

# Building Blocks for Fast and Scalable Photonic Computing



Samarth Aggarwal  
Lincoln College  
University of Oxford

Supervisor: Prof. Harish Bhaskaran

A thesis submitted in partial fulfilment of the requirements for the  
degree of

*Doctor of Philosophy*

Trinity 2023

This thesis is dedicated to  
Mom , Dad & Saumya  
for their continued love and support



उद्यमेन हि सिध्यन्ति कार्याणि न मनोरथैः ।  
न हि सुप्तस्य सिंहस्य प्रविशन्ति मुखे मृगाः ॥  
*-There is no substitute to hard work.*

# Acknowledgements

As my DPhil comes to an end, I look back at this incredible journey which was filled with a lot of happy moments and tough times, especially with a global pandemic. Throughout all these times, there have been a lot of people who supported me, and I want to convey my deepest gratitude to them.

First, I would like to thank my supervisor Prof. Harish Bhaskaran for giving me the opportunity to join the group and introducing me to the exciting world of phase change. Harish has always been a pillar of support academically and, most importantly, emotionally. I can't still forget the time he was there for me during the pandemic when things were not looking so good back at home. Once again, thanks to Harish for being so caring and supportive all these years.

Next, I would like to thank my mentor Xuan, who, from the first day, helped and guided me in all aspects of my DPhil. I am so grateful for answering my stupid questions and training me so well in everything. I can't thank you enough and express my gratitude to you. I would also like to thank Nick, who took me under his wings when Xuan left unexpectedly and since then has been there to guide me. You have been a really good friend and collaborator all these years.

I would like to give a special shout to Jason, as I had promised him in my first year. Since my first year, Jason has been pushing me to complete all my assignments and submissions on time. I really learnt a lot from you over these years and still cherish the time we worked late in the lab over weekends to finish our projects and all the office chats we had.

I would like to thank all the past and present group members from ANE team Bowei, Mengyun, Yuhan, Yu, Utku, Yi, Tara, Haavard, Sergei, Kairan, Sijing, Serena, Guoce, Eugene, Maxine, James, Wen, Mingde, Tangsheng, Claire, Anotonios, Johannes, Luci and Andy, for all the important discussions and fun times we had together. I would also like to thank Paul Pattison and Radka for keeping the cleanroom up and running all the time to get our work done. All the members of the IT team and especially Paul, solved all the IT-related problems. I would also like to thank all the collaborators, Bruno, Iman, Emmanuel, and Alessio, for helping with different aspects of my projects.

I would also like to thank all my teachers from school and IIT Roorkee, who helped me grow my interest in science. I would like to thank Prof Indranil Lahiri, who introduced me to research and allowed me to work in his group. Thanks to Prof Karl Stol, who also gave me a chance to work in his group for the summer. Special thanks to Dr Anu Agarwal, who introduced me to the exciting world of photonics and taught me everything about photonics and nanofabrication.

Throughout this arduous journey, my closest friends Rohan and Akshat have always been there. Thank you, Rohan, for all your input to my research and for

providing me with a home in London whenever I felt low. Thanks to Akshat for hours of weekly random talks and all the fun moments. Thanks to my dear friends Harshit, Dikshit, and Raman back in India for sharing all the stupid and fun-filled gossip, which kept my life interesting. Special thanks to all my friends from the Indian junta group: Aditya, Aditi, Ayush, Arya, Himani, Mayur, Mohita, Natasha, Varnika and Viveka. Thanks to all the flatmates: Fabian, Miranda, Rob, Conor and Leonard and friends from Clarendon and Lincoln College. You all made the time fly so quickly.

Finally, I would like to thank the most important people, my parents and my family. My parents sacrificed a lot while I was in Oxford. It was made extremely difficult due to the pandemic, and I am so grateful to them for all their support, regardless of everything going haywire around us. Special mention to my sister Saumya and my brother-in-law Priyanshu for taking care of Mom and Dad and all their love while I was away. I would also like to thank my cousins and extended family for all their love and support.

Samarth Aggarwal  
Oxford , 2023

# Abstract

The current state-of-the-art computing architecture is based on von-Neumann architecture, which involves shuttling off data between memory and the processing unit to carry out computation through the bus. Machine learning algorithms, on the other hand, require massive parallelisation and high data throughput, making von-Neumann's architecture highly inefficient. Therefore, there is a need for specialised hardware to meet the computational demands. Photonics has been proposed to be a promising platform for computing models which are highly distributed and parallel due to the inherent benefits of high interconnectivity and parallelism offered by multiplexing techniques such as wavelength division multiplexing (WDM) and mode division multiplexing (MDM).

So far, various photonic computing architectures have been demonstrated. However, current photonic systems are limited by slow memory cells without write and readout characteristics and efficient arrangement of such memory cells into a network for scalability. Another challenge is related to identifying the right photonic platform with both excellent optical properties and also allowing for the co-integration of electronic and photonic components to build a complete photonic computing system. To overcome these challenges, this thesis investigates functional materials to achieve ultrafast photonic memory operation. Next, a novel architecture is realised, which allows scalability of photonic networks. Finally a promising photonic platform based on silicon carbide for functional material integration to achieve memory device is investigated.

The first part of the thesis is dedicated to the development of an ultrafast integrated photonics memory device using single-element antimony (Sb). Using femtosecond pulses and modulating the number of pulses and their energies, a photonic device based on Sb can be reversibly switched between amorphous and crystalline phases. Each phase demonstrates different optical absorption, therefore, resulting in an optical memory with switching speeds of 500 MHz. Varying pulse energies can also result in up to 7 different memory levels with the potential application of optical analogue computing. Further, the Sb memory cell is shown to possess tuneable volatility, which has been exploited for building an all-optical tunable linear filter.

The next part of the thesis is dedicated to building a scalable photonic network using a reduced rank technique. This technique reduces the redundancy in neural networks and hence decrease the complexity of the entire network. This results in less aggressive scaling of the associated photonic circuitry and offers advantages in terms of a smaller footprint and a lower energy budget in the case of a scaled-up network. A large matrix with integrated photodetectors and modulators is designed using a process design kit (PDK) from a foundry. Next, to further

improve a photonic network's scalability and fabrication tolerance, an asymmetrical multimode Y-coupler design has been investigated using simulations. It is estimated that such a Y-coupler design has about a 10 per cent improvement in terms of coupling losses in the network and is tolerant to fabrication variations, therefore improving the reliability of the network.

Finally, a promising photonic platform based on amorphous silicon carbide (SiC) on an insulator is investigated. An integrated photonic platform is achieved with impressive linear and nonlinear properties. Ring resonators with quality factors approaching 0.4 Million are fabricated, corresponding to less than 1 dB/cm waveguide loss, demonstrating excellent linear optical properties. To exhibit third-order nonlinearity, degenerate four-wave mixing experiments are conducted. A conversion efficiency of -23 dB is obtained, showing excellent nonlinear properties of the platform. Finally, a memory cell device based on this SiC is realised. Dynamic switching characteristics of the platform are studied and compared with silicon and silicon nitride photonic platforms.

# Contents

|                                                                  |              |
|------------------------------------------------------------------|--------------|
| <b>List of Figures</b>                                           | <b>x</b>     |
| <b>List of Tables</b>                                            | <b>xxii</b>  |
| <b>List of Abbreviations</b>                                     | <b>xxiii</b> |
| <b>1 Introduction</b>                                            | <b>1</b>     |
| 1.1 Motivation . . . . .                                         | 1            |
| 1.2 Outline and Objectives . . . . .                             | 6            |
| 1.3 Key results . . . . .                                        | 9            |
| 1.4 Statement of originality . . . . .                           | 10           |
| <b>2 Integrated Photonic Neuromorphic Accelerators: A Review</b> | <b>11</b>    |
| 2.1 Memory Cell . . . . .                                        | 12           |
| 2.2 Photonic Network Architectures . . . . .                     | 19           |
| 2.3 Photonic Platforms . . . . .                                 | 29           |
| 2.4 Critical Analysis and Summary . . . . .                      | 35           |
| <b>3 Theoretical Concepts</b>                                    | <b>37</b>    |
| 3.1 Dielectric Waveguide . . . . .                               | 37           |
| 3.2 Grating Couplers . . . . .                                   | 41           |
| 3.3 Ring Resonators . . . . .                                    | 43           |
| 3.4 Four-Wave Mixing . . . . .                                   | 47           |
| <b>4 Device and Experimental Setup</b>                           | <b>50</b>    |
| 4.1 Design and Fabrication . . . . .                             | 50           |
| 4.2 Experimental Setup . . . . .                                 | 54           |
| <b>5 Antimony Phase change memory</b>                            | <b>64</b>    |
| 5.1 Background . . . . .                                         | 65           |
| 5.2 Memory cell: Design and Fabrication . . . . .                | 65           |
| 5.3 Switching of Sb Memory cell . . . . .                        | 69           |
| 5.4 Volatility in Sb Memory . . . . .                            | 76           |
| 5.5 Demonstration of Applications . . . . .                      | 80           |
| 5.6 Critical Analysis . . . . .                                  | 90           |
| 5.7 Chapter Summary . . . . .                                    | 91           |

|          |                                                                   |            |
|----------|-------------------------------------------------------------------|------------|
| <b>6</b> | <b>Scaling up Matrix Multiplication Network: Reduced Rank</b>     | <b>92</b>  |
| 6.1      | Background . . . . .                                              | 93         |
| 6.2      | Modelling, Design and Experiments . . . . .                       | 94         |
| 6.3      | Application Demonstration . . . . .                               | 104        |
| 6.4      | Scaling Up . . . . .                                              | 109        |
| 6.5      | Performance on Large Models . . . . .                             | 119        |
| 6.6      | Critical Analysis . . . . .                                       | 123        |
| 6.7      | Chapter Summary . . . . .                                         | 124        |
| <b>7</b> | <b>Scaling up Matrix Multiplication Network: Multimode System</b> | <b>126</b> |
| 7.1      | Background . . . . .                                              | 126        |
| 7.2      | Device Design . . . . .                                           | 128        |
| 7.3      | System Level Simulations . . . . .                                | 132        |
| 7.4      | Fabrication Tolerance Comparision . . . . .                       | 134        |
| 7.5      | Critical Analysis . . . . .                                       | 136        |
| 7.6      | Chapter Summary . . . . .                                         | 137        |
| <b>8</b> | <b>Amorphous Silicon Carbide photonic Platform</b>                | <b>138</b> |
| 8.1      | Background . . . . .                                              | 138        |
| 8.2      | Platform Development . . . . .                                    | 139        |
| 8.3      | SiC phase change memory . . . . .                                 | 149        |
| 8.4      | Critical Analysis . . . . .                                       | 155        |
| 8.5      | Chapter Contributions . . . . .                                   | 155        |
| 8.6      | Chapter Summary . . . . .                                         | 156        |
| <b>9</b> | <b>Conclusion and Outlook</b>                                     | <b>157</b> |
| 9.1      | Conclusions . . . . .                                             | 157        |
| 9.2      | Outlook . . . . .                                                 | 158        |
|          | <b>List of publications</b>                                       | <b>160</b> |
|          | <b>Appendices</b>                                                 |            |
| <b>A</b> | <b>Code Availibility</b>                                          | <b>164</b> |
| <b>B</b> | <b>DAC controller codes</b>                                       | <b>165</b> |
|          | <b>References</b>                                                 | <b>167</b> |

# List of Figures

|     |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |    |
|-----|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 1.1 | The block diagram for von Neumann architecture shows different essential components. The data from the input goes through the processing unit, where the program is loaded from memory, computed in ALU, and then stored back into memory and finally sent out. Source: Wikipedia creative commons . . . . .                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      | 2  |
| 1.2 | Comparison of various computing technologies based on electronics and photonics in terms of computing efficiency and computation speed. Source:[3] . . . . .                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      | 3  |
| 2.1 | Electric-field-effect based on photonic memories. (a) The strong Pockels effect of BaTiO <sub>3</sub> is used to create an integrated silicon photonic memory. Applying voltage results in a change in the effective refractive index, thus changing the resonance wavelength. Source: [22] (b) CMOS-based floating gate memories (i) schematic of the photonic memory with SiN as a floating gate, (ii) simulations showing the electric field interaction of optical mode. Source: [23] (iii-iv) demonstration of nonvolatile multilevel photonic memories with polysilicon as a floating gate. Source: [24, 25] (c) Photonic memories based on alternative floating gates (i) demonstration of photonic memory based on ITO as a floating gate. Source: [26] (ii) demonstration of graphene-based floating gate memory. (d) Implementation of semiconductor optical amplifier-based photonic memory on an InP platform. Source: [27] . . . . . | 13 |
| 2.2 | MEMS-based memories. (a) Using MEMS integrated with a photonic waveguide, a phase shifter is implemented. Source: [35]. (b) A MEMS-based optical switch. By changing the coupling distance of adiabatic couplers, over 50 dB extinction ratios can be achieved. Source: [36] . . . . .                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            | 15 |
| 2.3 | Phase change materials(GST) based photonic memories. (a) Heating the material above its melting point and rapid cooling results in amorphisation, while heating above the glass transition temperature and holding at this temperature with slow cooling results in crystallisation. (b-c) Amorphous and crystalline phases of GST show different optical and electrical properties, respectively. . . . .                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        | 17 |
| 2.4 | (a)Schematic of a phase change memory using waveguide demonstrating (b) reversible switching. Source:[7], (c) Up to 32 memory levels can be achieved using single-shot programming methods. Source:[8]                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            | 18 |



- 2.5 All optical spiking neural network (a) Illustrative model of an SNN  
b) All-optical spiking neural network using ring resonators for wave-  
length division multiplexing (WDM) and PCMs. Input signals of  
different wavelengths are modulated on the chip. The plastic synapses  
multiply the weights, and the intensity of different wavelengths of  
light is added using a ring resonator. If this is above the threshold  
for switching the PCM, there is a spiking output. Source:[59] . . . . 20
- 2.6 Demonstration of spiking neural network based on ring resonators  
and PCM. By changing the phase of the PCM, the critical coupling  
or resonance wavelength can be controlled. a) and b) Schematic  
diagram of a ring resonator with GST as PCM. c) and d) change  
in transmission on changing the phase of PCM in the through port  
and drop port. Since both ports have different responses, the same  
device can be used for positive and negative weights. Source:[60] . . . 21
- 2.7 By changing the port of the ring resonator connected to the circuit,  
both positive and negative weights can be achieved. As seen in Figure  
2.6, Since both the drop and through have different responses to the  
amorphisation pulse, the device can be used for both positive and  
negative weights. Source:[60] . . . . . 21
- 2.8 (a) Demonstration of the first on-chip all-optical deep neural network  
with N hidden layers. Using fundamental optical elements such as an  
optical interferometer and splitters, the weight matrix of a single layer  
is broken down into its single-value decomposition matrix. Using  
the different channels of waveguides as input vectors, the matrix  
multiplication operation is performed in the photonic circuit. The  
output of each layer is passed through a nonlinear optical unit, and  
the resultant signal is passed through another layer. (b) An N-layer  
DNN has been constructed for vowel recognition from speech data.  
Source:[13] . . . . . 23
- 2.9 a) Photonic waveguide model to achieve photonic multiplication using  
GST. Using a Pulse signal to change the phase of the GST, there is  
a change in transmittance. This is equivalent to multiplication, as  
explained in b). c) represents combining the multiplication scheme  
shown in a) and b) for vector-vector multiplication as explained  
by mathematics in d). Using a two-input one output device, the  
intensity of individual multiplication cells is added. Source:[69] . . . 24

|      |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |    |
|------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 2.10 | Phase change materials-based deep neural network implementations.<br>(a) Photonic neural network implementation exploiting WDM using frequency comb to modulate input signals and using a PCM-based crossbar array structure to carry out matrix-vector multiplication. Source:[15], (b) Photonic architecture using phase shifters to reduce I/O for PCM switching for computation. Source:[70], (c) Using the integrated photonic heaters, PCM weights can be programmed electrically while reading optically. Source:[71]. . . . . | 25 |
| 2.11 | DNN model showing vector-matrix multiplication. The input lasers (input vector – B) are distributed equally to the Micro-ring modulator module (Matrix A) using a multiplexer and splitter, as shown in Figure 2.11. Each ring, responsive to different wavelengths, can be individually set to the weight matrix. The photodetector sums the intensity of the light to get the output vector C. Source:[14] . . . .                                                                                                                  | 26 |
| 2.12 | SLM-based photonic computing implementations. Using SLM for a linear layer of a DNN and projecting this as an image through a nonlinear fibre first layer of DNN was implemented. Source [74] (b) Lithography-free implementation of a photonic neural network. By projecting high intensity modulated pump source on an III-V thin film, the photonic network was trained for vowel recognition. Source:[73]. . . . .                                                                                                                | 27 |
| 2.13 | Direct Electro-optic modulation based neural networks. (a) The input-modulated signals are multiplied directly using optical attenuators, and the optical signals are collected onto a photodetector and summed. This drives an optical modulator to achieve a nonlinear function. Source:[75] (b) Use of temporal, wavelength and spatial degree of freedom for convolution process. By shaping a frequency comb and using an EOM to modulate a time multiplex signal, convolution is performed. Source:[76] . . . . .               | 29 |
| 2.14 | Transparency window for various photonic platforms and the wavelength requirements depending on the application. Source:[77] . . .                                                                                                                                                                                                                                                                                                                                                                                                    | 30 |
| 2.15 | Typical silicon photonic foundry process node showing available active and passive components. . . . .                                                                                                                                                                                                                                                                                                                                                                                                                                | 31 |
| 3.1  | Schematic of planar waveguide with one-dimensional optical confinement, (b) Index profile corresponding to the planar waveguide. . . .                                                                                                                                                                                                                                                                                                                                                                                                | 38 |
| 3.2  | Ray-diagram showing the propagation of the guided ray in a slab waveguide structure. . . . .                                                                                                                                                                                                                                                                                                                                                                                                                                          | 39 |
| 3.3  | Cross-sectional schematic of a grating coupler with ray diagrams. .                                                                                                                                                                                                                                                                                                                                                                                                                                                                   | 42 |

|     |                                                                                                                                                                                                                                                                                                                        |    |
|-----|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 3.4 | Schematic of an all-pass ring resonator of radius $r$ and coupling parameters. . . . .                                                                                                                                                                                                                                 | 43 |
| 4.1 | A schematic of the photonic design patterns consisting of the waveguide, ring resonators and MZI, used to fabricate photonic components. . . . .                                                                                                                                                                       | 51 |
| 4.2 | Device Fabrication Process (clockwise). Starting with a silicon chip, an electron beam-sensitive photoresist is spin coated on the chip. The chip is then exposed with the desired pattern and developed. Finally, using RIE, silicon is partially etched, and the photoresist is removed using oxygen plasma. . . . . | 53 |
| 4.3 | Lift off process. The lithography step is used to open windows for material deposition. Using sputtering PCMs, thin films are deposited on the patterned chip. The chip is then dipped in a photo remover to get rid of the photoresist and the excess material with it. . . . .                                       | 54 |
| 4.4 | Experimental Setup for measuring transmission characteristics of waveguides. A laser source is sent across the photonic device. The input is polarisation controlled, and the output is recorded on a photodetector and recorded onto a computer using a DAQ. . . . .                                                  | 55 |
| 4.5 | Illustrative diagram showing the fibre array used to couple light in and out of the photonic chip. Grating couplers are used to input light into the device. . . . .                                                                                                                                                   | 56 |
| 4.6 | Experimental Setup for integrated photonic memory based on antimony using a femtosecond laser. Femtosecond pulses are captured using a pulse picker and amplified using EDFA to switch PCM. A CW laser is used to monitor the change in transmission. . . . .                                                          | 57 |
| 4.7 | Typical D-flip-flop timing diagram. At every clock rising cycle, the value of 'data' is obtained at the output. . . . .                                                                                                                                                                                                | 60 |
| 4.8 | (a) Timing diagram with pulse picker output to capture single and multiple pulses.(b) Experimental setup showing a train of pulses on an oscilloscope. (c) Experimental capture of a single pulse using a pulse picker. . . . .                                                                                        | 61 |
| 4.9 | Experimental setup for matrix multiplication experiments. The setup consists of two lines: inference and programming line. The programming line is used to set the weights for the matrix. After setting the weights, the inference line is used to carry out the computation. . . . .                                 | 61 |

|      |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |    |
|------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 4.10 | Experimental setup for four-wave mixing experiments. The setup consists of a pump and signal-tuneable laser sources. The light is combined using a 90/10 splitter, and the light is amplified using a cascaded EDFA system. After the device, the signal is split using another 90/10 splitter. One line is used to monitor the transmission of the devices using a photodetector. The other line is connected to an OSA to record the spectrum. . . . .                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | 63 |
| 5.1  | Schematic of SOI waveguide with 3 nm Sb, an amorphisation pulse (WRITE) or crystallisation pulse (ERASE) is used to switch the PCM.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    | 66 |
| 5.2  | Frequency domain simulations using COMSOL showing the E-field distribution in a silicon waveguide with 3 nm thick Sb. The colours represent the normalised (in the range of 0-1) intensity of the electric field. Crystalline Sb is highly absorptive and thus results in lower transmission. . . . .                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | 66 |
| 5.3  | Refractive index (n,k) of amorphous (black) and crystalline (red) antimony thin film from visible to near IR wavelengths . . . . .                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     | 67 |
| 5.4  | An SEM image showing the device structure, which includes a waveguide with grating couplers. The white dotted box shows the PCM deposited on the waveguide. . . . .                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    | 68 |
| 5.5  | AFM image confirms the thickness of Sb to be 3 nm. Profile lines at three regions have been plotted, confirming uniform deposition. . .                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                | 68 |
| 5.6  | Experimental results showing the effect of increasing Sb length on a waveguide on the loss (absorption of light) in as-deposited amorphous and hot plate annealed crystalline state. Increasing the length of Sb results in a higher contrast between amorphous and crystalline states.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                | 69 |
| 5.7  | Binary switching on a waveguide, a) Sb (length = 4 $\mu\text{m}$ ) on a waveguide switched using single, high energy ( $194 \pm 35\text{pJ}$ ) femtosecond pulse (800 fs) – (amorphisation pulse, red dashed vertical lines indicate this pulse). 100 low energy ( $45 \pm 9\text{ pJ}$ ) pulses (crystallisation sequence, blue dashed lines indicate when this is sent) crystallise the sample. b) Sb (length = 10 $\mu\text{m}$ ) switched between amorphous and crystalline phases as in a). As expected, longer length results in higher contrast. c) Multiple switching of the device in (b). d) Demonstration of single-pulse amorphisation and crystallisation for the same device as in (b-c) by fine-tuning the pulse energies; a single pulse is enough for binary switching. The red dashed lines represent the pulse number at which an amorphisation pulse of energy $91 \pm 7\text{pJ}$ is sent, while the blue dashed lines represent a crystallisation pulse of energy $59 \pm 11\text{pJ}$ . . . . . | 71 |

- 5.8 a) Multilevel programming in Sb. Four distinct levels of transmission are programmed (termed L0, L1, L2 and L3) using a single femtosecond (pulse width 800 fs) WRITE pulse (termed W1, W2 and W3) of energies  $145\pm 6$  pJ,  $150\pm 6$  pJ and  $160\pm 5$  pJ respectively. ERASE is achieved by sending 200 pulses of 100 pJ femtosecond pulses. b) Repeatedly reaching seven memory levels (L0-L6) by WRITE pulses W1-W6 (Pulse energies  $121\pm 7$  pJ,  $144\pm 5$  pJ,  $160\pm 5$  pJ,  $174\pm 4$  pJ,  $204\pm 4$  pJ and  $231\pm 4$  pJ respectively). c) The pulse energies corresponding to each level. . . . . 73
- 5.9 (a) Optical image of 4 port device used to measure the switching speed. Phase change material is deposited on the optical crossing. Using one of the ports as input for the pulse and the other as the probe, time-resolved switching characteristics are obtained. (b) Time-resolved switching dynamics of Sb, resulting in a rise time of 2 ns for a single sub-picosecond pulse on switching from crystalline to amorphous state. (c) Experimental comparison of the switching speed of Sb is with other known phase change materials such as GST and AIST using a single sub-picosecond pulse for amorphisation. . . . 75
- 5.10 a) Binary switching  $4\ \mu\text{m}$  long of Sb with 10 nm ITO capping on the using single, high energy ( $194\pm 35$  pJ) femtosecond pulse (800 fs) – (amorphisation pulse, red dashed vertical lines indicate this pulse). 100 low energy ( $45\pm 9$  pJ) pulses (crystallisation Pulse, blue dashed lines indicate this pulse sequence) crystallise the sample. b) Raman spectra of a capped and uncapped crystallised thin film of Sb on waveguide with their  $E_g$  and  $A_{1g}$  vibration modes. The circles denote the typical  $F_{2g}$  and  $A_g$  modes of  $Sb_2O_3$ , clearly showing the absence of any antimony oxide in the samples. . . . . 76
- 5.11 (a) Sub-millisecond readout for binary switching of  $10\ \mu\text{m}$  long of Sb. (b) Experiments in (a) were repeated multiple times, with less than 2% variation in memory levels. (c) Millisecond readout for intermediate memory levels obtained for different amorphisation pulse energies. Memory levels show a variation of 1% in readout over 100 ms time duration. . . . . 77

|      |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |    |
|------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 5.12 | (a)SEM image of the device used for long-term stability experiment. The probe signal is split using a Y-splitter into two arms one arm is used to measure the transmittance through Sb, and another reference arm is used to monitor the input power fluctuations and drift in the optical signal. The signal from the reference arm is used to normalise the transmission from the Sb arm and calculate the retention time of switched Sb, (b)Long-term stability of $4\mu\text{m}$ and $10\mu\text{m}$ long amorphous Sb, switched from crystalline phase using 409 pJ, a single femtosecond pulse. . . . .                                                                                                                                                                                                                             | 78 |
| 5.13 | a) Effect of the Probe Power on the maximum transmission achieved for a given Pump power. An increase in probe power results in a decreased maximum change in the readout. b) Effect of probe power on the retention time. Individual plots are normalised between 0 and 1, corresponding to the lowest (crystalline) and the highest (amorphous) transmission. The time required to reach 60% of the maximum change in readout is plotted. Increasing the probe power decreases the retention time. c) Effect of both Probe and Pump power on the maximum readout contrast. A high Pump and low probe power lead to the maximum change in the readout. d) Effect of Pump and Probe power on retention time (time required to reach 90% of maximum readout change). The volatility of Sb is controlled by varying either or both. . . . . | 79 |
| 5.14 | Impulse response of a leaky integrator is fitted with the switching response for the Sb filter device. An sb thin film of 5 nm is used. . .                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | 82 |
| 5.15 | Superposition of impulse response for Sb device. Experimentally obtained response (in black) is fitted to simulated responses (in blue and red). The summation of two simulated responses (in green) matches well with the experimentally obtained response. (a) The impulse response when two pulses are separated by a ‘long’ duration of 30ms. (b) Impulse response when two pulses are separated by a ‘short’ duration of 5 ms. . . . .                                                                                                                                                                                                                                                                                                                                                                                               | 83 |
| 5.16 | Illustration of the experimental setup used to achieve filtering operation consisting of pump and probe line. The fs-laser is modulated using an EOM, which is used to switch the Sb filter cell. The output is collected on a photodetector which is read on an oscilloscope. . .                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        | 85 |

|      |                                                                                                                                                                                                                                                                                                                                                                                                                          |     |
|------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 5.17 | (a) A clock signal with pulse width 25 ns and Frequency 3.3 MHz is used to downsample the femtosecond pulse train from 40MHz to 3.33 MHz. (b) RF signal, $ \sin(x) $ of 200 kHz is used as the RF-input signal, (c) Output of the RF mixer obtained that is sent to the EOM to modulate femtosecond laser pulses. (d) Downsampled and amplitude-modulated femtosecond pulses obtained that is used to switch Sb. . . . . | 86  |
| 5.18 | Sine wave responses of Sb filter cell (in blue) for various input sine wave (in red) of frequencies ranging from 4 kHz-400 kHz. . . . .                                                                                                                                                                                                                                                                                  | 87  |
| 5.19 | Frequency response analysis of the sine wave of data shown in Figure 5.18 is done to determine the pass-band and stop band of the filter. . . . .                                                                                                                                                                                                                                                                        | 88  |
| 5.20 | Demonstration of low pass filtering operation. A 4 kHz signal is amplitude modulated with various carrier frequencies, as reported in the graph. Increasing the carrier frequency results in the suppression of the amplitude of the carrier signal. . . . .                                                                                                                                                             | 88  |
| 5.21 | Frequency response analysis of the sine wave of data shown in Figure 5.20 is done to determine the pass-band and stop band of the filter . . . . .                                                                                                                                                                                                                                                                       | 89  |
| 5.22 | Simulation results showing the effect of retention time of Sb on the frequency cutoff for low pass filtering. . . . .                                                                                                                                                                                                                                                                                                    | 90  |
| 6.1  | Reduced rank architecture (a) Schematic of a photonic chip computing comparing the benefits of reduced rank architecture over conventional crossbar array architecture in terms of chip area and reduced I/O. (b) Kernels are factorised using rank reduction into two smaller matrices with reduced parameters. . . . .                                                                                                 | 94  |
| 6.2  | Schematic of the photonic implementation of the proposed architecture. Input signals $x_1$ - $x_7$ , encoded onto different wavelengths, are used as input vectors where two subsequent MVM results in the final output O. . . . .                                                                                                                                                                                       | 96  |
| 6.3  | Schematic diagram of a $3 \times 3$ equivalent device used in the experiments, showing signal flow. . . . .                                                                                                                                                                                                                                                                                                              | 97  |
| 6.4  | Cross-bar array schematic showing the distribution of optical power using a directional coupler for a $3 \times 3$ network. . . . .                                                                                                                                                                                                                                                                                      | 98  |
| 6.5  | The power transfer of a directional coupler from input to cross port on varying the coupling length of an SOI directional coupler. . . . .                                                                                                                                                                                                                                                                               | 99  |
| 6.6  | False coloured optical image of the fabricated device equivalent to $3 \times 3$ matrix but with a $33 \times$ reduction in size. (Inset) False-coloured SEM image of a unit cell showing waveguide interconnects and ‘active’ PCM memory. Zoomed in false coloured SEM image showing $1 \mu\text{m} \times 1 \mu\text{m}$ PCM (AIST) deposited on the waveguide. . . . .                                                | 100 |

|      |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |     |
|------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 6.7  | More than 5-bit data can be stored in PCM memory by changing the pulse energy amplitude. . . . .                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           | 101 |
| 6.8  | Schematic of the experimental flow used in Matrix-vector multiplication experiments. . . . .                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | 104 |
| 6.9  | Image Filtering (a) Input image of ‘Taj Mahal’ is used by pre-processing as an input vector for edge detection experiments. (b-c) Successful edge detection of vertical and horizontal filters, respectively, using the reduced rank photonic architecture. (d) Combining the edge detection results from (b) and (c) showing highlighted edges. (e) Computing error of expected vs measured value showing a Gaussian distribution with a standard deviation of 0.1. (f) Experimentally achieved weights using PCM for edge filter. . . . .                                                                                                                                                | 105 |
| 6.10 | (a) Sobel Filter is used for edge detection on an image of a London tube station. (b) edge detection results on software are compared with one processed on photonic hardware (c), showing excellent conformality of the two. (d) Computing error of expected vs measured value showing a normal distribution with a standard error deviation of 0.01. (e) Experimentally achieved weights using PCM for the Sobel filter . . . . .                                                                                                                                                                                                                                                        | 107 |
| 6.11 | Image classification using convolution neural network (a) Schematic of the convolutional neural network used to classify MNIST handwriting digits. The reduced rank photonic architecture with the kernel size equivalent to $3 \times 3$ , but with $33 \times$ reduction, is used. (b-c) Confusion matrix of image classification task for the model on software and experimentally obtained on the photonic hardware, showing an excellent agreement. (d) Error estimation of over 120K MVM by calculating expected vs measured results. A normal distribution fit reveals a standard error deviation of 0.013. (e) Experimentally achieved matrix for convolution experiments. . . . . | 108 |
| 6.12 | Effect of parameter reduction of matrix on the overall optical loss of the proposed photonic architecture. Parameter reduction results in lower losses in the system. . . . .                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | 110 |
| 6.13 | System level power requirements of a photonic computing network for different matrix sizes. . . . .                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        | 112 |
| 6.14 | The layout of the foundry chip design consists of various test structures and large-scale matrices. . . . .                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                | 113 |



|      |                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |     |
|------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 6.15 | EAM unit cell consisting of EAM as weighting element and two directional couplers to fan-in and fan-out optical signal. The Row-In signal is coupled to the EAM using DC-Splitter. This signal gets amplitude modulated by the EAM, and the resultant modulated signal is added with Column-In to get the Column-out using DC-combiner. . . . .                                                                                                                      | 114 |
| 6.16 | Design of an $8 \times 8$ equivalent crossbar array with EAM as a weighting element with integrated photodetectors and bond pads to read and write signal. . . . .                                                                                                                                                                                                                                                                                                   | 115 |
| 6.17 | PCM unit cell consisting of PCM as weighting element and two directional couplers to fan-in and fan-out optical signal. The Row-In signal is coupled to the PCM element using DC-Splitter. This signal gets amplitude modulated by the PCM memory, and the resultant modulated signal is added with Column-In to get the Column-out using DC-combiner. Using Pump-Input, the PCM can be switched, monitoring the weight using Probe-Input/Output couplers. . . . .   | 116 |
| 6.18 | Design of an $8 \times 8$ equivalent crossbar array with PCM as a weighting element with integrated photodetectors and bond pads to read photodetector signals. . . . .                                                                                                                                                                                                                                                                                              | 116 |
| 6.19 | A 40 analogue DAC controller evaluation board is used to control the EAM. The evaluation board is controlled using an Arduino-based microcontroller. . . . .                                                                                                                                                                                                                                                                                                         | 118 |
| 6.20 | Effect of reducing the number of parameters on the accuracy of matrix factorisation for different sizes of the matrix. Here the accuracy is defined as the difference between the Frobenius norm of the original weight matrix and the reduced rank matrix. Increasing the reduction factor results in lower accuracy. . . . .                                                                                                                                       | 119 |
| 6.21 | (a) Schematic of a feed-forward model used for MNIST handwritten digit recognition used for simulation in (b). (b) Effect of parameter reduction on the relative accuracy and programming energy of the system. Parameter reduction results in over $3 \times$ gain in computation density due to a smaller footprint without any loss in accuracy for the classification task with additional benefits of $2.5 \times$ gain in terms of programming energy. . . . . | 120 |
| 6.22 | (a) The deep learning model for the CIFAR-10 classification task. (b) Change in relative accuracy with decreasing the number of training parameters using the rank reduction method . . . . .                                                                                                                                                                                                                                                                        | 122 |

|      |                                                                                                                                                                                                                                                                                                                                                                                                            |     |
|------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 6.23 | Effect on the model accuracy for a) MNIST task with no reduction in training parameters, b) MNIST task with 60% pre-reduction in training parameters, c) CIFAR-10 task with no reduction in training parameters, d) CIFAR-10 task with 38% pre-reduction in training parameters. . . . .                                                                                                                   | 123 |
| 7.1  | Schematic diagram of the proposed asymmetric coupler with three ports. A single mode signal is coupled into the Bus input (port 2), resulting in a multimode output at Bus Output (Port 3). . . . .                                                                                                                                                                                                        | 129 |
| 7.2  | (a) Ey-field profile when $TE_0$ signal is injected at port 1. (b)-(e) Ey-field profile when $TE_0 - TE_3$ input is injected at Bus Input (port 2). (f) E-field magnitude profile when weakly supported $TE_4$ input is injected at Bus Input (port 2). . . . .                                                                                                                                            | 130 |
| 7.3  | Comparison of performance of the device as compared to a directional coupler and single mode Y-coupler with an increasing number of coupler nodes. . . . .                                                                                                                                                                                                                                                 | 132 |
| 7.4  | (a)Net System Loss (%) comparison of the device with an increasing number of nodes, (b)Broadband performance of the device with an increasing number of nodes. For wavelengths at 1500-1600nm, system losses are similar to that of single wavelengths. . . . .                                                                                                                                            | 133 |
| 7.5  | Eye Diagram for 32-node photonic network using the coupler design with a 10mW input signal modulated at 25 GHz, showing negligible bit error rate. . . . .                                                                                                                                                                                                                                                 | 133 |
| 7.6  | Schematic of a $3 \times 3$ photonic network with WDM using ring resonator fabricated in IMEC foundry. . . . .                                                                                                                                                                                                                                                                                             | 134 |
| 7.7  | (a)Variation in coupling power for a $3 \times 3$ photonic network based on directional coupler due to fabrication variation. (b)Variation in coupling power for a $3 \times 3$ photonic network based on ring resonator due to fabrication variation and resonance wavelength. . . . .                                                                                                                    | 135 |
| 7.8  | (a)Effect of waveguide width variation due to fabrication variations on the average transmission of the device. The waveguide widths are varied by 10%, and the change in the transmission is plotted. (b)Effect of waveguide thickness due to fabrication variations on the average transmission of the device. The etching depth is varied by 5%, and the change in the transmission is plotted. . . . . | 136 |
| 8.1  | The refractive index of SiC thin film obtained experimentally using ellipsometry. . . . .                                                                                                                                                                                                                                                                                                                  | 140 |

|      |                                                                                                                                                                                                                                                                                                                                                                                                  |     |
|------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 8.2  | Effective refractive index of SiC on insulator waveguides obtained using Lumerical. The waveguide width is varied from 500nm to 1500nm, keeping the waveguide thickness fixed at 265 nm. The effective refractive index for fundamental $TE_0$ and higher order mode at 1550 nm wavelength is plotted. The electric field distribution for $TE_0$ and $TE_1$ Mode is shown in the inset. . . . . | 141 |
| 8.3  | 2D FDTD Simulation setup used to simulate grating coupler parameters.                                                                                                                                                                                                                                                                                                                            | 142 |
| 8.4  | Experimental determination of grating coupler parameters for 1550 nm wavelength. Grating couplers with period $1.16 \mu m$ and fill factor of 0.50 show the best transmission with 7% coupling efficiency. . . .                                                                                                                                                                                 | 143 |
| 8.5  | Intrinsic Quality factors of Ring resonators deposited using ICPCVD and PECVD process. . . . .                                                                                                                                                                                                                                                                                                   | 144 |
| 8.6  | Intrinsic Quality factors of Ring resonators with varying waveguide widths from 500nm-1200nm deposited using ICPCVD (a) grown at $150^\circ C$ (b) grown at $250^\circ C$ . . . . .                                                                                                                                                                                                              | 145 |
| 8.7  | Four-wave mixing experiments demonstrating a conversion efficiency of -23 dB. . . . .                                                                                                                                                                                                                                                                                                            | 147 |
| 8.8  | Observation of emergence of Kerr frequency comb using amorphous silicon carbide ring resonators. . . . .                                                                                                                                                                                                                                                                                         | 148 |
| 8.9  | E-field profile of different absorption of input light for the $4 \mu m$ long crystalline and amorphous state of (a) AIST and (b) GST obtained using Lumerical FDTD simulations. . . . .                                                                                                                                                                                                         | 149 |
| 8.10 | (a)FEM thermal simulations for GST memory device using 5mW , 20ns pulse on changing the time from 5ns-20ns.(b) Plot for maximum temperature reached for GST device for a 20ns pulse and thus subsequent cooling rate. . . . .                                                                                                                                                                    | 151 |
| 8.11 | a)FEM thermal simulations for AIST memory device using 5mW, 20ns pulse on changing the time from 5ns-20ns.(b) Plot for maximum temperature reached for AIST device for a 20 ns pulse and subsequent cooling rate. . . . .                                                                                                                                                                        | 152 |
| 8.12 | SEM image showing fabricated SiC waveguide with 10 nm PCM and ITO as a capping layer deposited using RF sputtering. . . . .                                                                                                                                                                                                                                                                      | 152 |
| 8.13 | (a)Reversible switching of GST on SiC by varying the pulse width from 20-35ns for amorphisation. b) Reversible switching of AIST on SiC by varying the pulse width from 10-30ns for amorphisation. . .                                                                                                                                                                                           | 153 |
| 8.14 | a-b) Time-resolved dynamic switching response of GST and AIST, respectively, on SiC obtained by varying the pulse width and keeping the pulse amplitude constant c-d) Dead time of GST and AIST memory device for different pulse amplitude and pulse width. . . .                                                                                                                               | 154 |

# List of Tables

|     |                                                                                                                                                                                                              |     |
|-----|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 4.1 | Sputtering parameters and sputtering rates for various thin films used in the thesis. . . . .                                                                                                                | 54  |
| 6.1 | State of the art energy requirement of various photonic computing elements . . . . .                                                                                                                         | 111 |
| 6.2 | The energy required to set and store the weights for various weighting methods . . . . .                                                                                                                     | 111 |
| 6.3 | EAM Matrix Loss Estimation . . . . .                                                                                                                                                                         | 115 |
| 6.4 | PCM Matrix Loss Estimation . . . . .                                                                                                                                                                         | 117 |
| 6.5 | IMEC SiGe Photodetector characteristics . . . . .                                                                                                                                                            | 117 |
| 6.6 | EAM Matrix Photocurrent Estimation assuming 5mW input . . . .                                                                                                                                                | 118 |
| 6.7 | PCM Matrix Photocurrent Estimation assuming 5mW input . . . .                                                                                                                                                | 118 |
| 7.1 | Summary of advantages and disadvantages of various power coupling strategies for incoherent addition of light using single-mode Y-coupler, directional couplers and ring resonators. . . . .                 | 128 |
| 7.2 | Summary of the optimised design parameters for the asymmetric directional coupler obtained using particle swarm optimisation technique in Lumerical FDTD simulations. . . . .                                | 129 |
| 7.3 | Modal power distribution at Output Bus (Port 3) when single mode $TE_0$ light is injected at Signal Input port (Port 1), with a net transmission of 97.5%. . . . .                                           | 131 |
| 7.4 | Modal power distribution at Output Bus (Port 3) when various modes $TE_0$ - $TE_4$ are injected at Bus Input port (Port 2), with a complex power distribution across various modes. . . . .                  | 131 |
| 8.1 | Summary of group index( $N_g$ ) and propagation loss $\alpha$ for various SiC thin films deposited at various temperatures using PECVD and ICPCVD process obtained from the ring resonators of radius R. . . | 146 |
| 8.2 | Effect of changing waveguide width on propagation loss $\alpha$ and dispersion (D) for SiC thin films deposited using ICPCVD at 150° C and 250° C temperature, obtained from ring resonators. . . . .        | 146 |
| 8.3 | The refractive index of GST and AIST at 1550nm wavelength used in simulations. . . . .                                                                                                                       | 150 |
| 8.4 | Thermal parameters used in FEM simulations. Thermal conductivity(K), Heat capacity ( $C_p$ ) and Density( $\rho$ ) for various materials are reported. . . . .                                               | 150 |

# List of Abbreviations

|              |                                                                         |
|--------------|-------------------------------------------------------------------------|
| <b>ADC</b>   | Analogue-to-digital converter                                           |
| <b>AFM</b>   | Atomic force microscope                                                 |
| <b>AIST</b>  | $\text{Ag}_3\text{In}_4\text{Sb}_{76}\text{Te}_{17}$ phase change alloy |
| <b>AWG</b>   | Arrayed waveguide gratings                                              |
| <b>CMOS</b>  | Complementary metal-oxide-semiconductor                                 |
| <b>CW</b>    | Continuous wave                                                         |
| <b>DCNNs</b> | Deep Convolutional neural networks                                      |
| <b>DWDM</b>  | Dense wavelength division multiplexing                                  |
| <b>EAM</b>   | Electro absorption modulator                                            |
| <b>EBL</b>   | Electron beam lithography                                               |
| <b>EDFA</b>  | Erbium-doped fiber amplifier                                            |
| <b>EOM</b>   | Electro-optical modulator                                               |
| <b>FET</b>   | Field-effect Transistor                                                 |
| <b>FSR</b>   | Free spectral Range                                                     |
| <b>GST</b>   | $\text{Ge}_2\text{Sb}_2\text{Te}_5$ phase change alloy                  |
| <b>InP</b>   | Indium Phosphide                                                        |
| <b>I/O</b>   | Input-Output                                                            |
| <b>ITO</b>   | Indium Tin Oxide                                                        |
| <b>LN</b>    | Lithium Niobate                                                         |
| <b>MAC</b>   | Multiply and accumulate                                                 |
| <b>MVM</b>   | Matrix-vector multiplication                                            |
| <b>MZI</b>   | Mach-Zender interferometer                                              |
| <b>ONN</b>   | Optical neural networks                                                 |
| <b>PCMs</b>  | Phase change materials                                                  |
| <b>PD</b>    | Photodetector                                                           |
| <b>Sb</b>    | Antimony                                                                |
| <b>SEM</b>   | Scanning electron microscope                                            |
| <b>Si</b>    | Silicon                                                                 |
| <b>SiC</b>   | Silicon Carbide                                                         |
| <b>SiN</b>   | Silicon Nitride                                                         |

|                        |           |                                        |
|------------------------|-----------|----------------------------------------|
| <b>SiO<sub>2</sub></b> | . . . . . | Silicon Dioxide                        |
| <b>SLM</b>             | . . . . . | Spatial light modulator                |
| <b>SNN</b>             | . . . . . | Spiking neural networks                |
| <b>SOA</b>             | . . . . . | Semiconductor optical amplifier        |
| <b>SOI</b>             | . . . . . | Silicon on insulator                   |
| <b>TIA</b>             | . . . . . | Transimpedance amplifier               |
| <b>TE</b>              | . . . . . | Transverse electric                    |
| <b>TOPS</b>            | . . . . . | Tera operations per second             |
| <b>VCSEL</b>           | . . . . . | Vertical cavity surface emitting laser |
| <b>WDM</b>             | . . . . . | Wavelength division multiplexing       |

# 1

## Introduction

### Contents

---

|            |                                      |           |
|------------|--------------------------------------|-----------|
| <b>1.1</b> | <b>Motivation</b>                    | <b>1</b>  |
| 1.1.1      | Need for Active materials            | 3         |
| 1.1.2      | Photonic Network design              | 4         |
| 1.1.3      | Exploration of new photonic platform | 5         |
| <b>1.2</b> | <b>Outline and Objectives</b>        | <b>6</b>  |
| 1.2.1      | Objectives                           | 6         |
| 1.2.2      | Outline                              | 7         |
| <b>1.3</b> | <b>Key results</b>                   | <b>9</b>  |
| <b>1.4</b> | <b>Statement of originality</b>      | <b>10</b> |

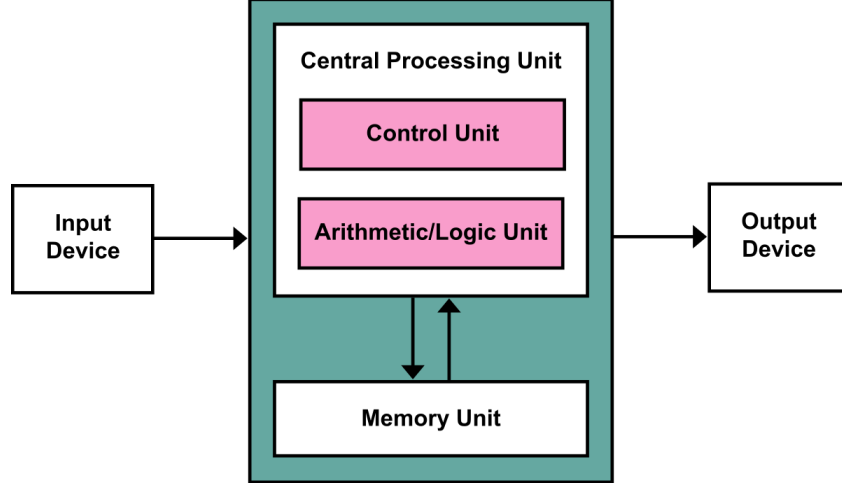
---

## 1.1 Motivation

In March 2016, Lee Sedol, the world's best GO player, lost 4-1 to an artificial intelligence machine programme, AlphaGo [1], developed by Google's Deep Mind team. Although humans lost this challenge to a machine, AlphaGo used more than 1 megawatt of power compared to only 20 watts used by the human brain [2]. While a machine might have been able to match a brain in terms of computing power or even exceed the power of a brain for this task, it was achieved in a power-inefficient manner.

Modern computers are based on the von Neumann architecture, which works efficiently for algorithmic instructions but runs into trouble for complex and abstract tasks such as the GO game [3]. This is because, for complex machine learning tasks, a large amount of data must be shuttled between the physically separated processor

and memory in the von Neumann architecture, as shown in Figure 1.1. This data movement consumes considerable energy and causes a bandwidth bottleneck [4].

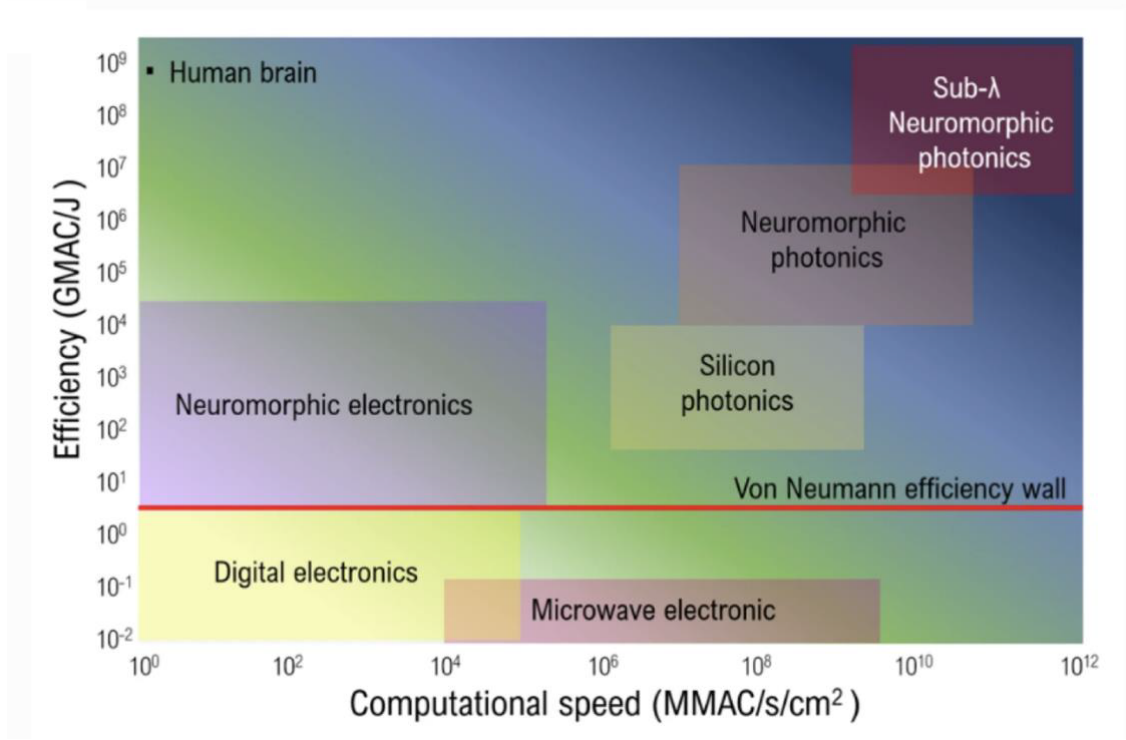


**Figure 1.1:** The block diagram for von Neumann architecture shows different essential components. The data from the input goes through the processing unit, where the program is loaded from memory, computed in ALU, and then stored back into memory and finally sent out. Source: Wikipedia creative commons

Brain-inspired in-memory neuromorphic architectures have been proposed to get away with these bottlenecks. With the slowing of Moore’s law, photonic computing has the edge over electronics for such machine learning architectures due to energy efficiency and superior computational bandwidth offered by wavelength multiplexing techniques [3]. Figure 1.2 shows a qualitative comparison of various computing technologies based on electronic and photonic in terms of computing speed and efficiency, thus showcasing the dominance of photonic neuromorphic computing.

So far, various photonic networks have been proposed but lack the necessary building blocks. Building a photonic computer requires various building blocks such as memory devices, lasers, modulators, photodetector and the photonic network itself [5]. The work herein focuses on such building blocks and ways of integrating them on a platform to achieve an on-chip photonic computing accelerator.





**Figure 1.2:** Comparison of various computing technologies based on electronics and photonics in terms of computing efficiency and computation speed. Source[3]

### 1.1.1 Need for Active materials

Common photonic platforms based on silicon (Si) and silicon nitride (SiN) are passive and offer limited tunability that can be used for implementing photonic hardware. This has necessitated the integration of active function materials with such platforms to achieve better tunability. With the growing need for computational power for machine learning tasks, there has been a growing interest in in-memory computing memories. Phase change materials (PCMs) are strong candidates for photonic memories, as they prove excellent tunability and are non-volatile. This means that once they are programmed, they retain their state.

However, conventional PCMs based on  $\text{Ge}_2\text{Sb}_2\text{Te}_5$  (GST) and  $\text{Ag}_3\text{In}_4\text{Sb}_{76}\text{Te}_{17}$  (AIST), though they have been shown for many photonic applications, including computing [6], suffer from two significant issues. The first is related to their switching speed which is on the order of 10-100 ns [7, 8]. The second drawback

is related to the use of the alloy material itself. The use of special alloys comes with its challenges of compositional integrity. Alloys undergo phase segregation when subjected to thermal cycling [9–11], a million cycles in the case of memory applications. This results in a change in the composition of the material and a degradation in the device’s endurance performance. Another challenge related to alloy compositional integrity is the difficulty in achieving the targeted material composition and maintaining the compositional uniformity over the large area of a wafer [12]. This often requires rigorous optimisation of material growth and deposition conditions.

This has resulted in the need for the exploitation of new active materials that can address the issues related to switching speed and the materials that can overcome the challenges related to the use of alloys. To address these issues, this thesis explores the use of elemental antimony (Sb) for memory applications and to achieve tunability in integrated photonic devices.

### **1.1.2 Photonic Network design**

One of the crucial building blocks of optical computing is the design of the photonic network. This involves arranging the photonic memory devices and the interconnection of these memory elements to perform computation. Thus far, various computing architectures have been proposed. Such hardwares are either based on brain-like architectures, which fall under the category of spiking neural networks (SNNs). The other implementations are GPU-like architectures, which involve a large number of matrix-vector multiplications (MVMs) and are categorised under deep neural networks (DNNs)-based architectures.

Various architectures to achieve both SNNs and DNNs have been demonstrated in integrated photonics. These networks have been shown to achieve various computationally intensive tasks such as image classification and speech recognition. Such photonic implementations have been projected to match the computational

output comparable to or even better than the current state-of-the-art electronic counterparts. However, such implementations have been limited to small networks and do not scale well with the increasing size of networks. Such network designs also do not address the critical issue related to the system's power requirements. The initial implementation of DNNs in integrated photonics was based on a tunable Mach-Zender interferometer (MZI) [13] mesh network involving the coherent addition of light sources. Such schemes of photonic networks have large footprints and require a large number of building blocks and therefore do not scale favourably with increasing the network size. Moreover, using phase shifters and heaters to control the phase of light comes with a considerable energy penalty.

More recent implementations based on the incoherent addition of light using ring resonator weight banks [14] suffer the same problem of the need for an active energy supply to the system to maintain their weights. Phase change memory provides an excellent platform to reduce the need for an active energy supply [15]. However, these networks are limited by photonic network losses and thus limit scalability. Thus, there is a need for the development of novel photonic network designs.

This thesis addresses these challenges by proposing a reduced-rank photonic network, resulting in a less aggressive network scaling with increasing input vectors. Furthermore, a multimode photonic solution is studied to reduce the photonic network losses due to the incoherent addition of light.

### **1.1.3 Exploration of new photonic platform**

So far, photonic platforms based on silicon, silicon nitride, and indium phosphide (InP) have been actively explored for implementing photonic computing. Photonic computing hardware consists of a large number of electronic components such as drivers to run lasers, transimpedance amplifiers (TIA) and analogue-to-digital converters (ADC) offered by silicon electronics [5].

On the contrary, in the optical domain, photonic platforms based on InP provide a number of active building blocks, such as lasers and amplifiers on a chip, which are lacking in silicon. Moreover, other emerging photonic platforms based on lithium niobate (LN) offer excellent modulation [16] owing to strong second-order nonlinearity and pocket effects, which are missing in silicon photonics. Nevertheless, silicon photonics remains a popular choice for implementing systems on chips due to the compatibility of silicon with CMOS technology. In reality, no single photonic platform provides all the functionalities. This has necessitated the cointegration of various such photonic platforms.

Ideally, one would like to use the active components offered by the InP platform and use the excellent linear optical properties offered by silicon photonics. However, working with such crystalline photonic platforms presents the challenge of cointegration due to the large lattice mismatch of silicon with InP and requires a complex fabrication process for cointegration [17].

Some of these challenges can be addressed by moving to another CMOS-compatible platform, such as silicon carbide. Silicon carbide has been used in power electronic applications, and in the photonic domain, crystalline silicon carbide was used to create the first LED in 1908. Recently there has been interest in CMOS-compatible amorphous silicon carbide platforms for photonic application[18]. This thesis addresses the opportunity offered by such a low-temperature amorphous silicon carbide platform, develops a low-loss platform, and explores the integration of such a platform with phase change material to increase the tunability.

## 1.2 Outline and Objectives

### 1.2.1 Objectives

This thesis addresses the challenges of building a scalable and fast photonic network. These hurdles are overcome through the exploration of functional materials for

ultrafast memory operations, proposing a novel architecture for scalable photonic networks and investigating the integration of silicon carbide as a potent photonic platform for memory devices.

First, a novel phase change integrated photonic memory based on elemental antimony with ultrafast write and read speeds up to 500 MHz is demonstrated. Next, the challenges related to the scalability issue of photonic networks are addressed. Departing from the conventional approaches of increasing the network sizes and thus increasing the number of control parameters, a reduced rank approach is explored, which results in a less aggressive scaling of the photonic network. To demonstrate the applicability of this design, convolution and image filtering tasks are performed.

Further, the challenges of cointegration of photonics and electronics are explored by proposing a low-temperature grown CMOS-compatible silicon carbide photonic platform. The photonic properties of this novel platform are explored by performing linear and nonlinear experiments, thus opening up many opportunities for further cointegration of such platforms to achieve maximum functionality on-chip.

### 1.2.2 Outline

The content of this thesis is outlined as follows:

- **Chapter 2:** A literature review has been carried covering various available memory technologies for integrated photonic computing. Next, various implementations of photonic networks are reviewed. Finally, a survey of the advances in various photonic platforms and the functionalities offered by such platforms are reviewed.
- **Chapter 3:** A theoretical background of the integrated photonic devices used in this thesis is presented. The chapter discusses the evolution of modes in waveguides. The working principle of a grating coupler is briefly discussed.

Finally, a background in use of a ring resonator for four-wave mixing to extract the nonlinear refractive index is developed.

- **Chapter 4:** Discussion of the device preparation and the description of the experimental setup used in various experiments. The first part of the chapter covers the design and fabrication of photonic devices used in the thesis in detail.

The experimental detail section contains the details of the transmission measurements used to characterise photonic devices. A pump-probe experimental setup used to switch PCM, then an experimental setup used to perform photonic matrix multiplication experiments is discussed. Finally, a description of the experimental setup used to conduct nonlinear experiments in the silicon carbide platform is discussed.

- **Chapter 5:** Experimental development and demonstration of antimony-based photonic memory and its application for signal processing is discussed in this chapter. The first section covers switching of antimony to achieve over eight memory levels, and then volatile characteristics are studied. Finally, an application in signal processing is demonstrated.
- **Chapter 6:** Development and experimental demonstration of a scalable large-scale photonic network using the reduced rank approach is demonstrated. Using the reduced rank technique, edge detection and classification tasks are achieved. A discussion on the scalability of such an architecture is presented.
- **Chapter 7:** The design and simulation of a multimodal device to improve the scalability of the photonic networks are discussed. This chapter consists of simulation results for developing an asymmetric Y-coupler to improve the efficiency of the addition of incoherent light. A study on the fabrication tolerance of such a device is presented.

