of implementing one action in the MPC case—with an analogue electrical circuit in 6 µs. The LP problem consisted of only two variables, x_1 and x_2, and four inequalities, which makes this example an extremely simple LP problem. More recently, Liu *et al.* showed in simulations that a memristor crossbar array can achieve sub-microsecond computing speeds for very small array sizes of less than 6×6 [247]. Furthermore, they also experimentally demonstrated a chip with 1.6 µs computing latency. In comparison, a digital QP solver would take 3.2 ms to solve the same problem, according to the authors [247].

For digital state-of-the-art implementations, such as Google's Tensor Processing Unit (TPU), the clock is limited to ~ 1 GHz [248], and this number rarely goes above a few GHz for other digital electronic hardware such as graphics processing units (GPUs) [249]. Although gigahertz-level clock cycles seem promising, the latency of the TPU scales with the number of clock cycles as $2N \times T$, where $N \times N$ is the matrix size and $T = 1/f$ is the time of one clock cycle [8], which limits this approach. Furthermore, such hardware is specialised for large-scale matrix operations and parallelism, but not necessarily for iteratively converging to a solution with recurring input to the same matrix, as we want in QP problems.

For the power requirements of electronic implementations, Liu *et al.* shows simulations of memristor crossbar arrays of various sizes that dissipate hundreds of milliwatts (~ 100 mW for 20×20 and ~ 500 mW for 100×100 crossbars) [247]. Their experimental circuit dissipated an average of 53.14 mW to complete all computations, which is an almost three times improvement compared with a digital baseline (151.8 mW) [247]. They also point out that the FPGA, a widely used platform for parallel MVM, can use more energy than the analogue crossbar array by more than two orders of magnitude.

5.4.2 Speed calculations for photonic hardware

The speed calculations of the PTC are presented in this section. These include the ToF of the signal traversing the circuit and the update time of the peripherals.

According to de Lima *et al.*, MPC is a good example of an application where fast processing (e.g., nanoseconds) is required but a slow learning rate (e.g., milliseconds) is acceptable [226]. Therefore, the learning (also called retraining), which practically means reconfiguring the elements of the weight matrix, does not have to be ultrafast. This aspect is essential to the success of PCM-based weight matrices because it helps avoid their major drawbacks (limited cyclability and high switching power).

Time of flight for optical circuits

The ToF is the time it takes the optical signal to traverse the circuit, given by $t = L/v$. In the photonic crossbar array, the ToF varies depending on the number of columns and rows traversed by the light. We define the ToF as the maximum propagation time among all possible optical paths, given by

$$t_{\max} = \frac{2L_{\max}}{v_{\mathrm{g}}} = 2L_{\max}\frac{n_g}{c} = 2\left(M \cdot w + N \cdot h\right)\frac{n_g}{c}, \tag{5.2}$$

where $L_{\max}$ is the maximum length the light travels in the matrix, v_{g}, c, n_g are the same as in Eq. (3.2), M and N are the number of rows and columns in the matrix, respectively, and w and h are the height and width of each unit cell in the matrix, respectively. The factor 2 accounts for the distance the light travels along the perimeter of the matrix before entering (due to routing). With $n_g = 4.2$, a 3×6 matrix with $40\,\mu\mathrm{m} \times 40\,\mu\mathrm{m}$ per unit cell will have a ToF of 10 ps, whereas light in a theoretical 225×225 matrix with the same parameters will take 0.5 ns to traverse the matrix once.

Update time for photonic peripherals

For small and medium-sized matrices, the ToF is not the limiting factor—instead, we are limited by the peripherals. For the architecture in Fig. 5.3, we need PDs, EO converters, and a non-linearity. In addition, we need to convert the photocurrent to useful modulation voltages, most conveniently achieved using a TIA. State-of-the-art PDs can reach bandwidths of above 100 GHz [250], while foundry-made PDs from the imec iSiPP50G silicon photonic platform are typically in the range from 20 to >50 GHz in imec's data sheets and recorded up to 67 GHz in the

literature [251]. The fastest amplitude modulators in silicon also reach speeds greater than 100 GHz [252]. In Ref. 253, the authors report the following numbers for different modulator technologies in the imec iSiPP50G silicon photonic platform: 30 GHz for a travelling-wave p–n junction modulator, 50 GHz for a microring modulator (MRM), and 50 GHz for SiGe EAMs.

For the non-linearity, there are several options. We can use an MRM similar to the one in Ref. 249, which incorporates the non-linearity and modulator in one device, where a laser supplies the light. The update frequency of this MRM was 30 GHz. Alternatively, we could use a membrane laser as the light source, ReLU function, and modulator in a single device [254]. Another option is to use an electronic non-linearity to control a linear optical modulator [255]. In theory, the non-linearity could also have been done all-optically [256, 257], but the output signal would most likely have to be amplified after each iteration. For the photocurrent conversion, we can use TIAs, which have bandwidths of up to 65 GHz in Ref. 258 and 97 GHz in Ref. 259. Finally, we need ADCs to read out the steady-state solution. In the memristor-based AIMC chip presented in Ref. 260, they assume 0.4 ns latency for 8-bit ADCs; however, it is important to note that this operation is only done once per time step in the MPC problem, not during the iterative phase.

Using the values from the previous paragraphs, we can estimate the time per iteration. For a 50 GHz PD, 65 GHz TIA, and a 30 GHz MRM acting as the non-linear activation function and EO converter, the total time is around 80 ps for a 6×3 matrix. In comparison, a Google TPU v1 would take around 10 ns [8]. For this design, we need to ensure that all signals are in sync (if not, the light traversing the shortest optical path would outpace that traversing the longest), which can be achieved using an electronic clock. Usually, the clock rarely exceeds a few GHz to prevent the digital electrical chip from overheating, but in contrast to digital electronics, photonic hardware is not the limiting factor, and the clock can be set to an update frequency of, for example, 12.5 GHz (equivalent of 80 ps) per iteration without performance degradation. For an MPC problem with 150 iterations per time step, the total computation time per control action is approximately 12 ns

(corresponding to 80 MHz) for a 6×3 matrix. Due to the longer ToF for a 40×40 matrix, the total time is doubled in this case.

5.4.3 Power calculations for photonic hardware

This section presents the power consumption of a photonic implementation of MPC, which depends on the matrix size and other factors, including speed–power trade-offs.

One great benefit of using PCM technology for MPC is that the weight matrices of the photonic neural network can be fixed across time steps, offering significant energy savings. As an example, the micro-ring weight banks used by Tait *et al.* need 5.2 mW per weight [178], so our PCM memories outperform their weight's power consumption as long as we reconfigure them less often than every 1.5 µs (assuming the PCM switching energies of Ref. 195).

As the weighting in this work is non-volatile, the other components of the photonic solver are responsible for the power consumption. The ultrafast TIAs have a power consumption of hundreds of milliwatts [258, 259], but this number can be reduced to sub-milliwatt levels if we sacrifice speed [261]. In the photonic platform developed by the LightMatter team [7], each TIA draws 41 mW. The power consumption of the integrated PDs, given by $P_{\mathrm{PD}} = I_{\mathrm{dark}} V_{\mathrm{bias}}$, is on the order of micro- to nanowatts and is therefore negligible compared with the power consumed by the TIAs used to convert the photocurrent. For the MRM acting as the EO converter, the power consumption is often given in units of fJ/bit, giving a power consumption of $\sim$1 mW for gigahertz modulation [56]. Additionally, we need optical light sources. Assuming a 10 % wall-plug efficiency [262], generating 10 mW of optical input power requires each of these devices to draw 100 mW. Finally, in the memristor-based AIMC chip presented in Ref. 260, the authors assume 0.6 µW leakage power per ADC, while the total power is significantly higher, 66 mW, for the high-speed ADCs used in the photonic platform in Ref. 7.

Considering all contributions, we can estimate the total power consumption of our circuit. For the MPC problem presented here, we need 9 PDs, TIAs, EO converters, ADCs, and light sources. The total power will mainly be determined by

the light sources and TIA, resulting in a total power consumption of around 2 W (assuming 100 mW per TIA and laser source). This power estimate is significantly higher than that of the memristor-based QP solver in Ref. 247. Because of losses in optical components and in the photonic computing architectures themselves [4], we need to supply significant optical power at the input to the matrices, which is detrimental to the photonic chip's power consumption. Two revealing examples highlighting this issue were presented recently. For the photonic platform developed by the LightMatter team, the lasers delivered 200 mW of optical power and required 1.25 W of cooling per laser [7]. In a similar vein, the laser modules employed by the Lightelligence team each consumed 0.4 W [8]. Nevertheless, because of the ultralow latency per computation, photonics already now outperforms the TPU in terms of energy efficiency (number of operations per power, given in TOPS/W) [8], with projected energy efficiencies surpassing the Google TPU v4 by two orders of magnitude [184].

5.4.4 Other practical aspects

In addition to the power and speed considerations discussed above, other aspects are worth noting regarding the implementation and performance of the photonic circuit. First, in the simulations presented in Fig. 5.4, I use only positive optical power values, which can be implemented by shifting all numbers by an offset. To avoid computational overhead per iteration, we need to know the global minimum across all iterations at the current time step and set it as our offset. Because we cannot necessarily know this offset before the time step, I have chosen it based on previous experience in this work. Alternative routes to optically represent negative values could be explored, such as using reference measurements. Alternatively, a phase shifter can be used in conjunction with the absorptive component at each crossing of the photonic crossbar array to represent negative numbers through complex-valued operation. However, this only works in coherent crossbar arrays [220], and renders the crossbar array architecture sensitive to phase variations [184]. Secondly, in Dong *et al.*, the researchers show that for the PTC in its current form, we are

limited to matrices of size $\sim 30 \times 30$ due to transmission constraints fundamental to the design [188]. In comparison, the processor developed by LightMatter (shown in Fig. 1 in Ref. 7) integrates four 128×128 PTCs. This limitation significantly reduces the application space of PCM-based photonic crossbar arrays for MPC.

5.5 Future work

At the time of writing this thesis, the PIC design in Fig. 5.5 has not yet been finalised by the foundry. Therefore, I have not done any experimental work in this project. As a result, there is much future work to be done before we can show PCM-based photonic crossbar arrays for solving MPC problems. Even though there was no physical chip to measure, I could have added more realistic conditions to the script in Appendix A.1.1 to simulate the effect of the physical components, such as implementing limited bit precision in the matrix core and read-out.

Although this chapter aims to solve a very specific defence-relevant problem, we can use the same approach to solve QP problems in general, which is extremely useful for many types of optimisation problems. In a recent paper from Microsoft Research, the authors present a computing and optimisation platform comprising a free-space (3-D) optical processor backed by analogue electronics for non-linearity and other computations [263]. The working principle of this platform is fundamentally the same as that of the PIC we have devised in this chapter, with the main difference being that it employs bulky components in free space, whereas we propose using integrated devices on a single chip. By using integrated photonics instead of free-space and fibre-coupled components, we get a smaller footprint and additionally immunity to external perturbations, such as vibrations. For future work, we can extend the problem space to encompass some of the problems they are able to solve in Ref. 263.

Finally, a comment on the current PTC design. With the photonic crossbar array in its current form, we are limited to $\sim 30 \times 30$ matrices due to transmission constraints inherent to the design [188]. This limitation means we need to think about the problem in a new light—in Section 7.4.2, I will discuss this limitation and a potential solution in more detail.

5.6 Chapter summary

In this chapter, I have provided an outline of photonic neuromorphic computing for sensor fusion and management, with a particular focus on our computing platform, the Kalman filter, and model predictive control (MPC). The benefits of our computing platform, which consists of a photonic crossbar array with phase-change weights, include non-volatile memories and ultrafast optical modulation and detection. Furthermore, as outlined in Section 2.3.4, PCM-based photonic tensor cores (PTCs) holds promise for use in extreme environments.

However, we are limited in how often we can update the crossbar array and in the experimental complexity of our circuits. These limitations are the main reasons why we saw it necessary to pivot from a general sensor management and sensor fusion problem to the more specific MPC problem; our current platform is suboptimal for sensor fusion algorithms like the Kalman filter but could be promising for MPC.

Here, I have presented an algorithmic approach and designs for photonic crossbar arrays to solve defence-relevant MPC problems (here: target tracking), and evaluated the operating performance of PTCs in terms of speed and power consumption. We estimate that for a simple MPC problem, our device architecture can find a steady-state solution in 12 ns (80 MHz)—in comparison, digital electronics typically operate on the kHz scale. In summary, I believe that this platform holds promise for beyond-electronics implementations of MPC and similar optimisation problems.

It is hard to fail, but it is worse never to have tried to succeed.

— Theodore Roosevelt

6

Non-volatile and small-footprint field-programmable coupler array

Contents

6.1 Introduction 130
6.1.1 Conceptual overview 131
6.2 Phase shifter with Sb_2Se_3 131
6.2.1 Design considerations 132
6.2.2 Integrated microheaters 133
6.2.3 Switching energy and multi-level tunability 137
6.2.4 Endurance 139
6.3 Building block: MZI tunable coupler 140
6.3.1 Optical measurements of MZI tunable coupler 142
6.4 Non-volatile FPCA: Optical measurements 143
6.4.1 nv-FPCA as a router 143
6.4.2 nv-FPCA as a ring resonator 144
6.4.3 nv-FPCA as a delay line 146
6.4.4 Loss estimates 148
6.5 Synthesised temporal differentiator 150
6.6 Discussion 150
6.6.1 Optimising the FPCA loss 152
6.6.2 Cyclability of Sb_2Se_3 153
6.6.3 Technology analysis 155
6.7 Future work 158
6.7.1 Foundry design 158
6.7.2 Future design 159
6.8 Chapter summary 162

In this chapter, I demonstrate a prototype of a non-volatile FPCA (nv-FPCA) utilising the newly discovered low-loss PCM Sb_2Se_3 as a phase shifter. As mentioned previously, the aim of this work is to showcase defence-relevant applications where PCMs would excel relative to other technologies. Here, I present a general-purpose platform useful not only for computing but also for optical signal processing, for example, in radar applications. The significant advantages of the developed technology lie in the strength and non-volatility of PCMs, which contribute to ultra-compact devices with no static power consumption.

This chapter is structured as follows, taking a bottom-up approach. After the introduction and a conceptual overview, I present the working principle of the Sb_2Se_3-based phase shifter. Next, I present the building block of the FPCA mesh, a 2×2 unitary gate consisting of a balanced MZI with Sb_2Se_3 as the phase shifter in each arm. Then, I demonstrate the usefulness of the FPCA by showing optical measurements and simulations of a device. Finally, I discuss current limitations and future work.

6.1 Introduction

In Section 2.4.3, I presented the work done on FPCAs in the literature. To date, two technologies have been demonstrated: platforms based on the TO effect and MEMS actuators. The prevailing technology for controlling the optical path and phase has been TO phase shifters in 2×2 couplers, which offer low loss and could, in theory, be lossless (however, the use of MZI structures for easy control over coupling and phase shifting inevitably introduces some loss due to the splitter and combiner). Furthermore, TO phase shifters are typically long ($\sim 1\,\text{mm}$) and power-hungry ($>1\,\text{mW}$). In summary, TO technology is highly used and reliable, but has suboptimal metrics.

A more recent approach is based on MEMS actuators. These phase shifters and couplers have low insertion losses ($<0.5\,\text{dB}$) and small footprints ($\sim 0.13\,\text{mm}$ each). For MEMS couplers, the possibility for continuous displacement makes them

more immune to fabrication variations than MZIs, and the crosstalk is therefore also negligible [18]. In an ideal scenario, the electrostatic actuators used to set the coupling and phase would have no leakage current and would therefore be non-volatile. However, in reality, the static power consumption and switching energy are on the order of a few femtowatts and picojoules, respectively.

The current limitations of FPCAs include footprint, power consumption, and loss. First, the behaviour of synthesised structures (e.g., ring resonators) is determined by the footprint of the individual building blocks (tunable couplers and phase shifters), which currently are inconveniently long. Secondly, the power required from voltage sources to maintain a constant bias—needed to keep a coupler or phase shifter in a given state—is likely to be orders of magnitude higher than the few femtowatts reported for MEMS couplers. Thirdly, the loss per coupler limits the total length the light can travel in the FPCA mesh. We wanted to develop an FPCA with the smallest possible footprint and zero static power consumption, without sacrificing loss performance compared with other technologies: the obvious candidates for this task are low-loss PCMs.

6.1.1 Conceptual overview

Figure 6.1 shows the basic concept of the FPCA in this work. A mesh of 2×2 MZI couplers can be used for a wide range of applications, for example, routing, interconnects, and filters (Fig. 6.1a). Each of the couplers is set to the bar or cross states, or in intermediate states, allowing for any coupling ratio (Fig. 6.1b). The MZI couplers use Sb_2Se_3 as non-volatile phase shifters, which can be switched between the amorphous and crystalline states (Fig. 6.1c) by sending different heat pulses (see Section 3.6). A schematic of a 2×2 MZI coupler is shown in Fig. 6.1d.

6.2 Phase shifter with Sb_2Se_3

The phase shifters used in the MZI couplers in this chapter are based on Sb_2Se_3 with integrated microheaters for electro-thermal switching. Based on the discussion in Section 2.3.2, we selected Sb_2Se_3 as the phase shifter material, as it has the

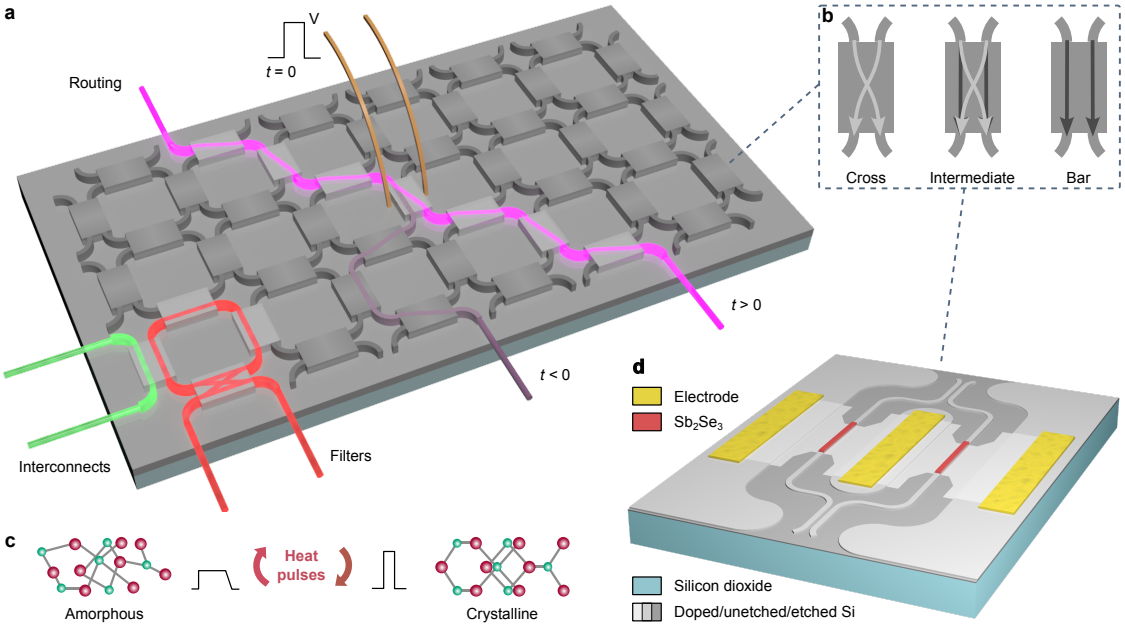

Figure 6.1: Conceptual overview of the FPCA with Sb_2Se_3-based MZI couplers. **a**, Conceptual illustration of an FPCA mesh with different optical paths that can be used for routing, interconnects, filters, and more. An electrical pulse is sent to an MZI at $t = 0$, changing the path of the pink light. **b**, Schematics showing the different states of the tunable couplers in the FPCA mesh; the cross, intermediate, and bar states. **c**, Schematic showing Sb_2Se_3 switching between the amorphous (disordered) and crystalline (ordered) states with different heat pulses. **d**, Schematic of the tunable coupler in this work: a balanced MZI with the PCM Sb_2Se_3 as the phase shifter in each arm of the MZI. The capping layer of $ZnS–SiO_2$ is not shown here.

shortest L_π, the lowest device loss, αL_π,[1] and the highest number of switching cycles reported for a low-loss PCM. Furthermore, we chose integrated electro-thermal heating of the Sb_2Se_3, as it is the preferred switching mechanism for large-scale systems with larger volumes of PCM (Section 2.3.3). In this section, I will discuss the overall design of these microheaters and the phase shifter.

6.2.1 Design considerations

Phase and amplitude modulators based on PCM with integrated microheaters come in many forms. They can have different heater technologies, for instance, based on graphene [139], ITO [199], bow-tie-shaped or bridged doped silicon [120], or p–i–n junctions [131]. Furthermore, the PCM can be deposited in one strip [120], an array of strips [131], or as sub-wavelength microdisks [199], among others. Due

[1] Although Pang *et al.* recently showed a comparable FOM for Sb_2S_3 [132].

to the ease of fabrication and our group's previous experience [195], we opted for doped-silicon heaters with a single patch of Sb_2Se_3.

As discussed in Section 2.3.2, to allow enough control of the light path in the FPCA, the Sb_2Se_3 patch must achieve at least a π phase shift. Furthermore, we are not really interested in the loss per unit length, but rather the more meaningful loss per π phase shift (αL_π), where α and L_π are functions of the Sb_2Se_3 thickness. Moreover, the total loss includes not only that of the Sb_2Se_3 but also that of the heater, meaning that large heater losses can be detrimental to the working principle of the FPCA for long heaters. Finally, we need to design the heater so that it heats the entire volume of the PCM without burning or delaminating the PCM patch.

In Section 4.2.1, we measured the loss of the heater to be $0.108\,\mathrm{dB\,\mu m^{-1}}$, which is considerable. Even if all other parts of the phase shifter were lossless, the doping loss of a 10 µm long heater would amount to more than a 1 dB loss on its own. Because initial simulations of the phase shifter using the basic ellipsometry model gave $L_\pi \approx 7\,\mu\mathrm{m}$ for 60 nm thick Sb_2Se_3, we designed heaters with 7 µm length. Only after the fabrication of these heaters did we realise that the actual L_π of 60 nm thick Sb_2Se_3 was 7.6 µm. Rather than depositing a thicker layer of Sb_2Se_3, we decided to stick with the 60 nm thick Sb_2Se_3 but have it extend slightly outside of the heater area, as the simulations shown in Section 6.2.2 showed the feasibility of this approach.

6.2.2 Integrated microheaters

To achieve programmability of the Sb_2Se_3 controlling the couplers, we engineered microheaters directly in the silicon through ion implantation (Fig. 6.2a). When we apply a voltage to the doped silicon, scattering between the electrical carriers and the crystal lattice converts electrical power into heat, raising the local temperature. This phenomenon is called Joule heating or resistive heating, with heating power given as $P = I^2R = V^2/R$. By applying carefully designed heat pulses, we can switch the Sb_2Se_3 patches on top of the waveguides to the desired state. The bow-tie shape of the heater (Fig. 6.2) is designed to effectively allow local Joule heating in the area around the PCM without causing thermal crosstalk with neighbouring devices.

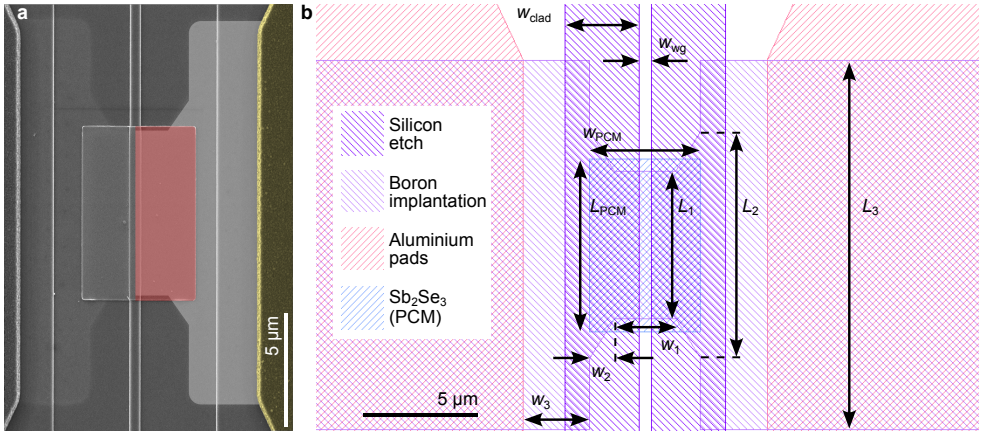

Figure 6.2: Sb_2Se_3-based phase shifter with integrated microheater and its structural parameters. **a**, Scanning electron micrograph of the phase shifter, false-coloured only in the right half, showing the electrodes (gold-coloured) and boron-implanted microheaters (light grey) with Sb_2Se_3 and a capping layer of $ZnS–SiO_2$ (red) on top. **b**, Structural parameters of the microheater. As outlined in Section 4.2, the steps of the fabrication procedure are silicon etching, ion implantation, deposition of metal pads, and deposition of PCM and capping layer. The dark purple design (silicon etch) is the half-etched SOI waveguide slab, and the white areas are 220 nm Si.

The fabricated phase shifter used in this work is shown in Fig. 6.2a, and a GDS design is shown in Fig. 6.2b. In this work, the structural parameters of the heaters are $w_{\text{clad}} = 3\,\mu\text{m}$, $w_{\text{wg}} = 0.5\,\mu\text{m}$, $w_{\text{PCM}} = 4.5\,\mu\text{m}$, $L_2 = L_1 + 3.2\,\mu\text{m}$, $L_3 = 15\,\mu\text{m}$, $w_1 = 2.5\,\mu\text{m}$, $w_2 = 1\,\mu\text{m}$, $w_3 = 2.7\,\mu\text{m}$. The lengths of the doped waveguide section and the PCM patch are $L_1 = 7.0\,\mu\text{m}$ and $L_{\text{PCM}} = 7.6\,\mu\text{m}$, respectively—these parameters were chosen based on the discussion in Section 6.2.1.

Simulations of integrated microheaters

To ensure that the entire volume of PCM reaches either the crystallisation temperature or the melting point with applied electro-thermal pulses, we conducted finite element method (FEM) simulations of the device in COMSOL Multiphysics. We used the *Electric Currents*, *Heat Transfer in Solids*, and *Electromagnetic Heating* physics interfaces to simulate Joule heating. The parameters used in the simulations are presented in Table 6.1. For the switching pulses in the simulations, the parameters were $t = 600\,\text{ns}$ and $V = 18.2\,\text{V}$ for the amorphisation pulse and $t = 100\,\mu\text{s}$ and $V = 7.0\,\text{V}$ for the crystallisation pulse. The heat capacity (C_p) of Sb_2Se_3

is given as [264]

$$C_\mathrm{p} = \frac{1000}{480.38}(123.93 - 0.01026T) \qquad T = 290 \ldots 888\,\mathrm{K}, \tag{6.1}$$

where $\frac{1000\,\mathrm{g\,kg^{-1}}}{480.38\,\mathrm{g\,mol^{-1}}}$ converts $\mathrm{J\,mol^{-1}\,K^{-1}}$ to $\mathrm{J\,kg^{-1}\,K^{-1}}$. In these simulations, the heat capacity alone is temperature dependent; all other parameters are held constant.

Table 6.1: Overview of the material parameters used for integrated microheater simulations. All parameters were taken from the default database of COMSOL Multiphysics, unless indicated by a footnote.

Property	Sb_2Se_3	SiO_2	i-Si	p^{++}-Si	Al	Unit
Thermal conductivity	0.22*	1.38	130	100†	238	$\mathrm{W\,m^{-1}\,K^{-1}}$
Heat capacity (C_p)	Eq. (6.1)	703	700	720‡	900	$\mathrm{J\,kg^{-1}\,K^{-1}}$
Density	5810§	2203	2329	2330‡	2700	$\mathrm{kg\,m^{-3}}$
Electrical conductivity	0.01¶	10^{-14}	10^{-12}	5000#	3.8×10^7	$\mathrm{S\,m^{-1}}$

* For amorphous Sb_2Se_3 [265]; † From Ref. 266; ‡ From Ref. 229; § From Ref. 267; ¶ From Ref. 268; # See Eq. (6.3).

In Fig. 6.3, we present COMSOL Multiphysics simulations of the microheater with Sb_2Se_3. Figure 6.3a shows the surface temperature of the device at the end of the pulses shown in Fig. 6.3b. The 2-D plots show that the entire volume of PCM is within the temperature contour lines representing the melting point $T_\mathrm{m} = 885\,\mathrm{K}$ and the crystallisation temperature $T_\mathrm{c} = 473\,\mathrm{K}$. Furthermore, the temporal plots show that the PCM remains at a constant temperature for a prolonged period during the crystallisation pulse and rapidly quenches at the end of the amorphisation pulse.

In the simulations, I assume an electrical conductivity of $\sigma = 5 \times 10^3\,\mathrm{S\,m^{-1}}$ for the doped silicon. This number is estimated from the formula

$$\sigma = q(\mu_n N_e + \mu_p N_h), \tag{6.2}$$

where, assuming $N_h = 2.9 \times 10^{18}\,\mathrm{cm^{-3}} \gg N_e$, we get

$$\begin{aligned}\sigma &= q\mu_p(N_h)N_h = 1.6 \times 10^{-19}\,\mathrm{C} \times 107.7\,\mathrm{cm^2\,V^{-1}\,s^{-1}} \times 2.9 \times 10^{18}\,\mathrm{cm^{-3}} \\ &= 50\,\mathrm{S\,cm^{-1}} = 5000\,\mathrm{S\,m^{-1}},\end{aligned} \tag{6.3}$$

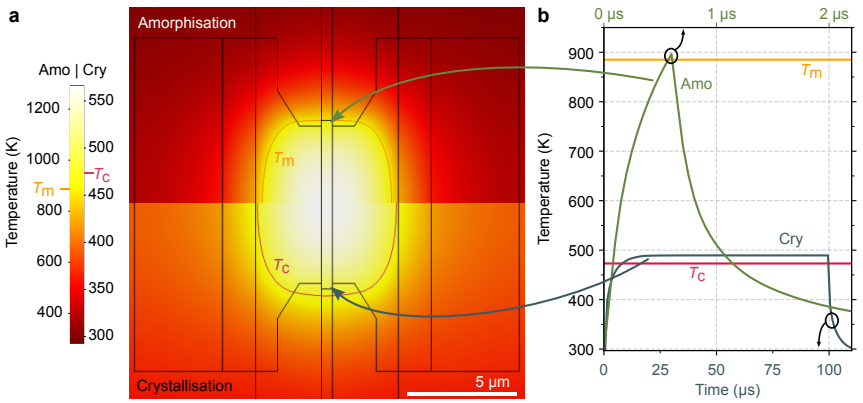

Figure 6.3: FEM simulations of Sb_2Se_3 phase shifter with integrated microheater. **a**, Microheater temperature at the end of the electrical pulse excitation shown in **b**, for amorphisation and crystallisation pulses. Note that identical colours in the two halves do not denote the same temperatures. **b**, Transient temperature at the top corner of the PCM patch (indicated by the arrows). The temperature reaches the melting point $T_\mathrm{m} = 885\,\mathrm{K}$ and crystallisation temperature $T_\mathrm{c} = 473\,\mathrm{K}$ for the amorphisation and crystallisation pulses, respectively. The pulse parameters are 600 ns and 18.2 V for the amorphisation pulse and 100 µs and 7.0 V for the crystallisation pulse.

with the relationship between $\mu_p(N_h)$ and (N_h) taken from Eq. 5 and Table 13 in Ref. 269. Because the hole density varies with the depth in the silicon, and the waveguide–slab and slab–side regions will have abrupt lateral transitions in the doping profile, I have somewhat arbitrarily assumed an effective hole density of $N_\mathrm{eff} = 2.9\times10^{18}\,\mathrm{cm}^{-3}$, which is the concentration found 90 nm into the material [270]. The maximum doping concentration occurs at a depth of around $d = 33\,\mathrm{nm}$ with $N_\mathrm{max} = 3.1\times10^{20}\,\mathrm{cm}^{-3}$ and the minimum concentration at $d = 120\,\mathrm{nm}$ with $N_\mathrm{min} = 9.5\times10^{15}\,\mathrm{cm}^{-3}$. All concentration values are taken from the Diffused Ion Implantation Profile Calculator by the Integrated Microfabrication Lab at Brigham Young University [270]. Instead of this *ad hoc* approach, we could have, for example, designed van der Pauw test structures to determine the exact electrical parameters [229] or modelled a varying electrical conductivity in the simulations.

Electrical characteristics

Figure 6.4 shows I–V sweeps of the microheaters in this work, with the heaters in the FPCA unit cell (presented in Section 6.4) shown in Fig. 6.4a. These microheaters have a resistance $R = (1.5 \pm 0.3)\,\mathrm{k\Omega}$, with the lowest recorded resistance being

1.0 kΩ. It should be noted that, similar to the heaters in the work by Ríos *et al.*, the devices show a roll-off effect at high voltages ($\gtrsim$6 V in Fig. 6.4b), owing to the temperature-dependent nature of carrier mobilities [120]. Furthermore, as mentioned in Section 4.5.2, the function generator used in this work assumes a 50 Ω impedance at the microheater, but because the actual resistance is significantly higher, the voltage from the function generator is higher by a factor $F(R) = \frac{2R}{R+50\,\Omega} \approx 1.9$, where R is the actual resistance of the microheater. In this work, when the function generator is used, the factor F is included in all quoted voltage values.

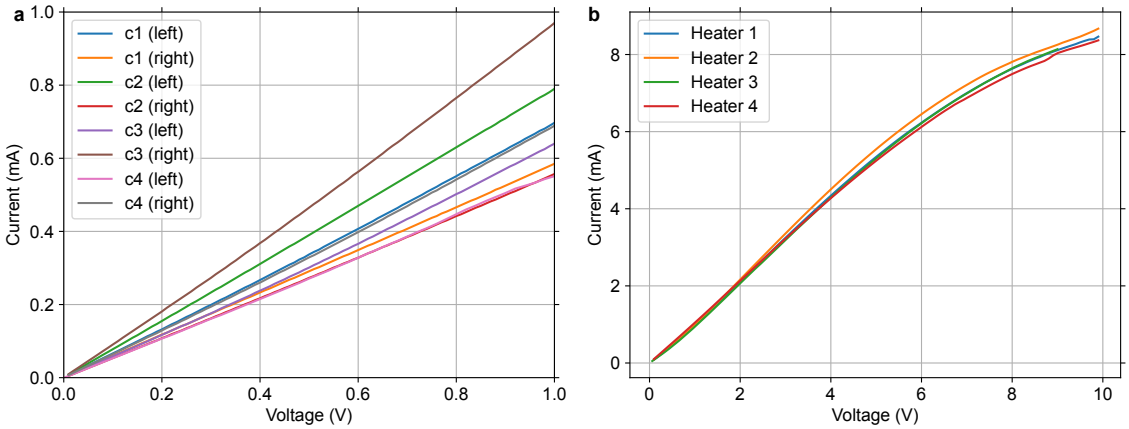

Figure 6.4: Current–voltage (I–V) sweeps. **a**, Sweeps of the microheaters in the FPCA unit cell in Section 6.4. **b**, Sweeps of heaters from a different device, showing saturation at large voltages.

6.2.3 Switching energy and multi-level tunability

Here, I show the switching energy and multi-level operation of our phase shifter. The energy consumed by a microheater, E, is given by

$$E = \int P(t)dt, \tag{6.4}$$

where $P(t)$ is the power consumed for time t. For amorphisation pulses, we use a single-step rectangular pulse, $E = \frac{V_{\text{amo}}^2}{R} t_{\text{amo}}$, where t_{amo} is the duration of the electrical pulse, and V_{amo} and R are, respectively, the voltage applied and the resistance of the microheater. For crystallisation, we sent trapezoidal pulses with energy given by

$$E_{\text{cry}} = \frac{V_{\text{cry}}^2}{R}\left(t_1 + \int_0^{t_2}\left(1 - \frac{t}{t_3}\right)dt\right), \tag{6.5}$$

where t_1 and $t_2 = 1000\,\mu s - t_1$ are the duration of the single-step rectangular section and the remaining time (limited by the pulse generator), respectively, and t_3 is the decay time from $V = V_{cry}$ to $V = 0$.

Switching measurements are presented in Fig. 6.5. In Fig. 6.5a, I show full amorphisation of the PCM. First, we completely amorphise the PCM patch with a total energy of 1.2 µJ (requiring 13 pulses of up to 0.1 µJ each, see Table A.1). This is significantly higher that the 176 nJ obtained by Ríos *et al.* for complete amorphisation [120]. Note, however, that Ríos *et al.* were able to fully amorphise the PCM patch with a single pulse, whereas we employed multiple pulses to ensure device integrity throughout the experiments. Also, the first four pulses alone changed the transmission from 100 % to 6.5 %, meaning that the large total energy is in part caused by the pulse sequences used. For optimised pulse parameters, we expect these numbers to decrease. In Fig. 6.5a, I also show a PCM patch set to >50 distinct levels by sending amorphisation pulses of varying energy, with pulse parameters found in Table A.2.

