All-optical
associative learning element

James TAN You Sian
Mansfield College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy

Trinity Term 2021

Dedicated to my parents

Tan Kok Eng (陳國應) and Lew Mui Lian (廖美蓮)

Acknowledgements

I would like to express my gratitude to Prof. Harish Bhaskaran, my DPhil supervisor, for his guidance and encouragement throughout the course of investigations. I have been helped by discussions with Zengguang Cheng, Johannes Feldmann, Prof. Wolfram Pernice, Prof. C.D. Wright; and profited from conversations with fellow members of Bhaskaran laboratory (including Xuan Li, Nathan Youngblood, Utku Emre Ali, Andrew Katumba, Nikolaos Farmakidis, Eugene Jia Hao Soh, Wen Zhou and MengYun Wang), Hazim Hanif and Gratianus Wesley. I am indebted to Paul Pattinson and Radka Chakalova for technical support.

The research described in this thesis was carried out while I received funding from the Yang di-Pertuan Agong scholarship, awarded by the Public Services Department of the Government of Malaysia. This research was supported via the Engineering and Physical Sciences Research Council Manufacturing Fellowship EP/J018694/1, the Wearable and Flexible Technologies (WAFT) collaboration (EP/M015173/1), the Chalcogenide Advanced Manufacturing Partnership (EP/M015130/1), and the EU-funded Fun-COMP project (#780848). To these bodies, I render my thanks.

Abstract

This thesis examines the idea that Pavlovian associative learning may be a building block in neural networks. With emphasis on hardware acceleration, on-chip optical associative learning device is conceived as an alternative to artificial neural networks (ANNs) hardware for high-speed applications. It is well known that 'conventional' ANNs, particularly in the form of modern deep neural networks (DNNs), are usually carried out using the backpropagation method. In the method, network weights are updated such that the net output better resembles the desired output after each forward-backward iteration, until convergence is met. A distinct approach is presented using associative learning as the basis of AI learning process. In formulating the architecture, associative learning is linked with supervised learning, based on their common goal of associating certain inputs with 'correct' outputs. The intuition leads to a reductionist framework to the problem, facilitating the transition from a single (or monadic) Pavlovian single *Input-Teacher* association to any arbitrary *n Input-Teacher* associations.

The hardware associative learning network is realised using silicon-based on-chip optical waveguides. Optical implementation of the learning network enables simultaneous parallel calculations using wavelength division multiplexing (WDM), which increases the capacity of AI information processing. This is adopted in the work described in this thesis. Monolithic integration of waveguide components, through which optical signals are channeled, augurs well for its adoption as artificial intelligence (AI) hardware accelerator. Machine learning using associative learning hardware network is demonstrated, with the network solving pattern and image recognition tasks. The patterns are randomised optical signals binary pattern, while the images are the discretised 72x72 patterns in the form of image pixels. Computational density of 118 TOPS/mm² is demonstrated, limited only by the available setup in the laboratory. The density is estimated to reach approximately 2.5×10^4 TOPS/mm², with 50 GHz modulation using 16 wavelengths. In solving the image recognition task, the hardware network is shown to be able to create a predictive model capable of generalising a set of patterns, instead of recognising only one particular pattern.

The optical associative learning device paves the way for further realisations of photonic intelligent systems that can infer both symbolically and numerically via associations. Further improvements in other relevant metrics (for example, volume and learning energy) are anticipated on different material platform and/or with other optimisation methods. These will be reserved for a future time when they shall have been more completely worked out. The device realisation potentially opens up new avenues of research in machine learning architectures and algorithms.

Abbreviations

AC	Alternating current
ADALINE	Adaptive Linear Neuron
ADC	Analogue-to-digital converter
AFG	Arbitrary function generator
AI	Analogue input
AI	Artificial intelligence
ALU	Arithmetic logic unit
AMLE	Associative Monadic Learning Element
ANN	Artificial neural network
ASIC	Application-specific integrated circuit
AO	Analogue output
APC	Angled physical contact
BNC	Bayonet Neill–Concelman
CAD	Computer aided design
CAM	Content addressable memory
CCD	Charge-coupled device
CMOS	Complementary metal-oxide semiconductor
CPU	Central processing unit
CW	Continuous wave
DAC	Digital-to-analogue converter
DAQ	Data acquisition device
DC	Direct Current
DFA	Direct Feedback Alignment
DI	Digital input
DiC	Directional coupler
DNN	Deep neural network
DO	Digital output
DRTP	Direct Random Target Projection
DYSTAL	Dynamically Stable Associative Learning
EDFA	Erbium doped fibre amplifier
ENN	Evolutionary neural network
EOM	Electro-optic modulator
FC	Ceramic Ferrule
FDTD	Finite-difference time-domain
FPGA	Field programmable gate arrays
FWHM	Full width at half maximum
GA	Genetic algorithm
GD	Gradient descent
GPIB	General purpose interface bus
GPU	Graphics processing unit
GST	$Ge_2Sb_2Te_5$
IDE	Integrated development environment
I/O	Input/output
LPCVD	Low pressure chemical vapour deposition
MADALINE	Many ADALINE
MLP	Multilayer perceptron
MMCX	Micro-miniature coaxial
MMI	Multimode interferometer

MRR	Microring resonator
MZI	Mach-Zehnder interferometer
MZM	Mach-Zehnder modulator
NRZ	Non-return-to-zero
OTF	Optical tunable filter
PC	Physical contact
PC	Polarisation controller
PCB	Printed circuit board
PCM	Phase change memory
PD	Photodetector
PECVD	Plasma-enhanced chemical vapour deposition
PFI	Programmable function input
PhC	Photonic crystal
PIC	Peripheral interface controller
PRBS	Pseudorandom binary sequence
PXI	Peripheral component interconnect extensions for instrumentation
Q	Quality factor
RAM	Random-access memory
RC	Resistance-capacitance
RIE	Reactive-ion etching
SCPI	Standard commands or programmable instruments
SEM	Scanning electron microscope
SF	Spectrally filtered
SMA	Subminiature version A
STDP	Spike-Timing Dependent Plasticity
TCP/IP	Transmission control protocol/internet protocol
TE	Transverse electric
TIR	Total internal reflection
TM	Transverse magnetic
TOPS	Tera operations per second
TPU	Tensor processing unit
USB	Universal serial bus
USBTMC	Universal serial bus test and measurement class
VOA	Variable optical attenuator
VISA	Virtual instrument software architecture
VXI	Versa module Eurocard extensions or instrumentation
WDM	Wavelength division multiplexing
WGM	Whispering gallery mode

Contents

List of figures

List of tables

Chapter 1

1. Introduction

This chapter introduces the author's research project. It explains the research background, relevant concepts to the research, research contribution, and structure of thesis organisation. The conceptual bases for the investigation are associative learning, and artificial intelligence hardware acceleration.

1.1. Research background

It is said that the human brain - which has strong learning, reasoning and understanding capabilities - is 'the most complex object in the known universe' [1]. There are active ongoing attempts to model the learning system [2-4]. One of the simplest forms of learning mechanism known to us is Pavlovian associative learning [5,6] where presentation of one input signal elicits the recall of another signal upon spatial and/or temporal co-occurrence. In this thesis, hardware devices that demonstrate this using light – on a chip – are introduced. Then, their scaled-up implementations to solve learning tasks are discussed.

1.1.1 What is learning?

One of the most important age-old problems that we still wrestle with is on deciphering the inner workings of our physical selves. This is particularly the case on how the brain exactly works. Today, we commonly attribute learning to the formation and processing of thoughts in the brain [7]. Learning mechanisms in nature have long inspired humans to replicate and build devices seemingly-endowed with intelligence [8,9]. Attempts to build such devices are based on the assumption that intelligence can be mechanised [10-13].

Throughout history, various constructs that give the impression to the casual observer that they are taking actions on their own volition have been built. An example is a certain automaton that follows fixed sequences of instructions. Because these actions are predetermined with no regards to cues from its environment [12-15], such a device is not a learning device. The modern computer, on the other hand, is a learning construct when it is programmed to transform a set of inputs (with real-world concepts) and process them to a set of outputs that respond to the inputs [16,17].

Common in learning constructs are entities with memory abilities, which can be classified into biological, psychological and artificial. Artificial learning entities are inspired by biological or psychological learning models. The different learning entities are mapped in **Fig. 1.1**. The author's research contributions to the broader body of learning entities are bordered green in the figure. These contributions are inspired by concepts related to associative learning. They are Pavlovian learning (known also as classical conditioning) [5], and Hebbian learning [8].

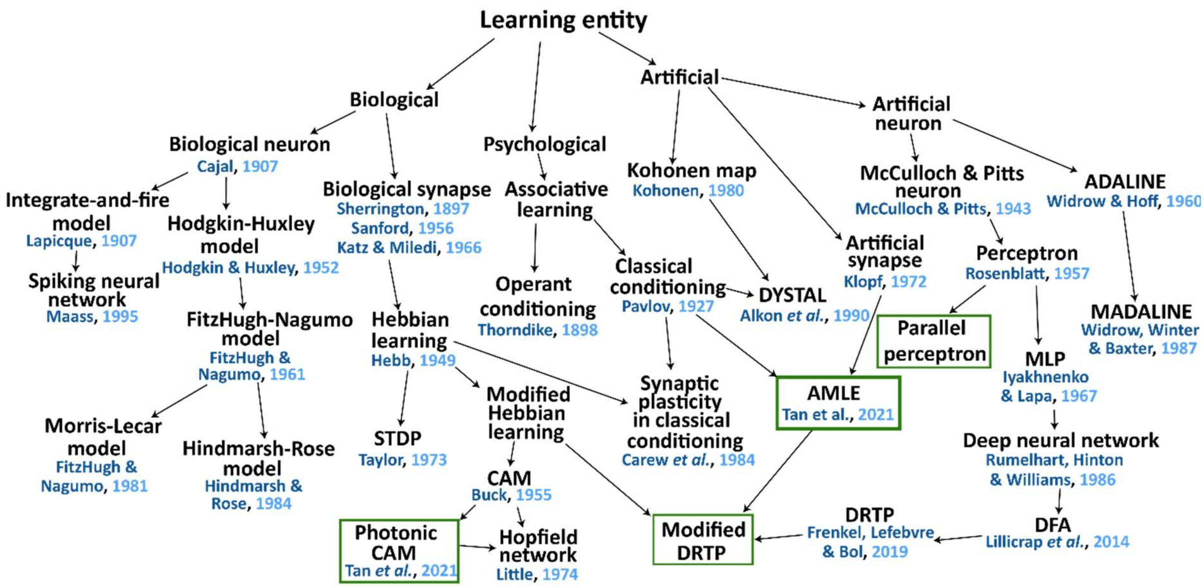

Fig. 1.1. Learning entities. Biological, psychological and artificial learning make up the three main learning entities we know of today. Other learning mechanisms are derived from these entities. For example, the basic concepts of content addressable memory (CAM) can be traced back to the neuroscientific (biological) theory of Hebbian learning, which explains that correlated activation of neural signal inputs lead to strengthening of the input-output connection. Green bordered boxes denote learning entities realised during the author's DPhil studies. Other abbreviated learning entities are Adaptive Linear Neuron (ADALINE) and Many ADALINE (MADALINE), Associative Monadic Learning Element (AMLE), Direct Feedback Alignment (DFA), Direct Random Target Projection (DRTP), Dynamically Stable Associative Learning (DYSTAL), multilayer perceptron (MLP), and Spike-Timing-Dependent Plasticity (STDP).

The neurological work by Cajal on neurons and Sherrington on neuron connections (synapses) forms the basis of most AI systems (see models in **Fig. 1.2**). Cajal proposed that neurons are not continuous but contiguous [18], while Sherrington emphasised that distinct neurons are connected via synapses [19]. Notable neuron models are shown in **Fig. 1.2a**. In the biologically-plausible leaky integrate-and-fire neuron model [20], the output evolves with $e^{-t/\tau}$ after time t when the weighted sum of the inputs x exceeds threshold y_{th}, where τ is the time constant. McCulloch and Pitts simplified the model by replacing the factor $e^{-t/\tau}$ with a constant [21]. Rosenblatt, in his perceptron model, considers weighted inputs in the model [22]. Later in **Chapter 7**, the concept of the parallel perceptron is discussed.

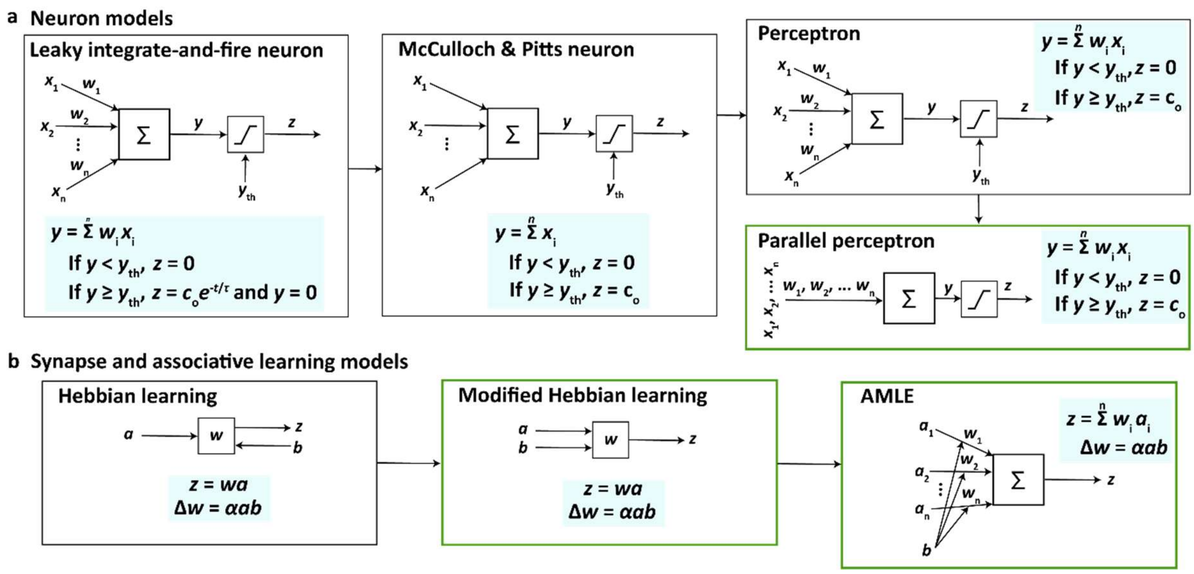

Fig. 1.2. Learning models. Learning models can in general be divided to neuron models [20-22] (**a**) , synapse [7] and associative learning [5] models (**b**). Whereas the output response in leaky integrate-and-fire neuron varies with $e^{-t/\tau}$ (where t and τ are respectively the time elapsed and time constant) after learning, the output response in other neuron models remain at constant c_o after the learning process. Green bordered boxes denote the learning platform introduced during the author's DPhil studies. In particular, the parallel perceptron model is applied to the photonic CAM discussed in **Chapter 7**. The AMLE, on the other hand, is comparable to the concept of Hebbian learning (and CAM – see **Fig. 1.1**) which states that the strengthening of input-output connections is due to the correlated activation of their inputs. In AMLE however, the input b 'supervises' the learning process such that any of the inputs a that is activated at the same time as input b would produce an output response similar to that from b. Although there are some similarities between Hebbian learning and AMLE, it is important to note that AMLE is inspired by the psychological process of Pavlovian associative learning, and not from Hebbian learning.

One synapse model (**Fig. 1.2b**), Hebbian learning – which attributes the increase in input-output connection weight to repeated and persistent stimulation of the inputs - is usually the reference to describe synapses. A modification to this can be made by organising the inputs to the same directions. This is labelled 'modified Hebbian learning' in **Fig. 1.2**. On a hardware level, such input reorganisation ensures unidirectionality. Compared to that involve bidirectional signals, this is more straightforward to manage. For example, on an optical hardware platform, the unidirectionality eliminates the need for optical isolators which is typically applied to manage bidirectional signals. An extension of these Hebbian synapse models is the Associative Learning Monadic Element (AMLE) device (**Fig. 1.2b**), which is inspired by Pavlovian associative learning [5], where the connection weights are modified when both the *Input a* and *Teacher b* are paired, with weighted sum of the inputs as the output.

1.1.2 Artificial intelligence and its physical implementations

The hardware built by the author models learning capabilities of living things in nature. Such ability is known as artificial intelligence (AI). It involves adapting, improvising and generalising from examples to apply in unfamiliar scenarios [23], for example by recognising objects, and responding to speech. For much of early history on AI, learning has been regarded as a mental discourse involving methodical manipulation of symbols [24-26] that can be explicitly framed into a set of *if-else* statements. However, the learning process can be based on approximations too: this is typically applied in human (and animal) problem-solving strategies [27] where carefully choreographed steps are employed to extract information from prior experiences with similar problems. The steps - however rational – are not guaranteed to exactly solve the problem, but can accurately approximate the desired solution. Indeed, in solving many AI tasks, a phletora of approximation tools (especially gradient descent optimisation [28,29]) are widely used today. Unlike their symbolic counterparts, such 'numeric' approaches often provide useful abstract representations of seemingly unapparent features in the input datasets [30,31].

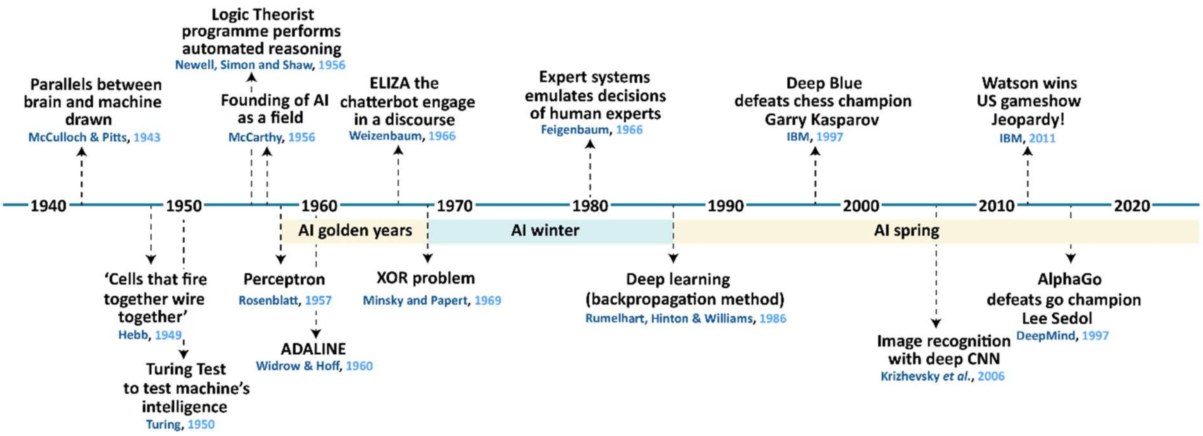

Fig. 1.3. Timeline of AI. Milestones in AI are indicated. The history of AI is marked by a period of increase [22], then decrease [32], and currently renewed [3,30,33] interests, funding, development and deployment of AI. These are respectively known as AI golden years, winter and spring. Abbreviated terms CNN and XOR respectively denote convolutional neural network and eXclusive OR Boolean logic operation.

Important milestones in AI are indicated in **Fig. 1.3**. The McCulloch and Pitts model [21] marks the advent of AI. Conventional AI systems today adopt the perceptron model (extension of McCulloch and Pitts model) in some way or another [30,33]. Minsky and Papert argued that the perceptron could only learn linearly (but not nonlinearly) separable functions [32], which sparked a period of reduced funding and interest in AI (known as 'AI winter'). This was found to be true with single-layer perceptrons, and multilayer perceptrons can address this limitation. Multilayer perceptron was developed by Ivakhnenko and Lapa in 1967 [34] and is commonly adopted in the form of the well-known 'deep learning' today [30].

'Deep learning' has been criticised as not biologically plausible or realistic [35-37]. There is currently a growing field to develop networks that more closely mimic biological neural network which can match the learning efficiency of conventional AI [38], for instance spiking neural network [39,40] – see conceptual map in **Fig. 1.1**. Although this may seem unnecessary given that networks using this model have already found extensive uses with high task accuracy [3,33], it is a justified endeavour considering the fact that the computation power consumption in the brain is about tenfold lower than a computer, and does not require huge number of iterations to solve detection problems [41-44] (compared to current state-of the-art AI systems [2,3]). We may perhaps find an equally or more powerful architecture or algorithm somewhere along the line of emulating the workings of a biological neural network.

1.1.3 Associative learning hardware

The perceptron model (in the form of deep learning) has gained acclaim for its inferencing accuracy when trained with huge amounts of data. It is part of a broader profusion of AI models we have today. Depending on requirements and/or applications, different AI models may be used. For example, the 'Random Forest' AI model can be used when it is important to select the most significant feature in a classification problem [45]. Such interpretability is absent in deep learning. As AI tasks in general are composed of interpenetrating needs and requirements, each of the various AI models with their own respective strengths complement one another. Important events on associative learning are summarised in **Fig. 1.4**.

From an elevated view, although AI models (indicated in **Fig. 1.4**) share the same attribute of associating input features [46-48], they can be categorised based on three attributes: association detection, inductive learning and deductive reasoning (details in **Fig. 1.5**). Association detection hardware look for correlated patterns [49-53] which respond to a set of input signals to associate input patterns that are similar, contrary, in close proximity (spatial) or in close succession (temporal). Inductive learning hardware extracts features to draw inferences. They predict rare events from temporal and sequential patterns of timestamped observations by making generalisations[51]. Hardware with complex generalisation capabilities (subset of inductive learning) infers relationships between seemingly unapparent or unrelated events by developing a simplistic yet realistic account of the set of data presented with no prior knowledge, even in the presence of multiple confounding factors [49,50,52,53]. On the other hand, hardware with deductive reasoning identifies how a system reaches a particular solution, normally by framing the problem in a well-understood form.

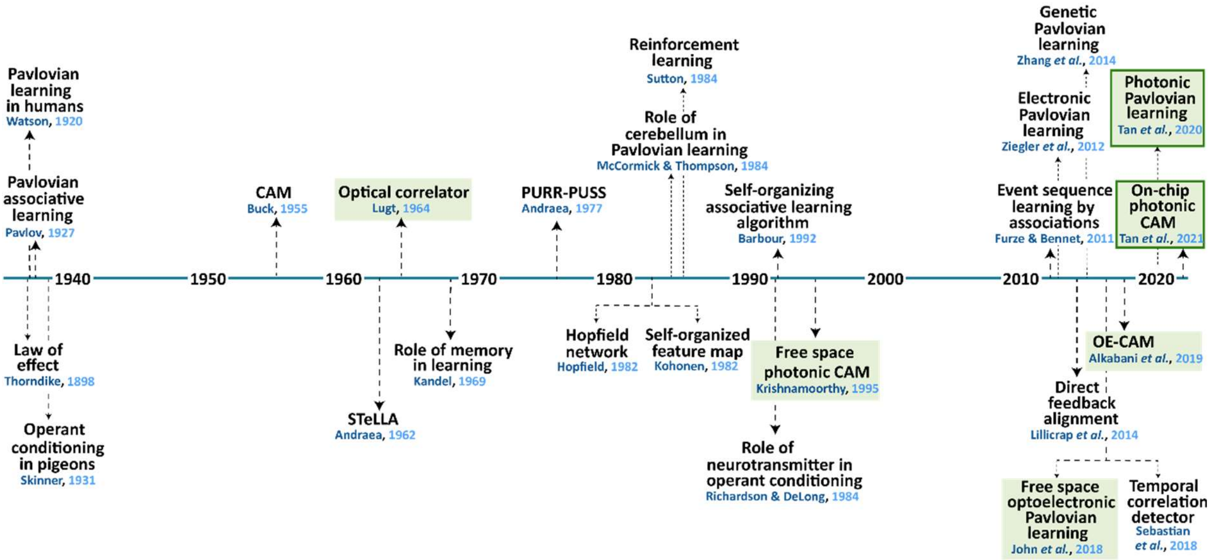

Fig. 1.4. Timeline of associative learning. Major forms of 'associative learning' are: operant and classical conditioning. Operant conditioning – studied in detail by Thorndike [54] and Skinner [55] – associates input with output signal during the learning process. This is in contrast to classical conditioning, in which learning occurs from association of input signals [5,56]. Inspired by these learning mechanisms (in some way or another), various associative learning systems have been realised. Optical or photonic implementations are shaded in grey. Green bordered boxes denote the learning platform introduced during the author's DPhil studies. Abbreviated term OE-CAM denotes opto-electronic content addressable memory.

Practical applications of such hardware include giving insights into the operations of asset operations [57]. With the ability to predict temporal event sequence via associations, the hardware provides predictive and prescriptive capabilities to improve situational performance and reliability. For example, to ensure the smooth running of assets and to increase efficiency in both energy and resource consumption [58], to aid prognosticating equipment failure, to avoid unplanned downtime cost, to identify imminent potential equipment failures, and to better balance supply with anticipated demand levels in industrial plants [59-61].

Today, there is no one single AI system that possess both learning and reasoning attributes. Whereas symbolic systems can explicitly reason how it arrives at the solution and are straightforward, numeric ones can solve problems with complex variations efficiently - but lacks interpretability and typically require many iterations to converge to a reliable solution. It can be prohibitively complex for symbolic systems anticipate all possibilities of complex variations. However, symbolic and numeric approaches need not be mutually exclusive. It is desirable to harness the advantages of both in a single system [62]. Realization for such a still-largely-elusive system (known as 'neurosymbolic AI' [63]) is expected to be highly disruptive.

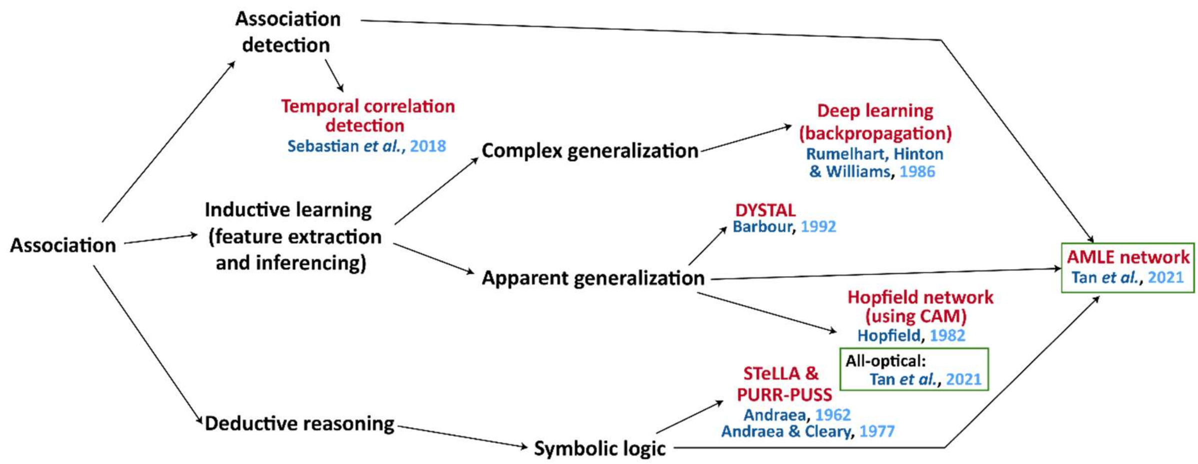

Fig. 1.5. Learning by associations: general domains. Learning mechanisms can be classified based on the ability to detect associations [53,64], inductively learn (complex generalisation, and apparent generalisation [64-67]) and deductively reason [64,68]. Green bordered boxes denote the learning platform introduced during the author's DPhil studies.

The processing of information in AI hardware has traditionally been carried out electronically [49-53,57-61], including hardware that implement associations [22,53,65,67,68]. Optical hardware that implements associations include AMLE network [64] and all-optical on-chip CAM [69]. These are summarised in **Fig. 1.6**. In recent years, optical AI processing has gained considerable interests, given the ability to simultaneously process input data via WDM for higher processing throughput [70-73]. For example, the commercial femtosecond Erbium-doped laser (C-fiber, Menlosystems GmbH) can readily provide 10^5 channels with 100 MHz mode spacing evenly distributed over a 75 nm spectral gain curve [74]. In the frequency comb, the phase-locked loop technique is used to stabilize the mode frequencies [75]. To narrow the mode linewidth of the comb (and thus reduce the possibility of coupling or crosstalk between the optical modes), chirp-profile optimization control is adopted [76,77].

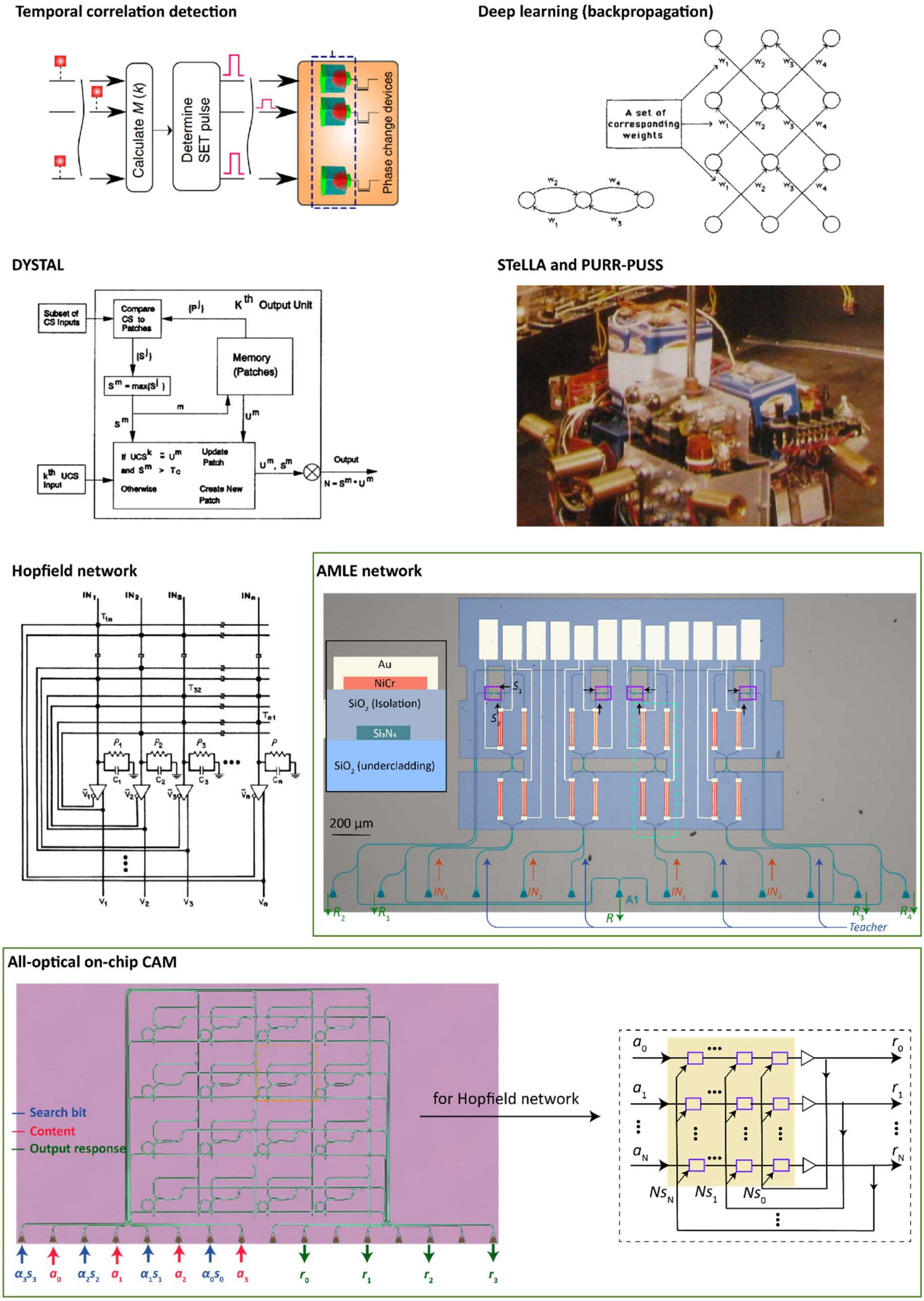

Fig. 1.6. System to implement associations. Temporal correlation detector [53], deep learning (using backpropagation algorithm) [30], DYSTAL [65], STeLLA and PURR-PUSS [68], Hopfield network [67], AMLE

network [64] and all-optical on-chip CAM [69]. Green bordered boxes denote the learning platform introduced during the author's DPhil studies.

1.2. Research design

A new perspective to AI, discussed in this thesis, pictures associative learning as the building block in learning architectures. The learning mechanism is based on the work by Ivan P. Pavlov, who demonstrated the recall of a signal upon temporal co-occurrence with another signal [5]. Inspired by this learning process, a hardware that computes explicit associations is realised. AI applications such as pattern recognition and image recognition are demonstrated using the hardware. The hardware runs on optical signals. The optical properties of phase change materials are suitable for memory element based on silicon optical waveguide. This has led to its adoption for the purpose of realizing the hardware.

1.3 Layout

This thesis is structured as follows:

- **Chapter 2** reviews the development of processing hardware, including the ongoing research on optical computing and AI hardware. It discusses approaches to address the need for high processing speed and throughput.
- **Chapter 3** describes an optical device that implements associative learning based on the concepts of Pavlovian associative learning. The concept is first described. Then, the two complementary on-chip layout designs to measure the device, and the steps to fabricate them are elucidated.
- **Chapter 4** introduces the experimental setups to measure the optical associative learning device. Then, the conception of two complementary on-chip layouts to address the erratic fluctuation of optical phases (due to measurement environment) is described.
- **Chapter 5** shows the experimental results, using the two different measurement setups that applies different methodologies described in **Chapter 4**. The complementary methods allow the author to cross validate the measurement results and extract important information about the learning process.
- **Chapter 6** discusses the scaled-up implementation of the associative learning device based on the similarities between the concept of supervised learning and associative learning. The scaled-up hardware is used to solve pattern and image recognition tasks.
- **Chapter 7** compares associative learning with the 'deep neural network' approach (conventional machine learning), and lays out open questions about associative learning. Then, other associative learning devices are reviewed.

- **Chapter 8** concludes the thesis by summarising key results, determines the current limitations of the device, and projects future directions and implementations to overcome them.

1.4 Main findings

Key results and new science in the work described in this thesis include:

- Experimental realisation of Pavlovian associative learning device
- Experimental schemes to observe photonic associative learning
- Scaled-up associative learning hardware that accelerates AI tasks (pattern and image recognition)
- Mapping of supervised learning concept in AI to Pavlovian associative learning
- Experimental methods to demonstrate very high computational density on associative learning hardware
- Experimental methods to build a model representation of training images for image recognition. Implementation of other related devices in the broader family of associative learning (mutual/two-way associative learning element, and content addressable memory CAM). In particular, a highly-scalable on-chip optical CAM architecture

We have discussed the different classifications of learning entities, and relevant learning models. Background of AI and associative learning are provided to help readers to grasp the big picture perspective of the field. The discussion is then narrowed down to associative learning, in which related works are arranged based on their ability to detect, learn and/or reason from an experiential process. As will be discussed in Chapter 6, the comparable concepts between supervised learning in AI and associative learning provides a more intuitive approach to AI. Finally, the research design, layout and main findings are outlined.

Chapter 2

2. Optical information processing

This chapter reviews the development of processing hardware, including the ongoing research on optical computing and AI hardware. It discusses approaches to address the need for high processing speed and throughput.

2.1 Computing hardware

Various methods have been used to aid computation in the past. They include counting methods (for example using fingers, tally stick, and markers on checkered cloth), and mechanical aids (such as weaving tools) [78]. As these tools increase in functionality and complexity, it became imperative to have a multi-purpose system that is able to perform a wide range of computations. This system – in the form of a general purpose computer – was conceived by Babbage in 1883 [16], with elements still applied in modern generic computers today: they are arithmetic logic units, conditional branching/loops, and memory.

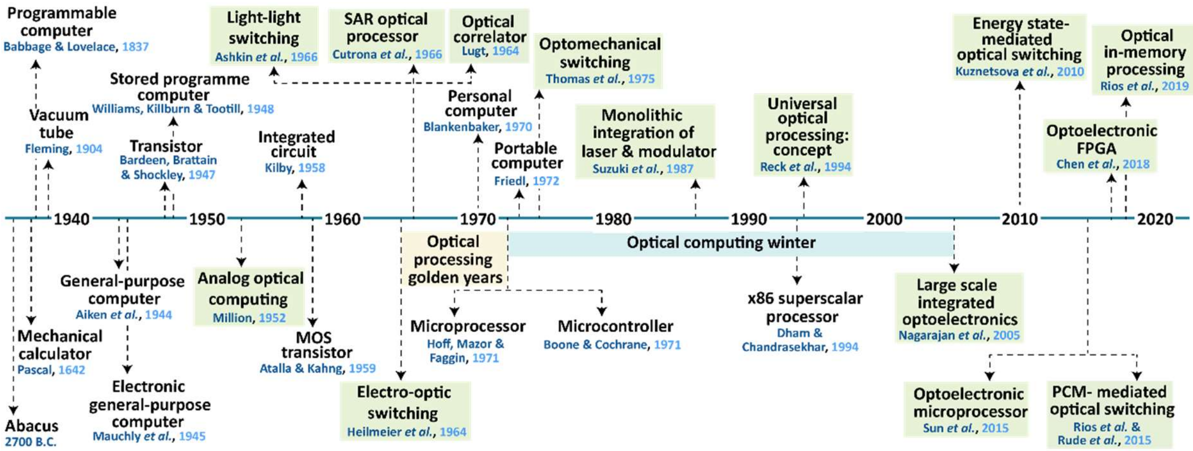

Fig. 2.1. Timeline of computer processing. The invention of the transistor ushered in the electronic age in 1947 [79,80]. The field of optical processing bloomed for a short period of time from 1964 [81] before slowing down in 1971 largely due to the proliferation of electronic microprocessors [82] and microcontrollers [83]. The realisation of large scale integrated optoelectronics in 2005 attracted renewed interests in optical processing [84]. Optical or photonic implementations are shaded in grey. Abbreviated terms MOS and SAR respectively denote metal oxide semiconductor, and synthetic-aperture radar.

However, Babbage's computing system was limited in terms of programmability, as it only accepts fixed programmed data as inputs by using mechanical punch cards [16,17]. Early electronic equivalents (Colossus in 1944 [85] and Eniac in 1945 [86]) also relied on fixed programmes. In these early electronic computers, new tasks could only be initiated by modifying the wiring and switches on the computer. Things took a turn in 1936, when Alan Turing envisioned manipulating symbols (data) for computation [87]: This hinted towards programmable computer which was explicitly conceived in 1945 by von Neumann [88] and later implemented in Manchester Baby in 1948 [89].

A timeline depicting key events in computer processing is provided in **Fig. 2.1**. The field of optical/optoelectronic processing developed alongside electronic computing. It experienced dramatic growth sparked by the use of liquid crystals to electrically modulate optical signals [81], but fell into decline with the inventions of electronic microprocessors [82] and microcontrollers [83]. There is current revival of interest in the field commonly attributed to two factors: efficient signal waveguiding using silicon structures [90], and the adoption of complementary metal oxide semiconductor (CMOS)-compatible solutions to photonic integrated circuit. The large scale integrated optoelectronic circuit reported in 2015 [84] – although electrically modulated/processed - demonstrates that on-chip circuit running on optical signals is indeed possible.

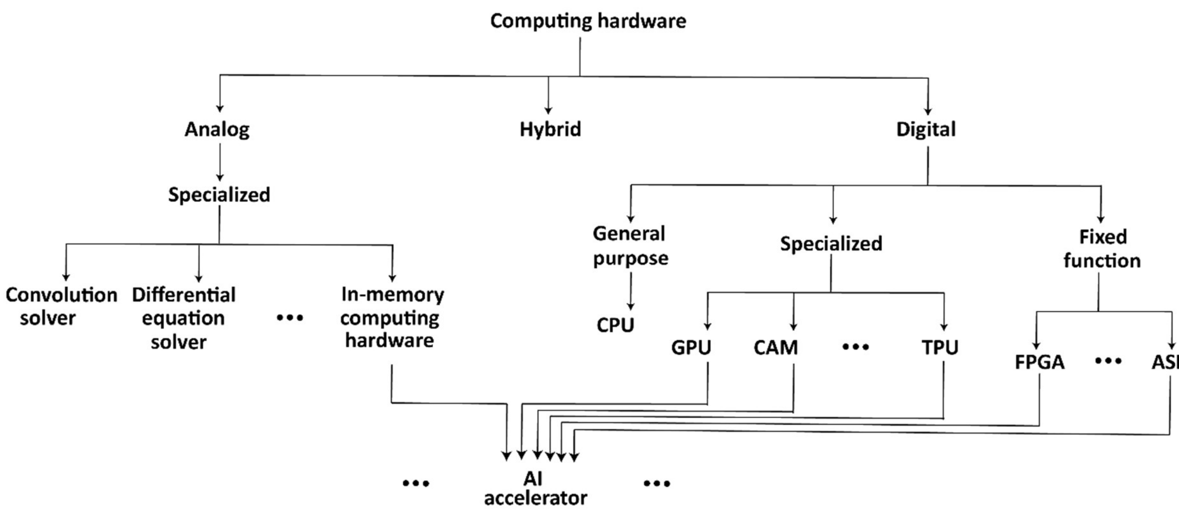

Fig. 2.2. Types of computing hardware. Computing hardware can be divided to analog, digital or a combination of both. General-purpose hardware can perform any types of computations. To expedite the computation process, specialised and/or fixed function hardware is used. One such hardware is the AI accelerator, which can take various forms (for example, in-memory computing hardware [91,92],

graphical processing unit (GPU) [93], content addressable memory (CAM) [94], tensor processing unit (TPU) [95], field-programmable gate array (FPGA) [96] or application-specific integrated circuit (ASIC) [97]).

Today, computing hardware can be classified into three types: analog, digital, and a hybrid of the two (details in **Fig. 2.2**). Analogue computers process continuous-valued data, while digital computers process discrete data (for example binary-valued data). Most commercial computers today are general purpose computers based on the 'von Neumann architecture' with the central processing unit (CPU) and memory unit as the main physical components. The CPU executes instructions provided by the programmes, while the memory unit stores general data and instructions. Overall, the computer follows the instruction cycle [88]:

i. Control unit <u>checks</u> programme counter for memory location where instruction is stored
ii. Control unit <u>fetches</u> instruction from location in memory unit
iii. Instruction register (in control unit) <u>decodes</u> instruction. Decoded instructions sent to arithmetic logic unit (ALU) for action to be executed
iv. Results from ALU <u>stored</u> in CPU register

During the instruction cycle, memories are used for different purposes. These memories can be classified based on their access speed. Primary storage has the highest access speed, while secondary storage is more distant to the CPU and is non-volatile. Other storages such as tertiary storage and off-line storage (for example, removable media drive and removable medium) do not form the main essential component of the computing process. Today, we use transistor memories [98-102] as both primary and secondary storage; and charged ion migration memories [103-107], spin-logic memories [108,109] and phase change memory (PCM) [110-117] as secondary storage.

In order to accelerate computation, certain computing tasks are offloaded onto specialised hardware components. Such specialised hardware include analogue accelerators (such as convolution solver, differential equation solver and in-memory computing hardware [91,92]), and digital accelerators (for example, graphical processing unit (GPU) [93] and tensor processing unit (TPU) [95]). Other ways to accelerate computation include specifying logic function in hardware (for example application-specific integrated circuit (ASIC) which has custom designed hardware logic functions [97], and field-programmable gate array (FPGA) which configures logic functions using programming languages after manufacturing [96]). These computing hardware can be made to perform certain functions more efficiently – for example to accelerate AI computations.