- **Chapter 8:** The development of a new photonic platform based on silicon carbide is presented in this chapter. A low-loss amorphous silicon carbide platform is showcased, and its nonlinear optical properties are studied. Finally, the integration of PCM memory devices is presented.
- **Chapter 9:** Future works and potential impact of the work demonstrated in the thesis are discussed.

### 1.3 Key results

The key results of this thesis are related to the development of building blocks for improving the speed, scalability, energy efficiency and integration of photonic computing.

In this regard, the thesis has focused on the development of the following :

- An ultrafast multilevel photonic memory based on elemental antimony has been demonstrated.
- To tackle the scalability challenges of photonic networks in terms of energy and footprint, a reduced-rank photonic network has been demonstrated.
- A foundry-fabricated chip has been designed, which consists of integrated photodetectors and modulation using such as electro absorption modulator (EAM). Efforts have been made to make an electro-optic control system for computing platforms.
- To tackle the power loss due to the incoherent addition of light, a multimode Y-coupler design is studied, which improves the coupling efficiency by up to 10% for large networks.
- A low-temperature process photonic platform of amorphous SiC with ultra-low loss and excellent nonlinear optical properties has been demonstrated.

## 1.4 Statement of originality

This thesis entails the original work performed by the author during their time as a DPhil student at the University of Oxford. All the results and experiments presented here have been obtained by the author solely. Wherever relevant, the collaborators' contributions have been acknowledged; this includes the fabrication of photonic computing chips, which have been fabricated by the foundry IMEC and the fabrication of silicon carbide chips fabricated at TU Delft.



# 2

## Integrated Photonic Neuromorphic Accelerators: A Review

### Contents

---

|            |                                       |           |
|------------|---------------------------------------|-----------|
| <b>2.1</b> | <b>Memory Cell</b>                    | <b>12</b> |
| 2.1.1      | Electric Field Based                  | 13        |
| 2.1.2      | Opto-Mechanical Device                | 15        |
| 2.1.3      | Thermo Optic Effect                   | 16        |
| 2.1.4      | Phase Change Material                 | 16        |
| <b>2.2</b> | <b>Photonic Network Architectures</b> | <b>19</b> |
| 2.2.1      | Spiking Neural Net architecture       | 19        |
| 2.2.2      | Deep Learning Architecture            | 22        |
| <b>2.3</b> | <b>Photonic Platforms</b>             | <b>29</b> |
| 2.3.1      | Silicon-Platform                      | 30        |
| 2.3.2      | InP-Platform                          | 32        |
| 2.3.3      | Silicon Nitride-Platform              | 33        |
| 2.3.4      | Promising photonic platforms          | 33        |
| <b>2.4</b> | <b>Critical Analysis and Summary</b>  | <b>35</b> |

---

Building a photonic neuromorphic accelerator which can outperform its electronic counterpart and meet the computational demands of the future requires some key building blocks. Such building blocks include the memory element to perform in-memory computing. It involves arranging these memory elements in a network to meet the interconnectivity requirements of a neuromorphic system and finally, a unified platform which can co-integrate optical and electrical components to build a system. This chapter presents a review of such building blocks to build an integrated photonic neuromorphic accelerator.

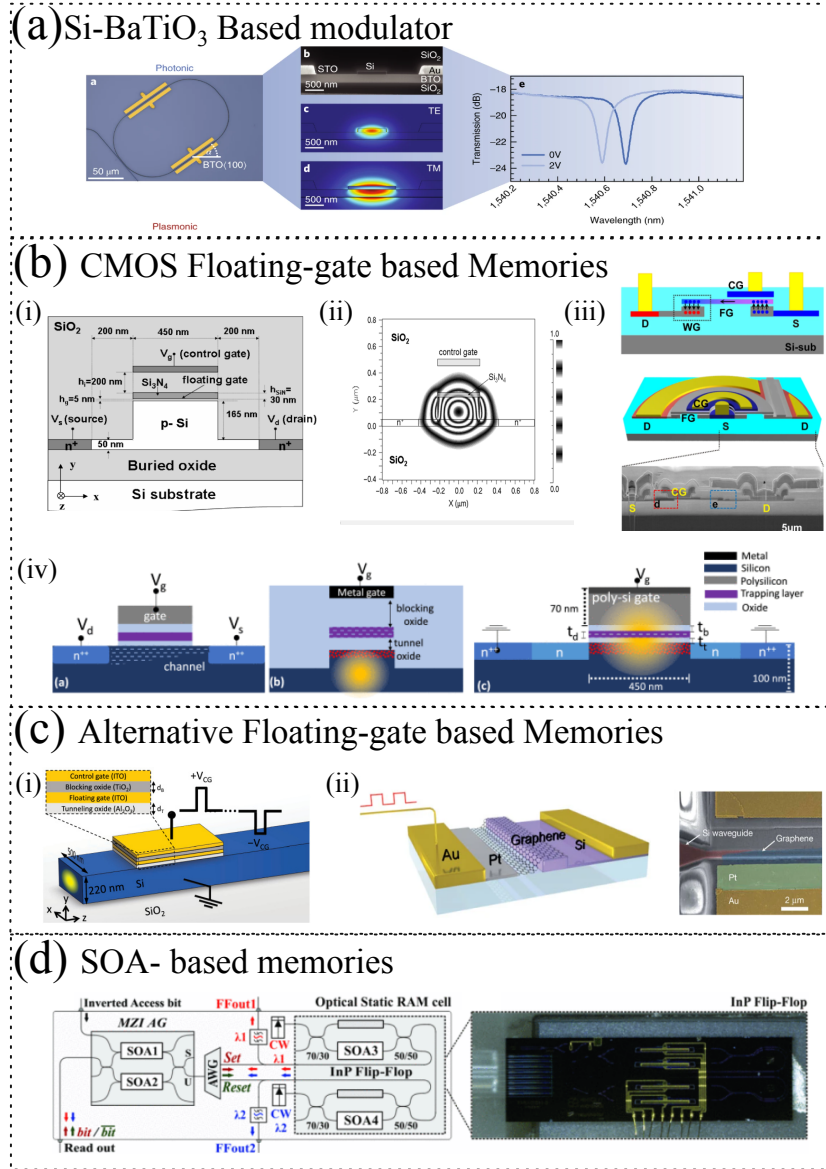
## 2.1 Memory Cell

The characteristics of a memory cell to build a neuromorphic accelerator are unique. One needs a memory device with a fast weight-set and readout speeds (close to GHz). The memory element should demonstrate high cyclability (up to million cycles) with good repeatability (4-5 bit precision) and low memory drifts (stability over decade) [19]. Ideally, such memory cells should offer long-term retention at high temperatures to allow for integration with CMOS processing [20]. Various volatile and nonvolatile methods of creating photonic memory cells or weight banks for neural networks have been demonstrated. On a physical scale, photonic memories work on the principle of introducing perturbation to the refractive index of the light propagation medium. Such change in refractive index can be either in the real part of refractive index  $n$ , which changes the phase of light or by introducing an amplitude change by tweaking the imaginary constant  $k$  of the material and often both  $n$  and  $k$ . Working of the memories based on amplitude modulation is straightforward and can be read directly as the change in light intensity. On the other hand, phase modulation can be read using light interference techniques, as in an MZI.

Some common optical modulation methods are based on effects in semiconductors such as the Pockels, Kerr, Franz Keldysh, free carrier plasma dispersion and thermo-optic effects [21]. Besides these effects intrinsic to the waveguiding medium, optical modulation can also be achieved by using functional materials such as phase change materials and perovskites.

The following sections discuss various methods of achieving modulation of light and generating photonic memories for neuromorphic computing on an integrated photonic platform reported in the literature. The terms *optical modulation* will be used for volatile schemes of introducing change in refractive index, and *photonic memory* will be used to describe their nonvolatile counterpart.

## 2.1.1 Electric Field Based



**Figure 2.1:** Electric-field-effect based on photonic memories. (a) The strong Pockels effect of BaTiO<sub>3</sub> is used to create an integrated silicon photonic memory. Applying voltage results in a change in the effective refractive index, thus changing the resonance wavelength. Source: [22] (b) CMOS-based floating gate memories (i) schematic of the photonic memory with SiN as a floating gate, (ii) simulations showing the electric field interaction of optical mode. Source: [23] (iii-iv) demonstration of nonvolatile multilevel photonic memories with poly-silicon as a floating gate. Source: [24, 25] (c) Photonic memories based on alternative floating gates (i) demonstration of photonic memory based on ITO as a floating gate. Source: [26] (ii) demonstration of graphene-based floating gate memory. (d) Implementation of semiconductor optical amplifier-based photonic memory on an InP platform. Source: [27]

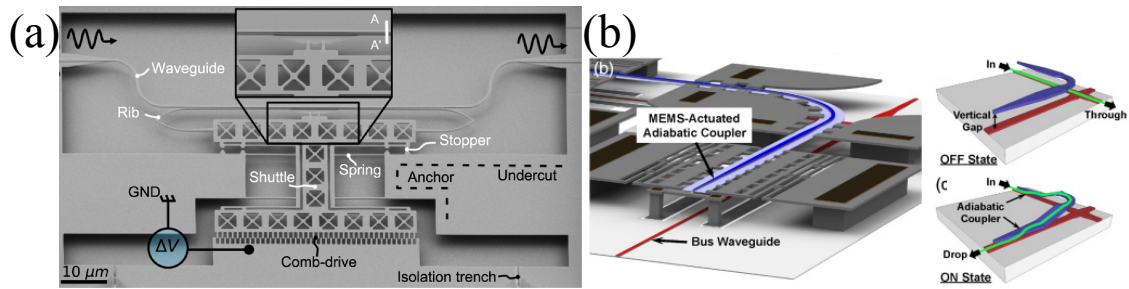
Electric field effects such as Pockels and Kerr have been demonstrated for optical modulation [16]. Non-centrosymmetric materials such as LN show strong Pockels effects [28]. Applying electric field results in a change in the refractive index of the waveguide. Modulation up to 100s of GHz has been demonstrated on such a platform. However, due to the presence of centrosymmetry, silicon does not demonstrate the Pockels effect. By integrating thin film material such as BaTiO<sub>3</sub>, Abel et al. [22] demonstrated strong Pockels effect-based modulators, Figure 2.1a.

Soref et al. [21] demonstrated free carrier injection can reduce the refractive index in silicon. They further showed other electric field effects, such as Keldysh and Kerr effects, are weak for silicon. Using the charge-induced dispersion principle, Lipson et al. [23] proposed a nonvolatile photonic memory by storing charges in a silicon dioxide floating gate, Figure 2.1b(i-ii). Song et al. [24] demonstrated a photonic memory cell with multilevel performance using polycrystalline silicon floating gate, Figure 2.1b(iii). They use a ring resonator to enhance the change in refractive index. Similarly, Olivares et al. [25] used light-doped polycrystalline silicon as a charge-trapping layer to reduce the voltage requirements due to metal waveguide insulation layers, Figure 2.1b(iv). Since these works, other CMOS-compatible processes have also been explored by storing charge in silicon nitride and hafnium oxide layers. Recently, more novel approaches have been proposed using ITO [26] and graphene [29, 30] to act as charge accumulation layers, Figure 2.1c.

Field-effect transistor (FET) based optical memories have been predicted to exhibit long-term charge storage at ultra-low programming energies of tens of pico-joules. However, the practical implementation of these memory cells uses resonant structures due to the weak plasma dispersion effect, which results in a large footprint of the devices. Moreover, the programming speed of these devices has been limited to microseconds, which needs to be further improved to compete with their electronic counterparts.

Another type of optical weighting technique that has been shown is based on a semiconductor optical amplifier (SOA) [27, 31–34]. Using a flip-flop configuration in an MZI array configuration, upto 5 Gb/s write and read speeds have been demonstrated. Such systems have been implemented on-chip in InP-based SOA devices [27], Figure 2.1d. They have demonstrated an energy penalty of about 2.3 dB for a write and 0.5 dB for a read operation. One drawback of the memory devices based on SOA is their large footprint ( $6 \times 2 \text{ mm}^2$ ) and use of multiple SOA. This limits the scalability of such a system.

### 2.1.2 Opto-Mechanical Device



**Figure 2.2:** MEMS-based memories. (a) Using MEMS integrated with a photonic waveguide, a phase shifter is implemented. Source: [35]. (b) A MEMS-based optical switch. By changing the coupling distance of adiabatic couplers, over 50 dB extinction ratios can be achieved. Source: [36]

Opto-mechanical memory cells introduce the perturbation in the guided mode effective index by mechanically changing the optical coupling between waveguides. This can be achieved by integrating MEMS structures with waveguide structures to achieve a tuneable phase shifter [35], Figure 2.2(a). Seok et al. [37] have shown photonic switches using MEMS-based design. By changing the vertical distance between adiabatic couplers, they showed over 50dB extinction ratio, Figure 2.2(b). Bageheri et al. [36] show by applying buckling forces, the coupling of a ring resonator can be changed, thus creating reversible nonvolatile memory elements.

Optomechanical devices provide for low loss monolithic integration solutions for achieving memory applications with small footprints [38]. However, these devices are limited in their operation speeds to the device's mechanical frequencies.

### 2.1.3 Thermo Optic Effect

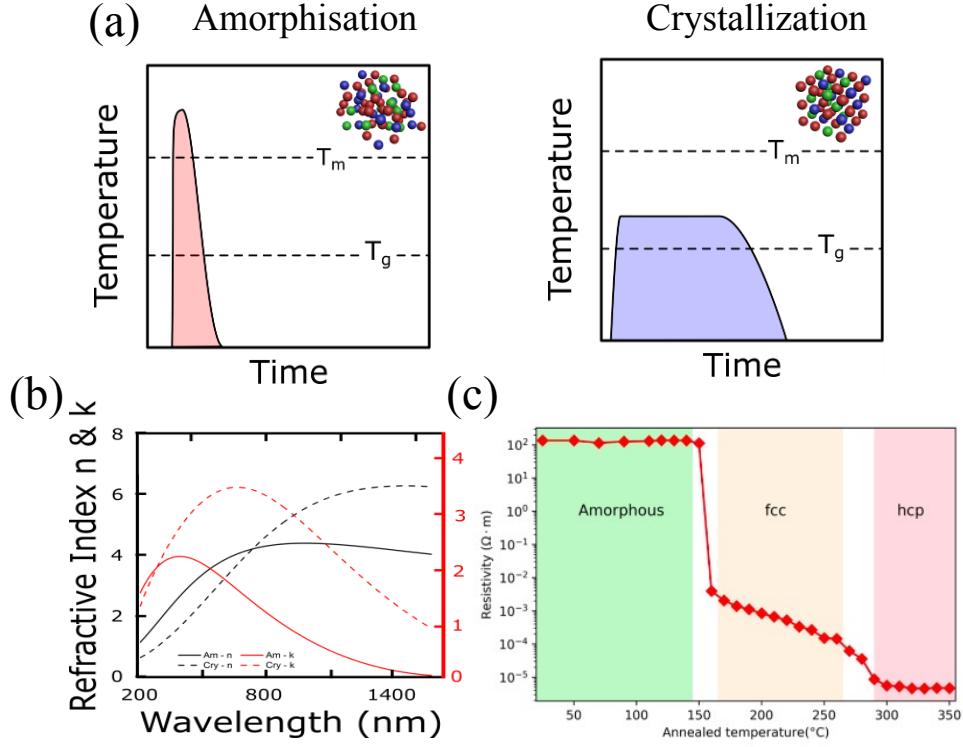
The refractive index of waveguiding material depends strongly on the temperature. Selective heating of a waveguide introduces a phase delay. This phase delay and interference effects can be used for amplitude modulation of light for a coherent light system. This weighting method has been used by Shen et al. [13] to implement a deep neural network. Such a modulation scheme can also be implemented in a resonant structure such as micro-ring resonators.

Ring resonators have a high sensitivity to such variations in the refractive index. Therefore, micro heaters can be used to introduce a phase delay and hence used to modulate the signals. Tait et al. [39–41] used microring resonators coupled with micro heaters to demonstrate photonic weight banks. By varying the current through the heater, they demonstrated over 5-bit data storage.

Such modulation techniques based on the thermo optic effect suffer from high energy requirements to power the heaters. Moreover, implementing a large-scale network requires many I/Os to control the heaters. The variation of temperature over the chip due to thermal crosstalk requires additional calibration steps and is, therefore, a challenge.

### 2.1.4 Phase Change Material

Phase change materials based on chalcogenide glasses have found application in data storage in CD and DVD drives [42]. However, recently they have been demonstrated for integrated photonic applications. Materials such as  $\text{Ge}_2\text{Sb}_2\text{Te}_5$  (GST) and  $\text{Ag}_3\text{In}_4\text{Sb}_{76}\text{Te}_{17}$  (AIST) have been shown to have a fast switching speed in tens of nanoseconds [43]. They show a high cyclability of up to 10 million cycles [44].

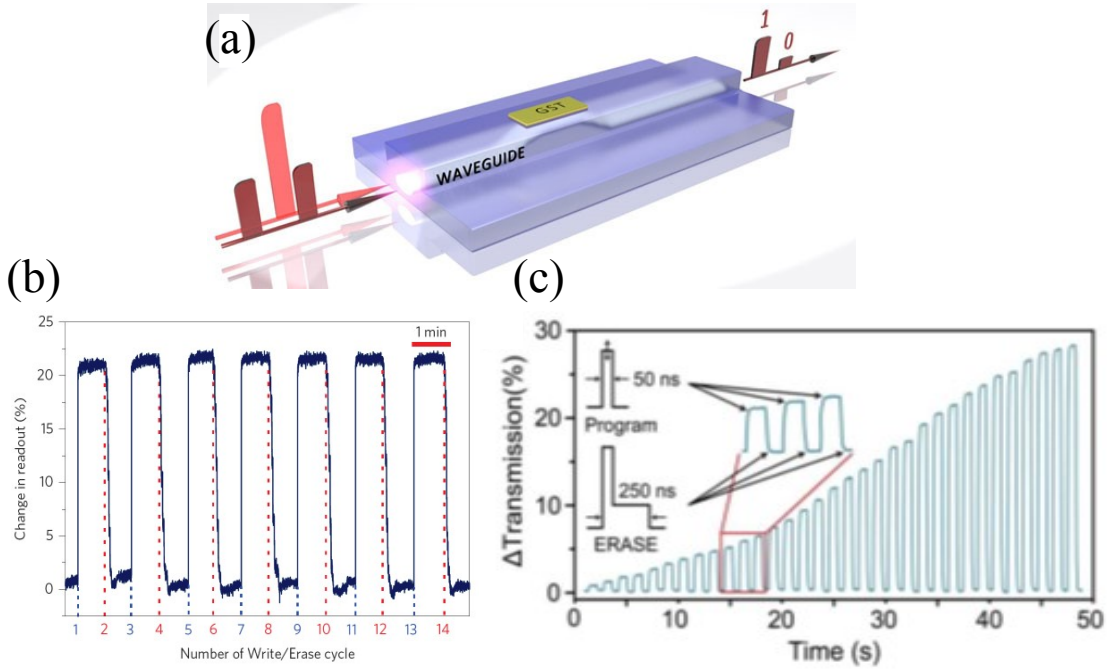


**Figure 2.3:** Phase change materials (GST) based photonic memories. (a) Heating the material above its melting point and rapid cooling results in amorphisation, while heating above the glass transition temperature and holding at this temperature with slow cooling results in crystallisation. (b-c) Amorphous and crystalline phases of GST show different optical and electrical properties, respectively.

The physical properties of phase change materials, such as the electrical conductivity and dielectric constant, cannot be described by purely ionic or covalent bonding. A special bonding mechanism, ‘metavalent’ bonding, has been coined to describe the physical properties of such materials[45]. These materials have properties intermediate of both metals and covalently bonded structures. Metalvalent bonds have been defined as multi-centre bonds with less than two electrons shared between bonding atoms[46].

GST exist in an amorphous state, metastable FCC and stable hexagonal crystal structure. In the FCC crystalline state of GST, Te occupies anion positions while Ge and Sb take octahedral vacancies. In the hexagonal phase, blocks of either Ge or Sb from alternative layers with Te anion. AIST is an alloy of Sb and Te with Ag and In impurities. Sb Te exists as the rhombohedral alternating structure of  $Sb_2Te_3$

and two Sb layers with Ag and In impurities occupying the Te position. [47] Phase change materials (PCMs), when heated above the recrystallisation temperature, can undergo recrystallisation; similarly, when the material is heated over melting point and cooled at rates faster than  $10^9$  °C/s can result in amorphisation [48, 49], as illustrated in Figure 2.3a. Conventional phase change materials such as GST and AIST have different refractive indexes and electrical conductivity [50]. In terms of refractive index, GST has different absorption coefficients, Figure 2.3(b-c). Therefore amplitude modulation can be achieved [7, 51] Figure 2.4(a-b). Up to 32 memory levels can be achieved using single-shot programming methods, as shown in Figure 2.4(c) [8].



**Figure 2.4:** (a) Schematic of a phase change memory using waveguide demonstrating (b) reversible switching. Source:[7], (c) Up to 32 memory levels can be achieved using single-shot programming methods. Source:[8]

Recently, there has been an interest in a new type of low-loss phase change materials such as  $\text{Sb}_2\text{Se}_3$  and  $\text{Sb}_2\text{S}_3$ , which show only change in the real part of the refractive index [52, 53]. These materials are of significant importance for introducing ‘phase delays’.



### **Advantages of PCMs**

Phase change material shows a nonvolatile performance. They show high stability over a long time [43]. Multiple memory levels can be easily achieved by growing the amorphous states [54]. Therefore allowing for multilevel memory operations, thus being suitable for photonic analogue in-memory computing.

### **Disadvantages of PCM**

Current PCMs are an alloy of special elements; this results in phase segregation [9] and thus hampers cyclability and long-term reliability. Current PCMs show strong thermo-optic effects, which slow down readout signals and thus limit the speed of operations [55]. Special alloys also result in challenges related to film homogeneity, thus possessing scalability challenges.

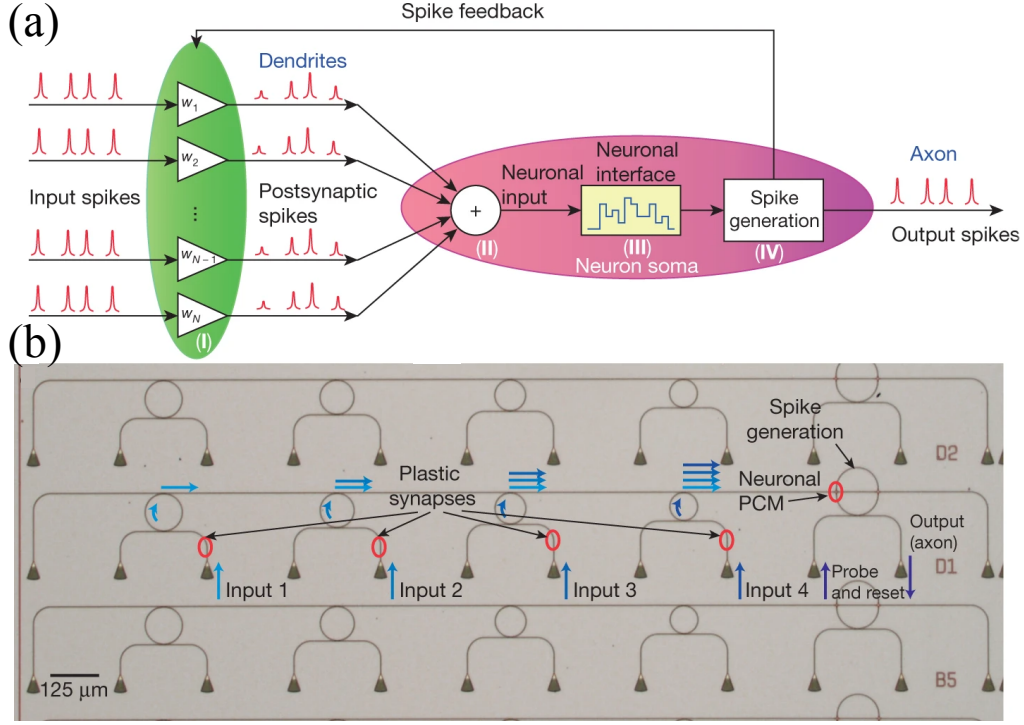
## **2.2 Photonic Network Architectures**

Optical Neural Networks (ONN) have been used to achieve many tasks that can be achieved by their electronic and von Neumann counterparts. Recently they have been shown to perform intensive jobs such as image recognition, handwriting recognition, and principal component analysis [40] [56]. These architectures have also been used to solve computationally intensive tasks such as subset sum problem [57], prime factorisation [58], and travelling salesman problem. Photonic computing architectures can be divided into Spiking Neural Networks (SNN), which are based on mimicking the brain-like computation methods and Deep Neural Network (DNN) or the GPU-based hardware.

### **2.2.1 Spiking Neural Net architecture**

Feldmann et al. [59] demonstrated an SNN based on PCMs. Optical PCMs such as GST can change their phase from amorphous to crystalline using the high enough power of optical pulses. Therefore, using the difference in the optical transmittance

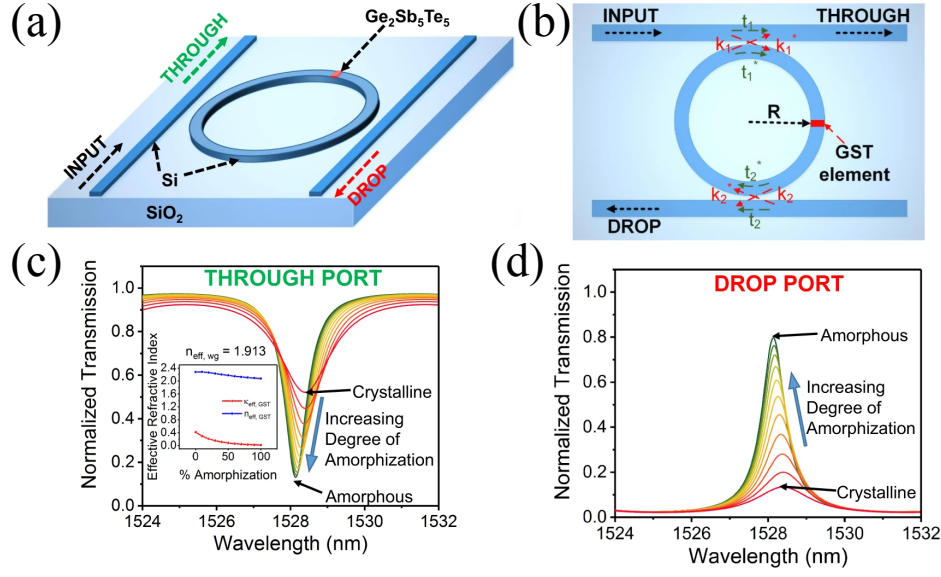
of the two phases, they created a ‘spike’ event whenever the sum of powers of input pulses is enough to carry out this phase transformation.



**Figure 2.5:** All optical spiking neural network (a) Illustrative model of an SNN (b) All-optical spiking neural network using ring resonators for wavelength division multiplexing (WDM) and PCMs. Input signals of different wavelengths are modulated on the chip. The plastic synapses multiply the weights, and the intensity of different wavelengths of light is added using a ring resonator. If this is above the threshold for switching the PCM, there is a spiking output. Source:[59]

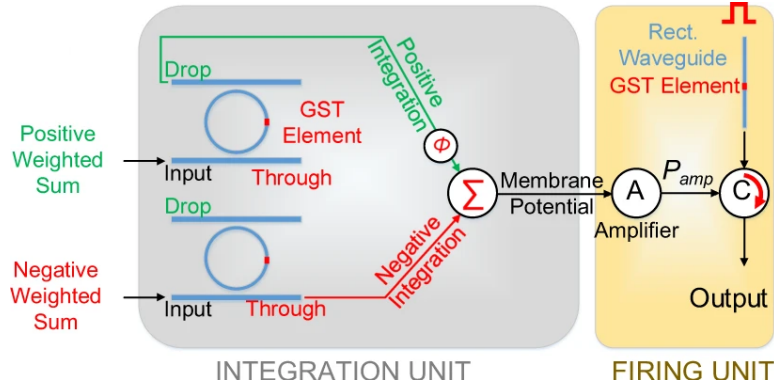
Using this property of PCMs, they could mimic the spiking event optically and eventually train their model for pattern recognition in supervised and unsupervised learning methods. Another similar work on integrating and firing SNN has been demonstrated by Chakraborty et al. [60] using PCMs. Here they use the PCMs combined with a ring resonator which is such as an optical bandpass filter. Using PCMs, they tune the allowed wavelength and create a spike event by tuning and detuning a particular wavelength.

One advantage of their model is that they can accommodate negative weights. As seen in Figure 2.6, the drop and through ports have opposite transmission characteristics. Hence by changing and adjusting which port feeds to the network,



**Figure 2.6:** Demonstration of spiking neural network based on ring resonators and PCM. By changing the phase of the PCM, the critical coupling or resonance wavelength can be controlled. a) and b) Schematic diagram of a ring resonator with GST as PCM. c) and d) change in transmission on changing the phase of PCM in the through port and drop port. Since both ports have different responses, the same device can be used for positive and negative weights. Source:[60]

they can mimic the structure as positive and negative weight, as explained in Figure 2.7.



**Figure 2.7:** By changing the port of the ring resonator connected to the circuit, both positive and negative weights can be achieved. As seen in Figure 2.6, Since both the drop and through have different responses to the amorphisation pulse, the device can be used for both positive and negative weights. Source:[60]

Other ways of preparing spiking neurons have been based on spiking laser sources. Recently, Robertson et al. [61] have shown pattern recognition using Vertical-Cavity

Surface-Emitting Laser (VCSEL). They have shown multiple input integration in a leaky-integrate and fire model of SNN at an ultrafast speed of around 200 ps.

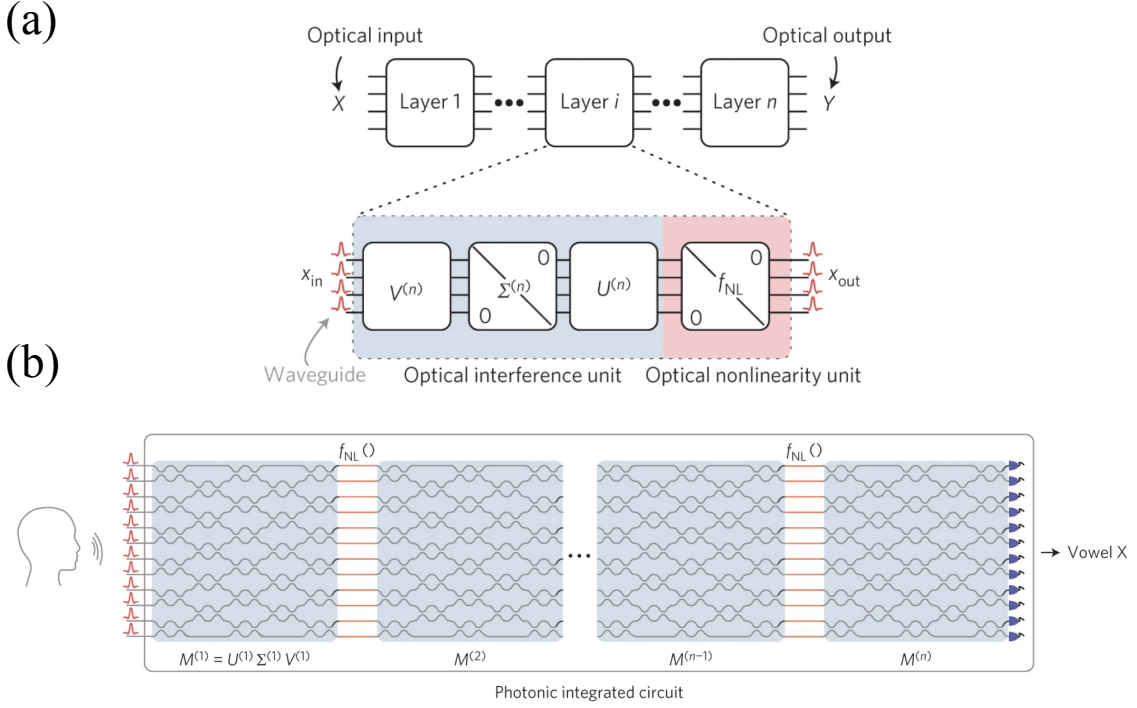
Besides photonic implementations based on SNN, other brain-inspired computing architectures using the Pavlovian associative learning principles are worth mentioning. Tan et al. [62] demonstrated a backpropagation-free learning architecture using an associative learning element based on PCMs. They demonstrated their associated learning element can be trained to recognise an image of a cat. Using the input and teacher input as an image of a cat, their system could learn the common features of a cat image and distinguish it from the image of a dog.

### 2.2.2 Deep Learning Architecture

#### Tuneable MZI mesh

The deep neural network architecture is based on matrix-vector multiplication. Free space optical computing based on interference has been shown as early as the 1980s [63, 64]. The first implementation of all-optical DNN based on integrated photonics has been demonstrated more recently by Shen et al. [13]. Using the Reck et al. [65] work on  $N \times N$  matrix factorisation using passive components such as splitters and interferometers, they built an optical network to perform such a large multiplication network as shown in Figure 2.8. Further integrating these splitters and interferometer with the thermo-optical modulator, they adjusted the synaptic weights and showed a 4-layer ONN. Even though the demonstration of the first on-chip ONN is innovative, their implementation uses a large number of active components (phase shifter) to perform the calculation, which increases the power budget of computation.

Similarly, other designs have been implemented using the coherent addition of light sources [66]. Xu et al. demonstrated a regression task on a chip using a similar coherent MZI addition of light. They placed two modulators in a push-pull



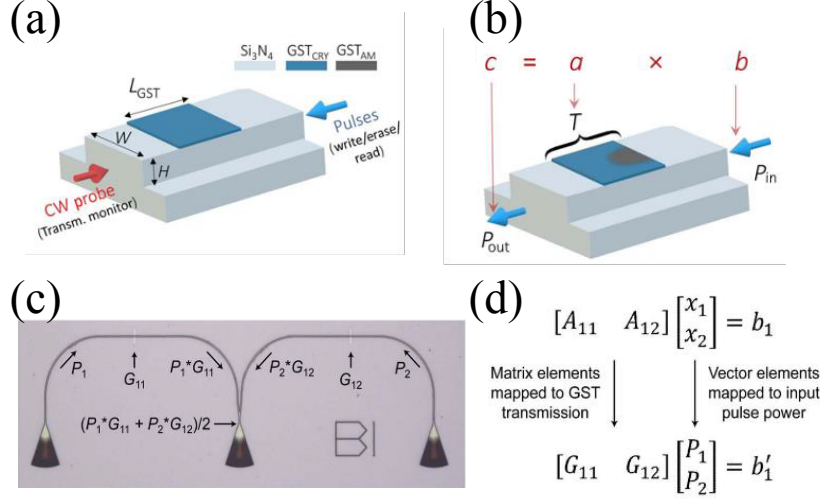
**Figure 2.8:** (a) Demonstration of the first on-chip all-optical deep neural network with  $N$  hidden layers. Using fundamental optical elements such as an optical interferometer and splitters, the weight matrix of a single layer is broken down into its single-value decomposition matrix. Using the different channels of waveguides as input vectors, the matrix multiplication operation is performed in the photonic circuit. The output of each layer is passed through a nonlinear optical unit, and the resultant signal is passed through another layer. (b) An  $N$ -layer DNN has been constructed for vowel recognition from speech data. Source:[13]

configuration to modulate the weight without introducing phase delay to achieve a dot product and show an image regeneration task.

Further works have been reported using these reconfigurable MZI mesh designs. These works have been focused on improving the scalability of the MZI mesh designs. Zhu et al. demonstrated one such design using the MZI mesh design [67]. They used a diffractive network to achieve parallel Fourier transformation and achieve convolution operation. One of the benefits of their design is their networks scale  $N$  times as compared to  $N^2$  for conventional DNN designs. Another way of improving the scalability of the MZI mesh array design has been demonstrated by Feng et al. [68]. They use the hadmard transformation technique to reduce the number of training parameters and showcase a  $7\times$  reduction in parameters while performing

image classification tasks.

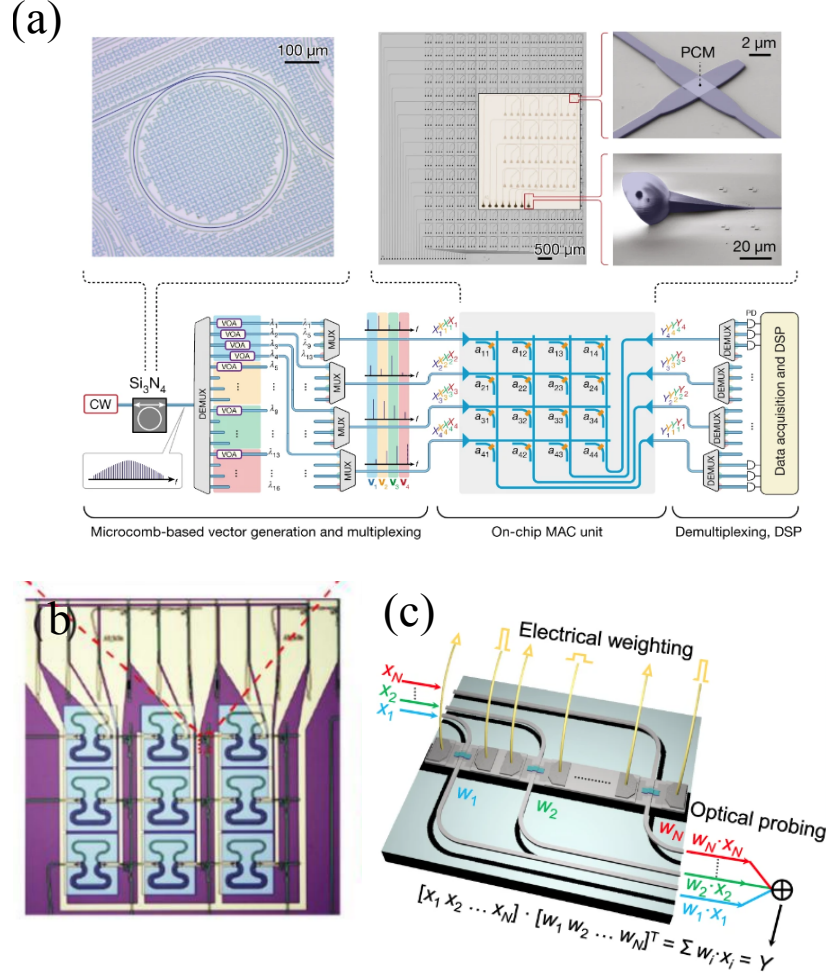
### Phase change memory



**Figure 2.9:** a) Photonic waveguide model to achieve photonic multiplication using GST. Using a Pulse signal to change the phase of the GST, there is a change in transmittance. This is equivalent to multiplication, as explained in b). c) represents combining the multiplication scheme shown in a) and b) for vector-vector multiplication as explained by mathematics in d). Using a two-input one output device, the intensity of individual multiplication cells is added. Source:[69]

To circumvent the usage of active devices and decrease the energy budget, matrix multiplication using nonvolatile optical memories has been suggested. In particular, one method which has shown immense promise is based on PCMs such as GST. Rios et al. [69] have shown photonic multiplication using PCMs. By changing the phase of the material from amorphous to crystalline, they achieve modulation in the transmission of the input signal. Furthermore, using this principle, they have demonstrated vector multiplications, shown in Figure 2.9.

As expected, the resolution of PCMs-based multiplication depends on the intermediate phases that can be achieved, i.e., the intermediate between amorphous and crystalline states. It is worth mentioning that all neural network multiplications are low-precision calculations. Over the years, up to 5-bit or 32 levels of PCM photonic memory [8] has been demonstrated. Feldmann et al.[15] have demonstrated



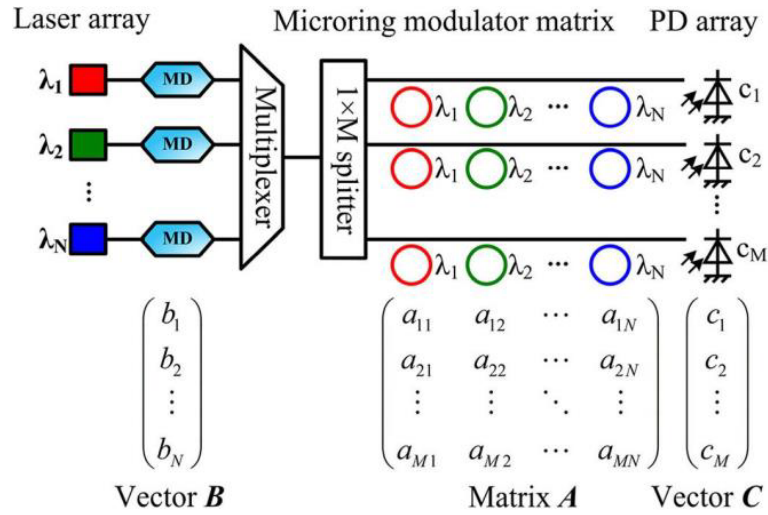
**Figure 2.10:** Phase change materials-based deep neural network implementations. (a) Photonic neural network implementation exploiting WDM using frequency comb to modulate input signals and using a PCM-based crossbar array structure to carry out matrix-vector multiplication. Source:[15], (b) Photonic architecture using phase shifters to reduce I/O for PCM switching for computation. Source:[70], (c) Using the integrated photonic heaters, PCM weights can be programmed electrically while reading optically. Source:[71].

a large-scale parallel photonic computing core using these preliminary results in phase change photonic computing, Figure 2.10(a). Using the benefits offered by a crossbar array and using an on-chip frequency comb, they demonstrated parallel computation using wavelength division multiplexing (WDM) techniques, demonstrating over  $10^{12}$  multiply-and-accumulate (MAC) operations per second. They demonstrated computationally intensive tasks such as edge detection and convolution on MNIST hand digit classification.



One of the limitations of such a PCM-based system is associated with switching PCM elements to achieve modulation. All optical switching of PCM has inherent benefits of low switching energy (300-1000 pJ). Therefore, Li et al. [70] demonstrated an on-chip photonic network on an SOI platform to overcome this. Using on-chip MZI heaters, they could concentrate input power to switch PCM memory cells, Figure 2.10(b). This reduced the number of input-output requirements of the system. They demonstrated pattern recognition tasks with their small  $3 \times 3$  matrix implementation. More recently, another work by Zhou et al. [71] has used an electro-optical approach to switch PCM memory cells. Using a doped silicon heater integrated into the photonic network, they could achieve write electrically, read optically approach to demonstrate convolution for classification tasks, Figure 2.10(c).

### Ring Resonator Modulators



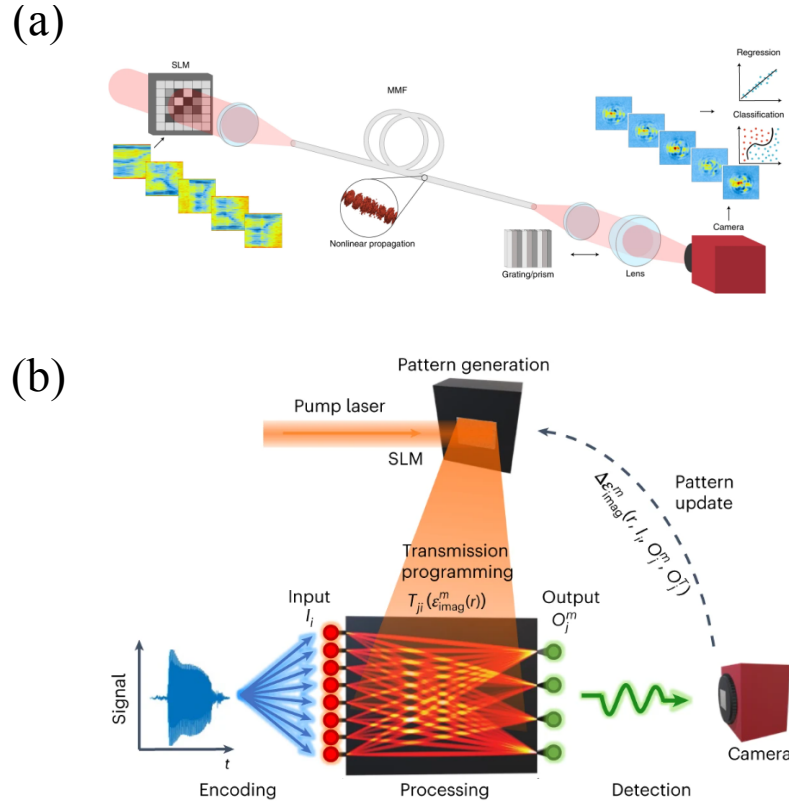
**Figure 2.11:** DNN model showing vector-matrix multiplication. The input lasers (input vector – B) are distributed equally to the Micro-ring modulator module (Matrix A) using a multiplexer and splitter, as shown in Figure 2.11. Each ring, responsive to different wavelengths, can be individually set to the weight matrix. The photodetector sums the intensity of the light to get the output vector C. Source:[14]

The Prucnal group at Princeton University has shown other non-coherent methods of implementing DNNs. Building upon their previous works on micro-ring



weights [14], they have demonstrated a large-scale neuromorphic network with WDM capabilities to enhance the parallel computation [72], Figure 2.11. Using this architecture, they have shown a large-scale photonic integrated circuit and performed machine learning tasks such as principal component analysis and pattern recognition [67, 73].

### Spatial Light Modulation



**Figure 2.12:** SLM-based photonic computing implementations. Using SLM for a linear layer of a DNN and projecting this as an image through a nonlinear fibre first layer of DNN was implemented. Source [74] (b) Lithography-free implementation of a photonic neural network. By projecting high intensity modulated pump source on an III-V thin film, the photonic network was trained for vowel recognition. Source:[73].

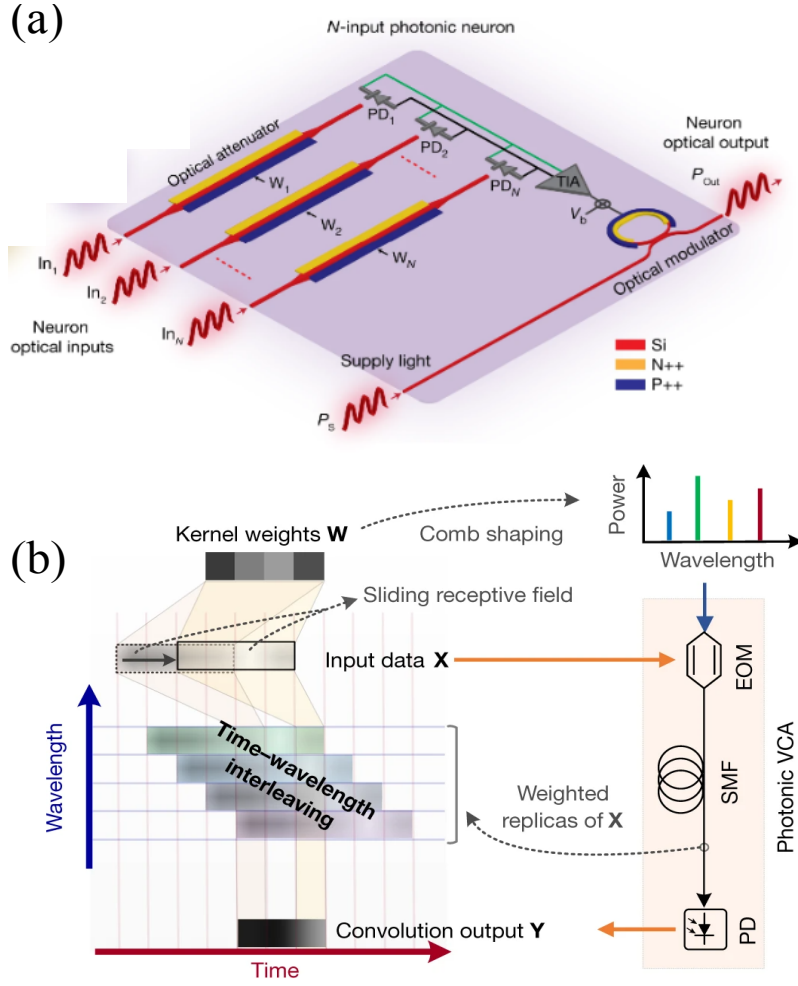
More recently, other methods of photonic computing have been achieved using spatial light modulators (SLM) as weighting layer [73, 74]. Tegin et al. [74] used an SLM as a weighting matrix and projected the output image onto a nonlinear fibre to implement a single layer of a DNN, as shown in Figure 2.12(a).

In a recent work by Wu et al. [73], they used an SLM to project a pump laser onto an InGaAsP thin film to generate a reconfigurable network to achieve a lithography-free method of achieving deep neural network. By projecting a pump image onto the III-V layer, they could change the refractive index of the layer, therefore, changing the interaction of the input signal in the network. By changing this projected image, they could train their network to achieve vowel recognition tasks with high accuracy as shown in Figure 2.12(b).

### **Direct Electro-optic Modulation**

Another way of implementing neural networks has been based on direct modulation of light using an electro-optic light attenuator [75]. Ashtiani et al. built a network with a light attenuator to achieve multiplication. Further, they routed the output to p-i-n photodetectors and added the photocurrent to achieve the accumulation step, performing MAC operation as shown in Figure 2.13(a). Their implementation involves multiple electrical-optical conversions. While their implementation is relatively straightforward, the scalability of the network is limited due energy penalty associated with such electro-optic conversions. Moreover, the large footprint of p-i-n photodetector and attenuators results in a large footprint.

Another work based on direct modulation of light has been demonstrated by Xu et al. [76]. They use temporal, wavelength and spatial degrees of freedom to achieve more than ten tera operation per second (TOPS). Direct modulation of frequency comb light using a spectral comb shaper to represent weight matrix and then with temporal modulating the light using electro-optic modulators to represent the input vectors. They could perform the MAC operations using a dispersive fibre to achieve temporal-wavelength intermixing and summing the output at a fast photodetector, as shown in Figure 2.13(b).

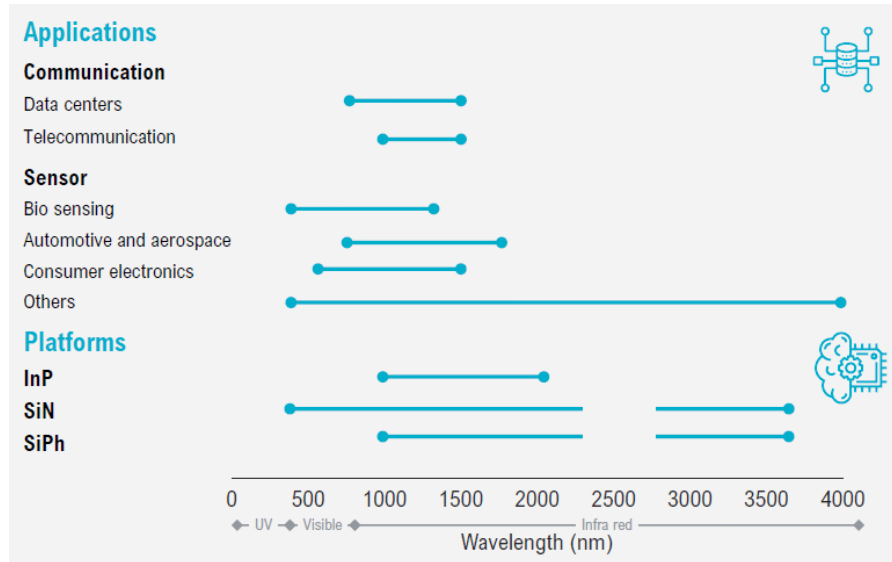


**Figure 2.13:** Direct Electro-optic modulation based neural networks. (a) The input-modulated signals are multiplied directly using optical attenuators, and the optical signals are collected onto a photodetector and summed. This drives an optical modulator to achieve a nonlinear function. Source:[75] (b) Use of temporal, wavelength and spatial degree of freedom for convolution process. By shaping a frequency comb and using an EOM to modulate a time multiplex signal, convolution is performed. Source:[76]

## 2.3 Photonic Platforms

A number of photonic platforms are available, such as those based on Si, SiN, and InP. Other promising platforms based on polymers and LN have emerged over time. Each platform has its benefits, and the choice of a photonic platform depends on the application. One key factor which decides the selection of a photonic platform is the working wavelengths. For applications such as sensing in the visible spectrum, SiN is preferred. For telecommunication communication, Si offers a large transparency

window in near-IR. Figure 2.14 summarises the transparency windows for popular photonic platforms.



**Figure 2.14:** Transparency window for various photonic platforms and the wavelength requirements depending on the application. Source:[77]

Similarly, when it comes to photonic computing, the requirements are unique. Photonic computing requires a platform that supports various building blocks such as modulators, lasers, photodetectors and associated electronic circuits. Another challenge associated with platform selection is compatibility with standard foundry processes. For scalability of the photonic systems, the availability of foundry processes is of utmost importance, especially to achieve device performance reliability. In terms of optical properties, an ideal platform should have excellent optical properties, both linear and nonlinear, with high mode confinement to reduce device footprint and availability of active components. In the following section, the capabilities and limitations of various photonic platforms will be discussed with respect to creating a common platform for photonic computing.

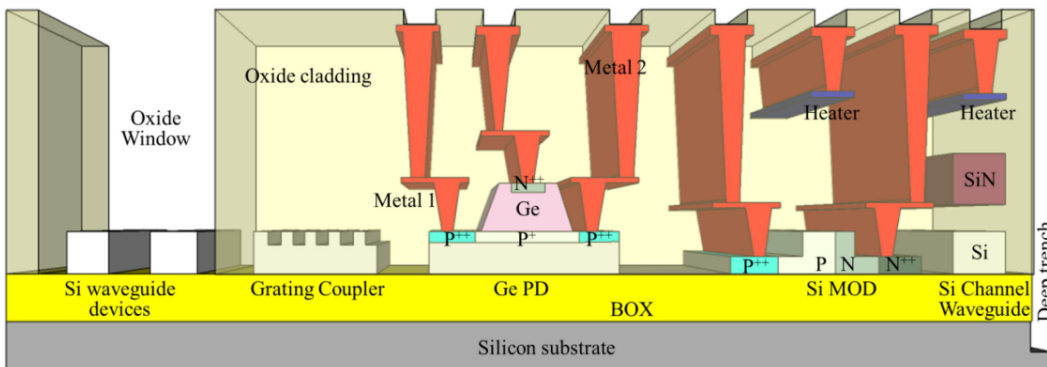
### 2.3.1 Silicon-Platform

Silicon photonic platform is based on the silicon on dioxide and is popularly termed SOI (silicon on insulator) platform. SOI- 220 nm platform is the most popular

platform for near-IR wavelength photonics [78]. SOI has emerged as one of the most popular photonic platforms owing to its transparency of silicon in near IR wavelength and ease of integration with the CMOS electronic foundries.

Silicon photonics benefits from high index contrast between the waveguide and the substrate material. This results in high optical mode confinements and thus benefits from small footprints and low loss [79, 80]. One of the key features required for high-density packaging of photonic components in a computing system. Silicon is an indirect bandgap material and therefore lacks active components such as laser sources and amplifiers. Nevertheless, silicon photonics offers high bandwidth; using arrayed waveguide gratings (AWG) [81, 82], a large number of optical signals can be multiplexed and thus proved a boost in data throughput for computing. Many loss wavelength division multiplexers [83] have been demonstrated in silicon.

Silicon lacks second-order nonlinearity, as offered by the lithium niobate platform. Therefore, the main modulation method in silicon is through carrier injection and thermo optic effects, as discussed in section 2.1. These effects are slow as compared to the speeds offered by LN. Nevertheless, p-i-n-based carrier injection silicon modulators have been demonstrated which can reach speeds close to 12 GHz [84], and charge depletion-based modulators can even reach speeds upto 50 GHz [84, 85]. The integration of germanium with a silicon photonic platform further opens up the



**Figure 2.15:** Typical silicon photonic foundry process node showing available active and passive components.

components available for photonic computing. Germanium photodetectors with high gain and high bandwidths up to 250 GHz [86] have been demonstrated in integrated platforms. The Si-Ge platform has also been used to implement CMOS-compatible modulators such as photodetectors [87] and electro absorption modulators (EAMs) [88, 89]. These advances in silicon photonics, therefore, make Si a promising platform of choice for photonic computing. Figure 2.15, shows a typical silicon photonic foundry process node with various active and passive components.

### 2.3.2 InP-Platform

Indium Phosphide-based photonic platform is one of the most mature platforms that has been in development since the 1980s [90]. III-V materials such as InP and GaAs show a direct bandgap and, therefore, can be used for gain medium in lasing and amplifier applications. InP and GaAs can be doped to achieve a tuneable bandgap, thus making the platform more exciting.

Development in InP photonics has been motivated by achieving monolithically integrated active and passive photonic computing elements [91]. As early as in the mid-1980s, integrated laser-modulator, combining distributed feedback lasers(dfbs) [92] and EAM, had been demonstrated [93]. This resulted in further development of the InP-GaAs photonic platform, as low loss waveguides [94] and passive components such as AWGs for wavelength multiplexing of the signals have been demonstrated [95, 96]. For modulation of the signal, EAM has been demonstrated with speeds approaching 40 GHz [97]. Monolithically integrated photodetectors up to 100 GHz have been demonstrated [98], thus proving an important platform for photonic computing.

Despite such advantages offered by the platform, the incompatibility with traditional CMOS processes still remains a challenge for the adoption of this platform. Moreover, the high cost of fabrication as compared to silicon photonics forms another bottleneck for the commercialisation of this platform.

### 2.3.3 Silicon Nitride-Platform

Over the past decade, SiN has become a platform of interest. One of the main reasons for the popularity of the SiN platform is the wide transparency window ranging from close to UV wavelengths up to mid-IR, Figure 2.14. Silicon nitride is CMOS compatible [99] and, like silicon, lacks a direct bandgap; thus cannot be used for lasers and other active components. The lower refractive index of silicon nitride ( $n=1.99$  at 1550 nm) compared to silicon ( $n= 3.67$  at 1550 nm) means less confined optical modes. This means larger waveguide size and, thus, larger bending radius requirements. On the other hand, the benefits of larger dimensions mean more tolerance to fabrication variation and low scattering losses [100].

Silicon nitride waveguide benefits from the ultra-low waveguide losses [101] and shows excellent Kerr nonlinear behaviour [102], thus making it a platform of choice for on-chip frequency combs generation [103, 104]. SiN has been integrated with graphene for photodetectors [105] and functional materials such as ferroelectric lead zirconate titanate (PZT) to demonstrate electro-optic modulators [106].

Silicon nitride has been shown to have ultra low propagation losses below 0.05 dB/cm using the damascene process [107]. Therefore due to the ultra-low propagation losses and CMOS compatibility, SiN remains a promising choice for passive optical components in integrated photonics.

### 2.3.4 Promising photonic platforms

There are various other promising photonic platforms which have been developed in past decade. In this section, two such platforms based on Lithium Niobate (LN) and other promising platforms based on polymers are discussed.

#### Lithium Niobate

Platforms based on LN have recently gained popularity due to their maturity in achieving high-quality waveguides. As described in the review article [108], the

development of LN photonics can be broken into three parts. The first involves the use of bulk crystals. LN crystal is a non-centrosymmetric material and has been known to show strong Kerr nonlinear effects [84]. The next development was using weakly confined LN modes. These waveguides were formed by achieving weak index contrast by irradiation with femtosecond laser [109] or diffusion out of lithium or titanium to induce weak refractive index contrast.

Lithium niobate is a hard material [110] to etch and to use the benefits of such platform hybrid waveguides using LN for modulation and silicon/silicon nitride as waveguiding material have also been demonstrated [111, 112]. As the LN photonics matured over the past 5 years, highly confined LN waveguides with propagation losses lower than 6 dB/m have been shown in visible wavelengths [113] and more than 10 million Q-factor ring resonators have been demonstrated [110]. To achieve modulation, ring modulators operating over 40 Gbps have been demonstrated on the LN platform [114, 115]. Since these developments, light sources based on Kerr frequency combs have also been demonstrated [116–118]. These developments in LN based photonic platform makes LN a promising platform for photonic computing.