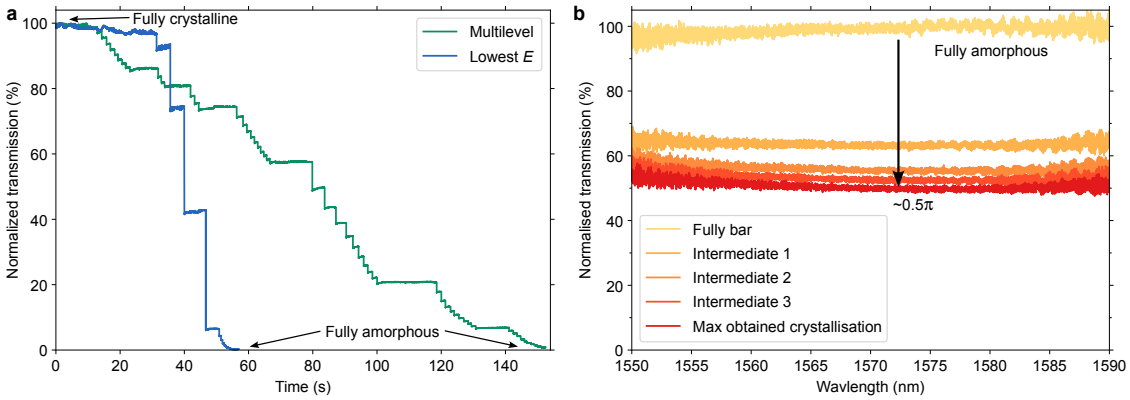

Figure 6.5: Switching of Sb_2Se_3 phase shifters with integrated microheaters. **a**, Normalised transmission from an MZI with one of the PCM patches sequentially switched from fully crystalline to fully amorphous, showing >50 distinct levels (green plot) and the lowest switching energy demonstrated (blue plot). Both PCM phase shifters of the 2×2 balanced MZI were initially in the crystalline state, and the transmission was measured from the cross port of the MZI. **b**, Normalised transmission spectra from an MZI with the phase shifter of interest initially set to the amorphous state and the other phase shifter in the crystalline state. The plots show that we can crystallise ∼50 % of the PCM over several intermediate steps. The transmission was measured from the bar port of the MZI.

Next, in Fig. 6.5, we present crystallisation measurements. Unfortunately, after

complete amorphisation of the phase shifter, we were only able to crystallise 50 % of the PCM volume, as seen in Fig. 6.5b. This limitation might be due to poor thermal engineering, material degradation, phase segregation, or partial delamination of the PCM caused by the high-temperature amorphisation pulses. The pulse parameters used for crystallisation are shown in Table A.3, with a total energy of 584 µJ over 16 pulses (of up to 50 µJ each). We assume that the total energy required for full crystallisation is twice this value, i.e., 1.16 mJ.

It is worth pointing out that, in the energy calculations in this chapter, we have accounted for the voltage factor $F(R)$ but not the saturation effect (shown in Fig. 6.4b). If we assume that the saturation of the devices shown in Fig. 6.5 was similar to that in Fig. 6.4b, then we can expect lower total energy consumption: at 8.5 V, the energy is 6 % lower than when using the resistance value obtained at 1 V, and above 10 V, the energy is at least 12 % lower. Due to the difficulty of accurately measuring this effect during short pulses, we have not included the saturation effect in the calculations in this section.

6.2.4 Endurance

For phase shifters based on PCM, one crucial metric is cyclability. In Fig. 6.6, we show experimental data of repeated switching of our PCM. The largest phase shift we were able to repeatedly show was for switching of ~ 13 % of the PCM volume (Fig. 6.6a), with the crystallisation step being the limiting factor. Due to drift issues, we did not attempt cycling until failure. Here, we find the switched volume from the equation

$$\Delta v = \frac{\Delta\varphi}{\pi} = \frac{2\arccos\left(\sqrt{T_1}\right) - 2\arccos\left(\sqrt{T_0}\right)}{\pi}, \tag{6.6}$$

where T_0 and T_1 are the normalised transmission levels before and after switching, respectively, and we have assumed that the total volume deposited is equivalent to a π phase shift. Because the bias tee can constrain pulses from the function generator of durations larger than ~ 10 µs, it was removed for the measurements shown in Fig. 6.6a.

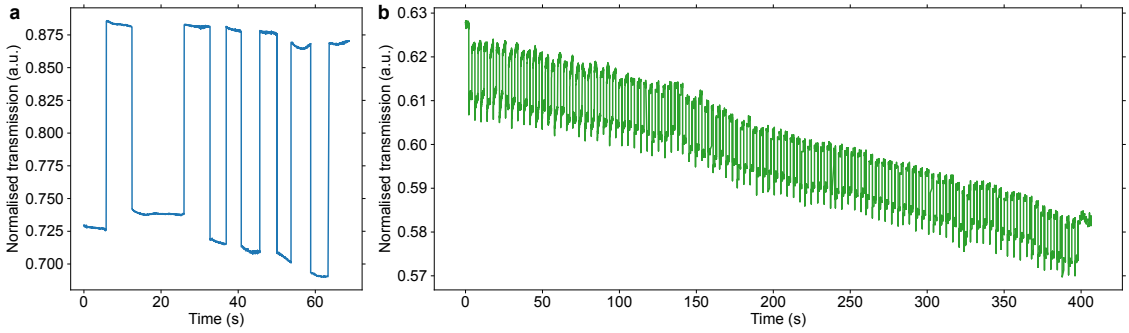

Figure 6.6: Cyclability tests of Sb_2Se_3 phase shifters with integrated microheaters. **a**, Repeated amorphous (600 ns, 18.7 V) and crystalline (200 µs, 7.4 V) rectangular pulses, switching ∼13 % of the PCM volume. **b**, Cyclability test of 100 repeated pulse sequences with rectangular amorphisation pulses (360 ns, 14.1 V) and crystallisation pulses (250 ns, 14.6 V), switching $\lesssim$1 % of the PCM volume.

Interestingly, we were able to demonstrate 100 repeated pulse sequences with sub-microsecond crystallisation pulses, as shown in Fig. 6.6b. Here, the amorphisation and crystallisation pulses are quite similar: 360 ns, 14.1 V and 250 ns, 14.6 V, respectively. However, only $\lesssim$ 1 % of the PCM volume is switched in each step, calculated from Eq. (6.6). Unfortunately, the switching contrast was not large enough for our ultrafast PD to read out the switching dynamics of the crystallisation pulse. I will discuss these measurements in larger detail in Section 6.6.2.

6.3 Building block: MZI tunable coupler

In this work, we use balanced MZIs for the 2×2 couplers that make up the mesh (see Fig. 6.1). We employ 2×2 DCs for the splitting and combining of the signals in the MZI and an Sb_2Se_3 patch in each arm for phase control. A scanning electron micrograph of an MZI is shown in Fig. 6.7a. Notice the size of the electrical pads relative to the phase shifters. Initial efforts went into using asymmetric DC switches or MMI switches instead of MZI switches, as they can significantly reduce the footprint of the coupler in both lateral directions [202, 271–273]. However, a downside of these switches is that they are more prone to fabrication variations than MZIs. Furthermore, MZIs allow for more relaxed constraints on the PCM; for MZIs, we can deposit a longer patch of PCM and switch only what we need

to achieve a complete π shift, while for DC switches or MMI switches, the exact parameters of the coupling region are more crucial [271].

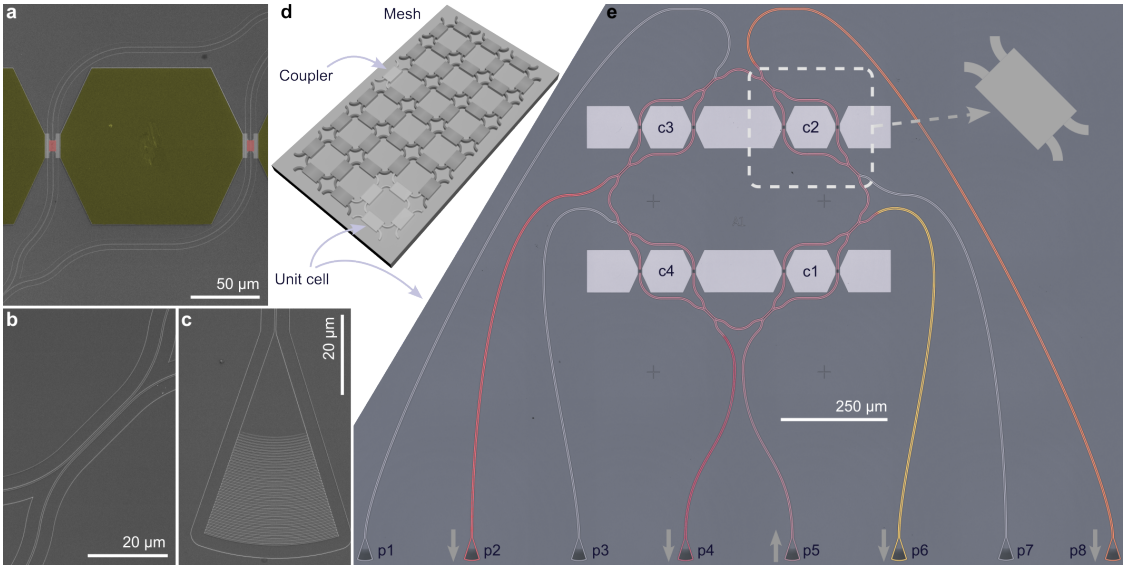

Figure 6.7: Images of FPCA components. **a**, False-coloured SEM image of a 2 × 2 MZI with an Sb_2Se_3-based phase shifter in each arm. The Sb_2Se_3 and capping layer are shown in red, doped Si in light grey, and Al pads in gold. **b**, SEM image of the 50:50 DC at the input and output of each MZI. **c**, SEM image of a grating coupler for coupling light onto the chip. **d**, Schematic showing the FPCA mesh. A single coupler and unit cell are highlighted. **e**, False-coloured optical microscope image showing the FPCA unit cell demonstrated in this work. The grating couplers are labelled p1, p2, etc., from left to right, and the MZI couplers are labelled c1–c4 counter-clockwise. The optical light paths are colourised: pink represents the input light from port p5 traversing the ring, and yellow, orange, salmon, and red represent the light exiting output ports p6, p8, p2, and p4, respectively. The inset shows one coupler as depicted in the schematic in **d**.

For 2 × 2 MZI switches, in addition to DCs, MMIs are also commonly used for the splitting and combining of the light [39]. MMIs offer wavelength- and polarisation-independent output leading to wide bandwidths, as well as superior fabrication tolerance, but at the cost of a larger footprint and loss [274, 275]. Here, we chose DCs at the MZI input and output (Fig. 6.7b) because they give more compact, low-loss structures. The main trade-off lies in the splitting ratio of the DC, which depends on the wavelength (see Fig. 4.10b).

6.3.1 Optical measurements of MZI tunable coupler

Before we show the nv-FPCA, we characterise the building block of the FPCA, which in this work is a balanced 2×2 MZI. Optical measurements of an individual MZI coupler are shown in Fig. 6.8. First, we show the tunable coupler in its extreme states, i.e., the cross state ($\Delta\varphi = 0$) and the bar state ($\Delta\varphi = \pi$). The spectra in Fig. 6.8a show low crosstalk between the ports, with maximum ER at a wavelength of $\lambda = 1582.5\,\text{nm}$ of $\text{ER}_{\text{bar}} = 22.6\,\text{dB}$ and $\text{ER}_{\text{cross}} = 27.6\,\text{dB}$ (limited by the photodetector's noise floor), with an ER of at least 20 dB over 15 nm. We also set the coupler to an intermediate state ($\Delta\varphi \approx 0.4\pi$) in Fig. 6.8b and send light into both input ports sequentially. The coupler exhibits high symmetry between the two input ports.

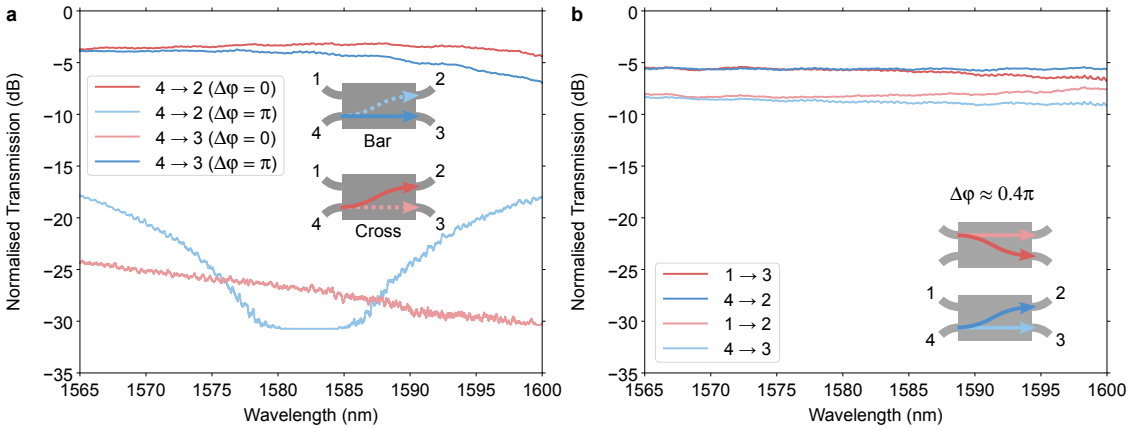

Figure 6.8: MZI coupler behaviour. **a**, Normalised transmission spectra of an MZI in the cross state ($\Delta\varphi = 0$) and bar state ($\Delta\varphi = \pi$) with light entering port 4. The inset schematics show the two settings. **b**, The MZI in an intermediate state ($\Delta\varphi \approx 0.4\pi$), measured at both outputs with light entering the inputs sequentially, as indicated by the insets.

The transmission spectra in this chapter are normalised to the transmission of a plain waveguide with etched grating couplers at the input and output (shown in Fig. 6.7c). Due to the limited bandwidth of the grating couplers on our SOI platform, the transmission spectra from all measurements, even those of plain waveguides, are enveloped by a Gaussian-like (i.e., bell-shaped) envelope. If the transmission of the grating couplers varies slightly between devices (for instance, due to less efficient optical free-space-to-waveguide conversion or a slightly shifted envelope), then the

normalisation might skew the plots at extreme wavelengths. This effect is visible in Fig. 6.8b, where the red measurements (input to port 1) are slightly slanted for the longest wavelengths, while the blue measurements (input to port 4) are flat in the entire wavelength range. This slant is most likely caused by the transmission envelope not matching that of the normalisation spectrum. Due to a similar wavelength-dependent variation in Fig. 6.8a, it is difficult to accurately estimate the device's loss. However, as a first estimate, the tunable coupler has an IL compared with the plain waveguide of $\mathrm{IL_{cross}} < 3.2\,\mathrm{dB}$ and $\mathrm{IL_{bar}} < 4.0\,\mathrm{dB}$ for $\lambda = 1582.5\,\mathrm{nm}$.

6.4 Non-volatile FPCA: Optical measurements

In this work, we demonstrate a square FPCA mesh unit cell consisting of four MZI couplers, as shown in Fig. 6.7. A schematic of an FPCA mesh, a unit cell, and an individual coupler is shown in Fig. 6.7d, while the demonstrated FPCA unit cell with four MZI couplers is presented in Fig. 6.7e. The optical light paths relevant to this section are colourised, with the pink colour representing the input light from port p5 traversing the ring and the yellow, orange, salmon, and red colours representing the light out of output ports p6, p8, p2, and p4, respectively. We proceed with demonstrating that the reconfigurable FPCA mesh can be used to synthesise three essential optical components: routers, ring resonators, and delay lines.

6.4.1 nv-FPCA as a router

First, we demonstrated routing of light to the different output ports of the device, as indicated by the schematics in Fig. 6.9. In the four figures, light is sent into port p5 and the FPCA is configured to route light out of ports p6, p8, p2, and p4, respectively. The corresponding configurations of the FPCA are: coupler c1 in the bar state (p6 has the highest output); c1 and c2 in the cross state (p8 has the highest output); c1 in cross, c2 in bar, and c3 in the cross state (p2 has the highest output); and finally, c1 in cross, c2 and c3 in bar, and c4 in the cross state (p4 has the highest output). The FPCA shows broadband behaviour, effectively routing light over a wide range of wavelengths. The extinction ratios between the different

output ports when the FPCA unit cell is in the configurations in Fig. 6.9a–d are >20 dB for wavelengths between 1581 nm and 1588 nm.

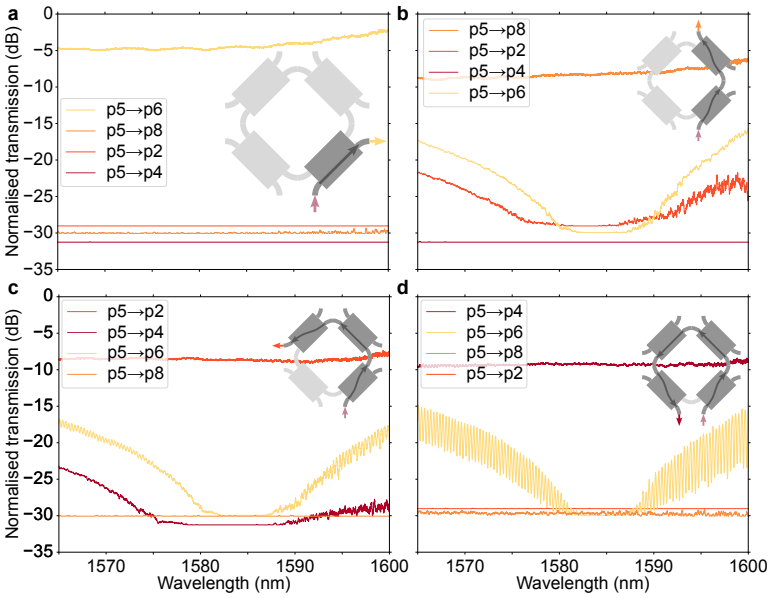

Figure 6.9: Routing of light out of different output ports of the nv-FPCA. **a**–**d**, Normalised transmission spectra from the different output ports of the FPCA unit cell, for when the c1 coupler is in the bar state (a), c1 and c2 are in the cross state (b), c1 in cross, c2 in bar, and c3 in cross state (c), and finally, c1 in cross, c2 and c3 in bar, and c4 in cross state (d). The insets in each figure show the configuration for each measurement.

6.4.2 nv-FPCA as a ring resonator

To further demonstrate the potential of our FPCA unit cell, we configured it to synthesise a ring resonator, i.e., when three of the tunable couplers guide light around the unit cell and the fourth is set to an intermediate state governing the coupling into the ring. In Fig. 6.10, we show transmission spectra for the synthesised ring resonator when the cross-coupling coefficient of the ring is close to 0 (bar state) and 1 (cross state), as well as several intermediate states, including close to critical coupling. In Fig. 6.10a,b, the colours represent the sequential tuning from the initial state (yellow) to the final state of the measurements (dark red) with several intermediate states. As seen in Fig. 6.10a, we can shift the resonance of the ring by simultaneously changing the PCMs of both arms of the guiding tunable couplers, i.e., changing φ in both arms of the tunable coupler but maintaining $\Delta\varphi$ between

the arms (as described in Fig. 6.10c). Here, the total achievable shift was about 0.25π. However, larger resonance shifts can be achieved by depositing longer PCM patches, enabling greater phase control.

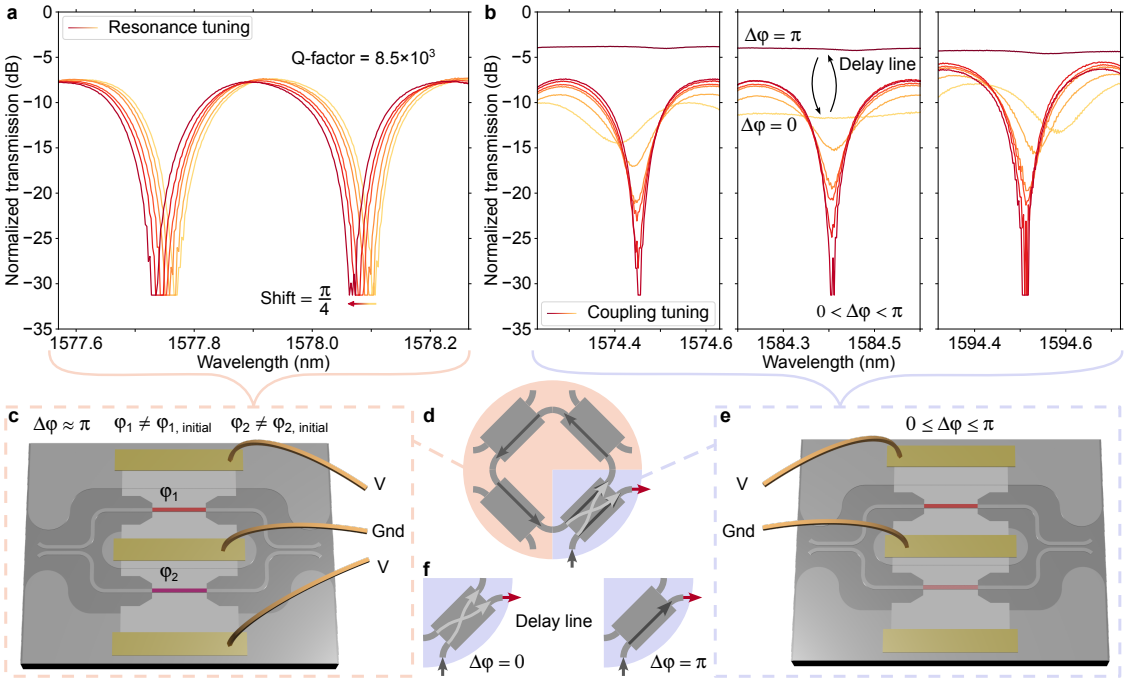

Figure 6.10: The nv-FPCA synthesised as a ring resonator and delay line. **a**, Normalised transmission spectra showing tuning of the resonance wavelength of the synthesised ring resonator near critical coupling. Tuning is achieved by changing the effective refractive index of the ring using the phase shifters in the couplers in the ring (peach-coloured in panel **d**). **b**, Normalised transmission spectra of the ring resonator at different states; with the input tunable coupler in the bar state ($\Delta\varphi = \pi$), in the cross state ($\Delta\varphi = 0$), and in several intermediate states ($0 < \Delta\varphi < \pi$), including close to critical coupling (sharpest resonance). In **a** and **b**, the colours represent the sequential tuning from the initial state (yellow) to the final state of the measurements (dark red) with several intermediate states. **c**, Schematic showing the coupling conditions of the guiding couplers in the ring (not the input coupler) for the transmission spectra in panel **a**. **d**, Schematic of the synthesised ring structure. **e**, Schematic showing the coupling conditions of the input coupler for the transmission spectra in panel **b**. **f**, Schematics showing how the FPCA unit cell can be used as a reconfigurable delay line by setting the input coupler to the extreme states.

Next, we show that both the resonance wavelength and transmission are changed when we switch the MZI used for coupling into the ring, as shown in Fig. 6.10b. In the figure, we go from the cross state through several intermediate states, including one close to critical coupling, to the bar state. Notice how Fig. 6.10a,b are similar to the figures in the work by Lian *et al.* [276], where the authors use a single

PCM element instead of MZIs for non-volatile phase modulation and coupling (although at the drop port) of a ring resonator.

We find the FSR and Q factor of the synthesised ring by fitting the troughs in the spectra to Lorentzian functions, as shown in Fig. 6.11. We used the data analysis software Origin to fit nine resonances and show here seven of the fits and the cumulative fit for the synthesised ring with the input coupler in the bar state (Fig. 6.11a), near critical coupling (Fig. 6.11b), and in the cross state (Fig. 6.11c), corresponding to the plots in Fig. 6.10b. Note that we were not able to fit the curve for the wavelength region around 1584 nm for the bar state plot in Fig. 6.11a. The synthesised ring resonator at near critical coupling has a loaded Q factor of 8.5×10^3, an FSR of 0.33 nm (40 GHz), and a ring extinction ratio of $\mathrm{ER}_{\mathrm{ring}} > 23.7$ dB for the resonances shown in Fig. 6.10a,b.

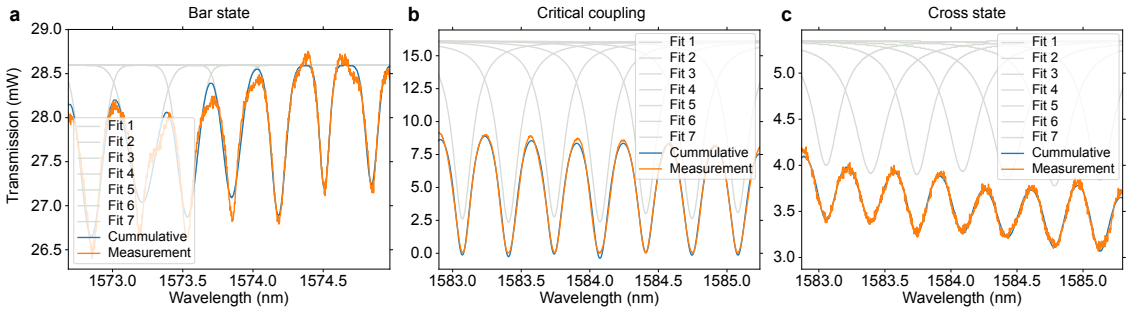

Figure 6.11: Lorentzian fits to the resonances of the synthesised ring resonator corresponding to the plots in Fig. 6.10b. **a**, When the in-coupler is in the bar state. **b**, When the ring is near critical coupling. **c**, When the in-coupler is in the cross state. The fit indices increase from left to right, with the cumulative fit based on nine resonances (one on each side of the shown plot). Note that the fit for the bar state is suboptimal and that we were not able to fit to the tiny resonances around 1584 nm.

6.4.3 nv-FPCA as a delay line

Finally, we show that we can synthesise a reconfigurable delay line using the FPCA. With the input tunable coupler (p5) in the bar state (cross-coupling set to 0), the light does not couple to the ring (marked with $\Delta\varphi = \pi$ in Fig. 6.10b,f). However, with the coupler in the cross state (coupling set to 1), almost all the light goes into, around, and out of the ring without interfering with itself (marked with $\Delta\varphi = 0$ in Fig. 6.10b,f). By alternating between these two extreme states, the FPCA effectively

functions as a non-resonant delay line. As can be seen in Fig. 6.10b, this is only true for wavelengths around 1584 nm, where the DCs at the input and output of the MZI have 50:50 splitting and the MZI therefore shows maximal extinction. With the input coupler in the cross state, the ER of the resonance at 1584.4 nm is only $\mathrm{ER}_{\mathrm{ring}} = 0.6\,\mathrm{dB}$, which is validated by simulations in Fig. 6.12. The synthesised delay line in this work has a group delay of 28 ps (the ToF through four additional MZIs), which can be increased by adding more unit cells in a larger FPCA mesh.

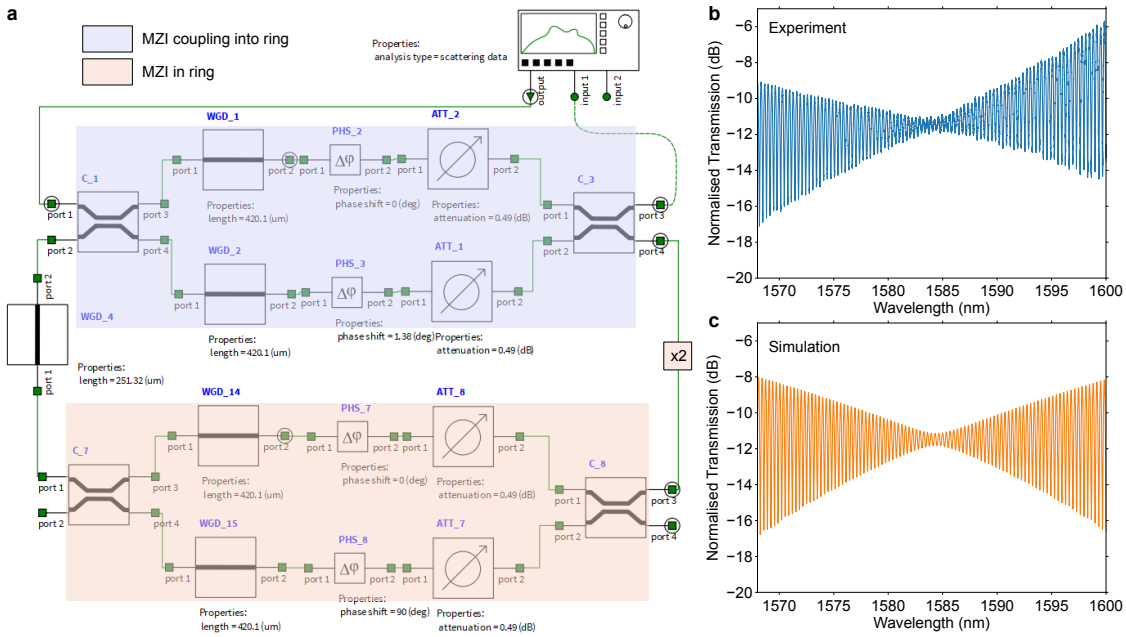

Figure 6.12: Synthesised delay line simulations. **a**, Setup used in the simulation software Lumerical INTERCONNECT. The purple rectangle indicates the MZI coupling light into the ring, whereas the peach-coloured rectangle indicates one of the MZIs forming the ring. Note that there are three peach-coloured MZI couplers, but two have been removed to save space. **b**, Normalised transmission spectrum of the cross-state measurement in Fig. 6.10b. The shape of the envelope is determined by the wavelength dependence on the coupling coefficient of the DCs. **c**, Normalised simulation data from Ansys Lumerical INTERCONNECT with imperfect DCs given by $k^2 = 4.5 \times 10^{-3}\,\mathrm{nm}^{-1} \times \lambda - 6.6298$. The similarity with the plot in **b** is striking.

In Fig. 6.12a, we show the setup used in Ansys Lumerical INTERCONNECT for simulations of the bow-tie effect seen in the cross-state measurement ($\Delta\varphi = 0$) in Fig. 6.10b. Note that there are three MZIs forming the ring (peach-coloured MZI couplers), but two have been removed to save space in the figure. In Fig. 6.12b, we show the cross-state measurement in Fig. 6.10b (also shown in Fig. 6.11c) over a

wider wavelength range. The shape of the envelope is determined by the wavelength dependence on the coupling coefficient of the DCs used in the MZI. In Fig. 6.12c, we show simulation data from Ansys Lumerical INTERCONNECT trying to replicate this behaviour, for the following parameters: the purple-coloured MZI in Fig. 6.12a is close to the cross state but has a tiny phase difference between the arms of 1.38°, the peach-coloured MZIs are in the bar state, the attenuation of each arm of the MZI is 0.49 dB, the total circumference of the synthesised ring is 1935.3 µm (same as the fabricated ring), and the DCs used for coupling into and out of the MZIs have a wavelength-dependent coupling given by $k^2 = 4.5 \times 10^{-3}\,\mathrm{nm}^{-1} \times \lambda - 6.6298$, which gives 50 % coupling at $\lambda = 1584.4\,\mathrm{nm}$ and ±10 % coupling at ±22.2 nm. The simulation data show good overlap with the experimental measurement.

6.4.4 Loss estimates

In Section 6.3.1, I estimated the IL of the tunable coupler compared with a plain waveguide to be $\mathrm{IL}_{\mathrm{cross}} < 3.2\,\mathrm{dB}$ and $\mathrm{IL}_{\mathrm{bar}} < 4.0\,\mathrm{dB}$ for $\lambda = 1582.5\,\mathrm{nm}$. From Fig. 4.2, we know that about 0.8 dB stems from the ion implantation. Furthermore, at least 0.1 dB can be attributed to each DC in the MZI [18, 277]. The remaining loss is due to the PCM (expected to be <0.8 dB), the longer optical path length of the waveguides, and the extra bends of the MZI.

It is challenging to precisely estimate the IL of the devices due to variations in grating couplers, the fibres of the fibre array, alignment, etc. Furthermore, due to limited chip space, we had few reference devices for optical loss measurements of waveguides and no devices to estimate bend loss. In the figures shown in this chapter, we have normalised the transmission to that of a plain waveguide and accounted for the slight difference in performance of the photodetectors used. One issue with the way the measurements are conducted is that the normalised transmission can appear tilted (as seen in Figs. 6.8 and 6.9) if the Gaussian-like shape of the transmission is slightly offset from that of the reference measurement, which makes it hard to estimate the loss of our devices. Furthermore, the different

fibres in the fibre array exhibit slightly different transmission, but this is difficult to accurately measure without extensive testing.

For the FPCA unit cell, the total system loss relative to a plain waveguide was 4.5 dB, 8.1 dB, 8.7 dB, and 9.3 dB at $\lambda = 1584\,\text{nm}$ for the configurations shown in Fig. 6.9a–d, respectively. Furthermore, the loss of the light traversing the entire delay line in Fig. 6.10b was 11.0 dB at the same wavelength. These numbers clearly show that the measured loss of the FPCA does not increase linearly with the number of MZIs.

We believe that the different path lengths of the connecting arms (shown in Fig. 6.7e) contribute significantly to the variation in estimated loss. The sample was annealed in a tube furnace at elevated temperatures, with impurities in the tube and gold alignment marks on the chip, which we expect contaminated the sample, leading to significant waveguide losses. This loss was not quantified through measurements.

Because the measurement parameters (waveguide length, grating couplers, optical fibres, and photodetector) are the same for the measurements in the delay line measurements in Fig. 6.10b, we can estimate the loss of our devices from these measurements. In the first case, the light only sees coupler c1 in the bar state, while in the second case, the light passes through coupler c1 in the cross state, then through the three other couplers in the bar state, and then again through coupler c1. The loss of coupler c1 can be estimated through the set of equations $\text{IL}_{\text{c1, bar}} + x = 4.5\,\text{dB}$ and $2\text{IL}_{\text{c1, cross}} + 3\text{IL}_{\text{bar}} + x = 11.0\,\text{dB}$, where x is excess loss from the waveguides, grating couplers, etc., and IL_{bar} is the insertion loss of the bar state couplers. Assuming $\text{IL}_{\text{bar}} = \text{IL}_{\text{cross}} + 0.4\,\text{dB}$ (due to the difference in PCM loss, with one arm crystalline and the other amorphous), the set of equations gives us $\text{IL}_{\text{bar}} = 1.83\,\text{dB}$ and $\text{IL}_{\text{cross}} = 1.43\,\text{dB}$. Here, IL_{cross} is the state with both PCM patches in the amorphous state and IL_{bar} with one patch in the crystalline state. In the case where both patches are in the crystalline state ($\text{IL}'_{\text{cross}}$, which is also the cross state but with different phase shift), we estimate the loss to be $\text{IL}_{\text{cross}} = 1.43\,\text{dB} + 0.8\,\text{dB} = 2.23\,\text{dB}$.

6.5 Synthesised temporal differentiator

Ring resonators can act as many different types of filters, as well as high-speed temporal differentiators, integrators, and Hilbert transformers [29, 30, 173, 205]. In Section 3.4.1, we introduced the concept of temporal differentiation, which is found in MWP systems and is used, for example, to detect signal edges in radar processing and image recognition [29]. Here, we wanted to demonstrate the power of our reconfigurable photonic platform for a real-world application by synthesising a ring resonator for temporal differentiation using our nv-FPCA platform.

Figure 6.13 shows simulations of the synthesised ring resonator in Ansys Lumerical INTERCONNECT. In Fig. 6.13a, we show the simulation setup, with high-speed oscilloscopes measuring the electrical signals from the PDs after the amplitude modulator and ring resonator, respectively. In the simulations, we select a probe wavelength at the bottom of a trough, here $\lambda = 1559.765\,\text{nm}$ (Fig. 6.13b), and set the ring parameters to those of the synthesised ring in our experimental setup: $L = 1935.26\,\mu\text{m}$, $n_{\text{eff}} = 2.5$, $n_{\text{g}} = 4.3$. The total ring loss is set to 11 dB. In Fig. 6.13c, we see that the output from the ring resonator (measured by oscilloscope OSC2 in Fig. 6.13a) gives an optical response at the start and end of the modulation pulse (measured by OSC1).

We attempted to measure the temporal response of the synthesised ring resonator experimentally, but were unable to capture the response signals. As the response time in the simulations was 28 ps (corresponding to the round-trip time of the synthesised ring), this pulse is too fast for the oscilloscope in our lab (4 GHz, with a 0 % to 90 % signal rise time of 100 ps), and even for the fast PD (12 GHz).