2.2 Optical computing hardware

Central to most computing operations is the switch. In early electronic computers, relays and vacuum tubes were used. As computers increased in complexity and functionality, it became apparent that energy-efficient switches (that modulate signals by a separate lower-power trigger signal) held the key to mass-efficient computers. The ubiquitous electronic devices we have today can be attributed to the monolithic integration of large numbers of these energy-efficient switches [118-120] commonly known as 'transistors' [79,80,121-126]. The transistors in today's commercial computers operate using electronic signals. The natural property of electrons to repel against each other facilitates the use of electrons as processing signals.

2.2.1 Optical processing

There is a surge of interest to use light instead of electrons to process information [70,73], given the increasing demand for information resources and services that in turn entails the need for faster signal processing. Research in optical processing is generally aimed at bringing about transformation similar to that in optical communications [73] which offers wide bandwidth, low latency and energy efficient signal propagation [127].

However, there is a caveat: light signals do not naturally interact in a linear medium [128]. Their field components would algebraically sum together ('interfere') and pass through one another as if unperturbed [129]. Although light can experience interaction in a nonlinear medium [128,130-132], this requires very high field intensity (> 10^8 V/m [133]) which makes this interaction approach unfeasible for many practical processing applications. To enhance nonlinearity, various optical field enhancement schemes have been suggested, for example using ultrashort pulses [134], and using near-zero index materials [135]. The search for the elusive reliable light-light interaction approach for processing remains an active ongoing research area [136,137]. **Table 2.1** compares the optical processing methods available.

Having good control of optical signals is the enabling factor to all-optical computing. To navigate the tradeoffs between optics and electronics, we commonly resort to optoelectronic conversions and/or processing. Such a balancing approach harnesses the natural ability of electrons to interact, and the significantly higher throughput from optical wavelength-division multiplexing (WDM) [138]. Common methods to control optical signals include applying electrical signals to tilt an array of microscopic mirrors for signal steering, and to shift the optical phase on the interferometric modulators [139], and photodetectors [140] (including 'linear light-light control' methods [141] which essentially relies on interferometric modulators to precisely control optical phases). Optoelectronic conversion has 30% energy conversion efficiency and suffers from conversion latencies, while optoelectronic processing is limited by electrical or radio frequency modulation speed (~ 100 GHz [139,142]) and consumes high energy (for example, about 10 – 20 mW for a thermal interferometric modulator [143]). In contrast, with light-light interaction (without requiring optoelectronic conversions and processing), the signal modulation speed could exceed THz range [73] with sub-fJ energy per bit.

Table 2.1. Comparison of optical processing methods.

Optical processing method	Processing speed	Power efficiency to trigger processing	Density
Nonlinear interactions	**Fast** ~ sub-ps Generally proportional to power per bit [144]	**Low** High optical power required to sustain optical nonlinearity ~ W per bit [145] Generally proportional to processing speed [144]	**Medium** For on-chip, at least several μm^3 [145] For free-space, ~ 100 μm^3 [146]
Non-volatile materials	**Medium** Conventionally ~ ns [117]	**Medium** ~ several mW (~ nJ energy with ns pulse) per bit [147,148]	**Medium** ~ tens of μm^3 [147,148]
Optoelectronic processing	**Medium** ~ 10 - 100 ps [139,142]	**Medium** Requires typically several tens of mWs per bit [143]	**Medium** ~ 0.1 mm^3 (~ several mm^2 area) [139]
Optoelectronic conversion	**Medium** For both electrical-to-optical converter (modulator) and optical-to-electrical converter (photoreceiver), ~ 10- 100 ps [139,140,142,149]	**Low** Limited by the device's external quantum efficiency (EQE) of ~ 30 % [150]	**Low** ~ 300 μm^3 [140]

A rather indirect way to process signals is by using memory elements [147,148]. For example, in a PCM memory element-deposited waveguide, the structural state of memory element changes when an optical trigger signal provide the essential signal state change. The modulation speed is limited by the PCM crystallisation speed [151] (typically in the few GHz range [113]) which is inversely related to the ratio of glass transition temperature over melting temperature [111,152].

2.2.2 Optical storage

Besides processing, information storage is essential in a computing hardware. Although optical signal information can be readily converted to electrical signals to be stored electrically, this introduces energy conversion loss and conversion latencies [138]. **Fig. 2.3** summarises key milestones in optical storage. To store light, the structure that naturally comes to mind is an optical cavity. Optical cavities are by definition structures that store light [153]. The earliest optical cavities are Fabry-Perot interferometer [154] and whispering gallery cavity [155-157]. The finest optical storage at telecommunications wavelength (~ 1.55 μm) today has cavity photon lifetime of about 0.1 μs (Q = 200 million at 1.52 μm [158]) by circulating light within a structure for as long as possible. However, the split-second lifetime in the structure is impractical for many storage tasks for computing. As light typically impedes confinement [130,159], there is currently no efficient optical cavity for storage [160].

Delay line buffers [161,162] and slow light buffers [163] are also used to delay signal retrieval for data 'storage'. In a hypothetical delay line buffer using standard single mode fibre with attenuation of 0.18 dB/km at 1.55 μm optical wavelength [164], only 1.5% of the initial signal power is left after 100 km: This would take just 0.5 milliseconds to traverse. In slow light buffers, signal information in the form of wave envelop (which contains the signal energy) is kept relatively still by tailoring the signal group velocity [165,166] such that it approaches zero. This requires judicious design of structures to support the high dispersion requisite using coupled resonators [167-170], metamaterials [171-173], atomic vapours [174].

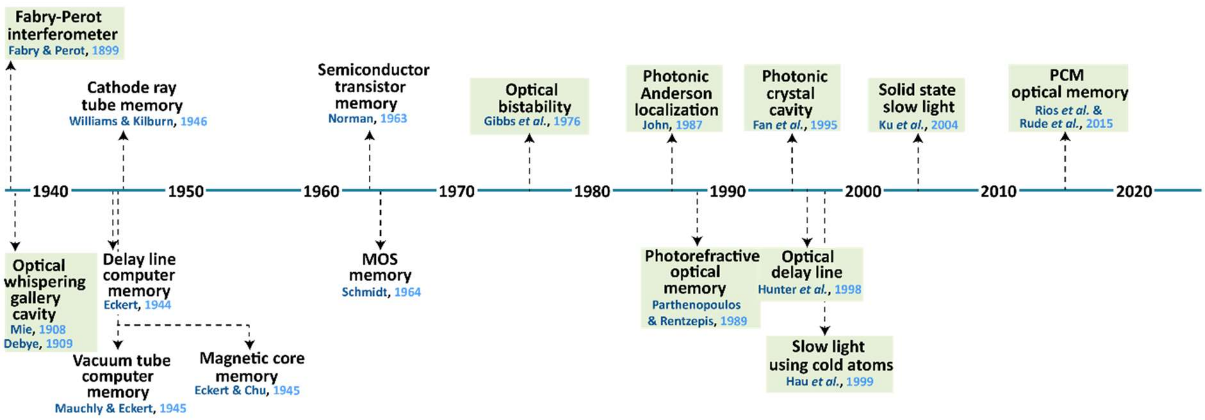

Fig. 2.3. Timeline of optical storage. Optical cavities were studied as early as 1878 [155], but not for computing purposes. Early computer memories were based on vacuum tubes [86]. Beginning 1976, other forms of optical storages were studied, ranging from optical bistability [175] to material-based optical memories such as photorefractive holographic material [176] and PCM [147,148]. Optical or photonic implementations are shaded in grey.

Due to difficulties in storing light, signal information has been stored as optical bistable states [177,178], interference condition states [134] and structural states of non-volatile material [147,148,179-181]. Both optical bistability and interference condition shifts for storage are induced using nonlinear effects. Storage in optical bistable states exploits nonlinear optical effects in lasers (e.g., cross gain modulation [182,183], cross phase modulation [184], injection locking [185,186]), while modulations to the interference condition occur due to refractive index shift [134,187]. Such nonlinear approaches require high energy to sustain optical nonlinearity [128,130,133]. To store information in the structural state of a non-volatile material, optical pulses generate heat in the material to induce reversible structural state in the material. To retrieve stored information, a separate optical signal can be sent to measure the transmission through the material [147,148,179-181]. Examples of such materials are PCM [147,148,188] and holographic storage material [179-181]. **Table 2.2** compares the optical storages.

Table 2.2. Comparison of optical storages.

Optical storage	Signal storage duration	Storage trigger and retrieval rate	Power efficiency		Density
			To sustain storage	To trigger or release storage	
Optical cavity	**Very short** ~ 0.1 µs in WGMs at 1.52 µm wavelength [158] ~ 0.1 s in superconducting Fabry-Perot cavity at 5.9 mm wavelength [189]	**Not available**	**High** No additional signal required to sustain storage. For superconducting Fabry Perot microwave cavity, additional energy is required to cool superconducting perfect electric conductor (PEC) mirrors (e.g., to $T \sim 1$K [189]).	**High** Comparable to cavity signal amplitude (~ mW range).	**Medium** Generally inversely proportional to storage duration and wavelength. Can range from ~ µm^3 to 0.1 mm^3 (with ~ µm to ~ 10 mm^2 area) for WGM at near-infrared wavelengths [158,190]. ~ 10 cm^3 for microwave Fabry-Perot cavity in [189]
Delay line buffers	**Very short** ~ µs on-chip [191] ~ ms using optical fibre [164]	**Not available**	**High** No additional signal required to sustain storage	**Not available**	**Low** ~ 0.1 cm^3 on-chip [191] ~ 0.1 m^3 using optical fibre
Slow light buffers	**Not available**	**Not available**	**High** For linear optical approaches [163,167], no additional signal required to sustain storage. For nonlinear optical approaches, additional signal may be required (e.g., to sustain gain medium [192]) For buffers using atomic vapours, additonal energy is required to cool the atoms (e.g., to $T \sim$ µK [174])	**High** Comparable to buffer signal amplitude (~ mW range).	**Low** For photonic buffers (e.g., coupled-waveguides or resonators), tens to hundreds of µm^3 [163,167,192] For atomic vapours, ~ hundreds of µm^3 [174] For microwave buffers, ~ several to tens of cm^3 [171,173]
Nonlinear bistable storage	**Long** Signal state is indefinitely retained so long as nonlinearity is sustained [177,182-186]	**High** Typically occurs in ~ ps [177,182-186]	**Low** High optical power required to sustain optical nonlinearity	**High** ~ mW range [177,182-186]	**Low** On-chip photonic lasers and optical amplifiers typically range in ~ tens of µm^3 to ~ mm^3 [182-186]
Nonlinear interference shift storage	**Long** Signal state is permanently modified via change in refractive index from optical nonlinear pulses [134]	**Not available** Rate of change of refractive index change in the photonic structure	**High** No additional signal required to sustain storage.	**Low** ~ GW range (2 mJ energy with ps pulses) [134]	**Low** Hundreds of µm^3 [134]
Non-volatile materials	**Long** ~ 10 years at room temperature [115]	**Medium** Conventionally ~ ns [117]	**High**	**Medium** For PCM, ~ several mW (~	**Medium** For PCM, ~ tens of µm^3 [147,148]

			No additional signal required to sustain storage.	nJ energy with ns pulses) [147,148] For photorefractive crystals, ~ several tens of mW [181]	For photorefractive crystals, ~ 0.1 mm^3 [181]
Electronically -mediated storage	**Long** Signal state is permanently set for as long as electrical signal is supplied.	**Medium** ~ 10- 100 ps [139,142]	**Medium** Continuous electrical signal supply required to sustain optical signal state.	**Medium** Typically several tens of mWs [143]	**Medium** ~ 0.1 mm^3 (~ several mm^2 area) [139]

2.2.3 Photonic integrated circuit

The field of photonic integrated circuit began in earnest in 1969. In a seminal paper [193], Miller envisioned microphotonic integration similar to electronic integration of circuit components onto a single chip [193]. The proposal came at a time when optical modes were extensively studied [194,195], when the high attenuation in glass optical waveguides was determined to be from impurities [196] and when the laser was invented by Maiman [197]. Notably, the microphotonic basic building blocks identified by Miller (lasers, modulators, photodetectors) still hold today. They are shown in **Fig. 2.4**.

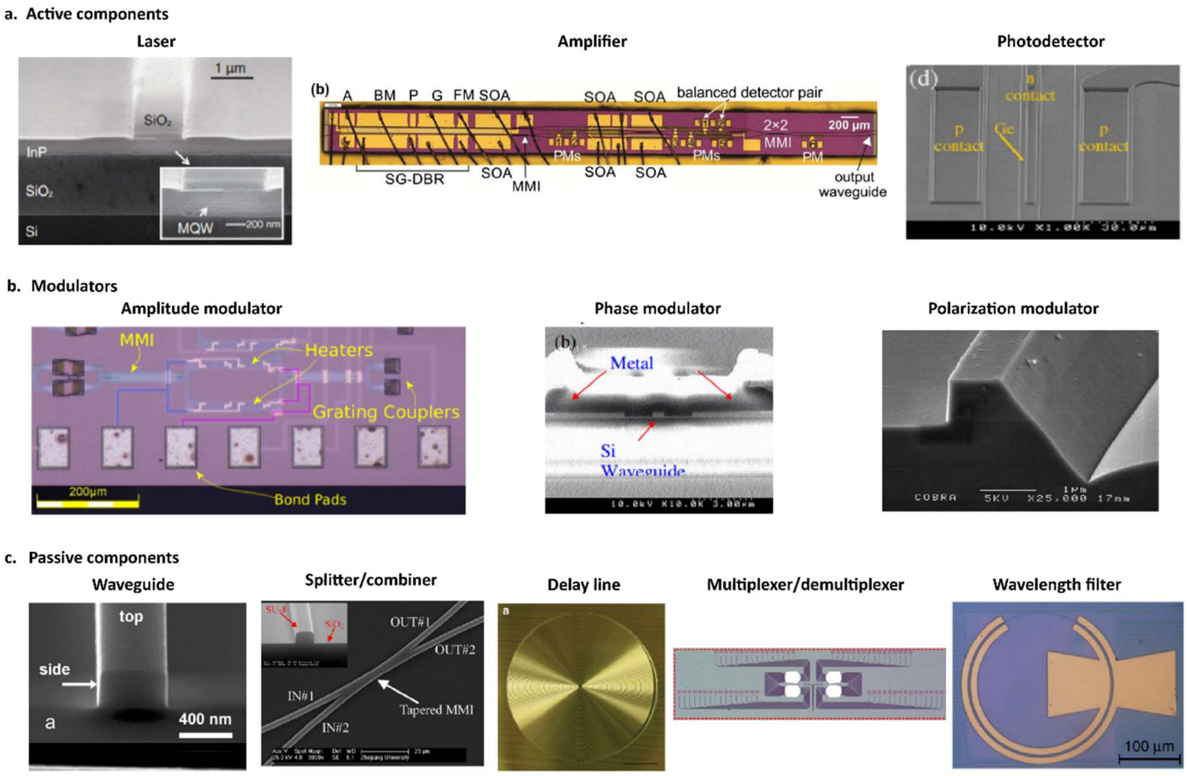

Fig. 2.4. Main components in integrated photonic circuit. Components can be divided into active components (**a**), modulators (**b**), and passive components (**c**). Common active components include lasers [198], amplifiers [199] and photodetectors [200]. Types of modulators include amplitude [201], phase [202],

and polarisation [203] modulators. Common passive components include waveguide [204], splitter or combiner [205], delay line [191], and multiplexer or demultiplexer [7,206] and wavelength filter [207].

It is desirable to have the components in **Fig. 2.4** monolithically integrated onto a silicon chip, using technology from the microelectronics industry. However, it is not straightforward to realise silicon-based laser and photodetector, as crystalline silicon has an indirect bandgap of 1.1 eV (~ 1.12 nm wavelength) [208]. As such, other materials have been considered. They are compared in **Table 2.3**. Hybrid and heterogenous integration of Si_3N_4, silicon, InP system, $LiNbO_3$ and polymer are adopted for cheap, practical and ubiquitious photonic system (details in **Fig. 2.5**). For example, due to the high cost of InP system, monolithic integration of InP photonic components is less usual. Instead, the material, which is typically used to construct lasers, is commonly co-integrated with silicon-based waveguides (Si_3N_4 and silicon) or $LiNbO_3$ modulators. It is usual to also co-integrate germanium (Ge) and cerium-substituted yttrium iron garnet (CE:YIG) on silicon-based platforms. Ge – with direct bandgap of 0.8 eV (i.e., at 1.55 μm wavelength) [209] – can be monolithically integrated on silicon for silicon-based photodetection [210], with the two-step Ge-on-Si heteroepitaxial growth process widely adopted to circumvent the 4% Ge-Si lattice mismatch problem [211]. Ce:YIG have been bonded to silicon-based waveguides to realize optical isolators and circulators [212].

Table 2.3. Comparison of materials for photonic integrated circuit.

Material	Density	Versatility				CMOS fabrication compatibility	Main features
		Laser	Amp	PD	Modulator		
Si_3N_4	**Medium** (n_{ref} ≈ 2.1 at 1.55 μm wavelength [213])	**Medium** Hybrid [214], heterogenous [215]. Generally insulating (≈ 5 eV indirect bandgap [216])	**Medium** Hybrid [214]	**Medium** Hybrid, heterogenous [217]	**Medium** Medium electro-optic (EO) coefficient (≈ 6 pmV^{-1} [218])	**Medium** (CMOS fab compatible [219], but not CMOS compatible when low pressure chemical vapour deposition (LPCVD) annealed at temperature > 1000°C [220])	Cheap (abundant:27 % of Earth's crust is silicon [221]). Process-dependent material properties [222].
Silicon	**High** (n_{ref} ≈ 3.44 at 1.55 μm wavelength [223])	**Medium** Hybrid [224], heterogenous [225], monolithic [226], epitaxial [227]. Crystalline silicon has indirect bandgap of 1.1 eV. Hydrogenated amorphous silicon has direct	**Medium** Hybrid [228], heterogenous [229]	**Medium** Hybrid [230], monolithic [231]	**High (when doped)** Monolithic, with high EO coefficient (≈ 20 pmV^{-1} [232]) and doping-induced losses.	**High** [219]	Cheap. Good for electronic-photonic integration.

		bandgap of 1.84 eV (visible wavelength) but with higher absorption [208]					
InP system*	**High** ($n_{ref} \approx 3.16$ at 1.55 µm wavelength [233])	**High** Direct bandgap of 1.344 eV energy [209]	**High** [234]	**High** High responsivity ($\approx$ 7.5 A W^{-1} at 1.55 µm wavelength [235])	**High** High EO coefficient ($\approx$ 20 pmV^{-1} [236]). But modulation temperature sensitive (band edge moves with temperature [237])	**Low** [219]	Good for laser integration. Expensive.
LiNbO$_3$	**Medium** ($n_{ref} \approx 2.21$ at 1.55 µm wavelength [238])	**Low** Direct and indirect bandgap of 3.5-4.7 eV (mid-infrared region [239,240])	**Low**	**Medium** Heterogenous and monolithic at $\approx$ 635 nm wavelength [241]	**Very high** High EO coefficient ($\approx$ 30 pmV^{-1} [242]). Based on Pockel's effect. Modulation at $\approx$ 50 – 100 GHz [142]	**Low**	Good for signal modulation. Low damage threshold [243]
Polymer	**Low** ($n_{ref} \approx 1.5$ at 1.55 µm wavelength [244])	**Medium** Hybrid [245]. Bandgap varies. Polymer lasers typically optically pumped [246]	**Medium** Hybrid [247]	**Medium** Hybrid [248]	**Varies** EO coefficient can be high ($\approx$ 300 pmV^{-1} [249])	**Low**	Good for integrating with silicon, silicon nitride and InP and GaAs platforms. Low reliability (thermal management)

*The InP system covers III-V semiconductors lattice matched to InP, for instance InGaAs , InGaAsP, InGaAlAs and InAlAsP [250,251].

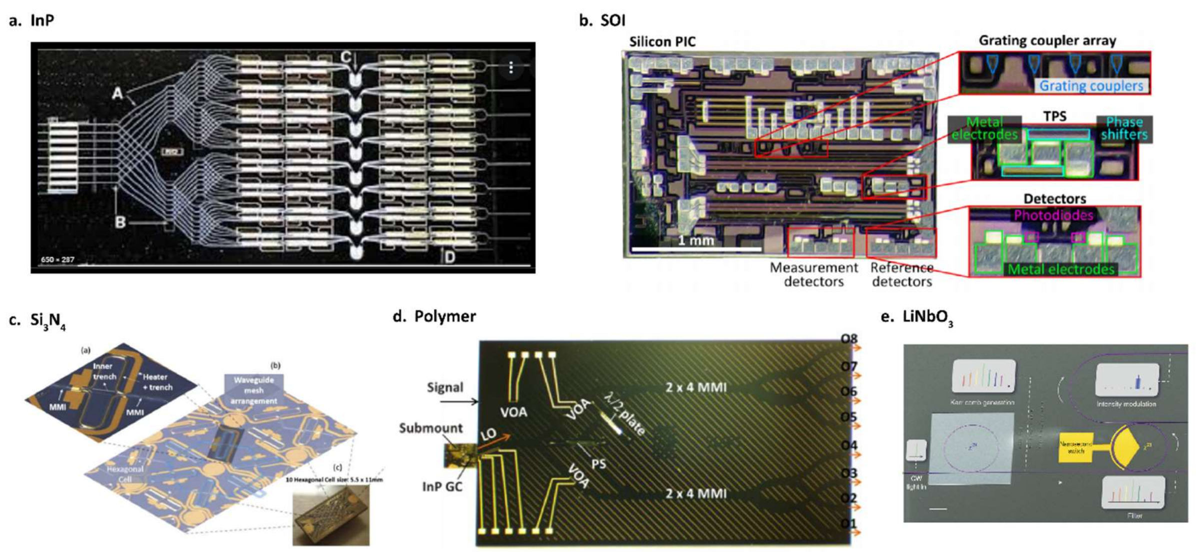

Fig. 2.5. Materials for photonic integrated circuit. Common materials are **a**, InP which is optically active [252], **b**, SOI [253], **c**, Si_3N_4 [254], **d**, polymer which provides a low cost alternative [255] to silicon-based

platforms in (**c**, **d**), and **e**, $LiNbO_3$ which has high electro-optical coefficient for optical signal modulation [256].

2.3 Optical purpose-built computing hardware

While there is currently no competitive optical alternative to electronic CMOS digital switches, this is not the case for analogue computations. Specialised analogue calculations can be readily performed using optical signals. An example is the use of light for differentiation calculations based on Beer-Lambert law $dI/dz = -\alpha I$, where one can carry out differentiation computations simply by measuring optical intensity I of an optical signal traversing across a medium of attenuation constant α after z distance [257,258]. Other reported optical analogue computations include basic operations (e.g., differentiation [259], integration [260,261]), mathematical transform operations (e.g., convolution [262], Fourier transform [263]) and other linear algebra calculations [264,265]. As design requirements increase in complexity, less-standard methods such as nonlinear optimisation [266] and metamaterials [267,268] are adopted (compared to traditional platforms such as optical lenses and filters, and later optical waveguides, and photonic crystals [257,259-262,264]) to enable complex simultaneous optimisation of interdependent optical parameters within a more spatially compact region.

Because optical analogue methods rely on physical phenomenon, they only accelerate certain computations. They are not suitable for computations that rely on processing [79,80,121-126] and storage units [98-117] in some way or another, or that is digital in nature [52,269-272]. For example, it is not straightforward to realise optical FPGAs [270,271] using analogue platforms as digital logic gates are inherent to the architecture. Thus, when implemented optically, such optical hardware accelerators have the same limitations as those in optical digital processing and optical storage, being generally limited to electrically-modulated optical switches [273] or optically-induced structural state change in material such as PCM [188] and ion-implanted silicon [274].

2.4 Optical AI hardware acceleration

One class of purpose-built computing hardware is the AI hardware accelerator which solves AI tasks more efficiently than is possible on software running on generic central processing units (CPUs) alone [71,72]. Such accelerators typically have distributed memories to represent AI network connection strengths with large parallel input data that can be conveniently handled using optical WDM [70-72]. Because reliable fully-optical CPU is presently absent, AI accelerators are today co-processed with an electronic CPU via DACs and ADCs, although other hybrid co-processing architecture may be considered. These are summarised in **Fig. 2.6**.

AI hardware accelerators can be classified based on its hardware reconfigurability: FPGA-based AI accelerators allow their logic block interconnects to be reconfigured [96], while non-reconfigurable AI accelerators are conventionally in the form of crossbar array for matrix [92,275] or tensor-multiplication [95,276,277] (since the connections between nodes in successive layers in deep neural networks (DNN) can be represented as matrix multiplication). Examples of such accelerators are GPUs and TPUs for matrix and tensor multiplications, respectively.

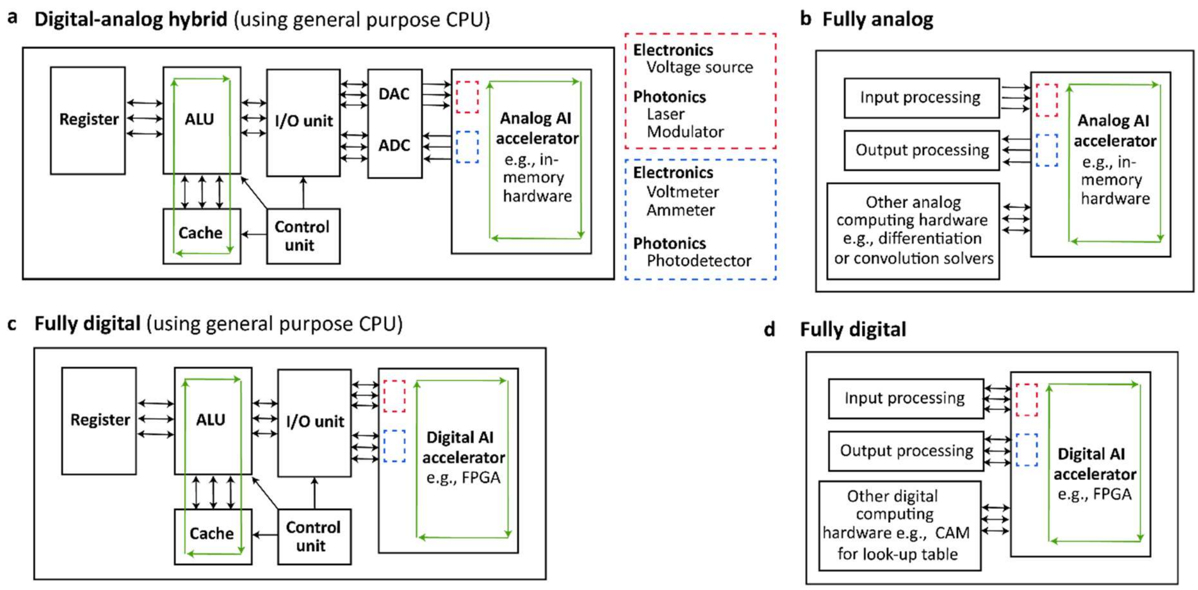

Fig. 2.6. Hybrid co-processing AI architecture. Architecture can be fully digital, fully analog, or hybrid. **a**, Conventionally, analogue AI accelerator is co-integrated with the digital CPU processor, which mainly consists of arithmetic logic unit (ALU), cache memory and register [95]. **b**, A distinct conceptual architecture is the fully analogue architecture, which replaces the digital CPU processor with an analogue input and output processing unit. **c, d**, In a fully digital architecture, the digital AI accelerator is co-integrated with a digital CPU or other input/output processing units (I/O unit). Green arrows denote processing. Red and blue boxes denote input and output sources to AI accelerator, respectively. Abbreviated terms DAC and ADC respectively denote digital-to-analog converter and analog-to-digital converter.

In electronic TPUs, the ability to process additional dimension of data is due to spatial partitioning (or multiplexing) [95]. However, it is bound to resistance-capacitor (RC) bandwidth-distance tradeoff [278], where data channels cannot be too closely spaced, lest RC delay becomes significant. The parallelisation can be streamlined using optically using WDM [276,277] to provide a convenient means to tensor multiplication, and the option to further extend the parallelism via spatial multiplexing. This, when combined with the AI accelerator implementations summarised in **Fig. 2.7** (using spectral filters [279,280], photorefractive crystals [281-283], semiconductor photodiodes [284], interferometric modulators [285] and free-space scattering layers [286]) provides increase in signal throughput.

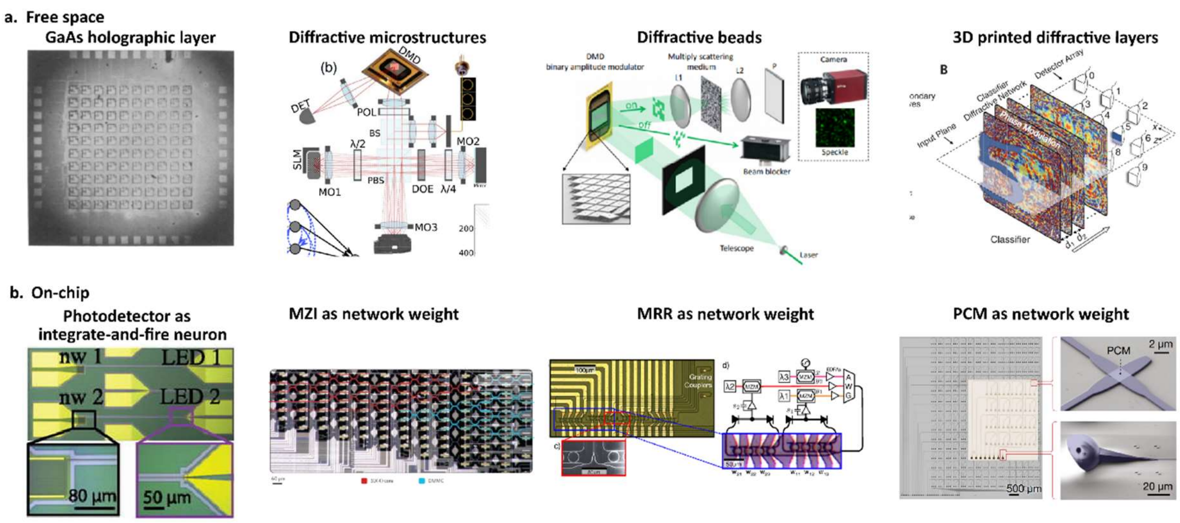

Fig. 2.7. Optical AI accelerator implementations. a, Free space approaches include using holographic layer [281], diffractive microstructures [287], diffractive beads [288] and 3D printed diffractive layers [286] as neural network layers. **b**, On-chip implementations include using photodetector as integrate-and-fire neuron [289], Mach-Zehnder interferometer (MZI) [285], microring resonator (MRR) [280] and PCM [276] as network weights.

There has been an increase in interest in optical/photonic AI accelerators, with important work in the field shown in **Fig. 2.8**. In conventional AI hardware accelerators, the network weights (represented largely by the matrix/tensor multipliers) are numerically approximated [28,29] to gradually [290,291] develop a 'perceived' model representation of the input dataset it receives [30]. Such numerically-approximated networks can be very adept at finding seemingly unrelated patterns in the datasets, but do not understand how they arrive at the solution [3,4,33]. A distinct form of hardware accelerator using symbolically-detected associations finds patterns in input datasets, as will be described later in this thesis. Current AI systems have either sophisticated learning capabilities (when solved numerically) or reasoning capabilities (when solved symbolically), but not both [292].

The realisation of an AI hardware accelerator that is skilled at both learning and reasoning tasks [63,293], and efficient in both computational effort and energy consumption is expected to be of great significance in the 'fourth industrial revolution' – which characterises the ongoing decentralisation and integration of decisions from machines, devices and people to self-optimise, selfconfigure and self-diagnose [294,295]. It demands enabling tools for the purpose. Prior industrial revolutions have had requisite enabling tools: steam engine [296-298], internal combustion engine [299,300], electronic transistor [79,80,121-126], and integrated circuit [118-120]. It is expected that the invention of strong learning and reasoning algorithms that run on powerful AI accelerators would epitomise the fourth industrial revolution that we are currently in. As with previous tools, the accelerator could project a transformative economic, social and cultural impact to shape the underpinnings of our modern society [301].

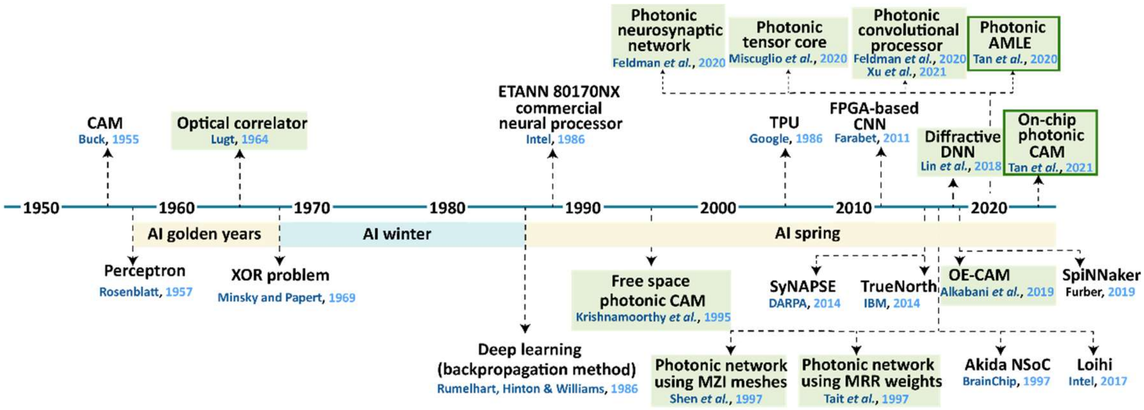

Fig. 2.8. Timeline of AI and optical/photonic AI acceleration. Learning models can be divided to neuron models [20-22], synapse [7] and associative learning [5] models. Optical or photonic implementations are shaded in grey. Green bordered boxes denote the learning platform introduced during the author's DPhil studies. Abbreviated terms DNN, NSoC, and TPU respectively denote deep neural network, neuromorphic system-on-chip, and tensor processing unit.

We have reviewed computing hardware in general, and placed the author's DPhil research in the context of AI hardware acceleration. Current progress in the field is provided, and motivations to optically implementing the acceleration are examined. While speed and throughput are the main motivations, it is shown that the implementations are not without difficulties, given the very nature of light that does not naturally self-interact and which impedes confinement. The different approaches that have been undertaken to overcome such limitations are discussed. Finally, real-world impact of the effort to our society is projected.

Chapter 3

3. Optical associative learning device: Theory, design and fabrication

This chapter describes the optical associative learning device. The concept of associative learning is first described. Then, the two complementary on-chip layout designs to measure the device, and the steps to fabricate them are elucidated.

3.1 Introduction to associative learning

Associative learning is one of the basic forms of learning. It pertains a relatively enduring change of behavior due to signal pairing(s) from the environment. In particular, a form of associative learning known as 'classical conditioning' associates input stimuli from the environment [5,302]. This learning process can be graphically described using the neural circuit simplified in **Fig. 3.1a** [303,304].

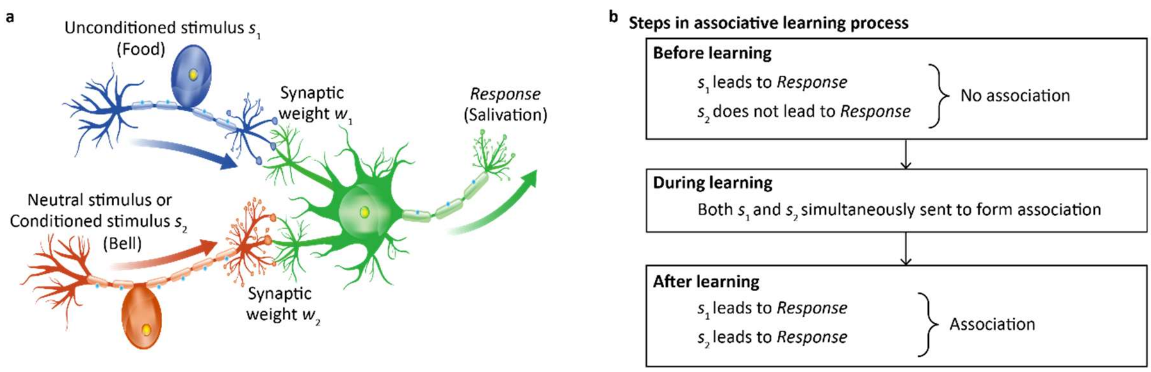

Fig. 3.1. Associative learning process. a, Simplified illustration of the neural circuitry for associative learning. After the stimulus s_2 is persistently paired with stimulus s_1, it is conditioned to the same response as s_1. **b**, Steps in associative learning process. Before learning, only s_1 input signal (Food) produces a response (Salivation). When both s_1 and s_2 signals are simultaneously sent, association is established between the two. After learning, s_2 input signal (Bell) comes to suggest s_1 input signal (Food), as both the input signals produces the response (Salivation).

In **Fig. 3.1a**, the motor neuron generates an action potential after receiving a sensory signal. The sensory signal can be either the natural stimulus ('unconditioned' stimulus (s_1)) or neutral/conditioned stimulus (s_2). The process of associating stimulus s_2 from a sensory neuron with the natural stimulus s_1 is the process of association. This learning mechanism, first described by Ivan Pavlov [5], showed that salivation in a dog can be 'stimulated' by associating the ringing of a bell with food. Once the association is established, the salivation response is triggered when either s_1 or s_2 is sent to the motor neuron through a synaptic weight (w_1 or w_2). Although it is straightforward to simulate the entire associative learning process (summarised in **Fig. 3.1b**) on a general-purpose computer using if-else statements, the learning process can be expedited when computed on a dedicated hardware device that naturally implements the associative learning process: this shall be discussed in this chapter.

3.1.1 Integrated optical waveguide

An associative learning hardware device processes input signals based on the concept of associative learning. When implemented optically, these signals obey the electromagnetic wave equation [305]. This equation encapsulates the optical field components (or modes) that are allowed to exist in the device media.

In the design of the device, optical waveguide forms the main component. To establish the qualitative aspects of the waveguide, we begin our analysis on a slab waveguide (1-dimensional waveguide), in which the optical signal travels in a direction parallel to a two-dimensional plane (x-y plane), and is uniform in the z-axis [306]. The signals within the plane are effectively decoupled as scalar field Ψ. In simple medium [305], $\Psi(x,y,z\,;\,t)$ satisfies

$$\nabla^2\psi = \frac{1}{c^2}\frac{\partial^2\psi}{\partial t^2} \tag{3.1}$$

where c is speed of light in vacuum. When the field propagates in the medium with refractive index n only in the x-direction ($\partial/\partial y = 0$), and has sinusoidal space- and time-dependence ($e^{-i\beta x}$ and $e^{i\omega t}$) with propagation constant β and angular frequency ω, the scalar $\Psi(x,y,z\,;\,t) \sim \Psi(z)e^{i\omega t}$ where $\Psi(z)$ satisfies

$$\left(\frac{\partial^2}{\partial z^2} + k^2 n_z^{\,2} - \beta^2\right)\psi = 0 \tag{3.2}$$

where n_z denotes refractive index of the medium as a function of z. The solution for $\Psi(z)$ is consequently

$$\psi(z) = \begin{cases} C_{11} \cos(K_1 z) + C_{12} \sin(K_1 z) & , \beta < kn_z \\ C_{21} e^{-K_2 z} & , \beta > kn_z \end{cases} \tag{3.3}$$

The field is normalised by C terms. The variables $K_1 = (k^2 n_z^2 - \beta^2)^{1/2}$ and $K_2 = (\beta^2 - k^2 n_z^2)^{1/2}$. To sustain waveguiding, the field in the guide has to be oscillatory [305-307]. This can be achieved using high refractive index medium as the guide core ($n_z \sim n_{core}$) placed between claddings of low refractive index media ($n_z \sim n_{cladding}$, where the field exponentially decays).

We now consider a slab guide with core height h, and core-cladding boundary at $z = 0$ and $z = h$. For simplicity, we assume vanishing field in the cladding and apply the boundary conditions $\Psi(z = 0) = \Psi(z = h) = 0$. It follows that the guide supports fields of propagation constant β with discrete mode m:

$$\beta_m = \sqrt{k^2 n_z^2 - \left(\frac{m\pi}{h}\right)^2} \quad , \quad m = 1,2,3,... \tag{3.4}$$

from which we can deduce the following:

- Discrete modes with distinct β_m exist in the waveguide
- The waveguide may support more modes when we:
  - Increase waveguide height h
  - Increase refractive index of waveguide core $n_z \sim n_{core}$
  - Decrease signal wavelength λ (given $k = 2\pi/\lambda$)
- There is a cutoff height h below which a mode ceases to exist (when $\beta_m < kn_z$).
- There is a threshold mode m beyond which the mode becomes radiative (when $\beta_m < kn_z$). The maximum allowed number of guided/non-radiative modes is $2h/\lambda$ (because $\sin \theta_m = m\lambda/2h$, where θ_m is the angle between β_m and kn_z)

In an actual waveguide, the field diminishes exponentially (not abruptly) into the cladding**.** The non-diminishing field can be compensated by slightly increasing the slab height by a small real value δh. We estimate this parameter by equating the critical angle for total internal reflection (which forms the basis of waveguiding): $\pi - \theta_c = \cos^{-1}(n_{cladding}/n_{core})$ with $\theta_{m=1} = \cos^{-1}(\beta_{m=1}/kn_{core})$:

$$\delta h = \frac{\pi}{k^2\left(n_{core}{}^2 - n_{cladding}^2\right)} - h \tag{3.5}$$

As cladding refractive index $n_{cladding}$ increases, waveguide field confinement decreases (due to δh increase).