### **Polymer Based Platform**

Polymer waveguides have found application as interconnect between the optical fibres and on-chip photonic components [119]. Polymer-based photonic platforms with losses below 0.1 dB/cm have been demonstrated [120]. Key features of polymer waveguides include flexibility and allow bending as opposed to the platform based on rigid substrates. Another key feature of such a platform is the ease of fabrication process [121]. Various methods of fabrication such as direct laser writing [122], two-photon absorption [123], hot imprinting [124] and standard lithography [120] steps have been used. [125]. Polymer-based ring resonator filters and modulators have been demonstrated with modulation speeds faster than 1 Gb/s [125]. For



multiplexing and demultiplexing applications, high-quality all polymer AWGs have also been demonstrated [126]. Thus showing the promising developments in polymer-based photonic platforms.

## 2.4 Critical Analysis and Summary

### Memory Cell

- Photonic memory cells for use in photonic computing needs to have a small footprint with a fast switching dynamics and high cyclability and should their retain memory states.
- PCM based photonic memory cell are a promising candidate due to ease of integration however face scalability issues due to incompatibility with standard foundry processes.
- For PCM memory cells, there is a need to reduce the switching energy while improving the cyclability of such devices to ensure reliability in achieving memory states.

### Photonic Network architecture

- Most of the implementations of photonic computing cores have been limited to only a few DNN layers with each architecture design having its challenges and scalability issues.
- Such photonic implementations only perform linear functions optically and rely on digital electronic processors for implementing nonlinear functions.
- One important issue that needs attention is the energy budget of the system. Most of the proposed solutions have focused on the computing output with little attention to the actual energy requirements to achieve those computation speeds.

## Photonic Platforms

- InP is a mature photonic platforms with a number of building blocks such are modulators and lasers available for use in photonic computing , but incompatibility with CMOS processes limit the scalability of this platform.
- Silicon photonic is widely available photonic platform due to compatibility with foundry processes. However silicon platform lacks the required active components needed for photonic computing. Moreover high two-photon coefficient, limits high power applications.
- The high processing temperatures for both InP and Si makes co integration of photonic platform with electronic difficult, thus making exploration of other photonic -platforms necessary.

# 3

## Theoretical Concepts

### Contents

---

|                                 |           |
|---------------------------------|-----------|
| <b>3.1 Dielectric Waveguide</b> | <b>37</b> |
| 3.1.1 Mode Analysis             | 37        |
| 3.1.2 Field Distribution        | 40        |
| <b>3.2 Grating Couplers</b>     | <b>41</b> |
| <b>3.3 Ring Resonators</b>      | <b>43</b> |
| 3.3.1 Introduction              | 43        |
| 3.3.2 Free-Spectral Range       | 45        |
| 3.3.3 Quality Factor            | 45        |
| 3.3.4 Propagation Loss          | 46        |
| 3.3.5 Ring Dispersion           | 46        |
| <b>3.4 Four-Wave Mixing</b>     | <b>47</b> |

---

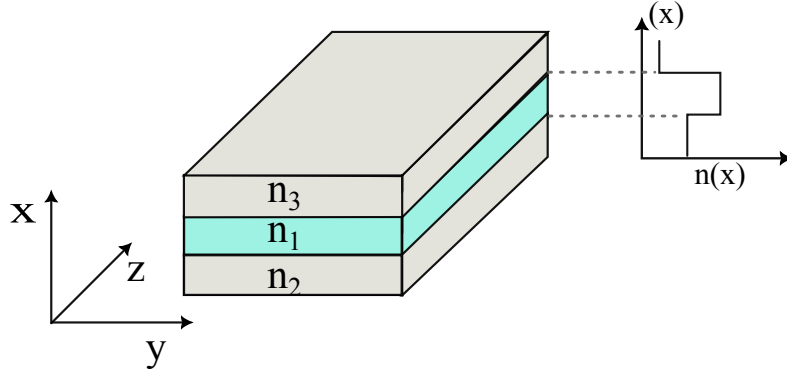
## 3.1 Dielectric Waveguide

Dielectric waveguides are an essential medium for guiding light in integrated photonic circuits. In this thesis, waveguides based on silicon on an insulator and silicon carbide on an insulator have been used. The following section covers the working principle of dielectric waveguides. The theoretical background presented here has been derived from the textbook by *Saleh* [127].

### 3.1.1 Mode Analysis

For simplicity, a planar waveguide is considered. A planar waveguide consists of an optical medium extending laterally (y-z plane) and having an index gradient longitudinally for optical mode confinement in only one direction (x-axis), as shown

in Figure 3.1. The high-index film ( $n_1$ ) is known as the core and is confined within low-index materials ( $n_3$  and  $n_2$ ), which are known as upper and lower cladding, respectively. The lower cladding will often be referred to as the substrate for waveguides. In this thesis, unless specified, the substrate is silicon dioxide.

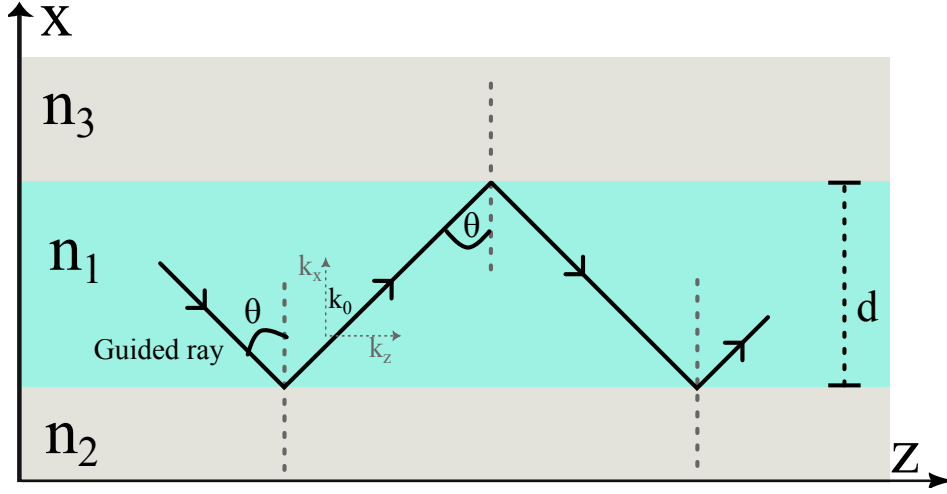


**Figure 3.1:** Schematic of planar waveguide with one-dimensional optical confinement, (b) Index profile corresponding to the planar waveguide.

A ray-optics approach is considered first to get a basic understanding of the mode principle. A dielectric structure with refractive indices,  $n_1 > n_2 > n_3$ , is shown in Figure 3.1. Assuming the wave in the slab is monochromatic with angular frequency  $\omega$  and of free space wavelength  $\lambda = \lambda_o/n$ , the propagation constant is written as  $k = n \frac{\omega}{c}$ .

For a wave to propagate losslessly, the incident angle  $\theta$  has to be larger than the critical angle  $\theta_{c1-3}$  and  $\theta_{c1-2}$  at both the interfaces ( $n_1 - n_3$  and  $n_1 - n_2$ ), for total internal reflection. The self-consistency condition that the wave reproduces after a round trip of bouncing back at the interfaces is imposed to determine the supported modes. Therefore, the reflected wave should constructively interfere with the incident wave.

In the x direction, in  $n_1$ , the wave propagates with  $k_x = n_1 k_o \cos\theta$  and  $k_z = n_1 k_o \sin\theta$ . For the slab thickness ‘d’, the path difference equals  $2d \cos\theta$ . Thus the phase shift in the two waves can be written as follows and should be a multiple



**Figure 3.2:** Ray-diagram showing the propagation of the guided ray in a slab waveguide structure.

of  $2\pi m$  for constructive interference.

$$2dn_1k_o\cos\theta + \phi_1 + \phi_2 = 2m\pi \quad (3.1)$$

$$2k_xd + \phi_1 + \phi_2 = 2m\pi \quad (3.2)$$

Where  $\phi_1$  and  $\phi_2$  represents the phase delay due to internal reflections at the interfaces, and  $m$  is an integer and is a function of  $\theta$ . For transverse electric fields [127],

$$\tan\frac{\phi}{2} = \sqrt{\frac{\cos^2\theta_c}{\cos^2\theta} - 1} \quad (3.3)$$

Since  $m$  can only assume discrete values, only certain values of  $\theta$  are allowed. For a simple case where the refractive index  $n_2 = n_3$ , the conditions for transverse electric (TE) modes can be written as :

$$\tan\left(\pi\frac{d}{\lambda}\cos\theta - m\frac{\pi}{2}\right) = -\sqrt{\frac{\cos^2\theta_c}{\cos^2\theta} - 1} \quad (3.4)$$

The solution for this equation would provide the values of  $\theta_m$  for allowed modes in the waveguides. In this thesis, the mode equations are solved using a Lumerical solver to find the correct parameters for fundamental mode waveguides.

### 3.1.2 Field Distribution

A more rigorous analysis is required to understand the electric field's distribution in a waveguide. The theoretical concepts presented in this section have been derived from the textbook by *Liu* [128].

As described above, the modes of the waveguide depend on the structure and dimensions of the waveguide. For a planar transverse mode, propagating in the  $z$ -direction, the electric ( $\mathbf{E}$ ) and magnetic fields ( $\mathbf{H}$ ) of a propagating mode for a planar mode can be written as follows:

$$\mathbf{E}_v(x, y, z, t) = \mathcal{E}_v(x) \cdot \exp(i\beta_\nu z - i\omega t) \quad (3.5)$$

$$\mathbf{H}_v(x, y, z, t) = \mathcal{H}_v(x) \cdot \exp(i\beta_\nu z - i\omega t) \quad (3.6)$$

where  $\nu$  is defined as the mode index,  $\mathcal{E}_v(x)$  and  $\mathcal{H}_v(x)$  are the mode electric and magnetic field profiles respectively and  $\beta_\nu$  is the propagation constant of a waveguide. Since the mode is confined only in the  $x$  direction, the mode field is independent of any  $y$ -terms.

For a linear, isotropic dielectric waveguide, Maxwell's equations can be written as follows:

$$\nabla \times \mathbf{E} = -\mu_o \frac{\partial \mathbf{H}}{\partial t} \quad \nabla \cdot (\epsilon \mathbf{E}) = 0 \quad (3.7)$$

$$\nabla \times \mathbf{H} = \epsilon \frac{\partial \mathbf{E}}{\partial t} \quad \nabla \cdot \mathbf{H} = 0 \quad (3.8)$$

Applying the Maxwell's equations for plane waveguide fields, described by equations (3.5)-(3.6), the electric  $\mathcal{E}$  and magnetic field  $\mathcal{H}$  components can be obtained. Since the electric field is invariant in  $y$ , any term with  $\frac{\partial}{\partial y} = 0$ .

$$-i\beta \mathcal{E}_y = i\omega \mu_o \mathcal{H}_x \quad -i\beta \mathcal{H}_y = i\omega \epsilon \mathcal{E}_x \quad (3.9)$$

$$i\beta \mathcal{E}_x - \frac{\partial \mathcal{E}_z}{\partial x} = i\omega \mu_o \mathcal{H}_y \quad i\beta \mathcal{H}_x - \frac{\partial \mathcal{H}_z}{\partial x} = i\omega \epsilon \mathcal{E}_y \quad (3.10)$$

$$\frac{\partial \mathcal{E}_y}{\partial x} = i\omega \mu_o \mathcal{H}_z \quad \frac{\partial \mathcal{H}_y}{\partial x} = i\omega \epsilon \mathcal{E}_z \quad (3.11)$$

Which can further be rearranged as following:

$$(k^2 - \beta^2)\mathcal{E}_x = i\beta \frac{\partial \mathcal{E}_z}{\partial x} \quad (k^2 - \beta^2)\mathcal{H}_x = i\beta \frac{\partial \mathcal{H}_z}{\partial x} \quad (3.12)$$

$$(k^2 - \beta^2)\mathcal{E}_y = -i\omega\mu_o \frac{\partial \mathcal{H}_z}{\partial x} \quad (k^2 - \beta^2)\mathcal{H}_y = i\omega\epsilon \frac{\partial \mathcal{E}_z}{\partial x} \quad (3.13)$$

Where  $k = \omega\sqrt{\epsilon_o\mu_o}$ . For a transverse electric field propagating in the z-direction,  $\mathcal{E}_z = 0$ , thus from the above equations,  $\mathcal{E}_x = \mathcal{H}_y = 0$ . Thus, all the field components can be calculated by knowing the field components  $\mathcal{E}_z$  and  $\mathcal{H}_z$ .

Further, using equations (3.11) and (3.13), the wave equation for a TE planar mode can be written as follows:

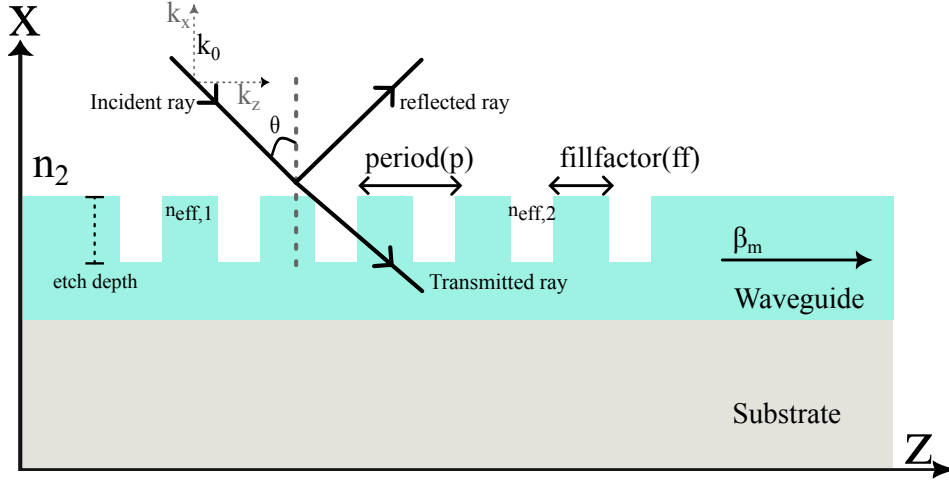
$$\frac{\partial^2 \mathcal{E}_y}{\partial x^2} + (k^2 - \beta^2)\mathcal{E}_y = 0 \quad (3.14)$$

Such coupled equations can be solved to get the field distribution in the waveguide. However, solving such equations for complex waveguide geometries, as used in this thesis, is not straightforward. Therefore, the mode profiles are solved using a *Lumerical* mode solver.

## 3.2 Grating Couplers

Various light coupling strategies have been proposed to couple light into the waveguide [129]. Some of such light coupling strategies are based on edge coupling. For edge coupling, the fibre is placed at the end of the chip facet. For efficient coupling of light, inverse tapers are used to match the fibres' spot size. Such a coupling strategy is quite efficient and provides broadband coupling of light. However, edge coupling techniques limit the accessibility of devices in the middle of the chip. Therefore vertical coupling strategies, such as using grating couplers are preferred.

A grating coupler changes the off-plane light to the in-plane waveguide direction. A grating coupler consists of a periodic structure of varying arrangements of different materials, as shown in Figure 3.3. A grating coupler diffracts the incoming light



**Figure 3.3:** Cross-sectional schematic of a grating coupler with ray diagrams.

into the waveguide propagation plane. The diffracted light must constructively add along the diffraction angle (Bragg condition) [130]. In other words, coupling incident light into the guided mode requires the wavevector component  $n_2 k_o \sin \theta$  to match the waveguide propagation constant  $\beta_m$ , where  $m$  is the mode number.

A grating coupler with a period ( $p$ ) modifies the wavevector of incident light by a phase factor of  $2\pi m/p$ , where  $m$  is the diffraction order. Thus the following phase matching condition can be satisfied:

$$n_2 k_o \sin \theta_i + 2\pi m/p = \beta_m \quad (3.15)$$

The propagation constant  $\beta_m$  for a grating coupler can be expressed as  $\beta_m = \frac{2\pi}{\lambda_o} n_{eff}$ . Thus Bragg's condition can be written as follows:

$$n_2 \sin \theta_i = n_{eff} - m \frac{\lambda}{p} \quad (3.16)$$

Here  $n_{eff}$  is expressed as :

$$n_{eff} = ff \cdot n_{eff,1} + (1 - ff) \cdot n_{eff,2} \quad (3.17)$$

Where  $ff$  is the fill factor of the gratings and  $n_{eff,1}$  and  $n_{eff,2}$  are the effective refractive index of the waveguide and the etched region.



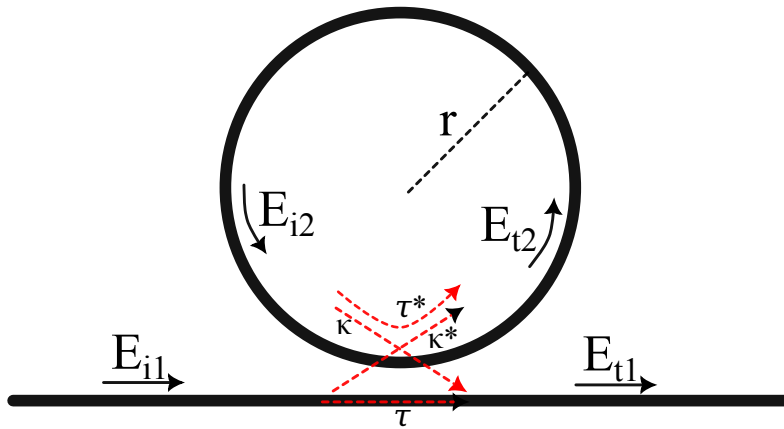
To improve the coupling efficiency and suppress the reflections due to Fabry-Perot cavity formation, apodised grating couplers [131] have been used in this thesis. For apodised grating coupler, the fill factor of the grating coupler is varied along the length. Therefore the grating couplers have been optimised using finite element analysis in *Lumerical* FDTD to find the correct grating parameters and have been discussed in relevant chapters.

Notably, the grating couplers have limited bandwidth and are polarisation-sensitive devices. Nevertheless, the ease of use and accessibility to devices makes them an ideal light coupling solution.

### 3.3 Ring Resonators

#### 3.3.1 Introduction

To enhance the optical power to study the nonlinear properties of silicon carbide substrate, ring resonators have been used in Chapter 8. A ring resonator consists of a plane waveguide coupled to a ring waveguide of radius ‘ $r$ ’ as illustrated in Figure 3.4. The theoretical approach described in [132, 133] have been used below. In this thesis, a ring resonator with only one coupled waveguide, also called an all-pass resonator, has been used. A ring resonator is excited in a single unidirectional mode.



**Figure 3.4:** Schematic of an all-pass ring resonator of radius  $r$  and coupling parameters.

The relation between the input and transmitted electric fields can be described

by the following transfer matrix:

$$\begin{pmatrix} E_{t1} \\ E_{t2} \end{pmatrix} = \begin{pmatrix} \tau & \kappa \\ -\kappa^* & \tau^* \end{pmatrix} \begin{pmatrix} E_{i1} \\ E_{i2} \end{pmatrix} \quad (3.18)$$

where the  $\tau$  and  $\kappa$  are the coupling coefficients of the ring with \* componets as their complex conjugates. The coupling parameters depend on the coupling gap and the coupling length of the ring and waveguide. For lossless coupling,

$$|\kappa|^2 + |\tau|^2 = 1 \quad (3.19)$$

For simplicity if the input field,  $E_{i1}$  is taken equal to 1 , then  $E_{i2}$  can be written as :

$$E_{i2} = a \cdot e^{i\theta} E_{t2} \quad (3.20)$$

where a is associated to the propogation loss ( $\alpha$ ) as  $a^2 = e^{-\alpha L}$ . The term  $e^{i\theta}$  is associated with the phase shift introduced. For a ring of radius r, the phase shift can be written as :

$$\theta = \frac{2\pi n_{eff}}{\lambda} \cdot 2\pi r = 4\pi^2 n_{eff} \frac{r}{\lambda} \quad (3.21)$$

Thus combining equations (3.18) - (3.20)

$$E_{t1} = \frac{-a + \tau \cdot e^{-i\theta}}{-a\tau^* + e^{-i\theta}} \quad (3.22)$$

$$E_{i2} = \frac{-a\kappa^*}{-a\tau^* + e^{-i\theta}} \quad (3.23)$$

$$E_{t2} = \frac{-\kappa^*}{1 - a\tau^* e^{i\theta}} \quad (3.24)$$

Thus the transmission of a resonator  $|E_{t1}|^2$  can be calculated as follows:

$$T = |E_{t1}|^2 = \frac{a^2 + |\tau|^2 - 2a|\tau|\cos(\theta)}{1 + a^2|\tau|^2 - 2a|\tau|\cos(\theta)} \quad (3.25)$$

At resonance condition,  $\theta = 2\pi m$ , where m is an integer value. From equation (3.21),  $\lambda_{res} = \frac{2\pi r \cdot n_{eff}}{m}$ . Thus the transmission can be written as:

$$T = |E_{t1}|^2 = \frac{(a - |\tau|)^2}{(1 - a|\tau|)^2} \quad (3.26)$$

At critical coupling conditions, the losses in the ring become equal to the light coupled into the ring, i.e.  $a = |\tau|$ . Thus at critical coupling condition, the transmission becomes 0.

### 3.3.2 Free-Spectral Range

One of the important ring resonator parameters is its free spectral range (FSR). This characterises the distance between the two resonance peaks. For a ring resonator, the FSR is given by:

$$FSR = \Delta\lambda = \frac{2\pi}{L} \left( \frac{\partial\beta}{\partial\lambda} \right)^{-1} \quad (3.27)$$

Where  $L = 2\pi r$ . For a waveguide, it is known that

$$n_{eff} = \frac{\beta}{k} \quad (3.28)$$

Thus,

$$\frac{\partial\beta}{\partial\lambda} = -\frac{\beta}{\lambda} + k \frac{\partial n_{eff}}{\partial\lambda} \quad (3.29)$$

$$\frac{\partial\beta}{\partial\lambda} = -\frac{k}{\lambda} n_g \quad (3.30)$$

Where  $n_g$  is defined as the group refractive index and is given as follows:

$$n_g = n_{eff} - \lambda \frac{\partial n_{eff}}{\partial\lambda} \quad (3.31)$$

This results in FSR  $\Delta\lambda$  to be given as follows:

$$FSR = \frac{\lambda^2}{n_g L} \quad (3.32)$$

### 3.3.3 Quality Factor

Another important parameter used in this thesis is the quality factor (Q). The quality factor is the measure of the sharpness of the resonance with respect to its central wavelength and is defined as follows:

$$Q - factor = \frac{\lambda_{res}}{FWHM} \quad (3.33)$$

Where FWHM is the full width at half maximum of the resonance peak, the Q-factor of a ring is associated with the number of round trips of the light inside the resonator before the power is reduced to  $1/e$  of the initial value. This is also called the loaded Q-factor for the ring. Since the coupling of the ring to the waveguide introduces losses to the ring, another term known as the intrinsic quality factor  $Q_i$  [134] is defined as follows:

$$Q_i = \frac{2Q_L}{1 \pm \sqrt{T}} \quad (3.34)$$

Where  $Q_L$  is the loaded Q factor, and T is the normalised transmission at resonance.

### 3.3.4 Propagation Loss

The propagation loss  $\alpha$  of the ring can be calculated using the intrinsic quality factor [125] as follows :

$$\alpha = \frac{2\pi n_g}{Q_i \lambda_o} \quad (3.35)$$

Using Equation (3.32) :

$$\alpha = \frac{\lambda_o}{Q_i \cdot FSR \cdot r} \quad (3.36)$$

The loss of the waveguide is generally reported as dB/cm, which can be calculated using the expression :

$$loss \left( \frac{dB}{cm} \right) = -10 \frac{1}{L} \cdot \log(e^{-\alpha \cdot L}) \quad (3.37)$$

### 3.3.5 Ring Dispersion

Experimentally, the free spectral range can be calculated from a transmission spectrum of a ring resonator. Thus by varying the ring length, the group index

can be calculated for the ring using equation (3.32). The average dispersion is defined as:

$$D = \frac{1}{c} \frac{\partial n_g}{\partial \lambda} \quad (3.38)$$

For third-order nonlinear experiments, it is important to achieve anomalous dispersion. This will be discussed in the next section.

### 3.4 Four-Wave Mixing

In this thesis, four-wave mixing experiments have been performed to measure the third-order nonlinear optical properties of silicon carbide waveguides. When a high-power optical light is injected into the waveguide, the linearity of waveguides is lost due to a third-order nonlinear phenomenon called the Kerr effect. Due to such nonlinear behaviour, additional wavelengths are generated when two or more wavelengths are injected. The theory for four-wave mixing used in the thesis has been developed from [135, 136].

In four-wave mixing, two photons from high power pump signal of frequencies  $\omega_{p1}$  and  $\omega_{p2}$  are mixed with a low power light (signal) of frequency  $\omega_s$  to generate an idler signal. From the conservation of energy, the idler frequency can be determined to be:

$$\omega_i = \omega_{p2} + \omega_{p1} - \omega_s \quad (3.39)$$

A special case arises when  $\omega_{p1} = \omega_{p2}$ . This is called degenerate four-wave mixing. In this thesis, degenerate four-wave mixing has been used. Thus the idler frequency  $\omega_i$  is expressed as:

$$\omega_i = 2\omega_p - \omega_s \quad (3.40)$$

Considering that the signal and the generated idler frequency have lower power than the pump frequency, the coupled amplitude equations is defined [135] as

follows:

$$\frac{\partial E_p}{\partial z} = -\frac{\alpha_p}{2} E_p + i\gamma |E_p|^2 E_p \quad (3.41)$$

$$\frac{\partial E_s}{\partial z} = -\frac{\alpha_s}{2} E_s + 2i\gamma |E_p|^2 E_s + \gamma E_p^2 E_i^* e^{-i\Delta\beta Z} \quad (3.42)$$

$$\frac{\partial E_i}{\partial z} = -\frac{\alpha_i}{2} E_i + 2i\gamma |E_p|^2 E_i + \gamma E_p^2 E_s^* e^{-i\Delta\beta Z} \quad (3.43)$$

Where  $E_{p,s,i}$  is the electric field,  $\alpha_{p,s,i}$  is the propagation loss for the pump, probe and idler frequencies.  $\Delta\beta$  is associated with the phase mismatch of the propagation constants is defined as follows:

$$\Delta\beta = \beta_s + \beta_i - 2\beta_p \quad (3.44)$$

The frequency conversion between the signal and idler can become more efficient and amplified exponentially if the phase matching condition is satisfied. In the equation above, the nonlinear coefficient term  $\gamma$  is the effective nonlinearity of the structure and is defined [137] as follows:

$$\gamma = \frac{n_2 \omega_p}{A_{eff} c} \quad (3.45)$$

where  $n_2$  is the nonlinear refractive index and  $A_{eff}$  is the effective mode area which is defined as [138]

$$A_{eff} = \frac{\left( \iint_{-\infty}^{\infty} \mathbf{S} \cdot d\mathbf{x} d\mathbf{y} \right)^2}{\iint_{-\infty}^{\infty} \mathbf{S}^2 \cdot d\mathbf{x} d\mathbf{y}} \quad (3.46)$$

Where  $\mathbf{S}$  is the Poynting vector. Therefore, a confined mode can result in an efficient nonlinear effect in the waveguide. The conversion efficiency  $\eta$  for a waveguide is defined as [139] :

$$\eta = \frac{P_{idler}}{P_{signal}} = (\gamma P_p L_{eff})^2 \quad (3.47)$$

Where  $P_p$  is the pump power and  $L_{eff}$  is the effective length given by :

$$L_{eff} = \frac{1 - e^{-\alpha L}}{\alpha} \quad (3.48)$$

The wavelength conversion efficiency in four-wave mixing can be enhanced using a cavity, such as in a ring resonator. Ring resonator can increase the intensity of the electric field and this enhancement is termed as Field Enhancement factor (FE) [137] and is given as follows:

$$FE = \left| \frac{\kappa}{1 - \tau \cdot \exp(-\alpha L/2 + ik_p L)} \right| \quad (3.49)$$

where ,  $\kappa$  and  $\tau$  are the ring coupling coefficients as described in Section 3.3 and are associated as  $\kappa^2 + \tau^2 = 1$ .  $k_p$  is the wavenumber for the pump. As discussed in Section 3.3 , for critical coupling conditions ring loss  $\alpha = \tau$  and  $k_p L = 2\pi m$ . Thus FE can be expressed as propagation loss parameter  $\alpha$  as follows:

$$FE = \frac{1}{\sqrt{1 - e^{-\alpha L}}} \quad (3.50)$$

The four-wave mixing conversion efficiency for a ring resonator is thus expressed as follows [137]:

$$\eta = \frac{P_i}{P_s} = |\gamma P_p L'|^2 FE_p^4 FE_s^2 FE_i^2, \quad (3.51)$$

$$L'^2 = L^2 \exp(-\alpha L) \left| \frac{1 - \exp(-\alpha L + i\Delta k L)}{-\alpha L + i\Delta k L} \right|^2 \quad (3.52)$$

Where  $L'$  is the effective length of the resonator,  $FE_j$  is the field enhancement for pump, signal, and idler, and  $\Delta k$  is the phase mismatch due to dispersion.

Thus using the four-wave mixing, the nonlinearity parameters for silicon carbide waveguide and ring resonators can be calculated, and the results will be discussed in detail in Chapter 8.

# 4

## Device and Experimental Setup

### Contents

---

|                                                   |           |
|---------------------------------------------------|-----------|
| <b>4.1 Design and Fabrication . . . . .</b>       | <b>50</b> |
| 4.1.1 Device Design . . . . .                     | 51        |
| 4.1.2 Device Fabrication . . . . .                | 52        |
| <b>4.2 Experimental Setup . . . . .</b>           | <b>54</b> |
| 4.2.1 Single device measurement setup . . . . .   | 55        |
| 4.2.2 Femtosecond laser switching setup . . . . . | 57        |
| 4.2.3 Matrix Multiplication . . . . .             | 61        |
| 4.2.4 Four-wave mixing . . . . .                  | 63        |

---

This chapter describes the details for fabrication of the PCMs memory cell device integrated with a waveguide. Then the experimental setup details for integrated switching of PCM using the pump-probe technique, focusing on femtosecond switching of PCM and thermo-optic effect measurements is discussed. Next, the experimental setup used to conduct the matrix-vector multiplication experiments on a reduced rank matrix is discussed. Finally, the experimental setup used to measure the nonlinear optical properties of the silicon carbide waveguides is discussed.

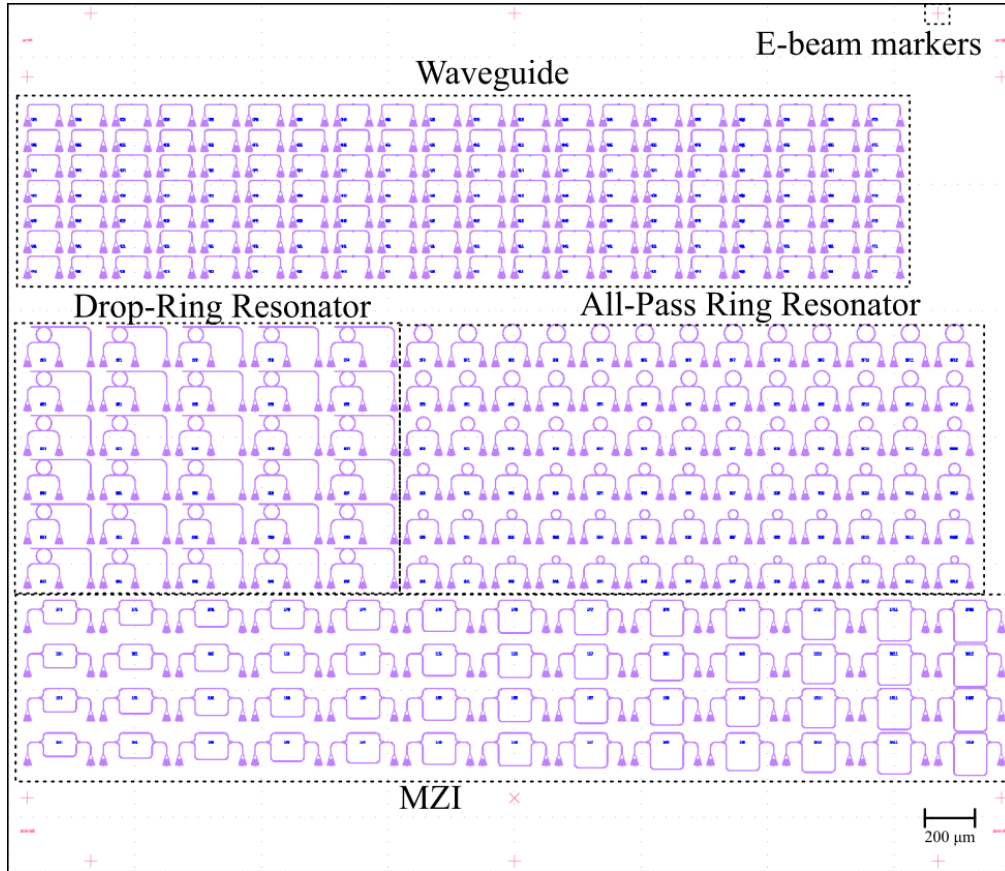
### 4.1 Design and Fabrication

Small-scale and single-element photonic devices used in the thesis have been fabricated in-house. However, large-scale implementations used in photonic computing have been developed with the help of IMEC foundry using their ISIPP-50G PDK. This excluded the devices used in silicon carbide experiments, where the collaborators at TU Delft have fabricated passive devices.



### 4.1.1 Device Design

Single devices used in photonic memory experiments consist of waveguide design with grating couplers to input light in and out of devices. Design patterns are drawn using python based design tool GDS-helpers. The design is exported as GDS-II cad files. A snapshot of one such CAD design consisting of various optical devices such as waveguides, ring resonators and MZI is shown in Figure 4.1. For designing large-scale matrix using IMEC PDK, python-based software IPKISS from Luceda has been used. The Python codes used to generate the photonic matrix have been added to appendix A.



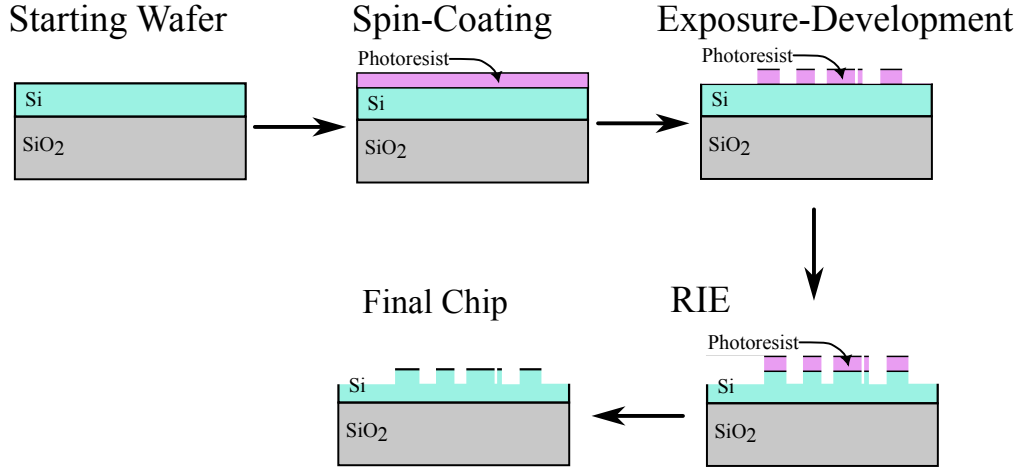
**Figure 4.1:** A schematic of the photonic design patterns consisting of the waveguide, ring resonators and MZI, used to fabricate photonic components.

### 4.1.2 Device Fabrication

The lithography methodology for SOI waveguides had been optimised by previous DPhil students in the group, and therefore the optimisation of the fabrication recipe has not been discussed in detail. The steps used in the device fabrication of a passive silicon waveguide are as follows (Figure 4.2)

1. Device fabrication of SOI devices begins with dicing a 6-inch SOI wafer into chips of size  $10 \times 10$  mm. A photoresist (S1813) is spun at 3000 rpm and baked at  $150^\circ$  C before dicing to protect the wafer from contamination.
2. The diced chips are then cleaned using MICROPOSIT Remover 1165. The chips are further cleaned in acetone and isopropyl alcohol (IPA) bath for 5 minutes each using ultrasound. The chip is dried by blowing nitrogen air. The dried chip is further heated on the hotplate for 10 mins at  $150^\circ$  C before the photoresist spin process.
3. An electron beam-sensitive, positive photoresist CSAR-62 (ARP-6200 series) is spun at 4000 rpm for 45 seconds. The photoresist is then prebaked at  $150^\circ$  C for three minutes.
4. The prepared chip is then exposed using electron beam lithography (EBL) at 100 pA and a dose of  $180 \mu\text{C}/\text{cm}^2$ . It is essential to write global markers alongside the waveguide patterns for the alignment process needed in the subsequent fabrication steps.
5. The exposed patterns are then developed in a three-step development process. First, the chip is developed in AR 600-546 developer for 30 seconds. Then the chip is washed in methyl isobutyl ketone (MIBK) for 15 seconds and washed in IPA for the final 15 seconds. The developed chips are dried using dry nitrogen air and inspected using an optical microscope.

6. The chips are then subjected to a dry etching process using reactive ion etching (RIE). To etch silicon, a mixture of  $\text{CHF}_3$  40 sccm and  $\text{SF}_6$  15 sccm are used with 150 W RF power for 42 seconds. The remaining photoresist is removed by a 5 min oxygen plasma cleaning in the RIE at 200W.



**Figure 4.2:** Device Fabrication Process (clockwise). Starting with a silicon chip, an electron beam-sensitive photoresist is spin coated on the chip. The chip is then exposed with the desired pattern and developed. Finally, using RIE, silicon is partially etched, and the photoresist is removed using oxygen plasma.

The fabricated waveguide devices are then tested for transmission using the setup described later. To fabricate devices for PCM switching experiments, the devices are subjected to a second round of lithography to open windows for PCM deposition on top of the waveguide using sputtering. The steps used for the fabrication of PCM memory devices are as follows (Figure 4.3):

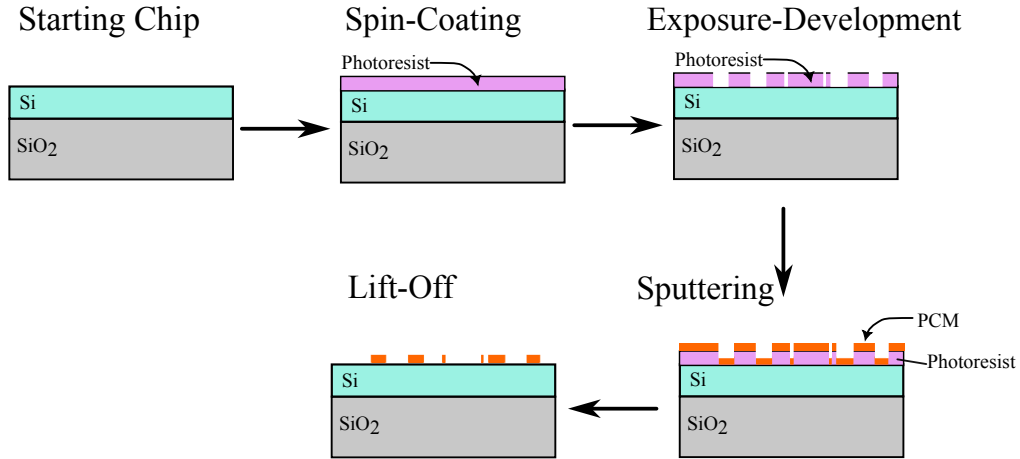
1. The chip with waveguide pattern is exposed with the pattern for PCM windows using EBL as described earlier. After developing the exposed patterns, the chip is ready for PCM deposition using sputtering. The sputtering rate of various thin films has been recorded in table 4.1.
2. After deposition of PCM, the chip is subjected to a liftoff process in photo remover REM-700. The photo remover is heated to  $60^\circ \text{C}$ , and the chip is

heated for 2-3 hours to ensure good liftoff. Finally, the chip is cleaned in acetone and IPA and dried using nitrogen air.

3. To crystallise the as-deposited PCM thin film on the waveguide, the fabricated device is placed on a hotplate.

**Table 4.1:** Sputtering parameters and sputtering rates for various thin films used in the thesis.

| Material | Pressure (mTorr) | RF Power (W) | Sputtering Rate (nm/min) |
|----------|------------------|--------------|--------------------------|
| GST      | 5                | 30           | 7.78                     |
| AIST     | 5                | 30           | 2.8                      |
| Sb       | 5                | 30           | 3.3                      |
| ITO      | 5                | 50           | 2.28                     |



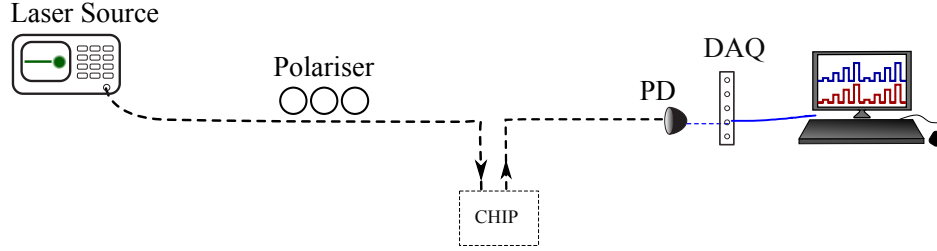
**Figure 4.3:** Lift off process. The lithography step is used to open windows for material deposition. Using sputtering PCMs, thin films are deposited on the patterned chip. The chip is then dipped in a photo remover to get rid of the photoresist and the excess material with it.

## 4.2 Experimental Setup

In this thesis, mainly four different setups have been used. These include the transmission measurement setup to assess the performance of devices. A pump-probe setup is used to switch antimony using a femtosecond laser, a setup to carry out matrix-vector multiplication and an experimental setup for nonlinear measurements.

Each experimental setup is discussed below. Wherever necessary, the details of the equipment used in the experimental setup are added to each section.

### 4.2.1 Single device measurement setup

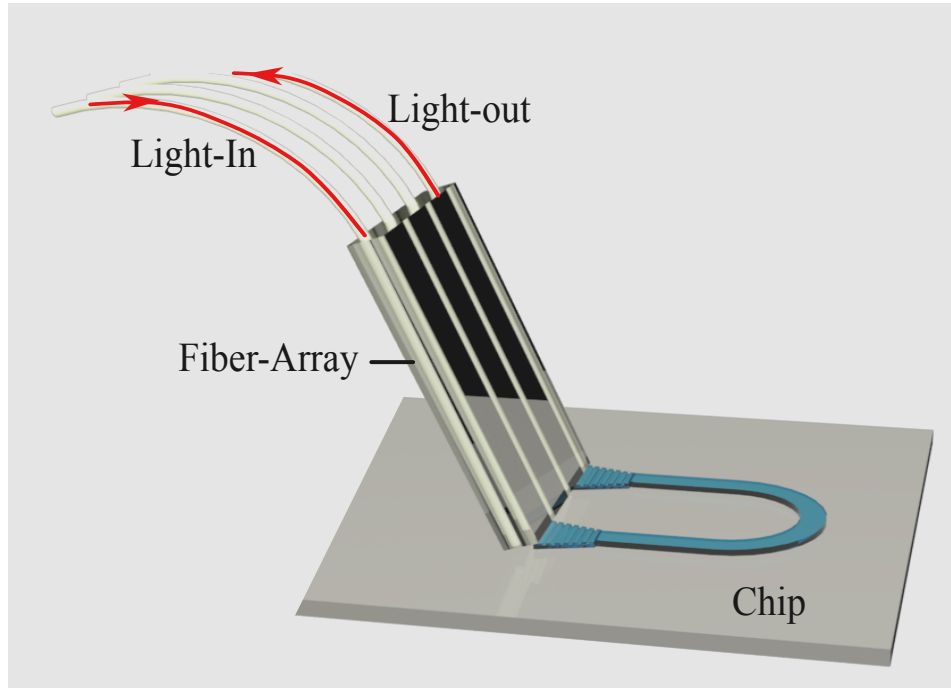


**Figure 4.4:** Experimental Setup for measuring transmission characteristics of waveguides. A laser source is sent across the photonic device. The input is polarisation controlled, and the output is recorded on a photodetector and recorded onto a computer using a DAQ.

A transmission measurement setup is used to test the performance of devices such as a waveguide, resonators and MZI, as shown in Figure 4.4. The setup comprises a C-L band tuneable laser from Santec (TSL-550). The signal from the laser is routed through an in-line polariser controller from Thorlabs (FPC032). The laser input is injected into the photonic chip using a fibre array, as shown in the schematic Figure 4.5. The output signal is collected on a 200kHz, tuneable gain photodetector from Newport (Model:2011-FC). A computer-controlled stage is used to align devices with the fibre array. A custom-built tube lens from Navitar is used to locate devices for measurements. The data is recorded on a computer with the help of a DAQ from national instruments.

#### Laser

A continuous wave (CW) tuneable laser from Santec, TSL-550, characterises photonic devices in near IR wavelengths (1500-1630 nm). The benefit of using the Santec laser is its fast wavelength sweep speed of up to 100 nm/s. The Santec laser has a wavelength resolution of up to 0.1 pm, which can be used to measure high-quality factor devices, up to a million, as discussed in later chapters. The Santec laser



**Figure 4.5:** Illustrative diagram showing the fibre array used to couple light in and out of the photonic chip. Grating couplers are used to input light into the device.

provides a high optical output of up to 13 dBm (19 mW) with a large signal-to-noise ratio. The laser can be tuned from a computer using the LabVIEW program developed by previous students in the group.

### Photodetector

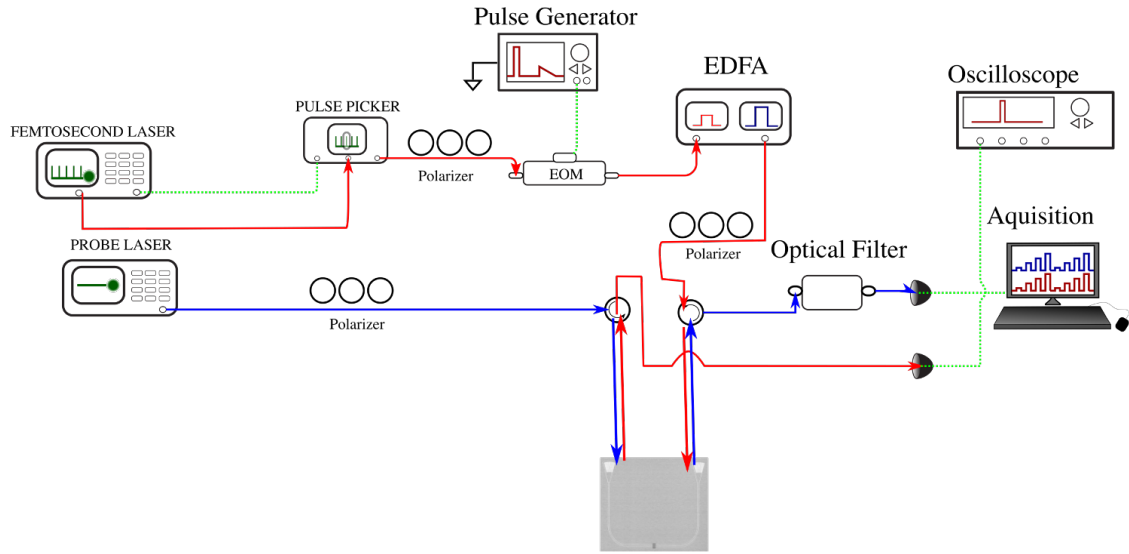
A low-bandwidth photodetector is used to monitor the transmission of the devices. The 2011-FC photodetector is an InGaAs tuneable gain photodetector with a high trans-impedance gain of  $18.8 \times 10^6$  V/A with a high signal-to-noise ratio. The photodetector has a responsivity close to 0.6 A/W for a 1550 nm wavelength.

### Polarisers

The in-line polariser uses stress-induced birefringence in a single-mode fibre to control the polarisation of the light. FPC032 in-line polarisers are three paddle polarisers consisting of a quarter-wave plate, half-wave plate and another quarter-wave plate in series to generate any desired polarisation for arbitrary input polarisation. A

quarter wave plate converts a circularly polarised to linearly polarised light and vice-versa. The role of a half-waveplate is to rotate the polarisation of the linearly polarised light. Finally, the third polariser converts the linearly polarised light back to circular polarisation.

#### 4.2.2 Femtosecond laser switching setup



**Figure 4.6:** Experimental Setup for integrated photonic memory based on antimony using a femtosecond laser. Femtosecond pulses are captured using a pulse picker and amplified using EDFA to switch PCM. A CW laser is used to monitor the change in transmission.

A pump-probe technique is used to switch antimony memory devices using the femtosecond laser pulses and observe the change in transmission. The experimental setup includes two optical lines, as shown in Figure 4.6. Red and Blue lines correspond to the Pump and Probe lines.

For the Pump line, a fibre-coupled femtosecond (fs) laser, centred at 1560 nm wavelength, with a pulse width of 800 fs and repetition rate of 40 MHz, is used. This signal is passed through a home-built pulse picker comprising an acoustic optical Modulator (AOM) to select a single or train of pulses. The amplitude of the pulses is modulated using an EOM and amplified using a high peak power Erbium-doped fibre amplifier (EDFA). The probe signal is a continuous laser centred at wavelength

1580 nm. Both the Pump and Probe signals are polarisation controlled to get maximum transmission through the device.

The two signals are injected in and out of the device using optical circulators. The output signal from the probe line is passed through an optical filter to block the signal from Pump and is used to monitor the change in transmission upon switching. Further, during the switching speed measurement experiments, the probe signal is connected to a 125 MHz photodetector from Newport (1811-FC) and connected to an oscilloscope to monitor the dynamic switching behaviour. The output of the femtosecond pulses is connected to another 125 MHz photodetector to monitor pulse characteristics using an oscilloscope.

### **Optical circulator**

A circulator is a device which allows light to propagate only in one direction. For the experiments in this thesis, a three-port fibre-coupled circulator from Thorlabs (6015-3-APC) is used. This allows low-loss transmission of light from port 1 to port 2 and port 2 to port 3. The signal flow in the other direction experiences high optical losses of over -50 dB.

### **Electro-optic modulator**

The optical pulses from the pulse picker are further amplitude modulated using an EOM. A 10 Gbps electro-optic modulator from Lucent (2623NA) is used for amplitude modulation of optical pulses. The EOM is controlled using an external arbitrary function generator from Tektronix (AFG3102C). An EOM works on the principle of phase shift induced due to the Pockel effect in a lithium niobate interferometer. Applying electrical voltage, the phase shift results in amplitude modulation due to the interferometer. The EOM has a  $V_\pi$  voltage of 3.8 V.



### **Erbium-doped fibre amplifier**

The pulses picked from the pulse picker are further amplified using a high peak power EDFA from Pritel (HPPPMFA-21-10). The EDFA provides a maximum gain of 20 dB for a 3 dBm input. By changing the pumping current for the EDFA, high-power pulses up to 189 mW with peak powers over 26 kW can be generated while maintaining pulse shapes. The EDFA features a monitor port to measure the output powers of the amplified pulses. A 99:1 tap line is used for the monitor port for pulse energy measurements.

### **Filter**

A manual optical filter OTF-320 from Santec is used to selectively read probe wavelengths and block stray light from the high-energy pump laser. OFT-320 offers a tuneable wavelength selection with a bandwidth of 0.5 nm.

### **Lasers**

Two different laser sources have been used for the measurements: pump and probe lasers. For the pump laser, a high-energy femtosecond laser from Pritel (FFL-TW-60MHz). The laser specification allows a power output of 10-43 mW depending on pulse width. Depending on the bandpass filter installed into the laser, the pulse width can be tuned from 140 fs-17 ps. A filter corresponding to 0.80 ps pulses is used for the experiments used in the thesis. The laser has a repetition rate of 40 MHz, with a centre wavelength of 1550 nm. Using a micrometre, the laser allows for tuneability of wavelength from 1530-1560 nm. The laser comprises an electrical sync-out for pulse picking function.

For probing, a CW tunable L-band laser from Keysight (N7711A) is used for the probe laser. The laser allows a power output of up to 19 mW with precise tunability of 0.1 pm. Keysight laser offers low noise operation, thus favouring its use as a probe laser.

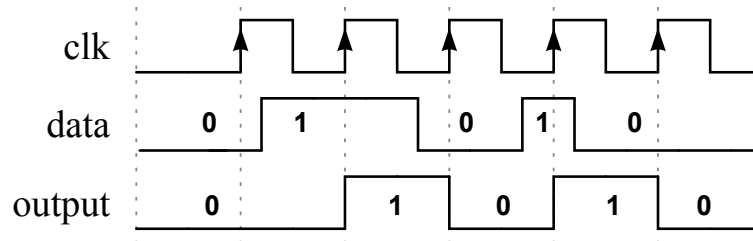
### Oscilloscope

An oscilloscope from Tektronix TDS7404B is used to capture electronic signals.

### Pulse generator

The control of the optical shutter for the pulse picker requires electrical signals. A 100 MHz pulse generator from Tektronix (AFG3102C). Arbitrary pulse shapes can be generated using a LabVIEW-controlled program.

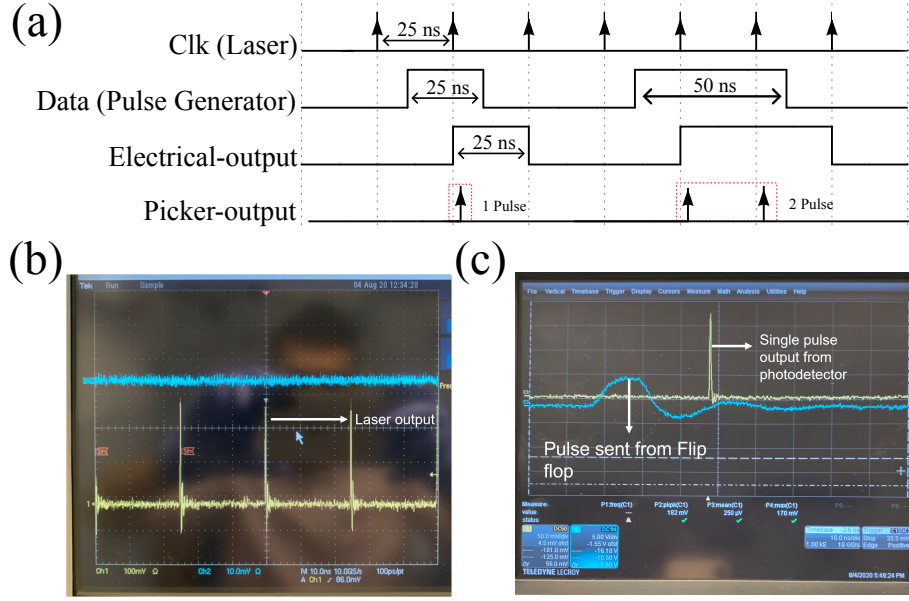
### Pulse Picker



**Figure 4.7:** Typical D-flip-flop timing diagram. At every clock rising cycle, the value of ‘data’ is obtained at the output.

PCM-switching using femtosecond pulses requires precise control over the number of pulses sent to the memory device. The output of a femtosecond laser is a train of fs-pulses with a period rate of 25 ns (40 MHz). A single pulse can be captured using an AOM as an optical shutter. A pulse generator signal controls the output signal from the AOM. However, capturing single or multiple pulses requires synchronisation of the optical shutter and electrical signal. A D-Flip-flop from Analog devices (HMC723LP3E) is used to achieve such synchronisation.

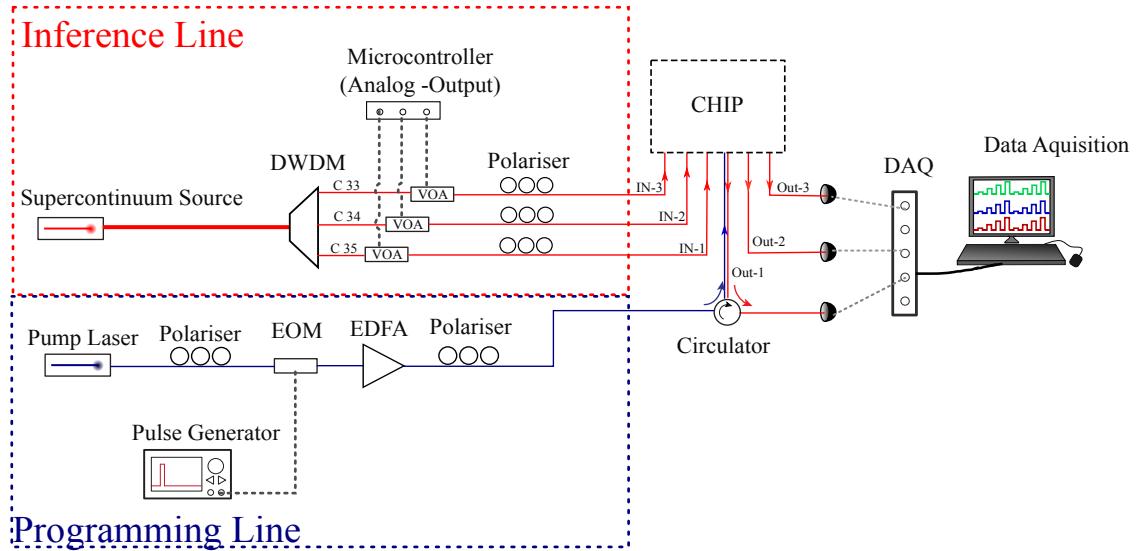
A D-flip-flop has two inputs, a clock signal and a data signal. Depending upon the value of data (0 or 1) at each clock cycle, the output signal is varied, as shown by the timing diagram in Figure 4.7. The diagram shows that the output signal synchronises with the clock, irrespective of the data signal. This property of synchronisation is exploited to build a pulse picker. Using the train of pulses from the laser as a clock source and by setting the data signal width to be a multiple of



**Figure 4.8:** (a) Timing diagram with pulse picker output to capture single and multiple pulses. (b) Experimental setup showing a train of pulses on an oscilloscope. (c) Experimental capture of a single pulse using a pulse picker.

the laser repetition rate, i.e. 25 ns, any arbitrary number of pulses can be captured as shown in Figure 4.8.

### 4.2.3 Matrix Multiplication



**Figure 4.9:** Experimental setup for matrix multiplication experiments. The setup consists of two lines: inference and programming line. The programming line is used to set the weights for the matrix. After setting the weights, the inference line is used to carry out the computation.

The experimental setup consists of inference and programming lines, as shown in Figure 4.9. The inference line is used to carry out the computation task. A supercontinuum light source is used, and using a dense wavelength division module (DWDM), C33-C35 is selected. Selected wavelengths act as input vectors, and the modulation is achieved using VOA, which is electronically controlled using a microcontroller. The input vectors' polarisation is controlled and sent into the photonic chip. The output of the matrix-vector multiplication is collected on photodetectors, which are recorded on a computer using DAQ.

The programming line is used to set the weights of the PCM memory cell. The programming line consists of a tuneable laser. A pulse shape is generated with the help of an electro-optic modulator (EOM). The obtained pulse shape is further amplified with the help of an EDFA. The amplified signal is sent to the chip through an optical circulator, which acts as a common channel for Pump line input and Probe line Out-1.

### **Supercontinuum light source**

A supercontinuum laser from NKT photonics, with a power output of 20 W, is used. The white laser is connected with a splitter to separate visible and NIR wavelengths.

### **DWDM**

A dense wavelength division module ITU grid from Gehzi (DWDM-100G-DEMUX) is used to demultiplex the light from a white laser source into 4 channels. Channel 33-35, corresponding to 1550.92 nm, 1550.12 nm and 1549.32 nm, are selected as input light sources for vector multiplication.