6.6 Discussion

In this work, we demonstrated a square FPCA unit cell with four MZI couplers. This design choice was made for several reasons. First, the size of the fibre array determines the maximum number of input/output ports, which in our lab is limited to 16. Consequently, we cannot reach the system complexity demonstrated by

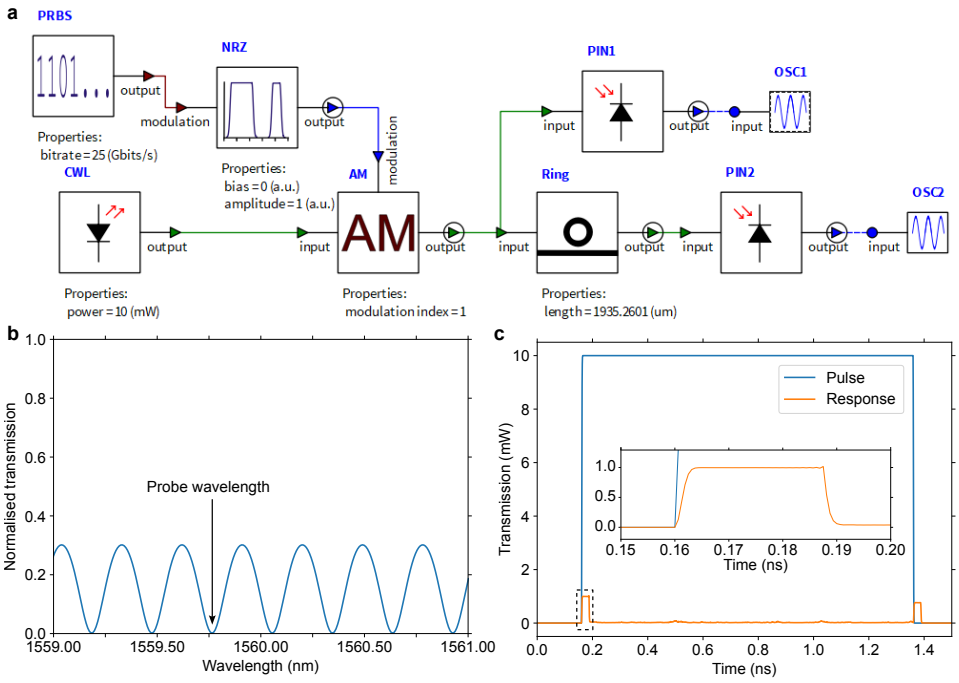

Figure 6.13: Simulation of a synthesised temporal differentiator in Ansys Lumerical INTERCONNECT. **a**, The simulation setup. **b**, Normalised transmission spectrum of the ring resonator. **c**, Temporal transmission at the input and output of the ring resonator with pulsed input. The inset shows the resonator response at the start of the pulse. PRBS: Pseudorandom binary sequence generator; NRZ: non-return-to-zero pulse generator; OSC: oscilloscope; PIN: p–i–n PD; CWL: CW laser; AM: optical amplitude modulator.

iPronics [170]. Nevertheless, it should be possible to measure all ports of a square mesh with four unit cells (organised in a 2×2 matrix). Moreover, not all MZI couplers at the edge of the mesh need to route light to a grating coupler; instead, pairs of couplers can be connected directly along the edge to close the open ends [278]. By closing several edges in this manner, even larger structures can be designed. However, as the system's complexity is already considerable, we initially opted for a single unit cell, given the limited chip area and the difficulty of reading out the state of MZI couplers that are not on the edge of the mesh.

In literature, three different FPCA configurations have been demonstrated so far: triangular, square, and hexagonal meshes. Although the square unit cell is the least promising of the three (e.g., in terms of the number of switching elements per unit area and losses per spatial resolution) [169], it provides the simplest heater design for electrical probing. In this work, we therefore designed a square unit cell.

6.6.1 Optimising the FPCA loss

As mentioned in the Introduction of this chapter, the three main limitations of current FPCAs are the loss, power consumption, and footprint of the tunable couplers. Here, I will discuss the former, while the latter two will be covered in Section 6.6.3.

To reduce the FPCA loss, we need to look at the individual contributions to the total loss. First of all, we should pick the PCM thickness that leads to the lowest loss for a π shift. As mentioned in Section 4.4.3, the authors in Ref. 132 present in their Fig. S3 a curve showing the loss vs. Sb_2S_3 thickness and pick the global minimum for their device. However, the optimal length for reducing overall system loss is not necessarily the one that minimises PCM loss, due to the non-negligible heater losses. Instead, we should pick the PCM thickness that minimises the combined loss of the heater and PCM.

A clear example of this sub-optimal minimisation was recently demonstrated. In the work by Chen *et al.*, they obtained an impressive loss of only $0.040\,\mathrm{dB}/\pi$ for their 10 nm thick and 77.81 µm long Sb_2Se_3 phase shifter (presented in their Supplementary Section S3) [278]. In comparison, we obtained $0.34\,\mathrm{dB}/\pi$ for 30 nm thick and 16.1 µm long Sb_2Se_3 and $0.80\,\mathrm{dB}/\pi$ for 60 nm thick and 7.6 µm long Sb_2Se_3. In other words, the loss of the hybrid waveguide in Ref. 278 is an order of magnitude smaller than that of our devices. However, Ref. 278 does not include the loss of the microheaters. In a previous work by the same group, they report a loss of $0.02\,\mathrm{dB\,\mu m^{-1}}$ for p–i–n microheaters similar to those used [279]. For the considerable length needed in their work, the total loss of the hybrid waveguide and microheater will therefore be $1.6\,\mathrm{dB}/\pi$; if we had used the same p–i–n heaters with our 60 nm thick Sb_2Se_3, we would only have a loss of $0.95\,\mathrm{dB}/\pi$.

Although the above discussion is technically valid, another consideration is the coupling-dependent loss in tunable couplers, which is crucial to system performance in large systems. For the phase shifters in this work, the difference in loss between the cross and bar states is estimated to be 0.4 dB. This coupling-dependent loss

means that almost 10 % of the light is lost in bar-state couplers relative to cross-state couplers. Compared with TO couplers, the coupling-dependent loss difference in our devices is large, while it is comparable to the 0.25 dB loss of the MEMS couplers in Kim *et al.* [18]. Importantly, for some synthesised structures, such as unbalanced MZIs, such loss can lead to power imbalance in the arms of the MZIs, which is detrimental to device performance. Therefore, one might be willing to trade off minimising the total system loss (i.e., $\alpha' L_\pi$, where $\alpha' = \alpha_{\text{cry}} + \alpha_{\text{heater}}$) for minimising the coupling-dependent loss (i.e., $\alpha_{\text{cry}} L_\pi$ with *only* the crystalline PCM loss included).

6.6.2 Cyclability of Sb_2Se_3

In this work, I showed limited cyclability for a small volume of Sb_2Se_3. However, this limited cyclability appears to be a common topic in the literature. When Ríos *et al.* demonstrated 125 switching cycles for Sb_2Se_3 deposited by thermal evaporation, the transmission contrast between the two output arms, ΔT, was only between 0 dB and 3 dB [120], meaning the switching occurred only for about 0.15π. In Ref. 139, they show 1000 switching cycles of Sb_2Se_3 in a ring resonator with a ring extinction ratio of >25 dB, but the transmission contrast is only achieved for 1.3 dB initially and 0.7 dB after 1000 cycles. Notice also how the authors change from 6.8 V to 6.4 V for the amorphisation pulses between the transmission spectrum in their Fig. 3d,e to limit the amount of amorphisation in the repeated switching experiments. Based on the values in their Fig. 3d, their switching gives about a 0.075π phase shift (found by comparing the wavelength shift with the FSR), which, at around −10 dB, would give a transmission change of ∼4 dB. The fact that the transmission contrast is <1.3 dB in their Fig. 3e indicates that the phase shift in the cyclability test is in fact likely smaller than 0.075π. In the recent work by Chen *et al.*, the authors demonstrate reversible switching of Sb_2Se_3, switching $\sim 0.23\pi$, but only over five cycles [278]. In Yang *et al.*, the authors show 3000 cycles for an Sb_2Se_3 patch and 1×10^4 cycles for an array of Sb_2Se_3 strips in an MZI. The switching contrast for the cycles is in the range ∼15 dB to 1.1 dB for the patch

and 4.5 dB to 2.2 dB for the array. Although the initial 15 dB is impressive, the final 1.1 dB contrast is tiny compared to the ER of the MZI of >25 dB, and is equivalent to an estimated phase shift of less than 0.1π.

More recently, Pang *et al.* demonstrated over 750 cycles with an MZI switch with Sb_2S_3 [132]. The ER of the switch was >18 dB, while the switching was ∼8 dB, giving an estimated phase shift of $\sim\pi/2$ (assuming the switching occurred in the linear region in the middle of the sinusoidal transfer function of the MZI). Furthermore, the authors were able to show both impressive cyclability and high tunability with a bit resolution of 7 bits, with a loss per π shift of only 0.25 dB. The low loss stems in part from a 45 nm thick layer of SiO_2 that serves as a screening layer between the doped Si and the Sb_2S_3. However, due to this layer and other design considerations, the required amorphisation voltage was as high as 27 V.

These numbers reveal that the reconfigurable platforms demonstrated in this and other works are not truly reconfigurable. For reconfigurable devices based on low-loss PCM to work as intended, we should ideally achieve repeated switching of at least a π phase shift. Therefore, in this work, we are hesitant to claim that we have achieved a specific number of switching cycles for our PCM.

There might be several reasons for this limitation. First, we used a shared sputtering system, which may have led to material degradation due to impurities during sputtering. Due to large heat pulses applied during amorphisation, the Sb_2Se_3 patch might have partially delaminated, or the atoms in the material might have segregated. In this project, we used microheaters that were proven to work for short patches of GST, but the heater geometry was not optimised for long patches of Sb_2Se_3. For future devices, better thermal engineering is expected to significantly improve heater loss, switching energy, and cyclability.

Based on this discussion, it is striking that reliable switching of large-volume, low-loss PCMs remains poorly understood and would likely require a significant scientific commitment. The level of investigation required is likely similar to that found in Ref. 280. Although not conducted in this thesis, such a work is extremely timely, interesting and important for the community.

Sub-microsecond switching

The exact mechanism behind the sub-microsecond crystallisation in Fig. 6.6b is not fully understood. One explanation could be that the crystallisation pulse heats the material just above the glass-transition temperature ($T > T_c$) but below the melting point ($T < T_m$), and that a small volume crystallises each time. In both Ríos *et al.* [120] and Yang *et al.* [131], the authors observed crystallisation with pulses down to 5 µs in length, but to the best of our knowledge, no one has reported sub-microsecond switching in literature. However, as commented by Ríos *et al.*, smaller pulses lead to smaller hotspots with $T > T_c$, which means the heat pulse is crystallising only small volumes of Sb_2Se_3 [120]. This is likely the explanation for what we observe. Because the switching contrast was insufficient for our ultrafast PD to resolve the switching dynamics of the crystallisation pulse, we were unable to confirm the dynamic behaviour of the sub-microsecond switching pulses experimentally.

6.6.3 Technology analysis

In the introduction to this chapter, I mentioned the currently existing FPCA technologies. In Table 6.2 (found at the end of this chapter), we present the main parameters of these FPCA demonstrations and compare their performance to our technology. In addition, we show the performance of a future PCM-based device with a thinner Sb_2Se_3 patch of 30 nm and p–i–n junctions for reduced overall loss (see Section 6.7.2). Notably, the PCM-based couplers in the nv-FPCA feature substantially shorter active lengths than those in other implementations and are non-volatile, removing the need for continuous electrical driving from relatively power-hungry voltage sources. A visual guide of this is shown in Fig. 6.14a.

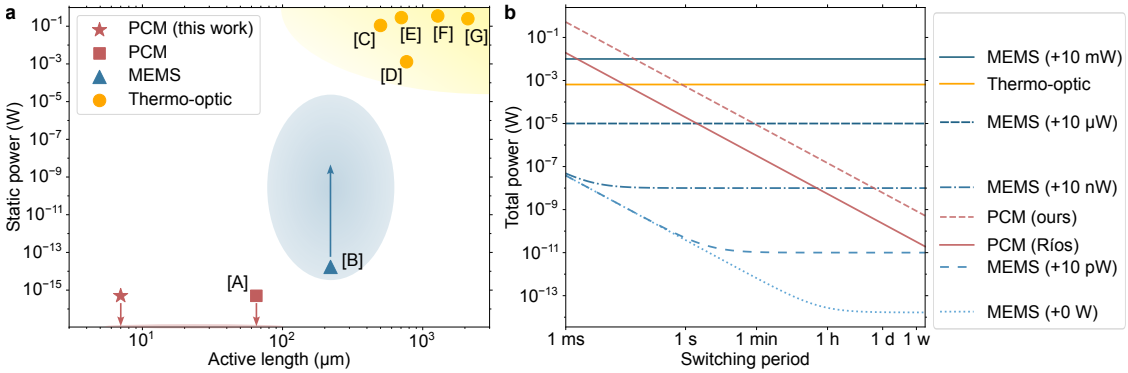

Figure 6.14: Power and active length comparison between different FPCA technologies. **a**, Static power consumption and active length of the coupling (and phase-shifting element, in the case of MEMS) of demonstrated photonic recirculating mesh networks. [A] is [278], [B] is [18], [C] is [169], [D] is [170], [E] is [168], [F] is [41], and [G] is [166]. For the MEMS devices, the power consumption of electronic control circuits (such as DACs) significantly exceeds the power of the devices themselves, meaning the actual static power consumption is substantially higher, as indicated by the blue arrow. **b**, Power consumption vs. switching period for different technologies. For the MEMS devices, the number in parentheses indicates the total power consumption of the DAC needed to set the state of the coupler and phase shifter.

Low-loss PCMs offer a compelling complementary alternative to the existing FPCA technologies shown in Fig. 6.14 and Table 6.2. PCMs allow for the smallest footprint of all optical materials platforms [92], which is crucial to improve the granularity of the FPCA mesh and to synthesise resonant structures with larger FSRs. PCMs also offer stable operation with long retention times [124], as well as wide bandwidth owing to the small wavelength dependence of Δn in the wavelength region of interest. Because each phase shifter needs not only PCM but also a heater, the insertion loss of even the most optimised devices is likely to remain at the higher end of the range compared with TO and TO platforms, but still within an acceptable range. A related issue is the phase-dependent loss between the crystalline and amorphous states, leading to a power imbalance in the MZI output. This drawback might be overcome using different coupler designs [273]. A major limitation at the moment, and the focus of significant ongoing research, is the material degradation and limited cyclability of low-loss PCMs in optical devices. However, we believe that as the technology matures, we will reach levels acceptable for infrequent switching applications such as FPCAs. The relatively high switching

energies and long reconfiguration times of low-loss PCMs are not ideal, but are also not detrimental to FPCAs, where stable, non-volatile couplers are preferred. Finally, as the technology matures, we also expect to see true CMOS compatibility, including integration into foundries and lower switching voltages.

The total power consumption of the nv-FPCA building blocks depends on the number of switching events per second. Of the other technologies presented in Table 6.2, the MEMS devices are by far the most energy efficient, with picojoule switching energy and femtowatt standby power dissipation. In comparison, PCM-based couplers have nano-to-millijoule switching energy but consume zero standby power. With the conservative switching energies demonstrated in Ríos *et al.* [120], PCM heaters require, on average, 19.3 µJ per switching event; therefore, for PCM-based couplers to be more energy efficient than MEMS couplers, the switching needs to happen less than every 36 years. However, this calculation assumes no power cost of the electrical driving of the MEMS devices. In reality, the actual power consumption of the MEMS couplers will be dominated by the driving electronics [85], such as the DACs setting the state of the couplers (requiring, e.g., from nanowatts [281, 282] to milliwatts [283–285]) and other control components, as indicated by the arrow in Fig. 6.14a.

In Fig. 6.14b, we show the total power consumed by the technologies as a function of the switching period. Assuming a total driving power consumption of 10 nW, the crossover point where PCM outperforms MEMS happens for switching once every 32 minutes, and already after less than 2 seconds for 10 µW. For applications where the mesh is set infrequently, the advantage of non-volatility becomes evident. An additional advantage of non-volatility is that a single DAC can be used to sequentially set the state of the phase shifters, whereas other technologies require one per phase shifter. Moreover, in the event of an electrical power loss, the mesh does not require reconfiguration.

The different FPCA technologies all aim to solve the same problem but may have quite different use cases. We foresee the following for the three different technologies. First, the low loss, large footprint, and high power consumption of the

TO platform make it suitable for most prototyping and lab applications with relaxed SWaP requirements, as well as for applications where optical performance matters. Secondly, the MEMS platform is a great alternative that reduces footprint and, especially, power consumption, and can be used in many applications, including cryogenic environments. Thirdly, the PCM-based platform will perform optimally in space- and power-limited systems in extreme environments with infrequent FPCA switching requirements.

Finally, it is worth commenting on the total power consumption shown in Fig. 6.14b. In the figure, we have included only the power consumption of the mesh (and electrical driving in the case of MEMS devices), but we have not included the laser power. This choice can be justified because a mesh that requires less power is always beneficial, so it is irrelevant to include laser power. For instance, as noted in Ref. 18, a low-power mesh (i.e., with low heat dissipation) but a high-power laser can still be used in cryogenic environments, provided the laser is outside. However, this argument is not valid for edge devices, where the device's total power must be taken into account. In such scenarios, we would need to compensate for optical loss in the mesh by increasing the laser's output power, which can increase overall power consumption. Furthermore, this compensation only works up to a limit, as we do not want to damage the devices by inputting large optical powers. In other words, for lossy FPCAs, the size of the mesh is limited by the loss per component.

6.7 Future work

In the previous sections, I discussed many of the limitations of this work and the challenges our technology faces. Here, I will not repeat these; instead, I will present some of the work done on future nv-FPCAs.

6.7.1 Foundry design

To synthesise more complex photonic structures than those covered in this section, I designed a PIC using the photonic design automation framework Luceda IPKISS and the PDK for the imec iSiPP50G silicon photonic platform, shown in Fig. 6.15.

In the design, we see an FPCA with four unit cells, with metal pads for electrical contacting at the top and grating couplers for optical input and output at the bottom. To ensure reciprocity, we did not use integrated photodetectors to read the signals in this design. Each unit cell contains balanced 2×2 MZI couplers with DCs in a square configuration. To set the state of the PCM, we use 20 µm long microheaters based on low-loss p–i–n junctions. A list of possible circuits with this four-unit-cell mesh design is presented in Appendix A.2.2. At the time of the completion of this thesis, the design had not yet been fabricated.

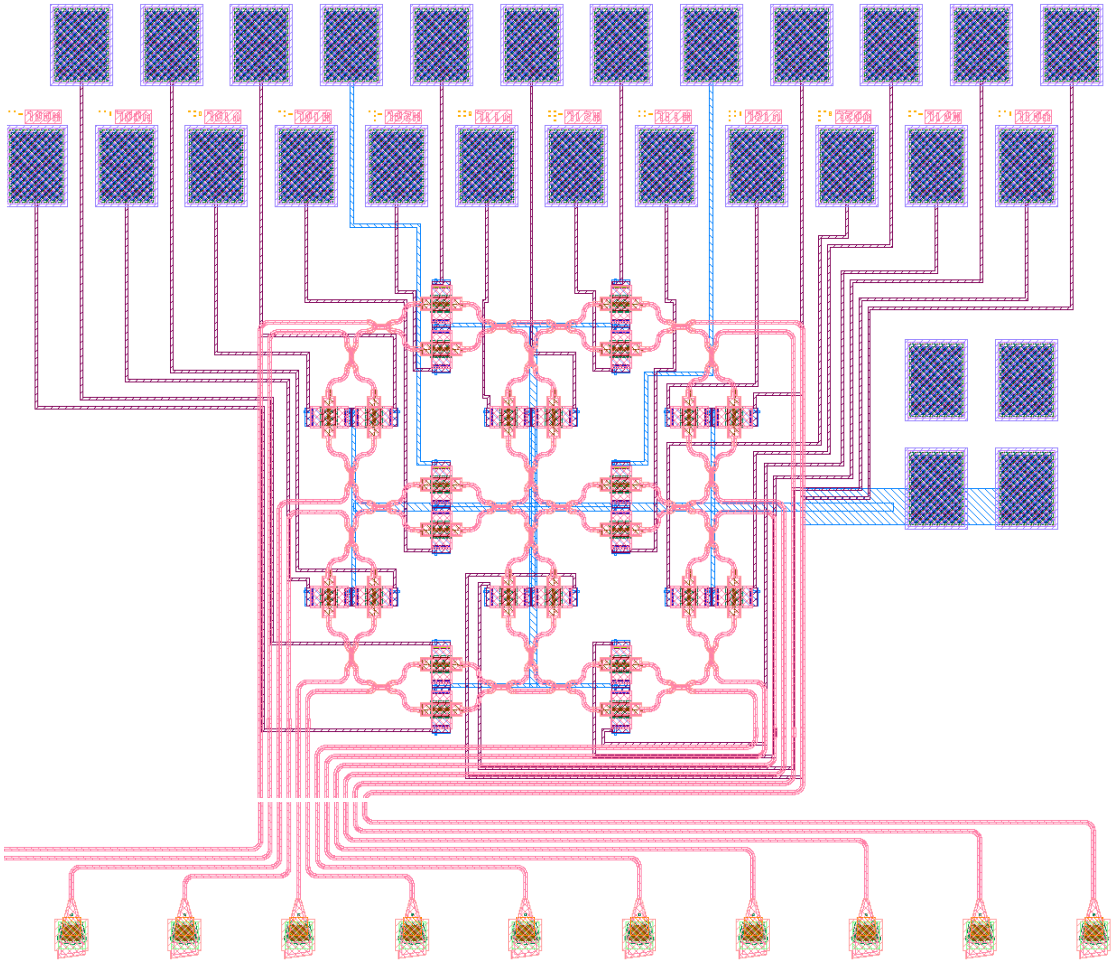

Figure 6.15: GDS design of a scaled nv-FPCA with four unit cells, for imec's foundry.

6.7.2 Future design

The miniaturisation of the MZI couplers in this work is restricted by the size of the electrical pad inside the tunable coupler, which we contact directly with

electrical probes. If we could instead use electrical vias to contact the microheater, as shown in Fig. 6.15, the optical path length of the PCM-based tunable coupler could be considerably shorter. In the following calculations, we assume that the straight part of the coupling section of the DC is $L_{\mathrm{DC}} = 8.9\,\mu\mathrm{m}$ to achieve 50:50 coupling and the bend radius and angle of the DC are $r = 10\,\mu\mathrm{m}$ and $\theta_{\mathrm{MZI}} = \pi/5\,\mathrm{rad} = 36°$, respectively, giving

$$L_{\mathrm{tot}} = 2L_{\mathrm{DC}} + 4r\theta_{\mathrm{MZI}} + 2r\theta_{\mathrm{bend}} + L_{\mathrm{PCM}} = 42.93\,\mu\mathrm{m} + \frac{20\pi}{n}\mu\mathrm{m} + L_{\mathrm{PCM}}, \qquad (6.7)$$

where $\theta_{\mathrm{bend}} = \pi/n$ is the bend angle into and out of the MZI coupler and n is the number of edges per unit cell ($n = 3$ for triangular lattice, $n = 4$ for square lattice, and $n = 6$ for hexagonal lattice).

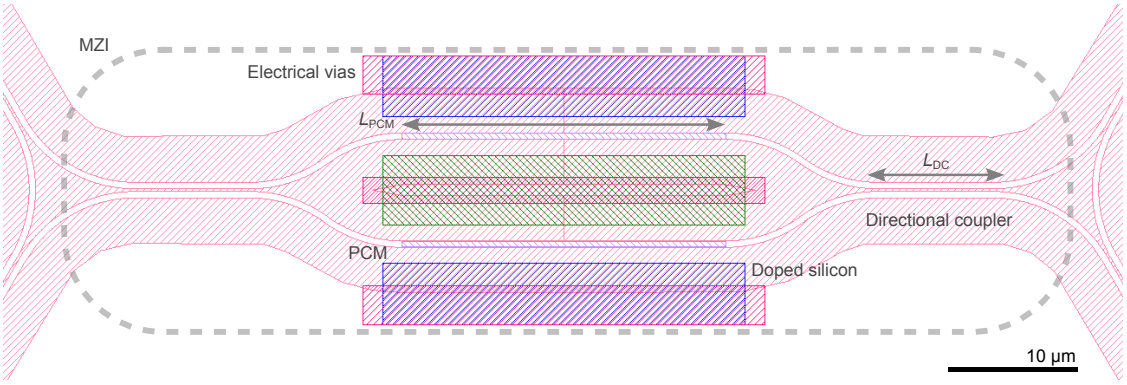

Figure 6.16: Future FPCA tunable coupler design with $L_{\mathrm{DC}} = 8.9\,\mu\mathrm{m}$, $r = 10\,\mu\mathrm{m}$, $\theta_{\mathrm{MZI}} = \pi/5\,\mathrm{rad}$, here shown for a hexagonal lattice arrangement and 25 µm long phase shifters, giving a total path length of 78.5 µm. The dark green and blue colours represent n- and p-doped silicon, respectively.

In this work, we used DCs for splitting and coupling of light in the MZI, which provide small footprints. One problem with DCs is that they feature larger dispersive profiles that lead to wavelength-dependent coupling, thereby giving varying splitting ratios for different wavelengths and limiting the bandwidth of interest [168]. Broadband alternatives include MMIs (at the expense of footprint and loss) [274, 275] and bent DCs [277]. In future devices, we should consider these trade-offs in greater detail.

For future systems, we can employ flattened meshes to further reduce the footprint, in which the mesh topology is collapsed to rows of horizontal photonic

switches. In such systems, there is no open space in the middle of each unit cell, which can significantly reduce the total area. Because the heating from each phase shifter is local (Fig. 6.3a), the risk of crosstalk between switches is limited (except for during switching, when the neighbouring waveguides could heat up). For a massive system with the switches shown in Fig. 6.16, the limiting factor could potentially be the routing of metal wires to the heaters. For most systems, however, electrical routing is not an issue, and I estimate that a unit cell could fit within an area on the order of $200\times50\,\mu m^2$ for a square mesh and $300\times100\,\mu m^2$ for a hexagonal mesh.

Length of the phase shifter

When using MZIs for combined coupling and phase control in an FPCA, the phase shifters need to allow for both full coupling between the bar and cross states and a global phase shift. If we assume that the MZI is at its quadrature point (i.e., we only need $\pi/2$ per arm to control the coupling) and that the obtainable phase shift of the MZI is π extra per phase shifter, the length of the straight section of the MZI must be $L_{3\pi/2}$. For 30 nm thick Sb_2Se_3, $L_{3\pi/2} = 25\,\mu m$ and the total MZI path length is 78.5 µm for a hexagonal lattice arrangement (as shown in Fig. 6.16). If we instead increase the PCM length to $L_{5\pi/2} = 41.8\,\mu m$ and use a square lattice arrangement, the total path length per MZI is 100.4 µm.

The inherent phase shift needed to set the MZI to the quadrature point can be obtained by adding a passive phase shifter to one of the arms, for example, an additional patch of PCM or a wider waveguide section [135, 286–289]. We find from Eq. (3.5) that the length needed for a $\pi/2$ phase shift is given by $L_{\pi/2} = 1.55\,\mu m/(4\Delta n_{\text{eff}})$. For 30 nm amorphous Sb_2Se_3, the change in refractive index relative to the bare Si waveguide is given by $\Delta n_{\text{eff}} = n_{\text{eff, amo}} - n_{\text{eff, Si}}$, which gives $L_{\pi/2} = 2.9\,\mu m$, as found experimentally in Section 4.4.1. Alternatively, for a different waveguide section, e.g., $w = 450$ nm instead of 500 nm, the change is given by $\Delta n_{\text{eff}} = n_{\text{eff}}(w = 450) - n_{\text{eff}}(w = 500) = 2.446 - 2.498 = -0.052$ from MODE simulations, which gives $L_{\pi/2} = 7.5\,\mu m$. Note that the length would need to be longer to accommodate adiabatic tapers to avoid loss. Importantly, by adding this

passive phase shifter to the bends of the MZI rather than the straight section, we can practically mitigate any increase in footprint caused by this design change.

Although full coupling control (between the bar and cross state) *and* at least 2π phase control is needed for universal optical 2×2 unitary gates [18, 166], for many real systems, there is no need for 2π phase control. Examples include interconnects, delay lines, beam splitters, synthesised ring resonators and MZIs, and unitary matrices. For these applications, limiting the phase control to the range 0 to π should be sufficient, because neighbouring devices can collectively provide a 2π phase shift in synthesised structures. A similar approach is used in the work by Kim *et al.* [18], where the tunable coupler and phase shifter are separated; there is only one 2π phase shifter per unit cell of the triangular lattice. However, for some of the most complex synthesised systems, for instance, multiple coupled rings used for the exploration of topological phenomena [290, 291], a 2π phase shift for each MZI is necessary [291]. In these cases, it is evident that the excess loss from the PCMs and heaters needs to be small.

6.8 Chapter summary

This chapter demonstrates a small-footprint, non-volatile field-programmable coupler array (nv-FPCA) that uses the low-loss phase-change material (PCM) antimony triselenide (Sb_2Se_3) as the active material. The building block of the FPCA is a 2×2 MZI tunable coupler with Sb_2Se_3 phase shifters in both arms. To control the state of the compact, non-volatile phase shifter, we utilise electro-thermal heating from integrated doped-silicon microheaters.

In this chapter, I have described the device design and electro-thermal simulations of the phase shifters. Furthermore, I have shown optical measurements of the MZI building block and the nv-FPCA. We demonstrate high extinction (>20 dB) over a wavelength range of 15 nm for a single coupler, with medium loss (estimated to be from 1.43 dB to 1.83 dB, depending on the state of the PCM). The switching energy required for full amorphisation and crystallisation of the 60 nm thick and 7.6 µm long Sb_2Se_3 patch is, respectively, 1.2 µJ and 1.2 mJ (estimated), but we

believe these numbers can be significantly reduced with optimised pulse sequences and thermal engineering. Furthermore, we present multi-level switching with over 50 distinct levels.

The nv-FPCA unit cell is used to realise three important optical components: routers, ring resonators, and delay lines. The synthesised ring resonator has a Q factor of 8.5×10^3 and a free spectral range (FSR) of 0.33 nm (40 GHz), which can be improved with lower loss and smaller footprint of the building blocks in future devices. Moreover, simulations show that the synthesised ring resonator can work as a temporal differentiator for edge detection.

The discussion contrasts PCM-based nv-FPCAs with TO- and MEMS-based platforms: PCM offers the smallest active lengths and zero static power (through its non-volatility) but presently suffers from high switching energies, long reconfiguration times, large and coupling-dependent loss, and, in particular, limited demonstrated cyclability—in this work, we were only able to crystallise 50 % of the PCM volume after full amorphisation. If the current limitations can be overcome, low-loss PCMs are expected to provide a complementary alternative technology to FPCAs based on TO and MEMS couplers for infrequent-switching applications. Finally, I present designs for scaled meshes and projected footprints in future systems.

In short, the work validates Sb_2Se_3-based, non-volatile tunable couplers as a promising route to compact, low-standby-power programmable photonic meshes, while identifying the key technical challenges (loss, switching energy, and cyclability) that future device and process improvements must address.

The material in this chapter is based on a manuscript published in *Science Advances*: **H. H. Toftevaag**, B. Dong, N. Farmakidis, A. Ortega-Gomez, and H. Bhaskaran, 'Non-volatile photonic field-programmable coupler array'. In: *Science Advances* 12.18 (May 2026), eaec7329.

Table 6.2: Performance comparison of demonstrated FPCA technologies.

Technology	Static power (P_π)*	Switching energy	Loss (dB)	Unit length	Active length	Non-volatility	Ref.
TO MZI (SiN)	$P_{\text{avg}} = 0.25$ W	N/A	N/A ($\lesssim$1)	3.45 mm	2.1 mm	No	[166]
TO MZI (Si)	110 ± 15 mW	N/A	0.6 ± 0.1	$\geq$0.98 mm	$\sim$0.5 mm	No	[169]
TO MZI (Si)	1.3 mW	N/A	0.48	0.81 mm	767 µm	No	[170]
TO DC (SiN)	0.35 W	N/A	**0.2**	$\sim$2.1 mm	1.28 mm	No	[41]
TO MZI (SiN)	0.290 W	N/A	2	1.3 mm	$\sim$0.7 mm	No	[168]
MEMS (Si)	$P_{\text{max}} < 17$ fW	**<44 pJ**	0.26–0.52†	**0.38 mm**	$150 + 100$ µm	No	[18]
PCM MZI (Si)	**0**	N/A‡	N/A‡	$\sim$1 mm	77.8 µm	**Yes**	[278]
PCM MZI (Si)	**0**	1.2 µJ / $\sim$1.2 mJ	1.4–1.8	0.48 mm§	**7.6 µm**	**Yes**	This work
PCM (future)	**0**	<108 nJ / 0.53 µJ¶	$\sim$0.5–1.0	**<0.1 mm**	<30 µm#	**Yes**	This work

* Here, the power is given as the electrical power required for a π phase shift (i.e., to tune the couplers from 0 % to 100 % coupling), P_π, unless otherwise stated: P_{avg} is the average power consumption per heater, P_{max} is the maximum standby power consumption.

† Loss range, assuming one phase shifter per unit cell.

‡ Not enough information is presented to estimate the values.

§ Affected by the size of the metal pads.

¶ Projected values based on those found in Yang *et al.* [131].

\# The length of the phase-change element and heater could be made shorter, but at the expense of smaller achievable phase tuning (done in this work).

There are many paths to the top of the mountain,
but the view is always the same.

— Chinese proverb

7

Non-uniformly quantised weights for analogue neural networks

Contents

7.1 Introduction 166
7.1.1 Uniform and non-uniform weight quantisation 167
7.2 Device design 167
7.2.1 Determining the device design through simulations . . . 168
7.2.2 Control of the device 175
7.3 Measurements of ring-assisted MZI weights 178
7.3.1 General behaviour 178
7.3.2 Wavelength dependence 180
7.3.3 Varying the shape of the curve 182
7.3.4 Neural network weight quantisation and accuracy 185
7.3.5 Discussion 190
7.4 Future work 191
7.4.1 Electrical tuning 191
7.4.2 System integration 192
7.4.3 Alternative photonic structures 193
7.4.4 Energy calculations 194
7.5 Chapter summary 196

This chapter discusses the operation and performance of the first non-uniformly quantised photonic neural network weight. The objective of this chapter is to demonstrate a concept that fundamentally changes how we approach hardware for analogue neural networks. Here, my focus is mainly on the weights, but I will also

briefly discuss the network design itself when used in photonic crossbar arrays.

7.1 Introduction

Neural networks, along with the LLMs built on them, are rapidly revolutionising the world. This is because they can learn intricate patterns from enormous datasets and produce human-like results by grasping context, facilitating advances in various fields, including speech processing and autonomous systems [7, 8, 292]. The downside is an enormous energy consumption for training and using the largest models [293], and even for smaller models, the storage and power requirements can be detrimental when deployed in resource-constrained edge devices [294, 295]. Solutions to these problems include, for example, software-based algorithmic advances such as quantisation [296, 297] and hardware-based implementations such as analogue light-based machine learning accelerators instead of digital electronic ones [4, 7, 8, 188].

Weight quantisation, that is, reducing the bit width of the neural network's weights (see Section 3.7.1), can lead to substantial energy savings in general, and especially in analogue systems, due to the way in which we set the weights. Although lower weight resolution generally means lower energy consumption, this reduction in bit-width typically also leads to lower neural network accuracy, thereby providing a trade-off between energy and accuracy. That being said, there are examples of neural networks employing weights with only three obtainable weight values (-1, 0, and 1; called *ternary quantisation*) that *outperform* full-precision models [211]. Although impressive, this approach is not a one-size-fits-all solution, and in reality, the problem at hand will determine the usefulness of the quantisation. With this in mind, our argument is not around quantisation *per se*, but around the non-uniform distribution of bits across the range. We know empirically that neural network weights typically follow a bell shape centred around zero (see Fig. 3.4c) [213, 214], and non-uniform quantisation allows us to leverage that. We endeavoured to devise a photonic component that inherently performs non-uniform quantisation—not in software, but in the weight itself; we call this concept non-uniform-in-weight quantisation.