In a channel waveguide with confinement in the *y*- and *z*-axis, the wave equation is

$$\left(\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}+k^2 n_{yz}{}^2-\beta^2\right)\psi=0 \tag{3.6}$$

where n_{yz} denotes refractive index of the medium as a function of y and z. The discrete propagation constants are given by:

$$\beta_{\mathrm{m}}=\sqrt{k^2 n_{\mathrm{yz}}{}^2-\left(\left(\frac{m\pi}{w}\right)^2+\left(\frac{n_{\mathrm{h}}\pi}{h}\right)^2\right)} \tag{3.7}$$

The optical signal (electromagnetic field) consists of electric *E* and magnetic *H* field components. On the boundary surface of an ideal waveguide, the *E* (*H*) fields tangential (normal) to the surface vanishes [306,307]. We cannot generally satisfy these conditions at the same time in a conventional waveguide. As a result, the waveguide supports two distinct types of modes: transverse electric (TE) and transverse magnetic (TM) which are respectively H_z - and E_z - dominant. The solutions for the wave equation are

$$\psi(y,z)=\begin{cases}\psi_0\, cos\left(\dfrac{m\pi y}{w}\right)cos\left(\dfrac{n_{\mathrm{h}}\pi z}{h}\right) & \text{TE mode}\\ \psi_0\, sin\left(\dfrac{m\pi y}{w}\right)sin\left(\dfrac{n_{\mathrm{h}}\pi z}{h}\right) & \text{TM mode}\end{cases} \tag{3.8}$$

which specifically indicate:

$$\begin{aligned}H_{\mathrm{x}}(y,z)&=H_{\mathrm{o,mn}}\, cos\left(\frac{m\pi y}{w}\right)cos\left(\frac{n_{\mathrm{h}}\pi z}{h}\right) && \text{TE mode}\\ E_{\mathrm{x}}(y,z)&=E_{\mathrm{o,mn}}\, sin\left(\frac{m\pi y}{w}\right)sin\left(\frac{n_{\mathrm{h}}\pi z}{h}\right) && \text{TM mode}\end{aligned} \tag{3.9}$$

Deriving from **(3.6)** using the usual method of separation of variables [307,308], we can retrieve the field components for TE mode:

$$E_x(x,y,z;t)=0 \tag{3.10a}$$

$$E_y(x,y,z;t)=\frac{i\omega\mu n_h\pi}{K_1^2 h}E_{o,mn}\,cos\left(\frac{m\pi y}{w}\right)sin\left(\frac{n_h\pi z}{h}\right)e^{i(\omega t-\beta x)} \tag{3.10b}$$

$$E_z(x,y,z;t)=-\frac{i\omega\mu m\pi}{K_1^2 w}E_{o,mn}\,sin\left(\frac{m\pi y}{w}\right)cos\left(\frac{n_h\pi z}{h}\right)e^{i(\omega t-\beta x)} \tag{3.10c}$$

$$H_x(x,y,z;t)=H_{o,mn}\,cos\left(\frac{m\pi y}{w}\right)cos\left(\frac{n_h\pi z}{h}\right)e^{i(\omega t-\beta x)} \tag{3.10d}$$

$$H_y(x,y,z;t)=\frac{i\beta m\pi}{K_1^2 w}H_{o,mn}\,sin\left(\frac{m\pi y}{w}\right)cos\left(\frac{n_h\pi z}{h}\right)e^{i(\omega t-\beta x)} \tag{3.10e}$$

$$H_z(x,y,z;t)=\frac{i\beta n_h\pi}{K_1^2 h}H_{o,mn}\,cos\left(\frac{m\pi y}{w}\right)sin\left(\frac{n_h\pi z}{h}\right)e^{i(\omega t-\beta x)} \tag{3.10f}$$

and for TM mode:

$$E_x(x,y,z;t)=E_{o,mn}\,sin\left(\frac{m\pi y}{w}\right)sin\left(\frac{n_h\pi z}{h}\right)e^{i(\omega t-\beta x)} \tag{3.11a}$$

$$E_y(x,y,z;t)=-\frac{i\beta m\pi}{K_1^2 w}E_{o,mn}\,cos\left(\frac{m\pi y}{w}\right)sin\left(\frac{n_h\pi z}{h}\right)e^{i(\omega t-\beta x)} \tag{3.11b}$$

$$E_z(x,y,z;t)=-\frac{i\beta n_h\pi}{K_1^2 h}E_{o,mn}\,sin\left(\frac{m\pi y}{w}\right)cos\left(\frac{n_h\pi z}{h}\right)e^{i(\omega t-\beta x)} \tag{3.11c}$$

$$H_x(x,y,z;t)=0 \tag{3.11d}$$

$$H_y(x,y,z;t)=\frac{i\omega\varepsilon n_h\pi}{K_1^2 h}H_{o,mn}\,sin\left(\frac{m\pi y}{w}\right)cos\left(\frac{n_h\pi z}{h}\right)e^{i(\omega t-\beta x)} \tag{3.11e}$$

$$H_z(x,y,z;t)=-\frac{i\omega\varepsilon m\pi}{K_1^2 w}H_{o,mn}\,cos\left(\frac{m\pi y}{w}\right)sin\left(\frac{n_h\pi z}{h}\right)e^{i(\omega t-\beta x)} \tag{3.11f}$$

From **(3.7)**, the normalized mode cutoff frequency is

$$\frac{\omega_m}{c}=\sqrt{\left(\frac{m\pi}{w}\right)^2+\left(\frac{n_h\pi}{h}\right)^2} \tag{3.12}$$

For nontrivial solutions, *m* and *n* cannot be simultaneously zero. The lowest TE and TM cutoff frequency is respectively when (*m*,*n*) = (1,0) or (0,1), and (*m*,*n*) = (1,1). For TE mode, it can be readily ascertained from **(3.10)** that when *m* = 1, and $n_h = 0$, the field component $E_x = E_y = H_z = 0$.

It is usual to more rigorously examine waveguide modes using numerical or graphical methods provided in [306] for greater accuracy. They involve theoretically solving transcendental equations which do not have closed-form analytical solutions (and are thus less intuitive).

Both silicon and silicon nitride have been widely used as waveguide material of choice due to their low optical losses at telecommunications wavelength (~ 1.55 µm). In particular, while silicon enables better potential for device integration with smaller footprint, silicon nitride can support optical power up to several Watts without introducing optical nonlinearity [222,309,310], has broader spectral transparency (from 0.4 µm to 4 µm wavelength range [223], compared to 1.1 µm to 4 µm on silicon), lower distributed backscattering (~ -35 dB/mm, compared to ~ -25 dB/mm on silicon) [311], higher versatility to stack layers (due to the use of plasma-enhanced, and low pressure chemical vapour deposition (PECVD and LPCVD) [312], compared to wafer bonding on silicon [313]) and has higher switching energy efficiency when deposited with the phase change material $Ge_2Sb_2Te_5$ (GST) [314]. Although both platforms can be adopted to realise the associative learning hardware device, silicon nitride is arbitrarily used as the waveguide material of the device. Later in **Chapter 7**, the realisation of a different type of photonic 'associative learning' hardware known as content addressable memory on silicon platform is discussed.

3.1.2 Integrated optical waveguide with memory element

In the previous subsection, an overview of optical waveguide modes was presented. With increasing m, the waveguide mode is more prone to radiate (due to decreasing β_m), and has higher propagation loss (due to increasing total internal reflections to sustain mode, as indicated by the increase in θ_m). Waveguides described in this thesis are of single-mode guidance (m = 1) unless specifically mentioned. As will be described, the associative learning device is made up of such waveguides placed in closed proximity.

An associative learning device additionally requires a memory hardware to store information about the different learning states (before and after learning). It is convenient to use the bistable optical properties of non-volatile phase change materials [110] for the purpose. Ellipsometry measurements indicate that at telecommunications wavelengths ($\lambda \approx 1.55$ µm), the refractive indices of the phase change material GST are $n_{ref} = 7.5+2.3i$ in the crystalline state and $n_{ref} = 3.8+0.1i$ in the amorphous state. The imaginary n_{ref} (which contributes to imaginary $\beta = 2\pi n_{eff}/\lambda$) indicate optical transmission attenuation. This can be ascertained by considering a plane wave with sinusoidal dependence $e^{i\beta x}$ that exponentially decays with increasing imaginary β.

The GST can be integrated onto an optical waveguide by depositing a thin film of it on the waveguide [147]. The deposition does not introduce abrupt change in field propagation, which would cause optical backreflection [306,315]. To switch the GST from amorphous to crystalline, the switching temperature has to be above the material's glass transition temperature (T = 125 °C) for a sufficiently long time (≈ hundreds of ns). In contrast, to switch from crystalline to amorphous state, the switching temperature has to be above the GST melting temperature (T = 600 °C) and then cooled rapidly (≈ 50 ns) [117]. Such required heat can be provided by intense laser pulses via a material with good heat conductivity (> 1 W m^{-1} K^{-1}) [147]. The GST is known to have high cycling endurance (~ 10^{12} cycles [115,316]), and long retention time (>10 years at room temperature [115]).

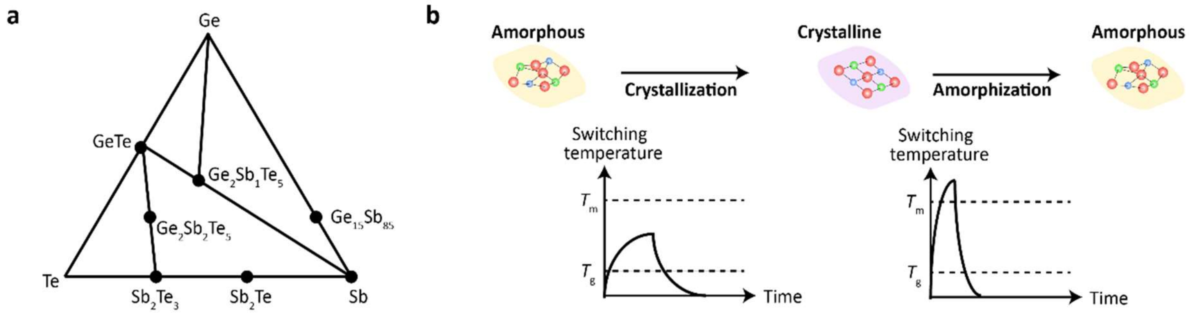

Fig. 3.2. **GST structural phase switching. a**, Ternary Ge-Sb-Te phase diagram with common phase change alloys indicated. Stoichiometric compounds along the red line i.e., GeTe-Sb_2Te_3 pseudo-binary line, is found to have rapid and repeated structural phase switching properties [111]. **b**, Structural phase switching from amorphous state to crystalline state requires the switching temperature to be above the material's glass transition temperature for a sufficiently long time. On the other hand, the switching from crystalline to amorphous state requires the switching temperature to be above the material's melting temperature and cooled rapidly after the trigger pulse is applied [117].

3.2 Theory and design of associative learning element

In this section, the optical associative learning device which consists of phase change materials-deposited optical waveguides is described. Previous associative learning devices have been developed on electronic [317-329], optoelectronic [330], and synthetic biological [331] platforms. They rely on multiple elements (for example, arithmetic and/or Boolean logic gate circuit components) [317-327], or large area thin film structures that depend on free space data inputs [330]. In contrast, the associative learning device described in this thesis is monadically realised, on-chip and triggered using all-optical signals. The device is designed, fabricated, and corroborated using coupled-mode theory and finite-difference time-domain (FDTD) numerical simulations.

3.2.1 AMLE concept and device design framework

We can identify two main roles of the simplified neural circuitry in **Fig. 3.1a**: firstly, to converge / associate the two inputs (s_1 and s_2 sensory inputs) and secondly, to store memories of these associations *in situ*. Such functionality can be realised optically using coupling between two modes. There are at least two ways to achieve this – using multimode interferometer (MMI) or directional coupler (DiC). MMI has broader waveguide width to support higher order modes to enable coupling, while DiC consists of two distinct waveguides placed in close proximity likewise to enable coupling. A comparison between the two is summarised in **Table 3.1.**

Table 3.1. Comparison of mode coupling approaches.
Factor taken into special consideration is framed in green.

Structure	Design simplicity	Bandwidth	Resistance to crosstalk	Polarisation tolerance	Spatial density
1. MMI	**Medium** Self-imaging method - conventionally used to design and analyse MMI – is less straightforward [332]	**High** Large MMI supermode guide separation prevents coupling between them [332]. Wavelength-independent for strong guided structures [333]	**Low** Crosstalk is sensitive to length [334]	**Low** Polarisation dependent [332]	**High** Can have 2-3 orders of magnitude shorter coupling length than that of DiC [334]
2. DiC	**High** Straightforward to design and analyse using coupled mode theory [167,335].	**Medium** A function of wavelength. Transmission variation $\Delta Tr/\Delta\lambda \sim$ several μm^{-1} [336]	**Medium** Crosstalk = $\cot^2((\pi/2)(L/L_c))$ Insensitive to length L within 10% of critical coupling length L_c [334]	**High** High polarisation tolerance when coupling length for TE and TM mode designed to be the same, via careful design of ridge waveguide width and gap [334]	**Medium** Higher coupling length with increasing etch depth (increasing from ridge to rib waveguides in DiC) [334]

Although both structures have their own sets of relative advantages and disadvantages, DiC is chosen over MMI for the device as it is more convenient to design: the associative monadic learning element (AMLE) consists of two cascaded coupled waveguides (or two cascaded directional couplers, **Fig. 3.3a**), which are designed to support only fundamental TE waveguide modes (H_z field component-dominant). These segments are denoted as I and II. They respectively associate s_1 and s_2, and store memories of the associations. The phase change materials thin film deposited on the lower waveguide is the well-known GST alloy. A thin capping layer of indium tin oxide (ITO) is deposited on the GST to prevent oxidation, and to help localise optically-induced heat to enable low-power GST structural phase switching [147,337]. The GST effectively modulates the optical field coupling between the waveguides. It can exist in two states (amorphous or crystalline), and can be in between these two states. These two states (and the fractional volumes between the two states) govern the amount of coupling between the waveguides.

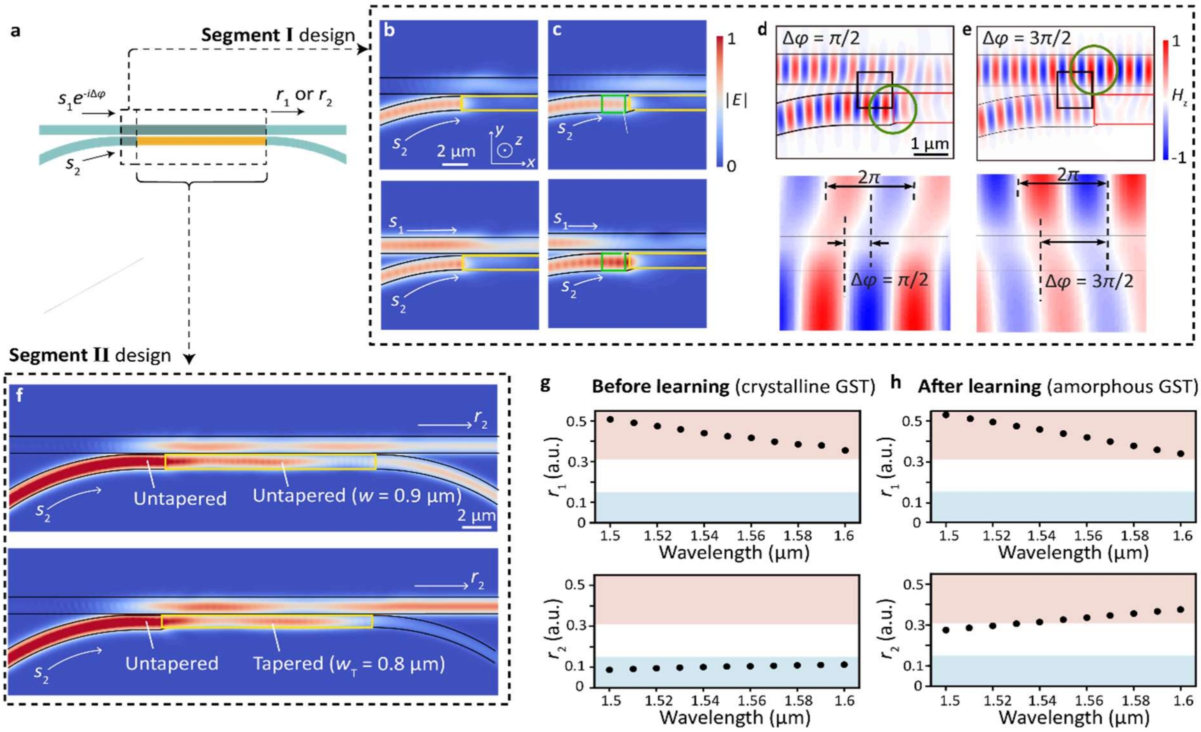

Fig. 3.3. AMLE design. a, Schematic of AMLE (boxed) which consists of segment I and II, with inputs s_1 and s_2, and outputs r_1 and r_2. GST cell in yellow. **b-e**, Segment I design. With segment I in (**c**), accumulated optical field $|E|$ at site prior to GST cell is significantly higher when *two* inputs are sent to AMLE at $\Delta\varphi = \pi/2$, compared to the case without the segment in (**b**). At $\Delta\varphi = \pi/2$ and at $\pi/2$ respectively, field tends to couple to lower waveguide (in **d**) and upper waveguide (in **e**), as indicated by the circles. Enlarged profiles (bottom plots of **d** and **e**) correspond to the boxed region in top plots. **f**, Segment II design. Field from s_2 in (**g**) is optimally coupled when GST-deposited waveguide is tapered. **g-h**, Plots of r_1 and r_2 at 1.5 - 1.6 μm wavelength range when GST is at crystalline (**g**; before learning) and amorphous (**h**; after learning) structural state. Red (blue) bands denote learned (no learned) response. The simulations are performed using the FDTD Solutions software from Lumerical Inc.

Due to the deposited GST on the bottom waveguide of segment II, the waveguide is tapered to enhance the coupling between the waveguides in the segment. The association between inputs s_1 and s_2 during the learning process occurs only when the two inputs are paired at a specific optical delay of the optical pulses Δt, resulting in a change in synaptic weight Δw between the s_1 and s_2 inputs with r_1 and r_2 outputs respectively. In the AMLE, the optical delay Δt is represented by the relative time delay introduced by the optical phase difference between the s_1 and s_2 inputs, i.e. $\Delta t = t_{s1} - t_{s2}$, where t_{s1} and t_{s2} are the times at which the respective optical field input signals s_1 and s_2 are referenced to the same optical phase.

The DiCs, made up of two optical single-mode waveguides in close proximity, allows optical energy exchange between the waveguide modes. The lower waveguide of segment I is the site where optical energy from s_1 and s_2 accumulate for GST structural phase switching to occur at the lower waveguide of segment II, thus regulating output field r_1 *and* r_2 respectively. We hereafter treat the optical modes in AMLE, which has the lossy component GST, using the coupled-mode equations

$$\frac{da}{dx} = i\kappa b \tag{3.13a}$$

$$\frac{db}{dx} = i\kappa a - \frac{\gamma b}{2} \tag{3.13b}$$

where a and b denotes normalised x-direction spatially dependent mode amplitudes of coupled upper and lower waveguides; κ is the coupling coefficient, and γ is the loss coefficient of mode b due to GST [335]. For **(3.13)** to be valid, the propagation constants between a and b has to remain the same even in the lossy region. The difference in waveguide propagation constants is compensated by tapering the waveguide with GST: this is comparable to those in passive parity-time symmetric directional couplers [338], which applies lossy material with diminishing real permittivity. Because segments I and II are respectively without and with GST, we first solve for modes in segment II and conveniently obtain the solution for segment I by simply letting $\gamma \to 0$, before cascading both matrices to solve for output field r_1 *and* r_2 with respect to input field s_1 and s_2 respectively.

For $\gamma/4\kappa \le 1$, given the $[-1, +1]$ range of a sine function, we let $\gamma/4\kappa = \sin\theta$ to arrive at

$$\begin{pmatrix} a \\ b \end{pmatrix} = \frac{e^{-\kappa x \sin\theta}}{\cos\theta} \begin{pmatrix} \cos(\kappa x \cos\theta - \theta) & i\sin(\kappa x\cos\theta) \\ i\sin(\kappa x \cos\theta) & \cos(\kappa x\cos\theta + \theta) \end{pmatrix} \begin{pmatrix} a_0 \\ b_0 \end{pmatrix} \qquad , \frac{\gamma}{4\kappa} \le 1 \tag{3.14a}$$

where a_0 and b_0 are the fields a and b at $x = 0$. For $\gamma/4\kappa > 1$, we let $\gamma/4\kappa = \cosh\theta$ given $[1, +\infty)$ range of a hyperbolic cosine function. Following the same procedure, we arrive at

$$\begin{pmatrix} a \\ b \end{pmatrix} = \frac{e^{-\kappa x \cosh\theta}}{\sinh\theta} \begin{pmatrix} \sinh(\kappa x \sinh\theta + \theta) & i\sinh(\kappa x \sinh\theta) \\ i\sinh(\kappa x \sinh\theta) & \sinh(\kappa x \sinh\theta - \theta) \end{pmatrix} \begin{pmatrix} a_0 \\ b_0 \end{pmatrix} \qquad , \frac{\gamma}{4\kappa} > 1 \tag{3.14b}$$

We readily obtain from **(3.14a)** the input-output relation of segment I by letting $\gamma \to 0$. The derivation steps from **(3.13)** to **(3.14)** are similar to those in [335,338].

To describe output field r_1 and r_2 as a function of input field s_1 and s_2, we multiply the 2×2 matrices that represent the segments: Segment I in **(3.14a)** after letting $\gamma \to 0$, and segment II in **(3.14a)** (when $\gamma/4\kappa \leq 1$) or in **(3.14b)** (when $\gamma/4\kappa > 1$). We next reduce the 2 × 2 matrix in segment II to a 1 × 2 matrix because output field $r = r_1$ or r_2 is taken from the upper waveguide of segment II.

It follows that we can relate the AMLE input and output using $r = M_{(II)} M_{(I)} s$ where $s = (s_1 \quad s_2)^T$ is the column vector that denote the respective inputs to AMLE while matrices $M_{(I)}$ and $M_{(II)}$ respectively describe the coupling tendencies in the cascaded segments of lengths $x = l_1$ and $x = l_2$:

$$M_{(I)} = \begin{pmatrix} \cos(\kappa l_1) & i\sin(\kappa l_1) \\ i\sin(\kappa l_1) & \cos(\kappa l_1) \end{pmatrix} \tag{3.15a}$$

$$M_{(II)} = \begin{cases} \dfrac{e^{-\kappa l_2 \cosh\theta}}{\sinh\theta} \big(\sinh(\kappa l_2 \sinh\theta + \theta) \quad i\sinh(\kappa l_2 \sinh\theta)\big) & , \dfrac{\gamma}{4\kappa} > 1 \text{ (before learning)} \\ \dfrac{e^{-\kappa l_2 \sin\theta}}{\cos\theta} \big(\cos(\kappa l_2 \cos\theta - \theta) \quad i\sin(\kappa l_2 \cos\theta)\big) & , \dfrac{\gamma}{4\kappa} \leq 1 \text{ (after learning)} \end{cases} \tag{3.15b}$$

in which $\theta = \sin^{-1}(\gamma/4\kappa)$ when $\gamma/4\kappa \leq 1$, and $\theta = \cosh^{-1}(\gamma/4\kappa)$ when $\gamma/4\kappa > 1$. Here, the inputs to the segments I and II are respectively at $l_1 = 0$ and $l_2 = 0$.

In **(3.15b)**, we classify $M_{(II)}$ based on the calculated ratio $\gamma/4\kappa$. This ratio can be estimated from the parameter values κ and γ retrieved based on eigenvalue splitting equation $\Delta\beta_{\pm} = \pm\, \kappa\, (1 - (\gamma/4\kappa)^2)^{1/2}$ which directly follows from **(3.13)** [335]. To retrieve the parameter values, we ran eigenmode simulations of 0.33 µm-thick silicon nitride Si_3N_4 optical waveguide on SiO_2 undercladding with nominal width of 0.9 µm and tapered width of 0.8 µm. As will be depicted later in **Fig. 3.5a**, the structural transition from nominal to tapered region is non-abrupt: it is linear within 0.5 µm transition length. From the simulations and eigenvalue splitting equation, we estimate the parameter values $\kappa = 0.157$ µm^{-1}, γ_{crys}

= 7.65κ and γ_{am} = 0.24κ , where γ_{crys} and γ_{am} are respectively the loss coefficient γ when GST is at crystalline and amorphous structural phases.

We have at this point described the parameters in the input-output relation of AMLE $r = M_{(II)} M_{(I)} s$ when *one* input is sent at a time (or when *two* non-coherent inputs are simultaneously sent, at different optical wavelengths λ). Next, we consider the case when *two* inputs of same optical field magnitude E_o and same wavelengths λ are sent to the AMLE. Here, the total optical field (from the inputs) coupled to the respective waveguides at l_1 in segment I is scaled by the product of matrix $M_{(I)}$ and column vector $(e^{-i\omega\Delta t}\ \ 1)^T$. In the column vector, $\Delta t = t_{s1} - t_{s2}$ is the relative time delay introduced by the optical phase difference $\Delta\varphi$ between the input field s_1 and s_2, where t_{s1} and t_{s2} are the times at which the respective optical field input signals s_1 and s_2 are referenced to the same optical phase. Δt is related to $\Delta\varphi$ from the relations $\Delta\varphi/2\pi = \Delta t/T$, where $T = 1/f$ is the period of the optical field. Because optical frequency $f = c/\lambda_{\text{eff}}$ where c is the speed of light in vacuum and $\lambda_{\text{eff}} = \lambda/n_{eff}$ is the effective wavelength in medium with n_{eff}, it follows that the following relations hold: $(\Delta\varphi/2\pi)\ \lambda_{\text{eff}} = c\Delta t$. Substituting the values when λ = 1.58 μm and n_{eff} = 1.59, this gives Δt = 0.825 fs when $\Delta\varphi = \pi/2$, and Δt = 2.475 fs when $\Delta\varphi = -\pi/2$ in our case.

As a result of the scaling by the column vector $(e^{-i\omega\Delta t}\ \ 1)^T$, the inputs to AMLE becomes $a_o = E_o e^{-i\omega\Delta t}$ and $b_o = E_o$. Consequently the input field coupled to the '*lower*' and '*upper*' waveguide in segment I (when '*two*' inputs a_o and b_o are both sent) are

$$\left|E_{upper}\right|_{\text{two}}^2 = \left|E_0 e^{-i\omega\Delta t} M_{(I)11} + E_0 M_{(I)12}\right|^2 = E_0^{\ 2}\left(1 - sin\left(2\kappa l_1\right) sin\left(\omega\Delta t\right)\right) \tag{3.16a}$$

$$\left|E_{lower}\right|_{\text{two}}^2 = \left|E_0 e^{-i\omega\Delta t} M_{(I)21} + E_0 M_{(I)22}\right|^2 = E_0^{\ 2}\left(1 + sin\left(2\kappa l_1\right) sin\left(\omega\Delta t\right)\right) \tag{3.16b}$$

This contrasts with the case when *one* input a_o is sent:

$$\left|E_{upper}\right|_{\text{one},a}^2 = \left|E_0 M_{(I)11}\right|^2 = E_0^{\ 2}\, cos^2\left(\kappa l_1\right) \tag{3.16c}$$

$$\left|E_{lower}\right|_{\text{one},a}^2 = \left|E_0 M_{(I)21}\right|^2 = E_0^{\ 2}\, sin^2\left(\kappa l_1\right) \tag{3.16d}$$

or when *one* input b_o is sent:

$$\left|E_{upper}\right|_{\text{one},b}^2 = \left|E_0 M_{(I)12}\right|^2 = E_0^{\ 2}\, sin^2\left(2\kappa l_1\right) \tag{3.16e}$$

$$\left|E_{lower}\right|_{\text{one},b}^2 = \left|E_0 M_{(I)22}\right|^2 = E_0^{\ 2}\, cos^2\left(2\kappa l_1\right) \tag{3.16f}$$

For the *two*-input case, **(3.16,a and b)** indicate that the field coupled to *upper* and *lower* waveguide in segment I are respectively 0 and $2E_0$ at $\kappa l_1 = \pi/4$ when $\omega\Delta t = \pi/2$. In contrast, for the *one*-input case,

(3.16,c to f) indicate that the field coupled to *upper* and *lower* waveguide in segment I are respectively 0 and E_0 when a_o is sent, and E_0 and 0 at when b_o is sent at $\kappa l_1 = \pi/2$. This implies that the beat length (length at which field from two signals is completely coupled to one of the waveguides) is $l_1 = \pi/4\kappa$ when *two* inputs are sent, and critical coupling length $l_1 = \pi/2\kappa$ when only *one* input is sent. The beat length (at $l_1 = \pi/4\kappa$) is thus half the critical coupling length (at $l_1 = \pi/2\kappa$) [335].

The fact the beat length is significantly shorter than the critical coupling length is crucial in our AMLE design. We recall that it is important to sufficiently distinguish the field accumulation at the region prior to GST cell upon *two*-input and *one*-input incidences, because AMLE associatively learns (via GST switching from field accumulation) only upon *two*-input incidence. This is achieved with segment I prior to GST cell to distinguish net optical field accumulated at the site prior to GST cell, as shown in the three-dimensional FDTD numerical simulations results in **Fig. 3.3, b-e**.

We have discussed how the associative learning process can be optically realised in our AMLE design by taking advantage of the difference between the coupling tendencies when *one*-input and *two*-input signals are sent to the device. This difference allows AMLE to associatively learn only upon *two*-input incidence. We next proceed to segment II, which regulates AMLE output based on GST structural phase difference. It is shown in **Fig. 3.6f** that tapering the bottom waveguide of segment II to match the waveguide propagation constants is essential for optimal waveguide coupling. With waveguide coupling optimised, we can regulate the AMLE output field *r*.

Having discussed the general design of the AMLE, we now delve into the AMLE design specifics. To estimate the coupling length of segment I (l_1) and segment II (l_2), we use the estimated κ and γ values obtained earlier ($\kappa = 0.157\ \mu\text{m}^{-1}$, $\gamma_{crys} = 7.65\kappa$ and $\gamma_{am} = 0.24\kappa$). On segment I, the design is based on two main considerations:

- AMLE 'associatively learns' via field accumulation at segment I (site prior to GST) only when *two* inputs are sent to AMLE
- Output *r* is measured by separately sending *one* input to AMLE at a time (or *two* inputs simultaneously when the inputs do not interfere). Regulation of output *r* is allocated to segment II which has the modulating element GST for the regulation.

We normalise the input field magnitude $E_0 = 1$, and compare the case when *two* inputs a_o and b_o are sent at $\omega\Delta t = \pi/2$ (from **(3.16b)**) and when *one* input b_o is sent (from **(3.16f)**). We determine $l_1 = 2\ \mu$m based on the above two considerations, such that '*two*-input coupling' to the site prior to GST is high (so that AMLE can 'associatively' learn when *two* inputs are sent, and not *one* input), and '*one*-input coupling' is low (so that we can allocate the regulation of output *r* to segment II). We note that by reciprocity, we can reach the same conclusion on the coupling extent (change in optical intensity $|E|^2$ as a function of l_1 relative to $|E|^2$ at $l_1 = 0$ at the lower waveguide) by plotting **(3.13d)** (when a_o is sent) instead of **(3.16f)** (when b_o are sent).

On segment II, the low '*one*-input coupling' in segment I allows us to approximate the ratio $\eta = |r_2/r_1|^2$ solely from segment II, that is $\eta \approx |M_{(II)12} / M_{(II)11}|^2$. This leads to $\eta_b \approx |sinh\ (\kappa l_2 sinh\theta_b) / sinh\ (\kappa l_2\ sinh\theta_b + \theta_b)|^2$ before learning (denoted by subscript 'b') and $\eta_a \approx |sin\ (\kappa l_2\ cos\theta_a) / cos\ (\kappa l_2\ cos\theta_a - \theta_a)|^2$ after learning (subscript 'a'), where θ_b is when $\gamma/4\kappa > 1$ and θ_a when $\gamma/4\kappa \leq 1$. We determine l_2 = 15 µm such that output r_2 upon s_2 input incidence transitions from a significantly low value ($\eta_b << \eta_a$) to that of s_1 ($\eta_a \approx 1$), with outputs due to s_1 remain relatively within the same intensity range (the difference of output intensity $R_1 = |r_1|^2$ before and after learning $|M_{(II)11b}|^2 - |M_{(II)11a}|^2 < 0.5$). The shorthand guide described above is corroborated with numerical simulations, as there may be slight deviations (for example due to slight presence of gap mode between the waveguides).

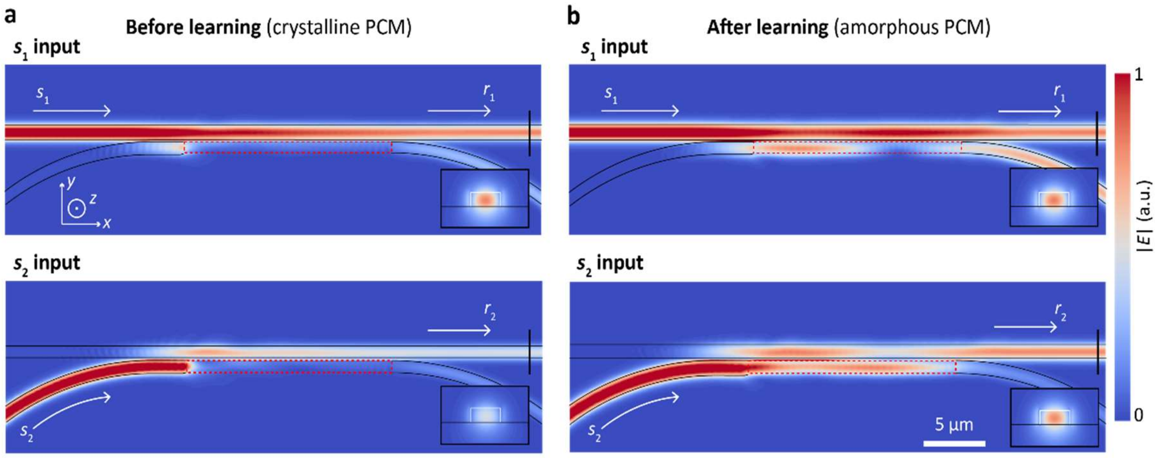

Fig. 3.4. AMLE input-output relations. Corresponding electric field profiles of AMLE before (**a**) and after (**b**) learning. Inset: Output cross-section electric field at locations denoted by black vertical bars. The colour bar is normalised. The simulations are performed using the FDTD Solutions software from Lumerical Inc.

In the AMLE design depicted in **Fig. 3.3, b-f**, we have considered the case when the optical wavelength of the inputs to AMLE is λ = 1.58 µm. To determine if the design is applicable across a broader wavelength range, the AMLE is numerically simulated using FDTD across a range of optical wavelengths from λ = 1.5 µm to λ = 1.6 µm at 10nm intervals (details in **Fig. 3.3, g** and **h**). The results indicate that the optical field tendencies in AMLE from λ = 1.54 µm to λ = 1.6 µm optical wavelength fits the learning tendencies of an associative learning device.

The design described above is verified using numerical simulations. As shown in **Fig 3.4a**, when the material is crystalline, there is no association between the inputs s_1 and s_2. However, when the two inputs (learning pulses) arrive at the same time, the material has sufficient absorption to amorphise the GST, changing the coupling between the waveguides. As the material amorphises, inputs s_1 and s_2 begin to 'associate' as shown in **Fig. 3.4b**.

3.2.2 Approaches to configure phase-dependent inputs to AMLE

The optical phase-dependence of AMLE when two inputs are sent implies the need for the optical phases to the inputs are stable. Because the slightest vibration and/or temperature change in the measurement environment can cause the optical phases to vary erratically, specific on-chip layouts are required to reduce or eliminate effects of environmental disturbances on phase control. Two different layouts are considered here: ring-based layout, and heater-based layout. Layout 1 enables the associative learning process to be observed in real-time while Layout 2 allows for the integration of multiple AMLEs using wavelength-multiplexed inputs.

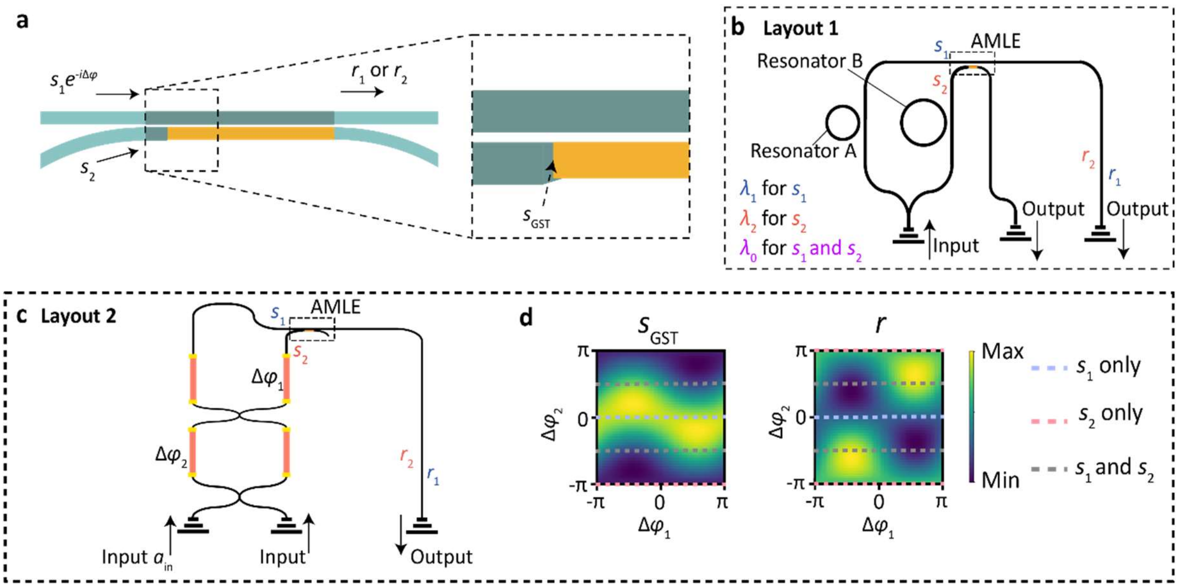

Fig. 3.5. Layouts to set Δ*φ* (*or* Δ*t*) to AMLE. a, Schematic of AMLE. Accumulated optical intensity at site prior to GST cell s_{GST} varies depending on optical phase difference between the inputs $\Delta\varphi$. **b**, Ring-based layout seamlessly integrated to AMLE. s_1 or s_2 input are sent to AMLE (dashed box) when optical signal to layout is of Resonator B (A) resonance wavelength λ_1 (λ_2). Both s_1 and s_2 inputs are sent (at $\Delta\varphi = \pi/2$) at wavelength λ_0. **c**, Heater-based layout seamlessly integrated to AMLE. Heating-induced $\Delta\varphi_1$ and $\Delta\varphi_2$ via voltage biasing on heaters modulates optical phases and amplitudes of s_1 and s_2

inputs. **d**, Plot of simulated s_{GST}, and AMLE output transmission r as a function of $\Delta\varphi_1$ and $\Delta\varphi_2$. Exact $\Delta\varphi_1$ and $\Delta\varphi_2$ to be induced is determined by measuring r.

Layout 1: Ring-based layout

To control the input signal combinations (s_1 only, s_2 only, and both s_1 and s_2) that make up the two inputs, wavelength-selective critically-coupled ring resonators are coupled to each of the inputs leading to the AMLE (shown in **Fig. 3.5, a and b**). This enables us to choose either or both the s_1 or the s_2 inputs to be active when the input to the on-chip layout is at the resonant wavelength of the bottom or top ring resonator, respectively. By using different wavelengths for s_1-only input, s_2-only input and both s_1 and s_2 inputs, simultaneous real-time monitoring of the AMLE can be carried out. The optical phase difference between the inputs vary sinusoidically across the optical wavelength spectrum when the optical path length difference between the waveguides leading to s_1 and s_2 inputs is unequal. As such, the optical phase difference to the inputs can be precisely controlled by judiciously selecting the optical wavelength to the layout.

Layout 2: Heater-based layout

Previous layout (Layout 1) enables real-time monitoring of AMLE which is important to characterise and understand the workings of AMLE. However, in many practical cases, multiple AMLEs need to be integrated on a single platform using multiple inputs. The use of rings with different wavelengths depending on AMLE inputs to be activated impedes such integration. Crucially, we require the use of same wavelength for both inputs to enable light accumulation from both inputs to the region prior to GST. This is because the optical wavelength of the two inputs to AMLE has to be equal for the pump/learning process. We use cascaded Mach-Zehnder Modulators (MZM) to enable wavelength-multiplexed inputs on large-scale platform because MZMs provide a reliable means to split the signals (to layout) equally with stable optical phases to the AMLEs.

The inputs to AMLE are controlled by heating (via voltage biasing) the thermo-optic heaters placed on two MZM waveguides to enable on/off regulation of input fields, and fix $\Delta\varphi$. The schematic of the layout is shown in **Fig. 3.5c**. Voltage biasing across the heater induces optical shift on top MZM $\Delta\varphi_1$ and bottom MZM $\Delta\varphi_2$ that respectively defines the optical phase difference to AMLE $\Delta\varphi$ (or optical delay of the phases Δt) and amplitudes of inputs s_1 and s_2. The exact modulation is determined by probing the output from grating coupler, which corresponds to the optical field of interest (that is, s_{GST} and r in **Fig. 3.5a**, **b**).

By applying the usual 1/√2 and j/√2 field splitting steps [339], we can express the optical field at the region prior to GST cell s_{GST}, and AMLE optical output field r, both as functions of input field to the layout a_{in}, as

$$s_{\mathrm{GST}} = \frac{j}{2} e^{j\beta(L_{12})} \left(e^{j\beta L_{\Delta}} \left(e^{j\Delta\varphi_2} - 1 \right) sin\,\kappa l_1 + e^{j\Delta\varphi_1} \left(e^{j\Delta\varphi_2} + 1 \right) cos\,\kappa l_1 \right) a_{\mathrm{in}} \tag{3.17a}$$

$$r = \frac{e^{-\kappa l_2}\, cosh\,\theta_{\mathrm{b}}}{2\, sinh\,\theta_{\mathrm{b}}} \left(e^{j\beta(L_{12}+L_{\Delta})} \left(e^{j\Delta\varphi_2} - 1 \right) P - e^{j\beta(L_1+L_2)+j\Delta\varphi_1} \left(e^{j\Delta\varphi_2} + 1 \right) Q \right) a_{\mathrm{in}} \tag{3.17b}$$

where $P = sinh\left(\kappa l_2\, sinh\,\theta_{\mathrm{b}} + \theta_{\mathrm{b}}\right) cos\,\kappa l_1 - sinh\left(\kappa l_2\, sinh\,\theta_{\mathrm{b}}\right) sin\,\kappa l_1$

$$Q = sinh\left(\kappa l_2\, sinh\,\theta_{\mathrm{b}} + \theta_{\mathrm{b}}\right) sin\,\kappa l_1 + sinh\left(\kappa l_2\, sinh\,\theta_{\mathrm{b}}\right) cos\,\kappa l_1$$

L_{12} and L_{Δ} respectively denote the sum of nominal arm lengths of both MZMs, and the remaining arm length of the top MZM. **Fig. 3.5d** plots **(3.17)**. From the figure, when r is minimum, accumulated optical intensity s_{GST} at site prior to GST cell is maximum (which is when $\Delta\varphi = \pi/2$). By simply measuring r, we can determine the exact $\Delta\varphi_1$ and $\Delta\varphi_2$ to be induced.

3.2.3 Design of layout components

The layouts consist of three main components external to the AMLE:

- **On-chip coupler** to couple light from fibre array to on-chip layout and vice versa.
- **Optical resonator**, in Layout 1 to activate the input(s) to the AMLE
- **Multimode interferometers and integrated waveguide heaters**, in Layout 2. These are essential to MZMs which modulate the input(s) to the AMLE

On-chip coupler

As the inputs to AMLE are not supplied on-chip, the measurement of AMLE requires efficient interface to enable optical signal to be sent to and from the on-chip layouts. The methods to achieve this can generally be classified into end-fire [340-357], adiabatic [358-362], diffraction gratings [363-372] and prism coupler [373-377] methods. They are compared in **Table 3.2**.

Table 3.2. Comparison of on-chip coupling approaches.
Factor taken into special consideration is framed in green.