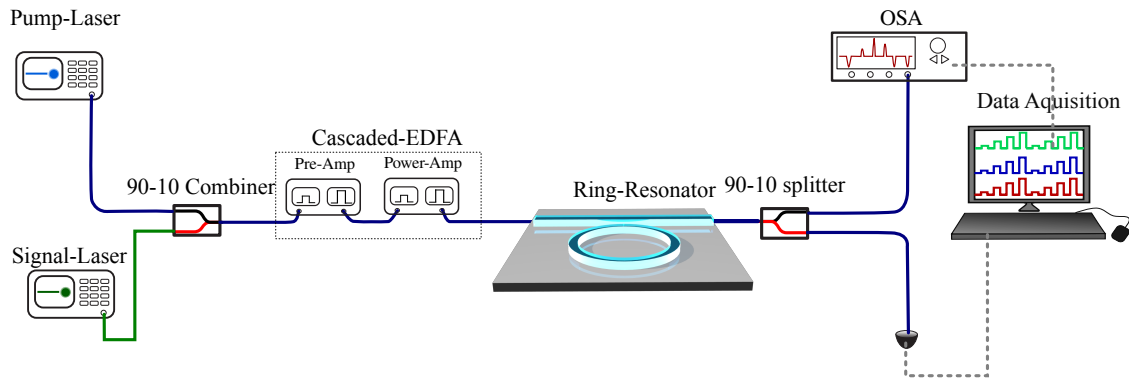
### **VOA**

A variable optical attenuator from Thorlabs (V1550A) is used. The VOA has a modulation speed of up to 1 kHz. This is used to encode the input vectors using

the LabVIEW program. On applying analogue voltage, the transmission response of the VOA can be varied, providing up to 30 dB of attenuation.

#### 4.2.4 Four-wave mixing

To test the nonlinear properties of SiC waveguide structures and get the nonlinear refractive index, four-wave mixing experiments have been performed.



**Figure 4.10:** Experimental setup for four-wave mixing experiments. The setup consists of a pump and signal-tuneable laser sources. The light is combined using a 90/10 splitter, and the light is amplified using a cascaded EDFA system. After the device, the signal is split using another 90/10 splitter. One line is used to monitor the transmission of the devices using a photodetector. The other line is connected to an OSA to record the spectrum.

Four-wave mixing experimental setup consists of two laser sources: Pump and Signal consisting of two tuneable CW Santec lasers, and the power of the laser is amplified using cascaded EDFA Pritel (FA-15 and FA-33) to generate a maximum of 4 W power. The Pump and Signal lines are added using a 90:10 splitter with 90% signal from Pump and 10% from the Signal. After the cascaded EDFA, the light is input to the ring resonators. The polarisation of the input light is controlled using polarisers. At the output end, power is split using a 90/10 splitter. 90% light is directed to an optical spectrum analyser (OSA) to capture the spectrum. The remaining 10% of the light is directed to a photodetector to record the transmission of the device.

# 5

## Antimony Phase change memory

### Contents

---

|            |                                                      |           |
|------------|------------------------------------------------------|-----------|
| <b>5.1</b> | <b>Background . . . . .</b>                          | <b>65</b> |
| <b>5.2</b> | <b>Memory cell: Design and Fabrication . . . . .</b> | <b>65</b> |
| 5.2.1      | Device design . . . . .                              | 65        |
| 5.2.2      | Device Simulations . . . . .                         | 66        |
| 5.2.3      | Device Fabrication . . . . .                         | 67        |
| 5.2.4      | Experimental Details . . . . .                       | 68        |
| <b>5.3</b> | <b>Switching of Sb Memory cell . . . . .</b>         | <b>69</b> |
| 5.3.1      | Hotplate switching of Sb . . . . .                   | 69        |
| 5.3.2      | All optical binary switching . . . . .               | 70        |
| 5.3.3      | Multiple Level Switching . . . . .                   | 72        |
| 5.3.4      | Switching Speed Measurement . . . . .                | 74        |
| 5.3.5      | Need for Capping Layer . . . . .                     | 75        |
| <b>5.4</b> | <b>Volatility in Sb Memory . . . . .</b>             | <b>76</b> |
| 5.4.1      | Retention time measurments . . . . .                 | 77        |
| 5.4.2      | Achieving Tuneable Volatility . . . . .              | 79        |
| <b>5.5</b> | <b>Demonstration of Applications . . . . .</b>       | <b>80</b> |
| 5.5.1      | Background Information . . . . .                     | 81        |
| 5.5.2      | Experimental Setup . . . . .                         | 84        |
| 5.5.3      | Frequency Response . . . . .                         | 85        |
| 5.5.4      | Envelope Detector . . . . .                          | 86        |
| 5.5.5      | Achieving tuneable-low Filter . . . . .              | 89        |
| <b>5.6</b> | <b>Critical Analysis . . . . .</b>                   | <b>90</b> |
| <b>5.7</b> | <b>Chapter Summary . . . . .</b>                     | <b>91</b> |

---

This chapter discusses the development of antimony-based integrated photonic memory on a silicon photonic platform. The objective here is to perform switching of single-element antimony using femtosecond pulses and demonstrate multilevel memory. The later section of this chapter demonstrates the application of antimony for signal processing applications. The data presented in this chapter has been

peer-reviewed and published in [140].

## 5.1 Background

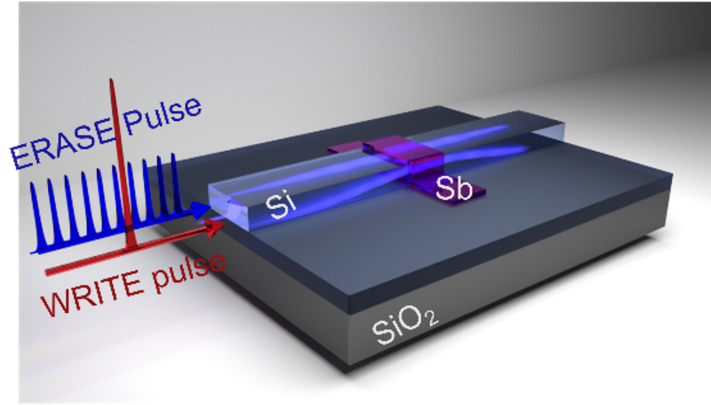
Antimony has been known to exhibit different electrical and optical properties in amorphous and crystalline phases [141]. In the crystalline phase, antimony has a rhombohedral crystal structure with  $a = b \neq c$  and  $\alpha = \beta = 90^\circ, \gamma = 120^\circ$ . However, synthesis of the amorphous phase from the crystalline phase using the melt-quench process was not possible due to the spontaneous recrystallisation of antimony. Recently it was shown that nanoconfined antimony could be switched electrically [142, 143] and optically [144]. As discussed in Chapter 2, conventional PCMs such as GST and AIST are limited in switching speeds and cyclability due to phase segregation. It has been known antimony has a fast recrystallisation rate and can even recrystallise explosively [145, 146]. Recent works with antimony have shown a recrystallisation speed of 5 ns with a high switching repeatability of 10 million cycles [147], thus demonstrating the potential of antimony as a PCM for analogue in-memory computing.

## 5.2 Memory cell: Design and Fabrication

The following section covers the design and experimental details for the development of antimony-based integrated photonic memory.

### 5.2.1 Device design

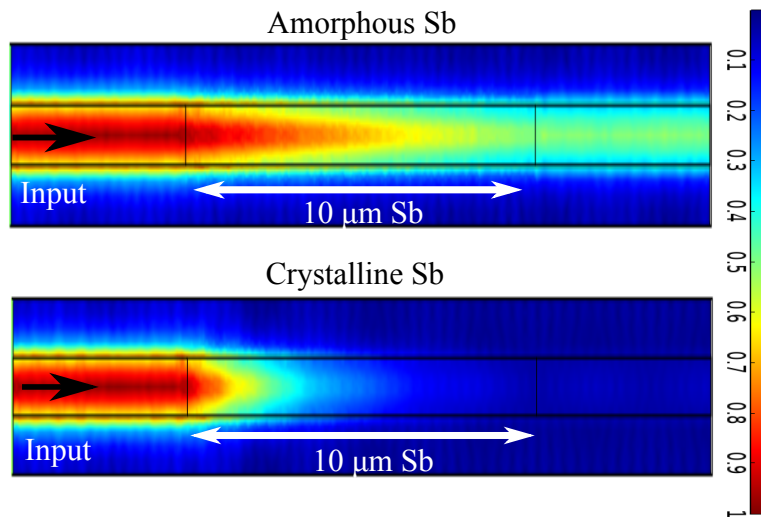
The first set of experiments investigates whether switching antimony in an integrated photonic platform is possible. A photonic waveguide on an SOI platform is designed to achieve this, with PCM evanescently coupled to the waveguide, as illustrated in Figure 5.1. A single-mode waveguide with an operating wavelength of 1550 nm is designed. Light is coupled into and out of the device using grating couplers designed to operate in wavelengths centred at 1550 nm.



**Figure 5.1:** Schematic of SOI waveguide with 3 nm Sb, an amorphisation pulse (WRITE) or crystallisation pulse (ERASE) is used to switch the PCM.

### 5.2.2 Device Simulations

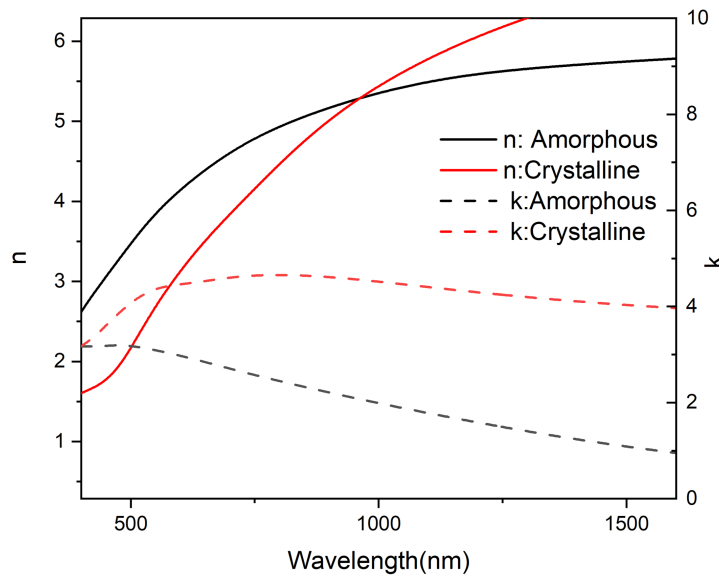
To check the interaction of thin film Sb with the light and to calculate the degree of transmission through the waveguide for Sb in both the amorphous and crystalline state, finite element modelling simulations are performed in COMSOL Multiphysics Frequency Domain. The simulation results show a change in transmission of a waveguide depending upto the change of antimony, Figure 5.2. This change in transmission arises due to the difference in absorption of the two states.



**Figure 5.2:** Frequency domain simulations using COMSOL showing the E-field distribution in a silicon waveguide with 3 nm thick Sb. The colours represent the normalised (in the range of 0-1) intensity of the electric field. Crystalline Sb is highly absorptive and thus results in lower transmission.



The light within the waveguide couples evanescently to the Sb film; the crystalline phase of nanometer-thin films of Sb absorbs more light when compared to its amorphous phase, as it has a higher attenuation coefficient  $k$  ( $k=2.81$  for crystalline and  $k=0.61$  for amorphous Sb at 1550 nm [144]). The refractive index of Sb has been obtained by previous students and reported in [144], is used in the simulation and is plotted in Figure 5.3.

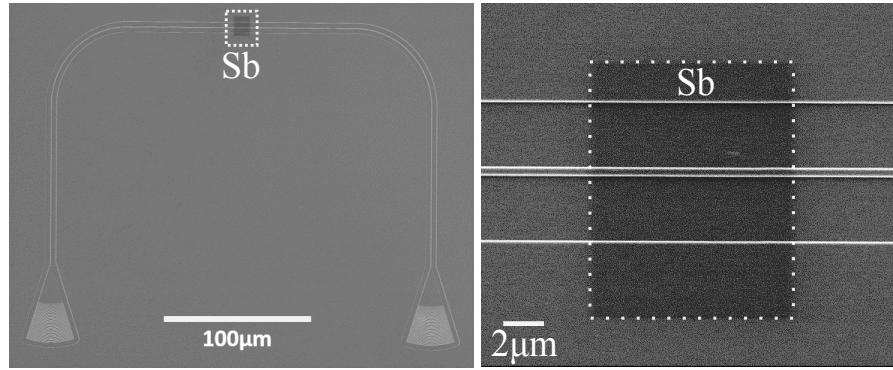


**Figure 5.3:** Refractive index ( $n,k$ ) of amorphous (black) and crystalline (red) antimony thin film from visible to near IR wavelengths

### 5.2.3 Device Fabrication

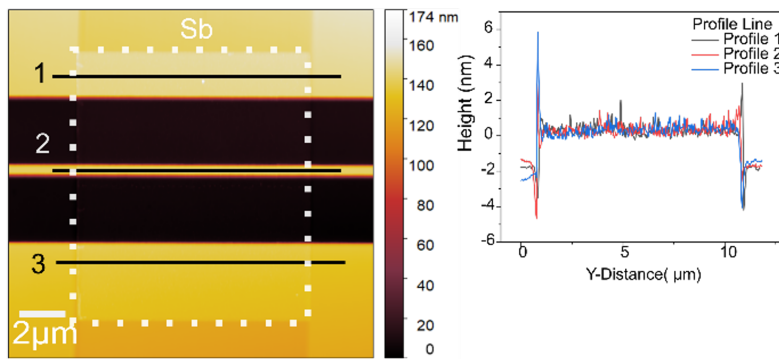
From the simulation results, it is expected to observe a change in the transmission of the waveguide, which depends on the solid phase of Sb. Therefore, a single-mode, partially etched SOI waveguide (450 nm wide on 220 nm SOI substrate with etch depth of 120 nm) is fabricated, following the fabrication steps described in Section 4.1. Following a lithography process to define windows for PCM deposition, thin films of antimony are deposited using RF sputtering, using the sputtering conditions described in Table 4.1.

From the literature, it is known that it is important to confine the thickness of the Sb films below 10 nm to obtain nonvolatile switching behaviour [142, 144].



**Figure 5.4:** An SEM image showing the device structure, which includes a waveguide with grating couplers. The white dotted box shows the PCM deposited on the waveguide.

The sputtering rate of Sb is optimised by calibrating the deposition rate of the RF sputtering system. TEM measurements performed by previous students from the group reveal as-deposited antimony to be amorphous [144]. Thus, Sb of thickness 3 nm is sputtered on top of the waveguide to achieve switching. Figure 5.4 shows the SEM images of the final fabricated devices and a zoomed-in SEM image showing sputtered Sb film on the waveguide. The thickness of the sputtered Sb is measured using an atomic force microscope (AFM), as shown in Figure 5.5.



**Figure 5.5:** AFM image confirms the thickness of Sb to be 3 nm. Profile lines at three regions have been plotted, confirming uniform deposition.

### 5.2.4 Experimental Details

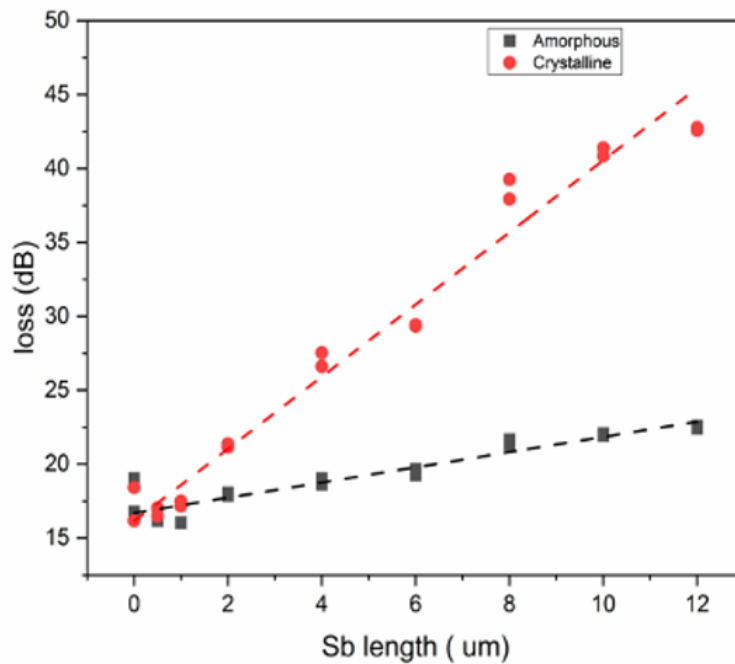
The details of the experimental setup used have been described in Chapter 4. Mainly two experimental setups have been used in the following section. The first setup is to measure the device's transmission. This has been used to obtain absorption

loss of antimony in amorphous and crystalline states as PCM lengths vary. The second setup is a pump-probe setup. This is used to achieve all-optical switching of antimony using femtosecond lasers and to measure the change in transmission of the device.

### 5.3 Switching of Sb Memory cell

This section covers the experimental results of achieving antimony memory cells. First, a static switching method based on a hotplate is demonstrated and then an all-optical switching of Sb is achieved using optical pulses. Further multiple memory levels are demonstrated.

#### 5.3.1 Hotplate switching of Sb



**Figure 5.6:** Experimental results showing the effect of increasing Sb length on a waveguide on the loss (absorption of light) in as-deposited amorphous and hot plate annealed crystalline state. Increasing the length of Sb results in a higher contrast between amorphous and crystalline states.

To measure the change in transmission of the photonic device and to measure

the absorption of antimony thin film in amorphous and crystalline states, different lengths of Sb were deposited. The transmission of the as-deposited Sb devices with varying lengths is noted. The sample is annealed on a hot plate at 230° C for 5 minutes to crystallise Sb thin film completely, and once again, the transmission of the devices with increasing lengths of Sb is noted.

As shown in Figure 5.6, the transmission measurements agree with simulation results, where the amorphous phase has lower loss than the crystalline phase. Moreover, there is a linear relation in increasing the length of Sb. Increasing the length of Sb results in higher absorption and hence a larger loss.

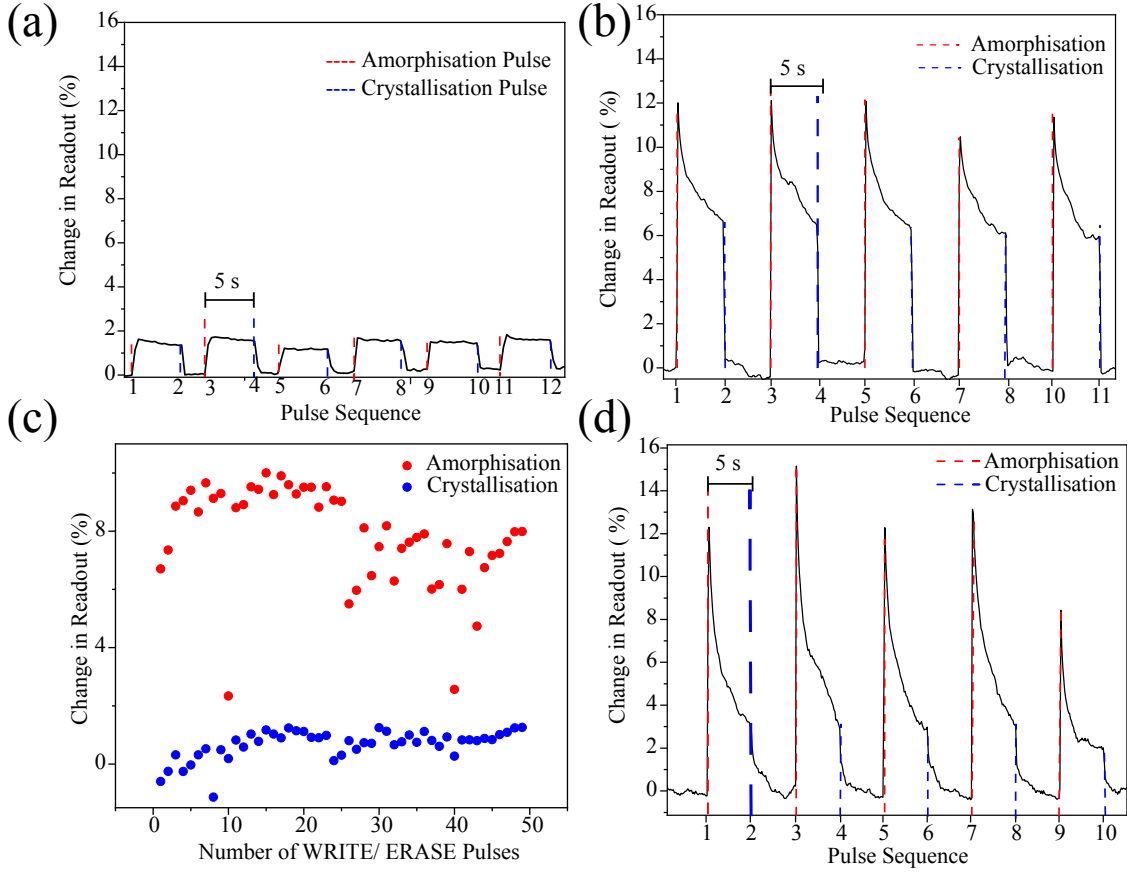
### 5.3.2 All optical binary switching

To switch antimony reversibly, the thin films are initially crystallised by annealing the device at 230° C for 5 minutes in the air on a hot plate; this allows the Sb to be more absorptive, enabling more efficient near-field absorption of light. Using the pump-probe technique, antimony memory cells are switched, and the change in the transmission is recorded.

Antimony can be switched to an amorphous state using a single femtosecond pulse (Write) of high energy of  $194 \pm 35$  pJ. This is recorded as increased transmission through the device using the probe laser, as this state is less absorptive. To crystallise (Erase) the material, One hundred low-energy pulses are sent, with individual pulses of energy  $45 \pm 9$  pJ (total energy 4.5 nJ).

A lower energy pulse results in a smaller volume above the recrystallisation temperature. Therefore, multiple pulses are used to achieve reversible switching. It is worth mentioning that for this work, the system's thermal design has not been optimised; therefore, the optimal pulse energies and pulse sequence required to switch materials have to be determined experimentally.

As shown in Figure 5.7(a), the material can be switched reversibly between the amorphous and crystalline states for lengths of Sb down to 4  $\mu\text{m}$ . Higher contrast



**Figure 5.7:** Binary switching on a waveguide, a) Sb (length = 4  $\mu\text{m}$ ) on a waveguide switched using single, high energy ( $194 \pm 35 \text{ pJ}$ ) femtosecond pulse (800 fs) – (amorphisation pulse, red dashed vertical lines indicate this pulse). 100 low energy ( $45 \pm 9 \text{ pJ}$ ) pulses (crystallisation sequence, blue dashed lines indicate when this is sent) crystallise the sample. b) Sb (length = 10  $\mu\text{m}$ ) switched between amorphous and crystalline phases as in a). As expected, longer length results in higher contrast. c) Multiple switching of the device in (b). d) Demonstration of single-pulse amorphisation and crystallisation for the same device as in (b-c) by fine-tuning the pulse energies; a single pulse is enough for binary switching. The red dashed lines represent the pulse number at which an amorphisation pulse of energy  $91 \pm 7 \text{ pJ}$  is sent, while the blue dashed lines represent a crystallisation pulse of energy  $59 \pm 11 \text{ pJ}$ .

in the readout signal of 8% can be achieved by increasing the length of Sb to 10  $\mu\text{m}$  as reported in Figure 5.7 (b). Here the contrast is defined as the percentage change in the readout signal at time  $t=5 \text{ s}$  after the amorphisation pulse. For consistency of results, unless mentioned, the length of Sb has been fixed to 10  $\mu\text{m}$  length for all subsequent experiments.

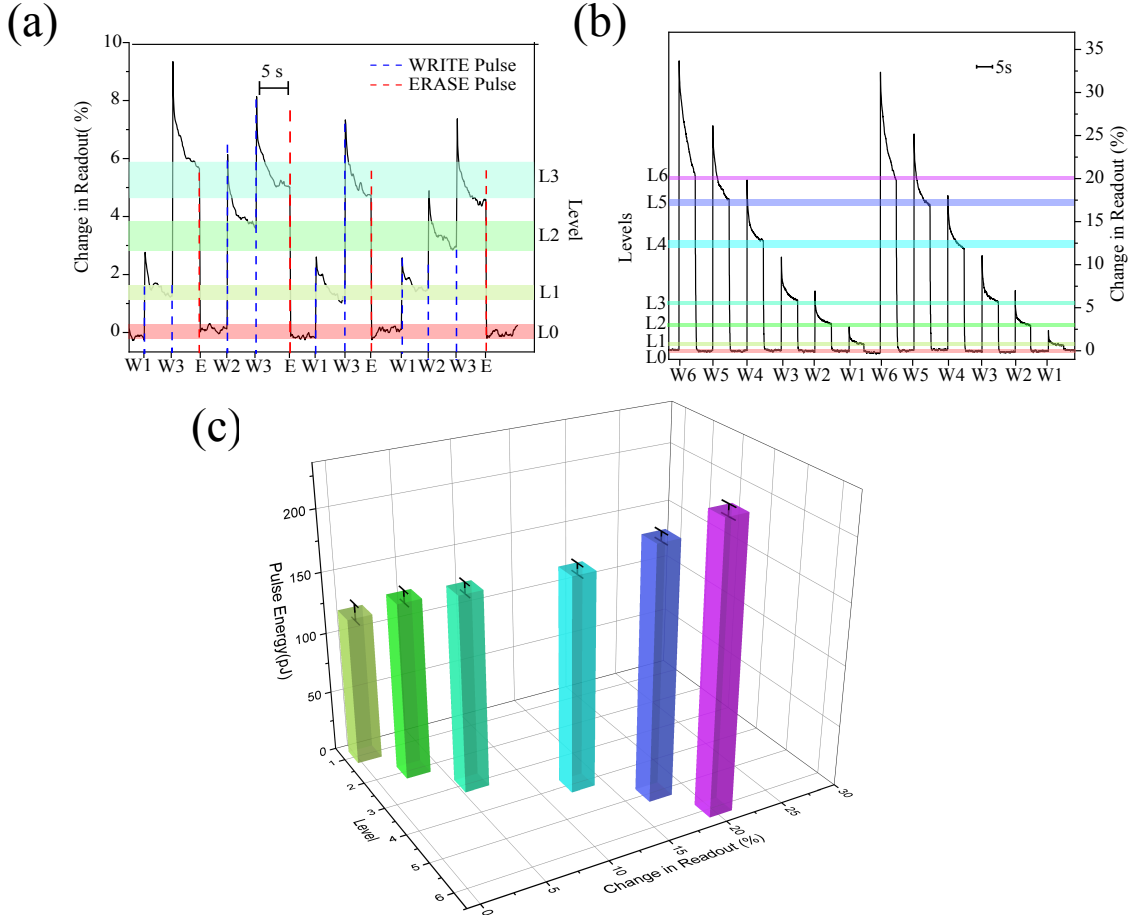
To test the reliability of switching Sb, the switching process is repeated for more than 50 cycles. Figure 5.7 (c) show the multiple switching of the Sb device with

variability in the switching contrast to be below 2%. This variability is attributed to the variation in the pulse energies. In principle, a single-element PCM should have superior cycling endurance; however, cyclability is a complex phenomenon. It is a function of not only the material but also the delamination of thin films, the determination of optimum switching pulse energies and the use of capping layers. This has not been explored in this thesis and requires further investigation.

For PCM, crystallisation remains the rate-determining step for switching speeds, as it is usually slow. From the literature, the crystallisation time for Sb has been reported to be less than five nanoseconds. Therefore, the Sb memory device should be able to crystallise using single fs pulses. To demonstrate this, a single low-power femtosecond pulse crystallises the material shown in Figure 5.7 (d). When the pulse energy is above the threshold ( $91 \pm 7 \text{ pJ}$ ) required to amorphise Sb, transmission changes, i.e. switching occurs. The transmission goes down for powers below the amorphisation threshold but above that required for crystallisation ( $59 \pm 11 \text{ pJ}$ ). Thus crystallisation occurs. However, as seen in Figure 5.7 (d), using a single crystallisation pulse does not form a stable memory level. Unlike multiple crystallisation pulses, a single crystallisation pulse results in a partially crystalline Sb. This intermediate crystalline level results in lower absorption of pulse energy and hence a lower contrast.

### 5.3.3 Multiple Level Switching

The subsequent experiment seek to verify whether it is possible to reach intermediate transmission states by partially switching Sb. Analogous to photonic phase change memories [54], increasing the switching volume results in higher contrast, as a larger area interacts with the near-field of the light transmitted within the waveguide. The volume of Sb amorphised or crystallised can be controlled by changing the power. Thus multilevel programming is achieved as a function of discrete transmission states by fine-tuning the pulse energy.



**Figure 5.8:** a) Multilevel programming in Sb. Four distinct levels of transmission are programmed (termed L0, L1, L2 and L3) using a single femtosecond (pulse width 800 fs) WRITE pulse (termed W1, W2 and W3) of energies  $145 \pm 6$  pJ,  $150 \pm 6$  pJ and  $160 \pm 5$  pJ respectively. ERASE is achieved by sending 200 pulses of 100 pJ femtosecond pulses. b) Repeatedly reaching seven memory levels (L0-L6) by WRITE pulses W1-W6 (Pulse energies  $121 \pm 7$  pJ,  $144 \pm 5$  pJ,  $160 \pm 5$  pJ,  $174 \pm 4$  pJ,  $204 \pm 4$  pJ and  $231 \pm 4$  pJ respectively). c) The pulse energies corresponding to each level.

Four distinguishable states are programmed by varying the pulse energy, as shown in Figure 5.8 (a). By sending a single pulse of fixed (800 fs) width and varying the pulse energy ( $145 \pm 6$  pJ,  $150 \pm 6$  pJ and  $160 \pm 5$  pJ), three distinguishable states can be achieved reliably. These states can be achieved arbitrarily and sequentially in ascending order by sending a WRITE pulse corresponding to the desired level. For example, in Figure 5.8 (a), a single pulse of 800 fs duration and 160 pJ energy is sufficient to reach level L3 from either of the lower levels, L1 or L2. Once at a higher level, the lowest level, L0, can be achieved by sending 100 low-energy pulses

of energy  $101 \pm 14$  pJ (ERASE).

Furthermore, by increasing the WRITE pulse energies (from  $121 \pm 7$  pJ to  $231 \pm 4$  pJ), seven distinguishable memory levels are achieved as shown in Figure 5.8(b). It is worth noting that the variation in the pulse energies reported here is because of the noise of the energy-meter sensitivity (Ophir PD10-IR). Increasing the switching energy increases the switching contrast; thus, a larger area is switched to an amorphous state. Therefore, a higher number of crystallisation pulses are sent for recrystallisation: 200 pulses in this higher contrast regime compared to 100 pulses in the binary regime.

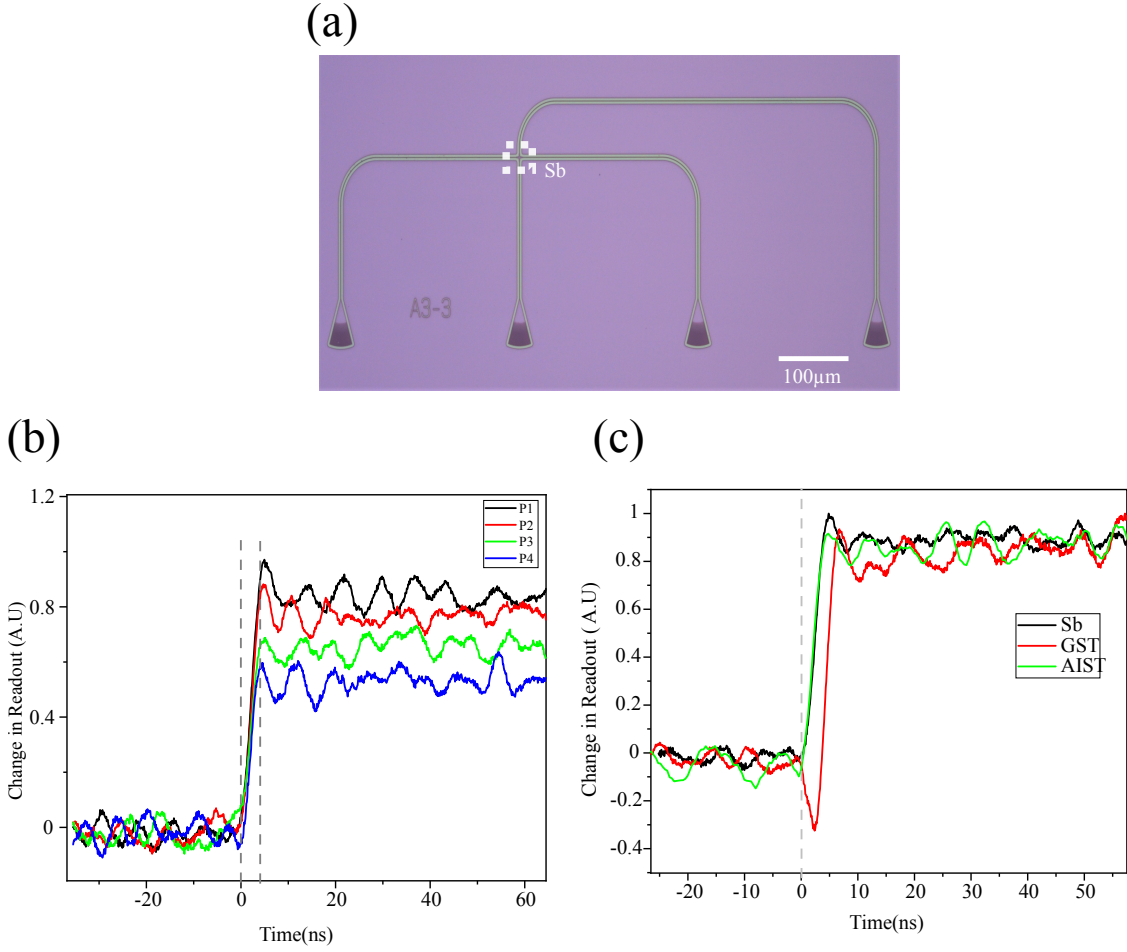
### 5.3.4 Switching Speed Measurement

Time-resolved switching experiments using a 4-port device, shown in Figure 5.9(a), are performed to calculate the switching speed of Sb on the waveguide. Probe signal output is connected to a 125 MHz photodetector to record the change in transmission on sending a ‘Pump’ pulse. A single write pulse switches the material from a crystalline state to an amorphous state. Due to the low signal-to-noise ratio, the change for ten readings is averaged and reported.

For switching, the signal rise time (from 10% of change to 90%) of 2 ns is observed, corresponding to an operation speed of up to 500 MHz, Figure 5.9(b). The experiments are repeated for different pump power to achieve different memory levels using a single sub-picosecond pulse of energy P1-P4, with decreasing pulse energies from P1-P4.

Furthermore, the switching speed of Sb is compared with other well-known phase change materials using the pump-probe technique. The switching speeds of Sb, AIST and GST are on the same timescales as shown in Figure 5.9(c). However, as shown in previous work on the optical switching of antimony [144], a very short femtosecond pulse is enough to amorphise Sb. Therefore the current switching speed for Sb is limited only by the photonic system rather than the material itself.



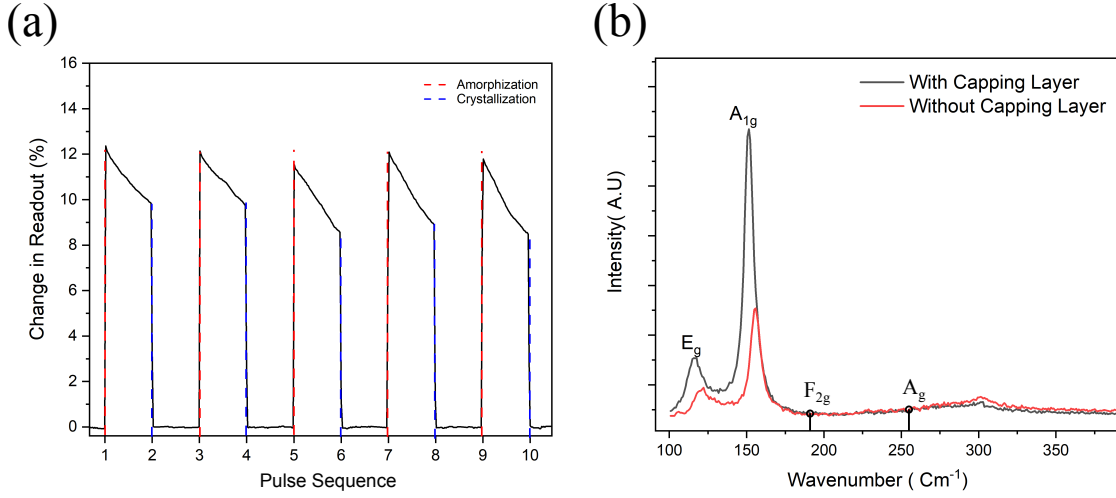


**Figure 5.9:** (a) Optical image of 4 port device used to measure the switching speed. Phase change material is deposited on the optical crossing. Using one of the ports as input for the pulse and the other as the probe, time-resolved switching characteristics are obtained. (b) Time-resolved switching dynamics of Sb, resulting in a rise time of 2 ns for a single sub-picosecond pulse on switching from crystalline to amorphous state. (c) Experimental comparison of the switching speed of Sb is with other known phase change materials such as GST and AIST using a single sub-picosecond pulse for amorphisation.

### 5.3.5 Need for Capping Layer

Phase change materials such as GST and AIST need a capping layer to prevent them from oxidation, thus resulting in their performance degradation. Therefore the performance of the Sb memory device is studied with an additional 10 nm capping layer of indium tin oxide (ITO).

Due to increased absorption of light due to the capping layer, a 4 μm long device is enough to get a 10% change in contrast as compared to a 10 μm long uncapped

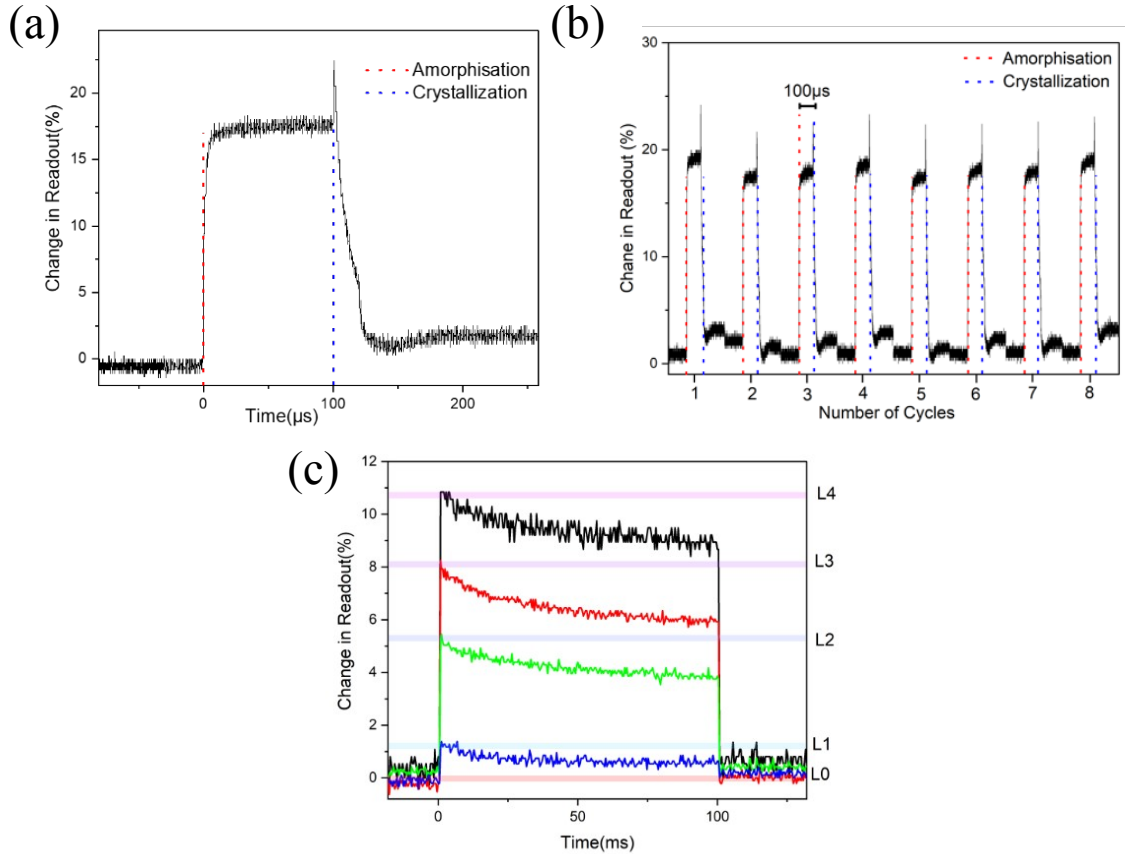


**Figure 5.10:** a) Binary switching  $4\ \mu\text{m}$  long of Sb with 10 nm ITO capping on the using single, high energy ( $194\pm35\ \text{pJ}$ ) femtosecond pulse (800 fs) – (amorphisation pulse, red dashed vertical lines indicate this pulse). 100 low energy ( $45\pm9\ \text{pJ}$ ) pulses (crystallisation Pulse, blue dashed lines indicate this pulse sequence) crystallise the sample. b) Raman spectra of a capped and uncapped crystallised thin film of Sb on waveguide with their  $E_g$  and  $A_{1g}$  vibration modes. The circles denote the typical  $F_{2g}$  and  $A_g$  modes of  $\text{Sb}_2\text{O}_3$ , clearly showing the absence of any antimony oxide in the samples.

device Figure 5.10(a). Further, capped and uncapped crystalline antimony samples are studied using Raman spectroscopy after annealing in air at  $230^\circ\text{C}$ . No antimony oxide ( $\text{Sb}_2\text{O}_3$ ) peaks at  $191\ \text{cm}^{-1}$  and  $255\ \text{cm}^{-1}$  were observed in the samples Figure 5.10(b). Thus proving no need for a capping layer. It is important to note that the oxide of antimony,  $\text{Sb}_2\text{O}_3$  could still be present in amorphous form or on the surface, which remains undetectable using Raman spectroscopy. Other techniques, such as XPS, can be used to detect the presence of any oxides; However is beyond the scope of this thesis.

## 5.4 Volatility in Sb Memory

This section covers the characterisation of the volatile behaviour in Sb memory cells by characterising the retention time of memory cells and next exploring methods of achieving tunable volatility.

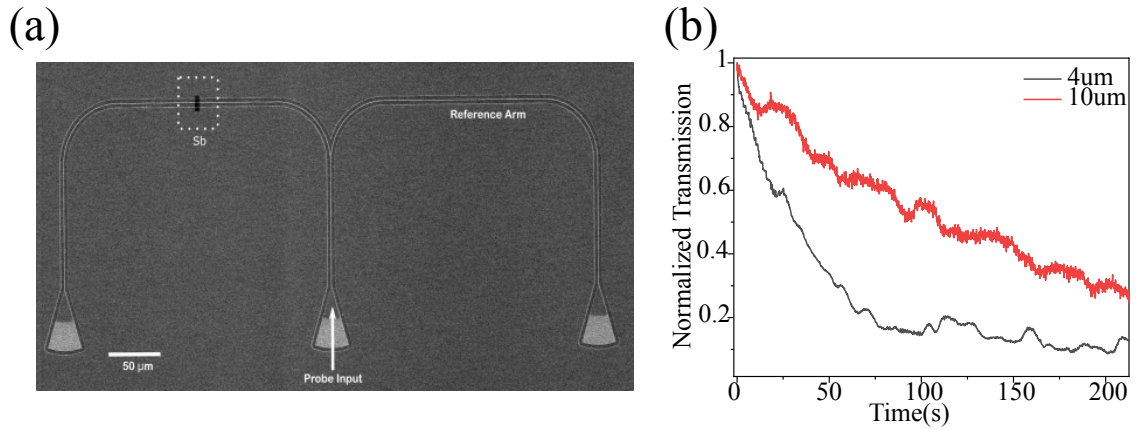


**Figure 5.11:** (a) Sub-millisecond readout for binary switching of 10  $\mu\text{m}$  long of Sb. (b) Experiments in (a) were repeated multiple times, with less than 2% variation in memory levels. (c) Millisecond readout for intermediate memory levels obtained for different amorphisation pulse energies. Memory levels show a variation of 1% in readout over 100 ms time duration.

#### 5.4.1 Retention time measurements

Antimony memory cells demonstrate volatility in tens of seconds scale, which can limit their functionality for analogue computing applications. Therefore the time scales associated with such volatility are assessed by measuring the stability and repeatability of achieving the binary memory levels in the sub-millisecond range. 10  $\mu\text{m}$  Sb device is amorphised using a single femtosecond pulse and subsequently crystallised using 100 low energy fs pulses after 100  $\mu\text{s}$ . The change in the transmission is recorded using an oscilloscope Figure 5.11(a). The switching is repeated multiple times and shows repeatability with less than 2% variation; multiple plots are stitched together and presented in Figure 5.11(b).

Next, the stability and repeatability of the device to achieve four intermediate memory levels in the millisecond range are assessed.  $10\ \mu\text{m}$  Sb device is amorphised and subsequently crystallised as before. The change in the transmission is recorded using an oscilloscope for different energy of amorphisation pulses 5.11 (c). The device shows the stability of intermediate memory over 100 ms with a variation readout of 1%. Thus with a clock speed of only 1 GHz, millions of operations are possible, sufficient to carry out calculations such as Tensor Core operations [15, 59, 70, 148] and associative learning [62].

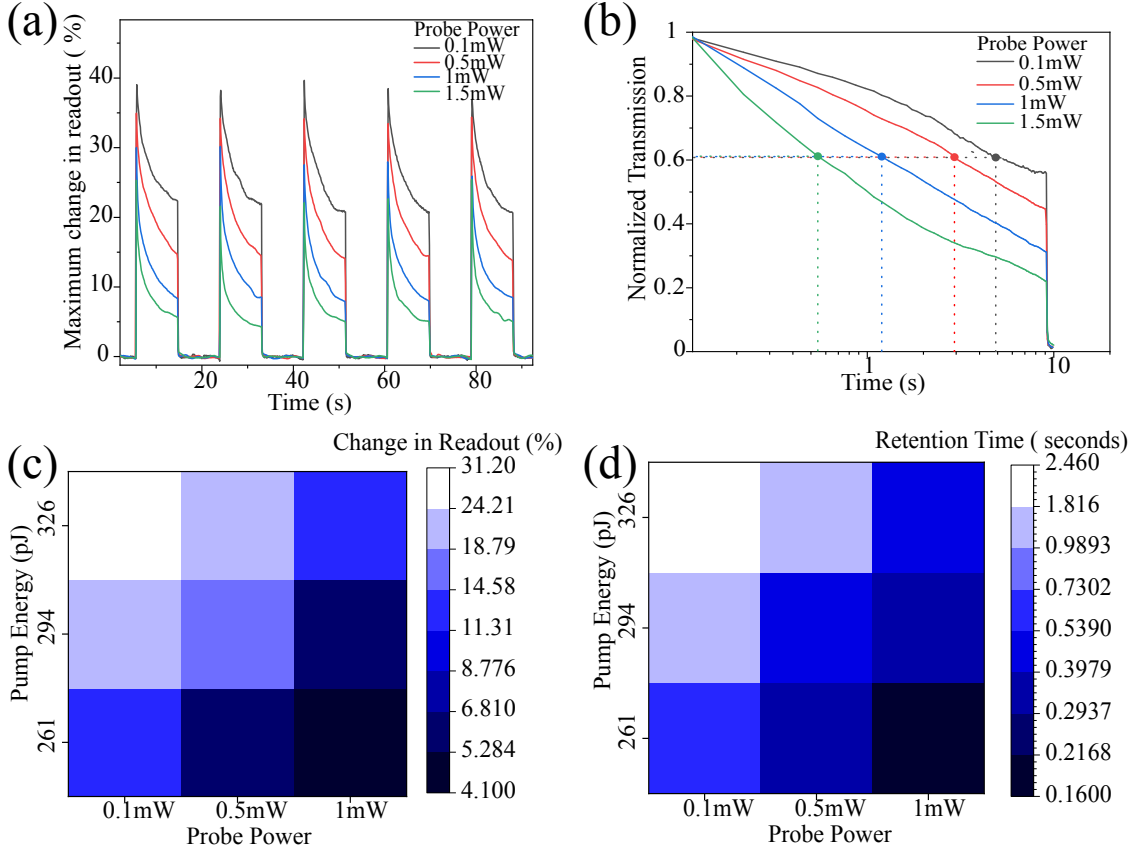


**Figure 5.12:** (a) SEM image of the device used for long-term stability experiment. The probe signal is split using a Y-splitter into two arms one arm is used to measure the transmittance through Sb, and another reference arm is used to monitor the input power fluctuations and drift in the optical signal. The signal from the reference arm is used to normalise the transmission from the Sb arm and calculate the retention time of switched Sb, (b) Long-term stability of  $4\mu\text{m}$  and  $10\mu\text{m}$  long amorphous Sb, switched from crystalline phase using 409 pJ, a single femtosecond pulse.

Next, the long-term switching stability of amorphous Sb for the memory device, in tens of second scale, is studied. A reference device without Sb accounts for stage drift and fluctuations in probe laser power over time, as shown in Figure 5.12(a). At time  $t=0$ , the memory cell is switched to the amorphous phase using a single femtosecond pulse of 409 pJ and the drop in transmission is monitored. The  $4\ \mu\text{m}$  and  $10\ \mu\text{m}$  long devices have a recrystallisation of 100 and 200 seconds, respectively, as reported in Figure 5.12(b). The intrinsic transient behaviour can be used to

implement leaky integrate and fire neurons [149, 150]. It can be used for applications requiring the combination of short and long-term plasticity in neurons [151].

### 5.4.2 Achieving Tuneable Volatility



**Figure 5.13:** a) Effect of the Probe Power on the maximum transmission achieved for a given Pump power. An increase in probe power results in a decreased maximum change in the readout. b) Effect of probe power on the retention time. Individual plots are normalised between 0 and 1, corresponding to the lowest (crystalline) and the highest (amorphous) transmission. The time required to reach 60% of the maximum change in readout is plotted. Increasing the probe power decreases the retention time. c) Effect of both Probe and Pump power on the maximum readout contrast. A high Pump and low probe power lead to the maximum change in the readout. d) Effect of Pump and Probe power on retention time (time required to reach 90% of maximum readout change). The volatility of Sb is controlled by varying either or both.

The effect of probe power on switching contrast and stability of memory states is investigated to achieve tuneable volatility. An increase in probe power decreases the maximum readout contrast, as reported in Figure 5.13(a). As in previous sections, a single femtosecond pulse for amorphisation and multiple low-energy pulses for

crystallisation. Further, the effect of probe power on the retention time of a memory state is studied. Figure 5.13(b) reports the fall in transmission from the amorphous state to 60% of the maximum transmission while changing the probe power. The transmission plots are normalised [0,1] for comparison.

An increase in probe power leads to a faster decay time. This behaviour is expected, and similar behaviour has been observed in phase change materials such as GST [152], where an increase in probe power results in recrystallisation and, therefore, shorter retention time. However, unlike GST, where higher probe power leads to higher contrast, for Sb, a lower contrast is observed.

Further, the effect of change in both pump and probe power on switching behaviour and the retention time is studied. These experiments result in two important observations. First, as shown in Figure 5.13(c), a higher pump power leads to an increased contrast in the readout and increased probe power results in decreased contrast.

Second, there is a direct correlation between contrast and retention time. As represented in Figure 5.13(c) and Figure 5.13(d), higher contrast results in higher retention time. Retention time is defined as the time required to reach 90% of the maximum contrast. The samples with similar contrast (probe power 0.1 mW and Pump energy 294 pJ) and probe (0.5 mW with pump energy 326 pJ) have a similar retention time of 1.1 seconds.

Understanding this direct correlation between contrast and retention time requires further analysis of recrystallisation dynamics using in-situ imaging techniques, which is beyond the scope of this thesis.

## 5.5 Demonstration of Applications

In this section, an Sb device has been demonstrated for application in signal processing. The sb filter cell has been shown to work as a low-pass filter, then the

frequency response is calculated and achieving a tuneable filter is proposed. This section of the thesis is currently under review for publication in a peer-reviewed journal.

### 5.5.1 Background Information

RF filters in the optical domain offer benefits such as the large bandwidth offered by photonics and the insensitivity to electromagnetic interferences [153]. Moreover, such signal processors can be embedded with optical sensors in IOT-based devices to preprocess the data and thus help reduce the computation demands. So far, there has been great interest in achieving such filtering operations[154–159]. However, most of the work involved the use of coherent light sources and often used ring resonator-based devices[160, 161].

The volatile switching behaviour of antimony has a response similar to that of a leaky integrator and, thus, should be able to perform signal processing. For a stream of sampled data, a leaky integrator adds up the input data while forgetting some of the previous data, ‘leaking’ it. A leaky integrator can thus be used to average a stream of data or acts like a low pass filter and is represented by the following equation :

$$y[n] = \alpha x[n] + (1 - \alpha)y[n - 1], \text{ where } 0 \leq \alpha \leq 1 \quad (5.1)$$

Where  $y[n]$  is the  $n^{\text{th}}$  data sample, and  $\alpha$  is the leaky parameter. Thus, for  $\alpha = 1$ , no averaging takes place, and thus reducing  $\alpha$  results in a stronger averaging. This can be further analysed by taking the Z-transform to get,

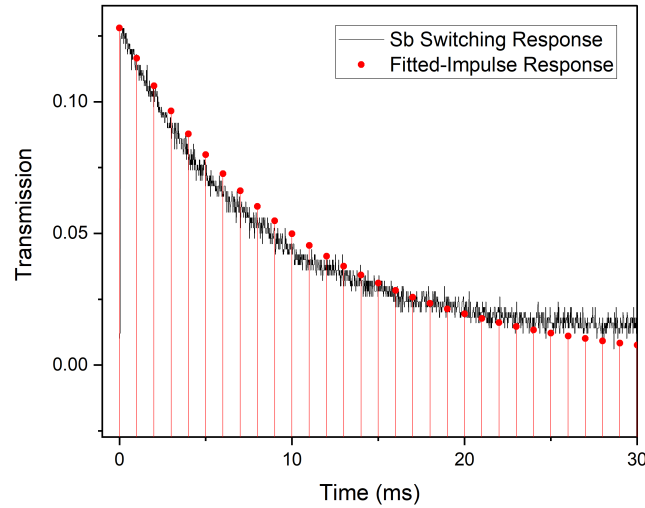
$$Y(z) = \alpha X(z) + (1 - \alpha)Y(z)z^{-1} \quad (5.2)$$

which can be rearranged to get the transfer function  $H(z)$  :

$$H(z) = \frac{Y(z)}{X(z)} = \frac{\alpha}{1 - (1 - \alpha)z^{-1}} \quad (5.3)$$

Thus the impulse response of a leaky integrator can be written as follows:

$$h[n] = \alpha(1 - \alpha)^n \quad (5.4)$$

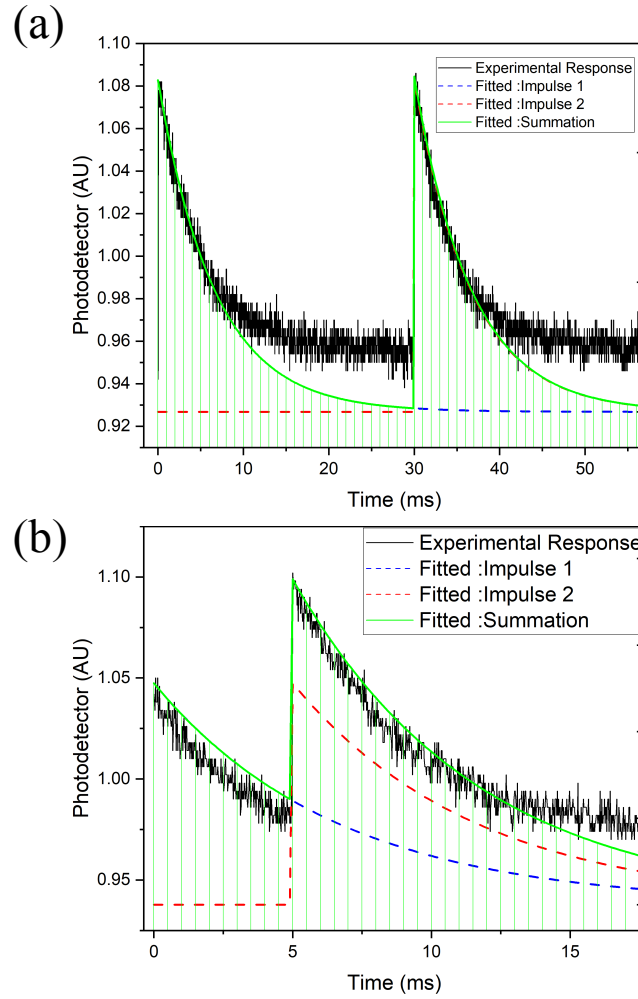


**Figure 5.14:** Impulse response of a leaky integrator is fitted with the switching response for the Sb filter device. An sb thin film of 5 nm is used.

The single pulse response of the antimony filter device is thus fitted with the impulse response of a leaky integrator, as shown in Figure 5.14, and shows good agreement. Next, the superposition property of the Sb device is investigated. For a linear filter with response  $H(s)$  and stimulus  $s_1$  and  $s_2$ ,  $H(s_1 + s_2) = H(s_1) + H(s_2)$ , i.e. the response of two stimuli is the summation of individual responses. Therefore two pulses separated in time are sent through the device. In the first case, the two pulses are separated by 30 ms, as shown in 5.15 (a). The second pulse is sent ‘long’ time after the first pulse. Thus the response of the device is similar to impulse responses separated by 30 ms. The impulse responses for 2 pulses are fitted and shown as dotted lines in blue and red.

In the second case, the two pulses are separated by a ‘short’ duration of 5 ms. As before, the impulse responses of 2 pulses separated by 5 ms are plotted as dotted lines in blue and red in Figure 5.15(b). The summation of the two





**Figure 5.15:** Superposition of impulse response for Sb device. Experimentally obtained response (in black) is fitted to simulated responses (in blue and red). The summation of two simulated responses (in green) matches well with the experimentally obtained response. (a) The impulse response when two pulses are separated by a ‘long’ duration of 30ms. (b) Impulse response when two pulses are separated by a ‘short’ duration of 5 ms.

impulses is shown in green. The summation of two impulse responses fits well with the transmission response of the Sb filter device. Thus showing the linearity of stimulus responses.