7.1.1 Uniform and non-uniform weight quantisation

Amplitude-modulating elements (e.g., TO MRR modulators or GST elements) in photonic crossbar arrays have different transmission levels depending on what the electrical input (e.g., Joule heating in TO heaters or electro-thermal heat pulses to set the state of PCM elements) is. This relation, the electrical-to-optical transfer function, determines the voltage values we need to input from the DAC to get a certain transmission, or weight value in the case of the photonic crossbar array. As an example, if we apply 0 V to get T_{min}, 0.5 V to get $\frac{T_{\text{max}}-T_{\text{min}}}{2} + T_{\text{min}}$, and 1 V to get T_{max}, we would say the electrical-to-optical transfer function is linear. In this scenario, if the electrical output from the DAC is limited to only a few voltage values and these values are uniformly spaced on the number line, then the obtainable transmission values of the amplitude-modulating weight would also be uniformly spaced. This is an example of uniform quantisation.

Non-uniform quantisation of photonic neural network weights can be achieved using two different approaches: (1) we design a DAC that outputs non-uniformly spaced voltage values and use a photonic device with a linear electrical-to-optical transfer function, or (2) we use a DAC with uniformly spaced values and a device with a non-linear electrical-to-optical transfer function. Because non-uniform output from DACs is not trivial and photonic devices often show non-linear electrical-to-optical transfer functions, I believe the second approach is the most feasible. Examples of non-linear transfer functions include those of MRMs and MZIs, which follow Lorentzian and sine curves, respectively. However, to get a suitable non-uniform quantisation in photonic neural network weights, we need to design the weights in a very particular manner—practically, the transmission response of the weight needs to reflect the bell shape of the global weight distribution shown in Fig. 3.4b.

7.2 Device design

Our initial hypothesis was that it should be possible to make a simple photonic device with a non-linear electrical-to-optical transfer function (i.e., optical transmission vs.

electrical input) that follows the quantile function of a bell curve, as seen in Fig. 3.2. From Section 3.4, we know that the phase of the light exiting the through port of an all-pass ring resonator will accumulate an effective phase delay resembling the CDF of a bell curve, with phase delay from 0 to 2π for detuning from $-\pi$ to π. More importantly, the phase shift accumulates as we go away from one resonance and towards another. When the detuning goes from 0 to 2π instead, the phase delay still changes only by 2π (from π to 3π), but the shape is now no longer resembling that of the CDF of a bell curve, but rather that of a quantile function! This accumulated phase is not valuable on its own, but, as shown in Section 3.5, the output of an MZI is determined by the phase difference between its arms. In this work, we designed a device by marrying the phase-dependent transmission response of an MZI with the phase response of an MRR—in other words, we designed a RAMZI.

7.2.1 Determining the device design through simulations

Our initial hypothesis—that the phase delay of a ring resonator and the phase-dependent transmission of an MZI together could give us the desired transmission curve—does not automatically mean that *any* RAMZI would suffice for this task. For example, Darmawan *et al.* presented various RAMZIs with single and multiple rings [206], but none of these designs are optimised for our task. Before designing and fabricating the photonic neural network weights in this work, we ran simulations to confirm the device behaviour. When simulating simple photonic devices, many researchers resort to the finite-difference time-domain simulation method (FDTD); however, in this chapter, I will present simulations using INTERCONNECT instead, as it allows repeated simulations with small design tweaks without wasting computational time and power.

In Fig. 7.1, I present different RAMZI designs and their simulated transmission spectra. Many different designs were considered, for instance, using different splitter structures (Y-branches vs. DCs) at the input and output of the MZI, adding a static $\pi/2$ phase shift in one arm, coupling light out of the add port of the ring

instead of the through port, and adding multiple rings and/or coupled rings. Some of these designs are shown in Fig. A.2.

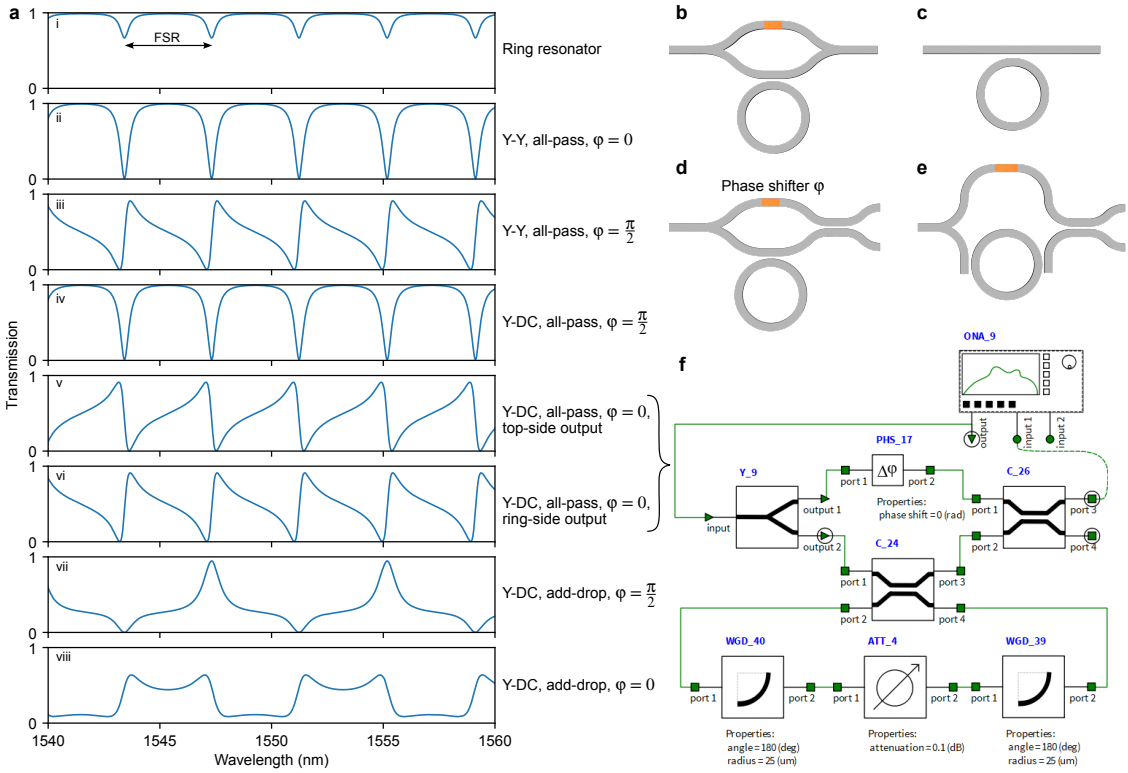

Figure 7.1: Optical simulations of various RAMZI devices in INTERCONNECT. **a**, Normalised transmission spectra from four different devices: (i) an all-pass ring resonator, shown in **c**; (ii,iii) the RAMZI shown in **b**, with $\varphi = 0$ (ii) and $\varphi = \pi/2$ (iii); (iv–vi) the RAMZI shown in **d**, with $\varphi = \pi/2$ (iv) and $\varphi = 0$ (v: top-arm output, vi: ring-arm output); (vii,viii) the RAMZI shown in **e**, with $\varphi = \pi/2$ (vii) and $\varphi = 0$ (viii). **b**, Schematic of a RAMZI with a Y-branch at both the input and output, an all-pass ring in one arm, and a phase shifter in the other. **c**, Schematic of an all-pass MRR. **d**, Schematic of a RAMZI with a Y-branch at the input, a DC at the output, an all-pass ring in one arm, and a phase shifter in the other. **e**, Schematic of the RAMZI in **d**, but with an add-drop ring instead of an all-pass ring. **f**, Simulation setup in INTERCONNECT for the transmission spectra in plot (v, vi) in **a**. The circuit is equivalent to the RAMZI in **d**. FSR: free spectral range; Y: Y-branch; DC: directional coupler.

In the simulations, we set the effective refractive index and the group index of the waveguides to 2.47 and 3.9, respectively. The ring has a radius of $r = 25\,\mu\text{m}$ and total loss of 0.3 dB, and the MZI arms are 10 µm long with 10^{-3} dB loss each. In Fig. 7.1, the coupling coefficients of all DCs (including the one coupling into the ring) are set to 0.5 for the entire spectrum. For the RAMZI in Fig. 7.1d,f, I will refer to the output on the side with the ring as the ring-arm output and

the other output as the top-arm output.

There are a few interesting things to note from the transmission spectra in Fig. 7.1. First, the RAMZI with a double Y-branch has identical output to the RAMZI with a Y-branch and a DC and a $\pi/2$ phase shift. In other words, using a DC instead of a Y-branch adds a $\pm\pi/2$ phase shift, in accordance with theory (Section 3.5.1). Secondly, the two outputs of the RAMZI in Fig. 7.1d (plots v and vi) are mirror images of each other. Thirdly, the add-drop ring can give the desired curve, but with two considerable drawbacks: the curve only repeats over two FSRs (as opposed to once per FSR for the all-pass ring) and the plateau of the curve is at 0.25, not 0.5, which is bad for the weight's behaviour.

The simplest design for our purpose consists of a balanced MZI with an all-pass ring resonator in one arm, and with either a Y-branch and a DC at the input and output, respectively, or with two Y-branches and a $\pi/2$ phase shift in one arm.[1] For ease of fabrication, we chose the former option.

Effect of device parameters

We proceed to uncover the behaviour of the RAMZI presented in Fig. 7.1d. The most important behaviour for the working principle of the neural network weight occurs when we apply a phase shift to the ring resonator. The phase shift leads to a redshift of the spectrum, as shown in Fig. 7.2a. By probing at a single wavelength, we get a non-linear phase-to-transmission curve with the same shape as the spectrum. Because the transmission of the RAMZI originates from the Lorentzian behaviour of the ring resonator, the curve takes the tangent-like shape described by Eq. (3.48). In Section 7.2.2, I present this phase-shifting mechanism in more detail and explain how it can be used to encode neural network weights. Interestingly, from Fig. 7.2b, we see that the same spectrum shift does not occur when we apply a phase shift to the MZI arm—instead, the desired tangent-like shape is no longer present for large phase shifts.

[1]For the desired behaviour, we need a $\pi/2$ phase difference between the arms, which can be introduced either by a phase shifter or through the DC.

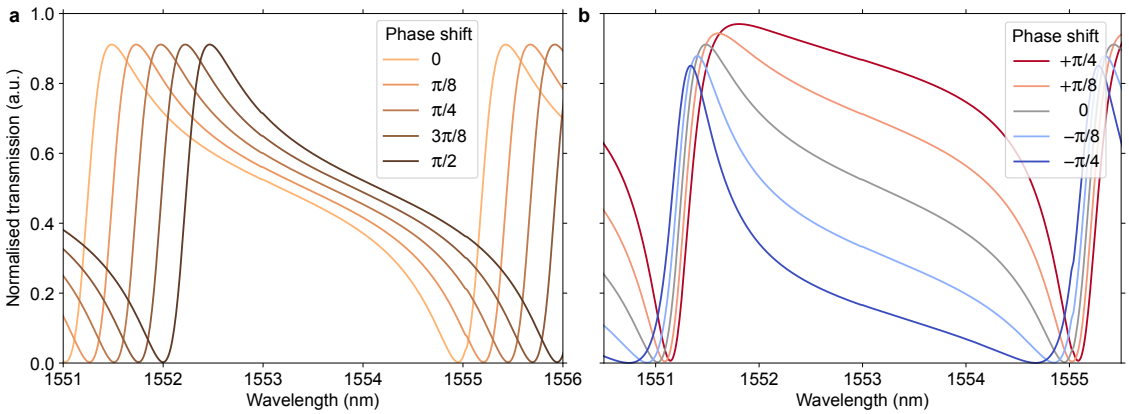

Figure 7.2: Impact of phase shifts on RAMZI behaviour. **a**, Normalised transmission spectra with different phase shifts in the ring. The plot redshifts with increasing phase, but the shape stays constant. **b**, Normalised transmission spectra with different phase shifts in the MZI arm. The plots change shape, but do not shift along the x-axis—instead, the troughs of the curve move back and forth and go back to their original position for every $\varphi = \varphi_{\text{initial}} + 2\pi m$, $m \in \mathbb{Z}$. In the simulations, the parameters are the same as in the simulation in Fig. 7.1f.

Next, in Fig. 7.3, we show how the coupling into the ring gives different tangent-like curve shapes. The first two rows of the figure show, respectively, the transmission and phase spectra for all-pass ring resonators with different cross-coupling coefficients. The third row shows the transmission spectrum from the ring-arm output of RAMZIs with the rings from the first two rows. Notice the change in the phase of the ring when going from barely overcoupled to critical coupling—instead of seeing an abrupt $+\pi$ phase shift [31], INTERCONNECT adds an abrupt $-\pi$ phase shift instead [298]. For the undercoupled ring, there is no abrupt jump, and the phase curve is continuous (compare with Fig. 1 in Ref. 298). Although the phase of the ring resonator looks distinctly different in the two plots in the middle of Fig. 7.3, both the ring resonator and the RAMZI transmission spectra show that there is virtually no difference in the device behaviour.

It is clear from the bottom row in Fig. 7.3 that changing the coupling into the ring changes the transmission curve of the RAMZI. As the coupling coefficient increases, the plateau of the curve becomes steeper, and overall transmission increases. Note that in the extreme case of $k = 1$, all the light couples into the ring and out again without interfering with itself, effectively setting the arm length

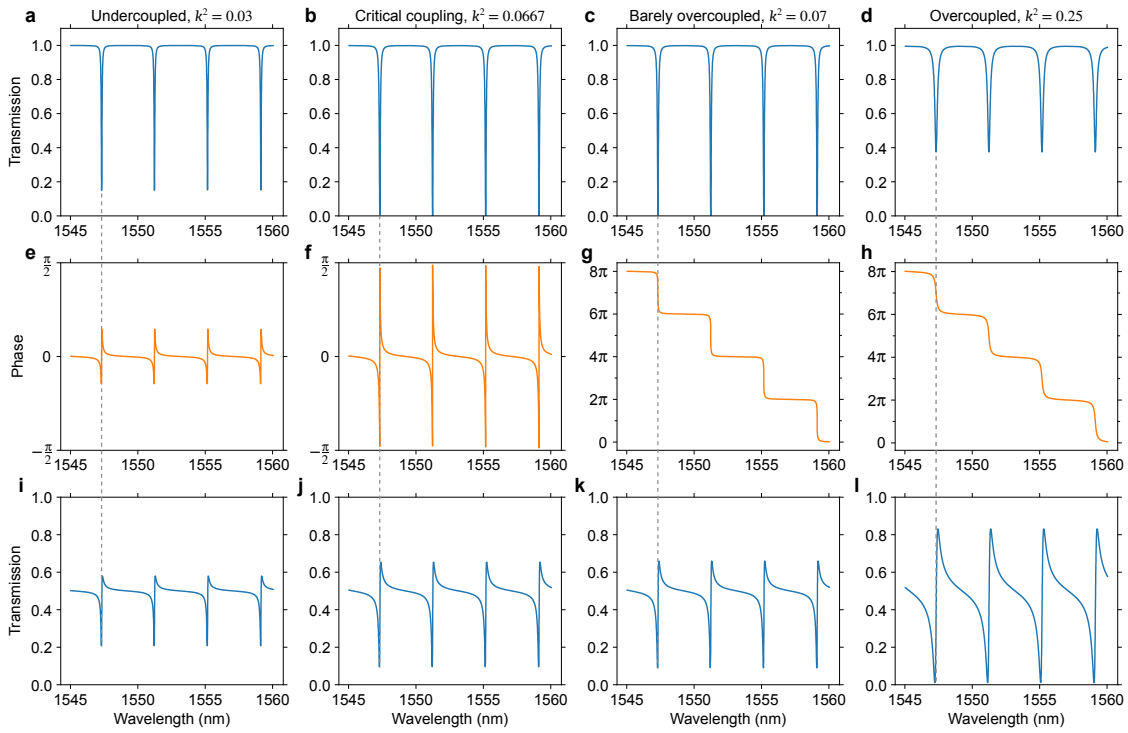

Figure 7.3: Optical simulations of the impact of the ring coupling on RAMZI behaviour in INTERCONNECT. **a**–**d**, Transmission spectra of an all-pass ring resonator with different cross-coupling coefficients. **e**–**h**, Phase spectra of an all-pass ring resonator with different cross-coupling coefficients. **i**–**l**, Transmission spectra of the RAMZI in Fig. 7.1d (ring-arm output) with different cross-coupling coefficients. The cross-coupling coefficients in the columns are, from left to right, $k^2 = 0.03$ (undercoupled), $k^2 = 0.0667$ (critical coupling), $k^2 = 0.07$ (barely overcoupled), and $k^2 = 0.25$ (overcoupled).

difference to $\Delta L = 2\pi r$ and turning the RAMZI into an unbalanced MZI (not shown in Fig. 7.3; see Fig. 7.4).

In addition to the effect of a phase shift and the ring coupling, I investigated the behaviour of an unbalanced RAMZI, i.e., when the arm length difference of the RAMZI $\Delta L \neq 0$. In these simulations, $\Delta L = 30\,\mu\text{m}$, still with $10^{-3}\,\text{dB}$ loss in each arm. The ring radius was $r = 75\,\mu\text{m}$ and the total loss of the ring was 0.3 dB. In Fig. 7.4, we clearly see the effect of the path length difference as a sinusoidal envelope on top of the balanced RAMZI spectrum. For the coupling coefficients in Fig. 7.4e–g, there are some interesting tangent-like resonances resembling those given by Eq. (3.48) (e.g., around 1545 nm), while none of the resonances for small coupling coefficients are suited for our application. Additionally, I have added the spectrum for $k^2 = 1$, which is when there is no resonance in the ring, and it can

be treated as an additional waveguide section, giving $\Delta L = 30\,\mu\text{m} - 2\pi r$. The other output (top-arm output) behaves in a similar way, as we would expect: the resonances are mirror images of those of the ring-arm output, and on top of that, the sinusoidal envelope is shifted (the envelope nodes become anti-nodes and vice versa). It is clear from Fig. 7.4 that the balanced option is a better fit for our purpose.

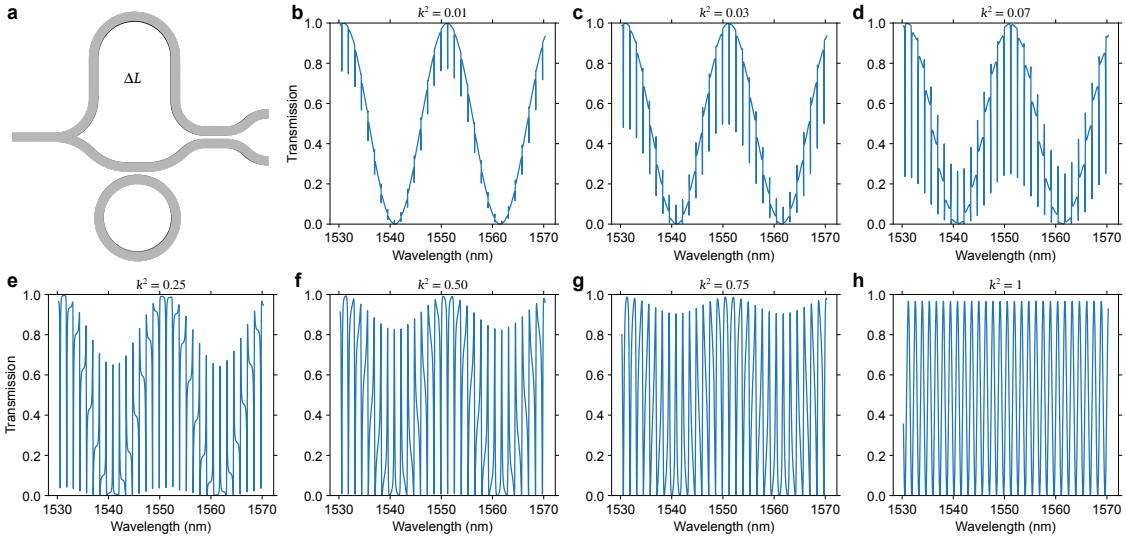

Figure 7.4: Optical simulations of an unbalanced RAMZI in INTERCONNECT. **a**, Schematic of an unbalanced RAMZI with path length difference ΔL. **b**–**h**, Transmission spectra of an unbalanced RAMZI (ring-arm output) with cross-coupling coefficients $k^2 = 0.01$ (b), 0.03 (c), 0.07 (d), 0.25 (e), 0.50 (f), 0.75 (g), and 1 (h).

I also investigated the effect of changing the ring circumference. In simulations with constant ring loss, the only resulting change is in the FSR, in agreement with analytical predictions. However, in a real device, changing the circumference can lead to greater loss, reducing the maximum achievable transmission. I have not shown these simulations here.

Finally, I examined the effect of introducing imperfect DCs for coupling into the ring and combining the light at the output of the MZI. In Fig. 7.5, I show the effect of a realistic wavelength dependence on the coupling coefficient of the DCs. The three plots show the transmission from a RAMZI with $k^2 = 0.5$ into the ring and 50:50 splitting at the output. First, both DCs provide 50:50 splitting at all wavelengths, then we introduce wavelength-dependent splitting in the output coupler, and finally, wavelength-dependent coupling in the ring coupler. We see

that the change in output splitting slightly tilts the spectrum: below 1550 nm, we observe lower transmission at the peaks than in the perfect case, while above 1550 nm, we observe higher transmission in the troughs (the transmission does not go to zero). Next, we observe that the slight variation in the coupling into the ring, k^2, alters the shape of the transmission curve—notice how the plateau of the purple curve is slightly flatter than the other curves below 1550 nm and steeper above 1550 nm, because the coupling coefficient is different. This behaviour is expected from the analytical expressions in Section 3.5.2 and is effectively the same as the one we see in Fig. 7.3i–l.

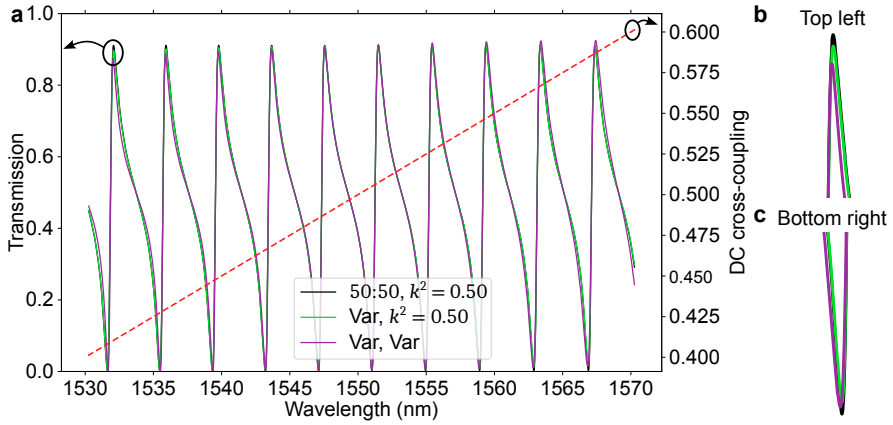

Figure 7.5: **a**, Optical simulations in INTERCONNECT of a RAMZI with imperfect DCs. All RAMZIs have DCs for 50:50 splitting at the output of the MZI (labelled 50:50) and 50 % cross-coupling into the ring ($k^2 = 0.50$), but where indicated by 'Var', the coupling of the DC has a wavelength dependence shown by the plot on the right y-axis. **b**,**c** Zoom-in on the top left peak and bottom right trough.

A comment on the obtainable transmission levels of the RAMZI is warranted. From the bottom row of Fig. 7.3, it is clear that a trade-off exists between the maximum transmission and the shape of the curve—flatter plateaus are accompanied by lower maximum transmission. In large-scale photonic neural network architectures, we want to minimise losses across all components; consequently, a limited maximum transmission is detrimental to the usefulness of the RAMZI weight in such systems. However, as mentioned in Section 3.5.2, using rings with very low loss ($a \approx 0.999$) enables us to achieve most curve shapes without sacrificing transmission. In other words, this trade-off is a practical limitation in our current, lossy rings, but not a fundamental drawback of our design.

7.2.2 Control of the device

The device presented in this chapter has a transmission response as described in Section 3.5.2 and the previous paragraphs, but for a photonic neural network, we want to actively change the device behaviour. We introduce a phase shifter in the ring to set the transmission of the RAMZI encoding the neural network weight value. When the phase is swept across one FSR of the ring, we can reach all transmission values between T_{min} and T_{max}—this curve is the electrical-to-optical transfer function. Furthermore, it is interesting not only to set the weight value, but also to change the transfer function of the RAMZI to better replicate the different weight distributions of the layers of a neural network [214]. The shape of the transmission curve is altered by modifying the coupling into the ring, which we can achieve in one of two ways: by changing the physical dimensions of the coupling region (e.g., the coupling length or gap of the racetrack resonator) or by changing the effective refractive index of the two waveguides in the coupling region (e.g., by heating).

In this work, we control the transmission level and shape of the curve using phase shifters. Specifically, we employ TO phase shifters made of nichrome heaters (Section 4.3.2), where Joule heating in the nichrome heaters causes a phase shift $\Delta\phi \propto \Delta T$ in the silicon waveguide, where ΔT is the change in temperature. The heat is given by $\Delta T \propto P_{\text{heater}} = V^2/R_{\text{heater}}$, where V is the applied voltage and R_{heater} is the resistance of the heater. Practically, we contact the nichrome heaters with Au pads and send current through electrical probes to the pads.

Controlling the weight value

The neural network weights in this work are set using phase shifters in the ring resonator. The phase shifter in the ring shifts the ring resonance and thus the transmission spectrum of the RAMZI, as seen in Fig. 7.2a. If we send in light with a wavelength λ close to a trough in the spectrum and apply a phase shift to the ring, the spectrum will redshift and the transmission will increase. By shifting the spectrum linearly, transmission will continue to increase and will essentially follow the tangent-shaped curve until it reaches its maximum value. In

the neural network, we will define the trough as -1, the midpoint of the plateau as 0, and the maximum value as 1.

In measurements, to find these values, we take the following approach. First, we sweep the phase from the minimum to the maximum transmission over one FSR, thereby tracing out the characteristic tangent-like transmission response. Next, we find the midpoint on the tangent-shaped curve, i.e., where the second derivative goes to zero, and label this T_0. Because the curve will not be perfectly symmetric around this midpoint, we introduce $T_{\text{ex}} = \min(T_{\text{max}} - T_0, T_0 - T_{\text{min}})$, where T_{max} (T_{min}) is the maximum (minimum) transmission value within that FSR. T_{ex} determines the distance between the midpoint and the obtainable maximum and minimum values of the neural network weight, which are given as $T'_{\text{max}} = T_0 + T_{\text{ex}}$, $T'_{\text{min}} = T_0 - T_{\text{ex}}$. With this approach, we ensure symmetry. In the neural network, we define $T'_{\text{max}} = 1$, $T_0 = 0$, and $T'_{\text{min}} = -1$.

In this chapter, we will emphasise the benefit of being able to send uniformly spaced voltage values from the DAC to the photonic weight that gives non-uniformly spaced transmission values. Crucially, this non-linear electrical-to-optical transfer function should be symmetric around 0. Unfortunately, with TO phase shifters, $\Delta\phi \propto V^2$, which means that the RAMZI curve becomes asymmetric around 0 with linear voltage input. Nevertheless, we chose TO phase shifters in this work because of their ease of integration. Importantly, the concept presented in this chapter is platform-agnostic, and we can therefore substitute the phase shifters with any phase shifter with a linear voltage-to-phase curve, for instance, those based on the Pockels effect. Finally, it is worth pointing out that although $\Delta\phi \propto V^2$ in this work, we can simply apply the voltage as $\sqrt{V}$ to get a linear 'voltage'-to-phase curve and see the exact device performance we expect. This simple trick demonstrates that using better-suited phase shifters with actual linear voltage-to-phase curves will be successful (I will continue this discussion in greater detail in Section 7.4.1).

One potential downside of the current design is that we always go from -1 to 1 when setting the weight value (or vice versa if the curve is mirrored, as is the case for the top-arm output), which means that most weight values around 0 will

require power (unless the weight is non-volatile). However, this limitation often applies to amplitude modulators in photonic neural networks, whether EAMs or ring resonators. Furthermore, with phase shifters based on the Pockels effect (see Section 7.4.1), both positive and negative refractive index changes can be achieved by controlling the direction of the electric field [299]. Therefore, we do not see this characteristic as a limitation of our device.

Controlling the curve shape

For non-uniform quantisation of the weight values in neural networks, it is beneficial to have tunability in the tangent-shaped electrical-to-optical transfer function, i.e., to be able to change the shape determined by γ' to better resemble the overall weight distribution of individual layers. We can achieve this tuning by modifying the coupling into the ring, k, which is parametrised as $r = \sqrt{1-k^2}$ in the equations for the ring transmission (t, Eq. (3.17)) and phase (φ, Eq. (3.22)). In practice, as mentioned in Section 3.3, we can modify the coupling by varying the coupler's physical length, adjusting the gap between the MZI arm and the ring, or altering the phase mismatch. In principle, the two former techniques can be made tunable by employing MEMS couplers to vary the length and/or gap of the coupling region; however, such implementations lie beyond the scope of this work. As a consequence, when we show different curve shapes with varying coupling length later in this chapter, the transmission curves are from different devices whose parameters are fixed at the time of fabrication; hence, they are not tunable in this work.

The third technique—altering the phase mismatch—can be achieved by changing the effective refractive index of one of the arms of the coupling region. In this work, we employ a phase shifter in the coupling region to modify the coupling condition, effectively realising a 2×2 DC switch.[2] A drawback of this approach is that applying a phase shift in the coupling region modifies not only the coupling condition but also the phase of the ring and, more problematically, the phase difference between the arms of the MZI. Consequently, the interference condition at the output of

[2]Although using, for instance, a 2×2 MZI switch is also possible, it would introduce more losses in the coupling region and lead to a larger footprint.

the MZI is altered, introducing an additional contribution to the overall transfer function (similar to that in Fig. 7.2b). To mitigate this effect, an equal phase shift can be applied to the other MZI arm, ensuring that $\Delta\varphi_{\text{arms}} = 0$ while the coupling condition into the ring is still tuned. We fabricated two closely related designs with a single variation, as shown in Fig. 4.4a,b: in the first design, the heater is placed above the coupling region to the ring, thereby heating only this section; in the second, the heater also extends over the other arm of the MZI. By heating both MZI arms simultaneously, it should therefore be possible to tune the coupling into the ring while preserving the relative phase between the MZI arms. Measurement results from both approaches are presented in Section 7.3.3.

7.3 Measurements of ring-assisted MZI weights

Here, I present optical measurements of the fabricated RAMZI devices, demonstrating their use for non-uniform quantisation in analogue neural networks. I begin with the general behaviour and introduce more computing-relevant results in the latter subsections.

7.3.1 General behaviour

Figure 7.6 shows a fabricated device and its general behaviour. In this device, shown in Fig. 7.6a, the coupling heater covered both arms (design 2). The transmission spectrum (Fig. 7.6b) is as expected, with tangent-shaped curves repeated for each FSR and with a sharp transition in transmission between them. The bell-shaped envelope is caused by the grating couplers. Furthermore, we notice that the grating couplers also introduce some roughness to the transmission spectrum.

In Section 7.2.2, we described how we can shift the transmission spectrum of the RAMZI using a phase shifter in the ring. We set the transmission level of the RAMZI weights in the analogue neural network using this approach. In Fig. 7.6c, I set the laser wavelength to 1578.15 nm and sweep the phase of the ring linearly across a resonance. The resulting curve follows Eq. (3.48) to an extremely high degree, as indicated by the curve fit. Additionally, the wavelength-dependent roughness

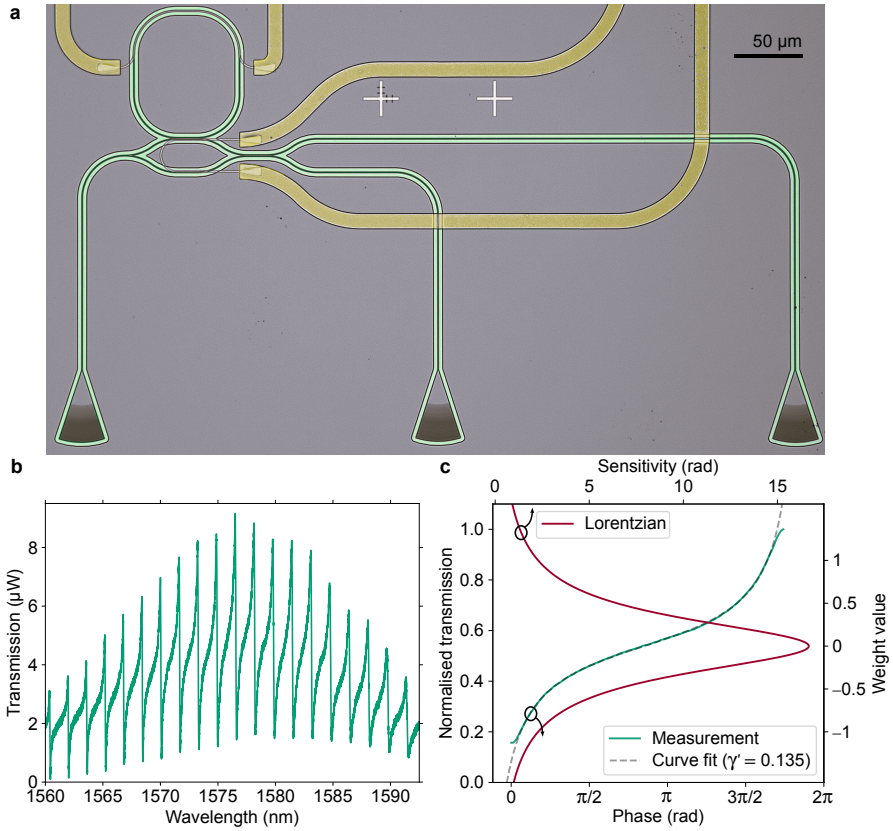

Figure 7.6: RAMZI weight transmission and curve fit. **a**, Optical microscope image of a RAMZI with ring circumference $C_{\text{ring}} = 275\,\mu\text{m}$ and ring coupling length $L_c = 14\,\mu\text{m}$. **b**, Transmission spectrum from the ring-arm output of a RAMZI with $C_{\text{ring}} = 400\,\mu\text{m}$ and $L_c = 42.6\,\mu\text{m}$ (corresponding to $k^2 \approx 0.01$). **c**, Transmission response of one of the resonances in **b**. The wavelength is set to 1578.15 nm (at a peak) and the phase is swept across a resonance. The dashed line is the curve fit with $\gamma' = 0.135$ (Eq. (3.48), for FSR $= 2\pi$), indicating excellent overlap, and the corresponding Lorentzian indicating the sensitivity of the transmission is shown on the top x-axis. The right y-axis shows the neural network weight value. Note that since the phase shift in the ring redshifts the spectrum, the transmission actually goes from T_{max} to T_{min} for a positive phase shift, but is mirrored around the y-axis here to imitate the wavelength sweep.

is no longer seen in the transfer function plot. The plot shown on the second x-axis indicates the sensitivity of the neural network weight (i.e., the derivative of the inverse of the curve), and takes the shape of a Lorentzian curve (Eq. (3.50)). Practically, the flatter the plateau of the quantile function, the sharper the resonance of the Lorentzian, as both are given by γ'. Furthermore, it is worth mentioning that $d = 7.06 = 2.25\pi$ in the curve fit, which is larger than in the ideal case ($d = 2\pi$). In the plot, one FSR of the RAMZI (equivalent to 1.64 nm at $\lambda = 1578$ nm) is 2π and the transmission (from bottom to top) spans 1.51 nm $= 1.85\pi$.

Interestingly, the spectrum in Fig. 7.6b is from the ring-side output, but is the mirror image of what is expected—it looks like the output from the top-side port. This behaviour is true for all devices in this work with ring circumference $C_{\text{ring}} = 400\,\mu\text{m}$ and ring coupling length from $L_{\text{c}} = 32.6$ to $44.6\,\mu\text{m}$, which means that the phase difference between the MZI arms is shifted by a π phase shift. All devices with $225\,\mu\text{m}$ to $275\,\mu\text{m}$ circumference and ring coupling length from $8\,\mu\text{m}$ to $20\,\mu\text{m}$ have the expected shape, which leads me to believe that the mirroring comes from the longer DCs, but I am not able to explain this behaviour.[3] In Fig. A.3, I show more transmission spectra of different devices.

7.3.2 Wavelength dependence

The RAMZI weight has many repeated resonances in the wavelength range presented in the figures above. If the resonances are identical, we can use the device with WDM schemes to achieve higher throughput [4]. Conversely, if there are variations in the curves between resonances, we can utilise these variations to tune the transmission curve simply by changing the wavelength, without changing the physical parameters of the coupling region. Ideally, we should be able to quantitatively determine what the parameters of the curves are from the transmission spectra of the measured devices. However, because of the jaggedness in the transmission spectra and the envelope from the grating couplers, a safer approach is to set the wavelength and sweep the phase shifter across the resonances sequentially.