Method	Coupling efficiency	Coupling bandwidth	Alignment tolerance	Polarisation tolerance	Spatial density
1. End-fire	**Medium** To enhance coupling, inverted tapers (efficiency ~ 3 dB or 50 % [352]), 2D spot size converters (efficiency ~ 1 to 3 dB, or 50 % to 80 % [353-355]), trident structures (efficiency ~ 3 dB or 50 % [356]), subwavelength gratings (efficiency ~ 1 dB or 80 % [378]), heterogenous cladding (efficiency ~ 0.6 to 3 dB or 50 to 88 % [342,346]) and vertically-curved cantilever waveguides (efficiency ~ 1.6 dB or 69 % [347]) have been used.	**High** (~ 100 nm – 1 µm [342,348-350,352,354-356,378])	**Low** ~ 0.1 - 5 dB/µm [342,350,352,356,378]. Sensitive to both position and incident angle of input signal. Due to reduced on-chip waveguide to fiber optical mode mismatch, alignment tolerance typically increases with coupling enhancement schemes (such as 2D spot size converters [348,354,355]) . Simultaneous multiple port coupling can be achieved using glass interposer [351] or photonic wire bonding [343 344-346,357].	**High** Generally insensitive to TM or TE polarisation [349,350,353,355].	**High** Facet size typically varies from several to tens of µm^2 [349,350,352-356,378]. Spatial density typically inversely related to coupling efficiency.
2. Adiabatic (evanescent)	**High** ≳ 0.22 dB (95 %) [359,360] Optical signal is seamlessly coupled to on-chip waveguide (in a directional coupler-like manner) via evanescent field.	**Low** ~ 20 nm [361]	**Medium** ~ 0.02 dB/µm along waveguide optical axis, ~ 0.5 dB/µm perpendicular to waveguide optical axis [359,360].	**High** Fibre taper adiabatically coupled to on-chip waveguide is not polarisation-sensitive [362].	**Medium** ~ several to tens of µm^2 to hundreds of µm^2 [358-360].
3. Diffraction gratings	**Medium** Uniform gratings can provide ~ 2.5 dB (55 %) efficiency [364]. To enhance coupling, gratings are apodised (efficiency ~ 0.8 dB or 83 %) or directionality – enhanced using overlays [365], interleaves [366,367] or reflectors [368,369] (efficiency ~ 0.5 to 1.6 dB, or 70 to 90 %).	**Low** ~ 20 to 50 nm [364,365,369]. Approach relies on wavevector-matching, which is wavelength dependent.	**High** Insensitive alignment requirement due to broad grating width (typically ~ 10 µm which is comparable to conventional fibre mode field diameter).	**Low** Incident angle varies depending on TM and TE modes, as these different modes have different wavevectors. Polarization-insensitive coupling approaches (such as non-uniform gratings [370,371] and 2D gratings [372]) typically come with tradeoffs in coupling efficiency (~ 7 dB, or 20% [370,372]) or structural	**Medium** Grating area is typically ~ 100 µm^2 [366,369-372].

				complexity (backside metal mirror [371]).	
4. Prism coupler (evanescent)	**High** ≳ 0.45 dB or 90 % [375]. Precise control of gap required for optimal coupling.	**Low** Approach relies on wavevector-matching, which is wavelength dependent.	**Medium** Sensitive to incident angle (~ 15 dB/deg [376]) of input signal. Angular aperture of ≲ 10^{-3} rad required [377].	**Medium** Incident angle varies depending on TM and TE modes, as these different modes have different wavevectors.	**Low** Ideal gap of 0.4 mm and length of 1 cm [375] for optimal coupling.

End-fire and adiabatic coupling approaches can enable high coupling efficiencies at broad optical bandwidth, but they have tight alignment tolerances. On the other hand, diffraction grating and prism coupling approaches are less alignment-sensitive with varying coupling efficiency. In the work described in this thesis, diffraction grating couplers are adopted in the layouts due to convenience - it has high alignment tolerance, is in-built onto waveguide structures and can thus fabricated in the same fabrication run.

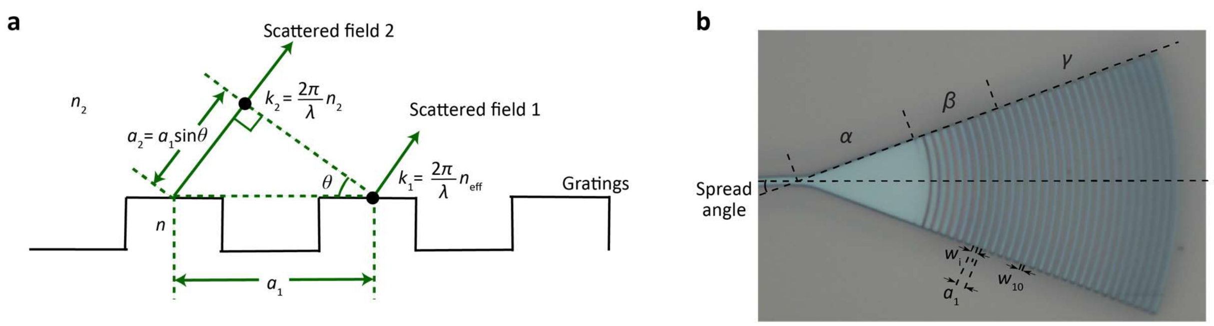

Fig. 3.6. Grating coupler. a, Schematic representation. Diffraction period a_1 chosen such that scattering from individual gratings constructively interfere at a certain spatial angle, that is when the optical phase difference $k_1a_1 - k_2a_2$ are integer multiples of 2π. **b**, Top-view schematic of apodised grating coupler, with period a_1. Segment β is apodised with increasing width w_i while segment γ has fixed width w_{10}. Apodisation enables a more Gaussian beam-like field scattering [379].

As depicted in **Fig. 3.6a**, grating couplers rely on gratings (of period a_1) to couple light into or out of on-chip optical components. The scattered field due to the gratings supports a plane wave-like optical field when they constructively interfere. This happens when the path length difference between the scattered field $k_2a_2 - k_1a_1$ are integer multiples of 2π:

$$k_2 a_2 - k_1 a_1 = 2m\pi \qquad , \quad m = 0, \ \pm 1, \ \pm 2 \tag{3.18}$$

where k_1 and k_2 are wavevectors of optical field at grating period a_1 and distance a_2. The field at a_1 and a_2 are respectively on gratings and in free-space (with effective refractive index n_{eff}). Consequently,

$$a_1 = \frac{m\lambda}{n_{\text{eff}} - n_1 \sin\theta} \tag{3.19}$$

The coupling efficiency can be optimised only for a specific polarisation and wavelength range due to its phase matching dependence (TM and TE waveguide modes have difference n_{eff} [306,339]). It is conventional to apply apodisation (as in **Fig. 3.6b**) to suppress backreflections [380].

Optical resonator

In Layout 1, two resonators are placed on the paths leading to the input(s) to the AMLE. The inputs are activated when the resonator is not of the resonance wavelength. From the optical resonator comparison summarised in **Table 3.3**, WGM resonators are straightforward to design, has high resistance to backreflection, and has high wavelength selectivity. Due to these factors, these resonators, in the form of ring resonators, are adopted in design of Layout 1.

The coupling between the waveguide and resonator can be simplified and analysed by representing resonator mode a and optical waveguide mode a_{in} in the following equations:

$$\frac{da}{dt} = -\left(i\omega_0 + \frac{\kappa_0 + \kappa_{\text{ex}}}{2}\right)a + \sqrt{\kappa_{\text{ex}}}\, a_{\text{in}} \tag{3.20}$$

$$a_{\text{out}} = 1 - \sqrt{\kappa_{\text{ex}}}\, a \tag{3.21}$$

after taking into consideration energy conservation and time-reversal symmetry [381]. ω_0 is the angular frequency when the wavelength is at resonance, κ_0 and κ_{ex} are the optical losses due to the resonator, and the waveguide-resonator coupling (a function of resonator-waveguide gap) respectively. The resonator—waveguide system is depicted in the schematics in **Fig. 3.7a**.

Table 3.3. Comparison of optical resonators.
Factor taken into special consideration is framed in green.

Resonator for on-chip filtering purpose	Design simplicity	Resistance to backreflection	Wavelength selectivity	Spatial density
1. Fabry-Perot (a special case of PhC resonator)	**Medium** For nanobeam approach, judicious design of hole dimensions required [382]. For Bragg grating approach, mirrors can be designed using Bragg reflectors [383].	**Low** Uncoupled signal is backreflected [154]	**High** Q ~ 1 million using nanobeam approach [382] and Q ~ 3.1k using Bragg grating approach [383] on SOI platform	**Medium** ~ 100 μm^2 [382] ~ 15 μm^2 [383]
2. WGM	**High** In the simplest case, simply requires a waveguide coupled to a microring resonator (as microdisk resonator supports higher-order radial modes with increasing radius) [384].	**Very high** Uncoupled signal simply propagates through the waveguide [384] Scattering due to fabrication imperfections on resonators [385] fabricated using standard microfabrication techniques unlikely to induce significant backreflection	**High** Resonator with ultrahigh quality factor Q (e.g., Q ~ 200 million [158]) requires combination of standard microfabrication techniques with selective reflow process [386]. Resonators fabricated using standard microfabrication techniques can have Q ~ 20k - 35k (with radius from 50 μm to 100 μm) on SOI platform [387], and Q ~ 60k (with radius of 80 μm) on silicon nitride platform [388]	**Low** For Q ~ 200 million, requires ~ 30 mm^2 For Q ~ 50k, requires 0.04 mm^2

By taking the resonator field as a plane wave $a = e^{-i\omega t}$, the derivative d/dt tends to $d/dt \rightarrow -i\omega t$ at steady-state. The transmission at waveguide output port $T = |a_{out}|^2$ can be written as

$$T = \left| 1 - \frac{\kappa_{ex}}{\frac{\kappa_0 + \kappa_{ex}}{2} - i(\omega - \omega_0)} \right|^2 \tag{3.22}$$

Critical coupling $T = 0$ can be achieved at the resonance angular frequency $\omega = \omega_0$, and when the losses are equal $\kappa_0 = \kappa_{ex}$. In general, for a certain mode number (number of field nodes), κ_0 decreases (higher quality factor) as the resonator radius increases. These are considered in the Layout 1 resonator design shown in **Fig. 3.7b**. On matching the resonator-waveguide propagation constants, the conditions involve transcendental equations provided in [389].

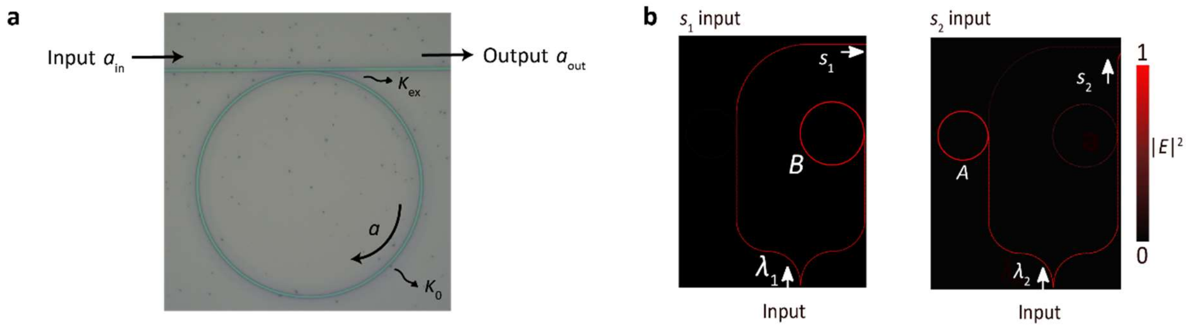

Fig. 3.7. Ring resonator-waveguide system. a, Optical micrograph of system. Field coupling between waveguide and resonator is governed by the matching of field losses (coupling loss κ_{ex} and resonator loss κ_o) and propagation constants (in the resonator and waveguide). These factors, along with the resonator resonant wavelengths, vary depending on the resonator radius. Waveguide is tapered at the coupling region to match the propagation constants. **b**, Profile of simulated optical intensity $|E|^2$ when input to layout is at λ_1 (left) and λ_2 (right), which activates s_1 (s_2) input to AMLE.

Multimode interferometers and heaters

In Layout 2, two cascaded Mach-Zehnder Modulators (MZMs) are used to modulate the inputs to the AMLEs. The MZMs consist of 2 input × 2 output multimode interferometer (MMI) structure to combine or split optical signals, with heaters to electrically modulate the MZMs. The MMIs are designed to split one input signal equally to two output signals, or combine two input signals of equal amplitudes (at certain optical phase difference) to one output signal (details in **Table 3.1**).

The positions of input waveguides determine the types of modes that can be excited [332]. Positioning the input waveguide at half the MMI width $W/2$ would excite symmetric/even modes (mode number m = 0, 2, 4, 6, …), while positioning it at $W/3$, or $2W/3$ would excite pairwise modes (m = 0-1, 3-4, 6-7, …). It is not possible to spatially position both input waveguides at $W/2$, thus the waveguides are positioned at $W/3$ and $2W/3$ in the MMI design. As it is sufficient to combine and split optical signals with an MMI that supports modes m = 0 and 1 only (without having to consider other modes which could complicate our analysis), the MMI width W is designed to support only these modes. The excitation of the two modes (m = 0 and 1) at the same optical wavelengths indicate that they can interfere/superimpose with one another. The 'spatial period' of such superimposed mode is known as beat length $L_\pi = \pi/(\beta_0 - \beta_1)$. As summarised in **Fig. 3.7a**, the propagation constant β_m of mode m can be retrieved from the dispersion equation that relates the lateral wavevector is $k_{wm} = (m+1)/W$, and wavevector in MMI $k_0 n$:

$$k_{wm}^2 + \beta_m^2 = k_0^2 n^2 \tag{3.23}$$

Applying Binomial expansion with $k_{wm}^2 \gg k_0^2 n^2$:

$$\beta_m \simeq k_0 n - \frac{(m+1)^2 \pi \lambda_0}{4n(W_{GH}+W)^2} \tag{3.24}$$

and plugging m=0 and m=1 into the equation, we retrieve:

$$L_\pi \simeq \frac{4n(W_{GH}+W)^2}{3\lambda_0} \tag{3.25}$$

where W is the MMI width and the increase in effective MMI width due to Goos-Hanchen shifts W_{GH} – comparable to **(3.5)** – is a function of mode polarisation (σ = 0 for TE, and σ = 1 for TM mode):

$$W_{GH} = \frac{\lambda_0}{\pi\sqrt{n^2 - n_{\text{cladding}}^2}} \left(\frac{n_{\text{cladding}}}{n}\right)^{2\sigma} \tag{3.26}$$

For the field in MMI to split equally when an input is sent to the 2×2 MMI, the MMI length has to be $L_\pi/2$. $W \sim 5.5$ µm is chosen to provide sufficient room for gap between the input and output waveguides positioned at $\sim W/3$ and $2W/3$. Although such large width supports about five modes, excitation at $\sim$ 1.55 µm effectively excites only the first two modes. Then, based on shorthand calculations from the above equations (W_{GH} = 0.26 µm and W = 5.8 µm) to yield $L_\pi/2 \sim 25$ µm), 3D numerical simulations of the structure is carried out. The results, shown in **Fig. 3.8c**, indicate field splitting using MMI. By reciprocity, the reverse applies - two signals (at certain optical phase difference) can be combined to a single output at $L_\pi/2$. The design is applied in MZMs on Layout 2 to split and combine signals to and from the MZM arms.

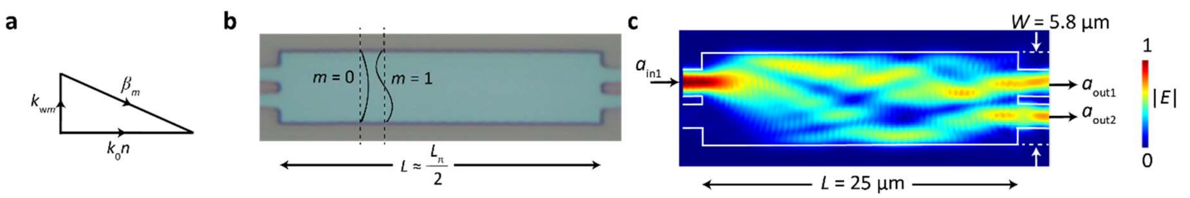

Fig. 3.8. 2×2 MMI components. a, Dispersion relation to retrieve propagation constant β_m from lateral wavevector k_{wm} and wavevector in MMI core $k_0 n$ (of refractive index n). **b**, Optical micrograph of MMI to highlight the excitation of m = 0 and 1 to split or combine signals. **c**, Simulated plot of cross-sectional field in MMI (as a splitter). The simulations are performed using the FDTD Solutions software from Lumerical Inc.

As will be described later in **Chapter 6** of the thesis, 1×2 MMI splitters are used in the scaled-up associative learning hardware. In the splitter, the input waveguide is placed at half the MMI width $W/2$ which excites symmetric/even modes (mode number m = 0, 2, 4, 6, …) [332]. The MMI is designed such that the input excites mode m = 0 and 2, as depicted in **Fig. 3.9a**. These modes interfere/superimpose with one another. Plugging m=0 and m=2 into **(3.24)**, we retrieve:

$$\beta_0 - \beta_2 \simeq \frac{8\pi}{3L_\pi} \tag{3.27}$$

Because it is unlikely for modes m = 0 and m = 2 to couple with one another, the field in the MMI can be adequately represented by a linear combination of guided modes:

$$\begin{aligned} \psi(x,y) &= c_0\psi_0(y)e^{j(\omega t-\beta_0 x)} + c_2\psi_2(y)e^{j(\omega t-\beta_2 x)} \\ &= e^{j(\omega t-\beta_0 x)}\left(c_0\psi_0(y) + c_2\psi_2(y)e^{j(\beta_0-\beta_2)x}\right) \end{aligned} \tag{3.28}$$

where c_0 and c_2 are normalised constants of guided modes Ψ_0 and Ψ_2. By substituting $\beta_0 - \beta_2$ into **(3.28)**, it is clear that the cross-sectional field at $L = 3L_\pi/4$ is identical to the input. The field is split equally at $L = 3L_\pi/8$. For $W \sim 5$ µm, the 50:50 splitting length is estimated to be at $3L_\pi/8 \sim 15$ µm. This is corroborated in the 3D numerical simulation results shown in **Fig. 3.9b**. By reciprocity, two signals (at certain optical phase difference) can be combined to a single output using the same design.

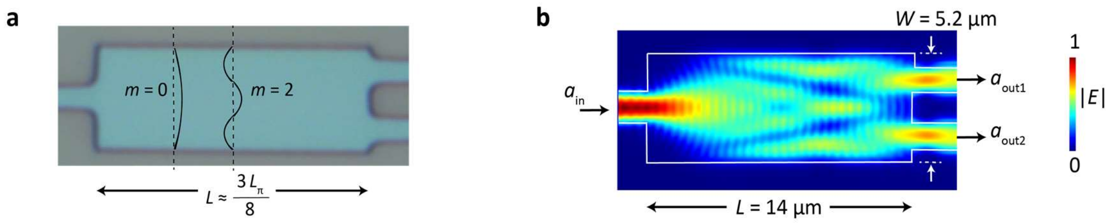

Fig. 3.9. 1×2 MMI splitter. a, Optical micrograph of MMI to highlight the excitation of m = 0 and 1 to split or combine signals. **b**, Simulated plot of cross-sectional field in 1×2 MMI (as a splitter).

On the MZM heaters, voltage applied across the heater deposited on one of the MZM arm introduces change in effective refractive index of arm waveguides to introduce optical phase shift essential for field modulation. The heating is enabled by the high thermo-optic coefficient $dn/dT = 2.51 \pm 0.08 \times 10^{-5}$ K^{-1} at 293 K (room temperature) [390] and fairly high thermal conductivity K = 43 Wm^{-1} K^{-1} [391] of Si_3N_4, and high resistivity ρ = 1100 nΩ m [392] of NiCr alloy. As the heaters are typically optically lossy (due to their negative relative permittivities $\epsilon < 0$), a layer of SiO_2 (which can thermally conduct

$K = 1.4$ W m^{-1} K^{-1} [393]) is added as isolation layer to reduce the losses. A balance between layer thickness and electrical driving voltage is required to effectively modulate the MZMs.

3.3 Fabrication of AMLE

Based on the design described in the previous section, the layouts seamlessly integrated to AMLE (**Fig. 3.10a**) are fabricated using the equipment shown in **Appendix A**. The geometrical dimensions for each fabrication layer are designed using the script-based functionalities of computer-aided design (CAD) software AutoCAD [394] and gdsHelpers [395] to enable convenient sweep of specific geometries by programming a loop counter. Libraries are created to easily call functions with basic geometrical shapes and elements.

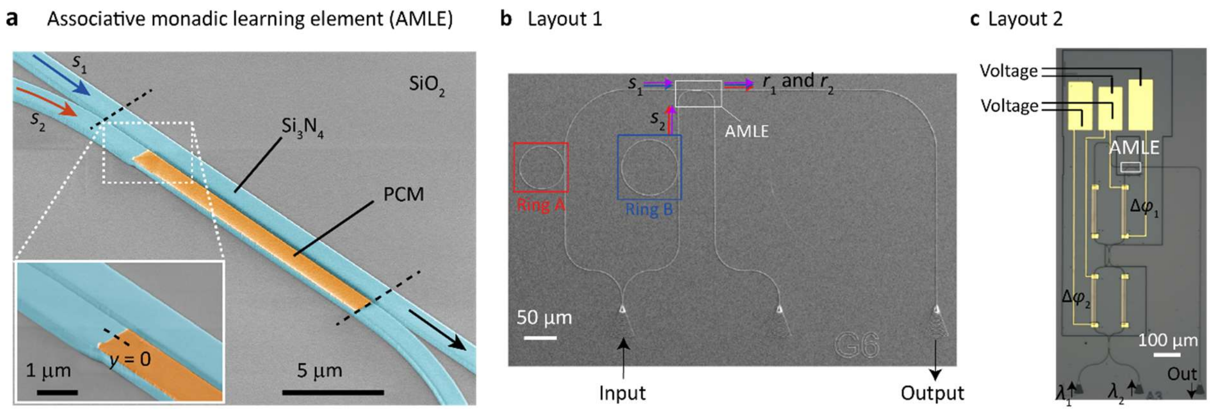

Fig. 3.10. AMLE seamlessly integrated to the layouts. a, SEM image (false coloured) of a fabricated AMLE, consisting of Si_3N_4 directional couplers (cyan), SiO_2 undercladding (grey) and GST cell (marigold). Inset shows detail of coupling area with GST. **b**, SEM image of the ring-based layout. Main components in boxes. **c**, Optical micrograph of heater-based layout.

The fabrication of AMLE on **Layout 1: Ring-based layout** (**Fig. 3.10b**) follows the following sequence:

a. Fabrication of markers and frames.

Gold markers are used to align the different fabrication layers: it can be viewed clearly on electron beam lithography writing tool (JEOL 5500FS; as gold is electrically conductive) and mask aligner (Karl Suss MJB3). Gold frame is fabricated in the same layer to ease identifying the devices on the sample. The steps to fabricate these are:

i. Si_3N_4 substrate (330nm Si_3N_4/3.3µm SiO_2) is cleaned.

ii. The substrate is spincoated and pre-baked with positive resists 495 PMMA at 4% and 950 PMMA at 2% (MicroChem Corp.) at 150°C for 5 and 15 minutes respectively.
iii. The substrate is patterned using electron beam lithography (JEOL 5500FS) at 50kV.
iv. The patterned resist is developed in IPA/MIBK:EMK (15:5:1; MicroChem Corp).
v. Au is deposited via thermal evaporation (Edwards 306 Vacuum Coater/Deposition Systems).
vi. The remaining resist is removed by Remover PG (MicroChem Corp.).

b. Fabrication of silicon nitride Si_3N_4 structure.

To define the Si_3N_4 waveguide structures, the following steps are carried out:
i. Si_3N_4 substrate (330nm Si_3N_4/3.3µm SiO_2) is cleaned
ii. The substrate is spincoated and pre-baked with Ti-Prime adhesion promoter (Microchemicals GmbH) at 120°C for 2 minutes, and ma-N2403 negative resist (microresist technology GmbH) at 90°C for 2 minutes.
iii. The substrate is patterned using electron beam lithography (JEOL 5500FS) at 50kV.
iv. The patterned resist is developed in MF-319 developer (Dow Chemical Company).
v. Reactive ion etching (PlasmaPro 80, Oxford Instruments) is performed in $CHF_3/O_2/Ar$ to fully etch the exposed Si_3N_4.
vi. The remaining resist is removed by mr-Rem 700 (micro resist technology GmbH).

c. Deposition of GST/ITO patches.

To define the GST cell with a thin ITO layer, the following steps are carried out:
i. The substrate is cleaned.
ii. The substrate is spincoated and pre-baked with positive resists 495 PMMA at 4% and 950 PMMA at 2% (MicroChem Corp.) at 150°C for 5 and 15 minutes respectively.
iii. Similar process as in a.iii. is performed to define GST/ITO patch patterns.
iv. The patterned resist is developed in IPA/MIBK:EMK (15:5:1; MicroChem Corp).
v. The substrate is sputtered with 10nm GST/10nm ITO using sputtering tool (Nordiko RF Sputter Deposition tool, Nordiko Technical Services Ltd.).
vi. The remaining resist is removed by Remover PG (MicroChem Corp.).

The fabrication of AMLE on **Layout 2: Heater-based layout** (**Fig. 3.10c**) follows the following sequence:

a. Fabrication of markers and frames.

Gold markers and frames enable good alignment and convenience to locate the devices on the layout. The steps to define the markers and frames are the same as those for Layout 1.

b. Fabrication of silicon nitride Si_3N_4 structure.

The optical waveguide structures on the layout has a Si_3N_4 core, with SiO_2 undercladding and air overcladding. To define the Si_3N_4 waveguide structures, steps defined for Layout 1 are carried out.

c. Deposition of SiO_2 isolation layer.

The SiO_2 isolation layer is deposited onto the Si_3N_4 waveguiding structure to create a gap between the waveguide on the MZM and the NiCr heaters used for the modulation. This gap is necessary to significantly reduce the optical loss from the NiCr metal alloys.

The steps to deposit the isolation layer are as follows:

i. The substrate is cleaned as in a(i).

ii. The substrate is spincoated and pre-baked with positive photoresist S1813 (Microresist Technology GmbH) at 120°C for 1 minute each.

iii. Photolithography is performed using mask aligner Karl Suss MJB3 (SUSS MicroTec SE) to define SiO_2 isolation layer patterns.

iv. The patterned resist is developed in MF319 (Microresist Technology GmbH).

v. SiO_2 layer (1 µm) is deposited onto the substrate using sputtering tool (Nordiko RF Sputter Deposition tool).

vi. The remaining resist is removed by mr-Rem 700.

d. Deposition of NiCr heater.

A layer of NiCr heater enables heating on the Si_3N_4 MZM waveguides, which induces optical phase shift on MZM waveguide arm to modulate the optical signals. In this fabrication layer, steps (i) to (iv) in b are repeated before depositing 200 nm NiCr using sputtering tool (Nordiko RF Sputter Deposition tool).

e. Deposition of Au (gold) electrodes.

The NiCr heaters require on-chip electrical pads so that electrical voltage across the heaters can be applied. In this fabrication layer, steps (i) to (iv) in c are repeated before depositing Au via thermal evaporation (Edwards 306 Vacuum Coater/Deposition Systems).

f. Deposition of GST/ITO patches.

The memory cell on the AMLE waveguides is GST with a thin ITO layer to prevent oxidation. To define the patches, the same steps as those for Layout 1 are carried out.

Note: In realising the heater-based layout, difficulties were faced in depositing the SiO_2 deposition layer. It is conventional for Si-based structures to deposit SiO_2 layer before buffered oxide etching the exposed oxide layer without affecting the Si layer. However, this is challenging for Si_3N_4-based structures (as in this work) as SiO_2/ Si_3N_4 etching selectivity is not substantial [396]. Thus, windows are opened on a thick S1813 photoresist to deposit the SiO_2 layer instead. As the deposition rate of SiO_2 is typically slow and requires high RF power (~ 4 nm/min at 150 W), the lift-process has been challenging, presumably due to the heat from sputtering process which 'bakes' the resist. It requires at least 4 days of heating at 80 °C in mr-Rem 700 for complete resist removal.

Reactive-ion etching (RIE) was attempted as well - using 3 µm-thick post/hardbaked (at 100 °C for 90 seconds) for both 495K and 900K A4 PMMA as etching mask to selectively etch SiO_2, but this was found to introduce high waveguide transmission losses. It was also not possible to use thicker PMMA resist (A8) as the resist tends to crack at the baking temperature. Due to cleanroom restrictions that forbade etching samples with chromium (to ensure RIE chamber remains 'uncontaminated'), no attempts were made to use chromium hardmask, although it could be a viable alternative [397].

On the modulation efficiency of integrated waveguide heaters on silicon nitride (Si_3N_4) platform, the temperature sensitivity of Si_3N_4 is $\sim 2.5 \times 10^{-5}$ K^{-1}. This is about a magnitude lower than the temperature sensitivity of silicon (which is $\sim 2 \times 10^{-4}$ K^{-1}). As a result, higher voltage is required to modulate the signals via the integrated heaters. Thermal isolation can be applied to reduce the modulation voltage [398].

We have studied the design of the optical associative learning device. Starting from the concept of Pavlovian associative learning, a design of the optical associative learning device is presented. A problem to observing the learning process is identified, and two different complementary layouts are proposed to solve the problem. Different choices of materials and approaches to realising the layout are carefully considered. Finally, the steps to experimentally realise the device and layouts are laid out. Problems encountered during the fabrication steps are highlighted.

Chapter 4

4. Experimental setup for AMLE characterisation

This chapter introduces the measurement setups to approach the erratic fluctuation of optical phases due to measurement environment, and describes in detail the equipment in the setups.

4.1 Measurement setup

The equipment configured to probe the AMLE state, and to induce learning process in the AMLE are described in this section.

4.1.1 Setup to probe AMLE state

It is important to probe the AMLE state to determine if the learning process occurs. In the probe setup, continuous-wave (CW) lasers provide the essential optical signals for the optical probing process. The lasers are coupled via single mode fibre to the polarisation controller where signal coupling to on-chip layout - which is polarisation dependent (due to the grating couplers [306,339]) – is optimised.

To couple signal to the layout, a fibre array is used. It consists of equispaced fibres slanted at an angle (8° in our case) for optimal coupling to the layout. The layout is configured such that the grating coupler pitch matches the fibre array pitch to enable coupling of signal into- and out of the layout. The output port of fibre array is then connected to a photodetector (PD) to detect the incident output optical signal. The PD is electrically connected to the data acquisition unit (DAQ) to analyse the results on computer.

To enable fibre array to be aligned with the on-chip grating couplers, a motorised stage is required. The stage is automatically controlled using a computer user interface. The location of grating couplers relative to fibre array ports is monitored using optical microscope and camera and viewed using a dedicated computer software.

4.1.2 Setup to induce AMLE learning

As the AMLE state transitions only when specific optical signal is sent, a separate (pump) setup to induce AMLE learning is required. In the setup, CW laser provides the optical signal. It is coupled via single mode fibre to the polarisation controller (PC) to optimise signal coupling to electro-optic modulator (EOM; which is polarisation-dependent [399]). The EOM receives electrical pulse from arbitrary function generator (AFG) or pulse generator, and transfer pulse shape it receives electrically to the optical pulse. A small fraction (in the author's case, 1% or 10%) of the optical pulse is sent to high-speed PD, which is connected to the oscilloscope to be analysed. The remaining fraction is amplified by erbium doped fibre amplifier (EDFA), spectrally filtered by optical tunable filter (OTF; to eliminate broadening effects from amplifier gain spectrum [400,401]), and sent to PC to ensure optimal coupling to the layout via fibre array.

4.1.3 Integrated setup

Although the setups (introduced in previous sections) can be separately configured, it is convenient to integrate them. This ensures smooth running of measurements without unnecessary interruptions.

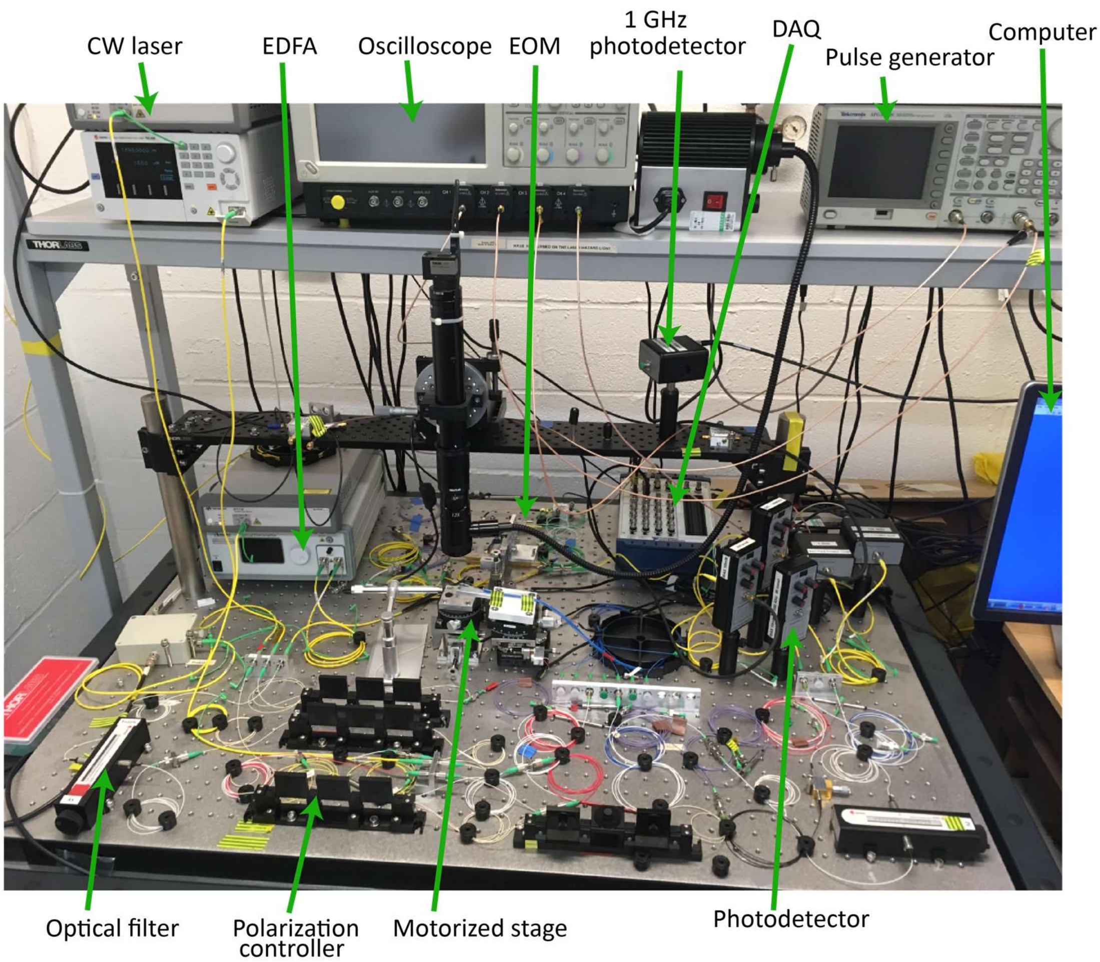

Fig. 4.1. Image of pump-probe integrated setup. Main components labelled.

There are at least two ways to integrate the pump and probe setups:

- Send pump and probe signals at different wavelengths to layout. This enables real-time probing of AMLE without signal interference. The ring-based layout (Layout 1) adopts this approach
- Use optical switch to temporally activate either pump or probe lines (but not both). This is particularly useful when we require using the same wavelengths for the signals (which would otherwise introduce signal interference). A downside of this approach is loss of real-time information. The heater-based layout (Layout 2) adopts this approach

Fig. 4.1 shows the integrated setup that forms the basis for other setups (with different configurations).

4.2 Optical measurement of AMLE (Layout 1: ring-based layout)

SEM image of the ring-based layout seamlessly integrated to AMLE, is shown in **Fig. 4.2**. The inputs to AMLE are regulated by two ring resonators (rings A and B). $\Delta\varphi$ can be precisely defined because the input signal to layout is split from a common point.

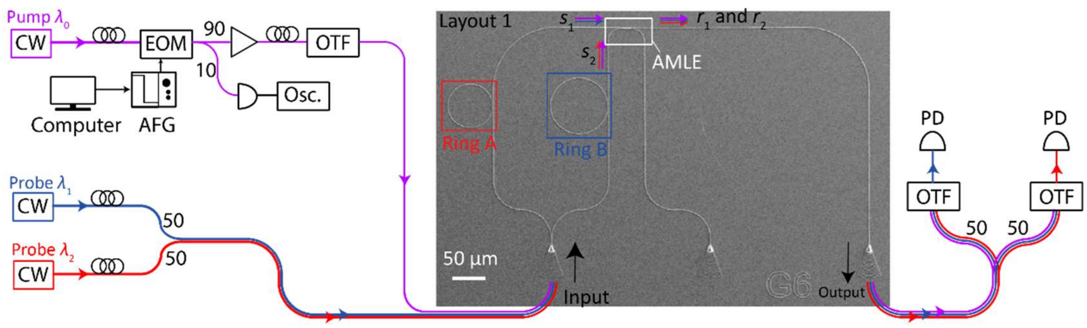

Fig. 4.2. AMLE measurement on ring-based layout (Layout 1). Schematic of experiment setup to measure AMLE. Two probe lines operating at wavelengths λ_1 (blue) and λ_2 (red), and one pump line at λ_0 (purple). λ_1 (λ_2) correspond to ring B (A) resonant wavelengths, to regulate s_1 (s_2) input to AMLE. Output readouts monitored in real-time.

The experiment setup to measure AMLE consists of two probe lines operating at wavelengths λ_1 (blue) and λ_2 (red), and one pump line at λ_0 (purple). Light sources are CW lasers (probe: N7711A, Keysight Tech., pump: TSL-550, Santec Corp.) operating at low power and coupled to single-mode fibres. Signal output is split at 50:50 ratio, filtered by optical tunable filter (OTF; OTF-320, Santec Corp.) respectively at λ_1 and λ_2, and detected by 200 kHz photodetector (PD; 2011-FC, Newport Spectra-Physics Ltd.). Pump signal, which induces the learning process, is modulated by an electro-optical modulator (EOM; 2623 NA, Lucent Tech.) via arbitrary function generator (AFG; AFG 3102C, Tektronix). The modulated signal is amplified by an erbium doped fibre amplifier (EDFA; AEDFA-CL-23, Amonics Ltd.) while monitored (on remaining 10%) using 1 GHz PD (1611FC-AC, Newport Spectra-Physics Ltd.) connected to an oscilloscope (TDS 7404, Tektronix).

4.3 Optical measurement of AMLE (Layout 2: heater-based layout)

Optical image of the fabricated heater-based layout seamlessly integrated to AMLE, is shown in **Fig. 4.3a**. The layout consists of two cascades of MZM to control the optical phase difference $\Delta\varphi$ and amplitudes of the inputs to AMLE (via $\Delta\varphi_1$ and $\Delta\varphi_2$ on the MZMs).

The experiment setup to measure AMLE consists of two probe/pump lines operating at λ_1 = 1550.92 nm (blue) and λ_2 = 1552.0 nm (red). The CW laser (N7711A, Keysight Tech.) inputs are combined at 50:50 ratio to an optical switch (GZ-1×2-SM, Gezhi) which selectively routes the signals to either probe or pump line. A second optical switch (GZ-1×2-SM, Gezhi) channels the signal (probe or pump) to an optical coupler to split the signal at 50:50 ratio before the signal is respectively filtered at λ_1 and λ_2, and coupled onto the chip. The output signal is filtered by OTF (OTF-320, Santec Corp.) and then detected by PD (2011-FC, Newport Spectra-Physics Ltd.). During pumping, an optical attenuator (V1550PA, Thorlabs Inc.) connected prior to the OTF is activated to null the signals to PD. The same optical pump approach as in previous ring-based layout (details in **Section 4.2**) is applied. The sourcemeters (Keithley 2400, Keithley 2614, Keithley Instruments) provide voltages to on-chip MZM waveguide heaters to induce optical shifts $\Delta\varphi_1$ and $\Delta\varphi_2$ on the MZMs that respectively defines $\Delta\varphi$ and amplitudes of inputs s_1 and s_2 to AMLE.

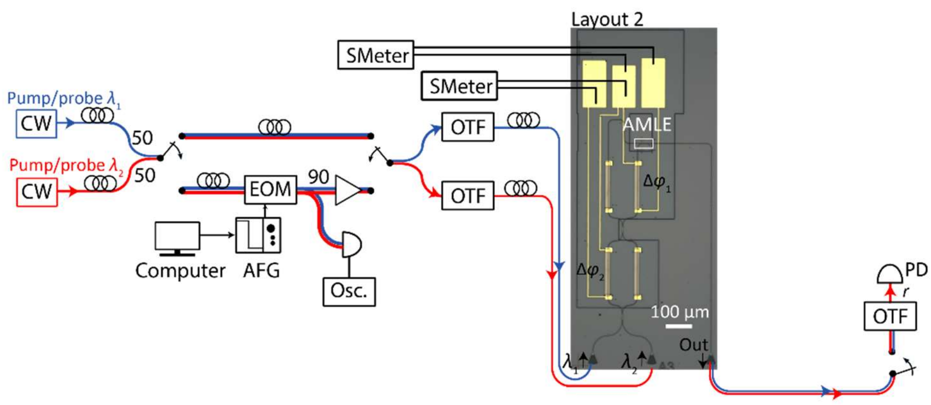

Fig. 4.3. AMLE measurement on heater-based layout (Layout 2). Schematic of experiment setup to measure AMLE. Voltage biasing on waveguide heaters (using source meters) modulates optical phases $\Delta\varphi$ and amplitudes of s_1 and s_2 inputs to AMLE (boxed). Two lines operating at wavelength λ_1 (blue) and λ_2 (red).

4.4 Components in measurement setup

The measurement setup described in prior sections contain optical, electrical and mechanical components. These are detailed hereafter.

4.4.1 Optical components

Laser. Two different lasers are used in the setup. They are TSL-550, Santec Corp. and N7711A, Keysight Technologies. Both are tunable lasers in telecommunications wavelength (TSL-550 at 1.5 – 1.63 µm, N7711A at 1.57 – 1.608 µm). Wavelength tuning using TSL-550 takes place almost instantaneously, while N7711A requires about five seconds. TSL-550 has maximum output power of 13 dBm, with 90 dB/0.1 nm signal-to-noise ratio and ±20 pm wavelength accuracy. N7711A has a maximum output power of 15.5 dBm, with 60 dB/0.1 nm signal-to-noise ratio and ±0.8 pm wavelength accuracy.

Optical fibre. The optical fibres in the setup are mainly single mode fibres (SMF-28e) with mode field diameter of 10.4 µm (*or* effective mode area of 85 μm^2). Most have 8°-angled polished ferrule connectors (FC/APC; P3-SMF28E-FC-5, Thorlabs Inc.), which prevents parasitic backreflections [402,403]. For equipment that require flat polished connectors, for example EOM 2623 NA, Lucent Tech., FC/PC to FC/APC single mode patch cables (P5-SMF28E-FC-1, Thorlabs Inc.) are used to enable APC to PC conversion. To split optical signals, 1x2 wideband SMF-28 fibre optic couplers are used (50:50 ratio using TW1550R5A1, Thorlabs Inc.; 90:10 ratio using TW1550R2A1, Thorlabs Inc.) These couplers have a ±100 nm bandwidth with insertion loss of ≤ 3.6 dB.