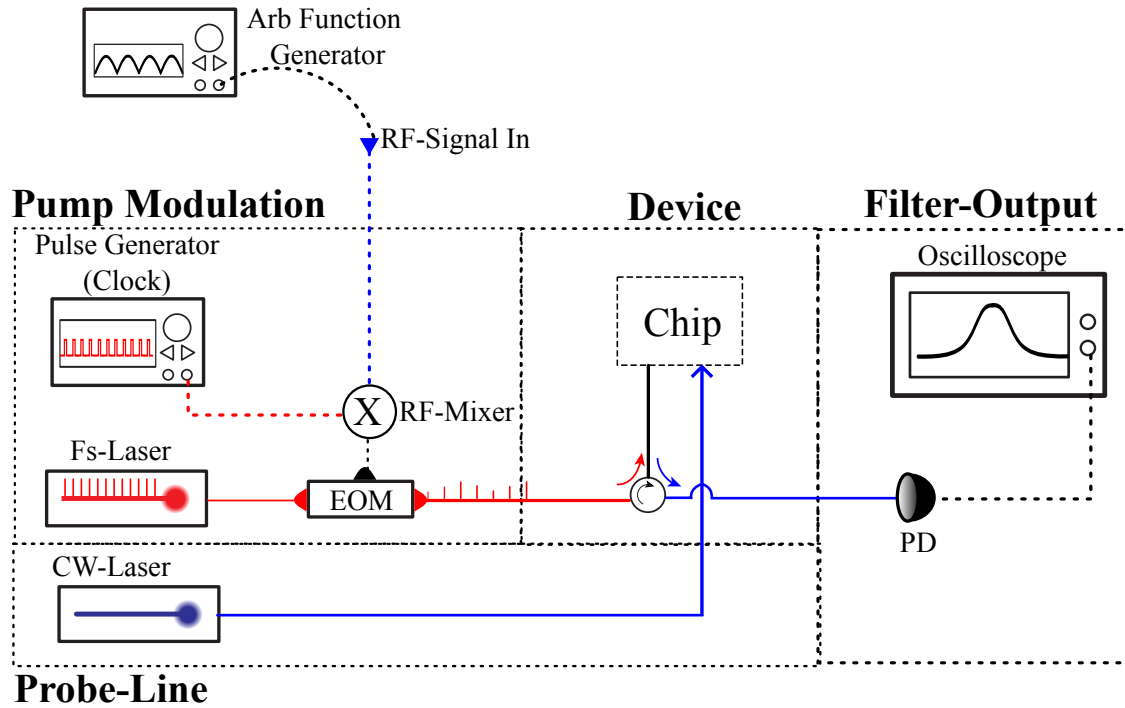
Any input signal can be discretised in time and represented as impulses with weighted amplitude and shifted in time. Thus, the response of such input signals through the Sb filter device can be predicted easily as the sum of the impulse response of Sb filter cells weighted with the impulse amplitude. Another consequence of this property is that the output at any time  $t$  is the weighted average of all the

previous impulse responses, therefore resulting in the smoothing of the input signal, or in other words, acts like a low pass filter. Therefore to test this property of a low pass filter for the Sb device, the transmission response for various sine waves of increasing frequency is studied in detail.

### 5.5.2 Experimental Setup

The experimental setup consists of a pump and probe setup as used before, with additional Pump modulation circuits as illustrated in Figure 5.16. A femtosecond laser is used to switch PCM, and a CW laser is used to monitor the change in transmission. To achieve the filtering task, the RF signal-In has to be encoded onto the femtosecond pulses. This can be achieved using an EOM. Thus, femtosecond pulses can be amplitude modulated. This would result in the discretization of the RF signal. Thus, each sample would be represented as  $X[n_i]$ . The femtosecond pulse laser used in the experiments has a frequency of 40 MHz, resulting in a sampling frequency of 40 MHz, which is too high for the Sb filter device. A clock divider circuit is needed to reduce the sampling frequency.

Therefore, to simplify the experimental setup, an RF mixer is used. A pulse generator is used to generate a clock signal, as illustrated in Figure 5.17a. This is multiplied with any arbitrary RF input signal, say a sine wave, as illustrated in Figure 5.17b. The clock signals get modulated with the input signal as shown in Figure 5.17c. This modulated signal is applied to the EOM, and thus the femtosecond laser can be both amplitude modulated and the sampling frequency can be controlled (Figure 5.17d). The modulated femtosecond pulses are then amplified using an EDFA and sent through the filter device to switch the PCM devices. Using a CW probe laser, the transmission of the device is monitored using a photodetector and recorded onto an oscilloscope.

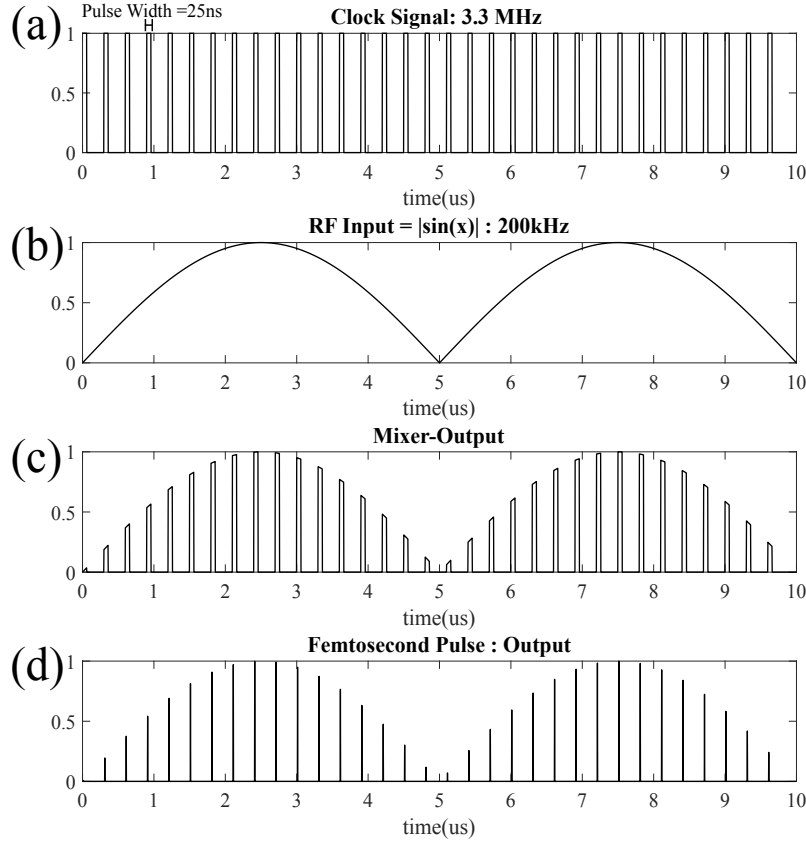


**Figure 5.16:** Illustration of the experimental setup used to achieve filtering operation consisting of pump and probe line. The fs-laser is modulated using an EOM, which is used to switch the Sb filter cell. The output is collected on a photodetector which is read on an oscilloscope.

### 5.5.3 Frequency Response

As discussed above, the Sb filter cell acts like a leaky integrator and thus can be used as a first-order low-pass filter. To characterise the performance of this device, the frequency response of the input signal consisting of rectified sine wave inputs with increasing frequencies is used. The response of the Sb device for various input signals (shown in red) is recorded in Figure 5.18 (shown in blue).

Increasing the input signal frequency results in a decrease in the amplitude of the output sine wave signal, thus performing as a low-pass filter. Next, the analysis for individual frequency response is performed by taking the fast Fourier transformation of the output response. The amplitude (in the dB) of the device output is plotted for various frequencies in Figure 5.19. This plot is used to determine the Pass Band and the Stop Band for the Sb device filter. Due to the low extinction ratio, 2 dB bandwidth is used here. A bandwidth of 300 kHz is obtained for the filter.

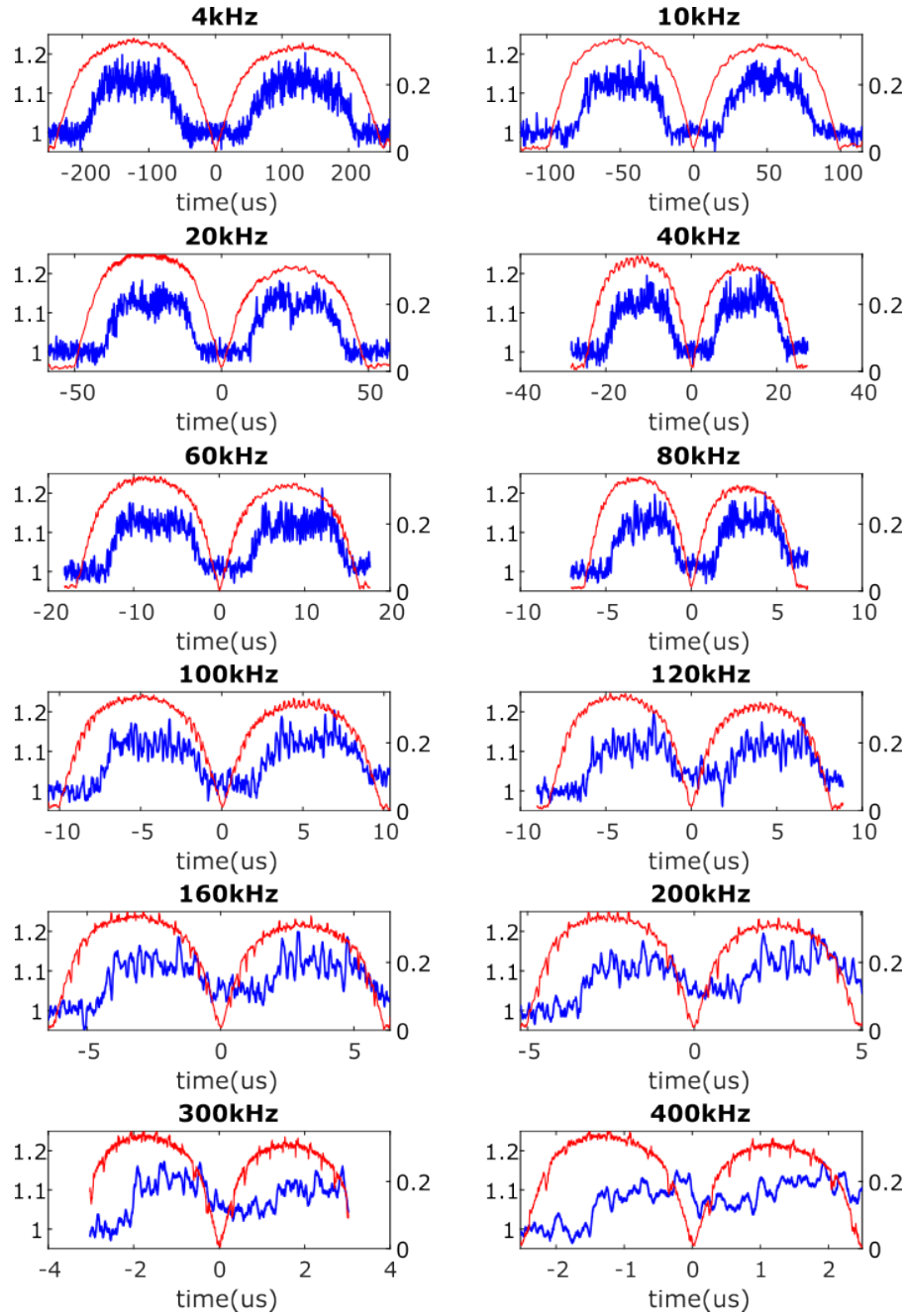


**Figure 5.17:** (a) A clock signal with pulse width 25 ns and Frequency 3.3 MHz is used to downsample the femtosecond pulse train from 40MHz to 3.33 MHz. (b) RF signal,  $|\sin(x)|$  of 200 kHz is used as the RF-input signal, (c) Output of the RF mixer obtained that is sent to the EOM to modulate femtosecond laser pulses. (d) Downsampled and amplitude-modulated femtosecond pulses obtained that is used to switch Sb.

Further, to showcase the low pass filtering property use of the Sb device for envelope detection is demonstrated in the next section.

#### 5.5.4 Envelope Detector

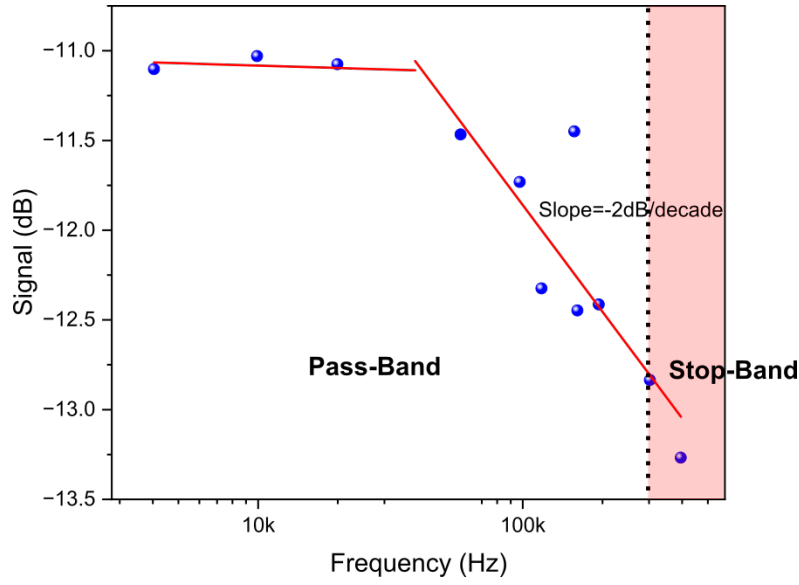
An envelope detector takes in a high-frequency amplitude modulated (AM) signal as input and returns the envelope of this signal, which is the low-pass output. Therefore, to demonstrate the performance of the Sb device as an envelope detector, a 4 kHz sine wave signal is AM with different carrier frequencies ranging from 40 kHz-2 MHz. As determined in the previous section, the Sb device has a 2 dB bandwidth of 300 kHz. Thus, frequencies above this range are filtered from the input signal. The response of the Sb device for various carrier frequencies is



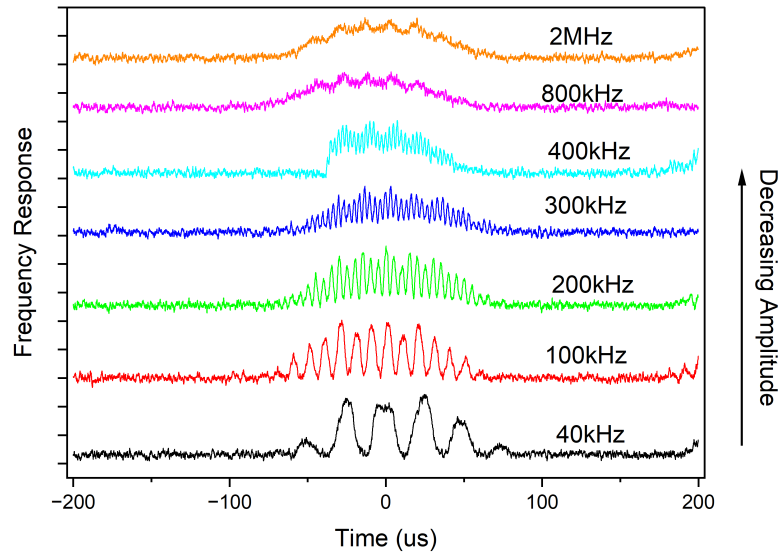
**Figure 5.18:** Sine wave responses of Sb filter cell (in blue) for various input sine wave (in red) of frequencies ranging from 4 kHz-400 kHz.

shown in Figure 5.20.

As expected, frequencies  $>300$  kHz are filtered from the 4 kHz sine wave signal, thus showing excellent performance as an envelope detector. To further test the performance of the envelope detector, as before, the amplitude of the carrier wave is plotted for various frequencies in Figure 5.21.

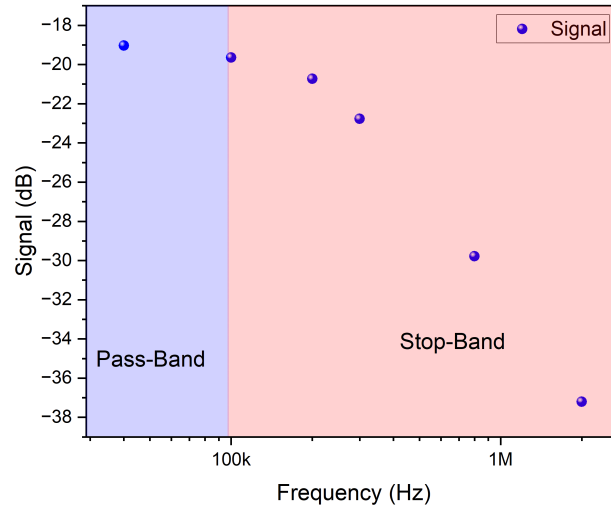


**Figure 5.19:** Frequency response analysis of the sine wave of data shown in Figure 5.18 is done to determine the pass-band and stop band of the filter.



**Figure 5.20:** Demonstration of low pass filtering operation. A 4 kHz signal is amplitude modulated with various carrier frequencies, as reported in the graph. Increasing the carrier frequency results in the suppression of the amplitude of the carrier signal.

An exponential suppression of the high-frequency signal is observed, thus demonstrating the excellent performance of the Sb device as an envelope detector.



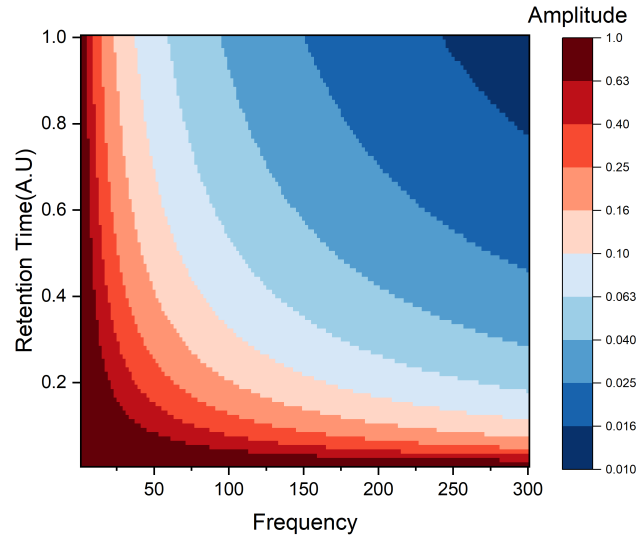
**Figure 5.21:** Frequency response analysis of the sine wave of data shown in Figure 5.20 is done to determine the pass-band and stop band of the filter

### 5.5.5 Achieving tuneable-low Filter

So far, the Sb device has been shown to demonstrate a low pass filter of fixed bandwidth. The frequency response depends on the leaky parameter of the filter. A ‘leaky’ filter would result in a higher cutoff frequency. For the Sb device, the degree of leakiness is dependent on the retention time of Sb. As discussed previously, the volatility of Sb can be tuned by varying the probing power.

Therefore to evaluate the effect of tuning the retention time (volatility) of the Sb device on the input sine wave signal, the response function is simulated for various frequencies. Frequency responses for various retention times are plotted in Figure 5.22.

Using fast Fourier transformation, the amplitude of the output response is determined. Increasing the retention time results in an increase in the low pass cutoff frequency. Thus Sb device can be used as a tunable low-pass filter. This kind of tuneable low-pass filter can also be used for the encryption of signals.



**Figure 5.22:** Simulation results showing the effect of retention time of Sb on the frequency cutoff for low pass filtering.

## 5.6 Critical Analysis

In this section, a critical analysis of antimony-based photonic memory is presented. Following are the shortcomings and scope of improvement of the work presented here :

- Implementation of a single-element phase change memory is exciting. The use of single-element-based memory cells should improve the ease of scalability of memory cells.
- The memory devices presented here have a retention time of memory states limited to a few seconds. Thus making use of demonstrated antimony-based memory cells for neuromorphic computing is quite challenging. Confinement of antimony to below 3 nm and in 3 dimensions could be a potential way of improving such retention times.
- In terms of switching speed, even with the use of femtosecond pulses, sub-nanosecond switching speed with antimony-based memory is not achievable.



There is only marginal improvement in switching speeds as compared to other PCM, such as AIST.

- In terms of reset speeds, about 100 pulses are used to achieve crystallisation. This also hampers the overall device performance in terms of speed.

## 5.7 Chapter Summary

This chapter has demonstrated that nanoscale thin films of antimony act as programmable elements on integrated waveguides. The memory device can be switched with a single femtosecond pulse with a switching speed corresponding to 500 MHz. Further, Sb can be programmed beyond two states and can be used to achieve up to 7 different discernible states, setting the stage for future improvements that could lead to higher-bit accuracy programming. Using single-element ultra-thin films allows for potential thickness scaling while enabling long-term cyclability because of the lack of phase separations induced by repeated phase transitions. This opens up many computing applications where high cyclability and ultra-fast speeds are required, such as efficient vector-matrix multiplications [15, 70, 148] and accumulative processing [59, 162–164]. One such application of the Sb device has been demonstrated to work a low pass filter for envelope detection. Further, the tuneable volatile property of Sb can be exploited to achieve a tuneable bandwidth low pass filter.

# 6

## Scaling up Matrix Multiplication Network: Reduced Rank

### Contents

---

|            |                                                       |            |
|------------|-------------------------------------------------------|------------|
| <b>6.1</b> | <b>Background . . . . .</b>                           | <b>93</b>  |
| <b>6.2</b> | <b>Modelling, Design and Experiments . . . . .</b>    | <b>94</b>  |
| 6.2.1      | Photonic-Mathematical Model . . . . .                 | 95         |
| 6.2.2      | Architecture Design . . . . .                         | 96         |
| 6.2.3      | Cross-Bar array design . . . . .                      | 98         |
| 6.2.4      | Device Fabrication . . . . .                          | 100        |
| 6.2.5      | Setting Weights . . . . .                             | 100        |
| 6.2.6      | Setting Baseline . . . . .                            | 102        |
| 6.2.7      | Experiment Details . . . . .                          | 103        |
| <b>6.3</b> | <b>Application Demonstration . . . . .</b>            | <b>104</b> |
| 6.3.1      | Image Filtering Task . . . . .                        | 104        |
| 6.3.2      | Classification Task . . . . .                         | 107        |
| <b>6.4</b> | <b>Scaling Up . . . . .</b>                           | <b>109</b> |
| 6.4.1      | Photonic Network Insertion Loss . . . . .             | 109        |
| 6.4.2      | System Level Energy Requirements . . . . .            | 110        |
| 6.4.3      | Foundry Assisted Chip . . . . .                       | 112        |
| <b>6.5</b> | <b>Performance on Large Models . . . . .</b>          | <b>119</b> |
| 6.5.1      | MNIST Performance . . . . .                           | 120        |
| 6.5.2      | CIFAR-10 Performance . . . . .                        | 121        |
| 6.5.3      | Pretraining Network with reduced parameters . . . . . | 122        |
| <b>6.6</b> | <b>Critical Analysis . . . . .</b>                    | <b>123</b> |
| <b>6.7</b> | <b>Chapter Summary . . . . .</b>                      | <b>124</b> |

---

This chapter discusses the development of a reduced-rank photonic architecture to improve the scalability of photonic networks in terms of footprint and programming energy. The data presented in this chapter has been peer-reviewed and published in [165]. The details for accessing the scripts and raw data used in this chapter are available in Appendix A.

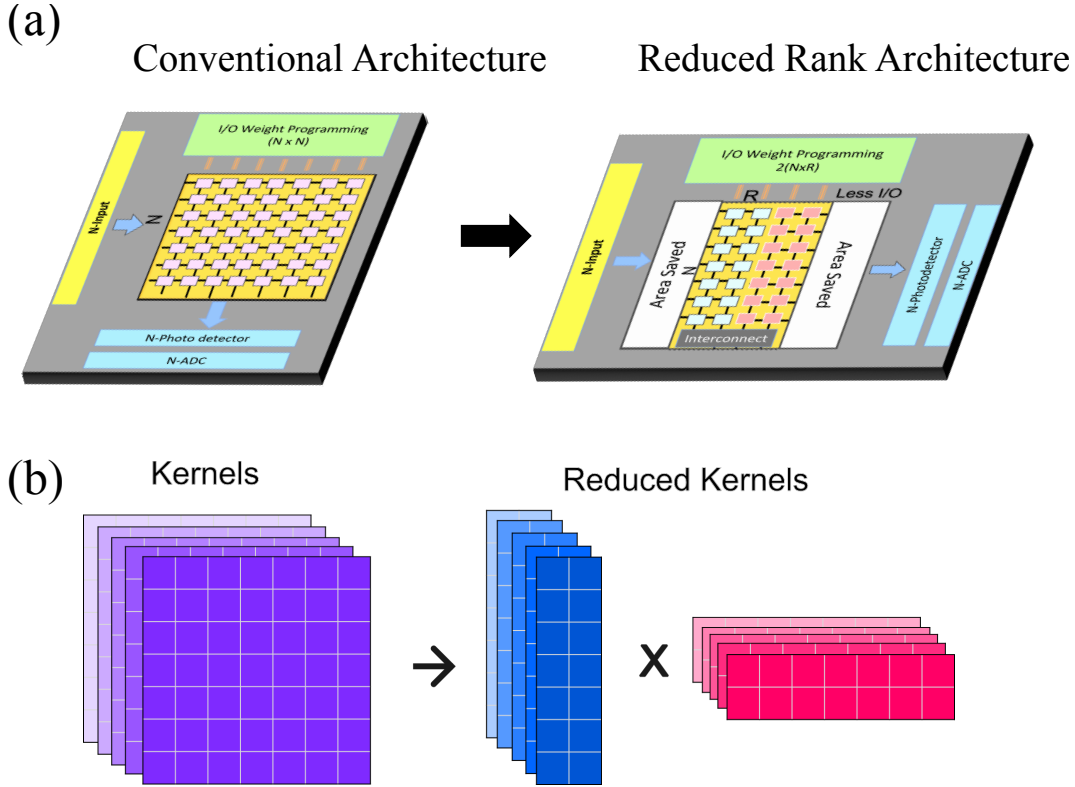
## 6.1 Background

Deep convolutional neural networks (DCNNs) have been extensively applied in complex tasks, including speech recognition [166], object classification and image filtering [167–170]. However, such neural networks are computationally intensive and take up the majority of memory and power consumption. This is particularly problematic in traditional von Neumann computing architectures, which are limited by the shuttling efficiency between the memory and the processing units. To circumvent this limitation, various in-memory computing hardware designs have been proposed to improve data movement [171–173]. With Moore’s law slowing down, photonic in-memory computational cores have been demonstrated with high data throughput and superior energy efficiency to their electronic counterparts [13, 14, 57, 62, 70, 76, 174, 175]. However, current photonic implementations face technological challenges related to the scalability of kernel sizes which are required to implement modern DCNN models.

Over the past few decades, various powerful DCNN models, such as ResNet50 and ResNet18, with tens of millions of trainable parameters, have been proposed. These large DCNN models require large-scale MVM operations. However, such DCNNs are usually over-parameterised [176], and therefore the redundancy can be reduced by rank reduction operations such as matrix factorisation. While such rank reduction operations have already been shown to achieve highly efficient DCNN tasks in software [177–179], hardware implementations of such rank reduction methods are, however power hungry with huge footprints [68, 180, 181]. Such implementations have been demonstrated using the traditional MZI mesh networks [182, 183]. However, such MZI mesh designs use large phase shifters. Such phase shifters allow a wide range of applications but at the cost of a larger footprint. Furthermore, they also frequently require an active energy supply increasing the overall energy budget.

## 6.2 Modelling, Design and Experiments

The reduced rank architecture is introduced in Figure 6.1(a). Conventional MVM photonic accelerators (size,  $N \times N$ ) consist of an  $N$ -input vector, which is processed through the computation core consisting of  $N^2$  weight elements. Each weight element is programmed to a corresponding weight, controlled using an  $N^2$  Input/Output (I/O) control system. This increases both the footprint and energy budget of the system.



**Figure 6.1:** Reduced rank architecture (a) Schematic of a photonic chip computing comparing the benefits of reduced rank architecture over conventional crossbar array architecture in terms of chip area and reduced I/O. (b) Kernels are factorised using rank reduction into two smaller matrices with reduced parameters.

The kernel matrix can be factorised and represented as a product of two reduced-rank matrices, as demonstrated in Figure 6.1 (b), resulting in an overall decrease in kernel parameters. This reduced rank architecture leverages the benefits of the factorisation process to gain benefits in terms of footprint and reduced number of

I/O required to control weight parameters and thus overall lower energy budget and higher computation density.

### 6.2.1 Photonic-Mathematical Model

For the photonic implementation of the reduced rank operation, a photonic-mathematical model is developed. To carry out matrix-vector multiplication of a weight matrix  $\mathbf{W}_{M \times N}$  and input intensity vector  $\mathbf{I}$ . The output vector  $\mathbf{O}$  can be thus written following :

$$\mathbf{O} = \mathbf{W}_{M \times N} \cdot \mathbf{I} \quad (6.1)$$

Where the output  $\mathbf{O}_i$  can be written as :

$$O_i = \sum_{k=1}^n w_{ik} \cdot I_k \quad (6.2)$$

The matrix  $\mathbf{W}_{M \times N}$  can be further factorised into two matrices  $\mathbf{U}_{M \times R}$  and  $\mathbf{V}_{R \times N}$ , and can be written as :

$$\mathbf{W}_{M \times N} = \mathbf{U}_{M \times R} \cdot \mathbf{V}_{R \times N} \quad (6.3)$$

Such that each element  $w_{ij}$  can be written as :

$$w_{ij} = \sum_{p=1}^r u_{ip} \cdot v_{pj} \quad (6.4)$$

From Equation 6.2 and 6.4

$$O_i = \sum_{k=1}^n \left( \sum_{p=1}^r u_{ip} \cdot v_{pk} \right) \cdot I_k \quad (6.5)$$

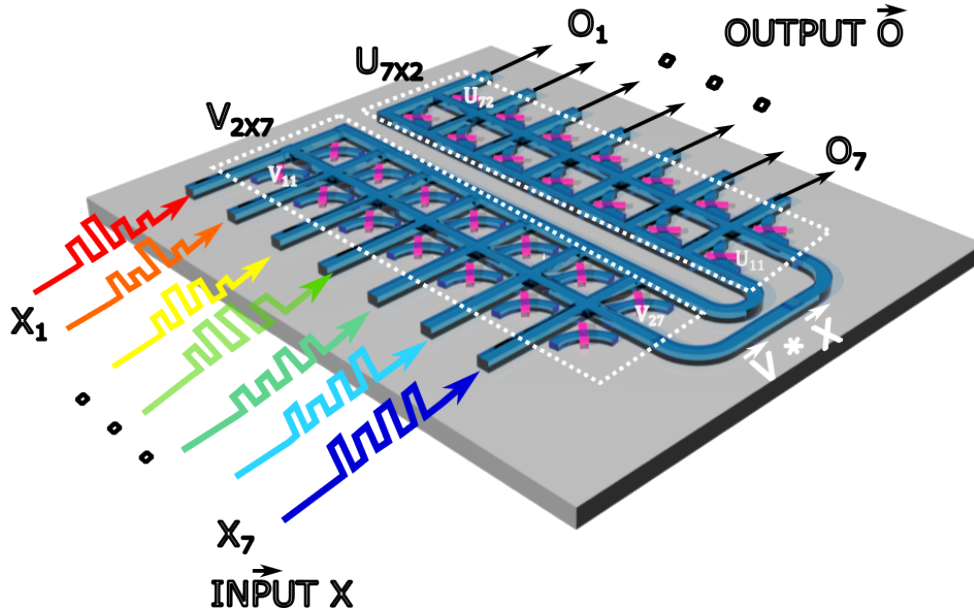
Rearranging the equation 6.5 gives:

$$O_i = \sum_{p=1}^r u_{ip} \cdot \left( \sum_{k=1}^n v_{pk} \cdot I_k \right) \quad (6.6)$$

Thus the term  $(\sum_{k=1}^n v_{pk} \cdot I_k)$  represents the intermediate output  $O_p^T$  of matrix-vector multiplication of matrix  $\mathbf{V}$  and vector  $\mathbf{I}$ . And the final output is the second matrix-vector multiplication of  $\mathbf{U}$  and  $\mathbf{O}^T$ . Thus, a photonic network representing two cascaded matrices is designed as explained in the section 6.2.2.

### 6.2.2 Architecture Design

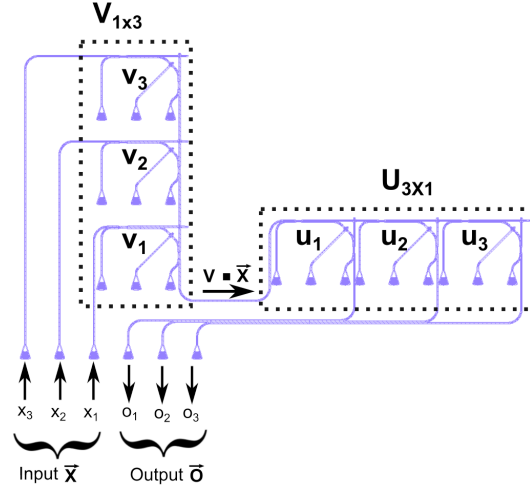
Figure 6.2 illustrates a schematic of the integrated photonic implementation of the reduced-rank photonic architecture. The schematic represents a computational core equivalent to a  $7 \times 7$  matrix but with a 40% reduction in size. The design is based on a crossbar array design [15] and has been explained in the next section. This architecture leverages the advantages of nonvolatile PCM photonic memory cells for high energy efficiency and high throughput operation.



**Figure 6.2:** Schematic of the photonic implementation of the proposed architecture. Input signals  $x_1$ - $x_7$ , encoded onto different wavelengths, are used as input vectors where two subsequent MVM results in the final output  $O$ .

Without loss of generality and for an easier understanding of the reduced rank design, an example of matrix-vector multiplication (MVM) of  $7 \times 7$  matrix is considered. A matrix  $\mathbf{W}_{7 \times 7}$  is factorised into two smaller matrices,  $\mathbf{U}_{7 \times 2}$  and  $\mathbf{V}_{2 \times 7}$ , of the reduced rank of ‘2’, resulting in a net 40% reduction in the size of the matrix. The MVM operation is mathematically equivalent to two inner products  $\mathbf{U} \cdot (\mathbf{V} \cdot \vec{X})$ .

To carry out the multiplication, the input vector  $\vec{X}$  ( $x_1$ - $x_7$ ) is pre-processed and normalised to  $[0,1]$ . The input vectors  $\vec{X}$  ( $x_1$ - $x_7$ ) are amplitude modulated and encoded on different wavelengths of the light  $\lambda_1 - \lambda_7$  (shown as different colours in the schematic) and used as the input signal to the photonic network. As discussed above, in this architecture, the input signal undergoes two matrix-vector multiplications. First, the input vector  $\vec{X}$  is multiplied by the matrix  $\mathbf{V}_{2 \times 7}$ . The resultant vector is then subsequently multiplied by the matrix  $\mathbf{U}_{7 \times 2}$ , therefore undergoing the second matrix-vector multiplication.



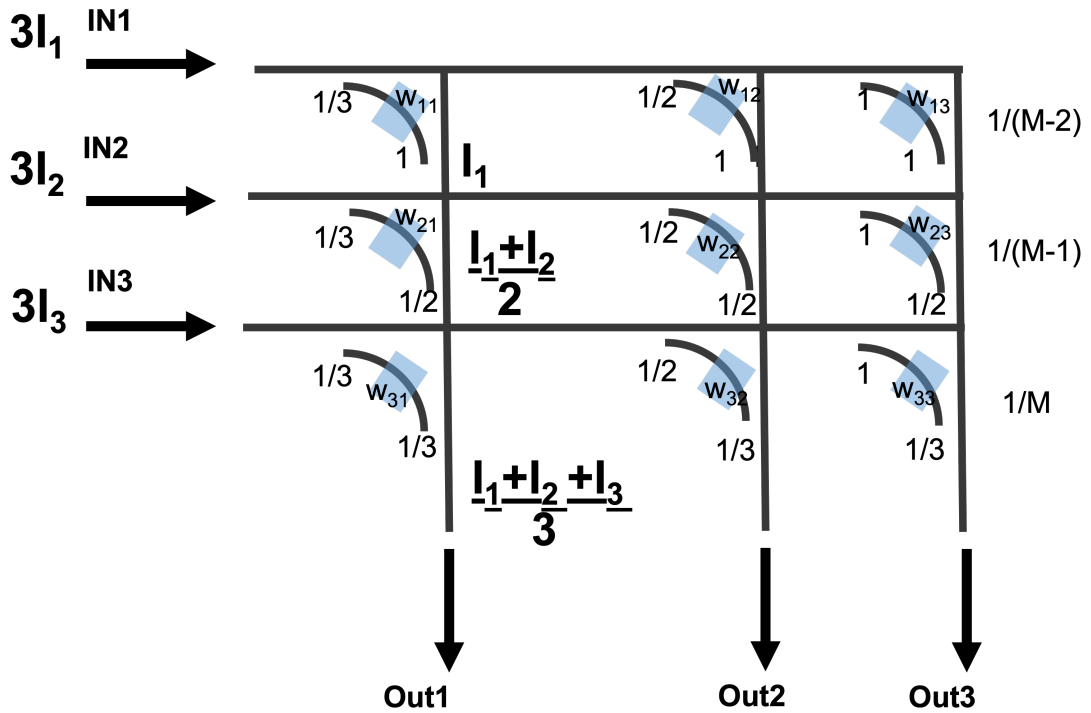
**Figure 6.3:** Schematic diagram of a  $3 \times 3$  equivalent device used in the experiments, showing signal flow.

For accurate computations, it is important to route light equally to the individual weight cells; therefore, for the first matrix  $\mathbf{V}$ , the input signal is split equally using directional couplers along the row of the matrix. Each signal then gets multiplied by the corresponding weight, stored in a photonic memory cell. Subsequently, the result of these multiplications is accumulated (added) onto a common bus waveguide (along the column) using directional couplers, which completes the first matrix-vector multiplication  $\mathbf{V} \cdot \vec{X}$ . This is then the input vector for the second multiplication, and the second output is computed in the same manner. Therefore,

the proposed architecture performs two multiply-accumulate operations, one after the other, in a single shot, which yields the complete MVM product  $\vec{O}$ .

A reduced rank processor equivalent to a  $3 \times 3$  matrix but with a  $33 \times$  reduction in training parameters is demonstrated in this thesis. Figure 6.3 illustrates the signal flow schematic of the actual device used in the subsequent experiments.

### 6.2.3 Cross-Bar array design



**Figure 6.4:** Cross-bar array schematic showing the distribution of optical power using a directional coupler for a  $3 \times 3$  network.

The crossbar consists of directional couplers to split input optical signals, as shown in the schematic in Figure 6.4. As mentioned above, for accurate computation of the matrix sum, it is important to distribute light equally through every weighting element and to have the equal contribution of each PCM weighting element onto the summation bus waveguide line.

Two sets of directional couplers are used in the crossbar design.  $DC_{IN}$  is used to split input light along the column and  $DC_{OUT}$  to add light from each row to a

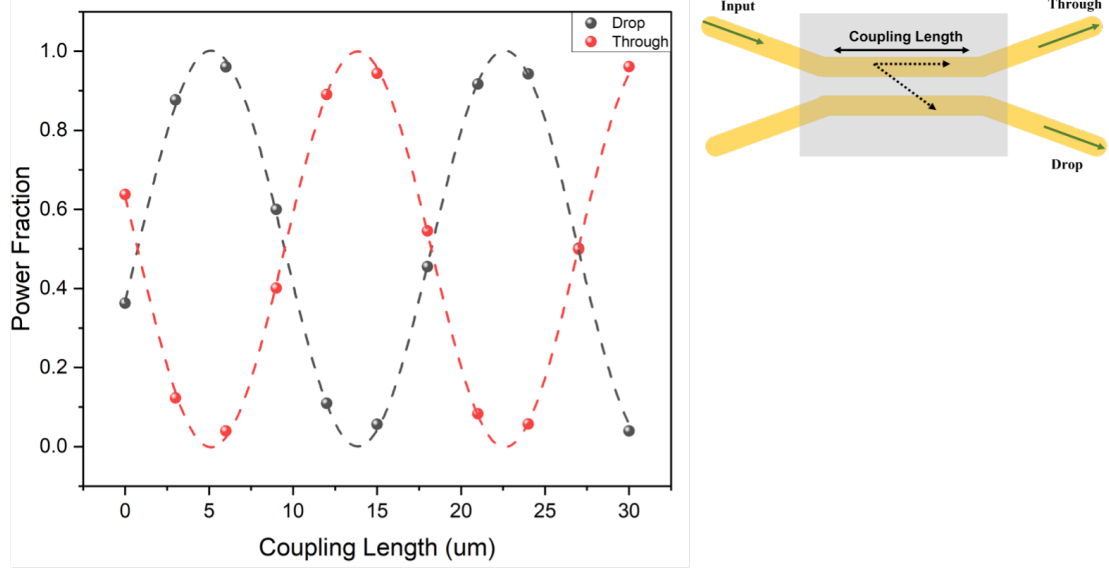


common bus. For a matrix of size  $n \times n$ , and  $i$  row and  $j$  column ( $i, j = 1 : n$ ), the ideal directional coupler ratio is given using the following relation :

$$Ratio_{ideal} = \begin{cases} \frac{1}{n+1-j} & DC_{IN}, \\ \frac{1}{i} & DC_{OUT}, \end{cases}$$

### Directional coupler calibration

The power splitting ratio of a directional coupler is dependent on the coupling lengths. Thus it is important to determine the relation between the coupling lengths and the coupling ratio experimentally. To do this, a directional coupler calibration is performed to obtain the power distribution fraction by varying the length of a directional coupler.

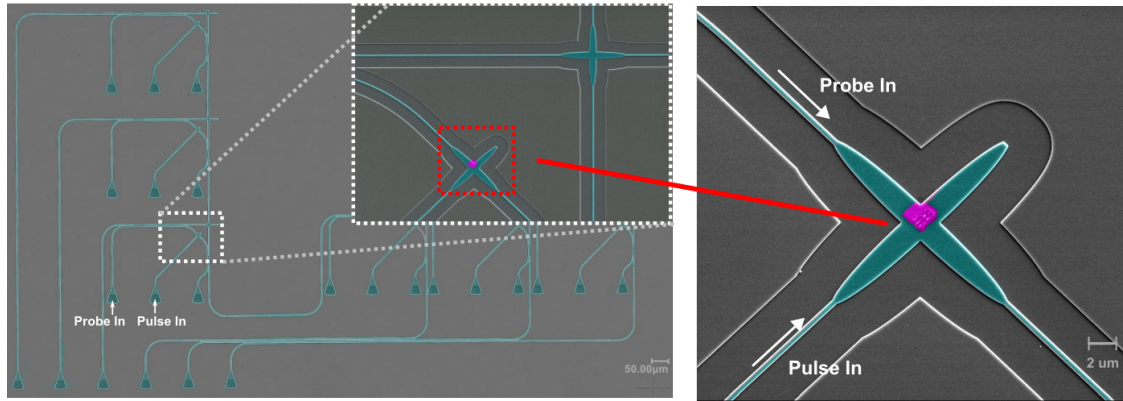


**Figure 6.5:** The power transfer of a directional coupler from input to cross port on varying the coupling length of an SOI directional coupler.

The power transfer function is fitted to a sine-squared function as shown in Figure 6.5. Thus the coupling length corresponding to the desired ideal coupling ratio can be back-calculated from the fitted curve.

### 6.2.4 Device Fabrication

The details of the device fabrication have been described in 4.1. The fabrication of the photonic network is on SOI wafers, with a 220 nm silicon device layer and a 3  $\mu\text{m}$  buried oxide. The waveguides are patterned using EBL and a positive photoresist (AR-P 6200). The patterned waveguide is then partially dry-etched (etch depth 120 nm). AIST is deposited using RF sputtering after patterning, followed by a lift-off process.



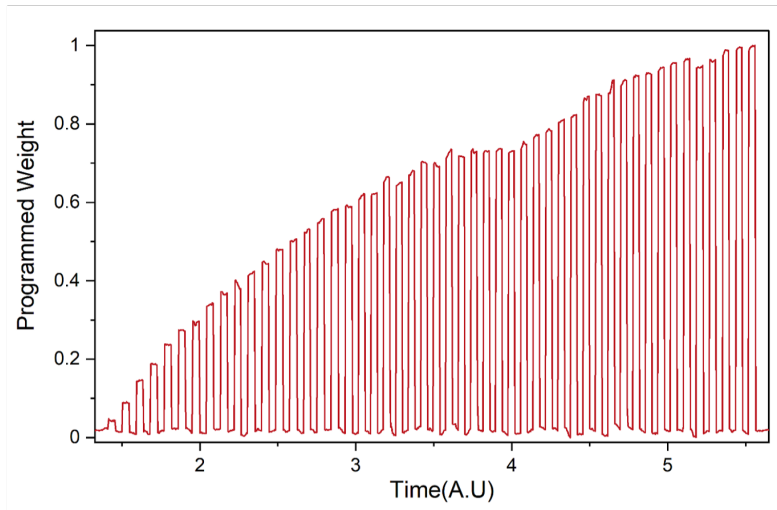
**Figure 6.6:** False coloured optical image of the fabricated device equivalent to  $3 \times 3$  matrix but with a  $33 \times$  reduction in size. (Inset) False-coloured SEM image of a unit cell showing waveguide interconnects and ‘active’ PCM memory. Zoomed in false coloured SEM image showing  $1 \mu\text{m} \times 1 \mu\text{m}$  PCM (AIST) deposited on the waveguide.

Figure 6.6 illustrates a false-coloured optical image of the fabricated device. The inset is a false-coloured SEM image of a single photonic memory cell using PCMs; this consists of a waveguide crossing to route signals across other memory cells and the  $1 \mu\text{m} \times 1 \mu\text{m}$  area of an in-memory active weight cell which in this case consist of the PCM AIST.

### 6.2.5 Setting Weights

To store weights, the PCM memory element-based memory cells are used as described by Rios et al. [7]. Using the single shot programming methods by Li et al. [8], individual weight matrices are set. PCMs have different refractive indexes depending on their phase. For conventional PCMs such as AIST and GST, the crystalline phase

is absorptive, and the amorphous phase is more transparent [184]. Therefore, by varying the fraction of the crystalline to the amorphous phase, different intermediate transmission states can be achieved. Pre-determined weights can be programmed within the matrix by sending a high-energy optical ‘Pulse’ (Figure 6.6). Individual memory cells can be encoded to better than 5-bit accuracy optically using 15 ns optical pulses and varying switching pulse amplitude, as shown in Figure 6.7.



**Figure 6.7:** More than 5-bit data can be stored in PCM memory by changing the pulse energy amplitude.

To set any weight in the range of  $[0, 1]$ , the crystalline state (lowest memory level) is mapped to a ‘0’, and the fully amorphous state (highest memory level) is mapped to a ‘1’, and therefore an intermediate amorphous state will correspond to a number between 0 and 1. Similarly to a set negative number, say in the range  $[-1, 1]$ , ‘-1’ can be mapped to the crystalline state (lowest memory level) and ‘+1’ can be mapped to the amorphous state (highest memory level); therefore all the numbers between -1 and +1 can be mapped to intermediate memory levels by varying the baseline. Details regarding setting the baseline are discussed in the following section.

### 6.2.6 Setting Baseline

As discussed in the previous section, photonic memory is dependent on the change in transmission of the photonic cell as the PCM is switched. Therefore it is important to set a baseline to normalise the change in the transmission and map it to the weight in the weight matrix. The initial transmission level is defined as the baseline transmission  $T_b$ , say 1 mW. After switching, the final transmission changes to  $T_f$ , say 1.4 mW. The percent change in transmission  $\Delta T = \frac{T_f - T_b}{T_b} \times 100$ , that is  $\frac{1.4-1}{1} \times 100 = 40\%$ . Now a 40% change in the transmission can be mapped to represent a '1'. Thus by measuring the baseline and the final transmission state, one can map the weight for all the weight elements in a matrix, which will be equivalent to  $\frac{T_f - T_b}{T_b \times \Delta T_{change}}$ , where  $\Delta T_{change}$  is the pre-decided maximum transmission contrast (40% in the previous example).

For a conventional matrix, such baseline values can be easily measured. For example, to measure the transmission of any weight element  $w_{ij}$ , the input  $i$  is set on and the transmission from photodetector  $j$  is read. However, reading such weights in the reduced-rank architecture is not as straightforward. Let  $T_b^W$  be the baseline transmission matrix and  $T_f^W$  be the final transmission matrix. Thus weight matrix achieved  $W_{achieved}$  is computed using the following element-wise operation :

$$W_{achieved} = \left( \frac{T_f^W - T_b^W}{T_b^W \times \Delta T_{change}} \right) \quad (6.7)$$

Since in the reduced rank approach, subsequent matrices are multiplied such that  $W = U \cdot V$ , thus each matrix baseline needs to be set separately.

$$W_{achieved} = \left( \frac{T_f^U - T_b^U}{T_b^U \times \Delta T_{change}^U} \right) \cdot \left( \frac{T_f^V - T_b^V}{T_b^V \times \Delta T_{change}^V} \right) \quad (6.8)$$

Which can be further rearranged to

$$W_{achieved} = \frac{1}{\Delta T_{change}^V \Delta T_{change}^U} \left( \frac{T_f^U T_f^V - T_b^U T_f^V - T_b^V T_f^U + T_b^U T_b^V}{T_b^U T_b^V} \right) \quad (6.9)$$

Therefore three baseline values need to be measured.  $T_b^U T_b^V$  represents when both U and V are not switched;  $T_f^U T_b^V$  represents the U matrix switched and V not switched,  $T_b^U T_f^V$  when the U matrix is not switched, and the V matrix is switched, and  $T_f^U T_f^V$  represents the actual computation step.

## 6.2.7 Experiment Details

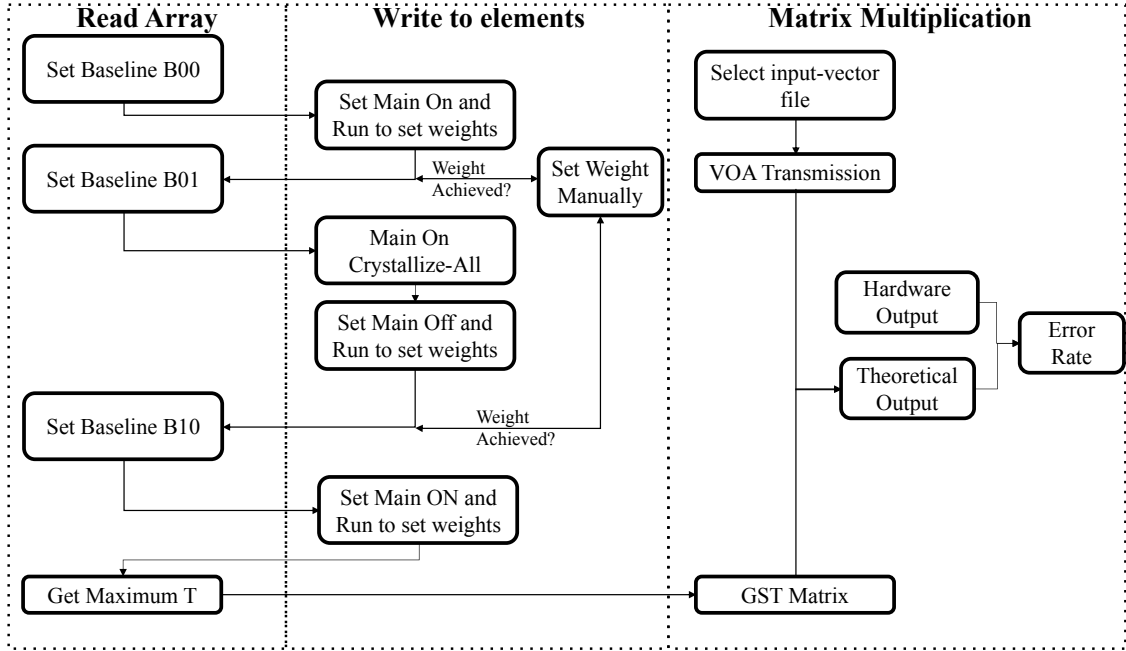
### Experimental Setup

The details of the experimental setup and the instruments used for programming and carrying out the computation have been described in Chapter 4.2.3. The experimental setup consists of two optical lines: one to set the weight (programming line) and another to carry out computation (inference line). Through the programming line, high-energy optical pulses are sent to set the desired weight of the matrix. The inference line consists of a supercontinuum light source. The inference light is split into input vectors and modulated by the input vector using a VOA. The modulated signals are sent through the chip to carry out the computation. The output of the computation is recorded on a computer using a DAQ.

### Experimental Flow

To carry out the photonic measurements, an automated LabVIEW program is developed to automatically set the desired weights and then further carry out the matrix-vector multiplication operation. The flowchart describing the experimental workflow is shown in Figure 6.8.

The LabVIEW program consists of three main windows, namely **Read Array** block to read the weight elements, **Write to Elements** block to set the weight, and **Matrix Multiplication** block to carry out the multiplication. As explained above, it is necessary to measure three baseline values. Thus using the Read array block, the inference line is activated, and the three baseline values are measured and stored by using different variations of  $\{100\}$  as the input vector.



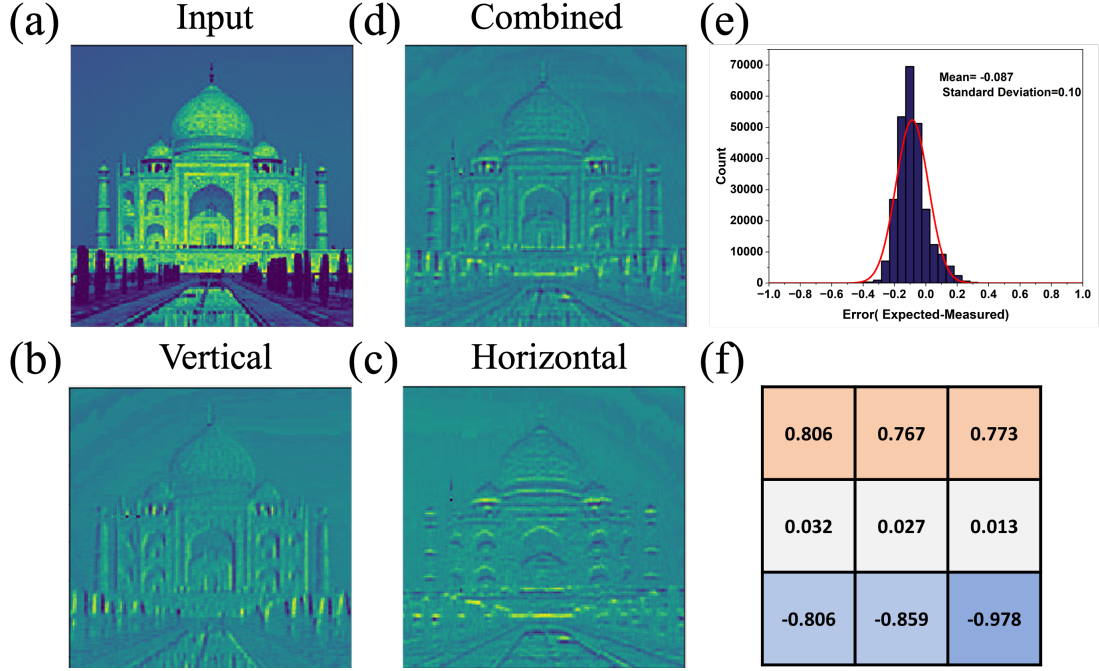
**Figure 6.8:** Schematic of the experimental flow used in Matrix-vector multiplication experiments.

The **Write to Elements** block activates the programming line since each matrix **U** and **V** have to be set separately. A toggle switch **Main** is used to select which matrix to switch. The desired weight of the matrix is set, and the program reaches the desired value automatically using iterative programming with a 1% error in achieving the targeted value. In case the system cannot reach the switching parameters automatically, the weights can be programmed manually. The final matrix is determined once all the baselines are measured and the matrix weights are set. The final achieved matrix is used to calculate the theoretical output and compared with the experimentally obtained values to calculate the error rate in **Matrix Multiplication** block.

## 6.3 Application Demonstration

### 6.3.1 Image Filtering Task

To demonstrate the utility of the architecture, the performance of this photonic hardware is tested to carry out image-filtering tasks using an edge detection filter.



**Figure 6.9:** Image Filtering (a) Input image of ‘Taj Mahal’ is used by pre-processing as an input vector for edge detection experiments. (b-c) Successful edge detection of vertical and horizontal filters, respectively, using the reduced rank photonic architecture. (d) Combining the edge detection results from (b) and (c) showing highlighted edges. (e) Computing error of expected vs measured value showing a Gaussian distribution with a standard deviation of 0.1. (f) Experimentally achieved weights using PCM for edge filter.

A vertical and horizontal edge detection filter is used, described mathematically as the matrices:  $\begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix}$ ,  $\begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix}$  respectively, on an image of ‘Taj Mahal’, Figure 6.9 (a-c). The input image is pre-processed and flattened as input vectors  $\vec{X}$  ( $x_1 - x_3$ ) to be fed through the photonic network with the filter as the kernel  $\mathbf{W} = \mathbf{U} \cdot \mathbf{V}$ . The vertical edge detection filter matrix is factorised into two smaller matrices of reduced rank ‘1’  $\mathbf{U} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$  and  $\mathbf{V} = \begin{bmatrix} 1 & 0 & -1 \end{bmatrix}$ , similarly the horizontal edge filter is reduced to  $\mathbf{U} = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$  and  $\mathbf{V} = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}$ .

Experimentally, input vectors are normalised to  $[0, 1]$  and optically encoded using variable optical attenuators (VOA). The optically encoded information is then multiplied by individual weights of the kernel and accumulated at the photodetector, therefore performing a MAC operation. The weights of the kernel between  $[-1, 1]$

are achieved by mapping the lowest memory level as ‘-1’ and the highest memory level as ‘+1’; therefore, any number in the range is represented by the corresponding intermediate memory state. A 10% switching contrast of the PCM memory cell between the lowest and highest level is used. Therefore, intermediate transmission represents numbers in the range of  $[-1,1]$ . The experimentally achieved edge filter used using PCM elements is shown in Figure 6.9(f).

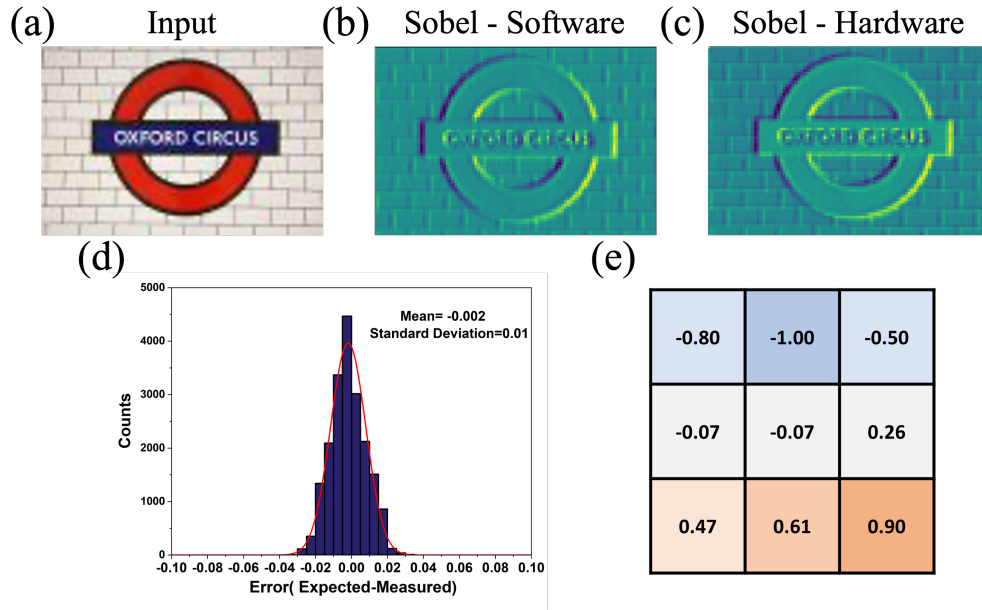
The image filtering results from horizontal and vertical edge filters are combined, as shown in Figure 6.9 (d), resulting in edge detection. To assess the computational accuracy, the error rate of the expected and measured output of over 40,000 MVM operations is plotted in Figure 6.9 (e). For computation, a normal distribution of the error is obtained with a mean error rate of -0.087 and a standard deviation of 0.10.

To show the ubiquity of the reduced rank network, another edge detection filter, a vertical Sobel filter, is implemented. This filter can be described mathematically as follows:  $\begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix}$  and is used to detect the left edge on a logo of an image of a London tube station signage (Figure 6.10(a)). The Sobel filter weight matrix  $\mathbf{W}$  is reduced into two low-rank matrices  $\mathbf{U}$  and  $\mathbf{V}$ , where  $\mathbf{U} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$  and  $\mathbf{V} = \begin{bmatrix} 1 & 0 & -1 \end{bmatrix}$ .

To improve the error rate, the PCM memory cells are switched to about 40% contrast to represent kernel weights between  $[-2,2]$ . The experimentally achieved Sobel filter used using PCM elements is shown in Figure 6.10(e).