Figure 7.7 shows the wavelength dependence of the non-uniform neural network weight. For each resonance, we sweep the phase shifter in the ring from top to bottom, and then normalise the data for each wavelength to $T'_{\text{max}} = 1$ and $T'_{\text{min}} = -1$ on the y-axis with corresponding values 0 and 1, respectively, on the x-axis. In Fig. 7.7a, the plot from Fig. 7.6b is presented again, but coloured and annotated with the measured γ' found in the phase sweeps in Fig. 7.7b. From the latter figure, we see a clear trend in the shape—the plateau becomes steeper at longer wavelengths, though there are some exceptions.

[3] Another, more obvious explanation could be that I confused the two ports during measurements. I have double-checked that this did, in fact, not occur.

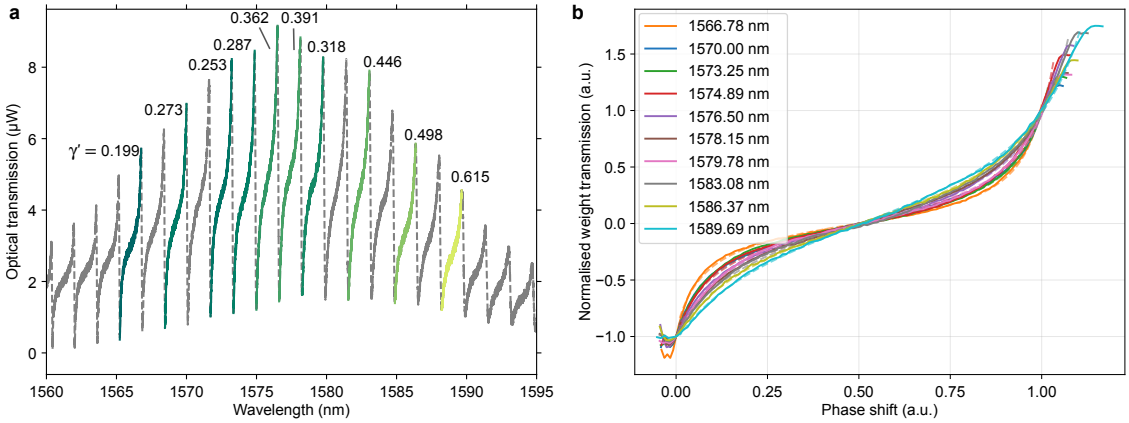

Figure 7.7: Wavelength dependence on the RAMZI curve shape. **a**, The transmission spectrum in Fig. 7.6b, coloured and annotated with the measured γ' from the phase sweeps. **b**, Neural network weight transmission vs. phase shift for different wavelengths, normalised symmetrically to -1 and 1 on the y-axis and 0 and 1 on the x-axis for each sweep. The fit of each curve is presented as a dashed line in the same colour.

It is well known that conventional DCs show wavelength-dependent coupling [277], and this is true also for our DCs. If we compare the simulation data in Fig. 7.5 with the transmission curves in Fig. 7.7b, it is obvious that a similar effect is happening: when the coupling into the ring increases with wavelength, so does the γ' parameter. If we want to mitigate this wavelength dependency, for example, to enable WDM schemes, we can use a curved DC or MMI in the coupling region instead.

Wavelength dependence of the phase control

So far in this section, I have not commented on the practical considerations of using a TO phase shifter to control the RAMZI weight. In this work, the heater response shows a wavelength dependence. For example, for the heater used in the device in Fig. A.4 with resistance of 1.14 kΩ, the power needed to sweep across the tangent shape varied between 52.2 mW and 55.8 mW for wavelengths 1515 nm and 1557 nm, respectively. In this particular example, there is a 2.8 % wavelength change (from 1515 nm to 1557 nm) and a 6.8 % change in the power required to sweep the tangent curve. In addition to the wavelength-dependent phase shift, $\Delta\varphi \propto P/\lambda$, the extra discrepancy could be explained by differences in the shapes of the tangent curves (how abrupt the transmission change at resonance is). For

WDM schemes, we ideally want the state of the phase shifter to be the same across wavelengths—in our current experiments, it is not.

7.3.3 Varying the shape of the curve

As mentioned in Section 7.2.2, we should be able to vary the shape of the curve by changing the coupling into the ring. First, I present results from different devices with varying coupling lengths and then from devices with a heater in the coupling region.

Changing the length of the coupling region

The length of the coupling region directly impacts the coupling into the ring (Eq. (3.16)). In Fig. 7.3, I showed INTERCONNECT simulations of the same device with different coupling conditions. Here, I show the change in the curve shape with different ring coupling lengths L_c (as indicated by the annotation in Fig. 7.8a). In Fig. 7.8b, I show one FSR each from four different transmission spectra. It is apparent that the jaggedness is much more pronounced in these spectra than in the phase sweep in Fig. 7.6c.

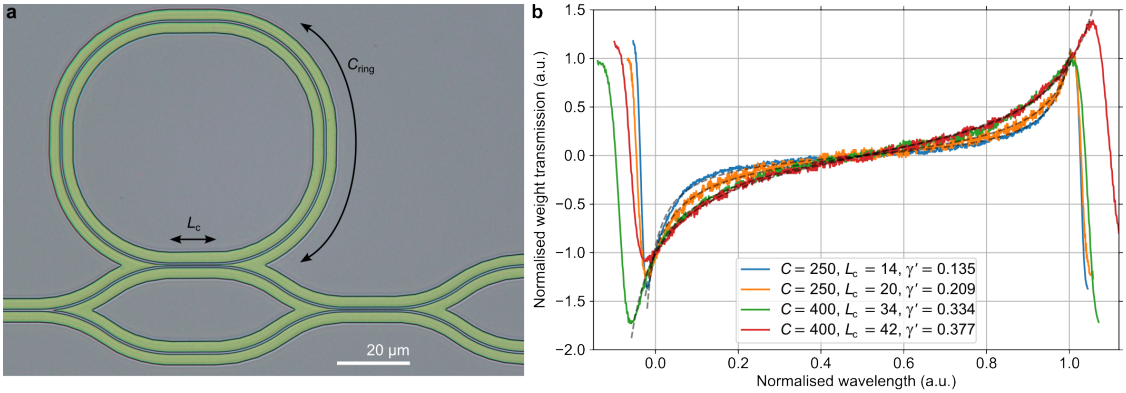

Figure 7.8: Coupling-length dependence on the RAMZI curve shape. **a**, Optical micrograph of a device without heaters, with ring circumference $C_{\text{ring}} = 250\,\mu\text{m}$ and ring coupling length $L_c = 12\,\mu\text{m}$. The results from this device are not shown in this work. **b**, Single resonances from the transmission spectra (transmission vs. wavelength) of different devices, normalised symmetrically to $(-1, 1)$ on the y-axis and $(0, 1)$ on the x-axis for the region of interest. The resonances shown are those found closest to $\lambda = 1576\,\text{nm}$ in each spectrum. The red plot ($C = 400$, $L_c = 42$) is from the resonance labelled 0.362 in Fig. 7.7a, and resonances from devices with $C = 250$ have been mirrored.

In theory, a clear correlation is expected between the coupling length and the shape of the curve. Although variations are visible in the plots in Fig. 7.8b, the

coupling condition calculated from Eq. (4.5) gives $k^2 = 0.39$, 0.01, 0.82, and 0.01 for $L_c = 14$, 20, 34.6, and 42.6 µm, respectively. This implies that the devices with $L_c = 20$ and 42.6 µm would be expected to exhibit similar curve shapes However, as I will discuss in Section 7.3.5, several practical factors influence the curve shape, making direct comparison between devices difficult.

Heating the coupling region

To get reconfigurable control of the shape of the tangent curve, we used phase shifters in the coupling region. As mentioned previously, we made two different types of designs for this coupling control: design 1 with a heater across the coupling region (Section 7.2.2a) and designs 2 and 3 with a heater spanning not only the coupling region but also the other arm of the MZI (Section 7.2.2b). Because heater design 2 did not produce results, possibly due to poor lift-off of the narrower nichrome heaters, we deposited design 3 on the same device. In Figs. 7.9 and 7.10, I show results from chips with design 1 and 3, respectively.

Figure 7.9 shows the effect of design 1. An optical micrograph of a fabricated device with design 1 is shown in Fig. 7.9a. The neural network weight transmission vs. phase shift is shown in Fig. 7.9b for varying coupling. In the figure, we see that the heater lifts the plateau of the curve, similar to when we apply a phase shift to one of the MZI arms in Fig. 7.2b.

In Fig. 7.9c, we investigate the effect of the varying coupling on the transmission spectrum. In the plots, we see two overlapping effects. First, the plateau of the tangent curve moves up and down with increasing phase shift (indicated by the applied voltages) and the curves become mirror images of each other (compare 1.0 V with 5.5 V and 2.0 V with 5.0 V). This effect stems from the increasing phase difference between the arms of the MZI. Secondly, we observe that the resonances redshift due to partial heating of the ring resonator (the ring waveguide in the coupling region).

In Fig. 7.10, I present the results from a device with heater design 3. Figure 7.10a shows the normalised neural network weight transmission vs. phase shift for varying

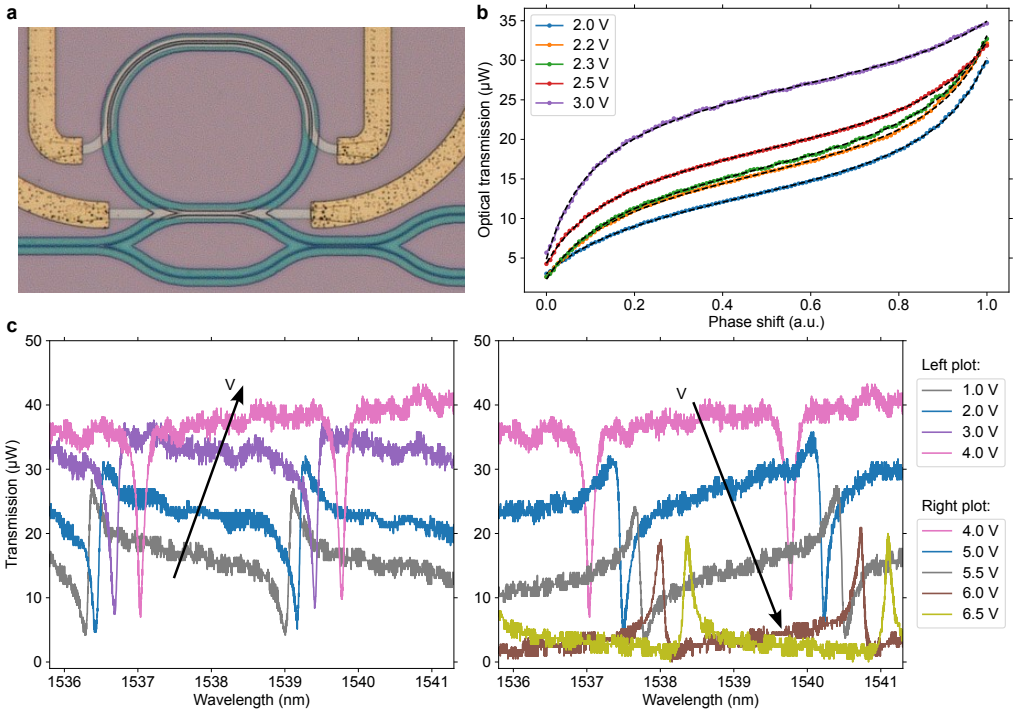

Figure 7.9: Effect of coupling heater on RAMZI transmission for design 1. **a**, Optical micrograph of a device with coupling heater design 1. The ring circumference and coupling length are $C_{\text{ring}} = 225\,\mu\text{m}$ and $L_c = 12\,\mu\text{m}$, respectively. The results from this device are not shown in this work. **b**,**c**, Transmission response for different couplings, indicated by the voltage applied to the heater, showing the plateau changing its position, for phase sweeps (b) and transmission spectra (c). The RAMZI has $C_{\text{ring}} = 225\,\mu\text{m}$ and $L_c = 15\,\mu\text{m}$. In **c**, the two panels show the same wavelength range.

coupling. The γ' and d of the curve fits were (0.4082, 1.328), (0.3927, 1.316), (0.3813, 1.315), and (0.3498, 1.272) for the curves with voltages 0.0 V to 11.2 V, respectively. The curves were measured at probe wavelengths 1576.53 nm, 1576.43 nm, 1576.46 nm and 1576.34 nm for voltages 0.0 V to 11.2 V, respectively. Transmission spectra of a device with varying coupling are presented in Fig. 7.10b. We see that the curves are redshifted by about $\pi/2$ between spectra, due to heating of the ring. From the plots, it is difficult to identify a clear change in the curve shape.

From Eq. (3.16), we have that the coupling across a DC is largest when $\sin^2(\xi L_c + C) = 1$, i.e., $L_c = \frac{1}{\xi}(\arcsin(\pm 1) + m\pi + C)$, where m is an integer and $\xi = 0.13$ and $C = 0.65$ from Eq. (4.5). In fact, it is only for the coupling lengths with $m \in \mathbb{N}$ (e.g., $L_c = 7.1\,\mu\text{m}$ and $31.2\,\mu\text{m}$) that we can achieve the maximum extinction ratio of our DC switch. In contrast, for $L_c = 42.6\,\mu\text{m}$, the initial change

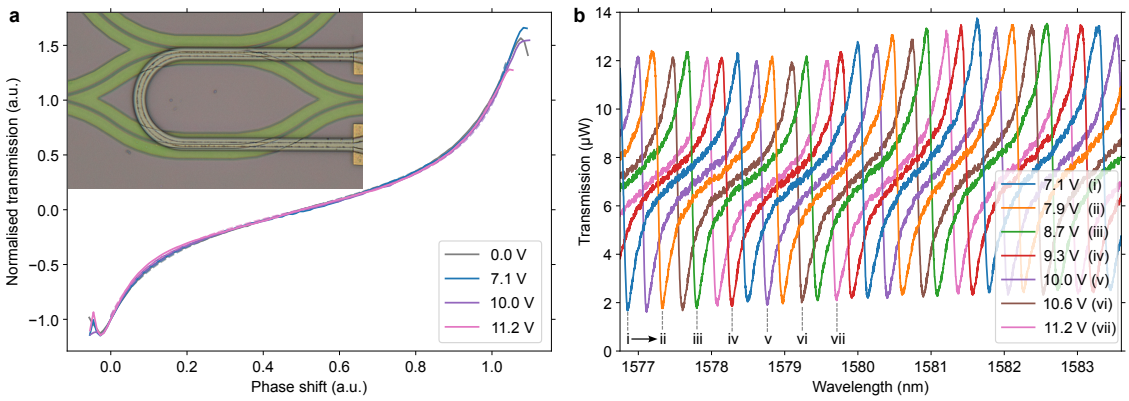

Figure 7.10: Effect of coupling heater on RAMZI transmission for design 3, with $C_{\text{ring}} = 400\,\mu\text{m}$ and $L_{\text{c}} = 42.6\,\mu\text{m}$. **a**, Normalised neural network weight transmission vs. phase shift for varying coupling, indicated by the voltage applied to the heater. The inset shows an optical micrograph of a device with coupling heater design 3 and ring coupling length $L_{\text{c}} = 12\,\mu\text{m}$. **b**, Measured transmission spectra of a RAMZI for different coupling voltages. The curves are redshifted by about $\pi/2$ between each spectrum.

in coupling across the DC is tiny for small Δn_{eff}; however, this claim assumes almost no coupling at $L_{\text{c}} = 42.6\,\mu\text{m}$, which seems not to be the case from the shape of the curves in Fig. 7.10. Therefore, it is not immediately clear whether the slight change in the shape of the curve is caused by measurement uncertainty or by the thermal gradient between the two arms of the coupling region.

7.3.4 Neural network weight quantisation and accuracy

In a neural network, we perform MAC operations by multiplying the input vector with the weight matrix and adding the products from each column. In each weight in the neural network, we multiply the light input, x, by the transmittance of the weight, w, to obtain $y = xw$. Here, I explain how to achieve this multiplication using RAMZIs as the amplitude-modulating weights, i.e., how the transmission of a RAMZI can be used to set the value of a non-uniformly quantised neural network weight. Furthermore, because of device imperfections in analogue systems, we also investigate the multiplication accuracy, i.e., the measured vs. expected product.

Non-uniform-in-weight quantisation

The RAMZI device can be used to represent an amplitude-modulating weight in any analogue photonic neural network, but its electrical-to-optical transfer function

is especially beneficial in quantised systems. In Fig. 7.11, I show how we approach this quantisation. For a given transfer function with midpoint $y = 0$ and extreme values $y = \pm 1$ on the y-axis, we divide the x-axis into $2^N - 1$ equal intervals between $x_{\min}$ and $x_{\max}$. This gives 2^N uniformly spaced points on the x-axis, where N is the number of bits. The y-values corresponding to each of these x-values are non-uniformly spaced on the y-axis. A non-uniformly quantised neural network weight with 3-bit precision is shown in Fig. 7.11a, for a curve symmetric around 0. We can also use only positive x-values to get non-uniformly spaced positive y-values between 0 and 1.

In Fig. 7.11b–f, we show the conversion between floating-point and quantised values, from Ref. 213 and our own work. These plots are obtained by taking the closest quantised number for each floating-point value. For example, in Fig. 7.11a, if we want the transmission value -0.9, we instead get the closest quantised value, which is -1, while for a transmission value -0.4, we get -0.34. In practice, we plot the array of quantised transmission values against itself in a step plot, with the midpoint of each step coinciding with the quantised value. First, in Fig. 7.11b, I show the quantisation of the two most extreme curves from Fig. 7.7b. It is clear that the quantisation levels in one of the plots are almost twice as dense for small float values as in the other, and vice versa for large float values. This difference between the transfer functions allows varying the non-uniformity of the quantisation, which is useful when the weight distributions across neural network layers vary.

Figure 7.11c–e are taken from Ref. 213 and show uniform quantisation, PoT quantisation, and APoT quantisation, respectively. The latter two methods are hardware-friendly quantisation algorithms that replicate the non-uniform weight distribution found in neural networks [213]. In Fig. 7.11f, I show the non-uniform-in-weight alternative, with 3- and 4-bit quantisation of the transmission curve presented in Fig. 7.11a. It is clear that our curve has much smoother steps than that of the APoT technique (indicated by the asterisks in Fig. 7.11e), without being confined to the ultrahigh, rigid resolution around 0 for the PoT technique.

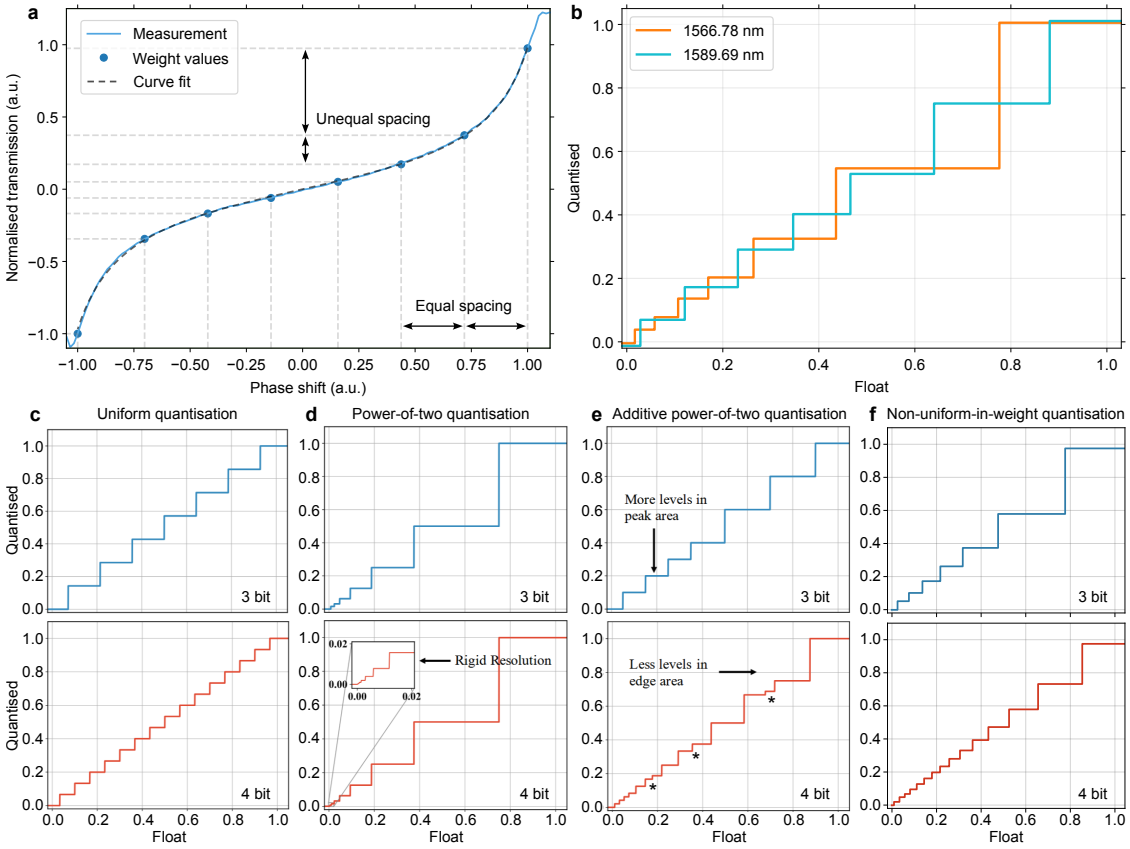

Figure 7.11: Non-uniform weight quantisation in photonic devices. **a**, Working principle of the non-uniform-in-weight quantisation, for 3-bit precision. The non-linear electrical-to-optical transfer function of the RAMZI converts uniformly spaced phase shifts to non-uniformly spaced transmission values. The curve is the same as the one labelled '1570.00 nm' in Fig. 7.7b. **b**, 3-bit quantisation of two of the curves shown in Fig. 7.7b. **c**–**f**, Quantisation with 3-bit (upper panels) and 4-bit (lower panels) precision, for uniform quantisation (c), power-of-two quantisation (PoT) quantisation (d), additive power-of-two quantisation (APoT) quantisation (e), and non-uniform-in-weight quantisation of the curve in **a** (f). In **e**, the asterisks indicate suboptimal quantisation levels. Figure **c**–**e** modified from Ref. 213 with permission.

A natural question arises when investigating Fig. 7.11a: why is there no weight value at 0? The whole point of non-uniform weight quantisation is to mimic the bell-shaped distribution of weights, which has most values close to 0. However, if the quantised values go from -1 to 1, there will be one value $-\delta$ right below 0 and another value δ right above, but no value at 0, simply because of the nature of even numbers (which bits are). A simple solution to this problem is to focus only on positive floating-point numbers, as done in the work by Li *et al.* [213]; the quantised values are automatically assigned to the range 0 and 1 with a value at 0. Because

works focused on software approaches to non-uniform weight quantisation also have no weights at 0 [300, 301], I will not delve into this matter further.

Calculation accuracy

We normally define the multiplication accuracy as the measured vs. expected result of the multiplication, for instance, $y_{\text{meas}} = 0.48 \times 0.49 = 0.24$ vs. $y_{\text{exp}} = 0.50 \times 0.50 = 0.25$, which has an error of $\epsilon = 0.24 - 0.25 = -0.01$. For amplitude-modulating weights with linear electrical-to-optical transfer functions in photonic crossbar arrays, the procedure is straightforward: we set the laser power to random levels between x_{min} and x_{max} and the electrical input (and consequently the transmission of the weight, $T(w) = w$) to random levels between w_{min} and w_{max}, and calculate the product $y_{\text{exp}} = xw$ and measure the output power y_{meas} for each set of values. However, for a non-linear electrical-to-optical transfer function, we cannot use the electrical input values directly. Two approaches can be used instead. First, we can use a lookup table to store pairs of x- and y-values from the electrical-to-optical transfer function. For the curve in Fig. 7.11a, for example, we have that the transmission is $T = -0.34$ when the phase shift (i.e., electrical input) is $w = -0.714$. By converting the desired transmission values to the corresponding electrical input, $T \rightarrow w$, we obtain the correct weight values. The second approach exploits the fact that the transmission response of the RAMZI can be approximated by a tangent function. However, since the transfer function gives the relationship between the electrical input and weight transmission, we must take the inverse of the transfer function to obtain the correct weight values, $w = T(w)^{-1}$.

In Fig. 7.12, I show the multiplication accuracy of RAMZI weights using the second approach described in the previous paragraph. I define the curve fit $Z_{\text{fit}}(x, T(w))$ as the expected values and compare them with the measured values $Z_{\text{meas}}(x, T(w))$. In Fig. 7.12a, I show the multiplication accuracy from the curve shown in Fig. 7.6c. The overlap is very high, as seen from the normalised error distribution of the plot, shown in Fig. 7.12b. For these measurements, I apply a single laser power ($x = 1\,\text{mW}$) and sweep the voltages rapidly.

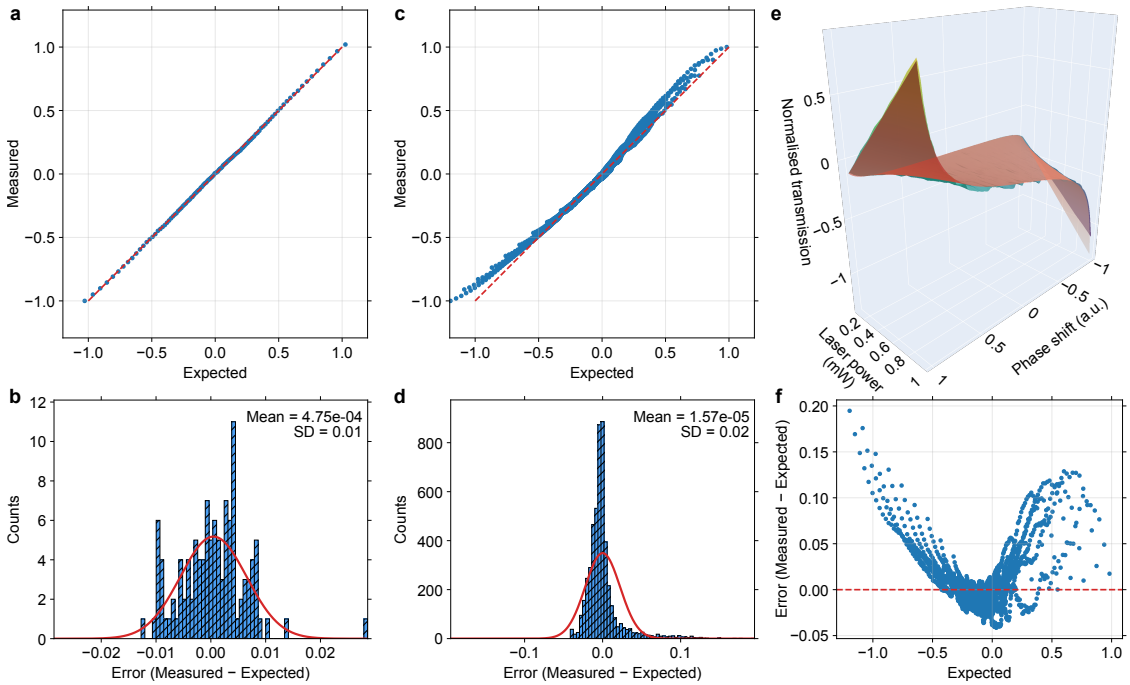

Figure 7.12: Multiplication accuracy of neural network weights. **a**,**b**, Accuracy (a) and normalised error distribution (b) of the device shown in Fig. 7.6c, for a single laser power and 114 voltage values. **c**,**d**, Accuracy (c) and normalised error distribution (d) of 4960 multiplication results, with 10 laser powers (from 0.1 mW to 1 mW) and 496 random voltage values. **e**, 3-D plot of the expected and measured planes for different laser powers and voltages. **f**, Error of the plot in **c**. The solid red curves in **b** and **d** are Gaussian fits to the histograms. SD: standard deviation.

More thorough calculation-accuracy measurements were conducted in collaboration with my colleague Jonathan Dearlove, with the analysis performed by him under my supervision. These results are shown in Fig. 7.12c–f. In Fig. 7.12c,d, I show plots equivalent to those in Fig. 7.12a,b, respectively. The measurement data and curve fit are shown in a 3-D plot in Fig. 7.12e, with the error plot shown in Fig. 7.12f. It is clear from these measurements that the discrepancy between the expected and measured values increases for extreme values, which is inevitable: the tangent function is only a good fit to the curve between two resonances. Although suboptimal, this behaviour is not detrimental to device functionality, as we are more concerned about the accuracy around 0 than at the extremes, and the error for small values is negligible. Furthermore, the error in Fig. 7.12c is considerably larger than that in Fig. 7.12a, probably due to a poorer fit to the measured data. Moreover, we see some waviness in the plots in Fig. 7.12e,f,

which is likely caused by drift in the laser power.

7.3.5 Discussion

In the fabricated devices shown in this chapter, several practical aspects affect the shape of the RAMZI transmission curve, making comparisons between devices difficult at times. First, it is worth noting that achieving good results across all devices was not trivial. For instance, for some of the devices with transmission shown in Fig. 7.8b, the ring heater was defective, and we were consequently not able to apply a phase shift. Instead, we measured the transmission spectrum. The transmission spectra in this work are all enveloped by the Gaussian-like transmission of the grating couplers, which affects the measured RAMZI curve. However, because I plotted data from the centre of the Gaussian envelope ($\lambda = 1576\,\text{nm}$), the impact of said envelope should be minimal on the plots in Fig. 7.8b. A similar problem is covered in Appendix A.3.3, in which a reflection in the light path resulted in a sinusoidal envelope across the transmission spectrum of some devices, which directly impacts the observed curve shape in the transmission spectrum.

In both subsections of Section 7.3.3, there are examples of discrepancies from the expected behaviour with coupling given by Eq. (4.5). In Fig. 7.8b, the length of the coupling region is $L_\text{c} = 14$, 20, 34.6, and 42.6 µm, which gives $k^2 = 0.39$, 0.01, 0.82, and 0.01, respectively, according to Eq. (4.5). From theory, we expect the shapes to reflect the coupling, but this does not seem to be the case. However, qualitatively, based on RAMZI spectra not shown here, it appears that devices with 16 and 38.6 µm coupling lengths have near-zero coupling into the ring. If we use, for instance, $k^2 = \sin^2(0.143L_\text{c} + 0.77)$ instead of the $\sin^2(0.13L_\text{c} + 0.65)$ found in Section 4.5.4, we get $k^2 \approx 0$ for $L_\text{c} = 16$ and 38.6 µm, and increasing coupling for $L_\text{c} = 14$, 20, 34.6, and 42.6 µm, with 34.6 and 42.6 µm being similar. This scenario is more similar to what we see qualitatively from the curve shapes in Figs. 7.6, 7.8 and 7.10. One explanation for the discrepancy between theory and practice is that Eq. (4.5) is not accurate for the devices fabricated in this project and may vary with fabrication parameters, such as etch depth. Another

explanation is that other effects, such as the envelope of the transmission spectrum, are more pronounced, as shown in Fig. A.4.

7.4 Future work

Thus far, I have presented a TO RAMZI weight for inherent non-uniform quantisation of neural networks. Here, I will discuss alternative technologies and designs, methods for integrating amplitude-modulating weights into photonic crossbar arrays, and the power consumption of our device.

7.4.1 Electrical tuning

The RAMZI weight shown in this chapter uses TO phase shifters to set the weight. As mentioned in Section 7.2.2, because the phase response of the TO heater in the RAMZI ring is linearly proportional to the heater power and therefore quadratically proportional to the heater voltage (i.e., $\phi \propto P \propto V^2$), by having a voltage source with linearly spaced values, we will not achieve the correct Lorentzian curve for the RAMZI transmission. As a result, we compensated for this effect by applying $V' = \sqrt{V}$ so that $P \propto V'$. By doing this voltage conversion, the whole argument of using non-uniform photonic weights in the first place is gone, as we need non-linearly spaced voltages from the DAC. Instead, we could have just used a non-linearly spaced DAC to achieve the Lorentzian-based weight sensitivity directly.

However, the concept presented in this chapter is platform-agnostic, and we can therefore substitute the phase shifters with any phase shifter with a linear voltage-to-phase curve. Such phase shifters include those based on the Pockels effect (as seen in Section 2.2.2). Furthermore, because the refractive index change introduced by the Pockels effect is linear for both positive and negative voltages, we can employ such phase shifters to achieve $w = 0$ at $V = 0$, with $w_{\max}$ at $V > 0$ and $w_{\min}$ at $V < 0$. Because most weight values are near $w = 0$, we can further reduce power consumption. The apparent downside of Pockels-effect devices is, of course, their significant footprint.

7.4.2 System integration

The weight presented herein will benefit any photonic neural network accelerator in which the weights are directly mapped to the physical hardware. Examples include photonic crossbar arrays (shown in Fig. 7.13a) [4, 190] and photonic dot-product engines [195, 249, 302]. Although the weights can also be used in forward-only MZI meshes, they do not provide any benefits, as the matrix is formed through singular value decomposition and therefore the distribution of matrix elements does not follow a bell curve [11].

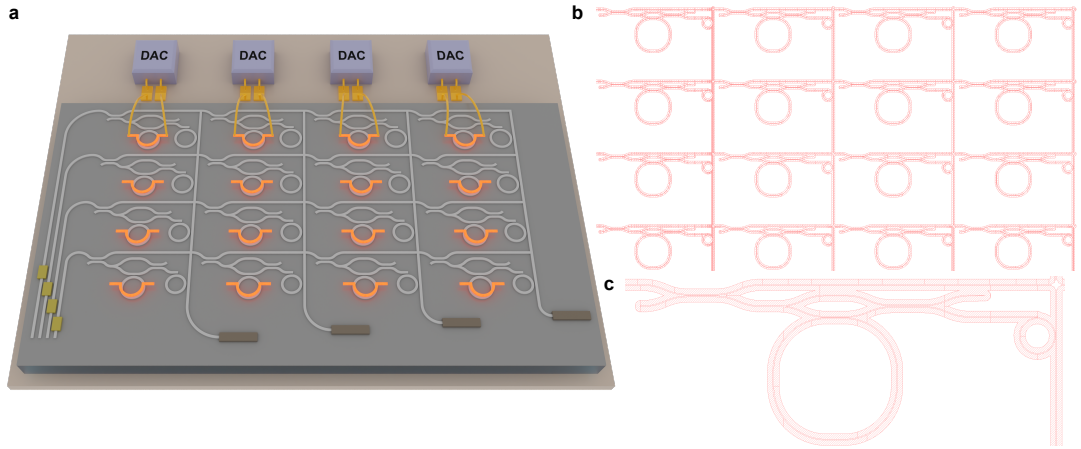

Figure 7.13: Photonic crossbar array with RAMZI weights. **a**, Schematic of a PTC with DACs for setting the RAMZI weight, with light modulators at the input of each row and PDs at the output of each column. **b**, GDS design of a 4×4 matrix, by Jonathan Dearlove. **c**, Zoom-in of one unit cell in **b**. The ring radii are 225 µm and 10 µm for the RAMZI and the output rings, respectively.

We can achieve WDM in a system using our RAMZI weights, although with certain caveats. Because of the ring-based resonance of the RAMZI weight, we cannot utilise dense WDM in the way we can for non-resonant amplitude modulators [4]. Still, in theory, if all resonances are identical, we can achieve WDM by employing one wavelength per RAMZI resonance and setting the wavelengths so that for a full voltage sweep, each tangent curve is symmetric around the curve's midpoint. In practice, however, I have shown that the tangent curve shapes are wavelength-dependent, stemming from the coupling condition of the DC, and that the phase shifter is inherently wavelength-dependent as well. Both of these limitations can

be overcome, for instance, by employing broadband couplers in the ring and using phase shifters with small wavelength dependence.

The reduced WDM capacity might even allow for scaling of the matrix. In Farmakidis *et al.* [184], the authors point out that for photonic crossbar arrays, we can employ resonant structures such as MRRs to trade off some WDM capacity with increased scalability. A photonic crossbar array with MRRs is shown in Ref. 191. Because our RAMZI weight is a resonant structure, we can use ring resonators at the output to couple to the bus waveguides in each column, without any extra WDM-related bandwidth loss. In Fig. 7.13b, a GDS design of a 4×4 matrix with RAMZIs as amplitude-modulating weights is presented, with one unit cell shown in Fig. 7.13c. Although the RAMZI is larger than, for instance, a PCM element, the unit cell size can still be comparable in size by optimising the design.

7.4.3 Alternative photonic structures

In future devices, to reduce the footprint and, potentially, the insertion loss of the amplitude-modulating weight, we can investigate different structures that produce similar transmission responses. One example is shown in Fig. 7.14a, in which two air holes constitute a low-finesse Fabry–Pérot cavity in the bus waveguide that leads to the spectrum shown in Fig. 7.14b when combined with a ring resonator [303]. Similarly, through modal coupling in the conveyor-belt coupler to the ring in Fig. 7.14c, we can achieve the spectrum shown in Fig. 7.14d. We see that the transmission spectra exhibit both Lorentzian, electromagnetically induced transparency (EIT), and Fano lineshapes—only the latter is of interest in our system. Consequently, in these devices, there are only a few useful wavelengths of interest for neural network weights.