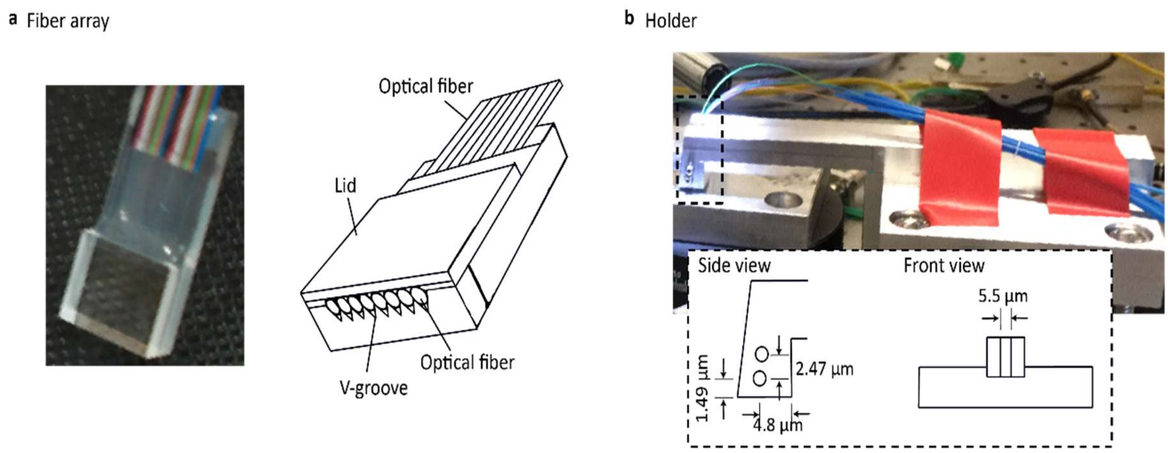

Fig. 4.4. Fibre array. a, Image (left) and schematic (right) of fibre array. Fibres are fastened to V-grooves and lid. **b**, Fibre array holder. Fibre array is held onto flat groove on the holder, and screws at the sides. It has an 8°-tilt. Inset: Schematic of holder.

Fibre array. In the fibre array, there are eight single mode fibres spaced 250 µm apart. The fibres are of Borofloat 33 G657A1 type, with 9 µm Si core, 125 µm SiO_2 cladding and 8°-angled polished connectors. They are coated with 242 µm Acrylate, and 900 µm Hytrel resin jacket for better protection, and assembled in v-grooves with 5 mm overall width. The fibre array is held tightly to its position using an aluminium fibre array holder (details in **Fig. 4.4**).

Polarisation controller. The polarisation controller (FPC030, Thorlabs Inc.) provides flexibility to adjust the fibre mode polarisation. This is required because the on-chip grating coupler and EOM are both polarisation-sensitive. In the controller, the fibre is spooled (mechanically bent) to introduce polarisation shift [404,405]. The number of spools is adjusted to mimic half-wave and/or quarter-wave plates (as in free space optics) [406,407] so that any desired polarisation can be attained. Although polarisation-maintaining fibres [408,409] (for example, P3-1550PM-FC-2-PM, Thorlabs Inc.) could instead be used, not all equipment is keyed to the required polarisation.

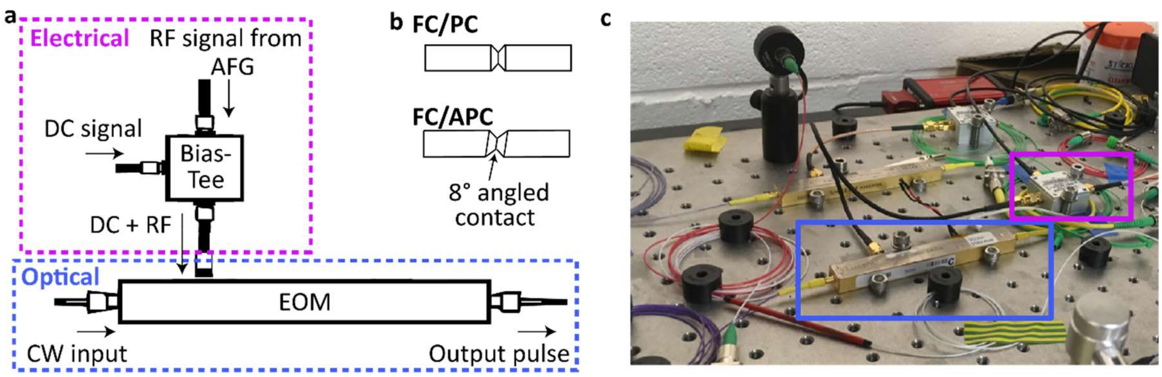

Fig. 4.5. EOM connected to bias-Tee. a, Schematic of connection. Bias-Tee enables EOM output to be referenced during modulation. **b**, The EOM 2623 NA has FC/PC connections, which requires intermediary FC/PC to FC/APC cables to connect to other setup components. **c**, Image of connection. EOM and bias-Tee respectively in blue and purple boxes.

EOM. The optical output signal from the EOM (2623 NA, Lucent Tech.) mirrors the RF signal it receives. The RF signal introduces optical phase shift on the MZI (in the EOM) to modulate the optical signal [399]. The EOM has an insertion loss of 3.7 dB, and can be amplitude modulated to 13 dB at modulation speed of up to 10 Gbits/s for 1.525 to 1.62 µm optical wavelength range. For maximum extinction ratio, a bias-Tee (ZFBT-4R2GW, Mini-Circuits, which combines DC and RF signals) is connected to the EOM so that the DC signal can reference the EOM output to the minimum (details in **Fig. 4.5a**). Because the EOM has FC/PC connectors - distinct from FC/APC connectors in the setup as depicted in **Fig. 4.5b**, intermediary FC/PC to FC/APC cables are required. **Fig. 4.5c** shows the overall connections. Optical pulses from the EOM, when amplified, introduces GST structural phase transition essential for AMLE learning.

EDFA. The EDFA (AEDFA-CL-23, Amonics Ltd.) amplifies optical pulse from EOM, so that the net optical pulse power is sufficient to switch the structural phase of GST for AMLE learning. It can receive input signals from -6 dBm (0.25 mW) to 3 dBm (2 mW), and amplify them up to 23 dB (0.2 W), for optical wavelengths between 1.528 to 1.562 μm, and 1.57 to 1.603 μm. The two main components in EDFA are erbium gain material doped in fibre, and a 980-nm wavelength internal pump laser to excite the erbium ions. Light amplification takes place when the input signal traverses through the excited erbium ions.

OTF. The OTF (OTF-320, Santec Corp.) selectively transmits light only one narrow wavelength, and is used for example to filter out white noise due to spontaneous emission in EDFA amplification. It has insertion loss of 7 dB, and 0.25 nm bandwidth at -3 dB optical power. Structurally, it is a mechanically-rotatable thin film with layers that form optical cavity to provide the narrow passband. The passband is tunable by tilting the incident angle to shift the resonance frequency via change in wave vector [130].

PD. The fibre-coupled PD (2011-FC, Newport Spectra-Physics Ltd.) allows optical transmission measurements (of up to 200 kHz modulation). Information about the optical power of incident light can be retrieved from the characterised responsivity [410] (0.85 A/W at around 1550 nm). It comes with in-built transimpedance amplifier [411] to increase signal-to-noise ratio. The fibre-coupled 1 GHz PD (1611-FC, Newport Spectra-Physics Ltd.) allows high-speed (1 GHz; 400 ps rise time) optical transmission measurements. Information about the optical power of incident light can be retrieved from the characterised current gain (~ 700 V/A at 1550 nm) responsivity (~ 1.04 A/ at 1550 nm).

Optical attenuator and switch. The optical attenuator (V1550PA, Thorlabs Inc.) nulls the pump signal to protect the PD. It can provide attenuation of up to 30 dB with an applied voltage of 5 V, and has a modulation speed of 1 kHz. When voltage is applied to the attenuator, a tilt to a micro mechanical mirror in the attenuator is induced to misalign the input signal [412]. The optical switch (GZ-1×2-SM, Gezhi) likewise tilt to a micro mechanical mirror in order to couple light to a different pathway. It has a switching speed is 125 kHz.

4.4.2 Electrical components

AFG and oscilloscope. The AFG (3102C, Tektronix) generates electrical pulses that defines the optical pulses sent to AMLE (via EOM). The oscilloscope (TDS 7404, Tektronix) graphically displays the optical pulses detected by PD. It has four separate channels – each can capture pulses of up to 4 GHz with 20 GS/s sample rate in real-time.

Source meter. The sourcemeters (K 2400, 2614, Keithley Inst.) provide and/or measure electrical signals through NiCr heaters on heater-based layout. K 2400 has ± 0.2 to 200 V and ±1µ to 1 A ratings. K 2614 has ± 0.2 to 200 V and ±0.1µ to 10 A ratings.

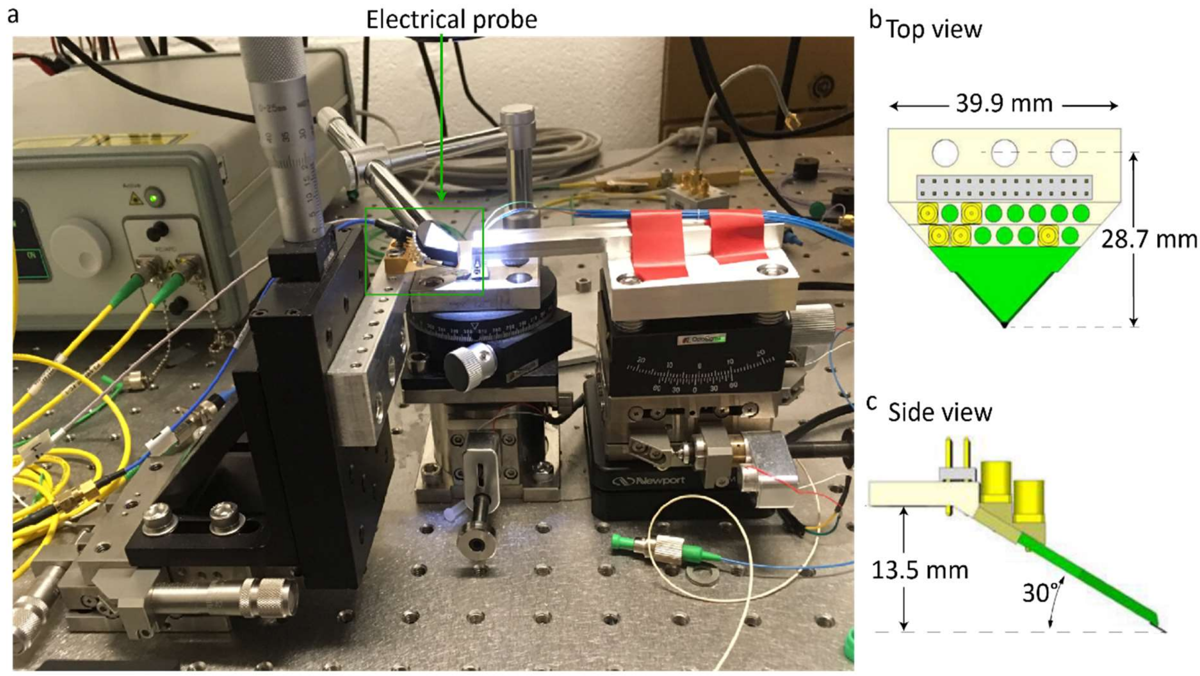

Fig. 4.6. Electrical probe. a, Photo of electrical probe (green boxed) supported by a home-built stage. The position of the stage is manually adjusted such that the electrical probe is on the on-chip electrical pads. **b**, Top view of electrical probe. **c**, Side view of electrical probe.

Electrical probe. The electrical probe (TITAN Multi-contact, MPI Corporation) provides electrical signals to on-chip electrical pads. It has 15 beryllium-copper probes (used due to low gold contact resistance ~ 45 mΩ) spaced at 125 µm, with contact width of 20 µm each. They are customised to ground, RF (up to 6 GHz) and logic (up to 500 MHz) signal configurations. To connect the multiprobe to source meter, RF cables with SMA connector and MMCX plug are used. **Fig. 4.6a** shows an image of the electrical probe, with top and side views depicted in **Fig. 4.6, b and c** respectively.

Microscope with charge-coupled device (CCD) camera. The microscope (Navitar 12x Ultrazoom) with CCD camera (DCC1240C, Thorlabs Inc.) is crucial to align the fibre array to the grating couplers. The microscope has a 3 mm fine focus and 0.58x to 7x zoom range, connected to a 24 W white LED light source to enable viewing. The CCD camera, connected to the microscope and computer, captures top view image of fibre array relative to grating couplers. It produces 1280 × 1024-pixel images

4.4.3 Mechanical component

Motorised stage. The motor (Picomotor 8742, New Focus; **Fig. 4.7, a and b**) moves the sample/chip stage. It can move in 4 axes (in relative or absolute positions) at ~ 30 nm resolution and 115 Kbps translation speed (which is about 100 μs character^{-1}). Details of the actuator is described in **Fig. 4.7, c and d**.

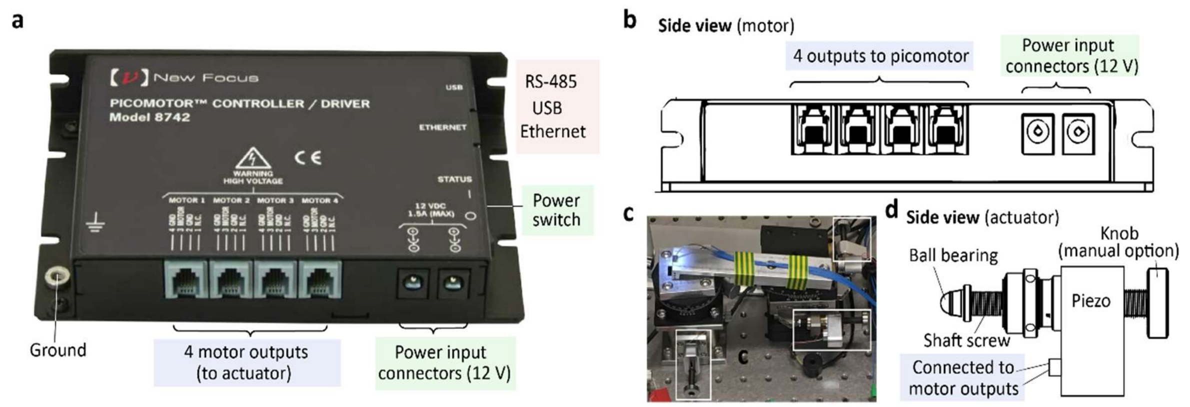

Fig. 4.7. Motorised stage. a, Image of motor. Red, green and blue shades indicate communications, power supply and actuator signals. **b**, Side-view schematic of motor. **c**, Image of on-chip stage, which has three linear actuators (in *x*,*y* and *z*-axis; boxed). **d**, Side-view schematic of actuator. When voltage is applied (blue shade), the piezo in the actuator changes in length thus rotating the shaft.

4.4.4 Data acquisition device

The data acquisition device (DAQ, NI-USB 6008; **Fig. 4.8, a and b**) centralises retrieving and sending signals to equipment from the computer. As BNC and SMA connections are mainly used in the setup, wires to/from the DAQ are connected to the respective solder connectors (shown in **Fig. 4.8, c and d**). To provide analogue and digital functions in real-time, Ni-DAQmx library is accessed using Virtual

Instrument Software Architecture (VISA) driver. Although other alternatives exist (for example, HiSLIP or VXI-11 over TCP/IP and USBTMC over USB), the VISA standard allows access to the library regardless of interfaces / bus systems (such as GPIB, VXI, PXI, Serial, Ethernet and USB), and is convenient to configure. The overall communication is established using Standard Commands for Programmable Instruments (SCPI) commands. Command line examples are shown in **Fig. 4.8e**.

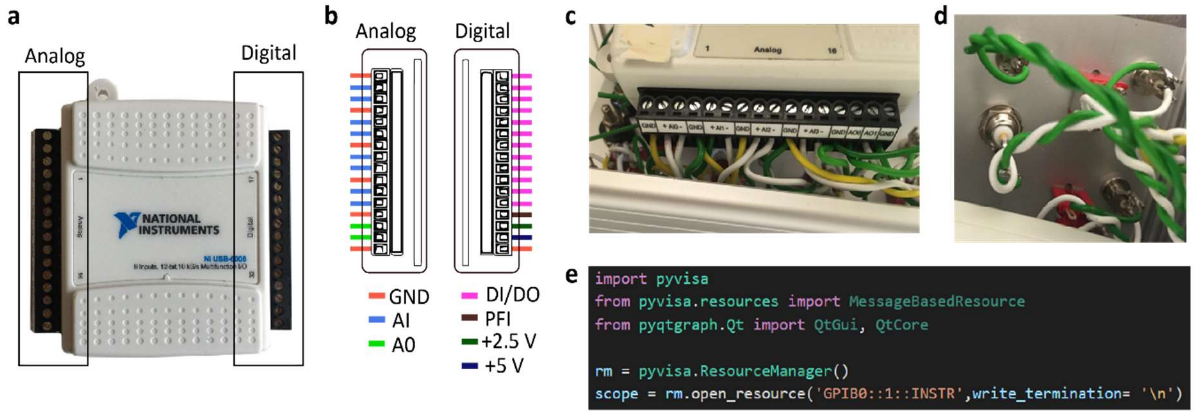

Fig. 4.8. Data acquisition device (DAQ). a, Image of DAQ, which supplies or receives analogue (left) and digital (right) signals. **b**, Analogue input and output lines are denoted as AI and AO respectively. Digital lines can be programmed as digital input (DI) or output (DO), or programmable function input (PFI; clocking input signal for synchronisation or trigger). Additionally, the DAQ provides high-purity reference voltage (+2.5 V) and a +5 V 200 mA power source. **c**, Image of wires connected to the DAQ, with wire shielding to reduce electrical interference. **c**, To reduce magnetic interference, the wires are twisted. The wires are connected to BNC and SMA solder connectors. **e**, On Python IDE, PyVISA and PyQt modules are called to respectively communicate with equipment and to create graphical user interface for the purpose.

We have analysed the experimental setup to observe associative learning using two separate layouts. These layouts are conceived to solve the optical phase stability problem. The setups are devised to prepare the pulse phase in order to trigger (pump), and to observe (probe) the learning process. The main setup components are scrutinised.

Chapter 5

5. Observation of all-optical associative learning

This chapter shows experimental results using the setups described in previous chapter. As will be described, these setups exhibit complementary features - allowing us to cross validate the measurement results and extract important information about the learning process.

5.1 Methods to observe associative learning

Previously in **Section 3.2.2**, two complementary layouts to observe the associative learning process are introduced. They are ring-based (Layout 1) and heater-based (Layout 2). Whereas Layout 1 enables the associative learning process to be observed in real-time (due to the different wavelengths used - which do not optically interfere - for the inputs to AMLE), Layout 2 allows for the integration of multiple AMLEs using wavelength-multiplexed inputs. A summary of the two layouts is depicted in **Fig. 5.1**.

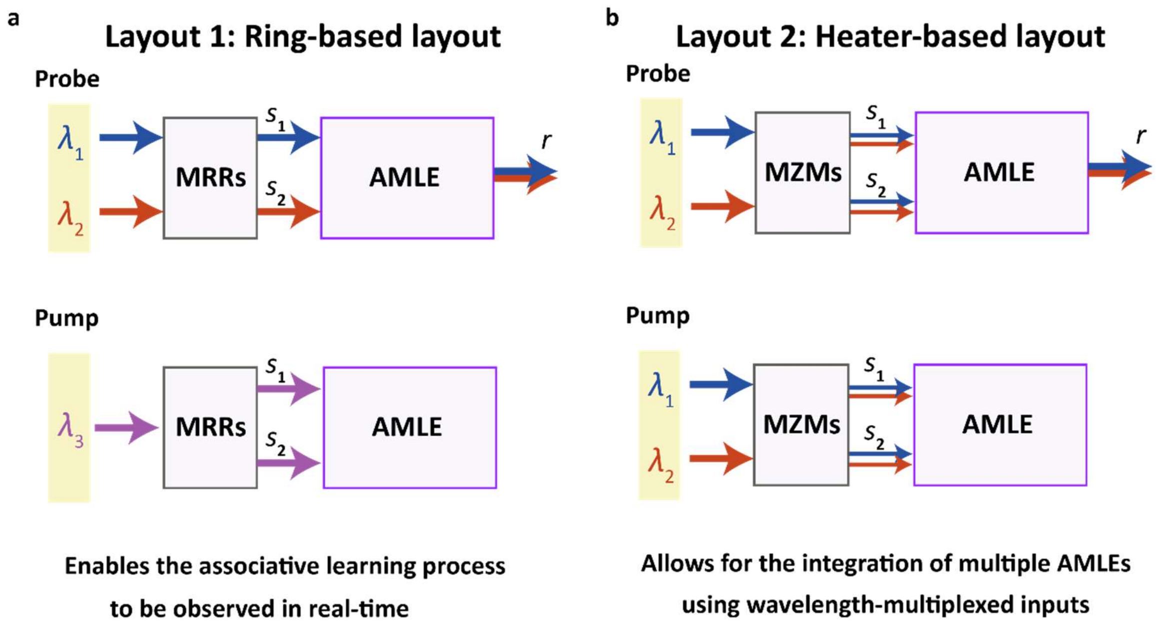

Fig. 5.1. Summary of layouts to observe associative learning. a, Layout 1 enables the associative learning process to be observed in real-time. Microring resonators (MRR) filters the wavelengths of the inputs to the AMLE. Probe and pump input wavelengths are different. **b**, Layout 2 allows for the integration of multiple AMLEs using wavelength-multiplexed inputs. Mach-Zehnder modulators

(MZMs) precisely define the amplitudes and optical phases of the inputs to AMLE. Probe and pump input wavelengths are the same.

5.2 Probing the associative learning process in real-time (on Layout 1)

It is important to characterise the AMLE in real-time to understand its dynamic response. To this end, different wavelengths are used for the pump and probe lines. This allows us to retrieve the process dynamics without introducing optical interference to the lines.

5.2.1 AMLE input settings

The ring-based layout is designed such that two ring resonators are coupled to the paths leading to s_1 and s_2 inputs. s_1 (s_2) input signal is sent to AMLE when input to the layout is of ring B (A) resonant wavelength λ_1 (λ_2). Numerical simulations, performed to confirm the signal propagation on the layout, corroborates this (results in **Fig. 5.2a**).

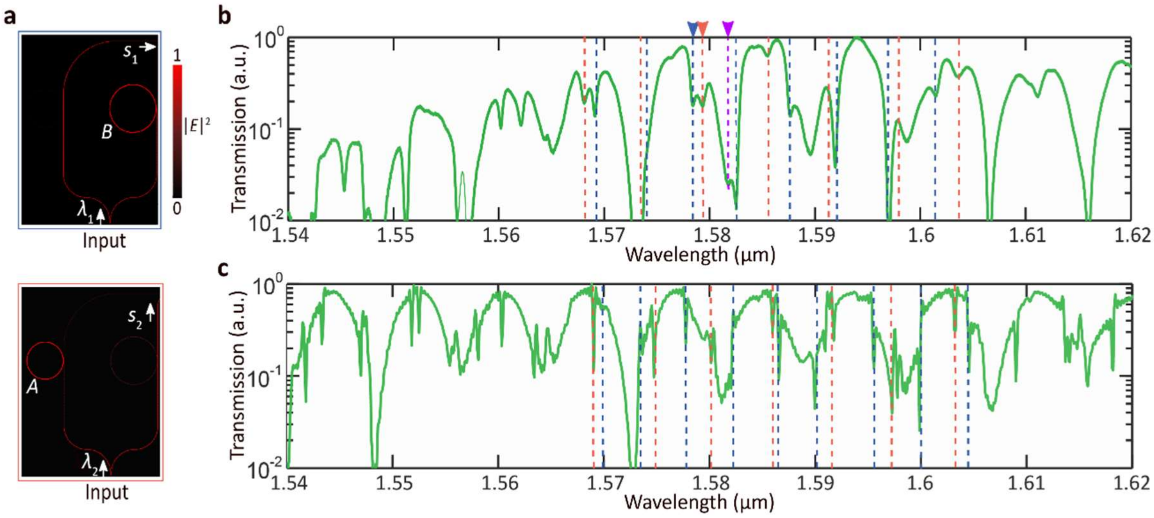

Fig. 5.2. Input settings (Layout 1). a, Profile of simulated optical intensity $|E|^2$ when input to layout is at λ_1 (top) and λ_2 (bottom), which activates s_1 (s_2) input to AMLE. **b** and **c**, Optical transmission spectrum in experiments (**b**) and simulations (**c**). Transmission spectrum envelope in experiments is due to broad resonance of grating couplers. Arrows denote λ_1 = 1578.3 nm (blue), λ_2 = 1579.4 nm (red) and λ_0 = 1581.9 nm (purple). Free spectral range in experiment corresponds well to that in simulations.

Fig. 5.2b highlights λ_2 and λ_1 in the transmission spectrum obtained from experiments (setup shown in **Fig. 4.2**), indicated by red and blue dashed lines respectively. Arrows denote the λ_2 and λ_1

wavelengths at which measurements are carried out when s_1 and s_2 probe inputs are sent to AMLE. Probe measurements are carried out at these wavelengths (λ_1 = 1578.3 nm and λ_2 = 1579.4 nm) after ensuring ring *A* and *B* are critically coupled to the waveguides (by measuring the waveguide-ring resonator structures of same dimensions adjacent to the layout). Ring A and B respectively have full width at half maximum (FWHM) of 0.21 nm and 0.39 nm, and extinction of -29 dB and -26 dB. The free spectral range in experiments (**Fig. 5.2b**) matches with that in simulations (**Fig. 5.2c**). Two-input pump signals (which induce learning) are sent at the non-resonant wavelength (λ_0 = 1581.9 nm). From the spectrum in **Fig. 5.2c**, $\Delta\varphi$ can be determined. The fringes in the spectrum are attributed to wavelength-dependence of $\Delta\varphi$.

5.2.2 Demonstration of associative learning

Having determined the input optical wavelengths after performing spectrum sweep shown in **Fig. 5.2**, the GST state in the AMLE is stabilised by applying a set of 'amorphising' (set) and 'crystallising' (reset) pulses. These switching pulses serve to prepare the AMLE for the learning process, by conditioning the phase change materials on the AMLE to achieve reproducibility. Here, the set of amorphising pulses are the consecutive 100 ns-wide pulses at 0.66 nJ, 0.87 nJ, 1.26 nJ and 1.45 nJ, while the set of crystallising pulses are the 100 ns-wide pulse at 0.43 nJ for ten times, followed by five 0.19 nJ 100 ns-wide pulses at 1 MHz repetition rate for ten times. As will be described, these sets of pulses are exactly the same as the pulses for the 'associative learning' and 'forgetting' process.

The starting point of the AMLE for the experiments is its crystalline state. The output transmission r_1 and r_2 of AMLE is probed in real-time using wavelengths λ_1 and λ_2. As shown in the bottom plot of **Fig. 5.3a**, the transmission readouts r_1 and r_2 remain the same for single input pulses at 1.45 nJ (pulse widths τ = 100 ns) as expected (for events 1 to 4). However, when inputs s_1 and s_2 are sent simultaneously at pump wavelength λ_0 with a fixed phase delay of $\pi/2$ at 0.66 nJ (τ = 100 ns) each in event 5, the transmission change Δr_1 and Δr_2 for the s_1 and s_2 probe readouts are $\sim -4\%$ and $\sim +4\%$ respectively. As the input pump pulse power is increased from 0.87 nJ (τ = 100 ns) to 1.45 nJ (τ = 100 ns) in events 6 to 8, probe readouts change by approximately −7% and +7% respectively, both of which are well above the transmission threshold of $r_{th} \sim 5\%$. Effectively, these experiments show that the two inputs can be 'taught' to associate with each other such that either triggers the response: the associations are learned.

These learned associations can also be reversed. A set of pulses at 0.43 nJ (τ = 100 ns) in event 9, followed by 0.19 nJ pulses (τ = 100 ns) in event 10 result in the 'forgetting' process, where the readouts r revert to the baselines ($r_1 \sim 0.14$ for s_1 input probe and $r_2 \sim 0$ for s_2 input probe). For the measurements, readout above $r_{th} \sim 0.05$ is designated as the learned state.

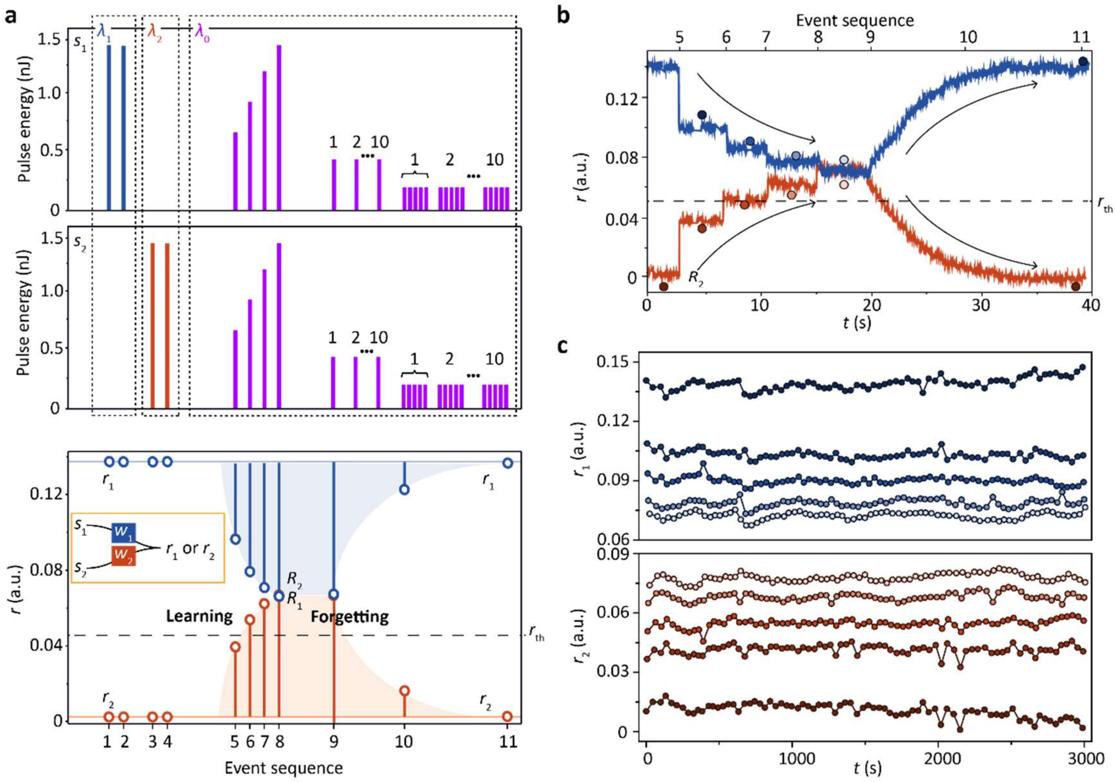

Fig. 5.3. Photonic Pavlovian learning process. a, Input-output relation of AMLE. s_1 and s_2 inputs denote 'Food' and 'Bell' inputs; the corresponding output transmission r (bottom plot) represents the transmission response r_1 and r_2. Blue, red and purple bars of the top and middle charts denote s_1, s_2 and both s_1-s_2 input incidences (λ_1, λ_2 and λ_0 are the wavelengths used to selectively address each one). The change in transmission response is attributed to the relative differences in both the real and imaginary refractive indices between GST at crystalline and amorphous structural states (values provided in **Section 3.1.2**). These real and imaginary refractive index contrast both lead to distinct coupling strength between the waveguides that constitute the AMLE, respectively due to changes in optical phase and optical absorption on the device. **b**, Corresponding real-time measurement of output probe transmission r of a single cycle learning and forgetting processes in **a**. **c**, Repeatability of the processes on AMLE over 80 cycles. The levels are denoted by filled circles of different colours that correspond to **b.**

Fig. 5.3b shows a single cycle of the real-time output readout of associative learning in events 5 to 8 and the forgetting process in events 9 to 11 of **Fig. 5.3a**. To test the repeatability of associative learning and forgetting processes, the AMLE is subjected through 80 learning cycles, examined over a period of 40 minutes. After 80 cycles (**Fig. 5.3c**), the individual learning weights are clearly identifiable with standard deviation of ± 0.69% in readout transmission.

To better understand the $\Delta\varphi$-dependence of AMLE, the accumulated optical field $|E|$ at the site prior to GST cell (depicted in **Fig. 5.4a**) is simulated with respect to $\Delta\varphi$. The results, summarised in **Fig. 5.4b**, clearly indicates that the field varies with $\Delta\varphi$. Based on the results in the figure, it is verified that the two inputs can be 'associated' when the accumulated field is sufficiently large. Input pump pulses (pulse width τ = 100 ns, from 1.3 nJ to 2.9 nJ) are sent to the AMLE and the output transmission r_2 is experimentally measured using the setup described in **Fig. 4.2**. The measurement results shown in **Fig. 5.4a** indicate that the two inputs are indeed 'associated' (manifested by high r_2) when the accumulated field is large (at $\Delta\varphi = \pi/2$). Although structural state switching of GST cell is possible when a single input signal (either s_1 or s_2) of sufficiently high optical energy is sent, this has to take into account the optical field coupling at the first segment of the device.

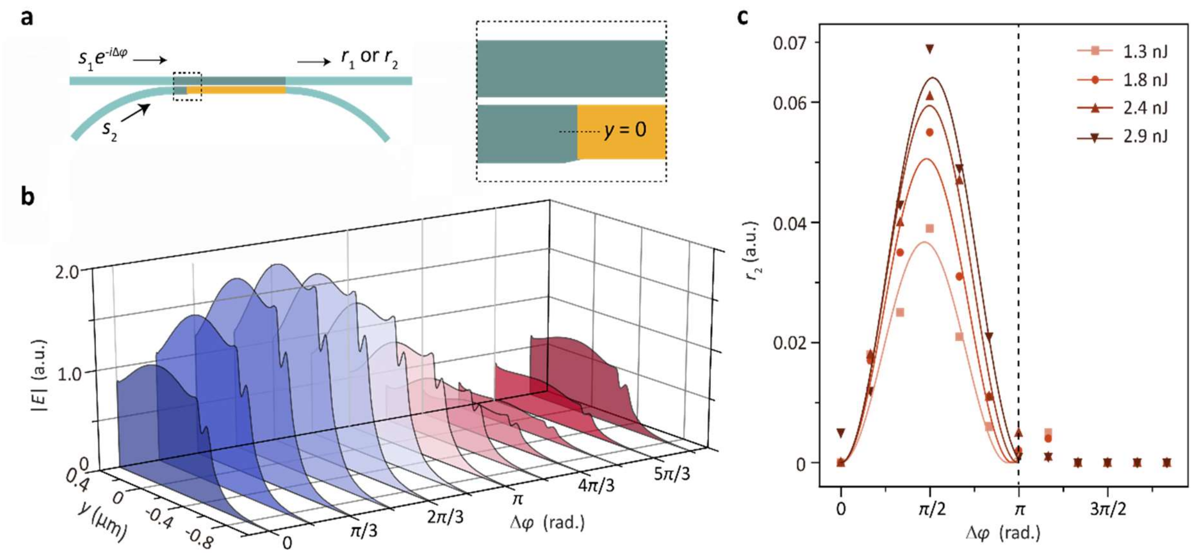

Fig. 5.4. $\Delta\varphi$-dependent change in learning readouts. a, Schematic indicates location y = 0 in AMLE. GST cell in yellow. **b**, Plot of simulated cross-sectional optical field $|E|$ across y with respect to $\Delta\varphi$. $|E|$ is maximum at $\Delta\varphi = \pi/2$ and minimum at $\Delta\varphi = 3\pi/2$. **c**, Plots of r_2 against $\Delta\varphi$ when input pump pulses are sent to AMLE at 1.3 nJ to 2.9 nJ (pulse width τ = 100 ns) in experiments is consistent with simulation results in (**b**), that r_2 is indeed dependent on $\Delta\varphi$.

5.3 Demonstration of associative learning on Layout 2

Previous demonstration of associative learning is carried out with different wavelengths for the input cases: λ_1 for s_1-only input incidence, λ_2 for s_2-only input incidence and λ_0 for both s_1 and s_2

input incidence. However, to scale the AMLEs up, it is desirable to be able to demonstrate associative learning process with the fixed wavelength(s) for the associated inputs. This section describes the measurement using on-chip heaters to achieve this.

5.3.1 AMLE input settings

The exact modulation to be induced can be determined by probing AMLE output transmission *r*. As indicated in **Fig. 3.5d** (which is plotted from **(3.14)** in **Chapter 3**), when *r* is minimum, accumulated optical intensity at site prior to GST cell s_{GST} is maximum (that is when $\Delta\varphi = \pi/2$). Before the experiments, the cases when the input(s) are sent are deduced using this scheme (details in **Fig. 5.5**).

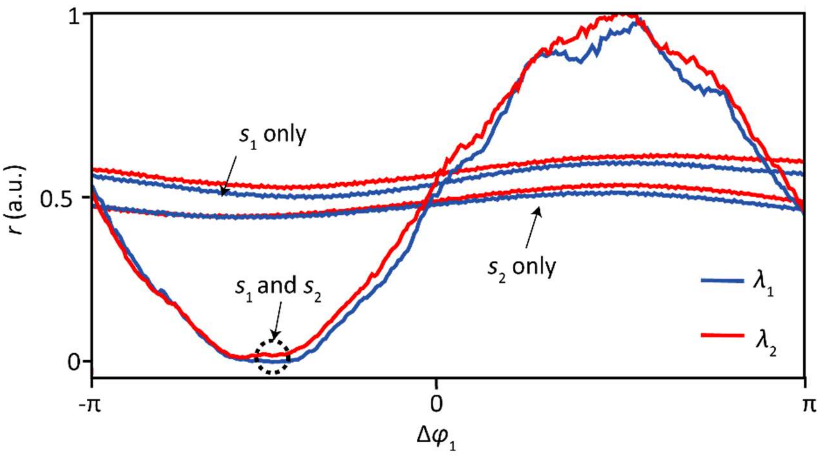

Fig. 5.5. Input settings (Layout 2). Plot of *r* as a function of $\Delta\varphi_1$ when s_1 only, s_2 only and both s_1 and s_2 are sent, at two different wavelengths (λ_1 = 1578.3 nm, λ_2 = 1579.4 nm) in our measurements. From the plot, the cases when the input(s) is sent is deduced based on the relations between s_{GST} and *r* in **Fig. 3.5d**.

5.3.2 Demonstration of associative learning

Using the setup in **Fig. 4.3**, three sets of measurements (summarised in **Fig. 5.6a**) are carried out. They correspond to the following configurations:

i. *Teacher* and *Input* signals as s_1 and s_2 (using single wavelength λ_1) respectively

ii. *Teacher* and *Input* signals as s_1 and s_2 (using two multiplexed wavelengths λ_1 and λ_2) respectively

iii. *Teacher* and *Input* signals as signal λ_1 and signal λ_2 respectively

The abstraction layer ('Abs.' in **Fig. 5.6a**) consists of MZMs with heaters to precisely define $\Delta\varphi$ and amplitudes of inputs to AMLE (details in **Fig. 5.4a**). **Fig. 5.6b, i** shows AMLE measurement results based on **Fig. 5.6a, i**. Transmission above $r_{th} \sim 5\%$ is designated as learned state. Transmission readout remains at baselines after s_1 pump input pulse (pulse width τ = 100 ns) is sent at 1.47 nJ in events 2 and 3, and s_2 pump input pulse (τ = 100 ns) at 1.47 nJ in events 4 and 5. Sending both s_1 and s_2 pump pulses together (at $\Delta\varphi = \pi/2$) at 1.47 nJ results in transmission probe change $\Delta r_1 \sim -8\%$ and $\Delta r_2 \sim +7\%$ respectively for s_1 and s_2 probe readouts. The same is observed in **Fig. 5.6b, ii** (based on scheme in **Fig. 5.6a, ii**), with $\Delta r_1 \sim -8\%$ and $\Delta r_2 \sim +7\%$ respectively in event 6.

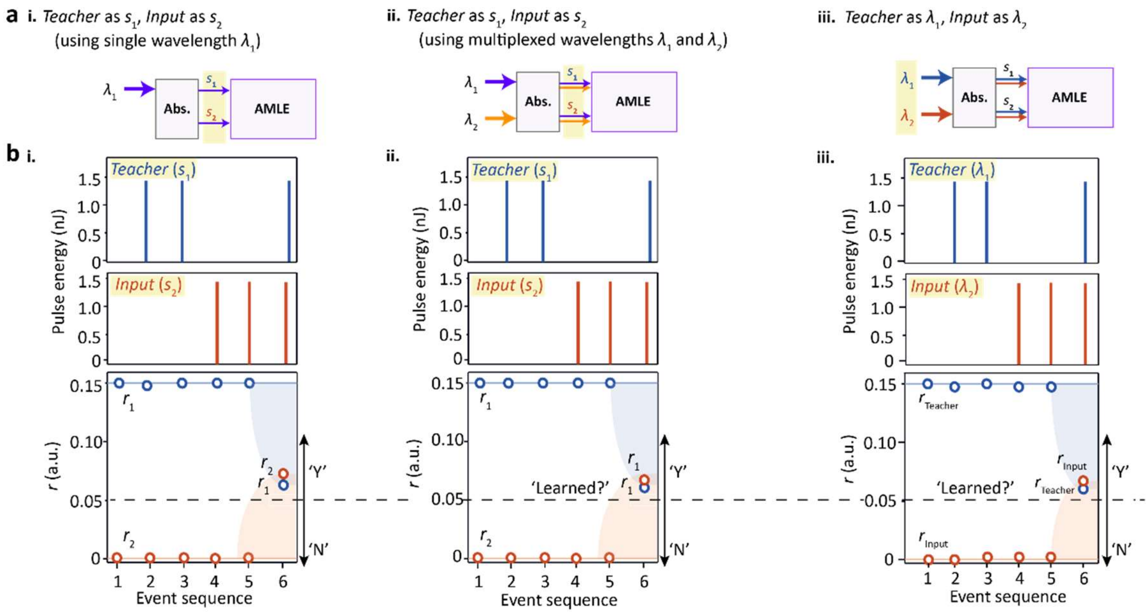

Fig. 5.6. AMLE measurement results on heater-based layout. a, Schematic of heater-based layout. Signal(s) to layout sent to AMLE via abstraction layer ('Abs.', consists of MZMs). With the layer, there are three ways to represent *Teacher* and *Input* signals (denoted by yellow highlights), using: (**i**) single wavelength λ_1, (**ii**) two multiplexed wavelengths λ_1 and λ_2, and (**iii**) wavelengths λ_1 and λ_2. **b**, Output transmission r (bottom plot) when *Teacher* (top blue chart) and *Input* (top red chart) signals are sent to AMLE in the experiments. Plots **i to iii** correspond to that in (**a**). Dashed lines indicate transmission threshold for learned response.