Next, the performance of the proposed photonic hardware to highlight the left edge is compared to edge detection on software Figure 6.10(a-c), therefore showing an excellent performance of the photonic hardware. As done previously, the computation accuracy is assessed by calculating the error of measured and expected MVM results of over 18,000 operations. As calculated in Fig 6.10(d), a normal error distribution is obtained. It is worth noting increasing the switching contrast



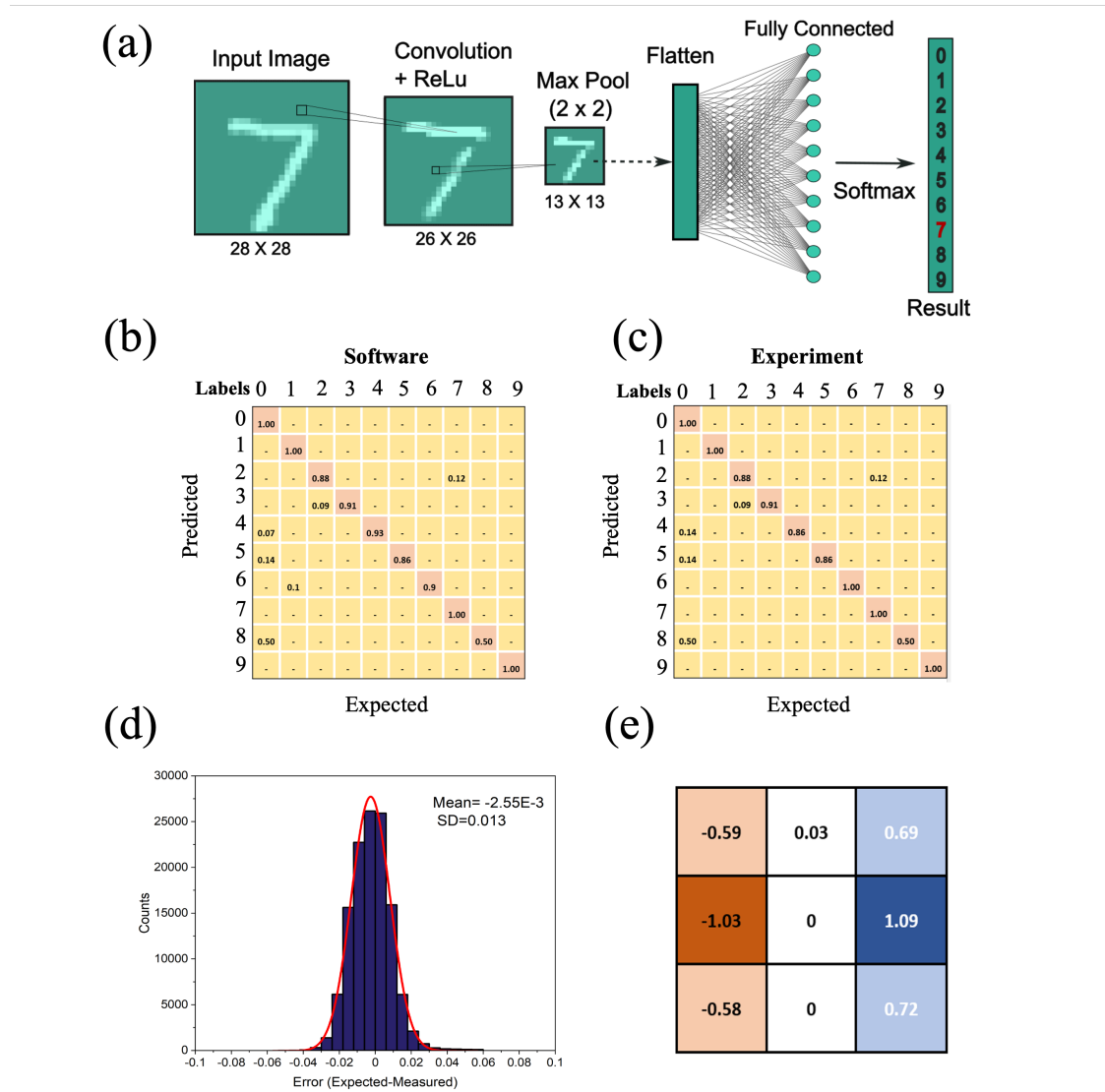


**Figure 6.10:** (a) Sobel Filter is used for edge detection on an image of a London tube station. (b) edge detection results on software are compared with one processed on photonic hardware (c), showing excellent conformality of the two. (d) Computing error of expected vs measured value showing a normal distribution with a standard error deviation of 0.01. (e) Experimentally achieved weights using PCM for the Sobel filter

from 10% to 40% helps in the reduction in the error by one order of magnitude. A mean error of  $2 \times 10^{-3}$  with a standard deviation of 0.01 is obtained. Therefore, for subsequent experiments, a switching contrast of 40% is used.

### 6.3.2 Classification Task

A DCNN model to demonstrate an image classification task on the MNIST digit dataset is constructed. Figure 6.11(a) shows use the schematic of the DCNN model using the low-rank matrix photonic core, which is a reprogrammable hardware-based convolution layer. The model is trained on software to achieve 96% accuracy. The photonic hardware is then programmed with the appropriate weights (by setting the state of the PCM photonic memories) and is then used as the convolution layer. The photonic kernel is set to the appropriate state as described previously. The experimentally achieved matrix using PCM memory cells is shown in Figure 6.11(e).



**Figure 6.11:** Image classification using convolution neural network (a) Schematic of the convolutional neural network used to classify MNIST handwriting digits. The reduced rank photonic architecture with the kernel size equivalent to  $3 \times 3$ , but with  $33 \times$  reduction, is used. (b-c) Confusion matrix of image classification task for the model on software and experimentally obtained on the photonic hardware, showing an excellent agreement. (d) Error estimation of over 120K MVM by calculating expected vs measured results. A normal distribution fit reveals a standard error deviation of 0.013. (e) Experimentally achieved matrix for convolution experiments.

A set of 100 images from the MNIST dataset are used for inferencing and calculating model accuracy. As before, the images are flattened as input vectors normalised to  $[0,1]$  for multiplication. Figure 6.11(b) and (c) show the confusion matrices for the image classification task obtained on software and experimentally

on the photonic hardware. For the classification task, an accuracy of 94% is achieved experimentally, which matches closely to the accuracy obtained on software, suggesting that this approach works very well. In Figure 6.11(d), a histogram of computation error for over 120,000 MVM operations is computed. A Gaussian distribution fit reveals a mean error of  $-2.55 \times 10^{-3}$  with a standard deviation of 0.013, approaching software-based accuracy.

## 6.4 Scaling Up

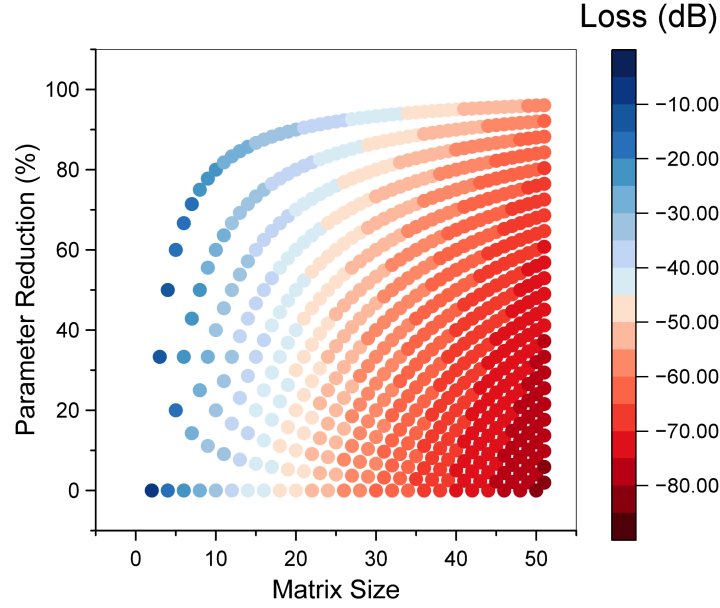
One of the limiting factors for scaling up photonic crossbar array architecture is the photonic insertion losses. For scaling up, it is necessary to carry out system-level energy calculations as well. The next challenge is related to the integration of active elements, such as photodetectors and weighting elements on-chip and interfacing with electronic circuits.

Therefore to address the issues related to scaling up of a photonic network, in this section, an analytical model is developed to calculate photonic network losses and overall system energy requirements. Next, a foundry-based photonic chip is designed, and the associated electronic circuitry is developed.

### 6.4.1 Photonic Network Insertion Loss

The sources of photonic network loss in the system are the insertion loss of the waveguide crossings, directional couplers, PCM absorption and that of the network due to the incoherent addition of light. For an  $M \times N$  weight matrix, which has been rank reduced and broken into two cascaded rank-reduced matrices,  $M \times R$  and  $R \times N$ , the final transmission at the output can be written as the following equation:

$$10 \times \log\left(\frac{P_{1 \rightarrow R \rightarrow N}}{P_{in}}\right) = (2R + M + N)\alpha_{DC} + 2\alpha_{PCM} + (2R + M + N + 4)\alpha_{cros} - 10\log(N \times M \times R) \quad (6.10)$$



**Figure 6.12:** Effect of parameter reduction of matrix on the overall optical loss of the proposed photonic architecture. Parameter reduction results in lower losses in the system.

Where  $P_{1 \rightarrow R \rightarrow N}$  is the output power when light is input through 1<sup>st</sup> row and output at R<sup>th</sup> column of the first matrix and obtained at the N<sup>th</sup> output of the second matrix;  $P_{in}$  is the input laser power in the first row;  $\alpha_{DC}$  is the insertion loss of a directional coupler which is estimated to be 0.1 dB per coupler [15];  $\alpha_{PCM}$  is the absorption loss of the PCM which is estimated to be 2 dB;  $\alpha_{cros}$  is the insertion loss of waveguide crossing which is estimated to be 0.12 dB [15];  $10\log(N \times M \times R)$  is related to the power lost due to incoherent addition of light and fan out loss to ensure equal splitting of the light in the network using directional couplers. Here the propagation loss of the silicon waveguide has been ignored, which is estimated to be less than 0.5 dB/cm. Reducing the size of the matrix through rank reduction results in a smaller size array and hence lower optical losses, as shown in Figure 6.12.

#### 6.4.2 System Level Energy Requirements

So far, a small-scale  $3 \times 3$  crossbar array with 3 input-output channels has been demonstrated. Nevertheless, this system can be scaled up, and therefore the

**Table 6.1:** State of the art energy requirement of various photonic computing elements

| Component                                                 | Power or energy | Bandwidth | Ref        |
|-----------------------------------------------------------|-----------------|-----------|------------|
| Light source (DFB frequency comb)                         | 0.7W            | NA        | [185, 186] |
| Microring unit for MUX and DE-MUX                         | N.A             | N.A       | [187]      |
| Transmitter (modulator, DAC, CMOS driver, digital banded) | 0.685 pJ/bit    | 40 Gb/s   | [188]      |
| Photonic tensor core unit cell                            | N.A.            | N.A.      | This work  |
| Receiver (photodetector, TIA, buffer)                     | 0.17 pJ/bit     | 25 Gb/s   | [189]      |
| ADC                                                       | 13 mW           | 12 GH     | [190]      |

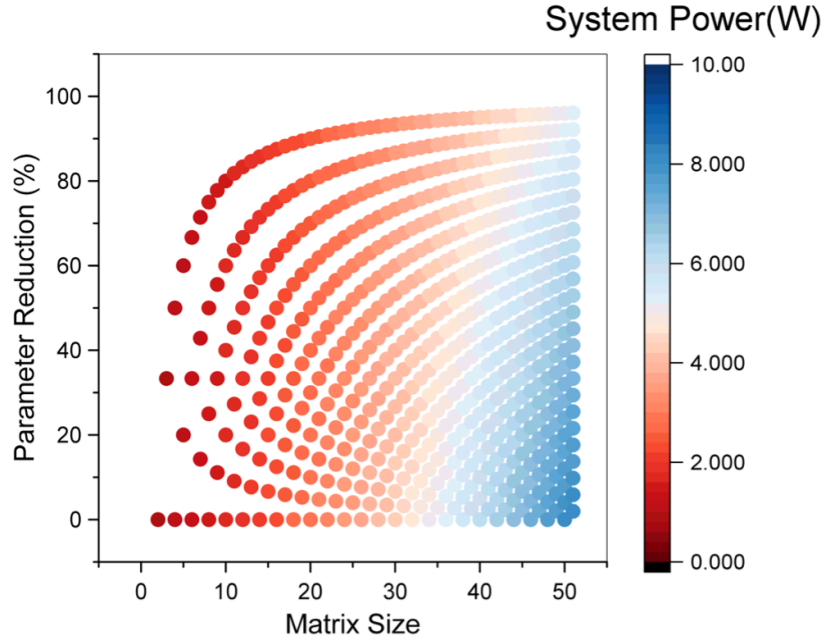
complete power requirements of the system, equivalent to the  $M \times N$  weight matrix, can be estimated. The sources of power consumption in a large-scale integrated network are the light source, the energy required for modulation of the input signal, the energy required to set and store the weight, the energy required at the receiver end and finally, the energy required by the ADC. Therefore, the total power is given by:

$$P_{lightsource} + M f_{DR} C h_{WDM} E_{Tx} + (MR + RN) E_{ST} + N f_{DR} C h_{WDM} E_{Rx} + N C h_{WDM} P_{ADC} \quad (6.11)$$

**Table 6.2:** The energy required to set and store the weights for various weighting methods

| Programming Element   | Energy/Power Requirement | Ref   |
|-----------------------|--------------------------|-------|
| Phase Change Material | 190pJ                    | [191] |
| Phase Shifters        | 10mW                     | [192] |
| EAM                   | 1.2mW / 12.8fJ/bit       | [88]  |
| Ring Modulators       | 5.2 mW                   | [14]  |

Where  $P_{lightsource}$  is the power consumption of the input light source,  $f_{DR}$  is the total data rate of each WDM channel,  $C h_{WDM}$  is the number of WDM channels,  $E_{Tx}$  is the energy required for encoding the signal onto wavelengths,  $E_{ST}$  is the energy required to set the weights and to keep the weights,  $E_{Rx}$  the energy required

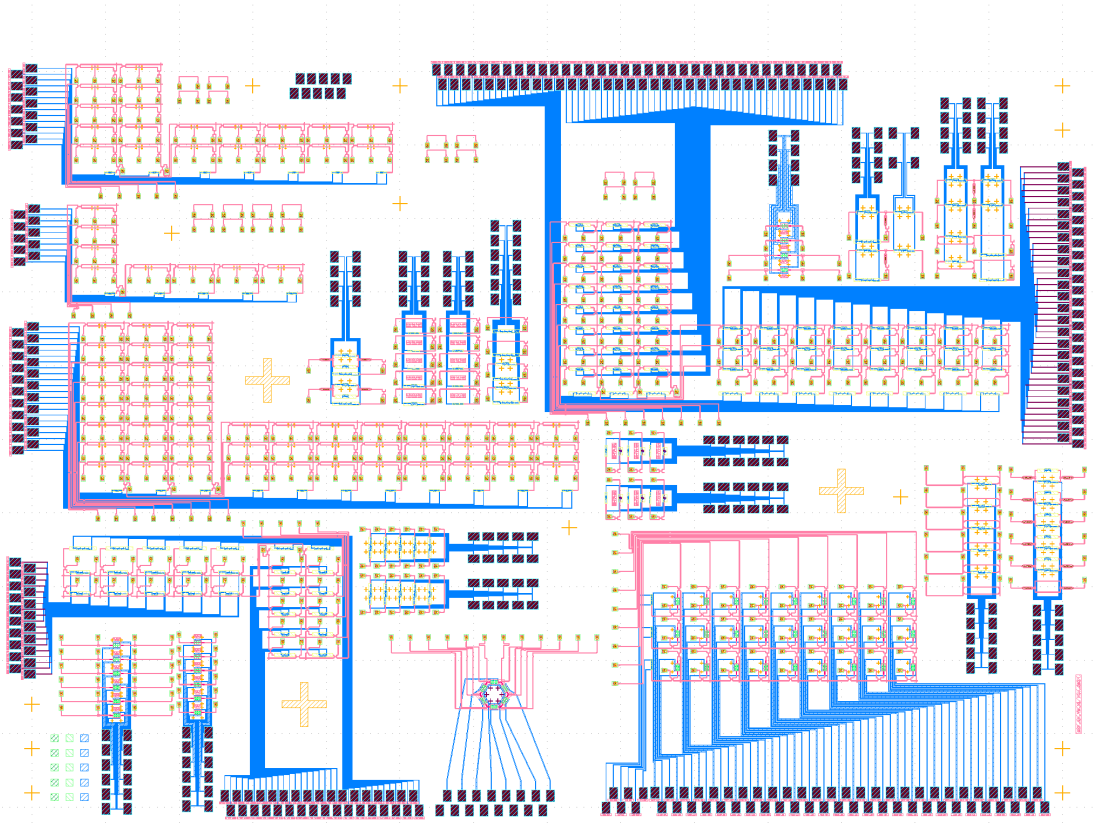


**Figure 6.13:** System level power requirements of a photonic computing network for different matrix sizes.

for optical transceiver circuit,  $P_{ADC}$  is the power requirements of ADC. The values are reported in Table 6.1 as follows. The energy required to set and store the weights for various weighting methods is reported in Table 6.2. Assuming a data rate of 10 GHz and 4 channels per DWDM, the energy requirements for the computation with a change in the size of the matrix and parameter reduction are computed in Figure 6.13.

### 6.4.3 Foundry Assisted Chip

So far, the reduced-rank architecture, which is fabricated in-house, has been demonstrated to show excellent computing performance. Due to the fabrication limitation, the matrix size has been limited to  $3 \times 3$  equivalent. Moreover, in the previous section, it has been analytically calculated that the reduced rank technique scales well in terms of both insertion losses and overall system-level energy. Next, the reduced rank design is further scaled up using IMEC foundry-assisted fabrication process.



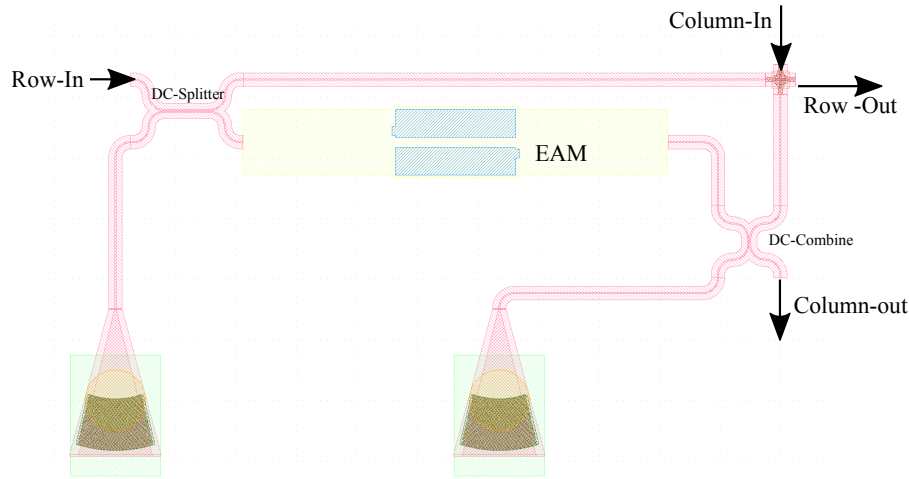
**Figure 6.14:** The layout of the foundry chip design consists of various test structures and large-scale matrices.

An integrated photonic chip with an integrated photodetector is designed. For scaling up the reduced rank photonic matrices, two designs are considered. The first design uses EAMs to act as weighting elements in the crossbar array. EAMs offer up to 40GHz of modulation. This can be used to achieve on-chip training. The second design consists of the PCMs as a weighting element, and the photonic design is similar to the design used in the experiments above.

Figure 6.14 shows the layout of the complete integrated photonic chip designed. The integrated photonic chip consists of various reduced-rank matrices with integrated photodetectors. The input signal is injected into the chip using grating couplers, and the computed results will be read through photodetectors. The photodetectors are then connected via metal wires to bond pads on the edge of the chip. The PIC will be wire bonded to a PCB to read the output signal.

Similarly, to control the EAMs, bond pads are used to supply analogue voltage to the EAMs.

### EAM Matrix Design



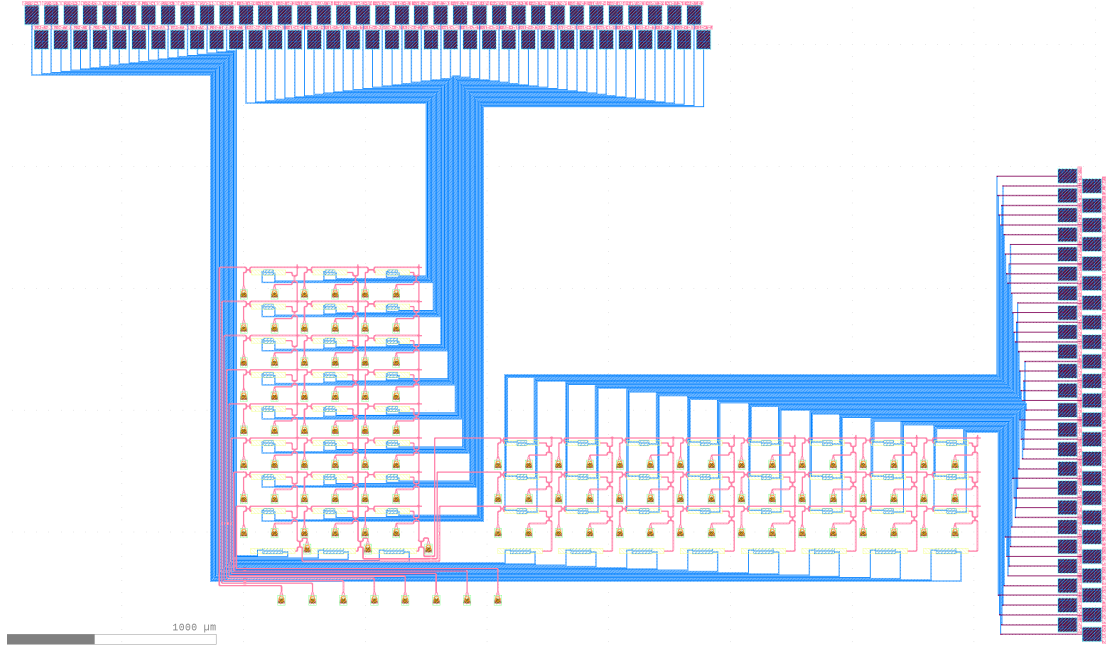
**Figure 6.15:** EAM unit cell consisting of EAM as weighting element and two directional couplers to fan-in and fan-out optical signal. The Row-In signal is coupled to the EAM using DC-Splitter. This signal gets amplitude modulated by the EAM, and the resultant modulated signal is added with Column-In to get the Column-out using DC-combiner.

The design of the EAM matrix cell is based on an EAM-unit cell arranged in a crossbar array. Each EAM acts as a weighting unit and consists of two directional couplers to fan in and fan out the input signal, 2 grating couplers to measure the weighting of EAMs and a waveguide crossing to route the signal along the row and column of the unit cell. The schematic of each EAM unit cell is shown in Figure 6.15.

Such unit cells are then arranged in a crossbar array structure which is further arranged as two cascaded reduced rank matrices, as discussed before. The design for the EAM matrix equivalent to  $8 \times 8$  matrix consisting of two cascaded  $8 \times 3$  and  $3 \times 8$  matrices is shown in Figure 6.16.

The 8 input signals are injected into the matrix using high-efficiency grating couplers. The light is then routed through the first matrix, where each input is weighted at each EAM-unit cell. The output of the first MVM is then routed





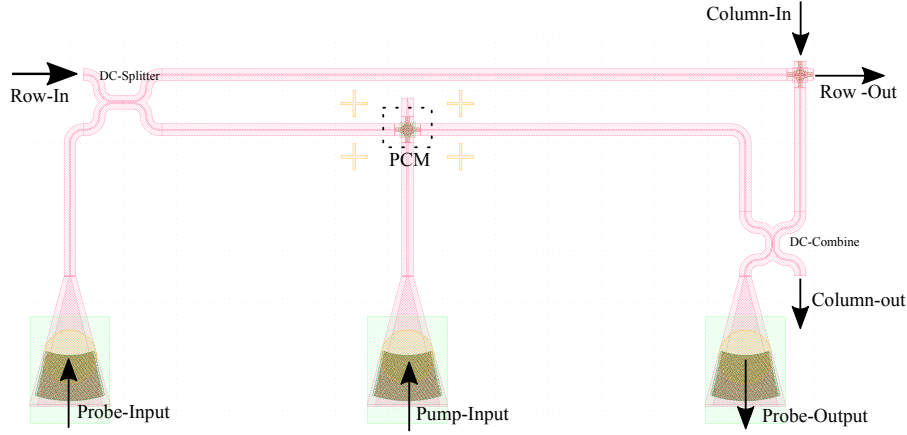
**Figure 6.16:** Design of an  $8 \times 8$  equivalent crossbar array with EAM as a weighting element with integrated photodetectors and bond pads to read and write signal.

through the second matrix, and the final output is collected at the integrated photodetector. The output product can be recorded as the photocurrent from the photodetectors, which can be accessed through the bond pads arranged at the edges of the chip.

**Table 6.3:** EAM Matrix Loss Estimation

| Source          | Number of Component | Loss         | 8X3X8   | 5X2X5   | 4X1X4   |
|-----------------|---------------------|--------------|---------|---------|---------|
| Grating Coupler | 1                   | 2.4 dB       | 2.4 dB  | 2.4 dB  | 2.4 dB  |
| Crossings       | $M+N+2R-4$          | 0.23 dB      | 4.14 dB | 2.3 dB  | 1.38 dB |
| EAM             | 2                   | 6 dB         | 12 dB   | 12 dB   | 12 dB   |
| DC              |                     | $0.9/MR^2N$  | 30 dB   | 20.5 dB | 12.5 dB |
|                 |                     | <b>Total</b> | 49 dB   | 37dB    | 29dB    |

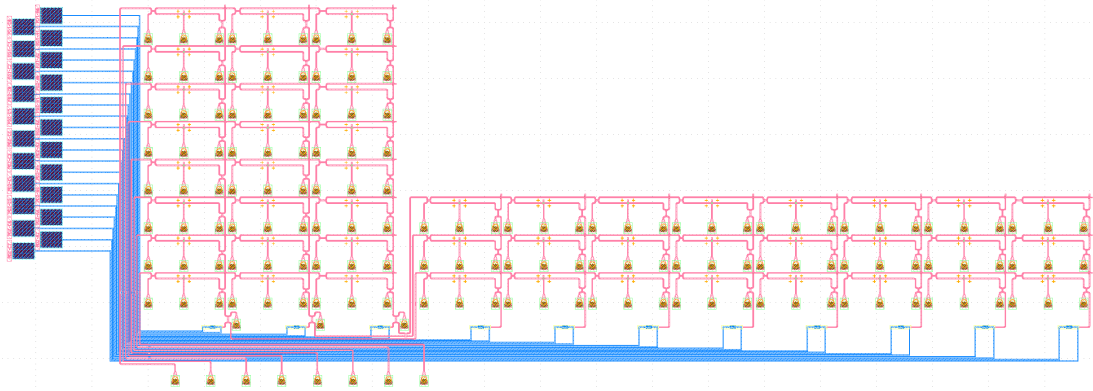
Using the network loss estimation described in Section 6.4.1, the highest network loss is calculated according to the insertion loss parameters provided by the foundry in Table 6.3 for different matrices.



**Figure 6.17:** PCM unit cell consisting of PCM as weighting element and two directional couplers to fan-in and fan-out optical signal. The Row-In signal is coupled to the PCM element using DC-Splitter. This signal gets amplitude modulated by the PCM memory, and the resultant modulated signal is added with Column-In to get the Column-out using DC-combiner. Using Pump-Input, the PCM can be switched, monitoring the weight using Probe-Input/Output couplers.

### PCM Matrix Design

The design of the PCM matrix cell is based on using PCM on a waveguide crossing. Similar to the EAM unit cell, PCM acts as a weighting element. The unit cell consists of directional couplers to fan-in and fan-out input signals. PCM is switched by sending high-energy pulses through the Pump-Input grating couplers, which monitor the weights using Probe-Input/Output grating couplers. The schematic of each PCM unit cell is shown in Figure 6.17.



**Figure 6.18:** Design of an  $8 \times 8$  equivalent crossbar array with PCM as a weighting element with integrated photodetectors and bond pads to read photodetector signals.

Such unit cells are then arranged in a crossbar array structure which is further arranged as two cascaded reduced rank matrices, as before. The design for the PCM matrix equivalent to  $8 \times 8$  matrix consisting of two cascaded  $8 \times 3$  and  $3 \times 8$  matrices is shown in Figure 6.18.

**Table 6.4:** PCM Matrix Loss Estimation

| Source          | Number of Component | Loss                       | 8X3X8  | 5X2X5   | 4X1X4   |
|-----------------|---------------------|----------------------------|--------|---------|---------|
| Grating Coupler | 1                   | 2.4 dB                     | 2.4 dB | 2.4 dB  | 2.4 dB  |
| Crossings       | $M+N+2R-2$          | 0.23 dB                    | 4.6 dB | 2.76 dB | 1.84 dB |
| PCM             | 2                   | $2 \text{ dB}/\mu\text{m}$ | 4 dB   | 4 dB    | 4 dB    |
| DC              |                     | $0.9/MR^2N$                | 30 dB  | 20.5 dB | 12.5 dB |
|                 |                     | <b>Total</b>               | 41 dB  | 29 dB   | 21 dB   |

As before, the network loss is estimated as described in Section 6.4.1. The highest network loss is calculated according to the insertion loss parameters provided by the foundry is calculated in Table 6.4 for different matrix sizes.

### Transimpedance Amplifier Selection

For the selection of the correct transimpedance amplifier (TIA), it is important to know the required bandwidth of the photodetectors and the maximum generated photocurrent to thus estimate the requirement for the gain of the TIA. SiGe photodetector parameters from the foundry are reported in Table 6.5.

**Table 6.5:** IMEC SiGe Photodetector characteristics

|              |         |
|--------------|---------|
| Responsivity | 0.8 A/W |
| Bandwidth    | 50GHz   |
| Dark Current | 40nA    |

Using these values, the photocurrents for various matrix sizes corresponding to a 5 mW single input for EAM and PCM matrix designs are reported in Table 6.6 and 6.7, respectively.

Using the photocurrent values, a maximum transimpedance gain of 100 K $\Omega$  is needed to read a minimum voltage of 5 mV. Thus a commercially available TIA

**Table 6.6:** EAM Matrix Photocurrent Estimation assuming 5mW input

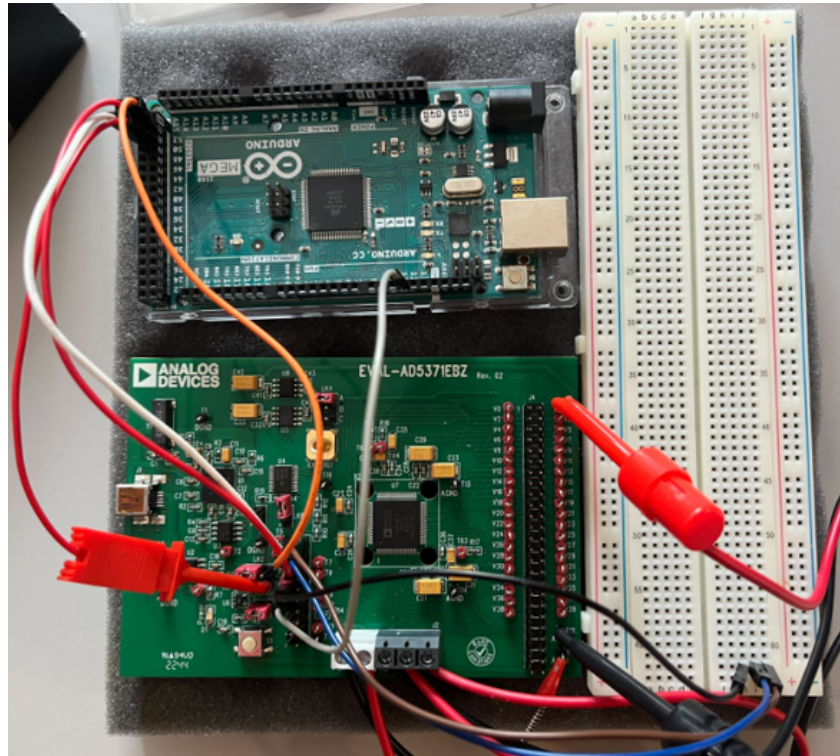
| Matrix Size | Max Loss (dB) | Power before PD(mA) | Photocurrent(nA) |
|-------------|---------------|---------------------|------------------|
| 8X3X8       | -49           | 6.29E-05            | 50               |
| 5X2X5       | -37           | 9.98E-04            | 798              |
| 4X1X4       | -29           | 6.29E-03            | 5036             |

**Table 6.7:** PCM Matrix Photocurrent Estimation assuming 5mW input

| Matrix Size | Max Loss (dB) | Power before PD(mA) | Photocurrent(nA) |
|-------------|---------------|---------------------|------------------|
| 8X3X8       | -41           | 3.97E-04            | 318              |
| 5X2X5       | -29           | 6.29E-03            | 5036             |
| 4X1X4       | -21           | 3.97E-02            | 31773            |

from Texas Instruments OPA3S2859 is chosen, which has a low input noise of  $2.2 \text{ nV}/\sqrt{\text{Hz}}$  and a bandwidth of up to 900 MHz.

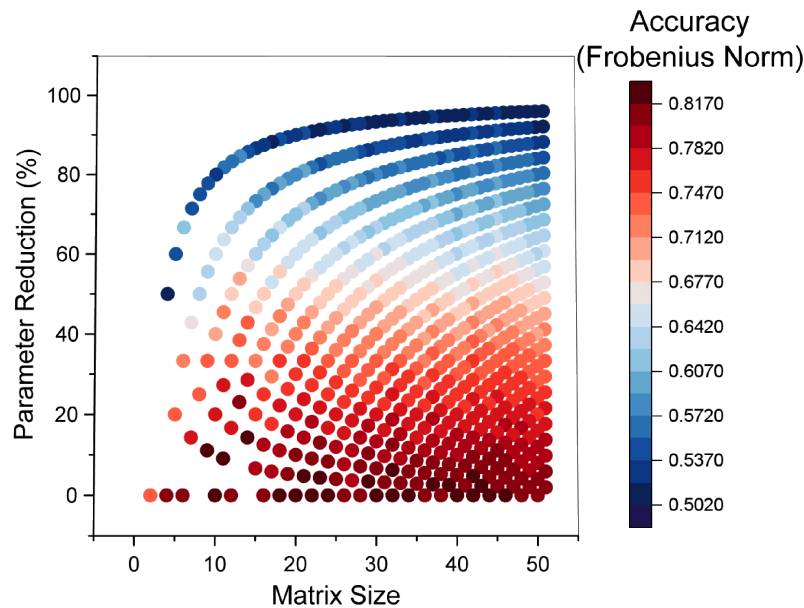
### Analog Output Controller

**Figure 6.19:** A 40 analogue DAC controller evaluation board is used to control the EAM. The evaluation board is controlled using an Arduino-based microcontroller.

EAM matrix design consists of 48 EAMs that need to be controlled using an

analogue output signal with over 9-bit precision. Thus a commercially available digital-to-analogue converter(DAC) solution from Analog devices AD5371EBZ is used. This device provides 40 independent analogue voltage output control. The evaluation board from Analog devices is controlled using an Arduino microcontroller using SPI communication protocol up to 50MHz speed. To set the desired voltage at any channel, a Python script has been developed to send the serial data to the DAC using Arduino. An image of the DAC board with Arduino is shown in Figure 6.19. The codes used to program the Arduino microcontroller and a Python script to get arbitrary voltage output can be found in Appendix B.

## 6.5 Performance on Large Models

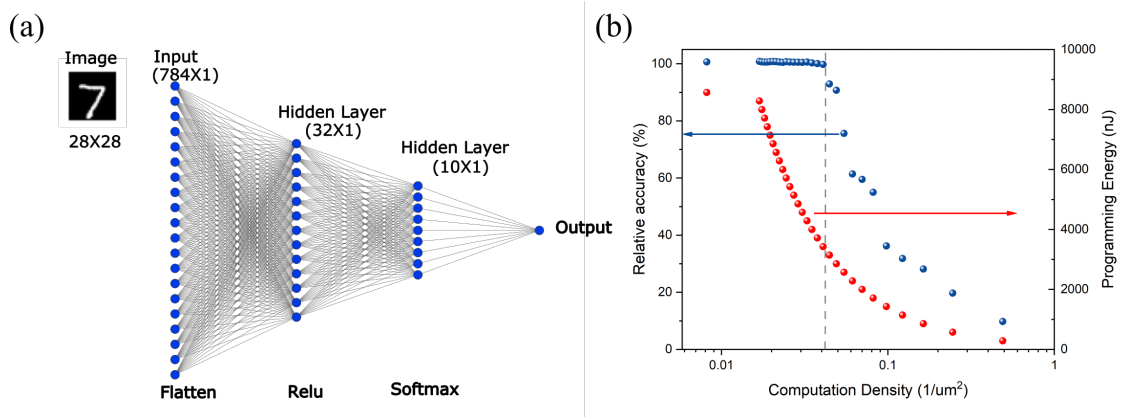


**Figure 6.20:** Effect of reducing the number of parameters on the accuracy of matrix factorisation for different sizes of the matrix. Here the accuracy is defined as the difference between the Frobenius norm of the original weight matrix and the reduced rank matrix. Increasing the reduction factor results in lower accuracy.

So far, it has been shown that the reduced rank technique works well for small matrices and shows excellent performance vis-à-vis software. In this section, the benefits of this technique for larger matrices are investigated.

As one would expect, the source of error in computation is due to the matrix factorisation process itself. Mathematically it is not feasible to have a reduced-rank matrix to represent the original weight matrix  $W_{m \times n}$ . As explained before, the weight matrix  $W_{m \times n}$  can be reduced to two matrices,  $U_{m \times r}$  and  $V_{r \times n}$ , where  $r$  is the targeted reduced rank of the matrix. The total number of parameters after rank reduction is  $mr + rn$ . Therefore to have advantages after rank reduction,  $r$  is chosen such that  $r < \frac{mn}{m+n}$ . The target rank  $r$  is varied, which is equivalent to reducing the number of parameters in the weight matrix. In Figure 6.20, the effect of parameter reduction (reducing the rank of the matrix) on the accuracy of representing the original matrix is simulated. As one would expect, reducing the number of parameters results in a reduction in accuracy.

### 6.5.1 MNIST Performance



**Figure 6.21:** (a) Schematic of a feed-forward model used for MNIST handwritten digit recognition used for simulation in (b). (b) Effect of parameter reduction on the relative accuracy and programming energy of the system. Parameter reduction results in over  $3\times$  gain in computation density due to a smaller footprint without any loss in accuracy for the classification task with additional benefits of  $2.5\times$  gain in terms of programming energy.

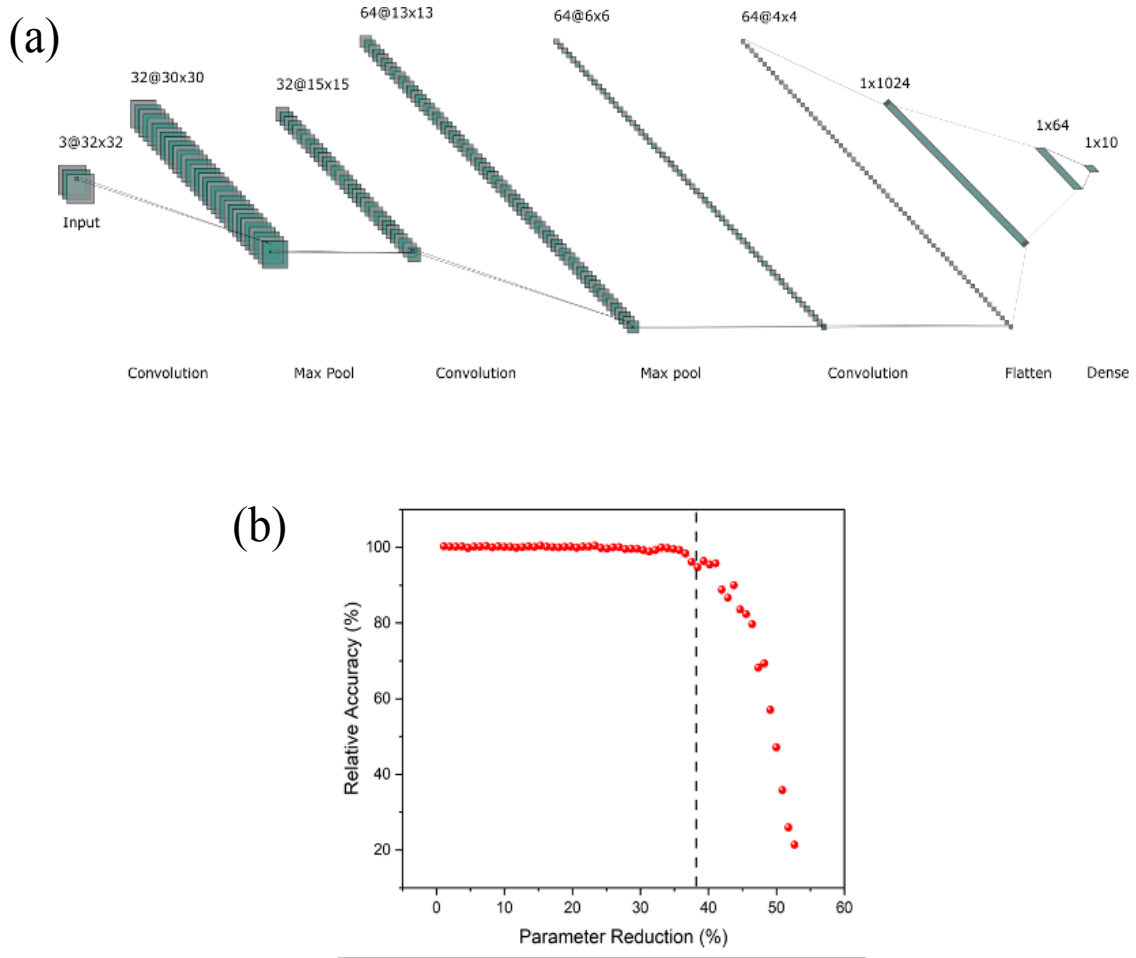
For practical applications of the reduced rank technique, one needs to optimise accuracy and programming energy reduction in the system. To find an optimum balance of parameter reduction, the effect of reducing the number of weight

parameters on the accuracy of the image classification task is calculated. To do that, a feed-forward DCNN with two hidden layers and over 25 thousand trainable parameters, as illustrated in Figure 6.21(a), is simulated. The model is trained for MNIST handwritten digit classification with over 91% accuracy. The weight matrix,  $\mathbf{W}$ , is then extracted from the trained model. Then the weight matrix is factorised using the rank reduction technique. To account for negative numbers in the trained matrix, a semi-negative matrix factorisation algorithm [193] is used. The relative accuracy of the classification task from a rank-reduced matrix network is plotted in Figure 6.21 (b) as a function of parameter reduction. It is observed that with a 60% reduction in the parameters, there is no significant loss in computation accuracy; this can be translated in  $>3\times$  improvement in computation density.

As far as programming energies are concerned: assuming the weight matrix based on PCM photonic memories as used in this thesis, a reduction in training parameters results in over  $2.5\times$  improvement in programming energy; thus, this architecture scales very favourably in terms of energy and computation density.

### 6.5.2 CIFAR-10 Performance

To check the performance of more computationally intensive tasks than MNIST, a classification task of the CIFAR-10 dataset is performed using a deep neural network. A DCNN model is built as illustrated in Figure 6.22a. The model has over 112K trainable parameters. Like before, the model is trained to over 70% accuracy and uses the rank reduction technique to reduce the parameters. Since the fully connected layers consist of the largest number of training parameters, the rank reduction operation is performed on this layer. It is observed that the rank reduction technique works really well. Even with a 40% reduction in the total number of parameters in the model (a much higher reduction in number for the fully connected layer), there is no drop in relative accuracy, as shown in 6.22b.

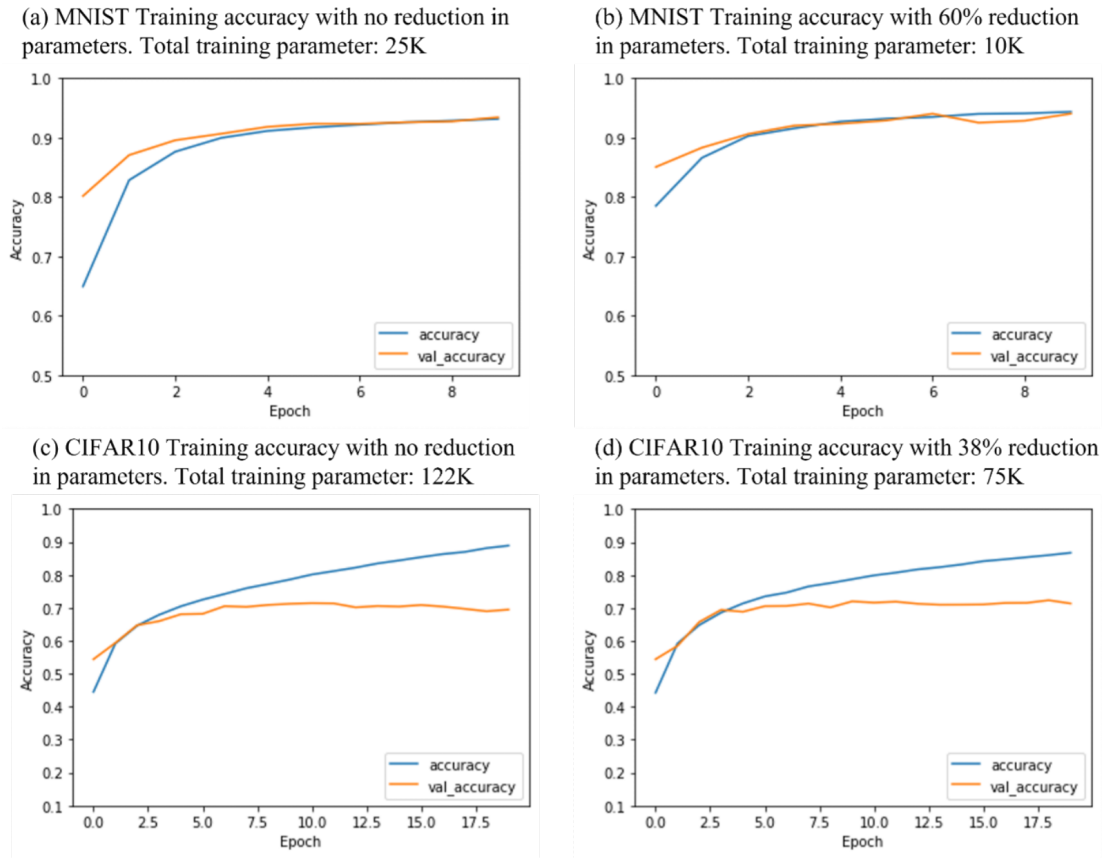


**Figure 6.22:** (a) The deep learning model for the CIFAR-10 classification task. (b) Change in relative accuracy with decreasing the number of training parameters using the rank reduction method

### 6.5.3 Pretraining Network with reduced parameters

So far, to calculate the performance of the MNIST and CIFAR-10 classification task, a network is trained first, and then the weight matrix is reduced. The rank reduction process itself is a computationally intensive task. This extra computation penalty can be overcome if a weight matrix with a reduced number of parameters is used during the training step itself. Therefore to test this, the MNIST and CIFAR-10 models used previously are simulated and trained from scratch. These models are trained with a lower number of parameters on a computer, with 60% less for MNIST and about 38% less for CIFAR-10. It is observed that the model





**Figure 6.23:** Effect on the model accuracy for a) MNIST task with no reduction in training parameters, b) MNIST task with 60% pre-reduction in training parameters, c) CIFAR-10 task with no reduction in training parameters, d) CIFAR-10 task with 38% pre-reduction in training parameters.

performance is as well and, at times, better if trained from scratch with a lower number of parameters, Figure 6.23.

## 6.6 Critical Analysis

In this section, a critical analysis of scaling up a photonic network using a reduced rank approach is presented. Following are the shortcomings and scope of improvement of the work presented here :

- The reduced rank approach improves the scalability in terms of overall footprint and memory cell switching energies. However, the use of a cascaded matrix introduces an extra network in the system due to the incoherent addition

of light. In photonic networks, the scalability is limited by the electronic component  $\sqrt{Hz}$  noise. Thus making the scalability of such reduced-rank architecture more difficult than a conventional crossbar design.

- Another shortcoming of such a reduced-rank approach is that the rank-reduction process is itself computationally costly, and not every matrix can be completely represented by reduced rank matrices. This severely limits the application of such an approach to a few limited applications.
- The architecture presented here uses a large number of I/O to program the weights of the PCM cell using all-optical switching techniques. As the network scales up, it becomes impractical to have such a large I/O for switching. Thus alternate approaches based on optoelectrical design, such as integrated photonic heaters, must be explored.

## 6.7 Chapter Summary

In this chapter, an MVM architecture for applications in neural networks to reduce kernel sizes without incurring significant losses in accuracy has been demonstrated. This architecture is experimentally implemented on a silicon photonic platform using phase change reconfigurable weights. By using the proposed architecture for image filtering and convolution tasks, a photonic hardware equivalent to  $3 \times 3$  kernel but with a reduction of 33% is demonstrated. In spite of the matrix size reduction and reduced complexity of the system, high accuracy in the computation is detected even due to the smaller number of required weights and resulting variability. Then a foundry-assisted photonic architecture with integrated photodetectors and EAMs for training the matrix is designed. Further, the scalability of such a network is analytically assessed in terms of optical losses and system energy. A use case of the proposed architecture is tested for MNIST, and CIFAR 10 classification task is assessed. It is estimated that for up to a 60% reduction in the kernel size, there is

no significant loss of information for the MNIST classification task. The proposed architecture thus not only improves scalability, complexity and programming energy but is also adaptable and can be readily implemented in diverse neuromorphic computing architectures.

# 7

## Scaling up Matrix Multiplication Network: Multimode System

### Contents

---

|            |                                                    |            |
|------------|----------------------------------------------------|------------|
| <b>7.1</b> | <b>Background . . . . .</b>                        | <b>126</b> |
| <b>7.2</b> | <b>Device Design . . . . .</b>                     | <b>128</b> |
| <b>7.3</b> | <b>System Level Simulations . . . . .</b>          | <b>132</b> |
| <b>7.4</b> | <b>Fabrication Tolerance Comparision . . . . .</b> | <b>134</b> |
| <b>7.5</b> | <b>Critical Analysis . . . . .</b>                 | <b>136</b> |
| <b>7.6</b> | <b>Chapter Summary . . . . .</b>                   | <b>137</b> |

---

In the previous chapter, it was discussed that the photonic network losses due to the incoherent addition of light are a major contributor to the overall system loss. This hinders the scalability of a photonic computing system. In this chapter, using simulations, a multimode system is presented for the efficient addition of light signals. The results presented in this chapter have been peer-reviewed and published in conference proceedings at the IEEE conference [194].

### 7.1 Background

A crossbar array architecture has multiple nodes comprising a number of splitters and couplers. Unlike electronics, combining optical power is not as straightforward as adding two voltages or currents. A coupler based on a single-mode Y-splitter (3 dB coupler) results in a 50% loss of optical power at each junction. This limits the number of nodes in a photonic network and therefore limits the scalability of photonic networks. In the case of a 3 dB splitter, the optical power of the first

input node drops exponentially  $(\frac{1}{2})^n$  as it passes through each node. The issue of unequal power loss at each node can be resolved using directional couplers and distributing the coupling losses equally amongst the coupling nodes, as shown in the cross-bar design in the previous chapter and in works based on a cross-bar array architecture [15, 70]. However, the overall optical loss of the system remains the same using both approaches.

Some of the major drawbacks of using directional couplers are their sensitivity to small fabrication errors, high sensitivity to wavelength [195] and a large footprint. Other passive approaches of combing optical signals through wavelength division multiplexing (WDM) using ring resonators [196] have similar issues to a directional coupler. They also have high sensitivity to fabrication errors and limited bandwidth. Table 7.1 summarises various coupling strategies with their advantages (in black) and disadvantages (in red).

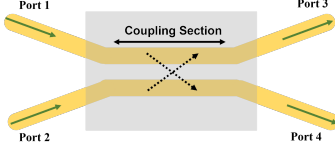
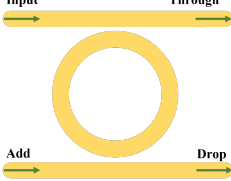
As it will be shown in the subsequent section, even for foundry-fabricated devices, there is a mismatch in the power coupling ratio of different optical signals due to fabrication variations across the wafer. Thus a Y-coupler design would yield a reliable and robust large-scale optical design with a smaller footprint.

A multimode coupler has been proposed for reservoir computing applications with a high number of nodes, as in neuromorphic computing hardware [197]. The coupling efficiency of such Y-couplers can be improved by considering asymmetric coupler design. Work on optical fibres has also demonstrated the advantages of using asymmetric multimode fibres for efficient signal coupling in a linear data bus [198].

In this chapter, an integrated photonic coupler design that uses principles of both multimode and asymmetric couplers is demonstrated. Using modal decomposition simulations, the scattering matrix (S-matrix) for various input-output modes is obtained. Using the S-matrix, the overall coupling efficiency of an asymmetric multimode Y-coupler is estimated. The coupling efficiency is then compared with a

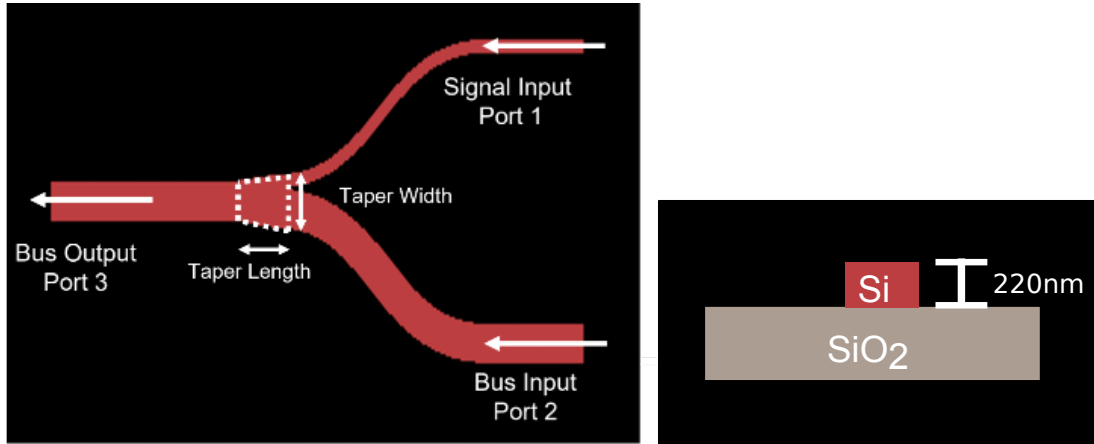
directional coupler and a single-mode Y-coupler to demonstrate a superior broadband performance with lower sensitivity to fabrication variations.

**Table 7.1:** Summary of advantages and disadvantages of various power coupling strategies for incoherent addition of light using single-mode Y-coupler, directional couplers and ring resonators.

| Y-coupler                                                                         | Directional Coupler                                                               | Ring Resonator                                                                      |
|-----------------------------------------------------------------------------------|-----------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|
|  |  |  |
| 3dB loss per coupler                                                              | Lower loss than Y-couplers                                                        | Efficient power coupling                                                            |
| Broadband                                                                         | Wavelength dependent coupling                                                     | Limited bandwidth                                                                   |
| No critical dimension                                                             | Critical fabrication dimensions                                                   | Wavelength dependent coupling                                                       |
| Small footprint                                                                   | Large footprint                                                                   | Large footprint                                                                     |
| High fabrication tolerance                                                        | High susceptibility to fabrication variations                                     | High susceptibility to fabrication variations - require additional active tuning.   |

## 7.2 Device Design

The design of the coupler is based on a standard 220 nm silicon-on-insulator (SOI) platform working at 1550 nm wavelength. As shown in the device schematic in Figure 7.1, the coupler has three ports: a single mode Port 1 (Input Signal), multimode Port 2 (Input Bus) and multimode Port 3 (Output Bus). At each node  $N_i$ , a Transverse Electric ( $TE_0$ ) signal with the wavelength  $\lambda_i$  is injected at port 1. A unique wavelength at each node ensures no interference between input signals. The interference issue can also be avoided by the use of an incoherent light source as an input signal. The Input Signal is coupled to the Input Bus from port 2 to obtain an Output at port 3. The input power from a single mode port 1 is converted to higher order modes at the output of port 3.



**Figure 7.1:** Schematic diagram of the proposed asymmetric coupler with three ports. A single mode signal is coupled into the Bus input (port 2), resulting in a multimode output at Bus Output (Port 3).

To achieve high coupling efficiency, the device is designed to have a high transmission for all higher-order modes. To achieve high average transmission of all incident modes, the taper length of the coupler has to be optimised to balance two conflicting effects [197]. Increasing the taper length decreases the scattering losses of the supermodes in the taper section. However, increasing the taper length decreases the mode conversion efficiency of higher-order supermodes in the taper to lower-order modes at the output.

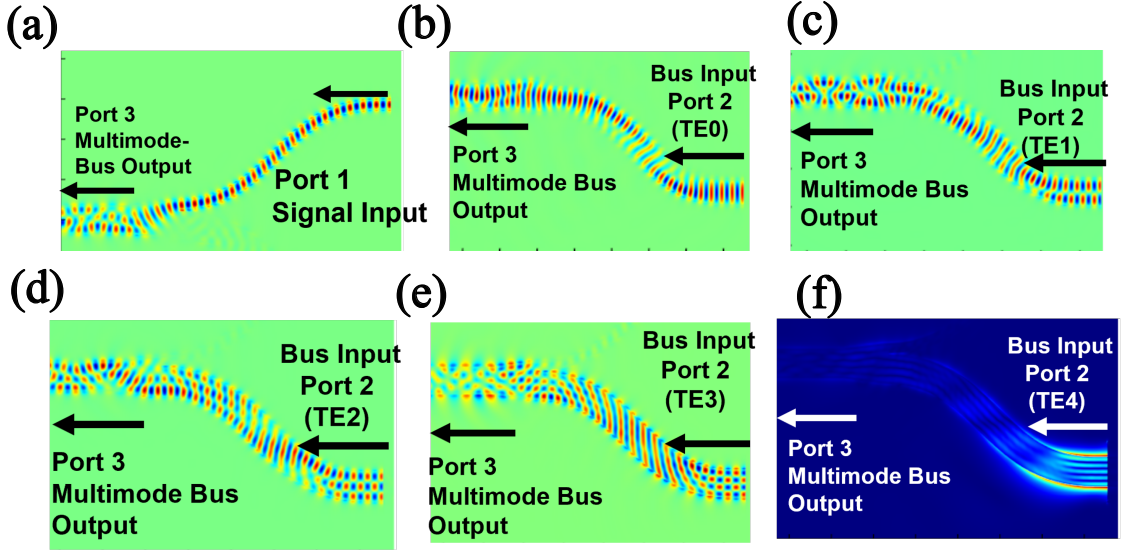
**Table 7.2:** Summary of the optimised design parameters for the asymmetric directional coupler obtained using particle swarm optimisation technique in Lumerical FDTD simulations.

| Device Parameters  | Value | unit   |
|--------------------|-------|--------|
| Bus Input Width    | 1460  | nm     |
| Signal Input Width | 500   | nm     |
| Taper Width        | 1960  | nm     |
| Taper Length       | 3870  | nm     |
| Bend Angle         | 5     | degree |

By using the particle swarm optimisation method in *Lumerical* FDTD, the geometry of the asymmetric Y coupler for average transmission is optimised by varying the waveguide width of the input bus waveguide, taper width and taper

length. The dimensions of the input and output multimode waveguides are kept equal to simplify the simulation optimisation. The final dimensions of the coupler are shown in Table 7.2.