More promising designs are shown in Fig. 7.15 for add-drop ring resonators. A ring with T-shaped bus waveguides is shown in Fig. 7.15a, which produces the desired curve across the entire transmission spectrum (Fig. 7.15b). A Mach–Zehnder optical switch with a cross-ring resonator coupled to both arms (also called a ring-resonator two-beam interferometer), shown in Fig. 7.15c, produces similar results

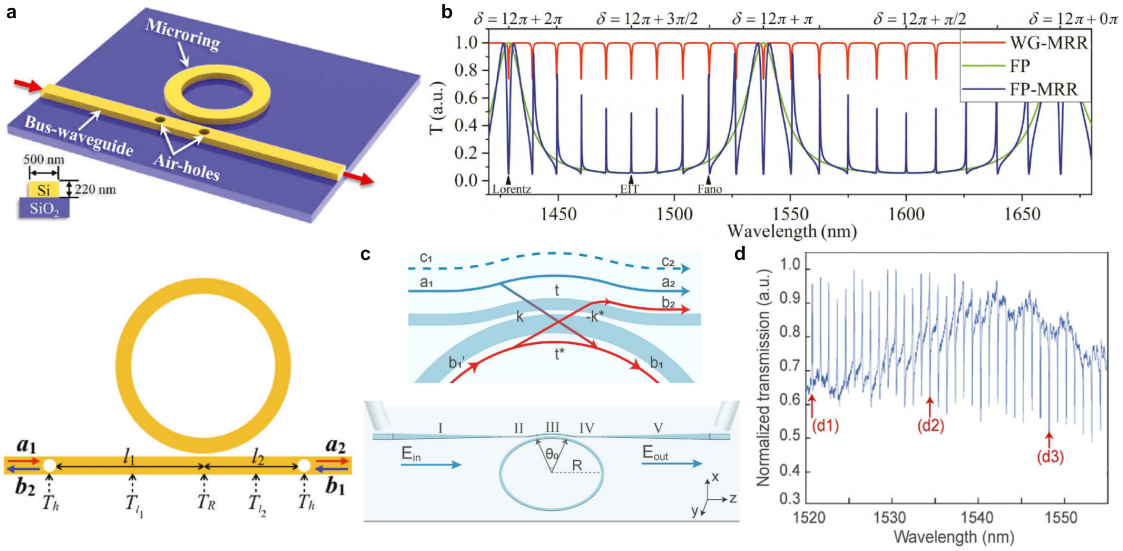

Figure 7.14: All-pass ring resonators showing Fano resonances. **a**, Schematics of the MRR with two air holes constituting a low-finesse Fabry–Pérot cavity. **b**, Calculated transmission spectra of a waveguide with the Fabry–Pérot cavity (labelled FP), an MRR coupled with a normal waveguide (WG-MRR), and an MRR coupled with the Fabry–Pérot cavity (FP-MRR). **c**, Schematic illustration of the modal coupling in the conveyor-belt coupler to the ring and experimental setup for a two-mode system. **d**, Normalised transmission spectrum for the device in **c**, showing EIT (d1), Fano (d2), and Lorentzian lineshapes (d3). Figure reproduced with permission from: **a**,**b**, Ref. 303, under a Creative Commons licence CC BY 4.0; **c**,**d**, Ref. 304, © 2024 Wiley–VCH GmbH.

for certain phase differences, δ (Fig. 7.15d). Especially the former design holds promise for small-footprint integration. However, from Ref. 305, it appears that there is a correlation between the size of the perturbation and the shape of the curve, as well as the insertion loss, indicating that certain curve shapes are associated with higher loss. Moreover, the integrated heaters used to shift the resonance in Ref. 305 do not produce a large enough phase shift. Further research is warranted to investigate the loss and phase control of these devices.

7.4.4 Energy calculations

Finally, I will comment on the benefits of using RAMZI weights for inherent non-uniform weight quantisation. First, quantisation alone can be extremely beneficial. For example, Zhao and Yuan claim that 4-bit weight quantisation can reduce the memory usage for model loading by nearly 75 % compared with half precision (FP16) [216]. In Fig. 6 of Ref. 307, the authors demonstrate that, in certain layers of

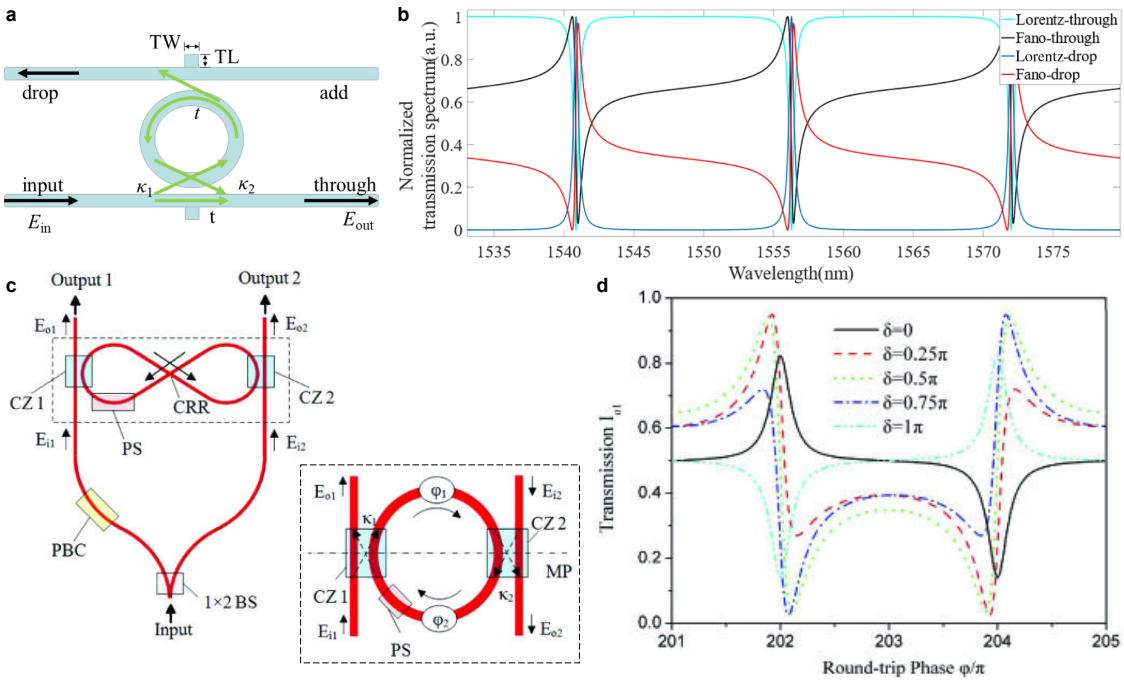

Figure 7.15: Add-drop ring resonators showing Fano resonances. **a**, Schematic showing an add-drop ring resonator with T-shaped bus waveguides. **b**, Calculated transmission spectra of the through and drop ports of a ring resonator with conventional bus waveguides (labelled 'Lorentz') and T-shaped bus waveguides ('Fano'). **c**, Schematic of an MZI with a dual-bus-coupled cross-ring resonator. The inset shows an equivalent representation of the dashed part of the structure, i.e., a dual-bus-coupled ring resonator. MP: mirror plane, CRR: cross-ring resonator, PS: phase shifter, BS: beam splitter, CZ: coupling zone, PBC: phase bias controller. **d**, Calculated transmission vs. round-trip phase φ for Output 1, for different phase differences δ. Figure adapted with permission from: **a**,**b**, Ref. 305, © 2022 Optica Publishing Group; **c**,**d**, Ref. 306, © 2009 Optica Publishing Group.

specific neural network architectures, the total energy consumption can be dominated almost entirely by weight-setting operations [307]. Although this dominance is not universal across all layers or models, quantisation generally reduces energy consumption. Furthermore, by reducing memory traffic and alleviating memory-bound bottlenecks, quantisation speeds up neural network inference [216]. The key to this project is to reap these benefits of quantisation while maintaining optimal representations of the weight information through the non-uniform behaviour, thereby avoiding accuracy loss in the network.

Photonic neuromorphic computing exploits the benefits of analogue computing, but is also bound by its limitations. One such limitation is the digital-to-analogue and analogue-to-digital conversions that occur at the interface between the photonic

chip and the input and output, for instance, to set the neural network's weights. As mentioned in Section 3.7.1, the power consumption of DACs grows exponentially with the number of bits [217, 218], meaning high-resolution weight setting is detrimental to device performance. Therefore, reducing the bit-width of the DACs can yield significant energy savings. In Ref. 213, the authors show that 3- and 4-bit quantised networks using the algorithm presented in Fig. 7.11e perform similarly to full-precision (FP32) networks. Assuming our device yields similar non-uniform quantisation, we can reduce the power consumption of the DACs by more than an order of magnitude compared with an 8-bit DAC.

Finally, a comment on the use of non-volatile memories is warranted. If the state of the amplitude-modulating weights in the matrix is set using non-volatile memories, there is no need for a constant voltage from the DACs, and the power consumption from weighting is drastically reduced. Nevertheless, since many non-volatile elements, for instance, phase-change memories, typically have limited weight resolution, it can still be beneficial to employ non-uniform weight quantisation using RAMZIs.

7.5 Chapter summary

In this chapter, I have shown the process of designing, simulating, fabricating, and measuring a photonic device for inherent non-uniform weight quantisation in neural networks. The device is a ring-assisted Mach–Zehnder interferometer (RAMZI) with an all-pass ring in one arm and a $\pi/2$ phase shift between the arms of the MZI, which gives an electrical-to-optical transfer function that follows the desired curve, i.e., the quantile function of a bell curve. Various simulations of the device behaviour were conducted, during which I investigated the effects of different designs and parameters to understand the device's working principle.

I then presented measurements from fabricated devices. Some devices performed exceptionally well, with excellent overlap with the expected curve. These devices can be used for non-uniform-in-weight quantisation, yielding smoother curves than those achieved with electronic implementations. Furthermore, the multiplication errors of these weights are, in ideal cases, close to zero.

For device tuning, we are currently limited by suboptimal heater design. Because the coupling region is shared by both the ring and the MZI, any modification to the coupling inevitably alters the behaviour of both the ring and the MZI. By carefully considering the device design, the latter effect was mitigated. However, the coupling heater design in this work was not optimised, resulting in only a small level of tunability. Further research is needed to fully exploit RAMZIs for photonic neural network applications.

Finally, I discuss some of the future work. Instead of using the thermo-optic (TO) metal heaters employed in this work, we can exploit the benefits of other, better optimised phase shifters, such as those based on the Pockels effect. Furthermore, in weight matrices employing RAMZIs as the amplitude-modulating weights, we can introduce ring resonators to couple the light, thereby reducing the loss. Because RAMZIs are resonant devices and therefore have limited wavelength-division multiplexing (WDM) capacity, this design alteration enables improved matrix scalability without further limiting the system's WDM capabilities. Additionally, I introduce several alternative structures that can replace the RAMZI design, and briefly discuss the benefits of non-uniform-in-weight quantisation, including reductions in memory, energy, and inference latency.

The material in this chapter is based on a manuscript currently in preparation:
H. H. Toftevaag, M. Zolfagharinejad, J. Dearlove, N. Farmakidis, A. Ortega-Gomez, W. G. van der Wiel, and H. Bhaskaran, "Non-uniform photonic neural network weight quantization" (in preparation).

In the end, everything will be okay.
If it's not okay, it's not yet the end.

— Unknown

8 Conclusion and outlook

This thesis investigates photonic hardware for neuromorphic and optimisation-oriented computing, with an emphasis on non-volatile programmability, energy efficiency, and architectural scalability. Across three main projects, I address algorithm–hardware co-design, device-level innovations, and circuit-level functionality, with the common objective of enabling beyond-electronics computing platforms for control, optimisation, and neural network applications.

The first project introduces photonic neuromorphic computing in the context of sensor fusion and sensor management, with particular attention to Kalman filtering and model predictive control (MPC). The proposed computing platform is based on photonic crossbar arrays with phase-change material (PCM) weights, combining non-volatility with ultrafast optical modulation and detection. While this architecture offers compelling advantages in speed and static power consumption, it also imposes practical constraints: in particular, high experimental complexity and limited reconfiguration frequency, and consequently restricted flexibility for algorithms that require frequent weight updates. These constraints motivated a pivot from general sensor fusion algorithms—such as Kalman filtering, which requires continuous updates of matrix elements—to MPC, where weight matrices can remain fixed or change infrequently.

Within this context, I present algorithmic formulations and photonic crossbar designs tailored to defence-relevant MPC problems, specifically target tracking. Performance estimates indicate that, for simple MPC problems, the proposed architecture can reach steady-state solutions within 12 ns (corresponding to update frequencies of approximately 80 MHz), substantially outperforming conventional digital implementations that typically operate on kilohertz timescales. This project establishes photonic MPC as a promising application domain for non-volatile photonic accelerators and demonstrates the importance of aligning algorithmic structure with hardware constraints.

The second project focuses on the experimental realisation of a non-volatile field-programmable coupler array (nv-FPCA) using the low-loss PCM Sb_2Se_3 as the active tuning material. The fundamental building block is a 2×2 MZI tunable coupler incorporating PCM-based phase shifters in both arms, controlled via electro-thermal microheaters. I describe the device design, supported by electro-thermal simulations, and present optical characterisation of both the individual couplers and the assembled nv-FPCA. The measured devices exhibit extinction ratios exceeding 20 dB over a 15 nm wavelength range, with moderate insertion losses that depend on the PCM state. Full amorphisation and crystallisation of the PCM patches require estimated switching energies of 1.2 µJ and 1.2 mJ, respectively, although these values are expected to decrease substantially with optimised pulse schemes and improved thermal engineering. Importantly, multi-level programming with more than 50 distinct states is demonstrated, highlighting the analogue programmability of the platform.

Using the demonstrated nv-FPCA unit cell, I synthesise key photonic components, including routers, ring resonators, and delay lines. The realised ring resonator exhibits a quality factor of approximately 8.5×10^3 and a free spectral range (FSR) of 40 GHz, with simulations indicating potential functionality as a temporal differentiator for edge detection. A broader comparison with thermo-optic and MEMS-based platforms reveals that PCM-based couplers offer unique advantages, mainly compact footprints and zero static power. However, our

phase-change devices are currently suffering from high switching energies, long reconfiguration times, loss variability, and limited demonstrated cyclability. Despite these challenges, this work validates low-loss PCM-based tunable couplers as a viable, complementary technology for infrequently reconfigured photonic meshes and outlines designs for future systems.

The third project explores photonic devices for inherent non-uniform weight quantisation in neural networks. I present the design, simulation, fabrication, and characterisation of a ring-assisted Mach–Zehnder interferometer (RAMZI) incorporating an all-pass ring resonator. By introducing a fixed $\pi/2$ phase shift between the MZI arms, the device realises an electrical-to-optical transfer function that closely follows the quantile function of a Lorentzian distribution. Experimental results show excellent agreement between measured and expected responses for selected devices, enabling smooth non-uniform-in-weight quantisation with near-zero multiplication error under ideal conditions.

The project also identifies practical limitations related to device tuning. Because the coupling region is shared by both the ring resonator and the MZI arm, adjustments to the coupling affect both subsystems simultaneously. While careful design can mitigate this interaction, the thermo-optic heater implementation used here provides suboptimal tunability. I therefore outline several directions for future work, including the use of more suitable phase shifters (e.g., Pockels-effect-based devices) and alternative coupling strategies for large-scale implementation. More broadly, this project demonstrates the potential of photonic devices to embed non-uniform quantisation directly in hardware, offering reductions in memory requirements, energy consumption, and inference latency. The beauty of working in our field lies in finding applications where photonics outperforms electronics. Sometimes, that can be in ways less intuitive than simply ‘because signals travel at the speed of light’—here, I have presented a concept that integrates non-uniform weight quantisation into hardware in a way inherently difficult for electronics.

8.1 Future research and broader impact

This thesis has explored integrated photonic switch technologies with a strong emphasis on aligning device characteristics, architectures, and applications. Rather than advocating photonics as a universal replacement for electronics, the work demonstrates that photonic hardware offers decisive advantages when its strengths (e.g., parallelism, low latency, electromagnetic interference (EMI), and non-volatile programmability) are deliberately matched to suitable computational tasks. The devices and system concepts presented therefore open several promising directions for future research, both at the device and architectural levels and at the application level.

A natural extension of the work on the photonic platform for solving MPC problems presented in Chapter 5 is to perform measurements on the designed system. Doing so would allow us to discover potential bottlenecks and difficulties, for instance, in the bit resolution of the PCM weights. However, even if we had performed measurements on the system, in its current form, the design lacks several necessary components for ultrafast operation. Ways to address this limitation in future designs include replacing the integrated thermo-optic modulators that encode the input vector with faster alternatives (e.g., EAMs), and implementing the activation function and bias directly on chip. Furthermore, we know that the overall system latency is primarily limited not by the time of flight of the photonic signals, but by the electronic peripherals. A key future research direction, therefore, lies in tighter electronic–photonic co-design with analogue electronics. With these future avenues, we can considerably reduce the footprint of analogue optical computers by miniaturising from free-space to integrated photonics systems.

A second clear future direction is the fabrication and experimental characterisation of scaled nv-FPCAs, as outlined in Chapter 6. In this work, a first-generation nv-FPCA based on the low-loss PCM Sb_2Se_3 demonstrates non-volatile, multi-level tunability with zero static power consumption in the photonic mesh. While our foundry-level designs indicate a clear path towards larger meshes, experimental validation at increased scale remains an important next step. Future work should

focus on demonstrating multi-cell FPCAs with realistic loss budgets and calibration strategies. Such experiments would allow investigation of how accumulated loss, phase errors, and device-to-device variability affect performance when synthesising more complex photonic circuits. In addition, scaled FPCAs would serve as a useful testbed for rapid prototyping of photonic signal processors and for exploring redundancy and fault tolerance in safety-critical systems.

Although low-loss PCMs enable compact and non-volatile photonic devices, their finite cyclability and energy-intensive switching remain important considerations. Future research should therefore investigate application-level lifetime models. In parallel, the inherent immunity of photonic circuits to electromagnetic interference, combined with the radiation-tolerant nature of PCMs, makes these technologies attractive for operation in extreme environments, including high-radiation or high-EMI scenarios. Experimental studies addressing radiation tolerance, thermal stability, and long-term drift would further clarify the suitability of photonic computing hardware for defence, aerospace, and industrial control applications.

Finally, a natural extension of the work in Chapter 7 is the experimental realisation of photonic crossbar arrays employing non-uniformly quantised weight elements. In this thesis, non-uniform weight quantisation is demonstrated at the device level using a RAMZI, showing for the first time that the optical transfer function of a carefully designed photonic structure can be matched to the global weight distribution of a neural network layer. Integrating such weight elements into a full photonic crossbar array would enable a direct hardware implementation of non-uniform quantisation, eliminating the need to emulate such behaviour digitally or rely on uniformly quantised optical weights.

Future research could investigate how the non-uniform photonic weights affect system-level accuracy, noise tolerance, and energy efficiency when deployed in inference-oriented photonic tensor cores. In particular, when combined with the inherently limited precision of analogue photonic computation, non-uniform quantisation may enable further reductions in the resolution of DACs without

sacrificing performance. More broadly, this approach invites deeper algorithm–hardware co-design, where neural network training explicitly accounts for the transfer functions and noise characteristics of photonic hardware. Finally, the concept of embedding non-uniform quantisation directly into photonic devices may influence future approaches to analogue in-memory computing, both optical and electronic.

Collectively, this work contributes to the ongoing shift towards application-driven photonic hardware, in which the goal is not to replace electronics universally but to complement them with specialised, high-performance photonic subsystems. By combining algorithmic insight with device- and circuit-level innovation, this work highlights both the opportunities and the remaining challenges on the path toward scalable, energy-efficient photonic computing architectures. As integrated photonics continues to mature, the devices and architectures introduced in this thesis are expected to inform the design of future low-power, high-speed computing and signal-processing platforms in both civilian and defence contexts.

Appendices

To achieve great things, two things are needed: a plan and not quite enough time.

— Leonard Bernstein

A Appendices

Here, I present additional information to the three main results chapters in this work.

A.1 Model predictive control

A.1.1 Python code

The code for the advanced example shown in Fig. 5.4 is presented here. First, I define the MPC problem and the target trajectory before computing the `P` and `G` matrices. Subsequently, the code iteratively finds solutions for `u` and `lam` at each time step.

```
# ------------------------- MPC configuration --------------------------
N          = 3          # horizon length
n_controls = N          # one alpha per time step in the horizon
alpha_max  = 5.0        # deg/s² angular acceleration limit
theta_hi   = 60.0       # allowed angles (0-60 deg)
steps      = 150        # total simulation time steps
dt_iter    = 0.1        # primal-dual integration step
max_iter   = 200        # inner-loop maximum iterations

# Make model more realistic - only allow positive optical values
offset = 0
if more_realistic:
    offset = 13.0

# ---------------------- Target motion generator -----------------------
jumps = [67, 92]  # introduce random jumps in the target trajectory
def generate_target(path_len=120):
```

```
    theta_target_act, theta_target_exp = [], []
    theta_act, theta_exp  = 0.0, 0.0
    joker, joker_old = 0.0, 0.0
    for k in range(path_len - 30):
        # jumps at k = 67 and 92
        if k == jumps[0]:
            joker = np.pi / 4
        elif k == jumps[0] + N - 1:
            joker_old = np.pi / 4
        elif k == jumps[1]:
            joker = 0  # equivalent of a np.pi / 4 jump
        elif k == jumps[1] + N - 1:
            joker_old = 0

        theta_act = 40 + 30 * np.cos(k/120 * 15 - joker)
        theta_exp = 40 + 30 * np.cos(k/120 * 15 - joker_old)
        theta_target_act.append(theta_act)
        theta_target_exp.append(theta_exp)
    for k in range(30):
        theta_target_act.append(50)
        theta_target_exp.append(50)
    return np.array(theta_target_act), np.array(theta_target_exp)

theta_target, theta_target_mpc = generate_target(steps)

# ------- 1-D angular dynamics (theta, omega) with control alpha -------
A = np.array([[1, 1],
              [0, 1]], float)
B = np.array([[0.5],
              [1.0]], float)
C = np.array([[1.0, 0.0]])

# ----- Prediction matrices H (controls→theta) and F (state→theta) -----
H = np.zeros((N, n_controls))
F = np.zeros((N, 2))
# Make F and H
for i in range(N):
    Ai = np.linalg.matrix_power(A, i + 1)
    F[i] = (C @ Ai).ravel()
    for j in range(i + 1):
        H[i, j] += (C @ np.linalg.matrix_power(A, i - j) @ B).item()

P = H.T @ H + np.eye(n_controls)     # quadratic cost matrix
print("P is \n", P)

# -------------------- Constraint matrix  G ----------------------------
# per horizon step:  alpha <= alpha_max,   theta_pred <= theta_hi,
G_rows = []
```

```
for i in range(N):
    row_acc       = np.zeros(n_controls); row_acc[i] = 1
    row_pos_upper =  H[i].copy()
    G_rows.extend([row_acc, row_pos_upper])
G = np.vstack(G_rows)
print("G is \n", G)

# ---------------------- Before the loop ------------------------------
u_log   = []          # to store offset-versions for plotting
lam_log = []

# ------------------- Closed-loop simulation --------------------------
x = np.array([0.0, 0.0])         # initial [theta, omega] for camera
theta_cam = []

for k in range(steps):
    # ---- store current cam angle
    theta_cam.append(x[0])

    # ---- build reference for the next horizon
    idx = np.clip(np.arange(k + 1, k + N + 1), 0, steps - 1)
    T   = theta_target_mpc[idx]      # predicted theta for next 3 steps

    # ---- cost linear term
    q = H.T @ (F @ x - T)

    # ---- bounds vector h (depends on current state through Z_free)
    Z_free = F @ x
    h = []
    for i in range(N):
        h.extend([alpha_max,                  # alpha  <= alpha_max
                  theta_hi - Z_free[i]])      #  theta_pred  <= theta_hi
    h = np.array(h)

    # ---- primal-dual iterations
    u = np.zeros(n_controls)                  # alpha sequence
    lam = np.zeros(G.shape[0])                # Lagrange multipliers

    # Find u and lambda iteratively
    for _ in range(max_iter):
        u_old = u  # so that I can abort early if at steady state

        # offset to make sure all input light values are positive
        u_old += offset             # lam is always > 0

        # Photonic matrix multiplication
        u   = (np.eye(3) - dt_iter * P) @ u_old - dt_iter * G.T @ lam
        lam = lam + dt_iter * G @ u_old
```

```
        # Electrical input to EO converter: q and h, as well as offset.
        # This is fine, as the offset is constant and is
        # simply baked into the electrical bias.
        u   -= (dt_iter * q + (np.eye(3) - dt_iter * P) @
                (np.ones_like(u_old) * offset))
        lam -= (dt_iter * h + dt_iter * G @
                (np.ones_like(u_old) * offset))

        if abs(u[0] - u_old[0]) < 1e-6:
            break  # break early if at steady state

        if more_realistic:
            # Introduce randomness (normally distributed)
            u   += np.random.normal(loc=0.0, scale=0.1, size=u.shape)
            lam += np.random.normal(loc=0.0, scale=0.1, size=lam.shape)

        lam = np.maximum(lam, 0)  # ReLU function

        # Store for plotting
        u_log.append(u_old.copy())
        lam_log.append(lam.copy())

    # ---- apply first acceleration and propagate cam plant
    alpha = u[0]
    x     = A @ x + B.flatten() * alpha

theta_cam = np.array(theta_cam)
time      = np.arange(steps)

# END OF SCRIPT
```

A.2 Field-programmable coupler array

A.2.1 Overview of switching parameters

Here, I present the switching parameters referenced in Section 6.2.3. These are for crystallisation (Table A.3) and for amorphisation, both the smallest total energy demonstrated (Table A.1) and multi-level switching (Table A.2).

A.2.2 List of synthesised circuits with four unit cells

The various circuits we can synthesise with four unit cells connected (as shown in Fig. 6.15) include ring resonators, delay lines, Sagnac interferometers, and a

Table A.1: Overview of switching parameters and transmission levels used for complete amorphisation of the device with the smallest total switching energy, as shown in Fig. 6.5a (blue plot).

t (ns)	V (V)	Transmission level after pulse (%)
400	15.2	93
425	16.0	74
450	16.5	43
500	16.7	6.5
500	16.7	4.3
500	16.7	3.1
500	16.7	2.1
500	16.7	1.4
500	16.7	1.0
500	16.7	0.75
500	16.7	0.60
500	16.7	0.42
500	16.7	0.28

Table A.2: Overview of switching parameters used for multi-level amorphisation of the device shown in Fig. 6.5a (green plot).

t (ns)	V (V)	# of pulses
375	14.9	10
375	15.2	3
395	15.2	3
400	15.4	10
425	16.4	10
425	17.0	10
435	17.2	11
435	16.8	3

balanced MZI. Examples of some of the obtainable circuits are shown in Fig. A.1.

For ring resonators, we can achieve both single all-pass and add-drop rings, as well as coupled rings. For single rings (Fig. A.1b–d), in addition to the input coupling switch, we can have either three, seven, eleven, or 15 unit lengths in the ring. We can also achieve single- or double-column CROW structures with two rings per column and single- or double-row SCISSOR structures with two rings per row, as well as combinations of these (Fig. A.1e–g).

To the best of my knowledge, we can achieve delay lines from one to 15 unit lengths, as well as designs with 18 unit lengths (Fig. A.1h–k). For Sagnac

Table A.3: Overview of switching parameters used for crystallisation of ∼50 % of the PCM volume, as shown in Fig. 6.5b. The time durations t_1, t_2, and t_3 refer to the parameters in Eq. (6.5).

Event(s)	t_1 (µs)	t_2 (µs)	t_3 (µs)	V (V)	# of pulses
Bar → Int1	300	700	1200	7.1	6
	300	700	1200	7.5	1
Int1 → Int2	300	700	1700	7.5	1
	300	700	1700	7.7	1
	300	700	1700	8.1	1
Int2 → Int3	400	600	1700	7.5	1
	400	600	1700	8.1	1
	500	500	1700	8.3	1
Int3 →	450	550	1700	8.5	1
Max crystallisation	500	500	1700	8.5	1
	500	500	1700	8.7	1

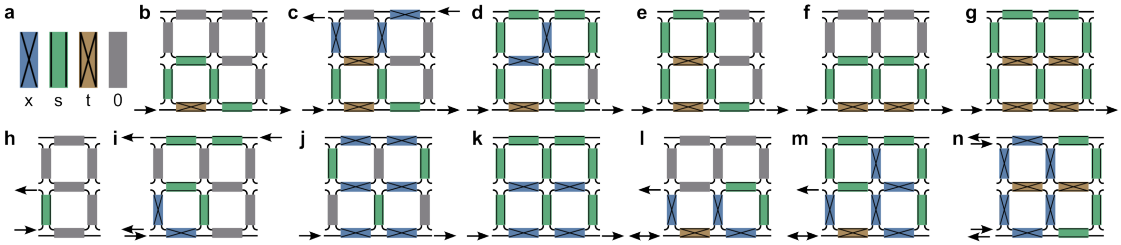

Figure A.1: Examples of synthesised circuits in a larger FPCA with four unit cells. **a**, Different switch states. x: cross; s: straight (bar); t: tunable; 0: inactive. **b**, An all-pass ring resonator. **c**, An add-drop ring resonator. **d**, An all-pass ring resonator with eight more unit lengths than in **b**. **e**, A coupled-resonator optical waveguide (CROW) with two rings. **f**, A side-coupled integrated spaced sequence of resonators (SCISSOR) with two rings. **g**, A combined CROW and SCISSOR structure with four rings. **h**, A delay line with a single unit length. **i**, Two different delay lines with two (four) unit lengths at the top (bottom). **j**,**k**, Delay lines with ten and 18 unit lengths. **l**,**m**, Sagnac interferometers with four and twelve unit lengths. **n**, A balanced MZI.

interferometers (Fig. A.1l,m), we can achieve a ring size of four, eight, or twelve unit lengths. Finally, we can also make a balanced MZI (Fig. A.1n), but not an unbalanced one.

A.3 Non-uniform neural network weight

A.3.1 More INTERCONNECT simulations of RAMZIs

Examples of more extensive INTERCONNECT simulations are shown in Fig. A.2. In the figure, I show several variations in the simulation settings, including the use

of Y-branches vs. DCs for splitting and combining light, the introduction of a phase difference in the MZI arms, and the use of add-drop rather than all-pass rings.

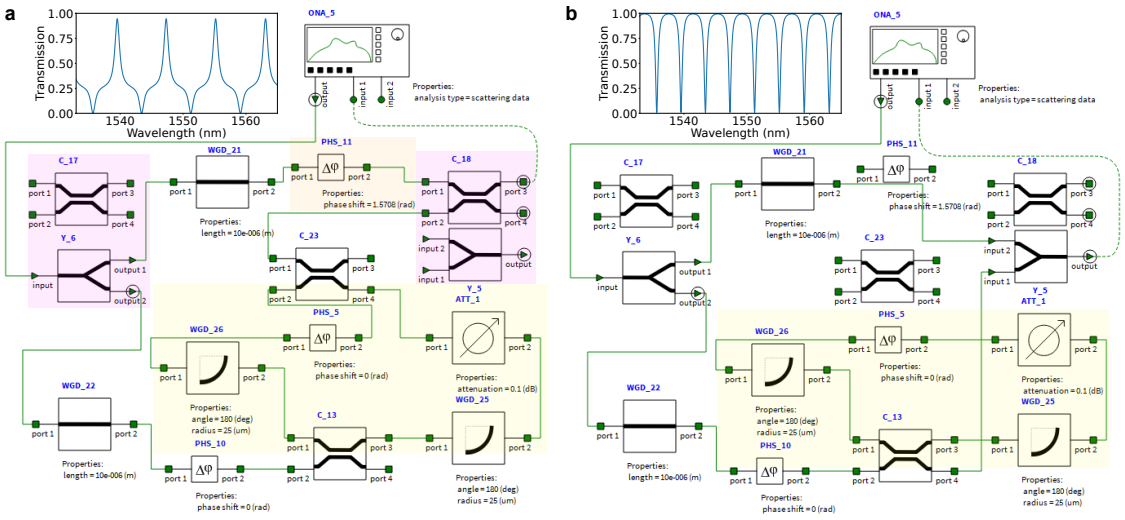

Figure A.2: More extensive INTERCONNECT simulations. **a**, INTERCONNECT simulation of a RAMZI with an add-drop ring and a $\pi/2$ phase shift in the other arm. **b**, INTERCONNECT simulation of a 1×1 RAMZI with an all-pass ring. The colours highlight the change of the devices in Fig. 7.1: purple for Y-branch vs. DC, orange for $\pi/2$ phase shift, and yellow for the ring.

A.3.2 More RAMZI transmission spectra

In Section 7.3, I comment on the fact that the output of the RAMZI presented is the mirror image of the analytical solution, which means that there is an additional π phase shift between the arms. In Fig. A.3, I show transmission spectra from different devices. Notably, all devices with a ring circumference of $C_{\text{ring}} = 400\,\mu\text{m}$ and a ring coupling length $L_{\text{c}} > 32\,\mu\text{m}$ have spectra with the curves being the mirror image of the expected behaviour.

A.3.3 Effect of reflection in the light path

In some of the devices, there is a reflection that arises from the short distance between the grating coupler and the Y-splitter, seen in Fig. A.4a,b for a device with ring circumference $C_{\text{ring}} = 250\,\mu\text{m}$ and $L_{\text{c}} = 18\,\mu\text{m}$. We included the Y-branch because it allows us to normalise to an internal reference port. The distance between

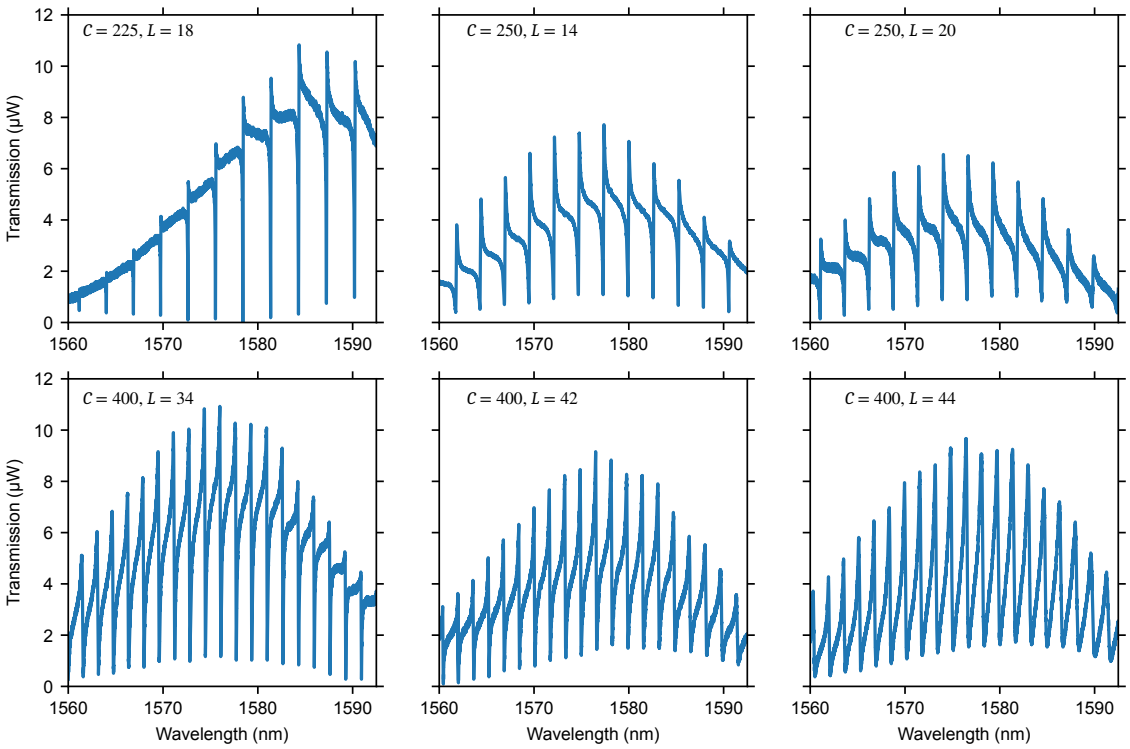

Figure A.3: Transmission spectra of RAMZI devices with the following ring circumference and coupling lengths: **a**, $C_{\text{ring}} = 225\,\mu\text{m}$ and $L_{\text{c}} = 18.0\,\mu\text{m}$ (corresponding to $k^2 \approx 0.02$); **b**, $C_{\text{ring}} = 250\,\mu\text{m}$, $L_{\text{c}} = 14.0\,\mu\text{m}$ ($k^2 \approx 0.38$); **c**, $C_{\text{ring}} = 250\,\mu\text{m}$, $L_{\text{c}} = 20.0\,\mu\text{m}$ ($k^2 \approx 0.01$); **d**, $C_{\text{ring}} = 400\,\mu\text{m}$, $L_{\text{c}} = 34.6\,\mu\text{m}$ ($k^2 \approx 0.82$); **e**, $C_{\text{ring}} = 400\,\mu\text{m}$, $L_{\text{c}} = 42.6\,\mu\text{m}$ ($k^2 \approx 0.01$); **f**, $C_{\text{ring}} = 400\,\mu\text{m}$, $L_{\text{c}} = 44.6\,\mu\text{m}$ ($k^2 \approx 0.03$).

the grating coupler and the Y-splitter is 17.5 µm, which leads to a wavelength-dependent sinusoidal envelope following $p = \frac{\lambda^2}{2Ln_{\text{eff}}}$. The calculated period of the envelope is approximately $p = 16.7\,\text{nm}$ for $\lambda = 1550\,\text{nm}$ and $n_{\text{eff}} = 4.1$. From Fig. A.4c, we see that the envelope has a similar period.