For the measurement schemes using λ_1 and λ_2 respectively as *Teacher* and *Input* signals in **Fig. 5.6b, iii** (based on scheme in **Fig. 5.6a, iii**), transmission remains at the baselines ($r_1 \sim 0.15$ and $r_2 \sim 0$) after *Teacher* and *Input* pump pulses are sent at 1.47 nJ each in events 2 and 3, and events 4 and 5 respectively. $\Delta r_1 \sim -8\%$ and $\Delta r_2 \sim +7\%$ (learning) are respectively observed when both *Teacher* and *Input* pump pulse are sent at 1.47 nJ in event 6. This results is consistent with that in **Fig. 5.6b, ii**.

Regardless of the optical wavelength used for the *Teacher* and *Input* pump pulse, the combined optical pulse has to be sufficiently high for the switching of GST structural state to occur. In subsequent large-scale pattern and image recognition experiments, this scheme is used, so that the heaters do not need to be continuously modulated to activate the different inputs during training.

We have presented the measurement results of the optical associative learning device. Two separate sets of measurements are carried out on two different layouts, respectively. The methods to determine the input configurations to the device are examined. Whereas Layout 1 enables real-time monitoring of the learning states of the device, Layout 2 provides the option to observe the learning process when the inputs to the device are of the same wavelength(s). The dependence of learning extent on the optical phase difference between the two inputs are experimentally characterised.

Chapter 6

6. Associative learning for supervised learning

This chapter discusses the scaled-up implementation of the associative learning device based on the similarities between the concept of supervised learning and associative learning. The scaled-up hardware is used to solve pattern and image recognition tasks.

6.1 Supervised learning concept

Up to this point, we have discussed associative learning as a single element. Associative learning (**Fig. 6.1a**) and supervised learning (a class of machine learning; **Fig. 6.1b**) are in essence comparable – both involving the pairing of input (*IN*) with the correct output (*Teacher*) to supervise the learning process. However, the exact network constituents that map the input to output in conventional supervised learning architectures are somewhat distinct – the *Teacher* signal propagates backward layer-by-layer to collectively adjust the network weights such that the actual output *R* better resembles the desired output (*Teacher*) after each learning iteration. To reinforce the relation between the two, a scaled-up design for AMLEs (illustrated in **Fig. 6.1c**) is required.

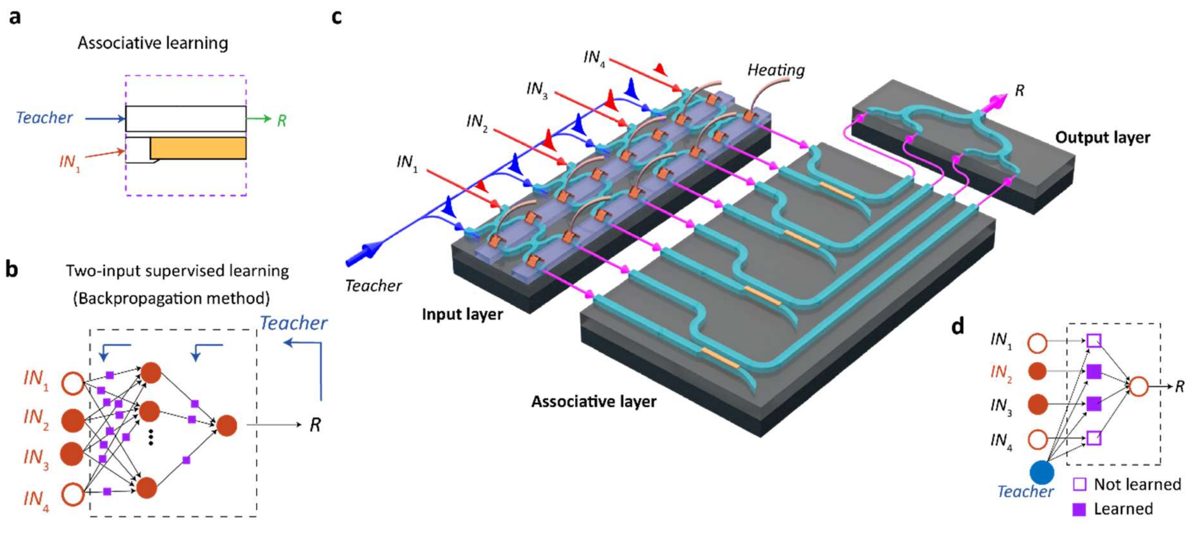

Fig. 6.1. Supervised Pavlovian associative learning. a, Pavlovian learning involves pairing the inputs (*IN*) with the correct outputs (*Teacher*) to supervise the learning process. **b**, Current conventional supervised learning networks use backpropagation. The network diagram depicts the network for supervised learning. **c**, Optical on-chip hardware diagram of supervised learning with two inputs using AMLE. Input signals IN_1 – IN_4 fed into the system during the learning process to be supervised by the *Teacher* input signal. IN_1 - IN_4 and *Teacher* inputs leading to the AMLE are controlled by thermo-optic (with NiCr heaters) MZMs. **d**, Network representation of the hardware diagram in **c**.

The model (schematic drawn in **Fig. 6.1c**, and in **Fig. 6.1d**) is implemented to an integrated chip. There are six fabrication steps in the hardware design, as discussed in **Section 3.3** of **Chapter 3.** To define the structures in these steps, photolithography is used (instead of electron-beam lithography) when the required thickness of resist window exceed 100 nm - such as that for markers and frame, isolation layer, thermo-optic heaters, and electric pads. The CAD and mask design are shown in **Fig. 6.2** and **Fig. 6.3**, respectively. The mask base material is soda lime glass (1.6 mm-thick) with low-reflective chrome plating (4 inch).

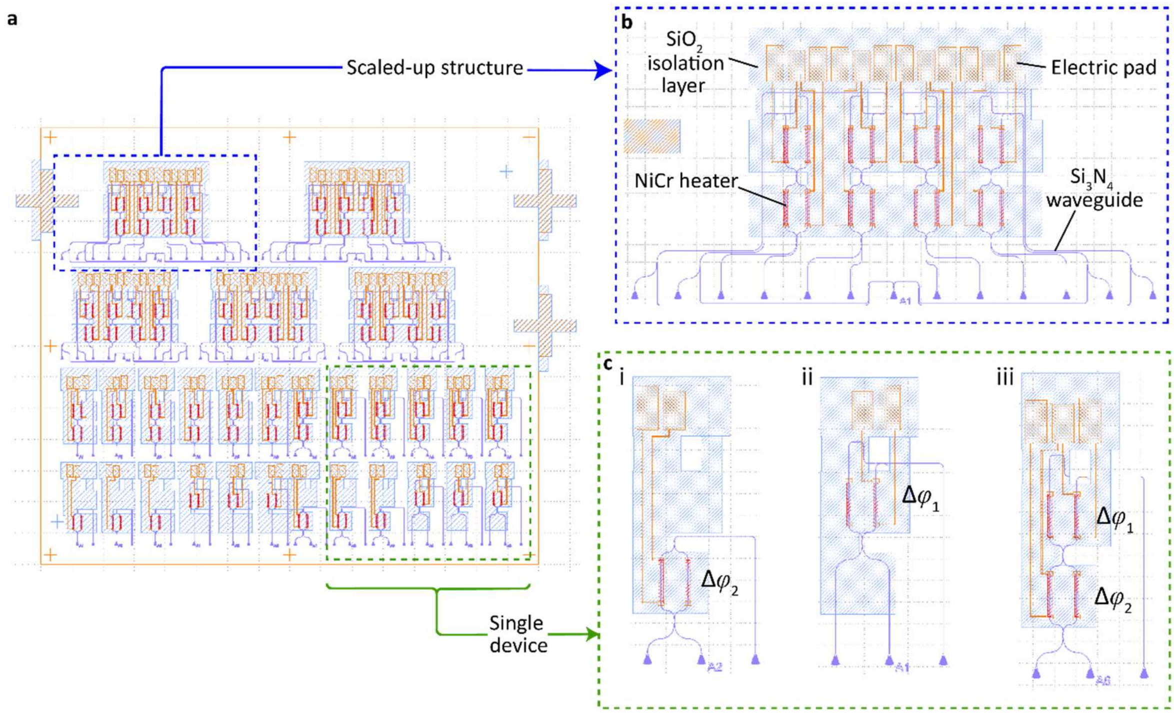

Fig. 6.2. CAD design. a, Single device and scaled-up structure design on a single chip. **b and c**, Zoomed-in view of scaled-up (**b**) and single device (**c**) structure design. The single device design include: i. bottom MZM cascade to characterise $\Delta\varphi_2$, ii. top MZM cascade to characterise $\Delta\varphi_1$, iii. Full single device, with both top and bottom MZM cascades.

The resulting hardware is shown in **Fig. 6.4a**. It consists of the associative layer (purple bordered boxes) between an input layer (input data and input *Teacher* label, denoted using red and blue arrows respectively) and an output layer (response *R*; green arrow). The NiCr thermo-optic heaters (green bordered boxes) enable the control of optical signal paths between the input layer and the associative layer. Optical micrographs with an enlarged view of the AMLE and thermo-optic heaters are shown in **Fig. 6.4b**, and **Fig. 6.4c** respectively, with the depiction of AMLE consistent with prior descriptions in **Fig. 3.2a**. The thermo-optic heaters, which control the optical phases and thus connections between the input layer to the associative layer, are modulated using external voltage sources via on-chip gold electrodes (see **Fig. 5.5** for details on the modulation)

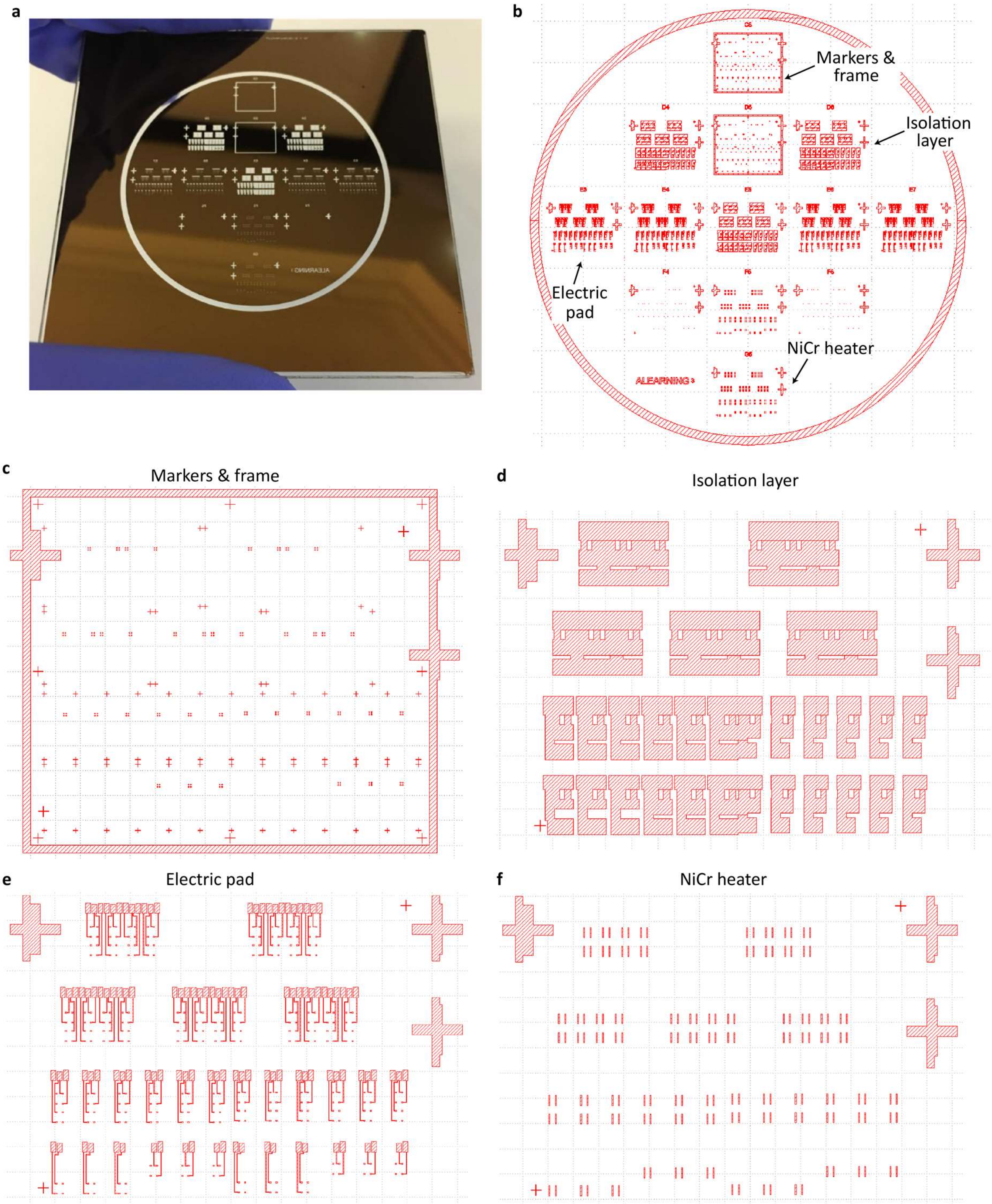

Fig. 6.3. Photolithography mask design. a, Optical image of mask. **b,** Optical image of mask. Design of structure from several layers on a single mask. **c-f**, Zoomed-in view of markers and frame (**c**), isolation (**d**), electric pad (**e**), and NiCr heater (**f**) layer design.

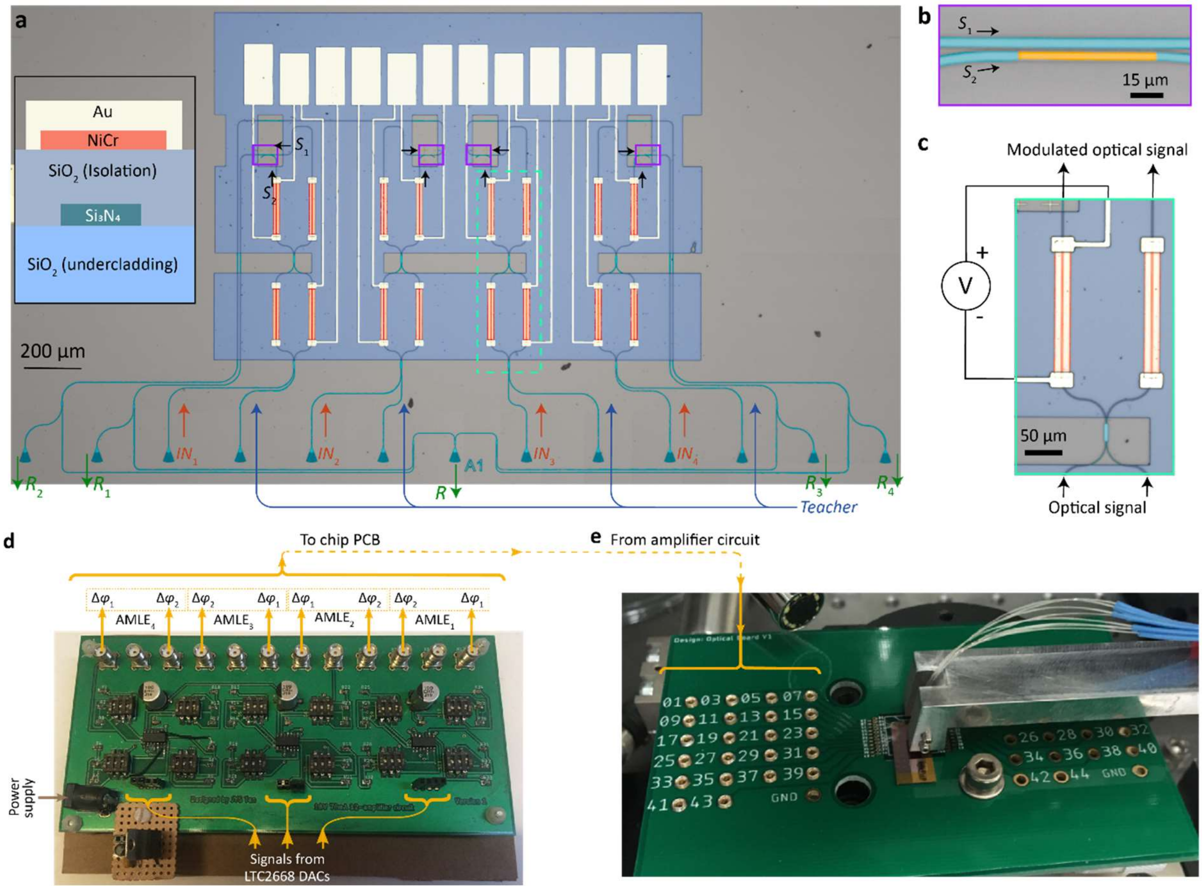

Fig. 6.4. Supervised associative learning hardware. a, Optical micrograph of supervised learning hardware (compare with schematic in **Fig. 6.1c**), which consists of four AMLEs (boxed, purple). The arrows show the optical input-output connections coupled to the on-chip hardware via grating couplers. **b, c**, Optical micrographs of a single AMLE and heaters respectively correspond to those in **a**. **d**, Image of amplifier circuit on PCB to increase voltage supplied from LTC2668 DACs to control $\Delta\varphi_1$ and $\Delta\varphi_2$. **e**, Image of chip PCB connected to amplifier circuit and wirebonded to electrical pads on AMLE hardware in (**a**).

Voltage biasing on the heaters (supplied via wire-bonded on-chip electrical gold pads, details in **Fig. 6.4a**) are delivered from the digital to analogue converter (DAC) microcontroller (LTC2668, Analog Devices [413]) amplified by electronic amplifier circuit on the printed circuit board (PCB) shown in **Fig. 6.4d**. The amplifier circuit, developed by the author based on circuit design shown in **Fig. 6.5**, conveniently and sufficiently ramp up the voltages on-demand from a computer connected to a microcontroller. Circuit simulations are run on this circuit design using circuit simulator software (LTSpice) to replicate the behaviour of an actual amplifier circuit. The amplifier circuit uses a voltage divider network to bias the operational amplifier in a non-inverting gain mode with resistors to form the required feedback loop, depending on the manual settings of the electro mechanical switch. Bypass capacitors are connected in parallel to the power supply to remove AC noise signals. Then, the circuit design is transferred to the PCB layout design software (Autodesk Eagle). The circuit connections are optimised using the functionalities provided by the software. The resulting optimised PCB layout is shown in **Fig. 6.6**.

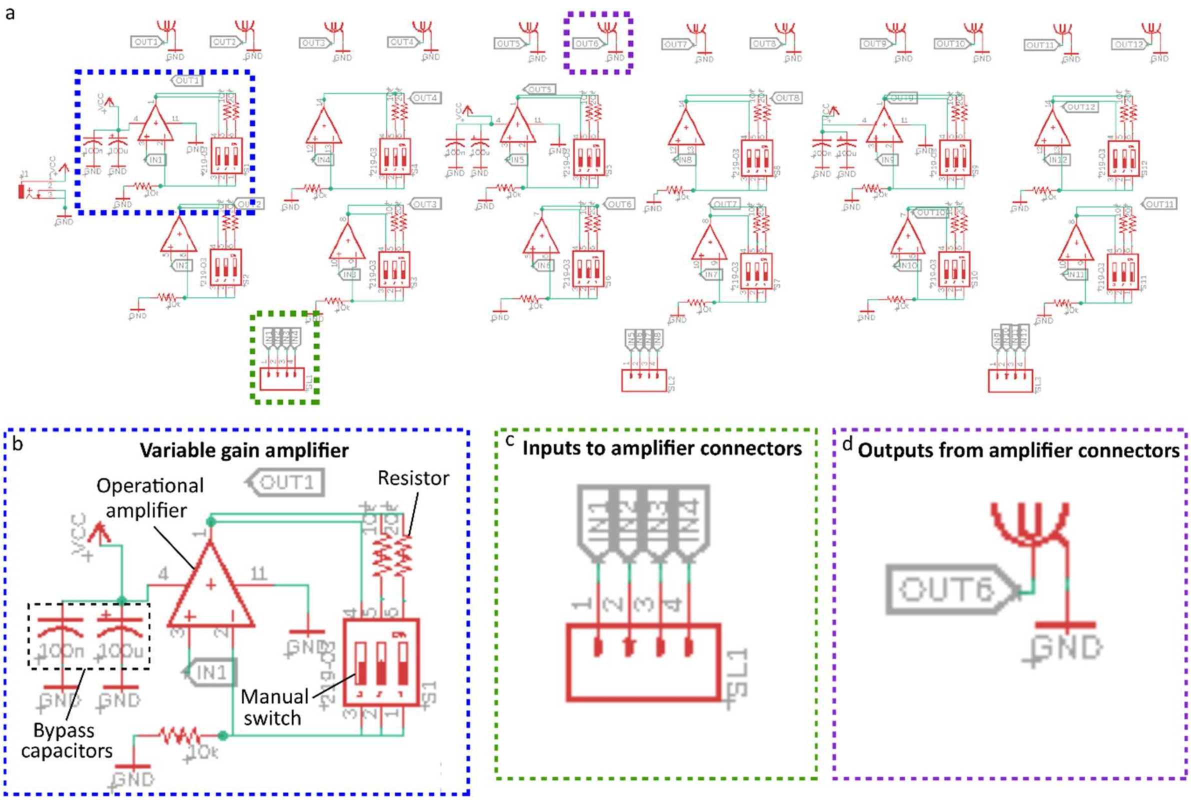

Fig. 6.5. Circuit design. a, Design as simulated on circuit simulator software. **b-d**, Zoomed-in view of the circuit main components: Variable gain amplifier, which consists of manual switch, operational amplifier, resistor, and bypass capacitors (**b**), inputs to amplifier connector (**c**) and outputs from amplifier connector (**d**).

For scaled-up experiments described in this chapter, measurement equipment in addition to that in **Chapter 4** is shown in **Fig. 6.7**. White CW laser source and spectral filter (Fianium WhiteLase Micro and SuperK Split, NKT Photonics) amplified using EDFA cascades (FA-15 and FA-33-IO, Pritel) provide input signals of broad wavelengths (for WDM) from a single source. The EDFA cascades supply amplification with higher gain – this is required as white laser typically operates at low power (optical signal power is shared with signals of many wavelengths). Other equipment includes pulse generator (HP 8133A, Hewlett Packard, which provides GHz-rate electrical modulations), and remote-controlled OTF (OTF-930, Santec Corp, which allows operating the OTF remotely from a computer).

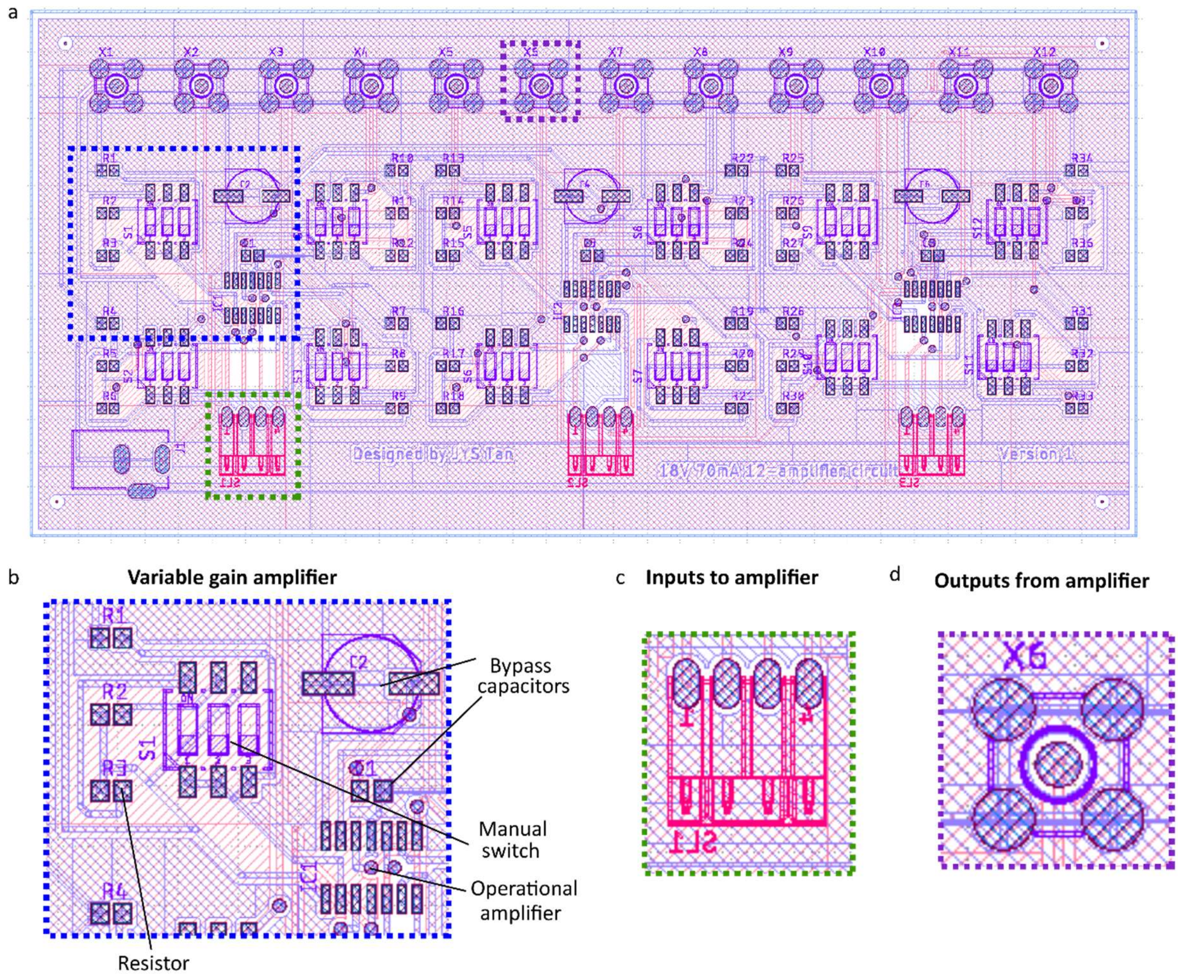

Fig. 6.6. PCB layout design. a, Layout design transferred from circuit design in **Fig. 6.5** - optimised using automation software. **b-d**, Zoomed-in view of the layout main components.

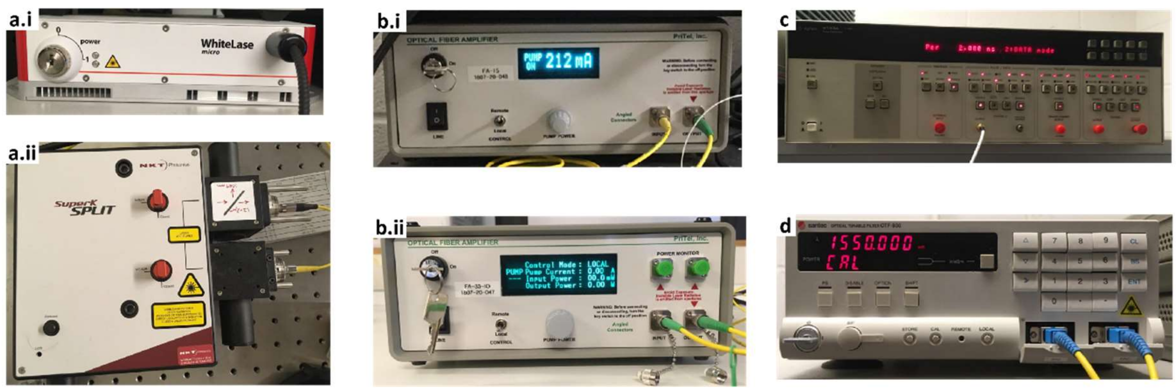

Fig. 6.7. Equipment for scaled-up experiments (in addition to that is discussed in Chapter 4). a, White CW laser - source (Fianium WhiteLase Micro, NKT Photonics) **(i)** and spectral filter (SuperK Split, NKT

Photonics) **(ii)**. **b**, Cascaded EDFA - (FA-15, Pritel) **(i)** and (FA-33-IO, Pritel) **(ii)**. **c**, Pulse generator (HP 8133A, Hewlett Packard). **d**, Remote-controlled OTF (OTF-930, Santec Corp.).

The measurement equipment in **Fig. 6.7** are detailed hereafter.

Optical components

White laser. The white laser (Fianium WhiteLase Micro) is used in the scaled up experiments to increase signal throughput (via wavelength division multiplexing (WDM)) over the signal channels in the experiments. The spectral shape and stability of white light spectra is attributed to nonlinear effects governed by optical dispersion in the optical medium, pulse length and peak power [414,415], which can be analysed via the nonlinear Schrodinger equation. Because it is conventional to solve the equation using numerical methods [416-418], it is less straightforward to retrieve the exact basis of the phenomenon. The supercontinuum effect involves a complex interplay of multiple nonlinear effects, which in combination leads to spectral broadening. These effects include self-phase modulation, cross-phase modulation, Raman scattering, four wave mixing, modulation instability, soliton formation and soliton fission [419]. To increase nonlinearity, structures with high effective area [420] (such as microstructured optical fibers - which also enables optical dispersion tailoring [419]) are typically used.

The optical output from the supercontinuum laser is typically pulsed, although continuous wave supercontinuum lasers have also been realised [421,422]. There are generally two approaches to generate supercontinuum optical signal: by sending optical pump input to a nonlinear medium in the normal or anomalous dispersion region. A comparison of the two approaches are summarized in **Table 6.1**. As indicated in the table, to generate a supercontinuum optical output with high spectral width, the pump signal is sent in the anomalous dispersion region; albeit with lower spectral flatness with less regular and consistent pulses. The Fianium WhiteLase Micro white laser is pumped at anomalous dispersion region to generate optical output with spectral range from 0.4 µm to 2 µm [423]. The optical output consists of a train of ~ 10 ps pulses with ~ 30 MHz repetition rate is produced at ~ 0.2 – 0.5 W total output power (~ mW/nm power density).

Spectral filter. The spectral filter (SuperK SPLIT) is used to split the optical signal free-space beam from the white laser to two spectral free-space outputs: Visible/Infrared region (400 – 830 nm / 915 – 2400 nm wavelengths) and near infrared/infrared region (400 to 1120 nm / 1180 to 2400 nm wavelengths. The optical signal from the filter is coupled from free-space to optical fibre by fitting it with a separate broadband optical fibre delivery.

Table 6.1. Comparison of supercontinuum generation from pump input sent in normal and anomalous dispersion region.

Parameter	Normal dispersion pumping	Anomalous dispersion pumping
1. Spectral width	**Medium** > 0.5 µm 2-dB spectral bandwidth [424]. Pump pulse induces self-phase modulation, which in turn induces a combination of four-wave mixing and stimulated Raman scattering to broaden the optical spectrum [424-426].	**High** ~ 1 µm 15-dB spectral bandwidth [427,428] Pump pulse induces Raman scattering that forms soliton to broaden the optical spectrum [429]. Spectral width can be further increased when the nonlinear medium has two zero dispersion wavelengths. The wavelength of generated solitons red-shifts to approach the second zero-dispersion wavelength. This effectively acts as a second pump pulse, thus extending the continuum [427,430].
2. Spectral flatness	**High** ~ ± 1 dB uniformity for 0.5 µm wavelength range (1.4 µm –1.8 µm wavelength) [424]	**Low** ~ ± 8 dB uniformity for 1 µm wavelength range (0.5 µm – 1.5 µm wavelength) [427,428,431]
3. Energy efficiency	**Varies** (generally depending on pulse width) ~ 60 % with ~ 0.1 W input pump power (low spectral uniformity (± 10 dB)) [432] ~ 30 % with ~ 0.35 W input pump power [424] ~ 35 % with ~ 1 W input pump power [426] Efficiency generally dependent on pulse width. Picosecond pulses generally enable low power input pump power [432].	**Varies** (generally depending on pulse width) ~ 75 % with ~ 50 W input pump power [428] ~ 40 % with ~ 20 kW input pump power [433] ~ 1 % with ~ 3 mW input pump power (on tapered fibers) [431] Efficiency generally dependent on pulse width. Picosecond pulses generally enable low power input pump power [428].
4. Pulse stability	**High** Low amplitude noise and timing jitter (e.g., amplitude noise is 1 % across 30nm optical bandwidth. Timing jitter is comparable to timing jitter of pump input [434]).	**Medium** Higher timing jitter (e.g., amplitude noise can vary between ~ 10 % - 30 % across 30 nm optical bandwidth. Timing jitter is about an order of magnitude higher than the timing jitter of pump input [434]).

EDFA. The pre-amplifier (FA-15 Pritel) and amplifier (FA-33-IO Pritel) are used to sufficiently amplify the white laser signal (which is typically weak, due to shared energy across the laser spectrum) for both probe and pump purposes. Compared to the Amonics EDFA introduced in **Chapter 4**, this pre-amplifier and amplifier cascade provides a higher amplified optical output. The amplification occurs in two stages: the pre-amplifier converts the weak signal from white laser into an output signal strong enough to be further amplified by the subsequent amplifier. The pre-amplifier can receive input signals from -30 dBm (0.001 mW) to 10 dBm (10 mW), and amplify them up to 15 dBm (31.6 mW), for optical wavelengths between 1.528 to 1.565 µm. The subsequent amplifier can receive input signals from -10 dBm (0.1 mW) to 20 dBm (100 mW), and amplify them up to 33 dBm (2 W), for optical wavelengths between 1.535 to 1.565 μm.

When the amplifier is fed with pulsed optical signals, the maximum optical pulse peak power P_{peak} that the amplifier can sustain defines the amplified pulsed signals - based on the relation: average power $P_{\text{average}} = P_{\text{peak}} \cdot f_{\text{repetition}} \cdot \tau_{\text{pulse}}$, where $f_{\text{repetition}}$ and τ_{pulse} are respectively the pulse repetition rate and temporal optical pulse width. The amplifier has been tested to be able to support $P_{\text{peak}} \sim 500$ W at $f_{\text{repetition}} \sim 10$ GHz and $\tau_{\text{pulse}} \sim 1$ ps. Taking the maximum P_{peak} as 500 W, when $f_{\text{repetition}}$ is decreased, τ_{pulse} has to increase (for example to $\tau_{\text{pulse}} \sim 300$ ps at $f_{\text{repetition}} \sim 30$ MHz). Such constraint on τ_{pulse} can be removed by limiting the power supplied to the amplifier to reduce P_{average} (for example, $P_{\text{average}} \sim$ 150 mW for $\tau_{\text{pulse}} \sim 10$ ps).

OTF. The OTF (OTF-930, Santec Corp.) enables remote tuning of filtered wavelength. It has insertion loss of 6 dB, and 0.3 nm bandwidth at -3 dB optical power. It has 80 nm tuning range, from 1.53 to 1.61 μm; and tuning resolution of 0.01 nm.

Electrical component

Pulse generator. The pulse generator (HP 8133A, Hewlett Packard) generates RF pulses at high modulation speed, to define the optical pulses sent to the AMLE (via EOM). The pulse generator supports 15 MHz to 3.35 GHz modulation speed, with less than 100 ps transition times, ± 100 ps width accuracy and ± 150 ps delay accuracy.

6.2 Pattern recognition

Rapid pattern recognition task is carried out on the AMLE scaled-up hardware chip in **Fig. 6.4a** to verify that once the associations are formed by the AMLEs, simultaneous parallel recognition can be achieved. The pattern-recognition experiment consists of two steps: pattern programming and pattern detection.

6.2.1 Network settings and pattern recognition results

Pattern programming

The experiment setup to programme a set of patterns to AMLEs (on the hardware shown in **Fig. 6.5a**) is depicted in **Fig. 6.8a**. *Input* signals IN_1 to IN_4, controlled by variable optical attenuator (VOA; V1550PA, Thorlabs Inc.), holds the pattern to be encoded to the AMLEs, and are represented as optical signals wavelength-multiplexed from the spectrally filtered (SuperK Split, NKT Photonics) laser (WhiteLase Micro, NKT Photonics).

A separate supervisory ENABLE *Teacher* signal from CW laser (TSL-550, Santec Corp.), likewise controlled by VOA, is distributed to each *Input* signals. The *Teacher* signal is centred at 1553.4 nm wavelength. The *Input* signal wavelengths are C33 to C36 (based on ITU-T G.694.1 spectral grids for dense wavelength division multiplexing WDM): λ = 1550.92 nm, 1550.12 nm, 1549.32 nm, 1548.51 nm. Although these wavelengths are used, the MZMs can be designed and set for many other wavelengths.

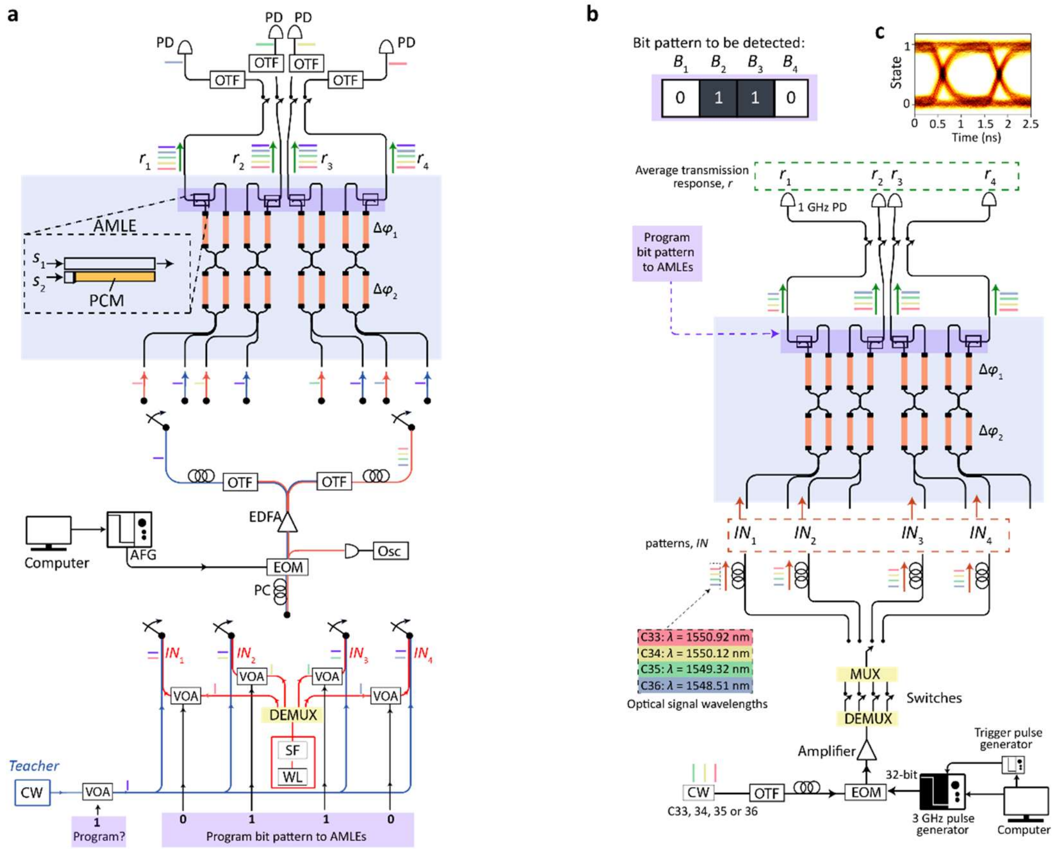

Fig. 6.8. Pattern recognition experiment – measurement setups. a, Schematic of measurement setup to programme bit patterns. ENABLE *Teacher* signal (blue) is distributed to each *Input* signals IN_1 to IN_4 (red; wavelength-multiplexed from spectrally-filtered (SF) white laser (WL)). *Input* is activated using VOA when bit pattern is '1'. **b**, Measurement setup for pattern recognition. Bit patterns to be detected '0110' are associated into the AMLE via a learning step. 1 GHz optical detection pulses defined by computer-generated pseudo-random bits at wavelengths from C33 to C36 (1548.51 nm to 1550.92 nm) are sent through AMLE devices and detected by photodiodes. Measured average output transmission R cumulatively summed to determine if pattern is recognised. **c**, Eye-diagram obtained from non-return-to-zero (NRZ) pseudo-random binary sequence (PRBS) pulses modulated at 1 GHz shows clear distinction between states '0' and '1'.

The selected signal from the combined signals is modulated by EOM (2623 NA, Lucent Technologies) via AFG (AFG 3102C, Tektronix), and amplified by EDFA (FA-15, Pritel and FA-33-IO, Pritel) while monitored using 1 GHz PD (1611FC-AC, Newport Spectra-Physics Ltd.) connected to an oscilloscope (TDS 7404, Tektronix). Amplified signal is split and filtered (OTF-320, Santec Corp.) to the respective *Teacher* and *Input* signals IN_1-IN_4, before they are sent to two 1×4 optical switches (GZ-1×4-SM, Gezhi) and coupled onto the AMLE hardware (via fibre array). To debug the signals sent to AMLE, probe debugging signals are sent through the AMLEs, coupled out of the hardware via fibre array, and filtered by OTF (OTF-930, Santec Corp.) to be detected by PD (2011-FC, Newport Spectra-Physics Ltd.). During pumping, optical switches (GZ-1×2-SM, Gezhi) are activated to null the signals to PD.

A set of amorphising and crystallising pulses defined in **Section 5.1.2** are sent to the AMLE. While the *Teacher* signal sent is continuous wave (CW), the *Input* signals – which define the bit pattern to be detected – are pulsed (as the white laser emits picosecond pulses). As such, the optical power effectively sent to the AMLEs (average power) has to take into consideration the pulse repetition rate and pulse width. The average power is given by $P_{\text{average}} = P_{\text{peak}} \cdot f_{\text{repetition}} \cdot \tau_{\text{pulse}}$, where P_{peak}, $f_{\text{repetition}}$ and τ_{pulse} are respectively the peak power, repetition rate (frequency) and temporal width of the optical pulses. It follows that for a train of optical pulses with 6.6 mW measured average power, 30 MHz repetition rate and 15 ps pulse width (considering pulse broadening effects on the measurement setup), the peak power of the pulse is ~ 15 mW. With both the *Input* optical pulse width τ_{pulse} (~ ps) and optical pulse repetition interval ($1/f_{\text{repetition}}$ = 1/30 MHz ~ 30 ns) being substantially lower than the pulse width defined by the EOM τ (100 ns), the pulses defined by the EOM would cover multiple pulses from the amplified white light signals. Thus, pulse synchronization between the white light laser and EOM is not required.

The GST cells on the AMLE are initially set to their crystalline states. The bit pattern '0110' is then programmed to the AMLEs by respectively sending the pump *Teacher* signal T = 1.47 nJ (τ = 100 ns) and pump *Input* signals IN_1 = 0, IN_2 = 1.47 nJ (τ = 100 ns), IN_3 = 1.47 nJ (τ = 100 ns) and IN_4 = 0 simultaneously – essentially associating the 'Program bit' with the 'Program 0110 bit pattern to AMLEs' signals within the AMLE hardware. After the pulses are simultaneously sent, the bit pattern to be detected – which is '0110' – is effectively stored in the PCM on the AMLE. The PCM structural states are respectively crystalline, amorphous, amorphous, and crystalline on the AMLE hardware. As in the single device measurement case (described in **Chapter 5**), heat from the optical pulses effectively induces GST structural phase switching on the AMLE for the associative learning process.