It is worth noting the small footprint of the asymmetric y-coupler device as compared to a directional coupler. The coupler device has a footprint of less than  $5 \mu m$  in total length as compared to a directional coupler, where the coupling lengths can vary from  $10\text{-}30 \mu m$  for a similar  $220 \text{ nm}$  SOI platform. This results in a smaller overall computing unit cell and hence can result in over 5 times improvement in compute density reported in the previous work [15].



**Figure 7.2:** (a)  $E_y$ -field profile when  $TE_0$  signal is injected at port 1. (b)-(e)  $E_y$ -field profile when  $TE_0 - TE_3$  input is injected at Bus Input (port 2). (f) E-field magnitude profile when weakly supported  $TE_4$  input is injected at Bus Input (port 2).

It is observed that five modes are supported in the bus waveguide, with the first four modes strongly guided and a fifth weakly guided mode. In Figures 7.2 (a-f), the E-field profile of various modes through ports 2 and 3 of the coupler device is plotted. Furthermore, using simulations, the transmission of various modes from both the signal input and bus input waveguides is calculated. The modal power distribution for single inputs is shown in Table 7.3 and Table 7.4. The modal power distribution can be represented as transfer matrix **T1** and **T2**, respectively.



**Table 7.3:** Modal power distribution at Output Bus (Port 3) when single mode  $TE_0$  light is injected at Signal Input port (Port 1), with a net transmission of 97.5%.

| Mode Name               | Modal Distribution |
|-------------------------|--------------------|
| $TE_0$                  | 0.101              |
| $TE_1$                  | 0.342              |
| $TE_2$                  | 0.352              |
| $TE_3$                  | 0.148              |
| $TE_4$                  | 0.030              |
| <b>Net Transmission</b> | 0.975              |

**Table 7.4:** Modal power distribution at Output Bus (Port 3) when various modes  $TE_0$ - $TE_4$  are injected at Bus Input port (Port 2), with a complex power distribution across various modes.

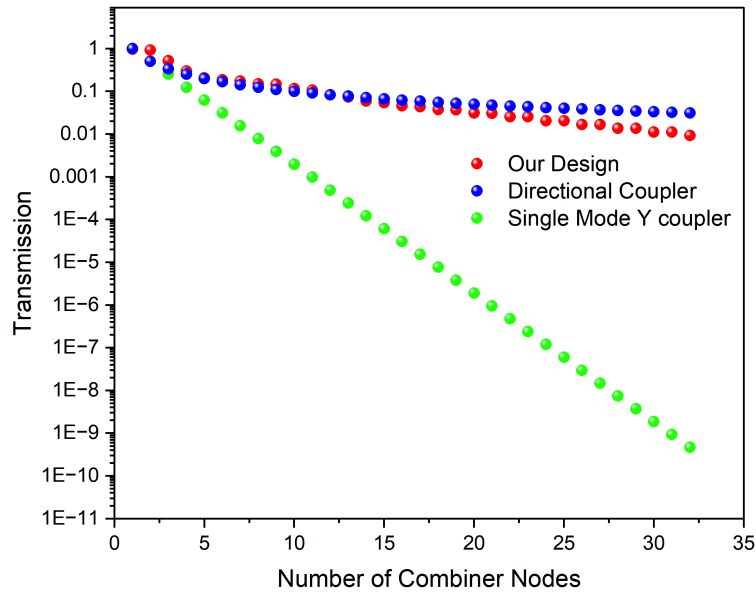
| Input $\rightarrow$     | $TE_0$ | $TE_1$ | $TE_2$ | $TE_3$ | $TE_4$ |
|-------------------------|--------|--------|--------|--------|--------|
| $TE_0$                  | 0.7769 | 0.0474 | 0.0668 | 0.0027 | 0.0015 |
| $TE_1$                  | 0.0534 | 0.4398 | 0.1086 | 0.0450 | 0.0020 |
| $TE_2$                  | 0.1175 | 0.1144 | 0.3862 | 0.0056 | 0.0005 |
| $TE_3$                  | 0.0393 | 0.3678 | 0.3610 | 0.0394 | 0.0058 |
| $TE_4$                  | 0.0038 | 0.0037 | 0.0318 | 0.7309 | 0.0958 |
| <b>Net Transmission</b> | 0.9908 | 0.9730 | 0.9543 | 0.8237 | 0.1057 |

In a multimode system, different modes have different group velocities in the waveguide. This results in a relative delay between signals of different modes and thereby making the transmission loss a complex function of phase velocities and, thus, difficult to calculate analytically. However, the power distribution transfer matrix (**T1** and **T2**) can be used to determine the approximate power drop with the increasing number of nodes. The approximate power distribution  $O_M$ , after M nodes, can be written analytically as :

$$O_m = (T_2)^{M-1} \times T_1$$

To account for the difference in group velocity due to the multimode system, the S-matrix of the coupler is calculated for various incident modes using Lumerical. Using the S-matrix, the transmission loss of the device for the increasing number of nodes is calculated. In Figure 7.3, the performance of the device with a directional coupler and single mode Y-coupler is compared. Assuming low insertion loss for a

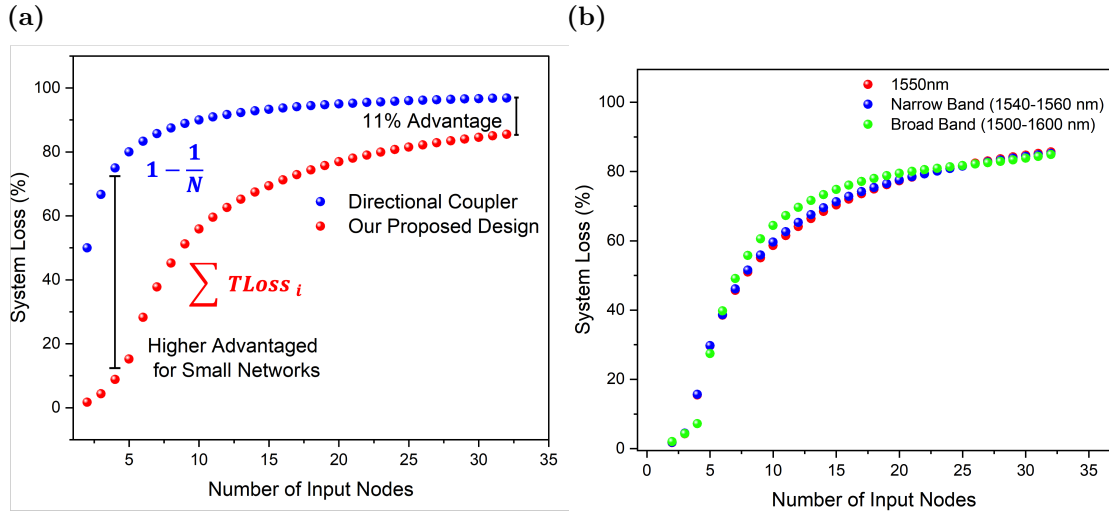
directional coupler, the transmission for a directional coupler is plotted as  $\frac{1}{M}$  for  $M$  nodes in the network. As expected, the transmission of an ideal single mode 3 dB Y-coupler drops exponentially  $\frac{1}{2^M}$  for an increasing number of nodes, while the transmission loss of the device is comparable to that of a directional coupler.



**Figure 7.3:** Comparison of performance of the device as compared to a directional coupler and single mode Y-coupler with an increasing number of coupler nodes.

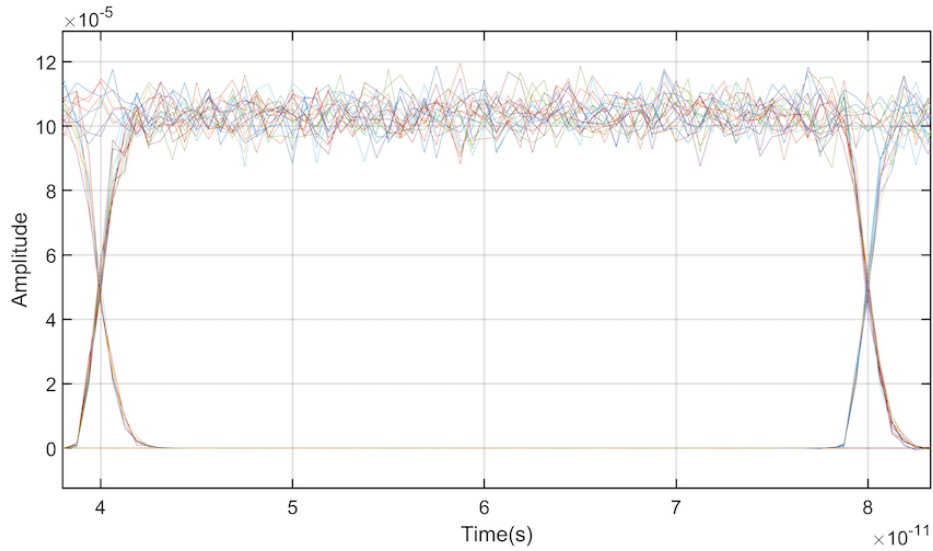
### 7.3 System Level Simulations

Having shown the higher performance of a single device in terms of optical loss as compared to a single mode Y splitter, a large optical crossbar array network is constructed to test the overall system loss and broadband optical performance of the device. Using Lumerical Interconnect simulations, a minimum improvement of 10% is shown in Figure 7.4a for larger networks, while more than 50% improvement is shown for smaller networks. As the number of nodes is increased, more optical power is lost in the system, as described above, thereby explaining the reason for such variations in efficiency.



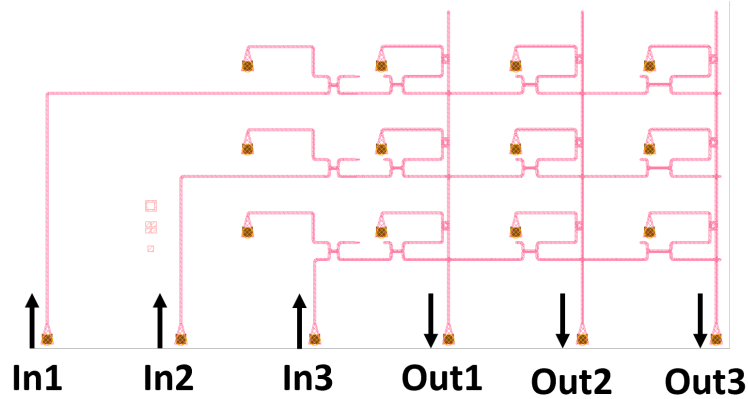
**Figure 7.4:** (a) Net System Loss (%) comparison of the device with an increasing number of nodes, (b) Broadband performance of the device with an increasing number of nodes. For wavelengths at 1500-1600nm, system losses are similar to that of single wavelengths.

Further system-level bandwidth performance simulations by varying the wavelength of the input signal are carried out. As shown in Figure 7.4b, there is no significant change in performance from the wavelength of 1500-1600 nm.



**Figure 7.5:** Eye Diagram for 32-node photonic network using the coupler design with a 10mW input signal modulated at 25 GHz, showing negligible bit error rate.

To test the effect of using a multimode device on the loss of computational bandwidth, calculations are performed in Interconnect. To set up the simulation, a

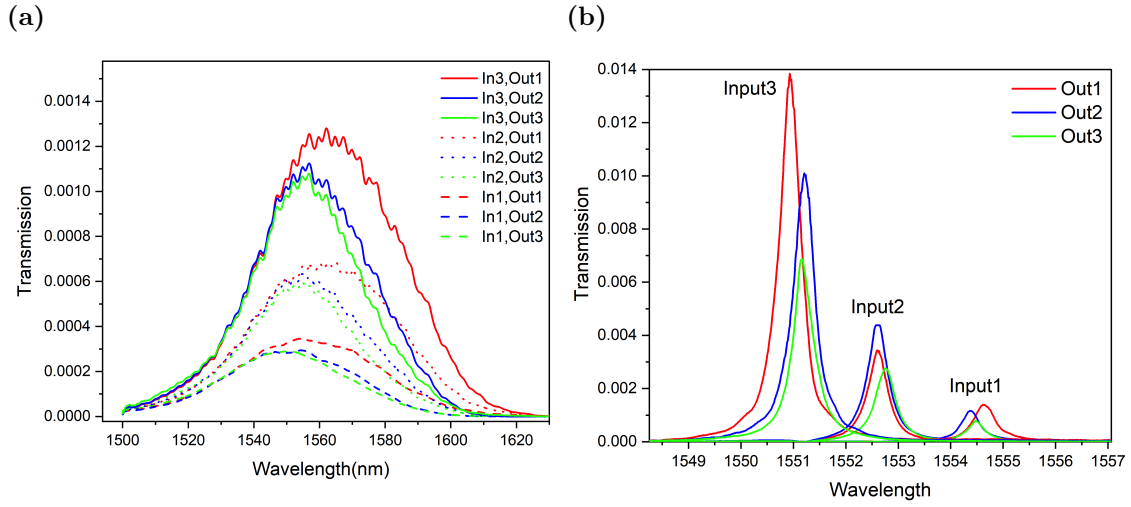


**Figure 7.6:** Schematic of a  $3 \times 3$  photonic network with WDM using ring resonator fabricated in IMEC foundry.

single source laser is modulated with a randomly generated signal. The output of the optical network is connected to the p-i-n junction photodetector. As shown by the eye diagram in Figure 7.5, there is no loss in computational bandwidth due to multimodal operation for a 32-node network with modulation up to 25 GHz with a very low bit error rate. Therefore using a design based on a multimode waveguide has no effect on data processing speeds.

## 7.4 Fabrication Tolerance Comparision

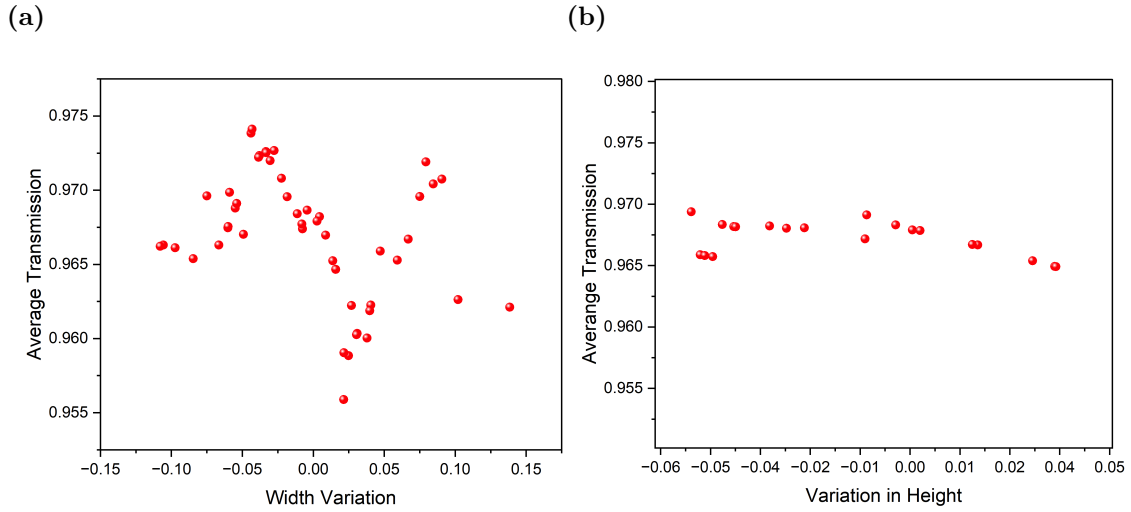
In this section, the advantages of using the proposed asymmetric Y-coupler network compared to designs based on directional couplers and WDM using ring resonators are discussed. For design using a directional coupler, varying the coupling length of the device results in a different power coupled into the drop port. A small variation in coupling gap or waveguide width from a variation in fabrication-related tolerances can result in a huge difference in the coupling coefficients of two coupled waveguides. A small  $3 \times 3$  optical network based on the crossbar array structure (Figure 7.6), as explained in the previous chapter, is fabricated in a foundry, IMEC. For a small device, the variation in the splitting of power from directional couplers due to fabrication variations across the wafer is observed in Figure 7.7a.



**Figure 7.7:** (a) Variation in coupling power for a  $3 \times 3$  photonic network based on directional coupler due to fabrication variation. (b) Variation in coupling power for a  $3 \times 3$  photonic network based on ring resonator due to fabrication variation and resonance wavelength.

Next, the photonic network design using the WDM technique with ring resonators to couple power is studied. As shown in Figure 7.7b, for the devices fabricated in a foundry, there is a variation in the resonance peaks resulting in the unequal coupling of optical power at each node. This shift in resonance can be corrected using active tuning methods such as microheaters, adding to the power budget of the network.

Having discussed the limitations of other photonic network architectures, the effect of fabrication variation on the performance of the device is studied. The sources of errors in fabrication are due to etching depth variation over a wafer and waveguide widths due to lithography variation. Therefore in the simulation, etching height and waveguide widths have been varied to account for fabrication variations. Assuming a 10% deviation in the fabrication process, simulations to test the performance of the proposed device have been proposed. As shown in Figure 7.8, the change in parameters has a limited effect on the performance of the device. These results are important for building a large, robust optical network.



**Figure 7.8:** (a)Effect of waveguide width variation due to fabrication variations on the average transmission of the device. The waveguide widths are varied by 10%, and the change in the transmission is plotted. (b)Effect of waveguide thickness due to fabrication variations on the average transmission of the device. The etching depth is varied by 5%, and the change in the transmission is plotted.

## 7.5 Critical Analysis

In this section, a critical analysis of scaling up the photonic network using a multimode system is presented. Following are the shortcomings and scope of improvement of the work presented here :

A multimode Y-coupler design eases the fabrication tolerance; however, the use of a multimode system comes with its own challenges.

- The primary drawback of the multimode design is due to challenges associated with the multimode waveguide itself. Using a multimode system limits the routing and bending of waveguides. Special design strategies have to be used to get over these challenges.
- Another challenge related to the multimode system itself is coupling light in and out using grating couplers. In a multimode, due to different effective indexes of different modes designing grating couplers matching Bragg's conditions for different modes can be quite challenging. However, 3D couplers based on the

reflection of light can be used to get away with these challenges. Another way of getting over light coupling challenges is by using integrated photodetectors on the chip.

- Photonic computing relies on wavelength division multiplexing methods to achieve high throughput. However, multiplexing and demultiplexing of signals using passive devices is challenging, resulting in making the implementation of a multimode system difficult.

## 7.6 Chapter Summary

In this work, a photonic device based on an asymmetric multimode Y-coupler to improve the scalability of a neuromorphic computing accelerator has been studied. This design has a high optical bandwidth with a small footprint and high tolerance to fabrication errors. For small photonic hardware, the proposed design has 50% better efficiency in terms of optical loss and for larger networks is 10% efficient. This architecture is compatible with foundry designs and can be readily implemented into other neuromorphic architectures.

# 8

## Amorphous Silicon Carbide photonic Platform

### Contents

---

|            |                                          |            |
|------------|------------------------------------------|------------|
| <b>8.1</b> | <b>Background . . . . .</b>              | <b>138</b> |
| <b>8.2</b> | <b>Platform Development . . . . .</b>    | <b>139</b> |
| 8.2.1      | Waveguide Parameters . . . . .           | 140        |
| 8.2.2      | Grating coupler sweep . . . . .          | 141        |
| 8.2.3      | Linear Optical Properties . . . . .      | 143        |
| 8.2.4      | Nonlinear Optical Properties . . . . .   | 147        |
| <b>8.3</b> | <b>SiC phase change memory . . . . .</b> | <b>149</b> |
| 8.3.1      | Simulation results . . . . .             | 149        |
| 8.3.2      | Experiments and Results . . . . .        | 151        |
| 8.3.3      | Dynamic Response . . . . .               | 153        |
| <b>8.4</b> | <b>Critical Analysis . . . . .</b>       | <b>155</b> |
| <b>8.5</b> | <b>Chapter Contributions . . . . .</b>   | <b>155</b> |
| <b>8.6</b> | <b>Chapter Summary . . . . .</b>         | <b>156</b> |

---

This chapter discusses the development of an amorphous silicon carbide photonic platform. Parts of the work presented in this chapter have been peer-reviewed and published in ACS photonics [199]

### 8.1 Background

Silicon carbide (SiC) has been a material of choice for optoelectronics, with the first demonstration of a LED using SiC. More recently, there has been a growing interest in SiC on oxide-integrated photonics platforms due to excellent linear and strong nonlinear properties [18, 200]. SiC platform offers a larger transparency window than SOI in visible and midinfrared wavelengths [201, 202]. Compared to SiN, it



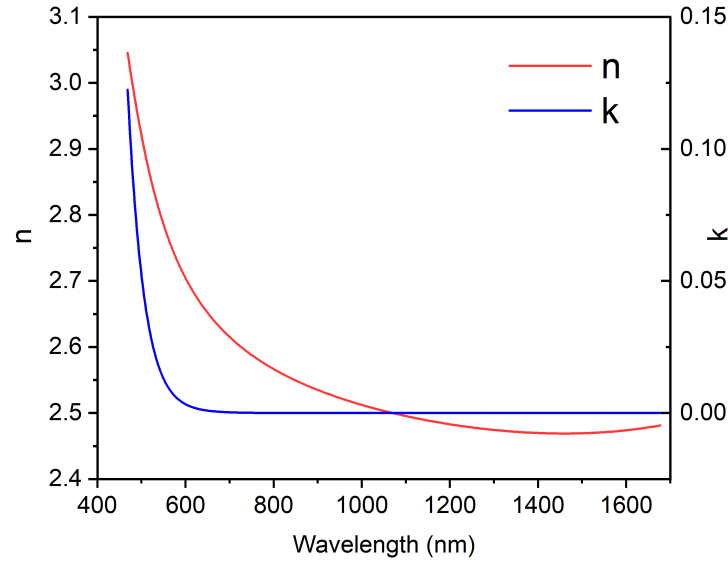
offers a smaller footprint due to a higher refractive index and larger contrast between the core and the cladding. Moreover, SiC devices are suitable for operations in harsh environments due to their excellent thermal [203] and mechanical properties. This makes SiC an ideal platform for sensing and onsite computing hardware, such as in space applications. SiC is already a platform of choice for optical quantum computing [200, 204]. Therefore, to tap the potential offered by this platform, there is an imminent need to increase functionality and tunability by co-integration with functional materials.

Phase change materials (PCMs) have opened up opportunities for reconfigurable integrated photonic devices and have been demonstrated in applications ranging from optical switch [191] to optical routing [205]. This study demonstrates the integration of PCM GST and AIST on low-temperature CVD-grown SiC waveguides. Further, time-resolved switching experiments are performed to study the dynamic response and obtain the switching speed of the memory device.

## 8.2 Platform Development

To study the linear and nonlinear properties of the amorphous silicon carbide platform, photonic waveguides are fabricated. SiC thin film of thickness 265 nm is grown on silicon dioxide using two different processes: plasma-enhanced chemical vapour deposition (PECVD) and inductive coupled chemical vapour deposition (ICPCVD) process. The development of the recipe for growing high-quality SiC has been performed by collaborators from TU Delft.

The deposition of amorphous silicon carbide films in PECVD is done using a mixture of SiH<sub>4</sub> (15 sccm), CH<sub>4</sub> (75 sccm) and Ar (285 sccm) with temperatures ranging between 300°C and 400°C. The chamber pressure was fixed at 1000 mTorr with a forward plasma power (PFW) of 20 W. For the deposition of ICPCVD a-SiC films, 15 sccm of SiH<sub>4</sub>, and 10 sccm of CH<sub>4</sub>, at temperatures ranging from



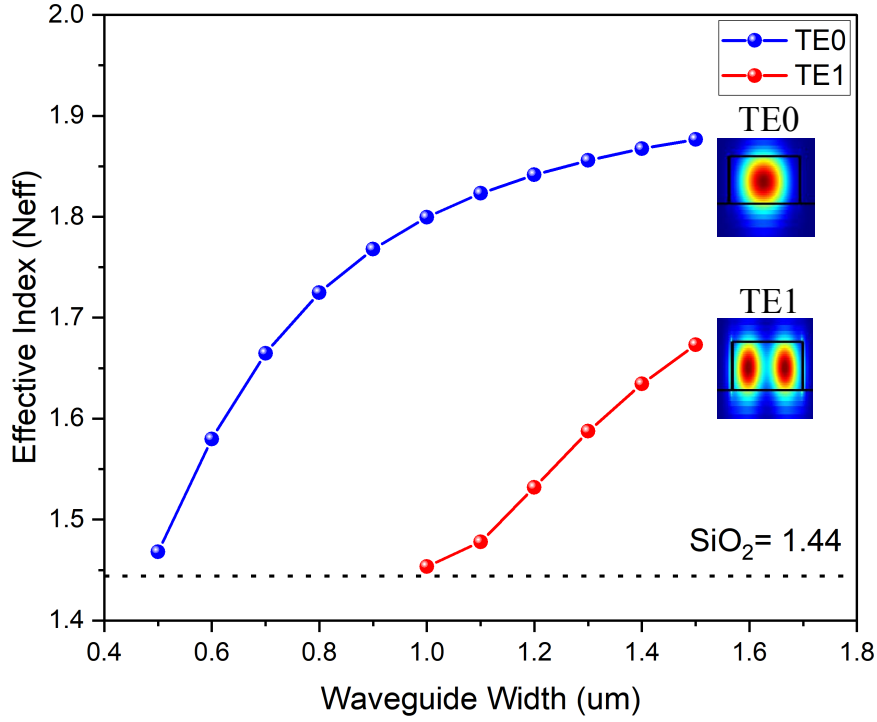
**Figure 8.1:** The refractive index of SiC thin film obtained experimentally using ellipsometry.

100°C to 400°C. Argon flow rate in the chamber is fixed at 10 sccm. The chamber pressure was kept at 2 mTorr for all recipes with a forward plasma power of 750 W. Deposition of SiC on an amorphous silicon oxide substrate at low temperature (100 °C- 400 °C ) results in amorphous silicon carbide film. It has been reported in the literature that for crystallisation of amorphous silicon carbide, annealing at temperatures higher than 600°C is required [206].

The work herein reported includes the design and evaluation of photonic devices to characterise the optical properties of the SiC platform.

### 8.2.1 Waveguide Parameters

First, the SiC waveguide on the insulator is developed. For ease of analysis, waveguides are designed for single-mode  $TE_0$  operation. To determine the waveguide parameters, mode simulations are performed in *Lumerical* by varying the waveguide width. Using the refractive index data obtained experimentally (Figure 8.1), the cutoff waveguide parameters are obtained as shown in Figure 8.2.



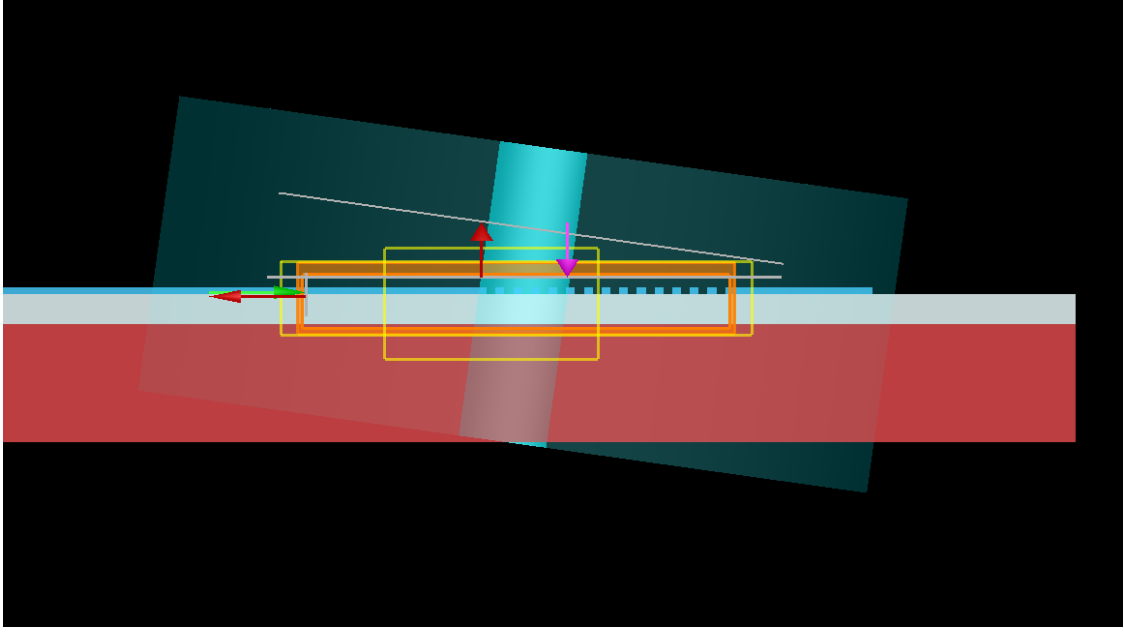
**Figure 8.2:** Effective refractive index of SiC on insulator waveguides obtained using Lumerical. The waveguide width is varied from 500nm to 1500nm, keeping the waveguide thickness fixed at 265 nm. The effective refractive index for fundamental  $TE_0$  and higher order mode at 1550 nm wavelength is plotted. The electric field distribution for  $TE_0$  and  $TE_1$  Mode is shown in the inset.

Due to the experimental limitation of growing thicker SiC films, the film thickness for SiC is fixed at 265 nm. From the simulations, a waveguide width of 500-900 nm allows for single-mode operation at 1550 nm wavelength. A waveguide width of 800 nm is selected for a confined single-mode  $TE_0$  mode.

Thus single mode waveguides at 1550 nm of width 800 nm are patterned using 100 kV EBL on ARP resist. The pattern is etched using reactive ion etching (RIE) with a mixture of  $SF_6$  and  $O_2$ .

### 8.2.2 Grating coupler sweep

This work uses grating couplers to couple light in and out of the waveguides. To determine grating parameters to couple light for 1550 nm, 2D FDTD simulations are set up as shown in Figure 8.3. Light from the fibre angled at  $8^\circ$  (corresponding



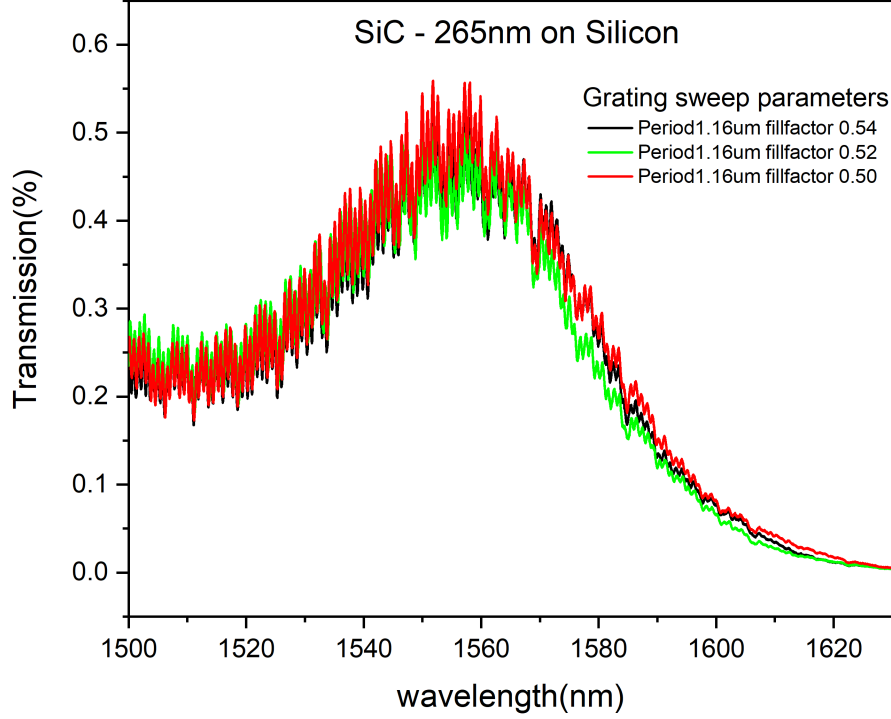
**Figure 8.3:** 2D FDTD Simulation setup used to simulate grating coupler parameters.

to the experimental condition) is used to input light to the grating couplers. As explained in Section 3.2, Bragg's condition needs to be satisfied to couple light into the waveguide. Thus the grating period is varied from 800 nm to 1200 nm to obtain optimum grating parameters.

Further, the fill factors of the grating coupler are varied to improve the grating coupler's efficiency. From simulations, a grating period of  $1.05 \mu\text{m}$  and a fill factor of 0.47 results in a coupling efficiency of 20%.

Next, using the grating coupler parameters as a reference, a grating coupler sweep is performed experimentally by changing both the grating period and the fill factor. The transmission of the grating couplers with the highest coupling efficiency at 1550 nm obtained is reported in Figure 8.4.

Grating couplers with period  $1.16 \mu\text{m}$  and fill factor of 0.50 show the best transmission with 7% coupling efficiency at 1550 nm wavelength. The discrepancy in experimental and simulation results is due to the limitation of 2D FDTD simulations to account for optical mode confinement along three dimensions.



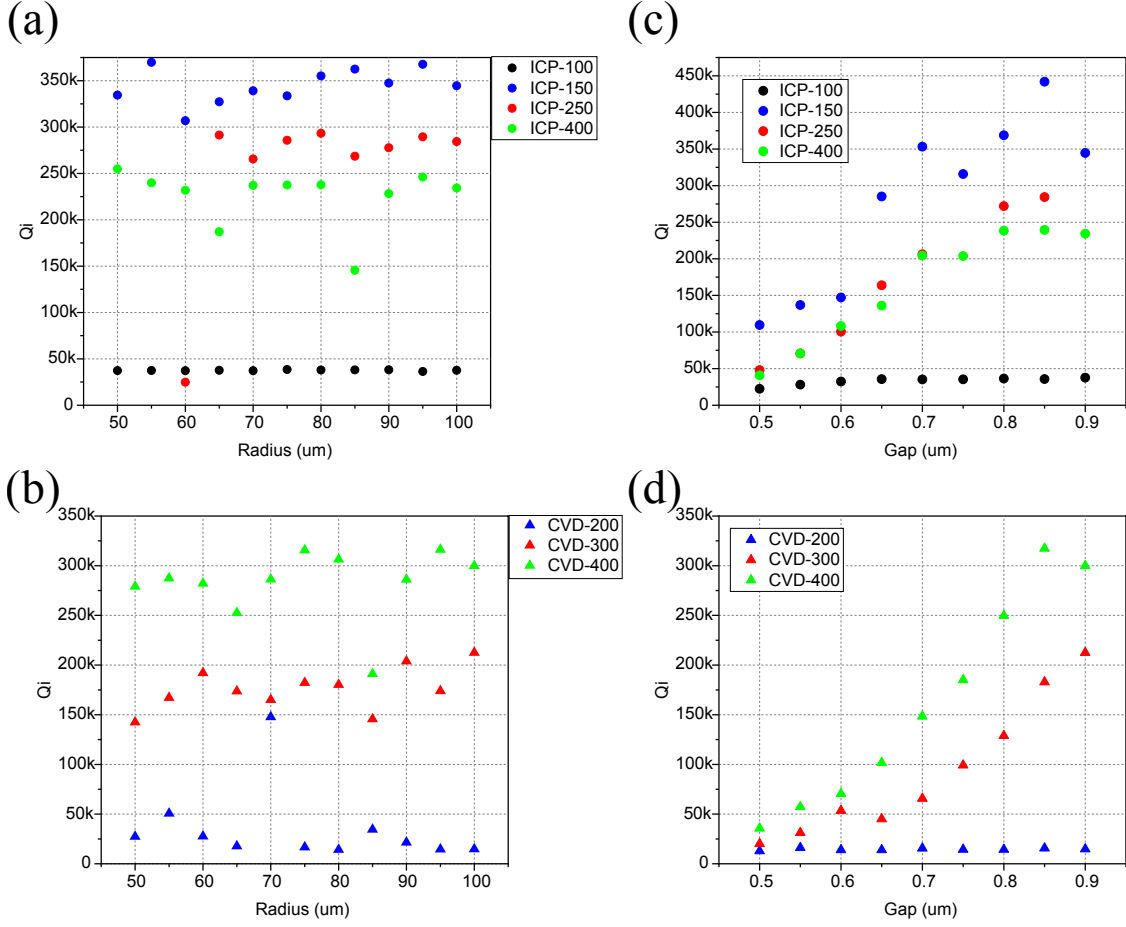
**Figure 8.4:** Experimental determination of grating coupler parameters for 1550 nm wavelength. Grating couplers with period  $1.16 \mu\text{m}$  and fill factor of 0.50 show the best transmission with 7% coupling efficiency.

### 8.2.3 Linear Optical Properties

After determining the waveguide and grating parameters, the linear properties, such as the optical loss of the SiC waveguides, are determined experimentally. To determine the propagation loss, SiC all-pass ring resonators are considered. As discussed in Section 3.3.4, the loaded quality factor of ring resonators can be used to determine the propagation loss.

The intrinsic quality factor ( $Q_i$ ) of ring resonators with waveguide width fixed at 800 nm for silicon carbide grown on an insulator using ICPCVD and PECVD processes at various growth temperatures is reported in Figure 8.5.

A ring resonator with a small ring radius introduces bending losses resulting in a drop in the quality factor. To determine the propagation losses, it is necessary



**Figure 8.5:** Intrinsic Quality factors of Ring resonators deposited using ICPCVD and PECVD process.

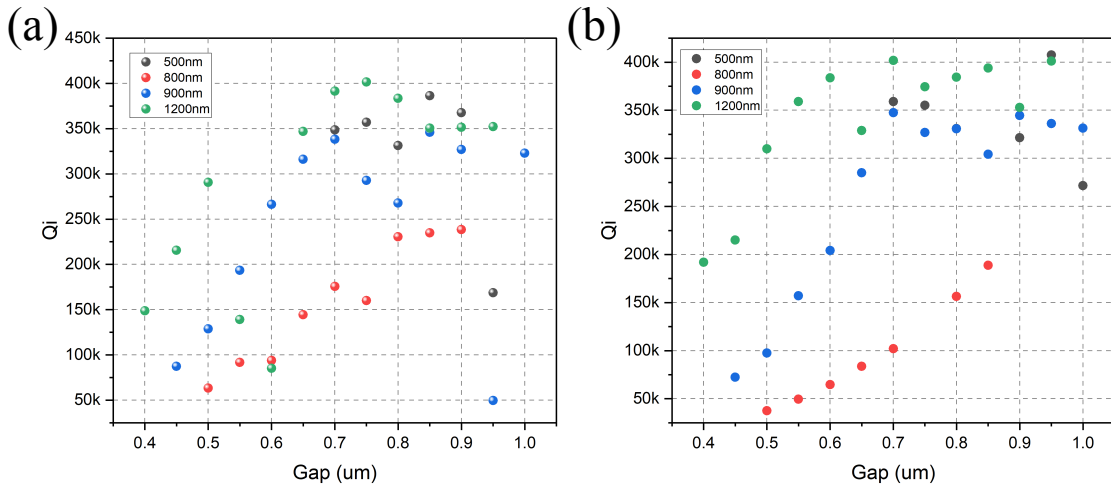
to choose ring resonators with sufficiently high ring radii to isolate ring bending losses. Therefore the effect of changing the ring resonator radius from 50-100  $\mu m$  on  $Q_i$  is determined for thin films grown at 100-400° C using ICPCVD in Figure 8.5(a). Similarly, for SiC, thin film grown using PECVD at 200-400° C is determined in Figure 8.5(b). It is determined that  $Q_i$  is independent of the ring radius for all devices.

Next, the coupling of the ring resonators with the input waveguide is varied by changing the coupling distance from 0.5-0.9  $\mu m$ , keeping the ring radius fixed at 100  $\mu m$  in Figure 8.5(c-d) for SiC growth using ICPCVD and PECVD respectively.

From the  $Q_i$ , the following observations are made :

1. The ICPCVD grown film at 150° C shows the best performance with quality factors approaching 250K.
2. For PECVD grown film, the films grown at 400° C shows the best performance with quality factors approaching 175K.
3. For PECVD films, increasing the film-grown temperature from 200° C to 400° C improves the device's performance in terms of optical losses.

Thus based on these observations, the thin films grown using ICPCVD at 150° C and 250° C are further improved, and the waveguide width is varied to change mode confinement. As before, the intrinsic quality factors of ring resonators with varying coupling gap is determined and plotted in Figure 8.6



**Figure 8.6:** Intrinsic Quality factors of Ring resonators with varying waveguide widths from 500nm-1200nm deposited using ICPCVD (a) grown at 150° C (b) grown at 250° C .

Ring resonators with more than 400K intrinsic quality factors can be obtained for 1200 nm wide waveguides. Increasing the waveguide width improves the confinement of the mode and, thus, a high-quality factor. Both Films grown at 150° C and 250° C show excellent performance in terms of linear optical properties.

### Propagation Loss

Using the relation between  $Q_i$  and propagation loss ( $\alpha$ ) as explained in Section 3.3.4, the propagation losses of the SiC waveguide are summarised in Table 8.1. Here the process ICP100 refers to the thin films grown using the ICPCVD process at 100° C . Further, the effect of varying the waveguide widths on the propagation loss is noted in Table 8.2.

**Table 8.1:** Summary of group index( $N_g$ ) and propagation loss  $\alpha$  for various SiC thin films deposited at various temperatures using PECVD and ICPCVD process obtained from the ring resonators of radius R.

| Process | $Q_i \times 10^4$ | R ( $\mu m$ ) | $\lambda$ (nm) | $N_g$ | $\alpha$ (dB/cm) |
|---------|-------------------|---------------|----------------|-------|------------------|
| ICP100  | 4.05              | 70            | 1572.57        | 2.24  | 22.14            |
| ICP150  | 44.2              | 100           | 1521.60        | 2.60  | 2.42             |
| ICP250  | 29.3              | 80            | 1552.01        | 2.48  | 3.41             |
| ICP400  | 25.5              | 50            | 1576.65        | 2.45  | 3.84             |
| CVD200  | 5.77              | 90            | 1548.88        | 1.87  | 13.47            |
| CVD300  | 21.35             | 100           | 1508.46        | 2.03  | 3.98             |
| CVD400  | 31.75             | 100           | 1512.88        | 2.32  | 3.03             |

**Table 8.2:** Effect of changing waveguide width on propagation loss  $\alpha$  and dispersion (D) for SiC thin films deposited using ICPCVD at 150° C and 250° C temperature, obtained from ring resonators.

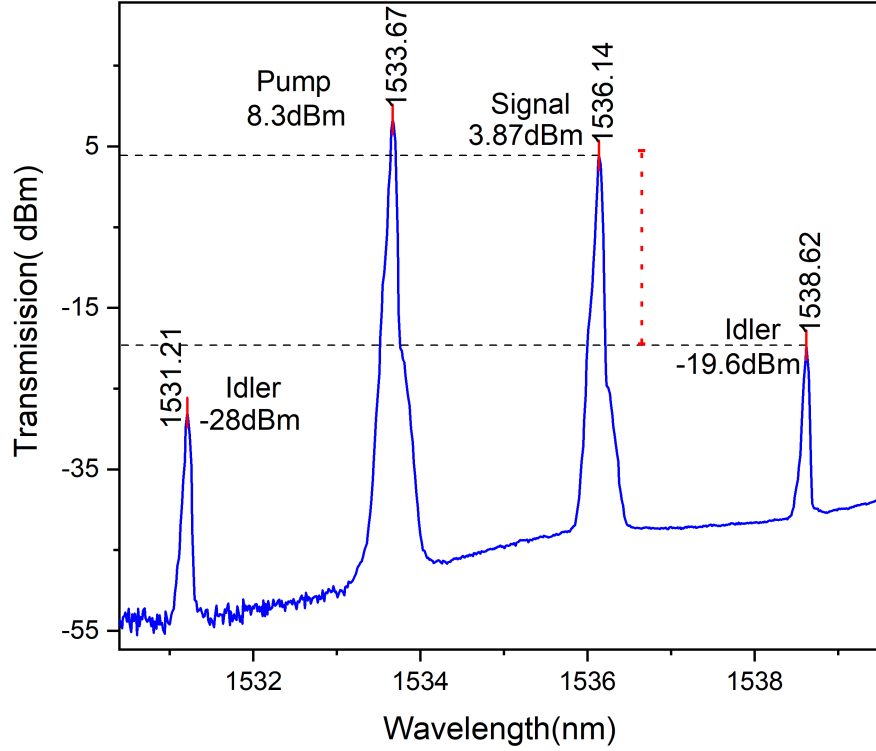
| Process | Width   | $Q_i \times 10^4$ | $\lambda$ (nm) | $N_g$ | $\alpha$ (dB/cm) | D ( $ps/nm \cdot km$ ) |
|---------|---------|-------------------|----------------|-------|------------------|------------------------|
| ICP150  | 500 nm  | 35.4              | 1570.90        | 1.99  | 0.98             | -5250                  |
| ICP150  | 900 nm  | 29.8              | 1528.72        | 2.51  | 1.50             | -5317                  |
| ICP150  | 1200 nm | 32.6              | 1571.24        | 1.95  | 1.04             | -4518                  |
| ICP250  | 500 nm  | 40.6              | 1572.12        | 1.99  | 0.85             | -5396                  |
| ICP250  | 900 nm  | 33.2              | 1533.99        | 2.52  | 1.35             | -5705                  |
| ICP250  | 1200 nm | 37.0              | 1558.81        | 2.55  | 1.21             | -5155                  |

Amorphous SiC waveguides with propagation losses less than 1 dB/cm have been obtained, thus showing an excellent linear optical property. Next, to characterise the SiC platform, the nonlinear properties are studied.



### 8.2.4 Nonlinear Optical Properties

Four-wave mixing experiments are performed to study the third-order nonlinear properties of SiC waveguide to extract the nonlinear refractive index. Using the setup described in Section 4.2.4. Using the four-wave mixing experiments as shown



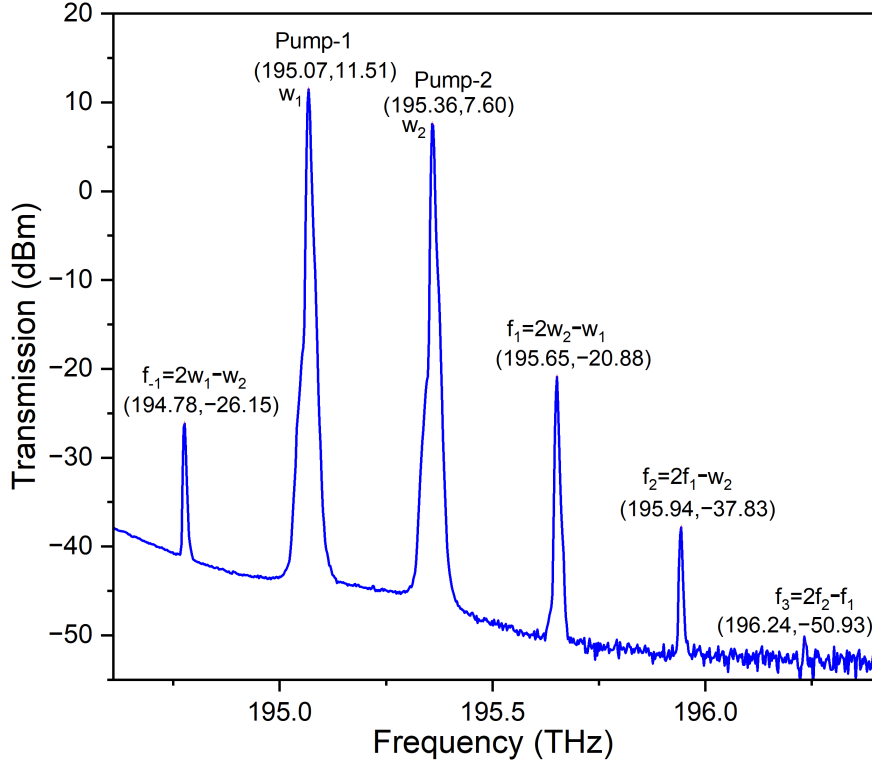
**Figure 8.7:** Four-wave mixing experiments demonstrating a conversion efficiency of -23 dB.

in Figure 8.7, the nonlinear refractive index of SiC is calculated. From equation 3.51, the conversion efficiency in the case of a ring resonator is related to nonlinear parameter according to the following relation :

$$\eta = \frac{P_i}{P_s} = |\gamma P_p L'|^2 F E_p^4 F E_s^2 F E_i^2, \text{ and}$$

$$L'^2 = L^2 \exp(-\alpha L) \left| \frac{1 - \exp(-\alpha L + i\Delta k L)}{-\alpha L + i\Delta k L} \right|^2$$

Using equation 3.50, a field enhancement factor of 5.35 is obtained. Since the pump and idler wavelengths are close, the phase mismatch term  $\Delta k$  can be assumed to



**Figure 8.8:** Observation of emergence of Kerr frequency comb using amorphous silicon carbide ring resonators.

be zero. Thus, an effective length  $L'$  of  $363 \mu m$  is obtained.

By rearranging the equations and using a conversion efficiency of -23 dB, the nonlinear coefficient  $\gamma$  is calculated to be  $33.22 W^{-1}/m$ . From equation 3.45, the nonlinear refractive index  $n_2$  is related to nonlinear coefficient  $\gamma$  as follows:

$$\gamma = \frac{n_2 \omega_p}{A_{eff} c}$$

Using the mode solver, the effective area  $A_{eff}$  is calculated to be  $0.137 \mu m^2$ . Thus resulting in the nonlinear refractive index  $n_2 = 1.18 \times 10^{-18} m^2/W$ . This is close to the nonlinear refractive index reported in similar work based on amorphous silicon carbide ( $n_2 = 4.8 \times 10^{-18} m^2/W$ ) [18].

Further, the pump and signal power is increased, resulting in the observation of frequency comb formation as shown in Figure 8.8.

The ring resonators are pumped with two high-power signals corresponding to frequency  $\omega_1$  and  $\omega_2$ . Due to four-wave mixing, this results in the formation of

idler frequencies  $f_{-1}$  and  $f_1$ . The idler frequency  $f_1$  further interacts with Pump signal  $\omega_2$ , resulting in the idler signal at  $f_2$  frequency. Further, frequencies  $f_1$  and  $f_2$  interact to produce frequency  $f_3$ .

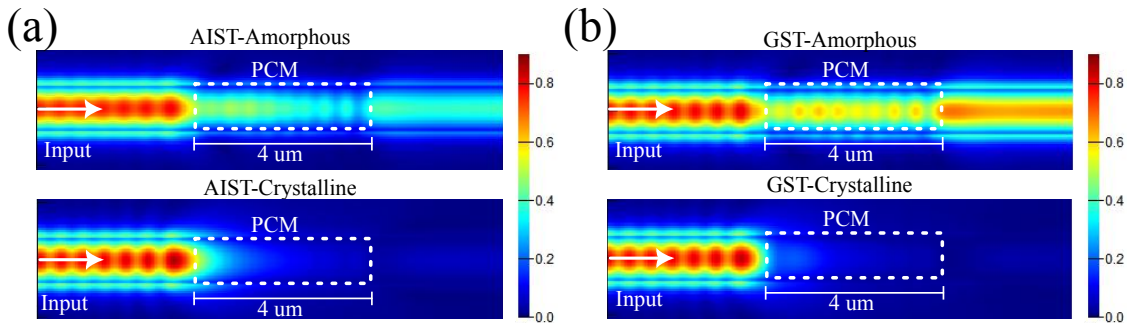
The conversion efficiency for four-wave mixing can be further improved by dispersion engineering. As noted above, the ring resonators have high dispersion, thus reducing the conversion efficiency.

Another reason for low conversion efficiency is related to achieving resonance conditions for the ring resonators for both the pump wavelengths. The ring resonators heat up when high-power light is coupled into the ring close to resonance wavelengths. Thus due to the thermo-optical effect, the resonance wavelength shifts.

### 8.3 SiC phase change memory

So far, the SiC platform with excellent linear and nonlinear optical properties has been discussed. This section studies the integration of phase change materials with the SiC platform. Here integration of two conventional phase change materials, AIST and GST, has been studied.

#### 8.3.1 Simulation results



**Figure 8.9:** E-field profile of different absorption of input light for the  $4\mu\text{m}$  long crystalline and amorphous state of (a) AIST and (b) GST obtained using Lumerical FDTD simulations.

First, using FDTD simulations, the interaction of 10 nm PCM (GST and AIST) is studied with SiC waveguides. To set up the simulation, a 265 nm thick, 800 nm wide SiC waveguide is considered with silicon dioxide as the substrate. 4  $\mu\text{m}$  long PCM is placed on top of the waveguide for near-field coupling. The refractive index for amorphous and crystalline AIST and GST used in the simulations is reported in Table 8.3.

**Table 8.3:** The refractive index of GST and AIST at 1550nm wavelength used in simulations.

| Refractive Index | Crystalline  | Amorphous    |
|------------------|--------------|--------------|
| AIST             | 6.53+i 2.86  | 4.29 +i 0.2  |
| GST              | 6.375+i 1.13 | 4.15 +i 0.05 |

Using simulations, the E-field profile is plotted for the amorphous and crystalline phases of AIST and GST in Figure 8.9 (a) and (b), respectively. As expected, crystalline PCM has higher absorption of light as shown in Figure 8.9. Thus switching PCM from an amorphous to a crystalline phase should result in a change in the transmission of the waveguide.

For all-optical switching, it is necessary that the ‘Pump’ pulses heat the PCM thin films above the melting point to achieve amorphisation through the melt-quench process. Thus using FEM simulations in COMSOL, thermal simulations are performed to simulate PCM switching on the SiC platform. The thermal parameters in Table 8.4 are used to set up the simulation [207]. A 5 mW 20 ns pulse is sent

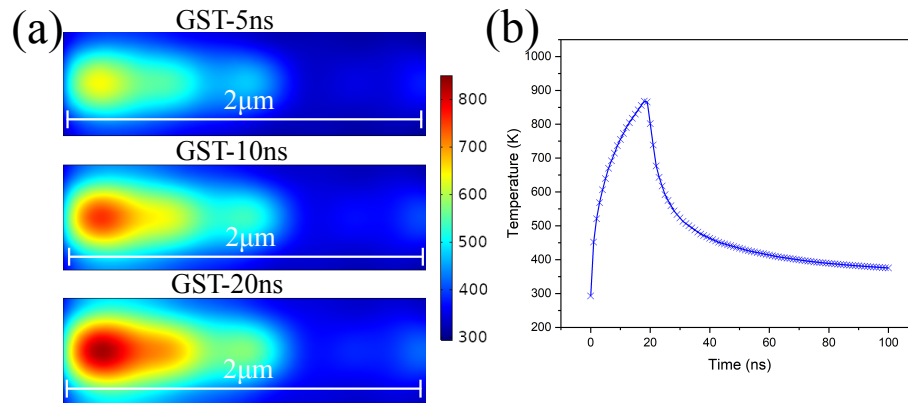
**Table 8.4:** Thermal parameters used in FEM simulations. Thermal conductivity(K), Heat capacity ( $C_p$ ) and Density( $\rho$ ) for various materials are reported.

| Material       | K ( $\text{W}/\text{m} \cdot \text{K}$ ) | $C_p$ ( $\text{J}/\text{kg} \cdot \text{K}$ ) | $\rho$ ( $\text{Kg}/\text{m}^3$ ) |
|----------------|------------------------------------------|-----------------------------------------------|-----------------------------------|
| a-SiC          | 1.8                                      | 700                                           | 2329                              |
| $\text{SiO}_2$ | 1.38                                     | 703                                           | 2203                              |
| AIST           | 3.2                                      | 270                                           | 6590                              |
| GST            | 0.57                                     | 250                                           | 6300                              |

through the memory device for AIST and PCM devices. The surface temperature of

the PCM thin film is noted. The temperature variation over time from 5-20 ns for GST and AIST devices is shown in Figure 8.10(a) and 8.11(a), respectively.

From simulations, two important observations are made. For AIST, a higher temperature can be achieved due to a higher absorption coefficient. Thus shorter pulses should be able to result in switching. In Figure 8.10(b) and Figure 8.11(b), the maximum temperature over time is plotted to compare the maximum temperature reached from AIST and GST devices and measure the cooling rate.



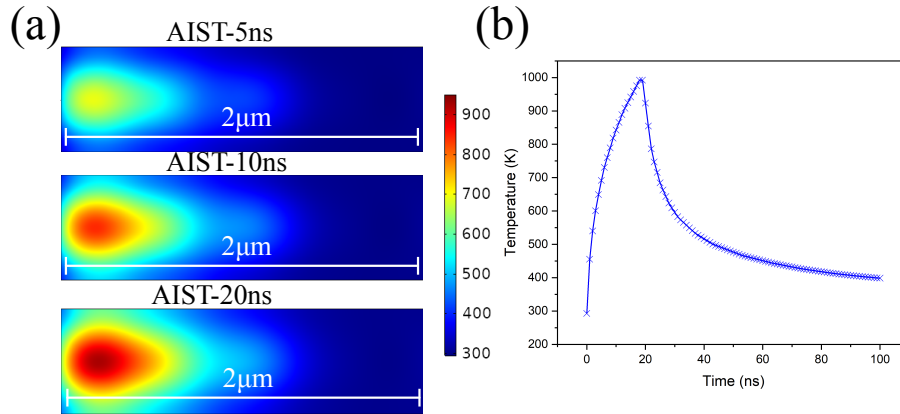
**Figure 8.10:** (a) FEM thermal simulations for GST memory device using 5mW, 20ns pulse on changing the time from 5ns-20ns. (b) Plot for maximum temperature reached for GST device for a 20ns pulse and thus subsequent cooling rate.

Second, with the increase in the pulse width, the area above the melting point increases. Thus by changing the pulse widths larger area can be switched. This is tested experimentally in the next section.

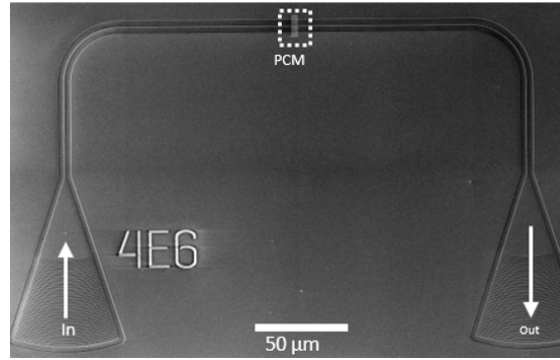
### 8.3.2 Experiments and Results

#### Device Fabrication

SiC waveguides are fabricated using PECVD, as described previously. Further, 10 nm thick PCMs (GST or AIST) are sputtered on a waveguide with 10nm ITO as a capping layer (Figure 8.12). The sputtering parameters used have been reported in Table 4.1.



**Figure 8.11:** a)FEM thermal simulations for AIST memory device using 5mW, 20ns pulse on changing the time from 5ns-20ns.(b) Plot for maximum temperature reached for AIST device for a 20 ns pulse and subsequent cooling rate.

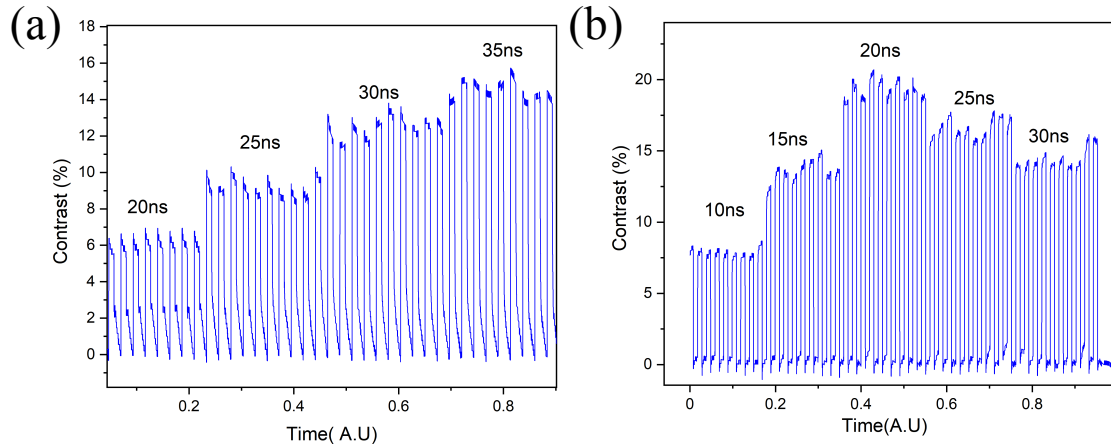


**Figure 8.12:** SEM image showing fabricated SiC waveguide with 10 nm PCM and ITO as a capping layer deposited using RF sputtering.