In Fig. A.4c, I show three curves in addition to the original transmission spectrum of the device in Fig. A.4a. The orange envelope is a Gaussian function, given as $G = Ae^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$, and the green and red curves are given by the equation

$$f(x) = G(1 + a\sin\left(\frac{2\pi x}{d} + \pi b\right) + c \tag{A.1}$$

where a, b, and c are fitting parameters and d is the period of the sine function. The red and green curves are fitted to the entire spectrum and the peaks, respectively. From these curve fits, we get that the periods of the two are $p = 16.87\,\text{nm}$ and 16.68 nm, respectively, in excellent agreement with the calculated period of the reflection.

Furthermore, I investigate the effect of this sinusoid on the observed shape of the RAMZI transmission (practically, the tangent curve fits). The curve in Fig. A.4d follows the analytical expression in Eq. (3.45) but with a sinusoidal modulation given as $1 + 0.5\sin(0.1x)$. The unmodulated analytical expression gives the same γ' and T_0 for all resonances fitted to Eq. (3.48), but, as we see from the fits in Fig. A.4d, the γ' and T_0 vary with the envelope.

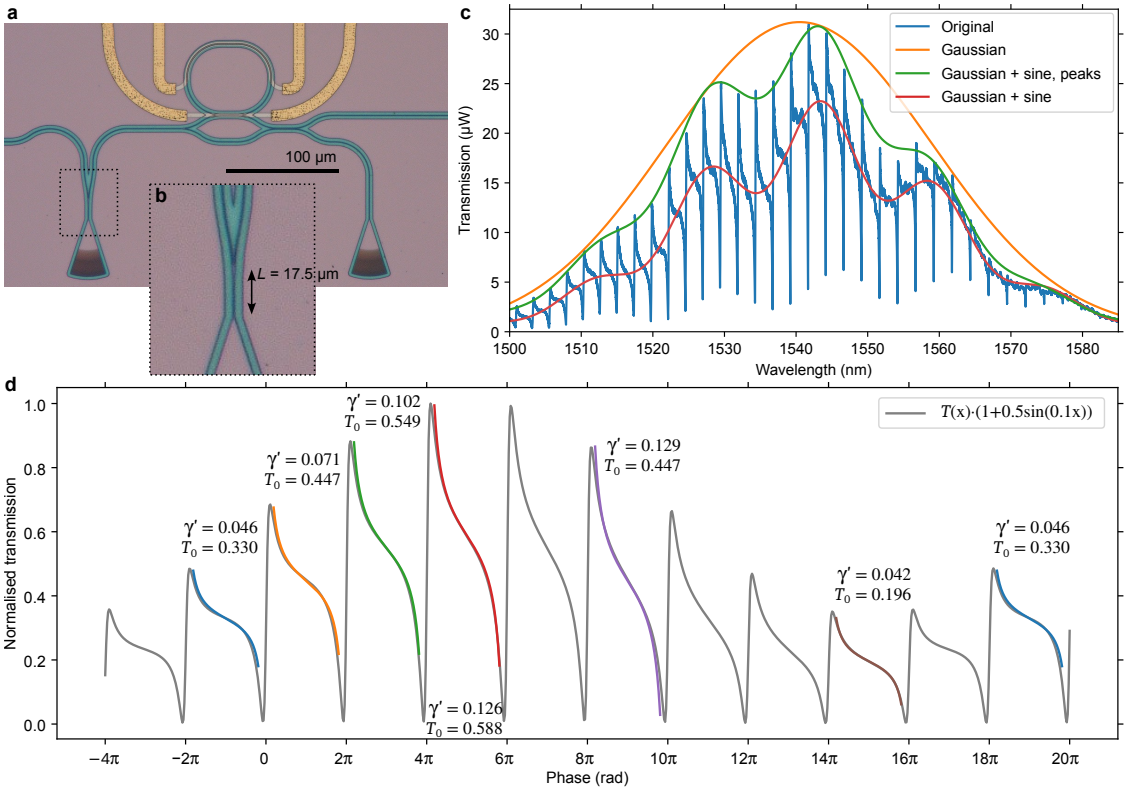

Figure A.4: Effect of reflection in the light path on the spectrum of a RAMZI. **a**, Optical microscope image of a device with ring circumference $C_{\text{ring}} = 250\,\mu\text{m}$ and $L_c = 18\,\mu\text{m}$, and a short distance between the grating coupler and the Y-branch to the reference port. **b**, Zoom-in from **a**. **c**, Transmission spectrum of the device in **a**, showing the Gaussian envelope and two curve fits, fitted to the peaks and entire spectrum, respectively. **d**, Calculated transmission from a RAMZI with a strong reflection in the light path. The fits (given by Eq. (3.48)) vary with the sinusoidal envelope and cycle over one period (see the blue fits).

Experimental transfer functions from the device with transmission presented in Fig. A.4c are shown in Fig. A.5. The experimental data backs up the analytical data.

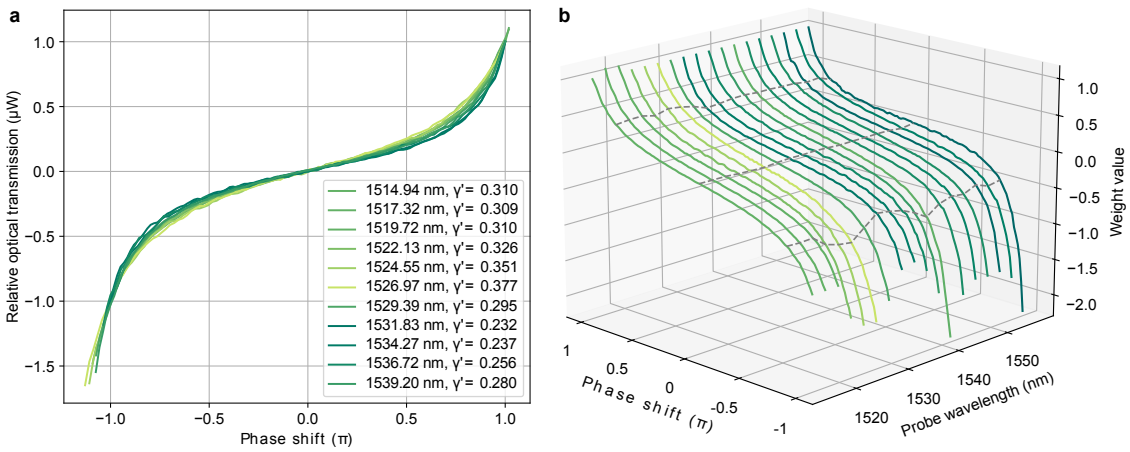

Figure A.5: Effect of reflection in the light path on the non-linear curves of a RAMZI. **a**,**b**, Normalised transmission vs. phase shift, shown in a 2-D (a) and 3-D plot (b).

References

[1] R.K. Ratnesh et al. 'Advancement and Challenges in MOSFET Scaling'. In: *Materials Science in Semiconductor Processing* 134 (Nov. 2021), p. 106002.

[2] Nicola Jones. 'How to Stop Data Centres from Gobbling up the World's Electricity'. In: *Nature* 561.7722 (Sept. 2018), pp. 163–166.

[3] Bhavin J. Shastri et al. 'Photonics for Artificial Intelligence and Neuromorphic Computing'. In: *Nature Photonics* 15.2 (Feb. 2021), pp. 102–114.

[4] J. Feldmann et al. 'Parallel Convolutional Processing Using an Integrated Photonic Tensor Core'. In: *Nature* 589 (Jan. 2021), pp. 52–58.

[5] Leonid Tsybeskov, David J. Lockwood and Masakazu Ichikawa. 'Silicon Photonics: CMOS Going Optical'. In: *Proceedings of the IEEE* 97.7 (July 2009), pp. 1161–1165.

[6] Hsuan-Tung Peng et al. 'Neuromorphic Photonic Integrated Circuits'. In: *IEEE Journal of Selected Topics in Quantum Electronics* 24.6 (Nov. 2018), pp. 1–15.

[7] Sufi R. Ahmed et al. 'Universal Photonic Artificial Intelligence Acceleration'. In: *Nature* 640.8058 (Apr. 2025), pp. 368–374.

[8] Shiyue Hua et al. 'An Integrated Large-Scale Photonic Accelerator with Ultralow Latency'. In: *Nature* 640.8058 (Apr. 2025), pp. 361–367.

[9] Joe Touch, Abdel-Hameed Badawy and Volker J. Sorger. 'Optical Computing'. In: *Nanophotonics* 6.3 (May 2017), pp. 503–505.

[10] Qihang Zhang et al. 'Broadband Nonvolatile Photonic Switching Based on Optical Phase Change Materials: Beyond the Classical Figure-of-Merit'. In: *Optics Letters* 43.1 (Jan. 2018), p. 94.

[11] Yichen Shen et al. 'Deep Learning with Coherent Nanophotonic Circuits'. In: *Nature Photonics* 11.7 (July 2017), pp. 441–446.

[12] Matthew Delaney et al. 'A New Family of Ultralow Loss Reversible Phase-Change Materials for Photonic Integrated Circuits: Sb_2S_3 and Sb_2Se_3'. In: *Advanced Functional Materials* 30.36 (2020), p. 2002447.

[13] Håvard Hem Toftevaag et al. 'Ultracompact Non-Volatile Field-Programmable Coupler Arrays'. In: *2025 Conference on Lasers and Electro-Optics (CLEO)*. May 2025, pp. 1–2.

[14] Junchao Song et al. 'Feasibility Study of a $Ge_2Sb_2Te_5$-Clad Silicon Waveguide as a Non-Volatile Optical on-off Switch'. In: *OSA Continuum* 2.1 (Jan. 2019), pp. 49–63.

[15] Zenghui Yu et al. 'Ultracompact Electro-Optical Modulator-Based $Ge_2Sb_2Te_5$ on Silicon'. In: *IEEE Photonics Technology Letters* 30.3 (Feb. 2018), pp. 250–253.

[16] Rui Chen et al. 'Broadband Nonvolatile Electrically Controlled Programmable Units in Silicon Photonics'. In: *ACS Photonics* 9.6 (June 2022), pp. 2142–2150.

[17] Shasha Liao et al. 'Integrated Optical Power Splitter With Continuously Adjustable Power Splitting Ratio'. In: *IEEE Photonics Journal* 12.6 (Dec. 2020), pp. 1–13.

[18] Dong Uk Kim et al. 'Programmable Photonic Arrays Based on Microelectromechanical Elements with Femtowatt-Level Standby Power Consumption'. In: *Nature Photonics* (Nov. 2023), pp. 1–8.

[19] Qixiang Cheng et al. 'Photonic Switching in High Performance Datacenters [Invited]'. In: *Optics Express* 26.12 (June 2018), pp. 16022–16043.

[20] Benjamin G. Lee and Nicolas Dupuis. 'Silicon Photonic Switch Fabrics: Technology and Architecture'. In: *Journal of Lightwave Technology* 37.1 (Jan. 2019), pp. 6–20.

[21] Xin Tu et al. 'State of the Art and Perspectives on Silicon Photonic Switches'. In: *Micromachines* 10.1 (Jan. 2019), p. 51.

[22] Richard Soref. 'Tutorial: Integrated-photonic Switching Structures'. In: *APL Photonics* 3.2 (Feb. 2018), p. 021101.

[23] Nathan Youngblood et al. 'Integrated Optical Memristors'. In: *Nature Photonics* (May 2023), pp. 1–12.

[24] Wencheng Yue, Yan Cai and Mingbin Yu. 'Review of 2×2 Silicon Photonic Switches'. In: *Photonics* 10.5 (May 2023), p. 564.

[25] Sajjad Abdollahramezani et al. 'Tunable Nanophotonics Enabled by Chalcogenide Phase-Change Materials'. In: *Nanophotonics* 9.5 (May 2020), pp. 1189–1241.

[26] Liangjun Lu et al. 'Low-Power 2×2 Silicon Electro-Optic Switches Based on Double-Ring Assisted Mach–Zehnder Interferometers'. In: *Optics Letters* 39.6 (Mar. 2014), p. 1633.

[27] Zheng Yong et al. 'Power-Efficient Silicon Nitride Thermo-Optic Phase Shifters for Visible Light'. In: *Optics Express* 30.5 (Feb. 2022), pp. 7225–7237.

[28] Dongdong Lin et al. 'Improving Performance of Silicon Thermo-Optic Switch by Combing Spiral Phase Shifter and Optimized Pulse Driving'. In: *IEEE Photonics Journal* 14.3 (June 2022), pp. 1–7.

[29] Hanke Feng et al. 'Integrated Lithium Niobate Microwave Photonic Processing Engine'. In: *Nature* 627.8002 (Mar. 2024), pp. 80–87.

[30] Ting Yang et al. 'Experimental Observation of Optical Differentiation and Optical Hilbert Transformation Using a Single SOI Microdisk Chip'. In: *Scientific Reports* 4.1 (Feb. 2014), p. 3960.

[31] W. Bogaerts et al. 'Silicon Microring Resonators'. In: *Laser & Photonics Reviews* 6.1 (Jan. 2012), pp. 47–73.

[32] Qianfan Xu, David Fattal and Raymond G. Beausoleil. 'Silicon Microring Resonators with 1.5-μm Radius'. In: *Optics Express* 16.6 (Mar. 2008), pp. 4309–4315.

[33] Tae Joon Seok et al. 'Highly Scalable Digital Silicon Photonic MEMS Switches'. In: *Journal of Lightwave Technology* 34.2 (Jan. 2016), pp. 365–371.

[34] Syotaro Abe et al. 'Time Response of a Microelectromechanical Silicon Photonic Waveguide Coupler Switch'. In: *IEEE Photonics Technology Letters* 26.15 (Aug. 2014), pp. 1553–1556.

[35] Sangyoon Han et al. 'Large-Scale Silicon Photonic Switches with Movable Directional Couplers'. In: *Optica* 2.4 (Apr. 2015), p. 370.

[36] Tae Joon Seok et al. 'Large-Scale Broadband Digital Silicon Photonic Switches with Vertical Adiabatic Couplers'. In: *Optica* 3.1 (Jan. 2016), p. 64.

[37] Peipeng Xu et al. 'Low-Loss and Broadband Nonvolatile Phase-Change Directional Coupler Switches'. In: *ACS Photonics* 6.2 (Feb. 2019), pp. 553–557.

[38] Ting Yu Teo et al. 'Comparison and Analysis of Phase Change Materials-Based Reconfigurable Silicon Photonic Directional Couplers'. In: *Optical Materials Express* 12.2 (Jan. 2022), pp. 606–621.

[39] Rui Chen et al. 'Non-Volatile Electrically Programmable Integrated Photonics with a 5-Bit Operation'. In: *Nature Communications* 14.1 (June 2023), p. 3465.

[40] Haibo Liang et al. 'Electro-Optical Phase-Change 2×2 Switching Using Three- and Four-Waveguide Directional Couplers'. In: *Applied Optics* 54.19 (July 2015), p. 5897.

[41] Daniel Pérez-López et al. 'Integrated Photonic Tunable Basic Units Using Dual-Drive Directional Couplers'. In: *Optics Express* 27.26 (Dec. 2019), pp. 38071–38086.

[42] Amged Alquliah et al. 'Reconfigurable Metasurface-Based 1×2 Waveguide Switch'. In: *Photonics Research* 9.10 (Oct. 2021), pp. 2104–2115.

[43] Wei Jia, Rajesh Menon and Berardi Sensale-Rodriguez. 'Unique Prospects of Phase Change Material Sb_2Se_3 for Ultra-Compact Reconfigurable Nanophotonic Devices'. In: *Optical Materials Express* 11.9 (Sept. 2021), pp. 3007–3014.

[44] Álvaro Rosa et al. 'High Performace Silicon 2×2 Optical Switch Based on a Thermo-Optically Tunable Multimode Interference Coupler and Efficient Electrodes'. In: *Optics Express* 24.1 (Jan. 2016), pp. 191–198.

[45] Mingxiao Li et al. 'Lithium Niobate Photonic-Crystal Electro-Optic Modulator'. In: *Nature Communications* 11.1 (Aug. 2020), p. 4123.

[46] Juan Zhang et al. 'Phase-Change Photonic Crystal Ring Resonator for Reconfigurable Directional-Coupler Switching'. In: *Photonics and Nanostructures - Fundamentals and Applications* 41 (Sept. 2020), p. 100798.

[47] Richard Soref and Joshua Hendrickson. 'Proposed Ultralow-Energy Dual Photonic-Crystal Nanobeam Devices for on-Chip N×N Switching, Logic, and Wavelength Multiplexing'. In: *Optics Express* 23.25 (Dec. 2015), p. 32582.

[48] Richard Soref. 'Resonant and Slow-Light 2×2 Switches Enabled by Nanobeams and Grating-Coupled Waveguides'. In: *2017 Progress In Electromagnetics Research Symposium - Spring (PIERS)*. May 2017, pp. 868–871.

[49] Giuseppe Cocorullo et al. 'Thermo-Optic Effect Exploitation in Silicon Microstructures'. In: *Sensors and Actuators A: Physical* 71.1-2 (Nov. 1998), pp. 19–26.

[50] Amir Arbabi and Lynford L. Goddard. 'Measurements of the Refractive Indices and Thermo-Optic Coefficients of Si_3N_4 and SiO_x Using Microring Resonances'. In: *Optics Letters* 38.19 (Oct. 2013), p. 3878.

[51] Michael R. Watts et al. 'Adiabatic Thermo-Optic Mach–Zehnder Switch'. In: *Optics Letters* 38.5 (Mar. 2013), p. 733.

[52] Joris Van Campenhout et al. 'Low-Power, 2×2 Silicon Electro-Optic Switch with 110-nm Bandwidth for Broadband Reconfigurable Optical Networks'. In: *Optics Express* 17.26 (Dec. 2009), p. 24020.

[53] Sheng Gan et al. 'A Highly Efficient Thermo-Optic Microring Modulator Assisted by Graphene'. In: *Nanoscale* 7.47 (Nov. 2015), pp. 20249–20255.

[54] Qi Li et al. 'Single Microring-Based 2×2 Silicon Photonic Crossbar Switches'. In: *IEEE Photonics Technology Letters* 27.18 (Sept. 2015), pp. 1981–1984.

[55] Zhipei Sun, Amos Martinez and Feng Wang. 'Optical Modulators with 2D Layered Materials'. In: *Nature Photonics* 10.4 (Apr. 2016), pp. 227–238.

[56] G. T. Reed et al. 'Silicon Optical Modulators'. In: *Nature Photonics* 4.8 (Aug. 2010), pp. 518–526.

[57] Graham T Reed and Andrew P Knights. *Silicon Photonics: An Introduction*. Chichester: John Wiley, Oct. 2004.

[58] Koen Alexander et al. 'A Manufacturable Platform for Photonic Quantum Computing'. In: *Nature* 641.8064 (May 2025), pp. 876–883.

[59] Chengli Wang et al. 'Lithium Tantalate Photonic Integrated Circuits for Volume Manufacturing'. In: *Nature* 629.8013 (May 2024), pp. 784–790.

[60] Anja Ulrich et al. 'Engineering High Pockels Coefficients in Thin-Film Strontium Titanate for Cryogenic Quantum Electro-Optic Applications'. In: *Science* 390.6771 (Oct. 2025), pp. 390–393.

[61] Georgios Sinatkas et al. 'Electro-Optic Modulation in Integrated Photonics'. In: *Journal of Applied Physics* 130.1 (July 2021), p. 010901.

[62] Wei-Ta Huang et al. 'Emerging Modulator Technologies in Silicon Photonics'. In: *IEEE Nanotechnology Magazine* 19.3 (June 2025), pp. 35–45.

[63] Jonathan Peltier et al. 'High-Speed Silicon Photonic Electro-Optic Kerr Modulation'. In: *Photonics Research* 12.1 (Jan. 2024), pp. 51–60.

[64] Mingbo He et al. 'High-Performance Hybrid Silicon and Lithium Niobate Mach–Zehnder Modulators for 100 Gbit s^{-1} and Beyond'. In: *Nature Photonics* 13.5 (May 2019), pp. 359–364.

[65] Yongqian Tang et al. 'High Performance Thin-Film Lithium Niobate Modulator on Silicon Substrate with a Thick Silica Buffer Layer'. In: *Optics Express* 33.10 (May 2025), pp. 20334–20344.

[66] Yunxiang Song et al. 'Integrated Electro-Optic Digital-to-Analogue Link for Efficient Computing and Arbitrary Waveform Generation'. In: *Nature Photonics* (Aug. 2025).

[67] Laurens Bogaert et al. 'SiGe EAM-Based Transceivers for Datacenter Interconnects and Radio Over Fiber'. In: *IEEE Journal of Selected Topics in Quantum Electronics* 27.3 (May 2021), pp. 1–13.

[68] C. H. Wu et al. 'Graphene Electro-Absorption Modulators Integrated at Wafer-Scale in a CMOS Fab'. In: *2021 Symposium on VLSI Circuits*. June 2021, pp. 1–2.

[69] Jochem Verbist et al. 'Real-Time and DSP-Free 128 Gb/s PAM-4 Link Using a Binary Driven Silicon Photonic Transmitter'. In: *Journal of Lightwave Technology* 37.2 (Jan. 2019), pp. 274–280.

[70] Dazeng Feng et al. 'High-Speed GeSi Electroabsorption Modulator on the SOI Waveguide Platform'. In: *IEEE Journal of Selected Topics in Quantum Electronics* 19.6 (Nov. 2013), pp. 64–73.

[71] Hannes Ramon et al. '70 Gb/s 0.87 pJ/Bit GeSi EAM Driver in 55 nm SiGe BiCMOS'. In: *2018 European Conference on Optical Communication (ECOC)*. Sept. 2018, pp. 1–3.

[72] P. De Heyn et al. 'High-Speed Germanium-Based Waveguide Electro-Absorption Modulator'. In: *2016 21st OptoElectronics and Communications Conference (OECC) Held Jointly with 2016 International Conference on Photonics in Switching (PS)*. July 2016, pp. 1–3.

[73] Takeshi Baba et al. '25-Gb/s Broadband Silicon Modulator with 0.31 V·cm $V_{\pi}L$ Based on Forward-Biased PIN Diodes Embedded with Passive Equalizer'. In: *Optics Express* 23.26 (Dec. 2015), pp. 32950–32960.

[74] Nicolas Dupuis et al. 'Nanosecond Photonic Switch Architectures Demonstrated in an All-Digital Monolithic Platform'. In: *Optics Letters* 44.15 (Aug. 2019), pp. 3610–3612.

[75] Wim Bogaerts et al. 'Programmable Photonic Circuits'. In: *Nature* 586.7828 (Oct. 2020), pp. 207–216.

[76] Liangjun Lu et al. '16×16 Non-Blocking Silicon Optical Switch Based on Electro-Optic Mach-Zehnder Interferometers'. In: *Optics Express* 24.9 (May 2016), pp. 9295–9307.

[77] Po Dong et al. 'Submilliwatt, Ultrafast and Broadband Electro-Optic Silicon Switches'. In: *Optics Express* 18.24 (Nov. 2010), pp. 25225–25231.

[78] Nicolas Dupuis et al. 'Design and Fabrication of Low-Insertion-Loss and Low-Crosstalk Broadband 2×2 Mach–Zehnder Silicon Photonic Switches'. In: *Journal of Lightwave Technology* 33.17 (Sept. 2015), pp. 3597–3606.

[79] Carlos Errando-Herranz et al. 'MEMS for Photonic Integrated Circuits'. In: *IEEE Journal of Selected Topics in Quantum Electronics* 26.2 (Mar. 2020), pp. 1–16.

[80] E. Ollier. 'Optical MEMS Devices Based on Moving Waveguides'. In: *IEEE Journal of Selected Topics in Quantum Electronics* 8.1 (Jan. 2002), pp. 155–162.

[81] Rohit Chatterjee and Chee Wei Wong. 'Nanomechanical Proximity Perturbation for Switching in Silicon-Based Directional Couplers for High-Density Photonic Integrated Circuits'. In: *Journal of Microelectromechanical Systems* 19.3 (June 2010), pp. 657–662.

[82] Jonathan Briere et al. 'Rotating Circular Micro-Platform with Integrated Waveguides and Latching Arm for Reconfigurable Integrated Optics'. In: *Micromachines* 8.12 (Dec. 2017), p. 354.

[83] Pierre Edinger et al. 'Silicon Photonic Microelectromechanical Phase Shifters for Scalable Programmable Photonics'. In: *Optics Letters* 46.22 (Nov. 2021), p. 5671.

[84] Pierre Edinger et al. 'A Bistable Silicon Photonic Mems Phase Switch For Nonvolatile Photonic Circuits'. In: *2022 IEEE 35th International Conference on Micro Electro Mechanical Systems Conference (MEMS)*. Jan. 2022, pp. 995–997.

[85] Samuel Gyger et al. 'Reconfigurable Photonics with On-Chip Single-Photon Detectors'. In: *Nature Communications* 12.1 (Mar. 2021), p. 1408.

[86] Syotaro Abe and Kazuhiro Hane. 'A Silicon Microring Resonator with a Nanolatch Mechanism'. In: *Microsystem Technologies* 21.9 (Sept. 2015), pp. 2019–2024.

[87] Syotaro Abe and Kazuhiro Hane. 'Variable-Gap Silicon Photonic Waveguide Coupler Switch With a Nanolatch Mechanism'. In: *IEEE Photonics Technology Letters* 25.7 (Apr. 2013), pp. 675–677.

[88] Ana Gonzalez and Diogo Costa. 'Programmable Chips for 5G Applications: The New Generation of Green and Smart Hardware for 5G'. In: *PhotonicsViews* 19.1 (Feb. 2022), pp. 38–41.

[89] Franck Chollet. 'Devices Based on Co-Integrated MEMS Actuators and Optical Waveguide: A Review'. In: *Micromachines* 7.2 (Feb. 2016), p. 18.

[90] Han Du, Fook Siong Chau and Guangya Zhou. 'Mechanically-Tunable Photonic Devices with On-Chip Integrated MEMS/NEMS Actuators'. In: *Micromachines* 7.4 (Apr. 2016), p. 69.

[91] Yifei Zhang et al. 'Myths and Truths about Optical Phase Change Materials: A Perspective'. In: *Applied Physics Letters* 118.21 (May 2021), p. 210501.

[92] Yi-Siou Huang et al. 'Optical Phase Change Materials'. In: *Annual Review of Materials Research* 55.1 (July 2025), pp. 255–283.

[93] Yoonjin Won et al. 'Phase and Thickness Dependent Modulus of $Ge_2Sb_2Te_5$ Films down to 25 nm Thickness'. In: *Applied Physics Letters* 100.16 (Apr. 2012), p. 161905.

[94] Matthias Wuttig. 'Towards a Universal Memory?' In: *Nature Materials* 4.4 (Apr. 2005), pp. 265–266.

[95] Matthias Wuttig and Noboru Yamada. 'Phase-Change Materials for Rewriteable Data Storage'. In: *Nature Materials* 6.11 (Nov. 2007), pp. 824–832.

[96] Peiman Hosseini, C. David Wright and Harish Bhaskaran. 'An Optoelectronic Framework Enabled by Low-Dimensional Phase-Change Films'. In: *Nature* 511.7508 (July 2014), pp. 206–211.

[97] Nathan Youngblood et al. 'Reconfigurable Low-Emissivity Optical Coating Using Ultrathin Phase Change Materials'. In: *ACS Photonics* 9.1 (Jan. 2022), pp. 90–100.

[98] Abu Sebastian, Manuel Le Gallo and Evangelos Eleftheriou. 'Computational Phase-Change Memory: Beyond von Neumann Computing'. In: *Journal of Physics D: Applied Physics* 52.44 (Aug. 2019), p. 443002.

[99] Carlos Ríos et al. 'In-Memory Computing on a Photonic Platform'. In: *Science Advances* 5.2 (Feb. 2019), eaau5759.

[100] Jacqueline Geler-Kremer et al. 'A Ferroelectric Multilevel Non-Volatile Photonic Phase Shifter'. In: *Nature Photonics* (May 2022), pp. 1–7.

[101] Paolo Pintus et al. 'An Integrated Magneto-Optic Modulator for Cryogenic Applications'. In: *Nature Electronics* 5.9 (Sept. 2022), pp. 604–610.

[102] Marijn P. G. Rombouts et al. 'A Sub-Picojoule per Bit Integrated Magneto-Optic Modulator on Silicon: Modeling and Experimental Demonstration'. In: *Laser & Photonics Reviews* 17.4 (2023), p. 2200799.

[103] Paolo Pintus et al. 'Integrated Non-Reciprocal Magneto-Optics with Ultra-High Endurance for Photonic in-Memory Computing'. In: *Nature Photonics* 19.1 (Jan. 2025), pp. 54–62.

[104] Haowen Shu et al. 'Significantly High Modulation Efficiency of Compact Graphene Modulator Based on Silicon Waveguide'. In: *Scientific Reports* 8.1 (Jan. 2018), p. 991.

[105] Hao Zhang et al. 'Highly Efficient Single-Layer Graphene Electro-Absorption Modulator'. In: *Optics Letters* 49.5 (Mar. 2024), pp. 1337–1340.

[106] Ipshita Datta et al. 'Low-Loss Composite Photonic Platform Based on 2D Semiconductor Monolayers'. In: *Nature Photonics* 14.4 (Apr. 2020), pp. 256–262.

[107] Haitao Chen et al. 'Waveguide-Integrated Optical Modulators with Two-Dimensional Materials'. In: *Journal of Semiconductors* 44.11 (Nov. 2023), p. 111301.

[108] Lei Wan et al. 'Highly Efficient Acousto-Optic Modulation Using Nonsuspended Thin-Film Lithium Niobate-Chalcogenide Hybrid Waveguides'. In: *Light: Science & Applications* 11.1 (May 2022), p. 145.

[109] Jayita Dutta et al. 'Near-Visible Low-Power Tuning of Nematic-Liquid Crystal Integrated Silicon Nitride Ring Resonator'. In: *ACS Photonics* 12.9 (Sept. 2025), pp. 4967–4980.

[110] Jin Tae Kim, Juhee Song and Chil Seong Ah. 'Optically Readable Waveguide-Integrated Electrochromic Artificial Synaptic Device for Photonic Neuromorphic Systems'. In: *ACS Applied Electronic Materials* 2.7 (July 2020), pp. 2057–2063.

[111] Bassem Tossoun et al. 'High-Speed and Energy-Efficient Non-Volatile Silicon Photonic Memory Based on Heterogeneously Integrated Memresonator'. In: *Nature Communications* 15.1 (Jan. 2024), p. 551.

[112] Zhen Chai et al. 'Ultrafast All-Optical Switching'. In: *Advanced Optical Materials* 5.7 (2017), p. 1600665.

[113] Kengo Nozaki et al. 'Sub-Femtojoule All-Optical Switching Using a Photonic-Crystal Nanocavity'. In: *Nature Photonics* 4.7 (July 2010), pp. 477–483.

[114] Patinharekandy Prabhathan et al. 'Roadmap for Phase Change Materials in Photonics and Beyond'. In: *iScience* 26.10 (Oct. 2023), p. 107946.

[115] M. Wuttig, H. Bhaskaran and T. Taubner. 'Phase-Change Materials for Non-Volatile Photonic Applications'. In: *Nature Photonics* 11.8 (Aug. 2017), pp. 465–476.

[116] Nikolaos Farmakidis et al. 'Plasmonic Nanogap Enhanced Phase-Change Devices with Dual Electrical-Optical Functionality'. In: *Science Advances* 5.11 (Nov. 2019), eaaw2687.

[117] Utku Emre Ali. 'Nanomechanical Resonators Using Functional Nanowires'. PhD thesis. University of Oxford, 2022.

[118] D. Loke et al. 'Breaking the Speed Limits of Phase-Change Memory'. In: *Science* 336.6088 (June 2012), pp. 1566–1569.

[119] Simone Raoux et al. 'Phase Change Materials and Phase Change Memory'. In: *MRS Bulletin* 39.8 (Aug. 2014), pp. 703–710.

[120] Carlos Ríos et al. 'Ultra-Compact Nonvolatile Phase Shifter Based on Electrically Reprogrammable Transparent Phase Change Materials'. In: *PhotoniX* 3.1 (Oct. 2022), p. 26.

[121] S. Lai and T. Lowrey. 'OUM - A 180 nm Nonvolatile Memory Cell Element Technology for Stand Alone and Embedded Applications'. In: *International Electron Devices Meeting. Technical Digest (Cat. No.01CH37224)*. Dec. 2001, pp. 36.5.1–36.5.4.

[122] I.S. Kim et al. 'High Performance PRAM Cell Scalable to Sub-20 nm Technology with below 4F2 Cell Size, Extendable to DRAM Applications'. In: *2010 Symposium on VLSI Technology*. June 2010, pp. 203–204.

[123] Daniel Lawson et al. 'Optical Switching beyond a Million Cycles of Low-Loss Phase Change Material Sb_2Se_3'. In: *Optical Materials Express* 14.1 (Jan. 2024), pp. 22–38.

[124] Geoffrey W. Burr et al. 'Phase Change Memory Technology'. In: *Journal of Vacuum Science & Technology B* 28.2 (Mar. 2010), pp. 223–262.

[125] Weiling Dong et al. 'Wide Bandgap Phase Change Material Tuned Visible Photonics'. In: *Advanced Functional Materials* 29.6 (2019), p. 1806181.

[126] Max Birkett et al. 'Band Gap Temperature-Dependence of Close-Space Sublimation Grown Sb_2Se_3 by Photo-Reflectance'. In: *APL Materials* 6.8 (May 2018), p. 084901.

[127] Chao Chen et al. 'Optical Properties of Amorphous and Polycrystalline Sb_2Se_3 Thin Films Prepared by Thermal Evaporation'. In: *Applied Physics Letters* 107.4 (July 2015), p. 043905.

[128] Rui Chen et al. 'Opportunities and Challenges for Large-Scale Phase-Change Material Integrated Electro-Photonics'. In: *ACS Photonics* 9.10 (Oct. 2022), pp. 3181–3195.

[129] Wim Bogaerts. *Programmable Photonics*. Mar. 2024. URL: https://www.optica-opn.org/home/articles/volume_35/march_2024/features/programmable_photonics/.

[130] Zhuoran Fang et al. 'Non-Volatile Reconfigurable Integrated Photonics Enabled by Broadband Low-Loss Phase Change Material'. In: *Advanced Optical Materials* 9.9 (Mar. 2021), p. 2002049.

[131] Xing Yang et al. 'Non-Volatile Optical Switch Element Enabled by Low-Loss Phase Change Material'. In: *Advanced Functional Materials* 33.42 (June 2023), p. 2304601.

[132] Jingzhe Pang et al. 'Nonvolatile Continuously Tunable Integrated Optical Router'. In: *Laser & Photonics Reviews* (July 2025), e00281.

[133] Chao Chen et al. 'Characterization of Basic Physical Properties of Sb_2Se_3 and Its Relevance for Photovoltaics'. In: *Frontiers of Optoelectronics* 10.1 (Mar. 2017), pp. 18–30.

[134] Julian E. Spallholz. 'On the Nature of Selenium Toxicity and Carcinostatic Activity'. In: *Free Radical Biology and Medicine* 17.1 (July 1994), pp. 45–64.

[135] Xing Yang et al. 'Phase Change Material Enabled 2×2 Silicon Nonvolatile Optical Switch'. In: *Optics Letters* 46.17 (Sept. 2021), p. 4224.