Pattern detection

After the pattern programming step, rapid pattern recognition task is carried out on the AMLE network chip in **Fig. 6.4a** to verify that once the associations are formed by the AMLEs, simultaneous parallel recognition can be achieved. As summarised in **Fig. 6.8b**, a series of randomised optical binary signals modulated at 1 GHz speed is sent to the AMLEs (with 0 mW for '0' and 0.7 mW for '1'), using different wavelengths (in the range 1548.51 nm to 1550.92 nm, C33 to C36 in the telecoms band) for each of the individual AMLE inputs – see **Fig. 6.9a**. The optical signal source prior to the EOM is the CW laser source (TSL-550, Santec Corp.) The 50 random signals sent to the devices are generated from the following numpy.random.seed in SciPy.org Python module [435]:

(IN_1, IN_2, IN_3, IN_4) at C33: (seed 1, seed 2, seed 1 inverse, seed 2 inverse)
(IN_1, IN_2, IN_3, IN_4) at C34: (seed 3, seed 4, seed 3 inverse, seed 4 inverse)
(IN_1, IN_2, IN_3, IN_4) at C35: (seed 5, seed 6, seed 5 inverse, seed 6 inverse)
(IN_1, IN_2, IN_3, IN_4) at C36: (seed 7, seed 8, seed 7 inverse, seed 8 inverse)

The AFG (AFG 3102C, Tektronix) specifically picks the continuous 32-bit 1 GHz pulses from the 3 GHz pulse generator (HP 8133A, Hewlett Packard). The EDFA (AEDFA-CL-23, Amonics Ltd.) amplifies the pulse from EOM (2623 NA, Lucent Tech.). The AMLE output pulses coupled from the layout are detected by 1 GHz fibre optic PD (1611FC-AC, Newport).

The output from each AMLE device in the scaled-up hardware is channeled to four photodiodes to generate the system outputs (or responses), R. The output transmission response of single AMLE r_i measured from the photodetectors is then summed to obtain the cumulative transmission response $\Sigma_i r_i$ for the different wavelengths, where R_i is the measured output transmission of i^{th} AMLE after normalising (between 0 and 1 range) using $r_i = (r_i - r_{base})/r_{base}$, that is the average value of two consecutive output transmission values using a trigger pulse (with r_{base} being the baseline of the optical transmission. The average value is taken to minimise the effects of electrical or optical noise. The corresponding output response of the AMLEs when the inputs (in **Fig. 6.9a**) are sent at the four different wavelengths C33 to C36 is shown in **Fig. 6.9b**. A high output response above the shown threshold indicates that the bit pattern '0110' is detected.

Having shown in the experiment how the associative learning hardware can detect a programmed bit pattern individually with respect to the optical signals sent at the different wavelengths, we now set the goal of systematically retrieving the cycle(s) at which the cumulative output response R is above a pre-determined learning threshold. In the experiment, the cumulative response R is recorded from the output pulses retrieved at the four optical wavelengths. The results, shown in **Fig. 6.9c**, pinpoints cycle 30 as the cycle at which the input signals from all four wavelengths match the bit pattern '0110', as the cumulative response R exceeds the learning threshold only at the cycle. This ability to simultaneously detect matching patterns could find useful practical applications in parallel object detection. The parallelization increases object detection throughput. The extent of latency reduction due to such throughput improvement scales linearly with the number wavelength-multiplexed optical carrier signals used in the detection process.

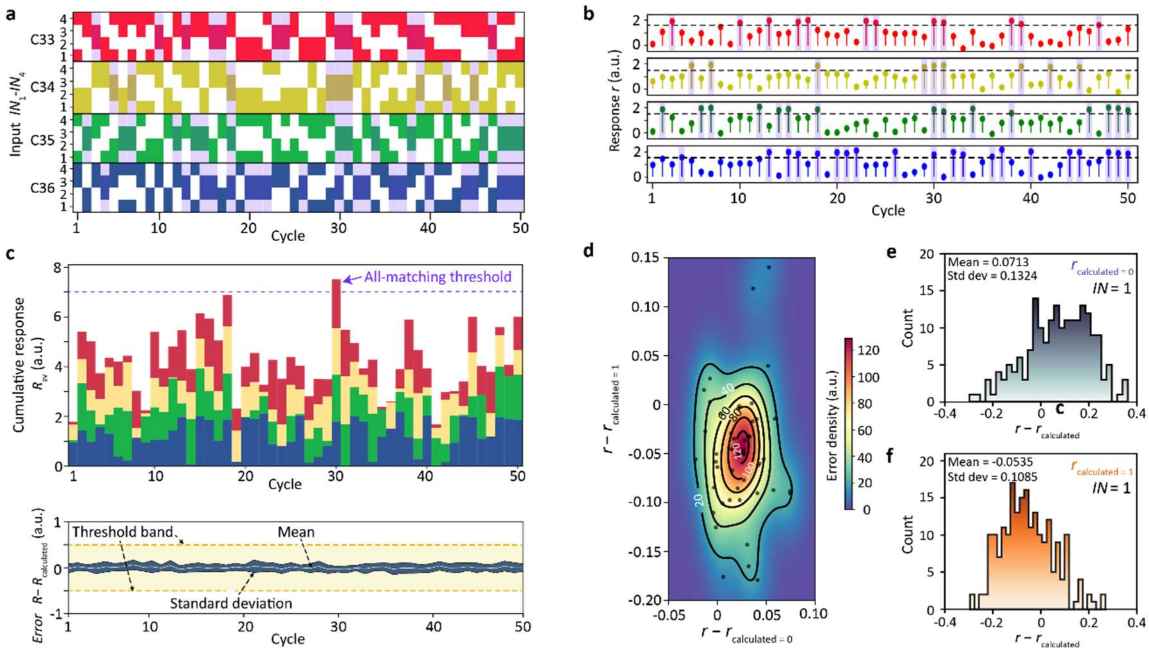

Fig. 6.9. Pattern recognition experiment results. **a**, Inputs IN_1 to IN_4 sent at the four different wavelengths i.e., C33 to C36 (input = '1' represented in red, yellow, green and blue respectively). Purple shades represent '0110' pattern sent to the devices. **b**. Output response r when inputs IN_1 to IN_4 are sent for the four different wavelengths (C33 to C36). Purple shades represent '0110' pattern detected. **c**, (Top) Cumulative response R obtained for the four respective wavelengths and detection cycles to recognize the same pattern '0110'. Values above cumulative transmission threshold $R = 7$ (the dashed black line) denote all-matching response (that is in cycle 30). Notably, the cumulative response R in cycle 18 is close to the threshold: upon closer inspection from **b**, the cycle has three (out of the total four) matching responses. (Bottom) Calculated error $R - R_{calculated}$ (difference between measured average transmission response and expected response) shows mean error is negligibly low with standard deviation well within the ±0.5 threshold band. **d-f**, Pattern detection error analysis. In (**d**), image plot of kernel density estimation of calculated *Error* = $r - r_{calculated}$ when $r_{calculated}$ is 0 (horizontal axis) and 1 (vertical axis). Distribution is skewed due to contributions from null signals when $IN = 0$. In (**e**) and (**f**), histogram of distribution counts of *Error* when $IN = 1$, with $r_{calculated} = 0$ and 1 respectively.

6.2.2 Pattern recognition error analysis

The bottom plot of **Fig. 6.9c** shows the performance of the AMLEs for detection predictions. The prediction error $Error = R - R_{\text{calculated}}$ is measured from the difference between output transmission R and the calculated transmission $R_{\text{calculated}}$. The error is within the ± 0.5 threshold bands. **Fig. 6.9, d-e** provides a more thorough analysis of the calculated *Error*. In **Fig. 6.9d**, the kernel density estimation of the *Error* is shown. The estimation is centered at around (0.03, − 0.05) with skewed distribution (elongated in the y-axis) because the null output signals from $IN = 0$ are considered in the plot. As shown in the distribution counts in **Fig. 6.9, e-f**, the *Error* is generally within ± 0.2.

6.3 Image recognition

Having demonstrated pattern detection, we next discuss how the AMLE-based hardware can achieve generalisation on an image recognition task using associative learning.

6.3.1 Scaling architecture of associative learning network

To demonstrate generalisation, the network architecture shown in **Fig. 6.10a** is experimentally implemented. The network is similar to those in **Fig. 6.1d**, which consists of three main network layers (that is the input layer, associative layer and output layer) [436].

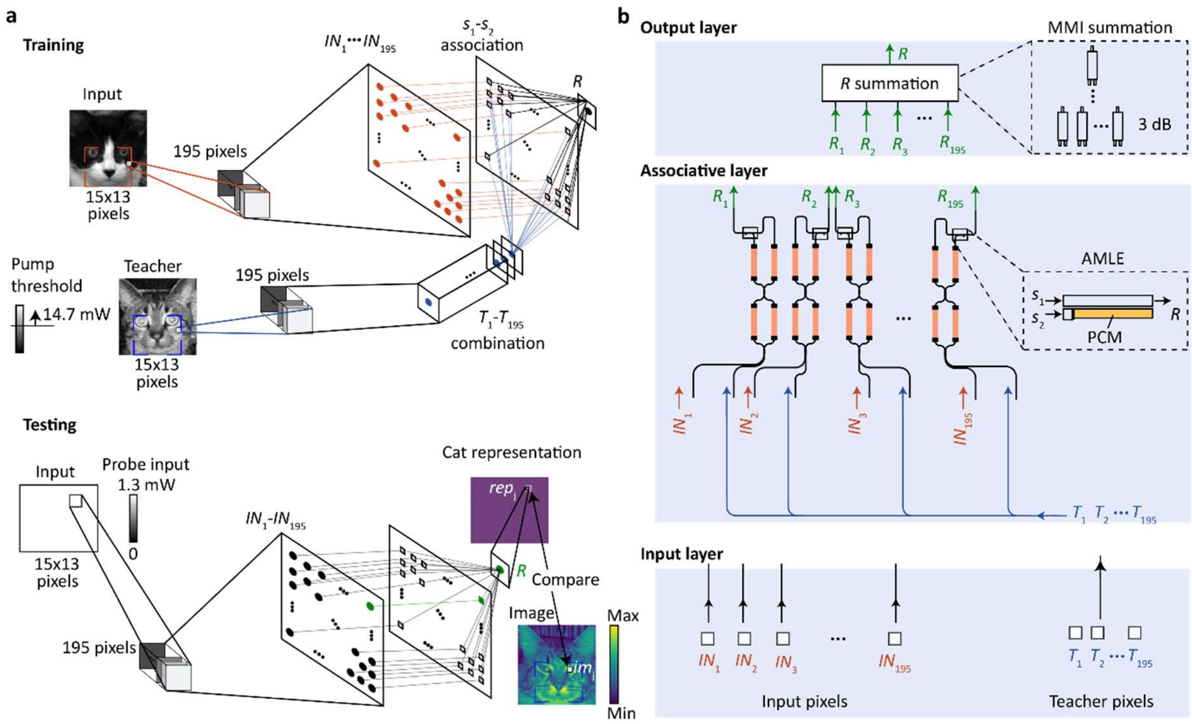

Fig. 6.10. Scaling architecture for image recognition using associative learning network. a, The general associative neural network composed of an input layer (IN_1-IN_{195}, T_1-T_{195}), associative layer (stimuli s_1-s_2 association) and output layer (transmission R). The input signal is the pattern (pixels from

image) to be classified, and the external teacher provides the desired set of output values chosen to represent the class of patterns being learned. During training, the input layer (IN_1-IN_{195}, T_1-T_{195}) is fed into the associative layer. Associative layer consists of AMLEs with states that are modified when both the input and external teacher signals are paired together. A model representation of the trained images is then obtained by sending a preprocessed input signal of blank white image to propagate through the layers. **b**, Optical on-chip implementation of the input layer, associative layer and output layer respectively. The associative layer, which consists of thermo-optic NiCr heaters, distributes the combination of the input and external teacher signals (from the input layer) as s_1 and s_2 inputs to the respective AMLEs. The output layer consists of a summation unit to sum up the output response from the AMLEs to form the net output.

During the training process, images to be trained are first pixelated to 15×13 input pixel data (195 pixels altogether). These data IN_1-IN_{195} are then sent to the associative layer, whilst simultaneously the *Teacher* signals are likewise preprocessed and sent to the associative layer. The change in the learning states of AMLEs is determined from their transmission readouts, when preprocessed *Input* data of maximum amplitude and the same dimensions as the training pixel data (blank white image) is fed through the layers. The transmission of these individual AMLEs is summed up at the output layer and rearranged to form a 15×13-pixel model representation. The complete computation is obtained by cumulatively adding the model representation from each training pair. Pixel-by-pixel comparison between the model (which generalises the training images) and measurement is then made to determine if the testing images are of the image to be detected.

6.3.2 On-chip implementation of associative learning network

The photonic implementation of associative learning network architecture for the image recognition is shown in **Fig. 6.10b**. Here, the GST elements on the AMLEs are first initialised to the crystalline state before the signals are sent to the on-chip photonic structure. During training, depending on the optical input signal and the input pump power sent to the AMLEs, the state of the GST cells on AMLEs either remain in crystalline, or structurally switch to amorphous state. This results in a change in optical probe output response of the AMLEs. In our experiment, because we only have 4 AMLEs, the 195 pairs of input pixels are rastered across the 4 AMLEs. The rastering process is carried out by sweeping horizontally from left-to-right four training pixels at one go and recording the measured probe output responses from the computer, before resetting the AMLEs to their crystalline states, and then repeating the process until the 195 pairs of input pixels are completely covered. As the rastering process is implemented through the 195 pairs of input pixels, the cat representation is built from the probing of the respective AMLE output responses. Due to limitations on the number of AMLEs (four in this case) in the associative learning hardware, the current image recognition system is dependent on an external memory from the computer to perform the post processing step of building the cat representation. Such dependence can be removed using a hardware with AMLEs of at least the same number of image pixels to be trained and tested. The overall detection speed is projected to be ~ µs using high-speed modulators [142] with a complete set of AMLEs on the hardware (195 AMLEs in this case), limited only by the phase modulation speed and computer processing latency. There is an inherent ~ 3 dB loss from the use of conventional MMI for the summation process, which is used in the on-chip realization of scaled-up associative hardware.

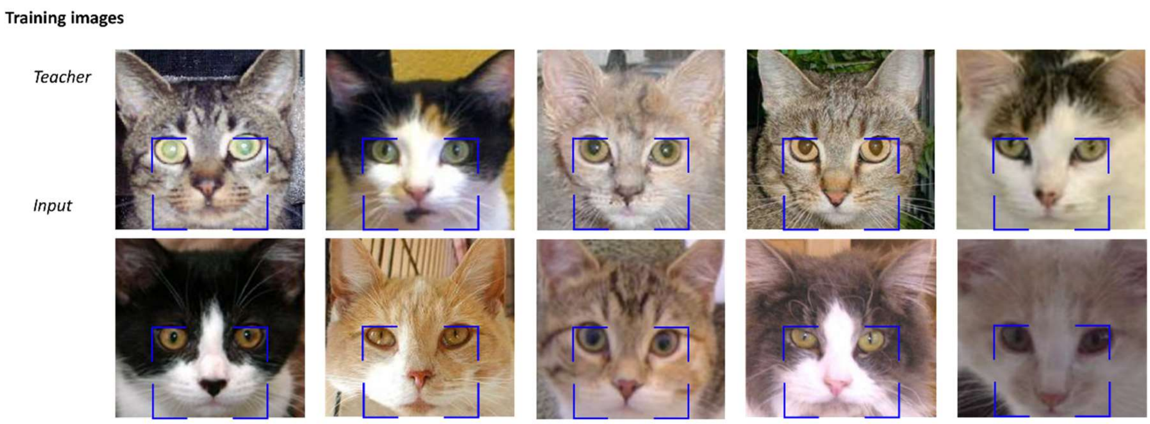

Fig. 6.11. Training the associative learning network to identify cat images. Bounding box corners indicate region of interest.

To examine image classification capabilities of the associative learning network, the 'Dogs vs Cats' dataset from fast.ai [437] (single dataset collected from CIFAR10, CIFAR 100 [2], Caltech 101 [438], Oxford-IIIT Pet [439] and Imagewoof [440]) is examined. Cat images used for the training process are shown in **Fig. 6.11**.

Fig. 6.12 shows the integrated pump-probe measurement setup for the image recognition experiment. *Input* signals IN_1-IN_{195} are represented as wavelength-multiplexed optical signals from the spectrally filtered (SuperK Split, NKT Photonics) supercontinuum laser (WhiteLase Micro, NKT Photonics). Another set of separate supervisory *Teacher* signals T_1-T_{195} is represented from CW laser (TSL-550, Santec Corp.) to each *Input* signals. These signal pairs are rastered across the 4 hardware AMLEs (shown in **Fig. 6.4a**), details of which are described earlier. The multiplexed wavelengths are C33 to C36 (λ = 1550.92 nm, 1550.12 nm, 1549.32 nm, 1548.51 nm) for the *Input* signals. The *Teacher* input signal is centred at 1553.4 nm.

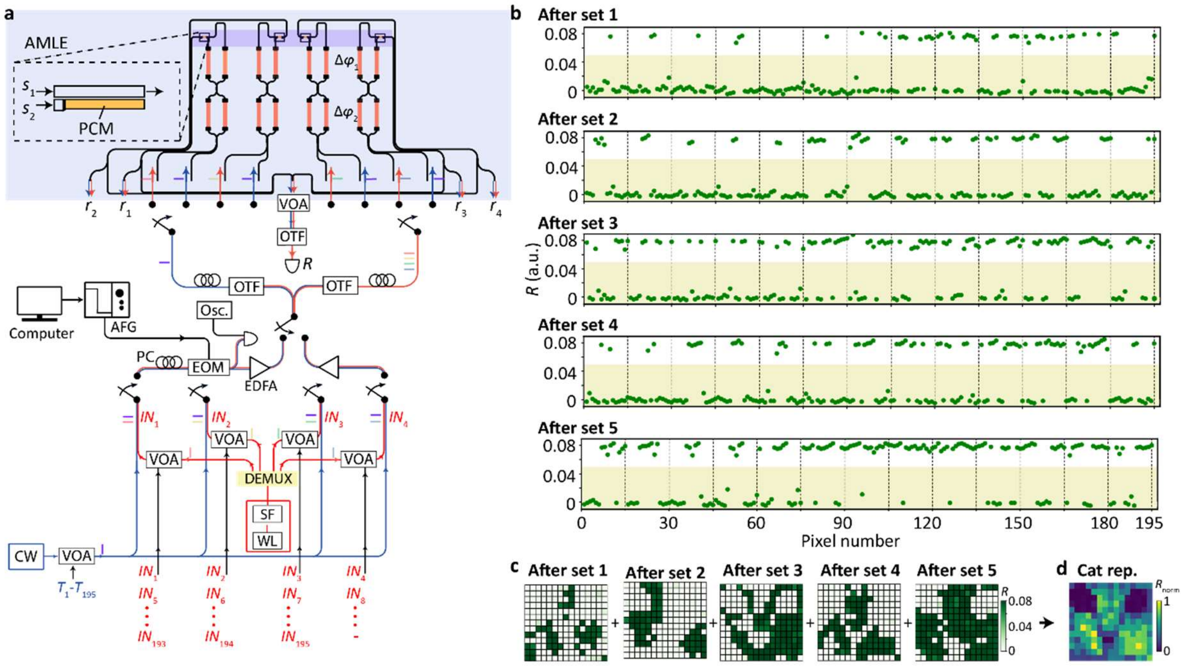

Fig. 6.12. Image recognition experiment – measurement setup and transmission data. a, Measurement setup. A set of supervisory *Teacher* signals (blue) is distributed to each of the *Input* signals (red; wavelength-multiplexed from spectrally filtered (SF) white laser (WL)) and switched between pump and probe lines. 195 pairs of *Teacher* and *Input* pixels are rastered across the four hardware AMLEs. **b**, Plot of measured output transmission R after the training process (for training sets 1-5 shown in **Fig. 6.11**). Transmission values are divided into that above and within the yellow bands, bordering at threshold R = 0.05. **c**, Transmission plots rearranged to form a 15×13-pixel image for each training sets. **d**, Normalised cat representation after cumulatively adding the transmission values in (**c**).

For both *Input* and *Teacher* training signals, the threshold for the grayscale pixels (normalised to unity) is set at 0.5 such that the pump signals for each pixel combinations are sent to AMLE at either 0 or 1.47 nJ (τ = 100 ns) for our case. These signals are combined and selectively routed to either probe line or pump line amplified by EDFA (FA-15, Pritel and FA-33-IO, Pritel). The routed signals are then channeled to an optical coupler each to split them at 50:50 ratio and filtered by OTFs (OTF-320, Santec Corp.) such that they now consist purely of *Teacher* or *Input* pump/probe signals. These signals are then connected to two 1×4 optical switches (GZ-1×4-SM, Gezhi) and coupled onto the on-chip hardware (via fibre array). The out-coupled output transmission R is sent to the PD (2011-FC, Newport Spectra-Physics Ltd.) through a VOA (V1550PA, Thorlabs Inc.) and OTF (OTF-930, Santec Corp).

6.3.3 Model representation of training images

After training, a blank white 15×13 pixelated image of maximum probe input power magnitude is fed to the on-chip hardware and measure *R* for each training set (five sets altogether, details in **Fig. 6.11**). Patterns (in the form of pixel amplitudes) are searched from the 15×13 pixel image sent to the network of monadic AMLEs in single steps. The *R* plots shown in **Fig. 6.12b** are then rearranged to form a 15×13-pixel image: these are shown in **Fig. 6.12c**, with upper and lower cutoffs respectively at $R = 0.08$ and $R = 0$. The images constitute the cat representation for the training sets. The net cat representation - depicted in **Fig. 6.12d** - is obtained by cumulatively summing up the transmission values *R* of the cat representations (for each pixel location) from the training iterations. After each training iteration, the network's ability to distinguish the appearance of a cat improves. The feature subtleties can then be captured from the net cat representation, giving us a valuable means to distinguish a cat from other objects.

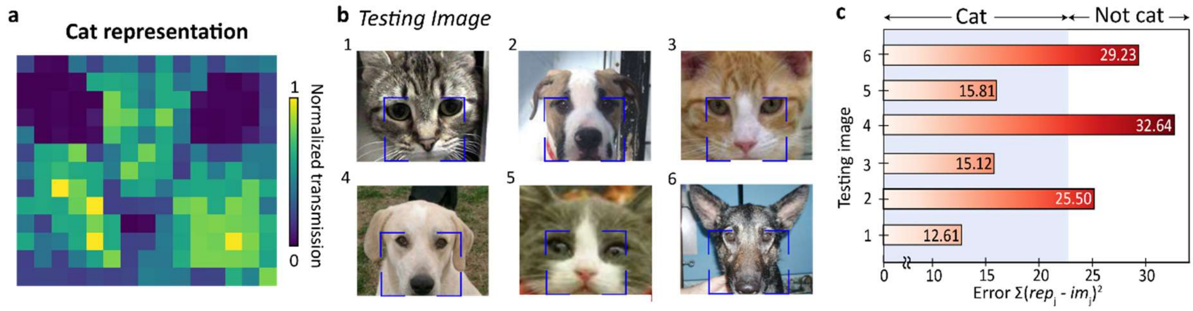

Fig. 6.13. Cat representation and testing images: comparison. a, Cat representation obtained from measurements (**Fig. 6.12d**). **b**, Testing images to test if the network can correctly classify pictures as cat and non-cat. **c**, The network successfully recognised cat and non-cat images based on the squared Euclidean distance $\Sigma\,(rep_j - im_j)^2$ measured over the testing images for each of the 15×13 pixels.

To test the model representation, the testing images shown in **Fig. 6.13b** are used to compare them (*im*) with the representation of the cat *rep* in **Fig. 6.13a**. The threshold $min(Error) + (max(Error) - min(Error))/2$, which is 22.625 for our case, is set to determine the *Testing Images* that resemble the model representation in **Fig. 6.12d**. The results, summarised in **Fig. 6.13b**, reveal that *Testing Image* 1, 3 and 5 resemble the cat representation. With the error (*Error* = squared Euclidean distance $\Sigma\,(rep_j - im_j)^2$) of Image 2, 4, 6 above the threshold, the network predicts that these images are not of a cat. The associative learning network thus accurately classifies images of cats from the model representation obtained from the training iterations.

In the experiment, the AMLE hardware is able to associate (but not de-associate) pixels. It is expected that a hardware that can both associate and de-associate will be of considerable practicality, given the behavioural tendency of biological beings to remove non-reinforced conditioned responses over time [441]. The ability to de-associate enables hardware systems to invalidate associations formed by erroneous and/or spurious signals: this would otherwise be mistakenly regarded as valid data signals. On a more far-sighted perspective, such ability to de-associate would enable hardware systems to dynamically learn associations – by shedding associations that would eventually turn out to be little to no relevance, while forming new associations of greater significance at a particular time of computation.

6.4 Evaluation metrics for AMLE

The application-specific system offers convergence in one training step due to its straightforward approach of detecting similarities without having to randomise network weights and perform back-propagation as in conventional neural networks. Relevant device-level evaluation metrics can be identified by contextualising the AMLE with a typical machine learning data load, which requires data to be transferred back and forth from data source to be run using cloud computing and/or supercomputers. For a more energy-efficient locally-run neural network, it is important to shrink the network for greater portability and reduce the energy consumption of the learning process. **Table 6.2** summarises the minimum active device volume (volume of GST) and learning energy of other associative learning devices [317-330].

Table 6.2. Comparison of active volume and learning energy in associative learning devices

Type	Active volume (μm^3)	Min. learning energy (nJ)	Ref.
Electronic			
▪ Memresistive			
i. Chalcogenide	0.12 – 10.5	4.7×10^4	317
	8	2.63	318
ii. Manganite	~ 0.1	1.35×10^3	319
	1.25×10^{10}	1.02×10^5	320
iv. Nickelate	4.7×10^3	7.20×10^5	321

	4.8×10^4	2.04×10^5	322
v. Metal oxide	$\sim 10 - 900$	$\sim 10^3 - 10^5$	323,328
vi. Organic	$\sim 0.1 - 0.5$	$\sim 10^3 - 10^4$	324,329
▪ Electrochemical	6×10^3	6×10^4	325
	9.6×10^5	125	326
▪ Memcapacitive	26.9	~ 30	327
Optoelectronic	1.62×10^3	2.1×10^3	330
Optical AMLE	0.12 (excluding volume of integrated waveguide heaters, which is 4.4×10^5 µm^3 in this work)	1.8 (excluding energy required to prepare the modulation state of the integrated waveguide heaters, which is 2.1×10^9 nJ in this work)	(**this work**)

The electronic and optoelectronic associative learning devices range from $\sim$ 0.1 to 10^{10} µm^3 in active volume and consume $\sim$ 2.63 to 10^5 nJ of energy per learning event. In comparison, the all-optical AMLE fares favourably relative to these devices in terms of dimensions and energy usage, with a low active volume at 0.12 µm^3 and minimum learning energy at only 1.8 nJ. The overall size of the single-element device is of dimensions 2 µm × 17 µm × 0.33 µm. The calculation steps to obtain the values in **Table 6.2** are:

a. Memristive

Ziegler, M., Soni, R., Patelczyk, T., Ignatov, M., Bartsch, T., Meuffels, P., Kohlstedt, H. An electronic version of Pavlov's dog. *Adv. Function. Mater* **22**, 2744–2749 (2012). doi:10.1002/adfm.201200244 [317]

Active volume calculations

From 'Section 4: Experimental Section' (Page 2748 and 2749 of manuscript),

Active area	=	2 µm^2 to 150 µm^2
Thickness of chalcogenide, $Ge_{0.3}Se_{0.7}$	=	60 nm to 70 nm

Thus, active volume	=	Area x thickness
	=	(2 µm^2 to 150 µm^2) × (60 nm to 70 nm)

$= \quad 0.12\ \mu m^3$ to $10.5\ \mu m^3$

<u>Minimum learning energy calculations</u>

Assume $I_{s2} = I_{s1}$

$I_F = I_m = 0.05$ mA

$E_{total} = E_{s1} + E_{s2}$

$= V_{s1} \times I_{s1} \times \Delta t + V_{s2} \times I_{s2} \times \Delta t$

$= (0.2\ V) \times (0.025\ mA) \times 3.75 + 0.3 \times 0.025 \times 3.75$

$= 47\ \mu J$

b. Electrochemical

Wan, C., Zhou, J., Shi, Y., Wan, Q. Classical conditioning mimicked in junctionless IZO electric-double-layer thin-film transistors. *IEEE Electron. Dev. Lett.* **35** (2014). doi:10.1109/LED.2014.2299796 [325]

<u>Active volume calculations</u>

From 'Section II. Experimental details' (Page 414 of manuscript)

Area of IZO film $= 150\ \mu m \times 1000\ \mu m$

Thickness of IZO film $= 40$ nm

Thus, active volume = Area x thickness

$= (150\ \mu m \times 1000\ \mu m) \times 40\ nm$

$= 6 \times 10^3\ \mu m^3$

<u>Minimum learning energy calculations</u>

From Figure 3b (Page 415 of main manuscript),

$V_{s1} = 2$ V

$V_{s2} = 2$ V

$I_{s1} = 100\ \mu A$

$I_{s2} = 20\ \mu A$

$n_{s1\ pulse} = 5$ pulses

$n_{s2\ pulse} = 5$ pulses

From 'Section III. Results and discussion' (Page 416 of main manuscript),

$\Delta t = 50$ ms

Thus, $E_{total} = E_{s1} + E_{s2}$

$= V_{s1} \times I_{s1} \times \Delta t \times n_{UCS\ pulse} + V_{s2} \times I_{s2} \times \Delta t \times n_{s2\ pulse}$

$= 2\ V \times 100\ \mu A \times 50\ ms \times 5$ pulses $+ 2\ V \times 20\ \mu A \times 50\ ms \times 5$ pulses

$= 6 \times 10^4\ nJ$

c. Memcapacitive

Wang, Z., Rao, M., Han, J.-W., Zhang, J., Lin, P., Li, Y., Li, C., Song, W., Asapu, S., Midya, R., Zhuo, Y., Jiang, H., Yoon, J. H., Upadhyay, N. K., Joshi, S., Hu, M., Strachan, J. P., Barnell, M., Wu, Q., Wu, H., Qiu, Q., Williams, R. S., Xia, Q., Yang, J. J. Capacitive neural network with neuro-transistors. *Nat. Commun.* **9**, 3208 (2018). doi:10.1038/s41467-018-05677-5 [327]

<u>Active volume calculations</u>

Estimating from the scanning electron micrograph in Figure 1a (Page 3 of manuscript),

Active area	=	$(18.98 \times 23.33)\ \mu m^2$
Active thickness	=	60.76 nm
Thus, active volume	=	Area $\times$ thickness
	=	$(18.98 \times 23.33)\ \mu m^2 \times 60.76$ nm
	=	$26.9\ \mu m^3$

<u>Minimum learning energy calculations</u>

From Figure 2 caption (Page 5),

Capacitance of synapses = 680 pF

From Figure 3c (Page 6),

$V_{s1} = V_{s2} = 2$ V

The current pulse magnitude is significant after 12th *UCS* and 11th *NS*/*CS* pulse

Thus, E_{total}	=	$E_{charging}$
	=	$0.5\ CV^2 \times n_{pulses}$
	~	$0.5 \times (680 \times 10^{-12}) \times (2\ V)^2 \times$ (12 pulses for s_1 + 11 pulses for s_2)
	=	30 nJ

Despite the relatively low AMLE active volume (of GST), compared to electronic hardware, the devices are still larger; however, the speed of photonic devices can lead to higher computational density. Experimentally, with two coupled waveguides (width = 0.9 µm each and length = 17 µm) in the AMLE, the area of our AMLE is calculated to be 2 µm × 17 µm. Thus, in the experiment with 1 GHz modulation speed with 4 multiplexed wavelengths, this gives computational density of (4 × 1 GHz) / ($34\ \mu m^2$) = 118 TOPS/mm². The 118 TOPS/mm² computational density is limited by the available setup in the lab. Scaling the device to more wavelength channels and modulating at higher speed potentially leads to significant gains in the compute density. Considering for example modulation at 50 GHz [142] and using 16 wavelengths, the device would deliver approximately (16 × 50 GHz) / ($34\ \mu m^2$) = 2.5×10^4 TOPS/mm². The overall net energy usage of the learning system is directly related to the number of iterations required to train a learning system. The energy-speed tradeoff – particularly evident in CMOS devices [442] – is another important evaluation metric that requires further investigation and research. The all-optical AMLE learns in ~ 100 ns, compared to ~ ms in a previous associative learning device [318]. Commercial photonic foundry services could facilitate further scale-up efforts.

We have mapped the concept of supervised learning in AI with Pavlovian associative learning, and presented a simple architecture to solving pattern and image recognition tasks. Equipment in addition to that is discussed in Chapter 4 are indicated (HP 8133A pulse generator, FA-15 and FA-33-IO optical amplifiers, DAC and the specially-designed peripheral amplifier circuit). High-speed (1 GHz) parallel recognition are experimentally demonstrated using WDM, and generalised patterns in the form of images by experimentally building a model representation of the training images presented. Comparison on active volume and minimum learning energy shows that the AMLE (device realised by the author) fares favourably compared to other reported associative learning devices. The computational density of the device is calculated to be ~ 118 TOPS/mm^2, and projected to reach ~ 2.5 × 10^4 TOPS/mm^2 using 16 wavelengths at 50 GHz detection modulation speed. This is several magnitudes higher than that in current state-of-the-art photonic tensor core that has 162 TOPS/mm^2 computational density using 16 wavelengths at 18 GHz detection modulation speed [276].

Chapter 7

7. Associative learning: Summary and outlook

This chapter compares associative learning and 'deep neural network' (a conventional machine learning approach), and lays out open questions about associative learning. Then, other associative learning devices are reviewed.

7.1 Associative learning for AI hardware acceleration

We have discussed how associative learning scaled-up hardware can be used to solve pattern recognition and image recognition tasks. To illustrate the general features of learning methods, the governing equations of deep neural network (using backpropagation method) and associative learning is derived in this section. After identifying the limitations in the 'symbolic' model applied in **Chapter 6**, a 'numeric' use of associative learning that enables extracting complex salient features in data is briefly discussed.

7.1.1 Deep neural network (using backpropagation method)

It is conventional to apply backpropagation method using deep neural network (DNN) architecture to solve supervised learning tasks. In the method, calculations proceed backwards across the network to update the network weights, such that the error gradient tends to zero. Ideally, after each forward-backward iteration, the output better resembles the desired output.

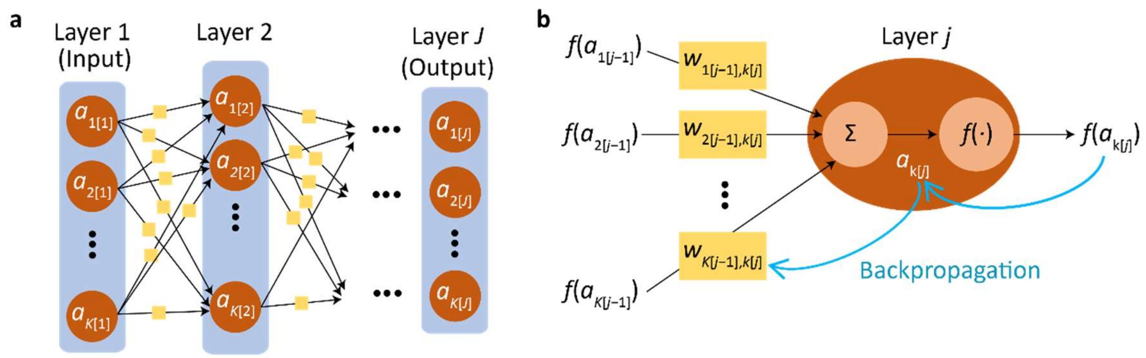

Fig. 7.1. Deep learning network architecture using gradient descent algorithm. a, Deep learning consists of at least one hidden layer of nodes (that is, at least two layers of weights). Nodes (red circle) in each layer are connected to those in subsequent layers with connection strengths represented by

weights (yellow). **b**, Model of deep learning network node. Except for those in input layer, each node is a nonlinear activation function $f(\cdot)$ of the weighted sum of the nodes in previous layer ($f(a_{\text{k [layer j-1]}}) \times w_{\text{node } k \text{ [layer } j]}$). Gradient of error function is computed by layer by layer.

As summarised in **Fig. 7.1**, the network output $f(a_{k'[j=J]})$ of the k^{th} node in layer j $a_{k[j]}$ is compared with target function $t_{k'[j=J]}$ for all $k'[j=J] \in \{1,2,\dots, K[j=J]\}$ at each forward iteration, before the weights are successively adjusted layer-by-layer in the backward direction (from layer J to layer 1). The calculation of error E at learning rate η for connection weight $w_{k[j],\, k[j+1]}$ between nodes $k[j]$ between $k[j+1]$ relies on differentiation, based on the update rule:

$$\Delta w_{k[j],\, k[j+1]} = -\eta \frac{\partial E}{\partial w_{k[j],\, k[j+1]}} \tag{7.1}$$

The derivative exploits the chain rule (to increase computational efficiency) [30]:

$$\frac{\partial E}{\partial w_{k[j],\, k[j+1]}} = \begin{cases} \dfrac{\partial E}{\partial a_{k[j]}} a_{k[j]} & ,\text{if } j = 1 \\ \dfrac{\partial E}{\partial a_{k[j]}} f(a_{k[j]}) & ,\text{if } j \neq 1 \end{cases} \tag{7.2}$$

with the right-hand side derivative of **(7.2)** in turn, given by

$$\frac{\partial E}{\partial a_{k[j]}} = f'(a_{k[j]}) \left(\sum_{k'=k[j'+1]}^{K[j'+1]} \cdots \sum_{k'=k[j']}^{K[j']} \prod_{j'=j+1}^{J} f'(a_{k'})\, w_{k',k'+1} \right) \sum_{k'=k[J]}^{K[J]} \left(f(a_{k'}) - t_{k'} \right) \tag{7.3}$$

7.1.2 Associative learning

The scaled-up architecture adopted in the experiments (summarised in **Fig. 7.2a**) is distinct from deep learning network. In the architecture, there are three layers (input j=1, associative j=2 and output j=3 layers, with each layer having K [j=1], K [j=2], K [j=3] number of nodes respectively). This is after an additional layer that maps the inputs to their input combination to the architecture is included to also consider nonlinear classification tasks, for completeness.

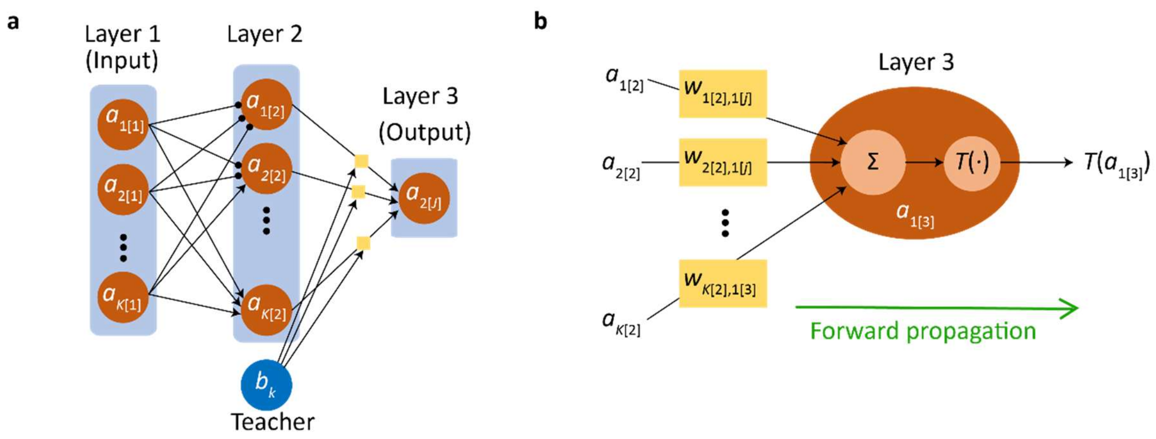

Fig. 7.2. Associative learning network. a, Associative learning consists of one hidden layer of weights. Outputs from layer j=2 nodes are sent together with the teacher signals b_k to the weights. Weights are adjusted when both layer j=2 outputs and b_k carries a signal. Node in layer j=3 outputs the thresholded weighted sum of layer j=2 outputs. **b.** Computation proceeds in forward direction.

Details of the layers are summarised as follows:

- <u>Input (j=1) layer</u>. This layer takes the combinations of all j=1 nodes $a_{k[j=1]} \in \{0,1\}$.

- <u>Associative (j=2) layer</u>. This layer 'associates' the signal received from input (j=1) layer with the *Teacher* signals. The number of nodes in layer j=2 is the total number of input combinations of all j=1 nodes (that is $K[j=2] = 2^{K[j=1]}$).

- <u>Output (j=3) layer</u>. This layer sums the respective layer j=2 weights $w_{k[j=2]}$ multiplied by their corresponding nodes $a_{k[j=2]}$.

The output node, depicted in **Fig. 7.2b**, takes the form

$$a_{k[j=3]} = \sum_{k'=k[j=2]}^{K[j=2]} w_{k'} a_{k'} \tag{7.4}$$

and is sent to the threshold function $T_h(\cdot)$ that produces an output if the weighted sum exceeds a certain value. When $a_{k[j=2]}$ (which is from *Input* $a_{k[j=1]}$) and *Teacher* $b_k \in \{0,1\}$ signals are simultaneously sent, change in $w_{k[j=2]}$ is induced. Its extent depends on the state of the current weight $w_{k[j=2]}$ relative to the maximum weight w_{max}:

$$\Delta w_{k[j=2]} = \begin{cases} \eta_{\text{train}}\ a_{k[j=2]}\ b_k & , w_{k[j=2]} < w_{\max} \\ w_{\max} & , \text{otherwise} \end{cases} \tag{7.5}$$

for learning rate η_{train}. Weight $w_{k[j=2]}$ is iteratively updated as

$$w_{k[j=2]}^{(m+1)} = w_{k[j=2]}^{(m)} + \Delta w_{k[j=2]} \tag{7.6}$$

or (with respect to net weight at m = 0):

$$w_{k[j=2]}^{(m=n)} = w_{k[j=2]}^{(m=0)} + \sum_{m=0}^{n} \Delta w_{k[j=2]} \tag{7.7}$$

For the nontrivial case when there exists $k'[j=2] \in \{1,2,\ldots, K[j=2]\}$ for which $\Delta w_{k'[j=2]} \neq 0$, the summation term for the particular $k'[j=2]$ reduces to $n\eta_{\text{train}}$. Given convergence when $w_{k'[j=2]}^{(m=n)} = w_{\max}$ and initial weight $w_{k'[j=2]}^{(m=0)} = 0$, the number of iterations to convergence $n = w_{\max}/\eta_{\text{train}}$. This indicates single-shot learning (n=1) when $\eta_{\text{train}} = w_{\max} = 1$.

7.1.3 Convergence comparison

Given the symbolic nature of the training process in the associative learning scaled-up model, the model would expectedly deliver significantly faster training convergence as we scale up the training task. For the sake of comparison, simulations on solving learning tasks are carried out, using the associative learning network and other conventional methods: deep neural network (DNN, using gradient descent algorithm GD) and evolving neural network (ENN, using genetic algorithm GA) [443].