### Switching Experiments

The as-deposited PCM photonic memory devices are annealed at 250° C to achieve a crystalline phase. The pump-probe technique, as described in Section 4.2.2, is used to switch the PCM memory device and monitor the change in transmission. However, here, instead of using an fs laser to switch PCM (as used in the previous chapter for antimony), the CW laser is modulated using an EOM to vary the pulse widths from 10-100 ns.

Sending high-energy ‘Pump’ pulses through the device results in near-field absorption of light, which heats up the PCM, as shown in the simulation results in the previous section. Starting with the crystalline phase, high-energy optical



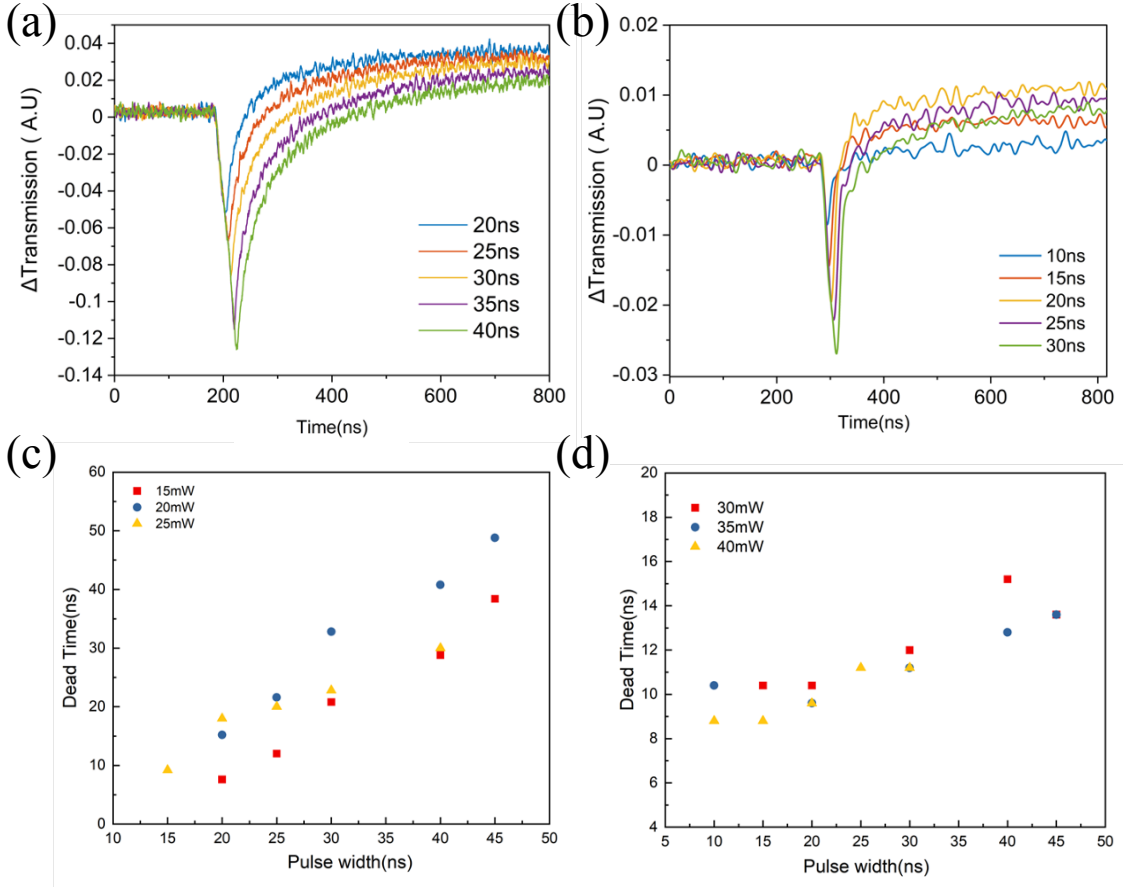
**Figure 8.13:** (a) Reversible switching of GST on SiC by varying the pulse width from 20-35ns for amorphisation. b) Reversible switching of AIST on SiC by varying the pulse width from 10-30ns for amorphisation.

pulses (15 mW) are sent through memory devices. Depending on the phase of PCM, the device's transmission is changed—partial amorphisation results in intermediate transmission levels. Hence pulse widths of the pump laser are varied to achieve such partial amorphisation.

For GST, multilevel memory operation can be obtained by varying the pulse widths from 20-35 ns and keeping the amplitude constant, Figure 8.13(a). AIST shows faster switching speeds than GST, and a 10 ns pulse is enough for amorphisation and varying the amorphisation pulse width from 10-30 ns results in multilevel memory operation, Figure 8.13(b). For AIST, increasing the pulse width beyond 20 ns reduces amorphisation due to partial recrystallisation and hence a lower switching contrast.

### 8.3.3 Dynamic Response

So far, reversible switching of GST and AIST has been achieved on the SiC platform. To understand the effect of pulse width and pulse energies on switching speed, experiments to capture the dynamic switching response of memory devices are performed.



**Figure 8.14:** a-b) Time-resolved dynamic switching response of GST and AIST, respectively, on SiC obtained by varying the pulse width and keeping the pulse amplitude constant c-d) Dead time of GST and AIST memory device for different pulse amplitude and pulse width.

Using the pump-probe technique as before, the PCM memory cells are switched. To record the high-speed change in transmission changes for the probe laser, a high-speed photodetector is used. The high-speed signals are recorded onto an oscilloscope for further analysis.

The dynamic switching response of GST and AIST switching on the SiC platform by varying the switching pulse width of fixed amplitude (15mW) is shown in Figure 8.14(a)-(b). Both GST and AIST show thermal-optical effects. This results in a drop in transmission before reaching the equilibrium state on switching and a delay in programming and readout, termed deadtime [7].

For GST increasing the pump width increases dead time (Figure 8.14c); 20



ns pulse of power 15 mW results in a dead time of 10 ns comparable to the time reported in SiN [208]. Moreover, AIST shows superior speed performance and increasing the switching amplitude and pulse width does not impact the dead time (Figure 8.14d).

## 8.4 Critical Analysis

In this section, a critical analysis of the development of the SiC platform is presented. Following are the shortcomings and scope of improvement of the work presented here :

- The linear properties of the SiC platform demonstrated here are exciting; however, it still lags behind state-of-the-art platforms in terms of optical losses.
- The conversion efficiency of four-wave mixing experiments is low and comparable to other work using a-silicon carbide. There is a need to explore dispersion engineering to improve conversion efficiency.
- For PCM integration, in terms of switching speeds, there is only a marginal benefit compared to SiN. As compared to silicon devices, the performance of SiC still lags behind. This is a limitation of an amorphous system where thermal conductivity is far lower than a crystalline system.

## 8.5 Chapter Contributions

The development of the growth of the SiC recipe and fabrication process has been performed by Bruno Lopez, Iman Zadeh and the team at TU Delft. The author of the thesis has contributed towards the design of all photonic devices for linear and nonlinear measurements and has performed the measurements independently. For PCM integration, the author has done all the secondary fabrication for PCM

deposition and performed the switching experiments with help from Yuhan He at Oxford.

## 8.6 Chapter Summary

In this chapter, a low-temperature CVD-grown amorphous silicon carbide platform has been studied with ultra-low propagation losses. Next, the third-order nonlinear refractive index of SiC is calculated using four-wave mixing experiments, and efforts have been made to generate frequency comb using this platform. Finally, to show the utility of such a platform for computing, PCM-based memory cells have been demonstrated. In terms of speed, the performance of the SiC device is comparable to SiN and is better in terms of device footprint. Integration of PCM is not limited to memory and computing applications but opens up a plethora of opportunities for programmable photonic elements for switching and routing in optical quantum computing applications.

# 9

## Conclusion and Outlook

### 9.1 Conclusions

In this thesis, three important building blocks for fast and scalable photonic computing have been studied. First, a memory device based on elemental antimony with switching speeds of 500 MHz is obtained. Such memory devices demonstrate multi-level memory operation with a retention time of tens of seconds. Thus such memory devices provide an ultrafast memory element for photonic computing.

Next, to improve the scalability of photonic networks, two approaches have been demonstrated. First, a reduced rank matrix approach is used to decrease the number of programmable elements, resulting in a less aggressive scaling of photonic networks. One of the limiting factors for photonic network scalability is optical power loss due to the incoherent addition of light. To tackle this challenge, a multimode system scheme is studied using simulations which show over 10% improvement in terms of system losses.

Finally, a low-temperature photonic platform based on silicon carbide is developed. This allows for the cointegration of various active photonic components. The linear and nonlinear characterisation is performed to demonstrate ultra-low loss waveguides and third-order nonlinearity of the same order as in silicon. To enhance the functionality of this platform, the integration of two PCM is studied, all-optical switching is achieved, and memory devices have been demonstrated. Using dynamic switching responses, it is shown silicon carbide performs similarly to

the SiN platform in terms of switching speeds. However, higher index contrast for SiC ensures smaller device footprints, thus enabling high-density packaging.

## 9.2 Outlook

### 9.2.1 Improving retention time for Sb

Single-element antimony-based memory devices provide an opportunity for scaling up PCM-based memory cells. However, the retention time of such memory cells needs to be further improved. As already mentioned in Chapter 5, various methods can be explored to address this. Further nanoconfinement of the Sb cells can result in a lowering of recrystallisation behaviour. Further atomically thin 2D antimonene [209] thin films could also be potentially integrated with photonic devices to further improve the retention time of memory states.

In terms of switching speeds, there is a need for analysis of switching pulse width and switching wavelengths to further improve switching speeds. In the GeSb system, it has been observed that femtosecond pulses can result in slower transformation time due to recalescence effects [210]. Thus a study of switching pulse width and device switching speed can help improve switching dynamics.

To improve switching energy-related performance and increase the functionality of Sb device, mixed mode nanogap devices as described by Farmakidis et al. [211, 212] can be tried. The nanogap devices can result in the confinement of the switching area to the nanoscale and will result in improvement in both switching energies and retention time. Such mixed-mode devices could also help in integration in neuromorphic computing where PCM is switched electrically and probed optically.

### 9.2.2 Scalability of Photonic Networks

The scalability of photonic networks has been limited due to network losses and larger footprint compared to their electronic counterparts. The reduced rank

approach tackles the footprint issues, however, at the cost of increased network losses. These network losses due to the incoherent addition of light can be reduced by using a multimode waveguide system. Another way of improving the scalability issue is by using integrated optical amplifiers in the system [213].

As mentioned in Chapter 7, using ring resonator-based WDM methods for the addition of light is an efficient way of implementing photonic networks but is limited by fabrication tolerances. Such networks based on ring resonators can be trimmed post fabrication [214] to achieve desirable wavelength tuning.

### **9.2.3 Cointegration of photonic platforms**

The low growth temperatures of the amorphous silicon carbide platform demonstrated in Chapter 8 offer opportunities for co-integration with other active platforms. The low growth temperature allows easy fabrication steps to be performed, such as lift-off on top of other photonic platforms. Such a platform could also enable 3D stacking of photonic networks allowing higher packaging. Amorphous silicon carbide can be easily grown on top of electronic circuits, thus making easy integration of optics and electronics.

# List of Publications

## Journal Articles

1. **Aggarwal, S.**, Dong, B., Feldmann, J., Farmakidis, N., ... & Bhaskaran, H. (2023) Reduced Rank Photonic Computing Accelerator, *Optica* 10, 1074-1080 (2023)
2. **Aggarwal, S.**, Milne, T., Farmakidis, ... & Bhaskaran, H. (2022). Antimony as a Programmable Element in Integrated Nanophotonics. *Nano Letters*, 22(9), 3532-3538.
3. Rodriguez, B\*, van der Kolk, B\*, **Aggarwal, S\***, Sharma, N., ... & Zadeh, I. E. (2023). High-quality amorphous Silicon Carbide for hybrid photonic integration deposited at low temperature.\*equal contribution *ACS Photonics* 2023, 10, 10, 3748–3754
4. B. Dong, **S. Aggarwal** ...& Bhaskaran, H. (2023) "Higher-dimensional processing using a photonic tensor core with continuous-time data". *Nature Photonics* (2023)
5. Farmakidis, N., J. S. Lee, J. Feldmann, M. Wang, Y. He, **S. Aggarwal**, B. Dong, W. H. P. Pernice, and H. Bhaskaran. "Scalable high-precision trimming of photonic resonances by polymer exposure to energetic beams." *Nano Letters* (2023).
6. Wang, M., Lee, J. S., **Aggarwal, S.**, Farmakidis, N., He, Y., Cheng, T., & Bhaskaran, H. (2023). Varifocal Metalens Using Tunable and Ultralow-loss Dielectrics. *Advanced Science*, 2204899.
7. Zhou, W., Farmakidis, N., Li, X., Tan, J., **Aggarwal, S.**, Feldmann, J., ... & Bhaskaran, H. (2022). Artificial Biphasic Synapses Based on Nonvolatile Phase-Change Photonic Memory Cells. *physica status solidi (RRL)–Rapid Research Letters*, 16(9), 2100487.
8. Farmakidis, N., Youngblood, N., Lee, J.S., Feldmann, J., Lodi, A., Li, X., **Aggarwal, S.**, Zhou, W., Bogani, L., Pernice, W.H. and Wright, C.D., 2022. Electronically reconfigurable photonic switches incorporating plasmonic structures and phase change materials. *Advanced Science*, 9(20), p.2200383.
9. Farmakidis, N., Swett, J. L., Youngblood, N., Li, X., Evangelini, C., **Aggarwal, S.**, ... & Bhaskaran, H. (2021). Exploiting rotational asymmetry for sub-50 nm mechanical nanocalligraphy. *Microsystems & Nanoengineering*, 7(1), 84.

10. Li, X., Youngblood, N., Zhou, W., Feldmann, J., Swett, J., **Aggarwal, S.**, ... & Bhaskaran, H. (2020, December). On-chip phase change optical matrix multiplication core. In 2020 IEEE International Electron Devices Meeting (IEDM) (pp. 7-5). IEEE.
11. **Aggarwal, S.**, Farmakidis, N., .... & Bhaskaran, H., (2023) " All optical tunable RF filter using elemental antimony" (In Review)
12. B. Dong, F.B Pluckelmann , **S. Aggarwal** ... & Bhaskaran, H. (2023) "Partial coherence enhanced parallelized photonic computing" (In Review)
13. J. S. Lee, Farmakidis, N., **S. Aggarwal** ... & Bhaskaran, H. "Spatio-spectral reconfiguration in silicon photonics" (In Review)

### Conference Presentations

1. **Aggarwal, S.** et al. "Antimony as a Tuneable Element in Integrated Photonics", In MRS Spring Meeting EL01.06.08 2023
2. **Aggarwal, S.** et al. Design for Robust and Efficient Neuromorphic Photonic Accelerator. In 2022 IEEE International Conference on Emerging Electronics (ICEE) (pp. 1-4). IEEE.
3. **Aggarwal, S.** et al. Ultrafast Switching in Integrated Photonics using Antimony. In CLEO: Science and Innovations (pp. SF2N-4). Optica Publishing Group.
4. **Aggarwal, S.** et al. "Tuneable silicon carbide photonics using phase change materials," Proc. SPIE PC12010, Photonic and Phononic Properties of Engineered Nanostructures XII, PC120100C (5 March 2022).
5. N. Farmakidis, N. Youngblood, W. H. P. Pernice, H. Bhaskaran, **Samarth Aggarwal**, "Engineering nanostructures at the interface between photonics and electronics," Proc. SPIE PC12010, Photonic and Phononic Properties of Engineered Nanostructures XII, PC120100O (5 March 2022) ( invited talk on behalf of H.Bhaskaran).
6. **Aggarwal, S.** et al. "Reconfigurable silicon carbide photonics using phase change materials. Poster presentation E/PCOS 2022. (Best poster award)

### Contributed Review Article

1. "Recent Advances and Future Prospects for Memristive Materials, Devices and Systems" ACS Nano 2023.

### Patents

1. Parallel in-memory photonic computing using continuous-time data representation B. Dong, **S. Aggarwal**, H. Bhaskaran, and X. Li (filed)





# Appendices



## Code Availability

The codes associated with the design of the Matrices using IMEC PDK can be downloaded from the following website:

[https://github.com/smrthagarwal/Thesis\\_Codes/tree/main/PDKdesign](https://github.com/smrthagarwal/Thesis_Codes/tree/main/PDKdesign)

The raw data and python scripts used to process the data in reduced rank matrix Chapter 6, can be downloaded from the following website:

[https://github.com/smrthagarwal/Thesis\\_Codes/tree/main/Reduced\\_Ranks](https://github.com/smrthagarwal/Thesis_Codes/tree/main/Reduced_Ranks)

# B

## DAC controller codes

Arduino codes to communicate with the DAC evaluation board.

```
1 # include <SPI.h>
2 const int BUFFER_SIZE = 3;
3 byte buf[BUFFER_SIZE];
4 void setup() {
5   Serial.begin(500000); // opens serial port, sets data rate to
      50000 bps
6   SPI.begin();
7   SPI.setBitOrder(MSBFIRST);
8   SPI.setDataMode(SPI_MODE1);
9   SPI.setClockDivider(SPI_CLOCK_DIV2); //divide the clock by 8
10  digitalWrite(SS, LOW);
11 }
12 void loop() {
13   if (Serial.available()){ // received information
14     digitalWrite(SS, LOW);
15     Serial.readBytes(buf, BUFFER_SIZE);
16     int a= buf[0];
17     int b= buf[1];
18     int c= buf[2];
19     SPI.transfer(a);
20     SPI.transfer(b);
21     SPI.transfer(c);
22     digitalWrite(SS, HIGH);
23   }
24 }
```

A sample Python code script to interface with Arduino.

```
1 from bitstring import BitArray
2 import numpy as np
3 import serial
4 import time
5
6 def voltage_to_serial_word(Vout, Voutx, Vref = 3.0, M = 16383, C
      = 8192, offset = 5461):
7     l = int(
```

```

8      (Vout*16384/(4*Vref) + offset + 8192 - C) * 16384/(M+1)
9      )
10     remainder = (Voutx % 8)
11     quotient = (Voutx // 8)
12     A2_A0 = remainder
13     A5_A3 = quotient + 1
14     output='0b11' + format(A5_A3, '#05b')[2:] + format(A2_A0, '
#05b')[2:] + format(1, '#016b')[2:] + '00'
15     b=BitArray(bin=output[2:])
16     b=b.uint
17     alpha=str(b)+ str('\r')
18     b = bytes(alpha, 'utf-8')
19     return b
20
21 def arduino_initialize() :
22     arduino = serial.Serial(port = 'COM5', baudrate=500000,
23     timeout=0 , write_timeout=0)
24     time.sleep(1)
25     return arduino
26
27 arduino = arduino_initialize()
28 alpha=str('\r')
29 ending = bytes(alpha, 'utf-8')
30 a=voltage_to_serial_word(-2,10)
31 d=arduino.write(a)
32 d=arduino.write(ending)

```

# References

1. Silver, D., Huang, A., *et al.* Mastering the game of Go with deep neural networks and tree search. *Nature* **529**, 484–489 (Jan. 2016).
2. *AI leaps forward, mastering the ancient game of Go / CEVA*
3. Goi, E., Zhang, Q., *et al.* Perspective on photonic memristive neuromorphic computing. *Photonix* **1**, 1–26 (2020).
4. Zou, X., Xu, S., *et al.* Breaking the von Neumann bottleneck: architecture-level processing-in-memory technology. *Science China Information Sciences* **64**, 1–10 (June 2021).
5. Shastri, B. J., Tait, A. N., *et al.* Photonics for artificial intelligence and neuromorphic computing. *Nature Photonics* **15**, 102–114 (Jan. 2021).
6. Wuttig, M., Bhaskaran, H., *et al.* *Phase-change materials for non-volatile photonic applications* 2017.
7. Ríos, C., Stegmaier, M., *et al.* Integrated all-photonic non-volatile multi-level memory. *Nature Photonics* **9**, 725–732 (Nov. 2015).
8. Li, X., Youngblood, N., *et al.* Fast and reliable storage using a 5 bit, nonvolatile photonic memory cell. *Optica* **6**, 1–6 (2019).
9. Yeoh, P., Ma, Y., *et al.* Thermal-gradient-driven elemental segregation in Ge<sub>2</sub>Sb<sub>2</sub>Te<sub>5</sub> phase change memory cells. *Applied Physics Letters* **114**, 163507 (Apr. 2019).
10. Xie, Y., Kim, W., *et al.* Self-Healing of a Confined Phase Change Memory Device with a Metallic Surfactant Layer. *Advanced Materials* **30**, 1705587 (Mar. 2018).
11. DeBunne, A., Virwani, K., *et al.* Evidence of Crystallization–Induced Segregation in the Phase Change Material Te-Rich GST. *Journal of The Electrochemical Society* **158**, H965 (2011).
12. Guerin, S., Hayden, B., *et al.* Synthesis and Screening of Phase Change Chalcogenide Thin Film Materials for Data Storage. *ACS Combinatorial Science* **19**, 478–491 (2017).
13. Shen, Y., Harris, N. C., *et al.* Deep learning with coherent nanophotonic circuits. *Nature Photonics* **11** (2017).
14. Tait, A. N., Lima, T. F. D., *et al.* Neuromorphic photonic networks using silicon photonic weight banks. *Scientific Reports* (2017).
15. Feldmann, J., Youngblood, N., *et al.* Parallel convolutional processing using an integrated photonic tensor core. *Nature* **589**, 52–58 (2021).
16. Mercante, A. J., Yao, P., *et al.* 110 GHz CMOS compatible thin film LiNbO<sub>3</sub> modulator on silicon. *Optics Express, Vol. 24, Issue 14, pp. 15590-15595* **24**, 15590–15595 (July 2016).
17. Yan, Z., Han, Y., *et al.* A monolithic InP/SOI platform for integrated photonics. *Light: Science & Applications* 2021 10:1 **10**, 1–10 (Sept. 2021).
18. Xing, P., Ma, D., *et al.* CMOS-Compatible PECVD Silicon Carbide Platform for Linear and Nonlinear Optics. *ACS Photonics* **6**, 1162–1167 (May 2019).

19. Youngblood, N., Ríos Ocampo, C. A., *et al.* Integrated optical memristors. *Nature Photonics* **17**, 561–572 (July 2023).
20. Lian, C., Vagionas, C., *et al.* Photonic (computational) memories: tunable nanophotonics for data storage and computing. *Nanophotonics* **11**, 3823–3854 (2022).
21. Soref, R. & Bennett, B. Electrooptical effects in silicon. *IEEE Journal of Quantum Electronics* **23**, 123–129 (1987).
22. Abel, S., Eltes, F., *et al.* Large Pockels effect in micro- and nanostructured barium titanate integrated on silicon. *Nature Materials* **18**, 42–47 (2019).
23. Barrios, C. A. & Lipson, M. Silicon photonic read-only memory. *Journal of Lightwave Technology* **24**, 2898–2905 (2006).
24. Song, J.-F., Luo, X.-S., *et al.* Integrated photonics with programmable non-volatile memory. *Scientific Reports* **6**, 22616 (Mar. 2016).
25. Olivares, I., Parra, J., *et al.* Non-Volatile Photonic Memory Based on a SAHAS Configuration. *IEEE Photonics Journal* **13**, 1–8 (2021).
26. Parra, J., Olivares, I., *et al.* Non-volatile epsilon-near-zero readout memory. *Optics Letters* **44**, 3932 (Aug. 2019).
27. Pitris, S., Vagionas, C., *et al.* WDM-Enabled Optical RAM at 5 Gb/s Using a Monolithic InP Flip-Flop Chip. *IEEE Photonics Journal* **8** (Apr. 2016).
28. Turner, E. H. HIGH-FREQUENCY ELECTRO-OPTIC COEFFICIENTS OF LITHIUM NIOBATE. *Applied Physics Letters* **8**, 303–304 (June 1966).
29. Jin, M., Wei, Z., *et al.* Silicon-Based Graphene Electro-Optical Modulators. *Photonics* **9** (2022).
30. Li, Y., Ping, H., *et al.* Nonvolatile silicon photonic switch with graphene based flash-memory cell. *Opt. Mater. Express* **11**, 766–773 (Mar. 2021).
31. Pleros, N., Apostolopoulos, D., *et al.* Optical static RAM cell. *IEEE Photonics Technology Letters* **21**, 73–75 (2009).
32. Bakopoulos, P., Vyrsoinos, K., *et al.* All-optical T-flip-flop using a single SOA-MZI-based latching element. *IEEE Photonics Technology Letters* **24**, 748–750 (2012).
33. Vagionas, C., Fitsios, D., *et al.* Optical RAM and flip-flops using bit-input wavelength diversity and SOA-XGM switches. *Journal of Lightwave Technology* **30**, 3003–3009 (2012).
34. Clavero, R., Ramos, F., *et al.* All-optical flip-flop based on a single SOA-MZI. *IEEE Photonics Technology Letters* **17**, 843–845 (Apr. 2005).
35. Khan, U., Quack, N., *et al.* Silicon photonic microelectromechanical phase shifters for scalable programmable photonics. *Optics Letters*, Vol. 46, Issue 22, pp. 5671–5674 **46**, 5671–5674 (Nov. 2021).
36. Bagheri, M., Poot, M., *et al.* Dynamic manipulation of nanomechanical resonators in the high-amplitude regime and non-volatile mechanical memory operation. *Nature Nanotechnology* 2011 6:11 **6**, 726–732 (Oct. 2011).

37. Wu, M. C., Seok, T. J., *et al.* Large-scale broadband digital silicon photonic switches with vertical adiabatic couplers. *Optica*, Vol. 3, Issue 1, pp. 64–70 **3**, 64–70 (Jan. 2016).
38. Errando-Herranz, C., Takabayashi, A. Y., *et al.* MEMS for Photonic Integrated Circuits. *IEEE Journal of Selected Topics in Quantum Electronics* **26** (Mar. 2020).
39. Tait, A. N., Nahmias, M. A., *et al.* Broadcast and Weight: An Integrated Network For Scalable Photonic Spike Processing. *Journal of Lightwave Technology* **32**, 4029–4041 (2014).
40. Tait, A. N., Chang, J., *et al.* Demonstration of WDM weighted addition for principal component analysis. *Optics Express* (2015).
41. Tait, A. N., de Lima, T. F., *et al.* Multi-channel control for microring weight banks. *Opt. Express* **24**, 8895–8906 (Apr. 2016).
42. Yamada, N. & Matsunaga, T. Structure of laser-crystallized Ge<sub>2</sub>Sb<sub>2+x</sub>Te<sub>5</sub> sputtered thin films for use in optical memory. *Journal of Applied Physics* **88**, 7020–7028 (Dec. 2000).
43. Raoux, S., Xiong, F., *et al.* Phase change materials and phase change memory. *MRS Bulletin* **39**, 703–710 (Aug. 2014).
44. Cheng, H. Y., Brightsky, M., *et al.* Atomic-level engineering of phase change material for novel fast-switching and high-endurance PCM for storage class memory application. *Technical Digest - International Electron Devices Meeting, IEDM* (2013).
45. Kooi, B. J. & Wuttig, M. Chalcogenides by Design: Functionality through Metavalent Bonding and Confinement. *Advanced Materials* **32**, 1908302. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/adma.201908302> (2020).
46. Maria, I., Arora, R., *et al.* Metavalent Bonding-Mediated Dual 6s<sup>2</sup> Lone Pair Expression Leads to Intrinsic Lattice Shearing in n-Type TlBiSe<sub>2</sub>. *Journal of the American Chemical Society* **145**. PMID: 37042625, 9292–9303. eprint: <https://doi.org/10.1021/jacs.3c02146> (2023).
47. Ahmed, S., Wang, X.-D., *et al.* Unraveling the optical contrast in Sb<sub>2</sub>Te and AgInSbTe phase-change materials. *Journal of Physics: Photonics* **3**, 034011 (June 2021).
48. Wuttig, M. & Yamada, N. Phase-change materials for rewriteable data storage. *Nature Materials* **6**, 824–832 (2007).
49. Ovshinsky, S. R. Reversible Electrical Switching Phenomena in Disordered Structures. *Physical Review Letters* **21**, 1450–1453 (Nov. 1968).
50. Guo, P., Sarangan, A., *et al.* A Review of Germanium-Antimony-Telluride Phase Change Materials for Non-Volatile Memories and Optical Modulators. *Applied Sciences* **9**, 530 (Feb. 2019).
51. Rios, C., Hosseini, P., *et al.* On-chip photonic memory elements employing phase-change materials. *Advanced Materials* **26**, 1372–1377 (2014).
52. Delaney, M., Zeimpekis, I., *et al.* A New Family of Ultralow Loss Reversible Phase-Change Materials for Photonic Integrated Circuits: Sb<sub>2</sub>S<sub>3</sub> and Sb<sub>2</sub>Se<sub>3</sub>. *Advanced Functional Materials* **30**, 2002447 (Sept. 2020).

53. Dong, W., Liu, H., *et al.* Wide Bandgap Phase Change Material Tuned Visible Photonics. *Advanced Functional Materials* **29**, 1806181 (Feb. 2019).
54. Ríos, C., Stegmaier, M., *et al.* Controlled switching of phase-change materials by evanescent-field coupling in integrated photonics [Invited]. *Optical Materials Express* **8**, 2455–2470 (2018).
55. Stegmaier, M., Ríos, C., *et al.* Thermo-optical Effect in Phase-Change Nanophotonics. *ACS Photonics* **3**, 828–835 (May 2016).
56. Lima, T. F. D., Tait, A. N., *et al.* Scalable Wideband Principal Component Analysis via Microwave Photonics. *IEEE Photonics Journal* (2016).
57. Xu, X.-Y., Huang, X.-L., *et al.* A Scalable Photonic Computer Solving the Subset Sum Problem. *Science Advances* **6** (Feb. 2020).
58. Nitta, K., Katsuta, N., *et al.* An optical parallel system for prime factorization. *Japanese Journal of Applied Physics* (2009).
59. Feldmann, J., Youngblood, N., *et al.* All-optical spiking neurosynaptic networks with self-learning capabilities. *Nature* **569**, 208–214 (2019).
60. Chakraborty, I., Saha, G., *et al.* Toward Fast Neural Computing using All-Photonic Phase Change Spiking Neurons. *Scientific Reports* (2018).
61. Robertson, J., Hejda, M., *et al.* Ultrafast optical integration and pattern classification for neuromorphic photonics based on spiking VCSEL neurons. *Scientific Reports* (2020).
62. Tan, J. Y. S., Cheng, Z., *et al.* Monadic Pavlovian associative learning in a backpropagation-free photonic network. *Optica* **9**, 792 (July 2022).
63. Liang, Y.-Z. & Liu, H. K. Optical matrix–matrix multiplication method demonstrated by the use of a multifocus hololens. *Optics Letters* (1984).
64. Farhat, N. H., Psaltis, D., *et al.* Optical implementation of the Hopfield model. *Applied Optics* (1985).
65. Reck, M., Zeilinger, A., *et al.* Experimental realization of any discrete unitary operator. *Physical Review Letters* (1994).
66. Xu, S., Wang, J., *et al.* Optical coherent dot-product chip for sophisticated deep learning regression. *Light: Science & Applications* 2021 10:1 **10**, 1–12 (Nov. 2021).
67. Zhu, H. H., Zou, J., *et al.* Space-efficient optical computing with an integrated chip diffractive neural network. *Nature Communications* 2022 13:1 **13**, 1–9 (Feb. 2022).
68. Feng, C., Gu, J., *et al.* A Compact Butterfly-Style Silicon Photonic–Electronic Neural Chip for Hardware-Efficient Deep Learning. *ACS Photonics* **9**, 3906–3916 (Dec. 2022).
69. Ríos, C., Youngblood, N., *et al.* In-memory computing on a photonic platform. *Science Advances* **5**, eaau5759 (Sept. 2022).
70. Li, X., Youngblood, N., *et al.* On-chip Phase Change Optical Matrix Multiplication Core in 2020 IEEE International Electron Devices Meeting (IEDM) **2020-Decem** (Institute of Electrical and Electronics Engineers Inc., Dec. 2020), 7.5.1–7.5.4.
71. Zhou, W., Dong, B., *et al.* In-memory photonic dot-product engine with electrically programmable weight banks. *Nature Communications* **14**, 2887 (May 2023).



72. Yang, L., Zhang, L., *et al.* On-chip optical matrix-vector multiplier in *Optics and Photonics for Information Processing VII* (2013).
73. Wu, T., Menarini, M., *et al.* Lithography-free reconfigurable integrated photonic processor. *Nature Photonics*, 1–7 (Apr. 2023).
74. Teğın, U., Yildırım, M., *et al.* Scalable optical learning operator. *Nature Computational Science* 2021 1:8 **1**, 542–549 (Aug. 2021).
75. Ashtiani, F., Geers, A. J., *et al.* An on-chip photonic deep neural network for image classification. **606**, 501–506 (June 2022).
76. Xu, X., Tan, M., *et al.* 11 TOPS photonic convolutional accelerator for optical neural networks. *Nature* **589**, 44–51 (Jan. 2021).
77. *Integrated photonics platforms compared: SiN, InP and SiPh*
78. Li, C., Guo, P., *et al.* Review of Silicon Photonics Technology and Platform Development. *Journal of Lightwave Technology*, Vol. 39, Issue 13, pp. 4374–4389 **39**, 4374–4389 (July 2021).
79. Dumon, P., Bogaerts, W., *et al.* Low-loss SOI photonic wires and ring resonators fabricated with deep UV lithography. *IEEE Photonics Technology Letters* **16**, 1328–1330 (May 2004).
80. Cardenas, J., Poitras, C. B., *et al.* Low loss etchless silicon photonic waveguides. *Optics express* **17**, 4752–7 (Mar. 2009).
81. Fukazawa, T., Ohno, F., *et al.* Very compact arrayed-waveguide-grating demultiplexer using Si photonic wire waveguides. *Japanese Journal of Applied Physics, Part 2: Letters* **43**, L673 (May 2004).
82. Liu, S., Hu, H., *et al.* Low-crosstalk silicon-photonics arrayed waveguide gratings integrated with micro-ring filter. *Frontiers in Physics* **10**, 1000 (Oct. 2022).
83. Dong, P. Silicon Photonic Integrated Circuits for Wavelength-Division Multiplexing Applications. *IEEE Journal of Selected Topics in Quantum Electronics* **22**, 370–378 (Nov. 2016).
84. Akiyama, S. & Usuki, T. High-speed and efficient silicon modulator based on forward-biased pin diodes. *Frontiers in Physics* **2**, 1–7 (Nov. 2014).
85. Rahim, A., Hermans, A., *et al.* Taking silicon photonics modulators to a higher performance level: state-of-the-art and a review of new technologies. <https://doi.org/10.1117/1.AP.3.2.024003> **3**, 024003 (Apr. 2021).
86. Lischke, S., Peczek, A., *et al.* Ultra-fast germanium photodiode with 3-dB bandwidth of 265-GHz. *Nature Photonics* **15**, 925–931 (Dec. 2021).
87. Splett, A., Zinke, T., *et al.* Integration of waveguides and photodetectors in SiGe for 1.3  $\mu\text{m}$  operation. *IEEE Photonics Technology Letters* **6**, 59–61 (Jan. 1994).
88. Srinivasan, S. A., Pantouvaki, M., *et al.* 56 Gb/s Germanium Waveguide Electro-Absorption Modulator. *Journal of Lightwave Technology* **34**, 419–424 (Jan. 2016).
89. Verheyen, P., Srinivasan, S. A., *et al.* Active Components for 50 Gb/s NRZ-OOK Optical Interconnects in a Silicon Photonics Platform. *Journal of Lightwave Technology*, Vol. 35, Issue 4, pp. 631–638 **35**, 631–638 (Feb. 2017).

90. Zhao, H., Pinna, S., *et al.* High-Power Indium Phosphide Photonic Integrated Circuits. *IEEE Journal of Selected Topics in Quantum Electronics* **25** (Nov. 2019).
91. Coldren, L. A., Nicholes, S. C., *et al.* High performance InP-based photonic Ics - A tutorial. *Journal of Lightwave Technology* **29**, 554–570 (2011).
92. Tohmori, Y., Suematsu, Y., *et al.* Wavelength tuning of GaInAsP/InP integrated laser with butt-jointed built-in distributed Bragg reflector. *Electronics Letters* **19**, 656–657 (Aug. 1983).
93. Suzuki, M., Noda, Y., *et al.* Monolithic Integration of InGaAsP/InP Distributed Feedback Laser and Electroabsorption Modulator by Vapor Phase Epitaxy. *Journal of Lightwave Technology* **5**, 1277–1285 (1987).
94. D, R. J., Kapon, E., *et al.* Low-Loss III-V Semiconductor Optical Waveguides. *IEEE JOURNAL OF QUANTUM ELECTRONICS* **21** (1991).
95. Zirngibl, M., Dragone, C., *et al.* Demonstration of a  $15 \times 15$  Arrayed Waveguide Multiplexer on InP. *IEEE Photonics Technology Letters* **4**, 1250–1253 (1992).
96. Smit, M. & Van Dam, C. PHASAR-based WDM-devices: Principles, design and applications. *IEEE Journal of Selected Topics in Quantum Electronics* **2**, 236–250 (June 1996).
97. Fukano, H., Yamanaka, T., *et al.* Design and Fabrication of Low-Driving-Voltage Electroabsorption Modulators Operating at 40 Gb/s. *Journal of Lightwave Technology* **25**, 1961–1969 (Aug. 2007).
98. Bach, H. G., Beling, A., *et al.* InP-based waveguide-integrated photodetector with 100-GHz bandwidth. *IEEE Journal on Selected Topics in Quantum Electronics* **10**, 668–672 (July 2004).
99. Romero-García, S., Merget, F., *et al.* Silicon nitride CMOS-compatible platform for integrated photonics applications at visible wavelengths. *Optics Express* **21**, 14036 (June 2013).
100. Sharma, T., Wang, J., *et al.* Review of Recent Progress on Silicon Nitride-Based Photonic Integrated Circuits. *IEEE Access* **8**, 195436–195446 (2020).
101. Liu, J., Huang, G., *et al.* High-yield, wafer-scale fabrication of ultralow-loss, dispersion-engineered silicon nitride photonic circuits. *Nature Communications* **12**, 2236 (Apr. 2021).
102. Krückel, C. J., Fülöp, A., *et al.* Linear and nonlinear characterization of low-stress high-confinement silicon-rich nitride waveguides. *Optics Express* **23**, 25827 (Oct. 2015).
103. Wang, L., Chang, L., *et al.* Frequency comb generation in the green using silicon nitride microresonators. *Laser & Photonics Reviews* **10**, 631–638 (July 2016).
104. Kim, S., Han, K., *et al.* Dispersion engineering and frequency comb generation in thin silicon nitride concentric microresonators. *Nature Communications* **2017 8:1** **8**, 1–8 (Aug. 2017).
105. Gao, Y., Tao, L., *et al.* Graphene-on-silicon nitride waveguide photodetector with interdigital contacts. *Applied Physics Letters* **112**, 211107 (May 2018).
106. Alexander, K., George, J. P., *et al.* Nanophotonic Pockels modulators on a silicon nitride platform. *Nature Communications* **2018 9:1** **9**, 1–6 (Aug. 2018).

107. Pfeiffer, M. H. P., Herkommer, C., *et al.* Photonic damascene process for low-loss, high-confinement silicon nitride waveguides. *IEEE Journal of Selected Topics in Quantum Electronics* **24** (July 2018).
108. Boes, A., Chang, L., *et al.* Lithium niobate photonics: Unlocking the electromagnetic spectrum. *Science* **379** (Jan. 2023).
109. Gui, L., Xu, B., *et al.* Microstructure in lithium niobate by use of focused femtosecond laser pulses. *IEEE Photonics Technology Letters* **16**, 1337–1339 (May 2004).
110. Zhang, M., Wang, C., *et al.* Monolithic ultra-high-Q lithium niobate microring resonator. *Optica* **4**, 1536 (Dec. 2017).
111. Chang, L., Pfeiffer, M. H. P., *et al.* Heterogeneous integration of lithium niobate and silicon nitride waveguides for wafer-scale photonic integrated circuits on silicon. *Optics Letters* **42** (2017).
112. Xu, Q., Wood, M. G., *et al.* Hybrid silicon and lithium niobate electro-optical ring modulator. *Optica*, Vol. 1, Issue 2, pp. 112–118 **1**, 112–118 (Aug. 2014).
113. Desiatov, B., Shams-Ansari, A., *et al.* Ultra-low-loss integrated visible photonics using thin-film lithium niobate. *Optica* **6** (2019).
114. Wang, C., Zhang, M., *et al.* Nanophotonic lithium niobate electro-optic modulators. *Optics Express* **26** (2018).
115. Wang, C., Zhang, M., *et al.* Integrated lithium niobate electro-optic modulators operating at CMOS-compatible voltages. *Nature* 2018 562:7725 **562**, 101–104 (Sept. 2018).
116. Wang, C., Zhang, M., *et al.* Monolithic lithium niobate photonic circuits for Kerr frequency comb generation and modulation. *Nature Communications* **10** (2019).
117. Zhang, M., Buscaino, B., *et al.* Broadband electro-optic frequency comb generation in a lithium niobate microring resonator. *Nature* 2019 568:7752 **568**, 373–377 (Mar. 2019).
118. Snigirev, V., Riedhauser, A., *et al.* Ultrafast tunable lasers using lithium niobate integrated photonics. *Nature* **615**, 411–417 (Mar. 2023).
119. Dangel, R., La Porta, A., *et al.* Polymer waveguides enabling scalable low-loss adiabatic optical coupling for silicon photonics. *IEEE Journal of Selected Topics in Quantum Electronics* **24** (July 2018).
120. Zuo, H., Hu, J., *et al.* Low loss, flexible single-mode polymer photonics. *Optics Express*, Vol. 27, Issue 8, pp. 11152–11159 **27**, 11152–11159 (Apr. 2019).
121. Hong Ma, B., K-Y Jen, A., *et al.* Polymer-Based Optical Waveguides: Materials, Processing, and Devices\*\*.
122. Krchnavek, R. R., Lalk, G. R., *et al.* Laser direct writing of channel waveguides using spin-on polymers. *Journal of Applied Physics* **66**, 5156–5160 (Dec. 1989).
123. Klein, S., Barsella, A., *et al.* One-step waveguide and optical circuit writing in photopolymerizable materials processed by two-photon absorption. *Applied Physics Letters* **86**, 1–3 (May 2005).
124. Rezem, M., Gunther, A., *et al.* Low-cost fabrication of all-polymer components for integrated photonics. *Journal of Lightwave Technology* **35**, 299–308 (Jan. 2017).

125. Rabiei, P., Steier, W. H., *et al.* Polymer micro-ring filters and modulators. *Journal of Lightwave Technology* **20**, 1968–1975 (Nov. 2002).
126. Keil, N., Yao, H. H., *et al.* Athermal polarisation-independent arrayed-waveguide grating (AWG) multiplexer using an all-polymer approach. *Applied Physics B: Lasers and Optics* **73**, 619–622 (Apr. 2001).
127. Saleh, B. E. A. & Teich, M. C. *Fundamentals of Photonics*, 2nd Edition 2007.
128. Liu, J.-m. in *Photonic Devices* 73–118 (Cambridge University Press, Apr. 2005).
129. Marchetti, R., Lacava, C., *et al.* Coupling strategies for silicon photonics integrated chips [Invited]. *Photonics Research* **7** (2019).
130. Cheng, L., Mao, S., *et al.* *Grating couplers on silicon photonics: Design principles, emerging trends and practical issues* 2020.
131. Bozzola, A., Carroll, L., *et al.* Optimising apodized grating couplers in a pure SOI platform to 05 dB coupling efficiency. *Optics Express* **23**, 16289 (June 2015).
132. Rabus, D. G. Ring resonators: Theory and modeling. *Springer Series in Optical Sciences* **127**, 3–40 (2007).
133. Bogaerts, W., de Heyn, P., *et al.* Silicon microring resonators. *Laser & Photonics Reviews* **6**, 47–73 (Jan. 2012).
134. Barclay, P. E., Srinivasan, K., *et al.* Nonlinear response of silicon photonic crystal microresonators excited via an integrated waveguide and fiber taper. *Optics Express*, Vol. 13, Issue 3, pp. 801–820 **13**, 801–820 (Feb. 2005).
135. Aso, O., Tadakuma, M., *et al.* Four-Wave Mixing in Optical Fibers and Its Applications (2000).
136. Agrawal, G. P. ( P. Nonlinear fiber optics, 592.
137. Hryniewicz, J. V., Wilson, R. A., *et al.* Wavelength conversion in GaAs micro-ring resonators. *Optics Letters*, Vol. 25, Issue 8, pp. 554–556 **25**, 554–556 (Apr. 2000).
138. Rukhlenko, I. D., Premaratne, M., *et al.* Effective mode area and its optimization in silicon-nanocrystal waveguides. *Optics Letters* **37** (2012).
139. Xing, P., Ma, D., *et al.* High efficiency four wave mixing and optical bistability in amorphous silicon carbide ring resonators. *APL Photonics* **5**, 76110 (July 2020).
140. Aggarwal, S., Milne, T., *et al.* Antimony as a Programmable Element in Integrated Nanophotonics. *Nano Letters* **22**, 3532–3538 (May 2022).
141. Hauser, J. J. Hopping conductivity in amorphous antimony. *Phys. Rev. B* **9**, 2623–2626 (Mar. 1974).
142. Salinga, M., Kersting, B., *et al.* Monatomic phase change memory. *Nature Materials* **17**, 681–685 (2018).
143. Kersting, B. & Salinga, M. Exploiting nanoscale effects in phase change memories. *Faraday Discussions* **213**, 357–370 (2019).
144. Cheng, Z., Milne, T., *et al.* Antimony thin films demonstrate programmable optical nonlinearity. *Science Advances* **7**, eabd7097 (Jan. 2021).
145. Bostanjoglo, O. & Schlotzhauer, G. Impulse stimulated crystallization of Sb films investigated by time resolved TEM. *physica status solidi (a)* **68**, 555–560 (1981).

146. Wickersham, C. E., Bajor, G., *et al.* Impulse stimulated "explosive" crystallization of sputter deposited amorphous (In,Ga)Sb films. *Solid State Communications* **27**, 17–20 (1978).
147. Hu, Y., Qiu, Q., *et al.* Ultrafast crystallization in nanoscale phase change film of monobasic antimony. *Applied Surface Science* **505**, 144337 (Mar. 2020).
148. Wu, C., Yu, H., *et al.* Programmable phase-change metasurfaces on waveguides for multimode photonic convolutional neural network. *Nature Communications* **12**, 96 (2021).
149. Rozenberg, M. J., Schneegans, O., *et al.* An ultra-compact leaky-integrate-and-fire model for building spiking neural networks. *Scientific Reports* **9**, 11123 (2019).
150. Dutta, S., Kumar, V., *et al.* Leaky Integrate and Fire Neuron by Charge-Discharge Dynamics in Floating-Body MOSFET. *Scientific Reports* **7**, 8257 (2017).
151. Sarwat, S. G., Kersting, B., *et al.* *Phase Change Memtransistive Synapse* May 2021.
152. Youngblood, N., Ríos, C., *et al.* Tunable Volatility of Ge<sub>2</sub>Sb<sub>2</sub>Te<sub>5</sub> in Integrated Photonics. *Advanced Functional Materials* **29**, 1807571 (Mar. 2019).
153. Capmany, J., Ortega, B., *et al.* A tutorial on microwave photonic filters. *Journal of Lightwave Technology* **24**, 201–229 (2006).
154. Norberg, E. J., Guzzon, R. S., *et al.* Programmable photonic microwave filters monolithically integrated in InP/InGaAsP. *J. Lightw. Technol.* **29**, 1611–1619 (2011).
155. Capmany, J., Gasulla, I., *et al.* Microwave photonics: the programmable processor. *Nat. Photon.* **10**, 6–8 (Jan. 2016).
156. Fandiño, J. S., Muñoz, P., *et al.* A monolithic integrated photonic microwave filter. *Nature Photonics* 2016 11:2 **11**, 124–129 (Dec. 2016).
157. Fandiño, J. S. & Muñoz, P. Photonics-based microwave frequency measurement using a double-sideband suppressed-carrier modulation and an InP integrated ring-assisted Mach–Zehnder interferometer filter. *Opt. Lett.* **38**, 4316–4319 (Nov. 2013).
158. Guan, B., Djordjevic, S. S., *et al.* CMOS compatible reconfigurable silicon photonic lattice filters using cascaded unit cells for RF-photonic processing. *IEEE J. Sel. Top. Quantum Electron.* **20**, 359–368 (July 2014).
159. Khilo, A., Spector, S. J., *et al.* Photonic ADC: overcoming the bottleneck of electronic jitter. *Opt. Express* **20**, 4454–4469 (Feb. 2012).
160. Guzzon, R. S., Norberg, E. J., *et al.* Integrated InP–InGaAsP tuneable coupled ring optical bandpass filters with zero insertion loss. *Opt. Express* **19**, 7816–7826 (Apr. 2011).
161. Marpaung, D., Morrison, B., *et al.* Si<sub>3</sub>N<sub>4</sub> ring resonator-based microwave photonic notch filter with an ultrahigh peak rejection. *Opt. Express* **21**, 23286–23294 (Oct. 2013).
162. Feldmann, J., Stegmaier, M., *et al.* Calculating with light using a chip-scale all-optical abacus. *Nature Communications* **8**, 1256 (2017).

163. Wright, C. D., Hosseini, P., *et al.* Beyond von-Neumann Computing with Nanoscale Phase-Change Memory Devices. *Advanced Functional Materials* **23**, 2248–2254 (2013).
164. Hosseini, P., Sebastian, A., *et al.* Accumulation-Based Computing Using Phase-Change Memories With FET Access Devices. *IEEE Electron Device Letters* **36**, 975–977 (2015).
165. Aggarwal, S., Dong, B., *et al.* Reduced rank photonic computing accelerator. *Optica* **10**, 1074 (Aug. 2023).
166. Abdel-Hamid, O., Mohamed, A.-r., *et al.* Convolutional Neural Networks for Speech Recognition. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* **22**, 1533–1545 (2014).
167. He, K., Zhang, X., *et al.* Deep residual learning for image recognition in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* **2016-Decem** (2016), 770–778.
168. Sladojevic, S., Arsenovic, M., *et al.* Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification. *Computational Intelligence and Neuroscience* **2016** (ed Van Hulle, M.) 3289801 (2016).
169. Cireşan, D., Meier, U., *et al.* Multi-column deep neural network for traffic sign classification. *Neural Networks* **32**, 333–338 (2012).
170. Guo, T., Dong, J., *et al.* Simple convolutional neural network on image classification in *2017 IEEE 2nd International Conference on Big Data Analysis, ICBDA 2017* (IEEE, 2017), 721–724.
171. Shin, D. & Yoo, H. J. The Heterogeneous Deep Neural Network Processor with a Non-von Neumann Architecture. *Proceedings of the IEEE* **108** (2019).
172. Sebastian, A., Gallo, M. L., *et al.* Computational phase-change memory: Beyond von Neumann computing. *Journal of Physics D: Applied Physics* **52** (2019).
173. Yao, P., Wu, H., *et al.* Fully hardware-implemented memristor convolutional neural network. *Nature* **577**, 641–646 (2020).
174. Zhou, H., Dong, J., *et al.* Photonic matrix multiplication lights up photonic accelerator and beyond Feb. 2022.
175. Wetzstein, G., Ozcan, A., *et al.* Inference in artificial intelligence with deep optics and photonics. *Nature* **588**, 39–47 (Dec. 2020).
176. Gan, J., Wang, W., *et al.* Compressing the CNN architecture for in-air handwritten Chinese character recognition. *Pattern Recognition Letters* **129**, 190–197 (Jan. 2020).
177. Kozyrskiy, N. & Phan, A.-H. *CNN Acceleration by Low-rank Approximation with Quantized Factors* 2020.
178. Luo, J. H., Wu, J., *et al.* ThiNet: A Filter Level Pruning Method for Deep Neural Network Compression in *Proceedings of the IEEE International Conference on Computer Vision* **2017-Octob** (2017), 5068–5076.
179. Jaderberg, M., Vedaldi, A., *et al.* Speeding up Convolutional Neural Networks with Low Rank Expansions. *ArXiv* **abs/1405.3** (2014).

180. Pai, S., Solgaard, O., *et al.* Scalable and self-correcting photonic computation using balanced photonic binary tree cascades. *arXiv preprint arXiv:2210.16935* (2022).
181. Banerjee, S., Nikdast, M., *et al.* Pruning Coherent Integrated Photonic Neural Networks. *IEEE Journal of Selected Topics in Quantum Electronics* **29**, 1–13 (2023).
182. Gu, J., Zhao, Z., *et al.* Towards Area-Efficient Optical Neural Networks: An FFT-based Architecture in 2020 25th Asia and South Pacific Design Automation Conference (ASP-DAC) (IEEE, Jan. 2020), 476–481.
183. Milanizadeh, M., Toso, F., *et al.* Coherent self-control of free-space optical beams with integrated silicon photonic meshes. **9**, 2196–2204 (Nov. 2021).
184. Gong, Z., Yang, F., *et al.* Phase change materials in photonic devices. *Journal of Applied Physics* **129**, 30902 (Jan. 2021).
185. Xiang, C., Jin, W., *et al.* Narrow-linewidth III-V/Si/Si<sub>3</sub>N<sub>4</sub> laser using multilayer heterogeneous integration. *Optica* **7**, 20–21 (2020).
186. Xiang, C., Liu, J., *et al.* Laser soliton microcombs heterogeneously integrated on silicon. *Science* **373**, 99–103 (July 2021).
187. Eid, N., Boeck, R., *et al.* FSR-free silicon-on-insulator microring resonator based filter with bent contra-directional couplers. *Optics Express* **24**, 29009 (Dec. 2016).
188. Moazeni, S., Lin, S., *et al.* A 40-Gb/s PAM-4 Transmitter Based on a Ring-Resonator Optical DAC in 45-nm SOI CMOS. *IEEE Journal of Solid-State Circuits* **52**, 3503–3516 (2017).
189. Saeedi, S. & Emami, A. A 25Gb/s 170uW/Gb/s optical receiver in 28nm CMOS for chip-to-chip optical communication in 2014 IEEE Radio Frequency Integrated Circuits Symposium (IEEE, June 2014), 283–286.
190. *Alphacore / Robust High-Performance Microelectronics*
191. Stegmaier, M., Ríos, C., *et al.* Nonvolatile All-Optical 1 × 2 Switch for Chipscale Photonic Networks. *Advanced Optical Materials* **5**, 1600346 (Jan. 2017).
192. Parra, J., Hurtado, J., *et al.* Ultra-low loss hybrid ITO/Si thermo-optic phase shifter with optimized power consumption. *Optics Express* **28**, 9393–9404 (2020).
193. Ding, C., Li, T., *et al.* Convex and Semi-Nonnegative Matrix Factorizations. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **32**, 45–55 (2010).
194. Aggarwal, S., Dong, B., *et al.* Design for Robust and Efficient Neuromorphic Photonic Accelerator. 2022 IEEE International Conference on Emerging Electronics, ICEE 2022 (2022).
195. Morino, H., Maruyama, T., *et al.* Reduction of Wavelength Dependence of Coupling Characteristics Using Si Optical Waveguide Curved Directional Coupler. *Journal of Lightwave Technology* **32**, 2188–2192 (2014).
196. Xu, Q., Schmidt, B., *et al.* Cascaded silicon micro-ring modulators for WDM optical interconnection. *Optics Express* **14**, 9431 (Oct. 2006).
197. Katumba, A., Heyvaert, J., *et al.* Low-Loss Photonic Reservoir Computing with Multimode Photonic Integrated Circuits. *Scientific Reports* **8**, 2653 (2018).

198. Henry, W. M. & Love, J. D. Asymmetric multimode Y-junction splitters. *Optical and Quantum Electronics* **29**, 379–392 (1997).
199. Lopez-Rodriguez, B., van der Kolk, R., *et al.* High-Quality Amorphous Silicon Carbide for Hybrid Photonic Integration Deposited at a Low Temperature. *ACS Photonics* **10**, 3748–3754 (2023).
200. Wang, C., Fang, Z., *et al.* High-Q microresonators on 4H-silicon-carbide-on-insulator platform for nonlinear photonics. *Light: Science & Applications* 2021 10:1 **10**, 1–11 (July 2021).
201. Fiorentino, G., Syed, W., *et al.* Artificial Dielectric Layer Based on PECVD Silicon Carbide for Terahertz Sensing Applications. *Procedia Engineering* **87**, 1497–1500 (Jan. 2014).
202. Wang, S., Zhan, M., *et al.* 4H-SiC: a new nonlinear material for midinfrared lasers. *Laser & Photonics Reviews* **7**, 831–838 (Sept. 2013).
203. Wei, R., Song, S., *et al.* Thermal conductivity of 4H-SiC single crystals. *Journal of Applied Physics* **113** (Feb. 2013).
204. Lukin, D. M., Dory, C., *et al.* 4H-silicon-carbide-on-insulator for integrated quantum and nonlinear photonics. *Nature Photonics* **14**, 330–334 (2020).
205. Matthew, D., Ioannis, Z., *et al.* Nonvolatile programmable silicon photonics using an ultralow-loss Sb<sub>2</sub>Se<sub>3</sub> phase change material. *Science Advances* **7**, eabg3500 (Dec. 2021).
206. Calcagno, L., Musumeci, P., *et al.* Crystallisation mechanism of amorphous silicon carbide. *Applied Surface Science* **184**. Proceeding of the European Materials Research Society 2001-Symposium F "Amorphous and crystalline Silicon Carbide: material and applications", 123–127 (2001).
207. Zhou, W., Farmakidis, N., *et al.* Artificial Biphasic Synapses Based on Nonvolatile Phase-Change Photonic Memory Cells. *Physica Status Solidi - Rapid Research Letters* **16** (Sept. 2022).
208. Li, X., Youngblood, N., *et al.* Experimental investigation of silicon and silicon nitride platforms for phase-change photonic in-memory computing. *Optica* **7**, 218–225 (Mar. 2020).
209. Ji, J., Song, X., *et al.* Two-dimensional antimonene single crystals grown by van der Waals epitaxy. *Nature Communications* 2016 7:1 **7**, 1–9 (Nov. 2016).
210. Wiggins, S. M., Bonse, J., *et al.* The influence of wavelength on phase transformations induced by picosecond and femtosecond laser pulses in GeSb thin films. *Journal of Applied Physics* **98** (Dec. 2005).
211. Farmakidis, N., Youngblood, N., *et al.* Electronically Reconfigurable Photonic Switches Incorporating Plasmonic Structures and Phase Change Materials. *Advanced Science* **9**, 2200383 (July 2022).
212. Farmakidis, N., Youngblood, N., *et al.* Plasmonic nanogap enhanced phase-change devices with dual electrical-optical functionality. *Science Advances* **5**, eaaw2687 (Nov. 2019).
213. Riemensberger, J., Kuznetsov, N., *et al.* A photonic integrated continuous-travelling-wave parametric amplifier. *Nature* 2022 612:7938 **612**, 56–61 (Nov. 2022).



214. Farmakidis, N., Yu, H., *et al.* Scalable High-Precision Trimming of Photonic Resonances by Polymer Exposure to Energetic Beams. *Nano Letters* **23**, 4800–4806 (2023).