[136] Xiaojun Chen, Jiao Lin and Ke Wang. 'A Review of Integrated Photonic Devices Using Sb_2Se_3'. In: *Advanced Physics Research* 4.1 (2025), p. 2400080.

[137] Joaquin Faneca et al. 'Towards Low Loss Non-Volatile Phase Change Materials in Mid Index Waveguides'. In: *Neuromorphic Computing and Engineering* 1.1 (Aug. 2021), p. 014004.

[138] Kunhao Lei et al. 'Magnetron-Sputtered and Thermal-Evaporated Low-Loss Sb-Se Phase-Change Films in Non-Volatile Integrated Photonics'. In: *Optical Materials Express* 12.7 (July 2022), pp. 2815–2823.

[139] Zhuoran Fang et al. 'Ultra-Low-Energy Programmable Non-Volatile Silicon Photonics Based on Phase-Change Materials with Graphene Heaters'. In: *Nature Nanotechnology* 17.8 (Aug. 2022), pp. 842–848.

[140] Carlos Ríos et al. 'Multi-Level Electro-Thermal Switching of Optical Phase-Change Materials Using Graphene'. In: *Advanced Photonics Research* 2.1 (2021), p. 2000034.

[141] Shilong Pan and Yamei Zhang. 'Microwave Photonic Radars'. In: *Journal of Lightwave Technology* 38.19 (Oct. 2020), pp. 5450–5484.

[142] Neetu Kanda et al. 'Radiation Hardness of $Ge_2Sb_2Te_5$ Thin Films to 80 MeV Si Ion Irradiation'. In: *Radiation Effects and Defects in Solids* 0.0 (Sept. 2021), pp. 1–10.

[143] Konstantinos Konstantinou et al. 'Origin of Radiation Tolerance in Amorphous $Ge_2Sb_2Te_5$ Phase-Change Random-Access Memory Material'. In: *Proceedings of the National Academy of Sciences* 115.21 (May 2018), pp. 5353–5358.

[144] A. L. Serra et al. 'Heavy Ions Radiation Effects on 4kb Phase-Change Memory'. In: *2020 20th European Conference on Radiation and Its Effects on Components and Systems (RADECS)*. Toulouse, France: IEEE, Oct. 2020, pp. 1–4.

[145] Giulio Terrasanta et al. 'Photonic Integrated Circuits for Optical Satellite Links: A Review of the Technology Status and Space Effects'. In: *International Journal of Satellite Communications and Networking* 43.3 (2025), pp. 210–228.

[146] Robert Gleixner. 'PCM Main Reliability Features'. In: *Phase Change Memory*. Ed. by Andrea Redaelli. Cham: Springer International Publishing, 2018, pp. 89–124.

[147] Diksha Thakur and Vir Singh Rangra. 'Challenges in Phase-Change Memory: A Focus on GST and In_2Se_3 PCM Materials'. In: *physica status solidi (b)* 262.9 (2025), p. 2400646.

[148] Xuan Li et al. 'Fast and Reliable Storage Using a 5 Bit, Nonvolatile Photonic Memory Cell'. In: *Optica* 6.1 (Jan. 2019), p. 1.

[149] David A. B. Miller. 'Attojoule Optoelectronics for Low-Energy Information Processing and Communications'. In: *Journal of Lightwave Technology* 35.3 (Feb. 2017), pp. 346–396.

[150] Erik J. Norberg et al. 'Programmable Photonic Microwave Filters Monolithically Integrated in InP–InGaAsP'. In: *Journal of Lightwave Technology* 29.11 (June 2011), pp. 1611–1619.

[151] Min Teng et al. 'Miniaturized Silicon Photonics Devices for Integrated Optical Signal Processors'. In: *Journal of Lightwave Technology* 38.1 (Jan. 2020), pp. 6–17.

[152] Peter J. Winzer, David T. Neilson and Andrew R. Chraplyvy. 'Fiber-Optic Transmission and Networking: The Previous 20 and the next 20 Years [Invited]'. In: *Optics Express* 26.18 (Sept. 2018), pp. 24190–24239.

[153] Steven Gringeri et al. 'Flexible Architectures for Optical Transport Nodes and Networks'. In: *IEEE Communications Magazine* 48.7 (July 2010), pp. 40–50.

[154] Martijn J. R. Heck. 'Highly Integrated Optical Phased Arrays: Photonic Integrated Circuits for Optical Beam Shaping and Beam Steering'. In: *Nanophotonics* 6.1 (Jan. 2017), pp. 93–107.

[155] David Marpaung, Jianping Yao and José Capmany. 'Integrated Microwave Photonics'. In: *Nature Photonics* 13.2 (Feb. 2019), pp. 80–90.

[156] Andrew J. Stark et al. 'Photonics for Electronic Warfare'. In: *2014 IEEE Avionics, Fiber-Optics and Photonics Technology Conference (AVFOP)*. Nov. 2014, pp. 3–4.

[157] Yu Li et al. 'Silicon and Hybrid Silicon Photonic Devices for Intra-Datacenter Applications: State of the Art and Perspectives [Invited]'. In: *Photonics Research* 3.5 (Oct. 2015), B10–B27.

[158] Andrea Annoni et al. 'Unscrambling Light—Automatically Undoing Strong Mixing between Modes'. In: *Light: Science & Applications* 6.12 (Dec. 2017), e17110–e17110.

[159] H. Aghaee Rad et al. 'Scaling and Networking a Modular Photonic Quantum Computer'. In: *Nature* 638.8052 (Feb. 2025), pp. 912–919.

[160] Jacques Carolan et al. 'Universal Linear Optics'. In: *Science* 349.6249 (Aug. 2015), pp. 711–716.

[161] Jacques Carolan et al. 'Variational Quantum Unsampling on a Quantum Photonic Processor'. In: *Nature Physics* 16.3 (Mar. 2020), pp. 322–327.

[162] Mark Dong et al. 'Programmable Photonic Integrated Meshes for Modular Generation of Optical Entanglement Links'. In: *npj Quantum Information* 9.1 (Apr. 2023), p. 42.

[163] Michele Spagnolo et al. 'Experimental Photonic Quantum Memristor'. In: *Nature Photonics* 16.4 (Apr. 2022), pp. 318–323.

[164] Gabriel J. Mendoza et al. 'Active Temporal and Spatial Multiplexing of Photons'. In: *Optica* 3.2 (Feb. 2016), pp. 127–132.

[165] Francesco Lenzini et al. 'Active Demultiplexing of Single Photons from a Solid-State Source'. In: *Laser & Photonics Reviews* 11.3 (2017), p. 1600297.

[166] Leimeng Zhuang et al. 'Programmable Photonic Signal Processor Chip for Radiofrequency Applications'. In: *Optica* 2.10 (Oct. 2015), pp. 854–859.

[167] Daniel Pérez et al. 'Reconfigurable Lattice Mesh Designs for Programmable Photonic Processors'. In: *Optics Express* 24.11 (May 2016), p. 12093.

[168] Daniel Pérez-López, Ana Gutiérrez and José Capmany. 'Silicon Nitride Programmable Photonic Processor with Folded Heaters'. In: *Optics Express* 29.6 (Mar. 2021), pp. 9043–9059.

[169] Daniel Pérez et al. 'Multipurpose Silicon Photonics Signal Processor Core'. In: *Nature Communications* 8.1 (Dec. 2017), p. 636.

[170] Daniel Pérez-López et al. 'General-Purpose Programmable Photonic Processor for Advanced Radiofrequency Applications'. In: *Nature Communications* 15.1 (Feb. 2024), p. 1563.

[171] Daniel Pérez, Ivana Gasulla and Jose Capmany. 'Field-Programmable Photonic Arrays'. In: *Optics Express* 26.21 (Oct. 2018), pp. 27265–27278.

[172] D. Pérez-López et al. 'Towards Field-Programmable Photonic Gate Arrays'. In: *Smart Photonic and Optoelectronic Integrated Circuits XXII*. Vol. 11284. SPIE, Feb. 2020, pp. 20–29.

[173] Weilin Liu et al. 'A Fully Reconfigurable Photonic Integrated Signal Processor'. In: *Nature Photonics* 10.3 (Mar. 2016), pp. 190–195.

[174] Weifeng Zhang and Jianping Yao. 'Photonic Integrated Field-Programmable Disk Array Signal Processor'. In: *Nature Communications* 11.1 (Jan. 2020), p. 406.

[175] Yaohui Sun et al. 'Field-Programmable Ring Array Employing AMZI-assisted-MRR Structure for Photonic Signal Processor'. In: *APL Photonics* 9.6 (June 2024), p. 066104.

[176] Changming Wu et al. 'Freeform Direct-Write and Rewritable Photonic Integrated Circuits in Phase-Change Thin Films'. In: *Science Advances* 10.1 (Jan. 2024), eadk1361.

[177] Forrest Miller et al. 'Rewritable Photonic Integrated Circuits Using Dielectric-Assisted Phase-Change Material Waveguides'. In: *Optics Letters* 48.9 (May 2023), pp. 2385–2388.

[178] Alexander N. Tait et al. 'Neuromorphic Photonic Networks Using Silicon Photonic Weight Banks'. In: *Scientific Reports* 7.1 (Aug. 2017), p. 7430.

[179] Tingzhao Fu et al. 'Photonic Machine Learning with On-Chip Diffractive Optics'. In: *Nature Communications* 14.1 (Jan. 2023), p. 70.

[180] Xiangyan Meng et al. 'Compact Optical Convolution Processing Unit Based on Multimode Interference'. In: *Nature Communications* 14.1 (May 2023), p. 3000.

[181] J. Feldmann et al. 'All-Optical Spiking Neurosynaptic Networks with Self-Learning Capabilities'. In: *Nature* 569.7755 (May 2019), pp. 208–214.

[182] Kristof Vandoorne et al. 'Experimental Demonstration of Reservoir Computing on a Silicon Photonics Chip'. In: *Nature Communications* 5.1 (May 2014), p. 3541.

[183] Peter L. McMahon et al. 'A Fully Programmable 100-Spin Coherent Ising Machine with All-to-All Connections'. In: *Science* 354.6312 (Nov. 2016), pp. 614–617.

[184] Nikolaos Farmakidis, Bowei Dong and Harish Bhaskaran. 'Integrated Photonic Neuromorphic Computing: Opportunities and Challenges'. In: *Nature Reviews Electrical Engineering* 1.6 (June 2024), pp. 358–373.

[185] Abu Sebastian et al. 'Memory Devices and Applications for In-Memory Computing'. In: *Nature Nanotechnology* 15.7 (July 2020), pp. 529–544.

[186] Samarth Aggarwal et al. 'Reduced Rank Photonic Computing Accelerator'. In: *Optica* 10.8 (Aug. 2023), pp. 1074–1080.

[187] Frank Brückerhoff-Plückelmann et al. 'Broadband Photonic Tensor Core with Integrated Ultra-Low Crosstalk Wavelength Multiplexers'. In: *Nanophotonics* (Feb. 2022).

[188] Bowei Dong et al. 'Higher-Dimensional Processing Using a Photonic Tensor Core with Continuous-Time Data'. In: *Nature Photonics* 17.12 (Dec. 2023), pp. 1080–1088.

[189] Shuhei Ohno et al. 'Si Microring Resonator Crossbar Array for On-Chip Inference and Training of the Optical Neural Network'. In: *ACS Photonics* 9.8 (Aug. 2022), pp. 2614–2622.

[190] Bowei Dong et al. 'Partial Coherence Enhances Parallelized Photonic Computing'. In: *Nature* 632.8023 (Aug. 2024), pp. 55–62.

[191] Shupeng Ning et al. 'Hardware-Efficient Photonic Tensor Core: Accelerating Deep Neural Networks with Structured Compression'. In: *Optica* 12.7 (July 2025), p. 1079.

[192] Shengping Liu et al. 'Thermo-Optic Phase Shifters Based on Silicon-on-Insulator Platform: State-of-the-Art and a Review'. In: *Frontiers of Optoelectronics* 15.1 (Apr. 2022), p. 9.

[193] Alex Malkevich. *Introducing NVFP4 for Efficient and Accurate Low-Precision Inference*. June 2025. URL: `https://developer.nvidia.com/blog/introducing-nvfp4-for-efficient-and-accurate-low-precision-inference/`.

[194] Naigang Wang et al. *Training Deep Neural Networks with 8-Bit Floating Point Numbers*. Dec. 2018. arXiv: `1812.08011 [cs]`.

[195] Wen Zhou et al. 'In-Memory Photonic Dot-Product Engine with Electrically Programmable Weight Banks'. In: *Nature Communications* 14.1 (May 2023), p. 2887.

[196] Apostolos Tsakyridis et al. 'Photonic Neural Networks and Optics-Informed Deep Learning Fundamentals'. In: *APL Photonics* 9.1 (Jan. 2024), p. 011102.

[197] Hanyu Zhang et al. 'Ultracompact Si-GST Hybrid Waveguides for Nonvolatile Light Wave Manipulation'. In: *IEEE Photonics Journal* 10.1 (Feb. 2018), pp. 1–10.

[198] Minh A. Tran et al. 'A Robust Method for Characterization of Optical Waveguides and Couplers'. In: *IEEE Photonics Technology Letters* 28.14 (July 2016), pp. 1517–1520.

[199] Changming Wu et al. 'Low-Loss Integrated Photonic Switch Using Subwavelength Patterned Phase Change Material'. In: *ACS Photonics* 6.1 (Jan. 2019), pp. 87–92.

[200] Chenran Ye et al. 'A Compact Plasmonic MOS-based 2×2 Electro-Optic Switch'. In: *Nanophotonics* 4.3 (Jan. 2015), pp. 261–268.

[201] Sitao Chen et al. 'Low-Loss and Broadband 2×2 Silicon Thermo-Optic Mach–Zehnder Switch with Bent Directional Couplers'. In: *Optics Letters* 41.4 (Feb. 2016), p. 836.

[202] Zhuoran Fang et al. 'Low-Loss Broadband Nonvolatile 2×2 Switch Based on Sb_2Se_3 for Programmable Silicon Photonics'. In: *Conference on Lasers and Electro-Optics*. San Jose, California: Optica Publishing Group, 2022, STh4G.6.

[203] Amnon Yariv and Pochi Yeh. *Photonics: Optical Electronics in Modern Communications*. 6. edition. The Oxford Series in Electrical and Computer Engineering. New York, NY: Oxford University Press, 2007.

[204] Choon Kong Lai et al. 'Methodology for Fabrication-Tolerant Planar Directional Couplers'. In: *IEEE Photonics Journal* 14.3 (June 2022), pp. 1–9.

[205] Fangfei Liu et al. 'Compact Optical Temporal Differentiator Based on Silicon Microring Resonator'. In: *Optics Express* 16.20 (Sept. 2008), pp. 15880–15886.

[206] S. Darmawan, Y. M. Landobasa and M. K. Chin. 'Phase Engineering for Ring Enhanced Mach-Zehnder Interferometers'. In: *Optics Express* 13.12 (June 2005), pp. 4580–4588.

[207] Mukesh Yadav et al. 'Spectral Shaping of Ring Resonator Transmission Response'. In: *Optics Express* 29.3 (Feb. 2021), pp. 3764–3771.

[208] Jinan Nijem et al. 'High-Q and High Finesse Silicon Microring Resonator'. In: *Optics Express* 32.5 (Feb. 2024), p. 7896.

[209] Zhe Han et al. 'Design of Ultrahigh-Q Silicon Microring Resonators Based on Free-Form Curves'. In: *Optics Express* 32.6 (Mar. 2024), p. 9553.

[210] Maximilian J. Müller et al. 'Tailoring Crystallization Kinetics of Chalcogenides for Photonic Applications'. In: *Advanced Electronic Materials* 8.8 (2022), p. 2100974.

[211] Chenzhuo Zhu et al. *Trained Ternary Quantization*. Feb. 2017. arXiv: `1612.01064 [cs]`.

[212] Weizhi Lu et al. *Ternary and Binary Quantization for Improved Classification*. Mar. 2022. arXiv: `2203.16798 [cs]`.

[213] Yuhang Li, Xin Dong and Wei Wang. *Additive Powers-of-Two Quantization: An Efficient Non-uniform Discretization for Neural Networks*. Feb. 2020. arXiv: `1909.13144 [cs]`.

[214] Jipeng Li, Xueqiong Yuan and Ercan Engin Kuruoglu. 'Exploring Weight Distributions and Dependence in Neural Networks With α-Stable Distributions'. In: *IEEE Transactions on Artificial Intelligence* 5.11 (Nov. 2024), pp. 5519–5529.

[215] Wei Huang et al. *BiLLM: Pushing the Limit of Post-Training Quantization for LLMs*. May 2024. arXiv: `2402.04291 [cs]`.

[216] Pengxiang Zhao and Xiaoming Yuan. *GANQ: GPU-Adaptive Non-Uniform Quantization for Large Language Models*. June 2025. arXiv: `2501.12956 [cs]`.

[217] Ye Li, Bertan Bakkaloglu and Chaitali Chakrabarti. 'A System Level Energy Model and Energy-Quality Evaluation for Integrated Transceiver Front-Ends'. In: *IEEE Transactions on Very Large Scale Integration (VLSI) Systems* 15.1 (Jan. 2007), pp. 90–103.

[218] Oscar Morales Chacón et al. 'Analysis of Energy Consumption Bounds in CMOS Current-Steering Digital-to-Analog Converters'. In: *Analog Integrated Circuits and Signal Processing* 111.3 (June 2022), pp. 339–351.

[219] Bartosz Bossy, Pawel Kryszkiewicz and Hanna Bogucka. 'Flexible, Brain-Inspired Communication in Massive Wireless Networks'. In: *Sensors* 20.6 (Jan. 2020), p. 1587.

[220] George Giamougiannis et al. 'A coherent photonic crossbar for scalable universal linear optics'. In: *Journal of Lightwave Technology* 41.8 (Apr. 2023), pp. 2425–2442.

[221] Lawrence M. Murray, Anthony Lee and Pierre E. Jacob. 'Parallel Resampling in the Particle Filter'. In: *Journal of Computational and Graphical Statistics* 25.3 (July 2016), pp. 789–805.

[222] Amin Jarrah, Abdel-Karim Al-Tamimi and Tala Albashir. 'Optimized Parallel Implementation of Extended Kalman Filter Using FPGA'. In: *Journal of Circuits, Systems and Computers* 27.01 (Jan. 2018), p. 1850009.

[223] Federico Castanedo. 'A Review of Data Fusion Techniques'. In: *The Scientific World Journal* 2013 (2013), pp. 1–19.

[224] J.Z. Sasiadek. 'Sensor Fusion'. In: *Annual Reviews in Control* 26.2 (Jan. 2002), pp. 203–228.

[225] Charles K. Chui and Guanrong Chen. *Kalman Filtering.* Cham: Springer International Publishing, 2017.

[226] Thomas Ferreira de Lima et al. 'Machine Learning With Neuromorphic Photonics'. In: *Journal of Lightwave Technology* 37.5 (Mar. 2019), pp. 1515–1534.

[227] Jan Marian Maciejowski. *Predictive Control: With Constraints.* 1. publ., [Nachdr.] Harlow: Prentice Hall, 2008.

[228] Nikolaos Farmakidis. 'Integrated Plasmonic Devices for Computing and Sensing'. PhD thesis. Oxford, UK: University of Oxford, Oct. 2021.

[229] Hanyu Zhang et al. 'Nonvolatile Waveguide Transmission Tuning with Electrically-Driven Ultra-Small GST Phase-Change Material'. In: *Science Bulletin* 64.11 (June 2019), pp. 782–789.

[230] Jaydeep Sarkar. 'Performance of Sputtering Targets and Productivity'. en. In: *Sputtering Materials for VLSI and Thin Film Devices.* Elsevier, 2014, pp. 171–195.

[231] Piero Orlandi et al. 'Tunable Silicon Photonics Directional Coupler Driven by a Transverse Temperature Gradient'. In: *Optics Letters* 38.6 (Mar. 2013), pp. 863–865.

[232] Michael J. Strain et al. 'Tunable Q-factor Silicon Microring Resonators for Ultra-Low Power Parametric Processes'. In: *Optics Letters* 40.7 (Apr. 2015), pp. 1274–1277.

[233] Manuel Le Gallo et al. 'A 64-Core Mixed-Signal in-Memory Compute Chip Based on Phase-Change Memory for Deep Neural Network Inference'. In: *Nature Electronics* 6.9 (Sept. 2023), pp. 680–693.

[234] Alfred O. Hero and Douglas Cochran. 'Sensor Management: Past, Present, and Future'. In: *IEEE Sensors Journal* 11.12 (Dec. 2011), pp. 3064–3075.

[235] G. W Ng and K. H Ng. 'Sensor Management – What, Why and How'. In: *Information Fusion* 1.2 (Dec. 2000), pp. 67–75.

[236] Yang Wang and Stephen Boyd. 'Fast Model Predictive Control Using Online Optimization'. In: *IEEE Transactions on Control Systems Technology* 18.2 (Mar. 2010), pp. 267–278.

[237] Aleksandra Lekić et al. 'Microsecond Nonlinear Model Predictive Control for DC-DC Converters'. In: *International Journal of Circuit Theory and Applications* 48.3 (2020), pp. 406–419.

[238] Arun Muraleedharan, Hiroyuki Okuda and Tatsuya Suzuki. 'Real-Time Implementation of Randomized Model Predictive Control for Autonomous Driving'. In: *IEEE Transactions on Intelligent Vehicles* 7.1 (Mar. 2022), pp. 11–20.

[239] Buyun Chen et al. 'A Memristor-Based Hybrid Analog-Digital Computing Platform for Mobile Robotics'. In: *Science Robotics* 5.47 (Oct. 2020), eabb6938.

[240] Shania Rehman et al. 'Analog–Digital Hybrid Computing with SnS_2 Memtransistor for Low-Powered Sensor Fusion'. In: *Nature Communications* 13.1 (May 2022), p. 2804.

[241] Pedro Freire et al. 'Artificial Neural Networks for Photonic Applications—from Algorithms to Implementation: Tutorial'. In: *Advances in Optics and Photonics* 15.3 (Sept. 2023), pp. 739–834.

[242] M. Rick et al. 'Autonomous Driving Based on Nonlinear Model Predictive Control and Multi-Sensor Fusion'. In: *IFAC-PapersOnLine*. 10th IFAC Symposium on Intelligent Autonomous Vehicles IAV 2019 52.8 (Jan. 2019), pp. 182–187.

[243] Trevor Bekolay et al. 'Nengo: A Python Tool for Building Large-Scale Functional Brain Models'. In: *Frontiers in Neuroinformatics* 7 (2014).

[244] Fatemeh Kiani et al. 'A Fully Hardware-Based Memristive Multilayer Neural Network'. In: *Science Advances* 7.48 (Nov. 2021), eabj4801.

[245] Isin Surekcigil Pesch et al. 'Multimodal Transistors as ReLU Activation Functions in Physical Neural Network Classifiers'. In: *Scientific Reports* 12.1 (Jan. 2022), p. 670.

[246] Sergey Vichik and Francesco Borrelli. 'Fast Solution of Linear and Quadratic Programs with an Analog Circuit'. In: *2014 American Control Conference*. June 2014, pp. 2954–2959.

[247] Zerui Liu et al. 'Ultrafast Hybrid Computing Systems Enabled by Memristor-Based Quadratic Programming Circuits'. In: *Advanced Functional Materials* 34.45 (Nov. 2024), p. 2401600.

[248] Norm Jouppi et al. 'TPU v4: An Optically Reconfigurable Supercomputer for Machine Learning with Hardware Support for Embeddings'. In: *Proceedings of the 50th Annual International Symposium on Computer Architecture*. Orlando FL USA: ACM, June 2023, pp. 1–14.

[249] Farshid Ashtiani, Alexander J. Geers and Firooz Aflatouni. 'An On-Chip Photonic Deep Neural Network for Image Classification'. In: *Nature* 606.7914 (June 2022), pp. 501–506.

[250] Yang Shi et al. '103 GHz Germanium-on-Silicon Photodiode Enabled by an Optimized U-shaped Electrode'. In: *Photonics Research* 12.1 (Jan. 2024), pp. 1–6.

[251] Hongtao Chen et al. '100-Gbps RZ Data Reception in 67-GHz Si-Contacted Germanium Waveguide p–i–n Photodetectors'. In: *Journal of Lightwave Technology* 35.4 (Feb. 2017), pp. 722–726.

[252] Changhao Han et al. 'Slow-Light Silicon Modulator with 110-GHz Bandwidth'. In: *Science Advances* 9.42 (Oct. 2023), eadi5339.

[253] Hong Deng et al. *Silicon Photonic Modulator Circuit with Programmable Intensity and Phase Modulation Response*. Feb. 2025. arXiv: `2502.20300 [physics]`.

[254] Naoki Takahashi et al. 'Optical ReLU Using Membrane Lasers for an All-Optical Neural Network'. In: *Optics Letters* 47.21 (Nov. 2022), p. 5715.

[255] Isin Surekcigil Pesch et al. 'Multimodal Transistors as ReLU Activation Functions in Physical Neural Network Classifiers'. In: *Scientific Reports* 12.1 (Jan. 2022), p. 670.

[256] Aashu Jha, Chaoran Huang and Paul R. Prucnal. 'Reconfigurable All-Optical Nonlinear Activation Functions for Neuromorphic Photonics'. In: *Optics Letters* 45.17 (Sept. 2020), pp. 4819–4822.

[257] Y. Peng et al. 'All-Optical Reconfigurable Low-Threshold Nonlinear Activation Functions for High-Precision Neural Network'. In: *Frontiers in Optics + Laser Science 2023 (FiO, LS)*. Tacoma, Washington: Optica Publishing Group, 2023, FW6E.3.

[258] Mohammad Mahdi Khafaji, Guido Belfiore and Frank Ellinger. 'A Linear 65-GHz Bandwidth and 71-dBΩ Gain TIA With 7.2 pA/√Hz in 130-nm SiGe BiCMOS'. In: *IEEE Solid-State Circuits Letters* 4 (2021), pp. 76–79.

[259] Giannino Dziallas et al. 'A 97-GHz 66-dBΩ SiGe BiCMOS Low-Noise Transimpedance Amplifier for Optical Receivers'. In: *IEEE Microwave and Wireless Components Letters* 31.12 (Dec. 2021), pp. 1295–1298.

[260] Piergiulio Mannocci et al. 'A Universal, Analog, In-Memory Computing Primitive for Linear Algebra Using Memristors'. In: *IEEE Transactions on Circuits and Systems I: Regular Papers* 68.12 (Dec. 2021), pp. 4889–4899.

[261] Robert Polster et al. 'A TIA for Optical Networks-on-Chip in 65 nm CMOS'. In: *2015 IEEE Optical Interconnects Conference (OI)*. Apr. 2015, pp. 109–110.

[262] Di Liang et al. 'High-Performance Quantum-Dot Distributed Feedback Laser on Silicon for High-Speed Modulations'. In: *Optica* 8.5 (May 2021), pp. 591–593.

[263] Kirill P. Kalinin et al. 'Analog Optical Computer for AI Inference and Combinatorial Optimization'. In: *Nature* (Sept. 2025), pp. 1–8.

[264] Otfried Madelung. *Semiconductors: Data Handbook*. 3rd ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004.

[265] Kiumars Aryana et al. 'Optical and Thermal Properties of $Ge_2Sb_2Te_5$, Sb_2Se_3, and Sb_2S_3 for Reconfigurable Photonic Devices [Invited]'. In: *Optical Materials Express* 13.11 (Nov. 2023), pp. 3277–3286.

[266] Xuanhui Fan et al. 'Systematic Investigations on Doping Dependent Thermal Transport Properties of Single Crystal Silicon by Time-Domain Thermoreflectance Measurements'. In: *International Journal of Thermal Sciences* 177 (July 2022), p. 107558.

[267] Jacek Gosciniak. 'Nonvolatile Plasmonics Based on Optically Reprogrammable Phase Change Materials'. In: *IEEE Photonics Journal* 14.3 (June 2022), pp. 1–8.

[268] Sanjeev Gautam et al. 'Electrical Properties of A-Antimony Selenide'. In: *Pramana* 50.1 (Jan. 1998), pp. 25–33.

[269] W R Thurber et al. *The Relationship between Resistivity and Dopant Density for Phosphorus-and Boron-Doped Silicon*. Tech. rep. NBS SP 400-64. Gaithersburg, MD: National Bureau of Standards, 1981, NBS SP 400–64.

[270] *Diffused Ion Implantation Profile Calculator and Graph*. 2024. URL: `https://cleanroom.byu.edu/implantcal`.

[271] Chunmeng Song et al. 'Compact Nonvolatile 2×2 Photonic Switch Based on Two-Mode Interference'. In: *Optics Express* 30.17 (Aug. 2022), pp. 30430–30440.

[272] Kun Shang et al. 'Non-Volatile 2×2 Optical Switch Using Multimode Interference in an Sb_2Se_3-Loaded Waveguide'. In: *Optics Letters* 49.3 (Feb. 2024), pp. 722–725.

[273] Khoi Phuong Dao, Juejun Hu and Richard Soref. 'Design of an Ultra-Compact, Energy-Efficient Non-Volatile Photonic Switch Based on Phase Change Materials'. In: *Journal of Optical Microsystems* 4.03 (May 2024).

[274] Thuy Tran Thi Thanh et al. 'Design and Demonstration of a High-Performance, Compact 2×2 Optical Switch Based on Optimized RI-MMI Couplers for Fast C-band Switching in Scalable Photonic Networks'. In: *Optics Communications* 596 (Dec. 2025), p. 132462.

[275] Qiyuan Yi et al. 'Silicon Photonic Flat-Top WDM (de)Multiplexer Based on Cascaded Mach-Zehnder Interferometers for the 2 µm Wavelength Band'. In: *Optics Express* 30.15 (July 2022), pp. 28232–28241.

[276] Chuanyu Lian et al. 'Phase and Amplitude Trimming of Photonic Integrated Circuits Using Phase Change Materials'. In: *2022 IEEE Photonics Society Summer Topicals Meeting Series (SUM)*. July 2022, pp. 1–2.

[277] Ahmed H. El-Saeed et al. 'Low-Loss Silicon Directional Coupler With Arbitrary Coupling Ratios for Broadband Wavelength Operation Based on Bent Waveguides'. In: *Journal of Lightwave Technology* 42.17 (Sept. 2024), pp. 6011–6018.

[278] Rui Chen et al. *NEO-PGA: Nonvolatile Electro-Optically Programmable Gate Array*. June 2025. arXiv: `2506.18592`.

[279] Jiajiu Zheng et al. 'Nonvolatile Electrically Reconfigurable Integrated Photonic Switch Enabled by a Silicon PIN Diode Heater'. In: *Advanced Materials* 32.31 (2020), p. 2001218.

[280] Carlos Rios et al. 'Controlled Switching of Phase-Change Materials by Evanescent-Field Coupling in Integrated Photonics [Invited]'. In: *Optical Materials Express* 8.9 (Sept. 2018), pp. 2455–2470.

[281] Iffa Binti Sharuddin and L. Lee. 'An Ultra-Low Power and Area Efficient 10 Bit Digital to Analog Converter Architecture'. In: *2014 IEEE International Conference on Semiconductor Electronics (ICSE2014)*. Aug. 2014, pp. 305–308.

[282] Noor A.B.A. Taib et al. 'A Low Power Op Amp for 3-Bit Digital to Analog Converter in 0.18 μm CMOS Process'. In: *Research Journal of Applied Sciences, Engineering and Technology* 5.8 (Mar. 2013), pp. 2592–2598.

[283] *8-CHANNEL, 12-/10-/8-BIT, 2.7-V TO 5.5-V LOW POWER DIGITAL-TO-ANALOG CONVERTERS WITH POWER DOWN AND INTERNAL REFERENCE*. Nov. 2008. URL: https://www.ti.com/lit/ds/symlink/tlv5632.pdf?ts=1754894385154.

[284] *10-Bit, Low-Power, Quad, Voltage-Output DAC with Serial Interface*. 2008. URL: https://www.mouser.co.uk/datasheet/2/609/MAX5741-3470420.pdf.

[285] Hua Fan et al. 'A 4-Channel 12-Bit High-Voltage Radiation-Hardened Digital-to-Analog Converter for Low Orbit Satellite Applications'. In: *IEEE Transactions on Circuits and Systems I: Regular Papers* 65.11 (Nov. 2018), pp. 3698–3706.

[286] David González-Andrade et al. 'Ultra-Broadband Nanophotonic Phase Shifter Based on Subwavelength Metamaterial Waveguides'. In: *Photonics Research* 8.3 (Mar. 2020), p. 359.

[287] Xiheng Ai et al. 'On-Chip Broadband Mach-Zehnder Interferometer Based on a Broadband Taper-Section Phase Shifter'. In: *Optics Express* 32.20 (Sept. 2024), p. 35551.

[288] Lei Qiao, Weijie Tang and Tao Chu. '32×32 Silicon Electro-Optic Switch with Built-in Monitors and Balanced-Status Units'. In: *Scientific Reports* 7.1 (Feb. 2017), p. 42306.

[289] Yating Wu and Tao Chu. 'Ultralow-Crosstalk Silicon Electro-Optic MZI Switch Based on Cascaded Phase Shifters with Arm Loss Equivalence'. In: *Laser & Photonics Reviews* 19.19 (2025), e02129.

[290] Mehmet Berkay On et al. 'Programmable Integrated Photonics for Topological Hamiltonians'. In: *Nature Communications* 15.1 (Jan. 2024), p. 629.

[291] Tianxiang Dai et al. 'A Programmable Topological Photonic Chip'. In: *Nature Materials* (May 2024), pp. 1–9.

[292] Ashish Vaswani et al. *Attention Is All You Need*. Aug. 2023. arXiv: 1706.03762 [cs].

[293] David Patterson et al. *Carbon Emissions and Large Neural Network Training*. Apr. 2021. arXiv: 2104.10350 [cs].

[294] Jungwook Choi et al. *PACT: Parameterized Clipping Activation for Quantized Neural Networks*. July 2018. arXiv: 1805.06085 [cs].

[295] Yang Zhang et al. 'Memristive Quantized Neural Networks: A Novel Approach to Accelerate Deep Learning On-Chip'. In: *IEEE Transactions on Cybernetics* 51.4 (Apr. 2021), pp. 1875–1887.

[296] Sanghyun Seo and Juntae Kim. 'Efficient Weights Quantization of Convolutional Neural Networks Using Kernel Density Estimation Based Non-uniform Quantizer'. In: *Applied Sciences* 9.12 (Jan. 2019), p. 2559.

[297] Fengfu Li et al. *Ternary Weight Networks*. Nov. 2022. arXiv: 1605.04711 [cs].

[298] Iosif Demirtzioglou et al. 'Frequency Comb Generation in a Silicon Ring Resonator Modulator'. In: *Optics Express* 26.2 (Jan. 2018), pp. 790–796.

[299] S. Hadi Badri et al. 'Compact Modulators on Silicon Nitride Waveguide Platform via Micro-Transfer Printing of Thin-Film Lithium Niobate'. In: *Scientific Reports* 15.1 (Apr. 2025), p. 11681.

[300] Young Min Kim et al. 'Non-Zero Grid for Accurate 2-Bit Additive Power-of-Two CNN Quantization'. In: *IEEE Access* 11 (2023), pp. 32051–32060.

[301] Eunhyeok Park, Junwhan Ahn and Sungjoo Yoo. 'Weighted-Entropy-Based Quantization for Deep Neural Networks'. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. July 2017, pp. 7197–7205.

[302] Alexander Sludds et al. 'Delocalized Photonic Deep Learning on the Internet's Edge'. In: *Science* 378.6617 (Oct. 2022), pp. 270–276.

[303] Linpeng Gu et al. 'A Compact Structure for Realizing Lorentzian, Fano, and Electromagnetically Induced Transparency Resonance Lineshapes in a Microring Resonator'. In: *Nanophotonics* 8.5 (May 2019), pp. 841–848.

[304] Tingge Yuan et al. 'Frequency-Space Selective Fano Resonance Based on a Micro-Ring Resonator on Lithium Niobate on Insulator'. In: *Laser & Photonics Reviews* 18.11 (2024), p. 2400457.

[305] Yingjie Xu et al. 'T-Shaped Silicon Waveguide Coupled with a Micro-Ring Resonator-Based Fano Resonance Modulator'. In: *Applied Optics* 61.31 (Nov. 2022), pp. 9217–9224.

[306] Fan Wang et al. 'Fano-Resonance-Based Mach-Zehnder Optical Switch Employing Dual-Bus Coupled Ring Resonator as Two-Beam Interferometer'. In: *Optics Express* 17.9 (Apr. 2009), pp. 7708–7716.

[307] Tien-Ju Yang et al. 'A Method to Estimate the Energy Consumption of Deep Neural Networks'. In: *2017 51st Asilomar Conference on Signals, Systems, and Computers*. Oct. 2017, pp. 1916–1920.