In the simulations, the *Inputs* of associative learning network are mapped to their combinations before they are paired to their corresponding *Teacher* signals. The DNN has one hidden layer of nodes (that is, two hidden layer of weights), with the number of hidden layer nodes arbitrarily set as the round down of 1.5 times the number of inputs (that is, 3, 6 and 10 hidden nodes when the number of inputs is respectively 2, 4 and 7). Logistic sigmoid is used as the nonlinear function, as is convention for binary classification problems. To promote convergence, the weights in DNN are randomised using the random generator function seed (0) to seed (19) from Numpy [435]. The ENN parameters in the simulations are based on default settings in [443]. **Table 7.1, a-c** shows the training data used in the comparison with two, four and seven *Input* (*IN*) signals (which has 2^2 = 4, 2^4 = 16 and 2^7 = 128 signal combinations). The corresponding expected outputs R are randomly generated using the random generator function seed(0) from NumPy library [435].

Table 7.1. Training data for the comparison. Randomly generated [435] output R (actual R) for two (**a**), four (**b**) and seven (**c**) *Input* (*IN*) signals.

a

IN		***R***
1	**2**	
0	0	0
0	1	1
1	0	1
1	1	0

b

IN				***R***
1	**2**	**3**	**4**	
0	0	0	0	0
0	0	0	1	1
0	0	1	0	1
0	0	1	1	0

IN				***R***
1	**2**	**3**	**4**	
0	1	0	0	1
0	1	0	1	1
0	1	1	0	1
0	1	1	1	1

IN				***R***
1	**2**	**3**	**4**	
1	0	0	0	1
1	0	0	1	1
1	0	1	0	1
1	0	1	1	0

IN				***R***
1	**2**	**3**	**4**	
1	1	0	0	0
1	1	0	1	1
1	1	1	0	0
1	1	1	1	0

c

IN							***R***
1	**2**	**3**	**4**	**5**	**6**	**7**	
0	0	0	0	0	0	0	0
0	0	0	0	0	0	1	1
0	0	0	0	0	1	0	1
0	0	0	0	0	1	1	0
0	0	0	0	1	0	0	1
0	0	0	0	1	0	1	1
0	0	0	0	1	1	0	1
0	0	0	0	1	1	1	1
0	0	0	1	0	0	0	1
0	0	0	1	0	0	1	1
0	0	0	1	0	1	0	1
0	0	0	1	0	1	1	0
0	0	0	1	1	0	0	0
0	0	0	1	1	0	1	1
0	0	0	1	1	1	0	0
0	0	0	1	1	1	1	0
0	0	1	0	0	0	0	0
0	0	1	0	0	0	1	0
0	0	1	0	0	1	0	0
0	0	1	0	0	1	1	1
0	0	1	0	1	0	0	0
0	0	1	0	1	0	1	1
0	0	1	0	1	1	0	1
0	0	1	0	1	1	1	0
0	0	1	1	0	0	0	0
0	0	1	1	0	0	1	1
0	0	1	1	0	1	0	1
0	0	1	1	0	1	1	1
0	0	1	1	1	0	0	1
0	0	1	1	1	0	1	0
0	0	1	1	1	1	0	1
0	0	1	1	1	1	1	0
0	1	0	0	0	0	0	1
0	1	0	0	0	0	1	0
0	1	0	0	0	1	0	1
0	1	0	0	0	1	1	1
0	1	0	0	1	0	0	0
0	1	0	0	1	0	1	1
0	1	0	0	1	1	0	1
0	1	0	0	1	1	1	0
0	1	0	1	0	0	0	0
0	1	0	1	0	0	1	1
0	1	0	1	0	1	0	0

IN							***R***
1	**2**	**3**	**4**	**5**	**6**	**7**	
0	1	0	1	0	1	1	1
0	1	0	1	1	0	0	1
0	1	0	1	1	0	1	1
0	1	0	1	1	1	0	1
0	1	0	1	1	1	1	1
0	1	1	0	0	0	0	0
0	1	1	0	0	0	1	1
0	1	1	0	0	1	0	0
0	1	1	0	0	1	1	1
0	1	1	0	1	0	0	1
0	1	1	0	1	0	1	1
0	1	1	0	1	1	0	1
0	1	1	0	1	1	1	0
0	1	1	1	0	0	0	1
0	1	1	1	0	0	1	0
0	1	1	1	0	1	0	0
0	1	1	1	0	1	1	1
0	1	1	1	1	0	0	1
0	1	1	1	1	0	1	0
0	1	1	1	1	1	0	1
0	1	1	1	1	1	1	0
1	0	0	0	0	0	0	1
1	0	0	0	0	0	1	0
1	0	0	0	0	1	0	0
1	0	0	0	0	1	1	0
1	0	0	0	1	0	0	0
1	0	0	0	1	0	1	0
1	0	0	0	1	1	0	1
1	0	0	0	1	1	1	1
1	0	0	1	0	0	0	0
1	0	0	1	0	0	1	0
1	0	0	1	0	1	0	0
1	0	0	1	0	1	1	1
1	0	0	1	1	0	0	1
1	0	0	1	1	0	1	0
1	0	0	1	1	1	0	1
1	0	0	1	1	1	1	0
1	0	1	0	0	0	0	0
1	0	1	0	0	0	1	1
1	0	1	0	0	1	0	0
1	0	1	0	0	1	1	1
1	0	1	0	1	0	0	1
1	0	1	0	1	0	1	1

IN							***R***
1	**2**	**3**	**4**	**5**	**6**	**7**	
1	0	1	0	1	1	0	1
1	0	1	0	1	1	1	1
1	0	1	1	0	0	0	1
1	0	1	1	0	0	1	0
1	0	1	1	0	1	0	1
1	0	1	1	0	1	1	1
1	0	1	1	1	0	0	0
1	0	1	1	1	0	1	0
1	0	1	1	1	1	0	1
1	0	1	1	1	1	1	0
1	1	0	0	0	0	0	0
1	1	0	0	0	0	1	1
1	1	0	0	0	1	0	1
1	1	0	0	0	1	1	0
1	1	0	0	1	0	0	1
1	1	0	0	1	0	1	0
1	1	0	0	1	1	0	0
1	1	0	0	1	1	1	1
1	1	0	1	0	0	0	0
1	1	0	1	0	0	1	0
1	1	0	1	0	1	0	0
1	1	0	1	0	1	1	1
1	1	0	1	1	0	0	1
1	1	0	1	1	0	1	0
1	1	0	1	1	1	0	1
1	1	0	1	1	1	1	0
1	1	1	0	0	0	0	0
1	1	1	0	0	0	1	0
1	1	1	0	0	1	0	0
1	1	1	0	0	1	1	0
1	1	1	0	1	0	0	1
1	1	1	0	1	0	1	0
1	1	1	0	1	1	0	1
1	1	1	0	1	1	1	0
1	1	1	1	0	0	0	1
1	1	1	1	0	0	1	1
1	1	1	1	0	1	0	1
1	1	1	1	1	1	1	1
1	1	1	1	1	0	0	1
1	1	1	1	1	0	1	0
1	1	1	1	1	1	0	1
1	1	1	1	1	1	1	1

To compare the number of n iterations to learn the patterns defined in **Table 7.1**, the learning rate is standardised at $\eta_{train} = 0.1$ and mean squared error MSE = Σ (actual R – measured R)^2 measured at each iteration. **Fig. 7.3a** shows the MSE after n iterations. The plot indicates that n to convergence is respectively about 10^1, 10^4 and 10^2 for associative learning, DNN and ENN. For more complex patterns (for example, in pattern c case), n to convergence for ENN increases significantly.

Next, for a more holistic comparison, the simulations are extended to the cases when the DNN has one to three hidden layers of nodes, while also varying the learning rate from η_{train} = 0.1 to 1. The number of first, second and third hidden layer DNN nodes are arbitrarily set as the round down of 1.5, 2 and 1.5 times the number of inputs. For instance, for the case of seven inputs, there are 3, 6 and 10 nodes respectively for the first, second and third hidden layers. The simulation results in **Fig. 7.3b** confirm the orders of magnitude fewer n to convergence of associative learning network (~ 10^1 iterations) relative to the other conventional architectures and algorithms (~ 10^2 to 10^4 iterations for DNN and ENN).

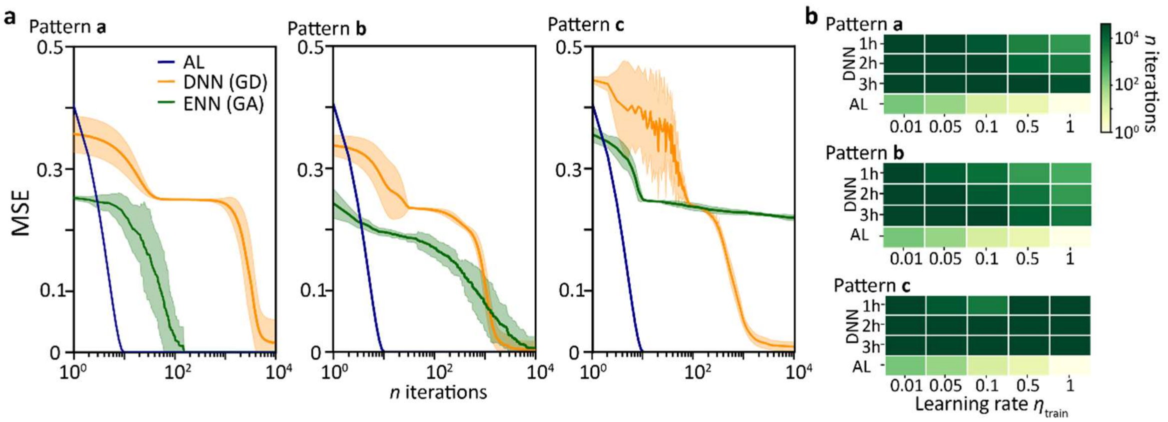

Fig. 7.3. Number of iterations comparison. a, Plots of mean squared error (MSE) against n number of iterations to convergence (to successfully learn pattern in **Table 7.1**), using associative learning network, DNN (gradient descent (GD) algorithm) and ENN (genetic algorithm (GA)). Learning rate set at η_{train} = 0.1. Mean values and standard deviations (for 20 cycles) represented as solid curves and error bands. **b,** Plots of n iterations to convergence for associative learning network and DNN with varying number of hidden layers (from 1 to 3, labelled 1h - 3h) and learning rates (from 0.01 to 1). Single-shot learning using associative learning network when η_{train} = 1, regardless of number of inputs.

As with most (if not all) other symbolic approaches [63,293], the associative learning network described here may be advantageous for problems that require quick detection, and that do not require complex feature extractions which may not be needed for many less complex machine learning tasks. A network that scales linearly with number of inputs (without mapping them to their combinations) can readily be applied for practical applications such as pattern and image recognition (details in **Chapter 6**), but such direct approach does not address nonlinear classification problems. While self-organising mappings [65,436] can be used to identify and remove redundant weights, the scaling extent varies depending on the training data (which can still scale nonlinearly). The nonlinear scaling for nonlinear classification problems may be circumvented by introducing numerical methods to the network. This would allow us approximate the weights within a limited number of nodes, and/or by adding more layers to the network.

7.1.4 Extracting high-level features with associative learning

DNN using backpropagation method has shown outstanding feature extraction performances [444]. Having deep layers (**Fig. 7.4a**) enables the network to progressively extract higher-level features. Such complex features is not attainable using the associative learning scaled-up model applied in **Chapter 6**. In the spirit of extracting complex features, the direct random target projection (DRTP) [445] is placed in the context of associative learning. The method is simplified from DNN based on the fact that it is not required to use exactly the same parameters in **(7.1)** to update the weights during the training process, as information is distributed across the network. For example, convergence can be likewise reached by projecting a separate set of modified weights $T_{k', k'+1}$ in the same direction as the actual weights $w_{k', k'+1}$ [445]:

$$0 < \sum_{k'} T_{k',k'+1} \, w_{k',k'+1} \tag{7.8}$$

Consequently, **(7.3)** simplifies to

$$\frac{\partial E}{\partial a_{k[j]}} = f'(a_{k[j]}) \left(\sum_{k'=k[j'+1]}^{K[j'+1]} \cdots \sum_{k'=k[j']}^{K[j']} \prod_{j'=j+1}^{J} f'(a_{k'}) \, T_{k',k'+1} \right) \sum_{k'=k[J]}^{K[J]} \left(f(a_{k'}) - t_{k'} \right) \tag{7.9}$$

where we have substituted $w_{k', k'+1}$ with the *Teacher* signals $T_{k', k'+1}$. Its significance can be appreciated by noting that the exact $w_{k', k'+1}$ values need not be explicitly measured at each iteration. Instead, a constant predetermined set of $T_{k', k'+1}$ values can be used for the update: this is depicted in **Fig. 7.4b**.

Interestingly, the condition in **(7.8)** need not be rigidly followed afterall [445]. Any $T_{k', k'+1}$. having random values can be used, as there are other contributing terms in **(7.9)** that can drive the error *E* to global minimum (notably, while retaining higher-level feature extraction capabilities [446] and order of iterations to convergence [445,447]). All in all, although the weight terms have been simplified, the node terms still have to be calculated in each iteration.

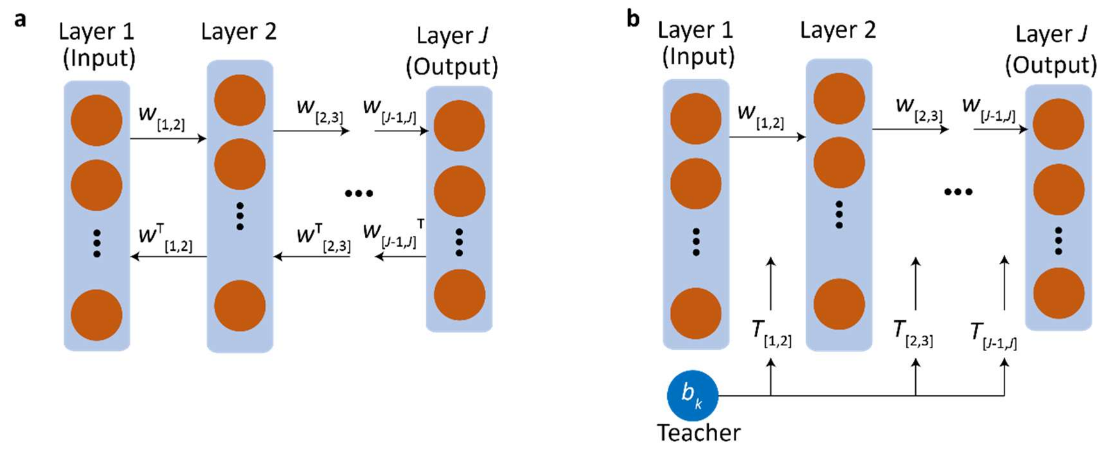

Fig. 7.4. Applying associative learning in DNN. **a**, Simplified depiction of DNN architecture (compare **Fig. 7.1a**). At each iteration, weights $w_{[j,j+1]}$ have to be calculated (in forward direction) before they are refined during training (in backward direction). **b**, Modified DNN architecture (modified DRTP) to simplify (**a**) [445], after applying associative learning. An *a priori* determined static set of *Teacher* signals $T_{[j,j+1]}$ are instead sent during training without having to calculate the weights $w_{[j,j+1]}$.

One straightforward simplification to the DRTP introduced here by the author is to use discrete rather than continuous randomised values for the *Teacher* signals. The efficacy of such simplification should come to no surprise, given the frequent use of 'regularisation' in conventional networks [448]. Because the method to recognising the patterns is numeric in nature, the initial conditions of the weights (and *Teacher* signals) can tolerate deviations from ideal conditions to some extent. To test the simplified approach, the 'modified DRTP' is tasked at recognising the expected output Out_{ex} = [1,1,0,0,0,0,0,0,0,1,0,0,1,0,1,1] for a four-input case, using one hidden layer of nodes, with the number of nodes arbitrarily set as the round down of 1.5 times the number of inputs (which is 6), and using logistic sigmoid nonlinear function. The network weights are randomised arbitrarily using the random generator function seed (19) from Numpy [435] with high and low values of -0.5 and 0.5. The *Teacher* signals take the binarised initial weights values, such that if the value exceeds 0, it is assigned 0.5, else -0.5. The convergence results of DNN and modified DRTP shown in **Fig. 7.5, a** and **b** are compared, after varying the learning rates from 0.005 to 0.1. The plots indicate that although the convergence is higher for the modified DRTP case, the order of iterations are similar with the DNN case. Further improvements to the simplification is anticipated. On a hardware level, the AMLE hardware depicted in **Fig. 6.2** can be readily used to implement the network in **Fig. 7.4b** by outsourcing the loss calculations electronically via ADC and DAC (as described in **Fig. 2.6**).

Next, the network feature extraction is studied using 9 hidden layer nodes. Features extracted from the modified DRTP network, summarised in **Fig. 7.5, c-e**, indicate complex representations where there is always subtle exceptions to perceived generalisations. It can be prohibitively difficult to acquire such complex representations via symbolic AI means (for example, that in **Chapter 6**).

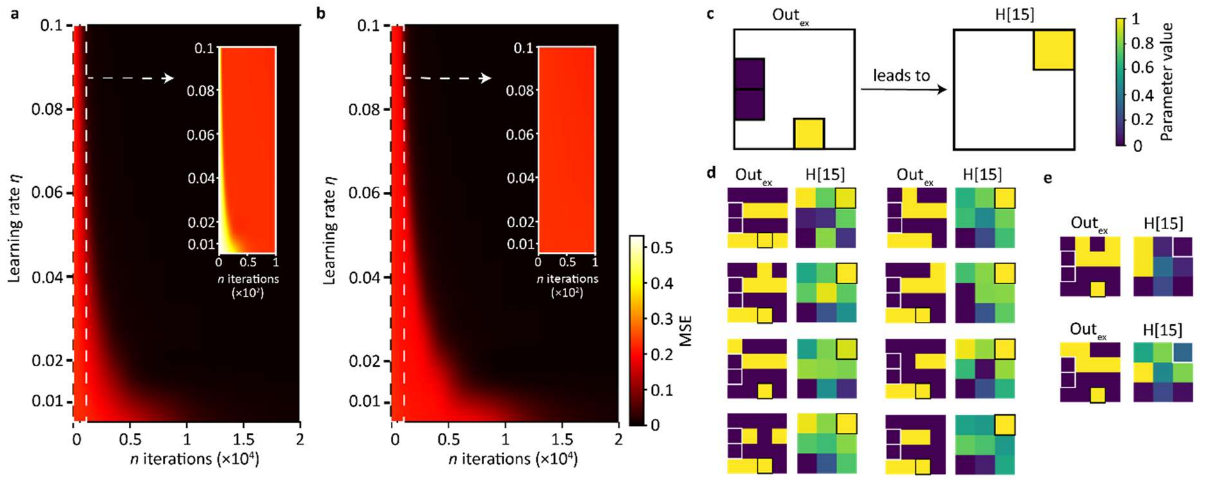

Fig. 7.5. Modified DRTP: Convergence and feature extraction. a-b, Convergence comparison between DNN (**a**) and modified DRTP (**b**). The two approaches are within the same order of convergence. **c-e**, Feature extraction from hidden layer nodes. Although the expected output Out_{in} = [x,x,x,x,0,x,x,x,0,x,x,x,x,x,1,x] generally leads to the shown feature in third hidden layer node when *Input* = [1,1,1,1] (details shown in (**d**)), there are subtle exceptions to such generalisations (as shown in (**e**)). It can be prohibitively complex to take into account all details in the generalisation using if-else conditions. Such complex features can be readily extracted using numeric AI approaches (such as DNN and modified DRTP – as shown here). In the plots, *H*[15] represents the hidden layer nodes with such *Input* combinations, while x denote 'don't-care term'.

7.2 Other associative learning devices

Besides Pavlovian associative learning (which underpins AMLE), there are other related 'associative learning' concepts. Implementations of such concepts include Y-combiner with memory cell, mutual-two way associative learning and photonic content addressable memory which are summarised hereafter.

7.2.1 Y-combiner with memory cell

Although signals can be correlated using Y-combiner with memory cell, the criteria that forms the basis of the devices are different. To compare, we shall consider two designs: 'Y-combiner 1' (with GST cell on bottom input waveguide) and ‘Y-combiner 2’ (with GST cell on output waveguide). We recall that the ‘Pavlovian associative learning device’ needs to fulfil the following criteria:

- Before learning, the two inputs lead to different outputs
- After learning, the two inputs lead to same output (one of which remains the same).
- The learning process only occurs when two inputs are sent to device.

On Criteria 3, the accumulated field when both s_1 and s_2 are sent to the device has to be significantly larger than the case when only s_1 or s_2 is sent, in order to ‘associate the two inputs’.

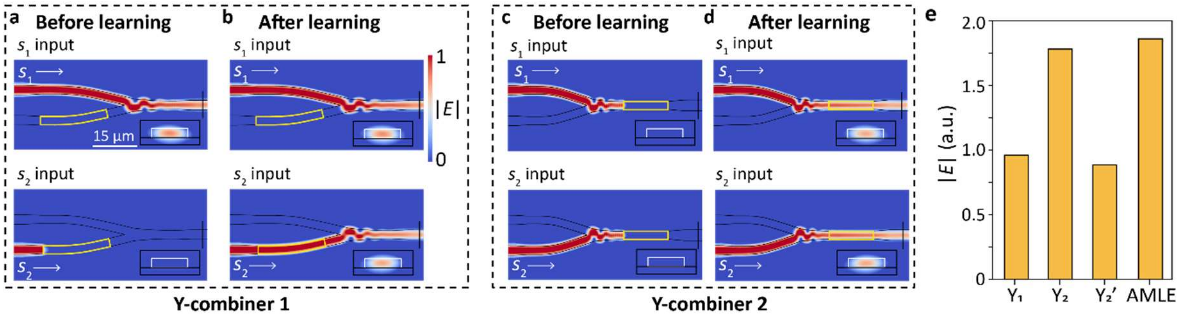

Fig. 7.6. Comparison between AMLE and GST-deposited Y-combiners. a-d, Optical field $|E|$ profile of Y-combiner 1 (**a**,**b**) and Y-combiner 2 (**c,d**) at 1.58 µm wavelength when GST cell (yellow outline) are at crystalline (**a** and **c**; before learning) and amorphous (**b** and **d**; after learning) structural state. Inset: Output cross-section $|E|$ profiles at locations denoted by black vertical bars. Colour bar is normalised. **e**, Bar plot of normalised $|E|$ profile at site prior to GST cell for Y-combiner 1 (both inputs at 1.58 µm wavelength, labelled Y_1), Y-combiner 2 (both inputs at 1.58 µm wavelength, labelled Y_2), Y-combiner 2 (top input at 1.55 µm wavelength and bottom input at 1.58 µm, labelled Y_2’), and AMLE (both inputs at 1.58 µm wavelength). $|E|$ normalised to amplitude of single input.

FDTD simulations are carried out for the comparison. The length and thickness of GST cell and ITO capping layer is 15 µm and 10 nm each. The tapered width of the GST-deposited Y-combiners is 0.8 µm, while the width of input and output waveguides to the Y-combiner is 1.3 µm. The thickness of the waveguide structure is 0.33 µm. These dimensions are consistent with that in AMLE. To test Criteria 1 and 2, input s_1 or s_2 is sent to the Y-combiners. From the simulation results in **Fig. 7.6, a-d**, only Y-combiner 1 satisfies both the criteria. To test Criteria 3, further simulations are performed to measure the field intensity at the site prior to GST cell when both s_1 and s_2 are sent to the devices. The results, summarised in **Fig. 7.6e**, indicate that Y-combiner 2 and AMLE satisfies the criteria. Summarising the overall results, Y-combiner 1 can satisfy Criteria 1 and 2 (but not Criteria 3), while Y-combiner 2 can satisfy only Criteria 3 (but not Criteria 1 and 2). The AMLE fits all the criteria of Pavlovian associative learning.

7.2.2 Mutual two-way associative learning ($s_1 \rightleftarrows s_2$)

In Pavlovian associative learning, after the learning process, input s_2 comes to suggest the same output as that from input s_1. This reflects one-way learning $s_2 \rightarrow s_1$. A distinct two-way associative learning can be realised: inputs s_1 and s_2 come to suggest the same output as those from s_2 and s_1 respectively.

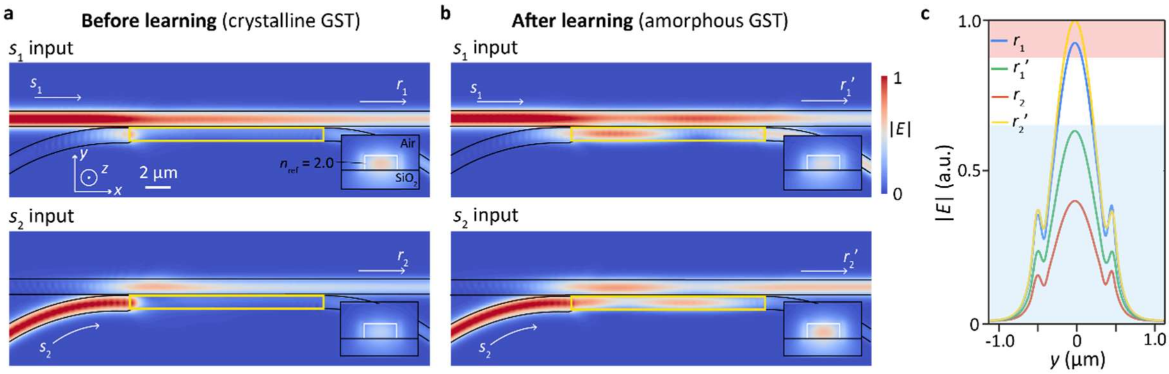

Fig. 7.7. Mutual/two-way associative learning $s_1 \rightleftarrows s_2$. a-b, Profile of simulated optical field $|E|$ in device before (**a**, crystalline GST) and after (**b**, amorphous GST) learning. Resulting outputs r_1' (r_2') is comparable with r_2 (r_1). Inset: Cross section of output field at locations denoted by black vertical bars. Colour bar is normalised to enhance output visual contrast. **c**, Corresponding output $|E|$ cross-sectional curve, showing distinct contrast between r_2 and r_1' (blue band) with r_1 and r_2' (red band). The transmission tendencies are signatures of $s_1 \rightleftarrows s_2$.

The device is designed by reducing the device refractive index (n_{ref} = 2) and segment I length (l_1 = 1.5). **Fig. 7.7, a**-**b** shows input $s_1 \rightleftarrows$ input s_2 association in our 3D FDTD simulations. The corresponding plot of cross-sectional output field $|E|$ in **Fig. 7.7c** corroborates the association.

7.2.3 Photonic CAM for accelerating search operations

Content addressable memory (CAM) is a hardware conventionally used for very-high searching applications. It is traditionally used to search for data in parallel on the basis of data content rather than by specific address or location as in standard memory structures. However, its applications extend beyond that. They include correlations, parallel computations, combinatorial optimisation, and specific machine learning tasks [67,94,449]. In all cases, the operation of a CAM is based on the comparison of symbols or bits, which are referred to generally as content and search symbols/bits.

To retrieve the address location of the content stored in CAM, it compares a search bits with the preloaded contents of the CAM blocks, and generate a given output signal depending on how many matching bits are there between the search bits and preloaded contents. So far, commercial CAMs are electronic and/or magnetic. Although optical CAMs have been realised optically, they are implemented on free space [283,450], opto-electronic [451]or nonlinear platforms [452].

Towards the end of the author's DPhil studies, the author conceived an all-optical on-chip CAM architecture shown in **Fig. 7.8a**. Each content word is wavelength-multiplexed such that when they are sent to each CAM row, the content word in each row is selectively assigned to the associative element (purple boxes) in each column, using resonators to selectively filter the content depending on the respective wavelengths of the content. The search data is likewise wavelength multiplexed such that each search bit, assigned to the different CAM columns, are sent equally to the CAM rows.

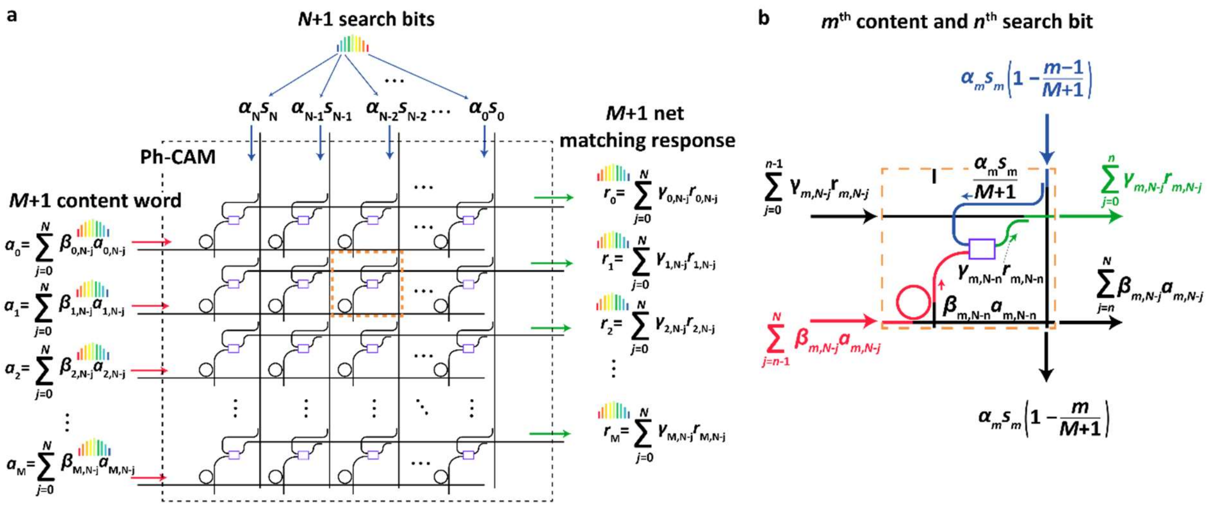

Fig. 7.8. Photonic CAM architecture. a, Schematic of architecture. Search data in each *n* column is divided equally (by *M*+1 rows) to the respective matrix cell in the column using a set of directional couplers. Content in each *m* row is sorted and distributed (based on their designated wavelengths) to the respective matrix cells using a set of optical resonators. Information in the content and search data can be masked when queried to infer associations, i.e., to search for content which have some part (but not all parts) in common with one another. Output from each matrix cell is collectively coupled via set of directional couplers to a shared output line. This combines the signal with those from other matrix cells for the respective row, to form a resulting matching response r_m. **b**, Schematic of CAM array unit that corresponds to that in **(a)**. Input signals from content (red arrow) and search data (blue arrow) are filtered in red and blue boxes respectively before being sent to matrix cell (purple). Remaining signals channeled to other units (in the next row or column; via black outgoing arrows) for the same process. Output signal from associative element combined with that of preceding

column (black incoming arrow) in green box. The resulting output matching response (green arrow) combined with output signals from other units.

The elements in the CAM matrix are the associative element, which has memory cells. When implemented using phase change material as the memory cell, the memory state is set at crystalline state before the searching process. Then, signal(s) from content and/or search bits are respectively sent. The memory switches to amorphous state when the accumulated signals are sufficiently high, to result in a change in optical transmission response of the individual CAM elements (**Fig. 7.8b**). The net matching response is the sum of the changes from the elements the respective rows. Depending on applications, here are at least three different methods to run the hardware:

- Send both the content and search bits simultaneously to the CAM. To take into account matching 0 bits, content and search bits are temporally toggled during the searching operation.
- Represent each content and search bits in two matrix rows and columns respectively. Here, the second rows and second columns of each content and search bits are respectively the toggle of the first. The content and search bits are simultaneously sent during the searching operation. This enables one-shot searching capability, without having to temporally toggle the bits.
- Only the content is first sent to the CAM such that the content is programmed into the CAM elements. Here, the signal from the content has to be sufficiently high to switch the phase change memory on the elements. The searching speed is not limited by the memory switching speed.

Other 'associative elements' can be used besides the AMLE, for example a simple 2×1 MMI with memory cell. The photonic CAM layout using 2×1 MMI fabricated on silicon-on-insulator (SOI) platform is shown in **Fig. 7.9**. The fabrication steps - summarised in **Appendix B** - are comparable to that is provided in **Section 3.3** in **Chapter 3**.

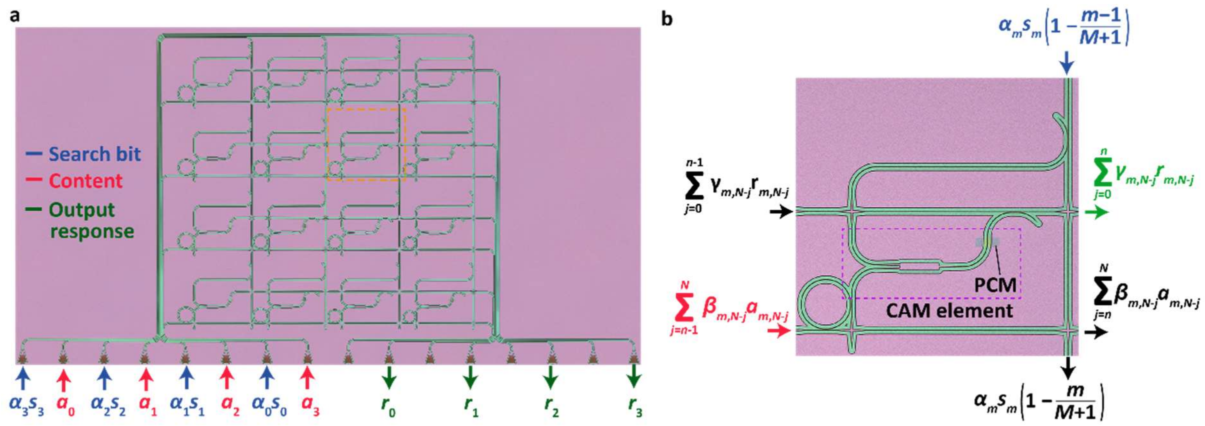

Fig. 7.9. CAM design layout. **a,** Photonic CAM with 4 rows and 4 columns. The layout corresponds to the architecture in **Fig. 7.7a**. **b,** CAM array unit based on schematic in **Fig. 7.8b**.

The design rule of thumb of the CAM components are as follows:

a. **Waveguide crossing**. The local half angle of the tapers in waveguide crossing is chosen such that the diffraction spreading of the fundamental mode is sufficiently large to form the first self-image at the centre of the crossing's MMI structure (*or* centre of MMI crossing, where another intersecting structure is perpendicularly placed), and such that the second self-image forms at the output of the crossing. This typically results in an MMI crossing width that supports the first five MMI modes [453]. Field profile of finalised structure is shown in **Fig. 7.10a**. Due to mirror symmetry, when two separate signals (regardless of wavelengths) are simultaneously sent to the two intersecting waveguides respectively, the signals would propagate through the intersecting junction seemingly unperturbed (in contrast to that in directional couplers DiCs).
b. **Combiner**. There are at least three possible options to realise the combiner:
    i. **Y-combiner**. Two separate waveguides are seamlessly merged to form a single output. Conventionally, when a single input or two non-interfering inputs are sent to the Y-combiner, the signal transmission through the Y-combiner is ~ 50%, as indicated in **Fig. 7.6, c and d**. Less seamless taper design – such as that in **Fig. 7.10b** –supports radiative modes at the output waveguide walls. This leads to increasing signal transmission loss as output waveguide length increases.
    ii. **2×1 MMI**. The structure is the spatial inverse of 1×2 MMI splitter (design approach in **Chapter 6**). Similar to Y-combiner, when a single input or two non-interfering inputs are sent to the MMI, the signal transmission through the structure is ~50%, as indicated in **Fig. 7.10c**.
    iii. **Cascaded 2×2 - 2×1 MMI**. The structure capitalises on the abilities of 2×2 MMI to split equally when a single input or two non-interfering inputs are sent (design approach in **Chapter 3**), and of 2×1 MMI to combine with ~100% transmission when two inputs of the same optical phase difference are sent (design approach in **Chapter 3**). The optical phase of the q^{th} output of the 2×2 MMI is $(2 - q)q\pi/2$ [332,454], which are 0 and $\pi/2$ respectively. Provided that the input optical phases to the 2×1 MMI is compensated such that they are both equal, the signal transmission through the cascaded 2×2 - 2×1 MMI is ~ 100% when a single input or two non-interfering inputs are sent to the overall structure, as depicted in **Fig. 7.10d.**
c. **Optical wavelength filter**. It is convenient to use ring resonator as wavelength filter. When coupled to two separate perpendicular waveguides, the resonator enables signals of certain wavelengths to be routed to a perpendicular direction. Signal coupling between the resonator and waveguides can be better understood by modifying from **(3.17)** and **(3.18)** in **Chapter 3** - using coupled mode equations to describe the field coupling of resonator mode a to optical input a_{in} and output a_{out} waveguide modes:

$$\frac{da}{dt} = -\left(i\omega_0 + \frac{\kappa_0 + \kappa_{ex1} + \kappa_{ex2}}{2}\right)a + \sqrt{\kappa_{ex1}}\,a_{in} \tag{7.10}$$

$$a_{out} = 1 - \sqrt{\kappa_{ex2}}\,a \tag{7.11}$$

ω_0 is the angular frequency when the wavelength is at resonance, κ_0 and (κ_{ex1}, κ_{ex2}) are the optical losses due to the resonator, and the (input, output) waveguide-resonator coupling of respectively [381,455]. Following the derivation steps described prior to **(3.19)** in **Chapter 3**, the transmission at waveguide output port $T = |a_{out}|^2$ can be written as

$$T = \left| 1 - \frac{\sqrt{\kappa_{ex1}\kappa_{ex2}}}{\frac{\kappa_0 + \kappa_{ex1} + \kappa_{ex2}}{2} - i\left(\omega - \omega_0\right)} \right|^2 \tag{7.12}$$

It is straightforward to show from **(7.12)** that waveguide-resonator critical coupling can be attained when $\kappa_o = \kappa_{ex1} = \kappa_{ex2}$ and $\omega = \omega_o$.

d. **Optical splitter with variable splitting ratio**. It is convenient to use directional couplers (DiCs) to split optical signals at a certain splitting ratio. The input-output relations of a DiC is encapsulated in **(3.12a)** in **Chapter 3**; from which the transmission of coupled signal (as function of coupling length *l*) is

$$T = sin^2\left(\kappa l\right) \tag{7.12}$$

where κ is the coupling coefficient of the coupled waveguides [335]. In the CAM layout (**Fig. 7.8** and **Fig. 7.9**), DiCs are used to:

i. Distribute search bit signals to the CAM elements in each row equally. This is achieved by setting the m^{th} row DiC coupled signal transmission to be $T = (m+1)/(M+1) \times 100\%$, where M is the total number of rows in the CAM.

ii. Accumulate the constituent output responses from the CAM elements in each column. This is achieved by setting the n^{th} column DiC coupled signal transmission to be $T = (N - n + 1)/N \times 100\%$. The use of DiC in the architecture leads to signal loss due to the symmetric coupling of the directional coupler. This can be mitigated by replacing the DiCs with an asymmetric couplers - structures that channels the output response signal, while also allowing output response signals from the columns before to transmit through; for example that in [456].

For $N = 4$ and/or $M = 4$, the relevant T are 100%, 75%, 50% and 25% (intensity profile $|E|^2$ in **Fig. 7.10f**).

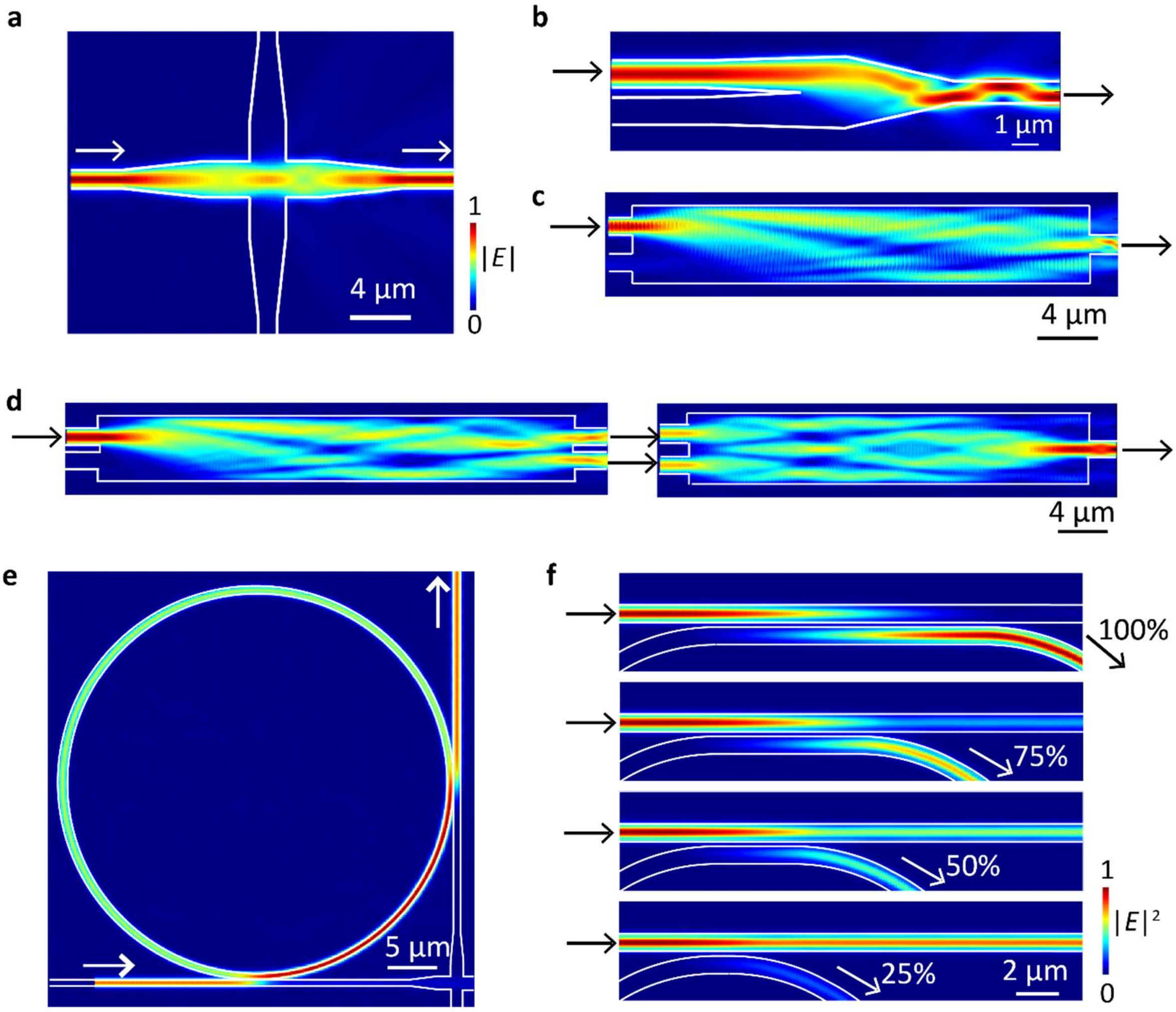

Fig. 7.10. CAM components on SOI platform. a, Optical field $|E|$ profile of waveguide crossing. **b-d,** $|E|$ profile of combiners. Conventional 2x1 MMI combiner (based on 1x2 MMI design described in **Fig. 6.2** in **Chapter 2**, for example in **c**) has transmission ~ 50%. Other structures with seemingly 100% transmission (such as that in (**b**)) excites radiative modes at the output waveguide walls: this causes signal loss. To circumvent such loss, cascade of 2x2 and 2x1 MMI can be used (**d**). **e**, $|E|$ profile of ring resonator coupled to two waveguides. **f**, Optical field intensity $|E|^2$ profile of optical splitter (directional coupler). Signal energy transfer is a function of coupling length.

One common application of CAM is to implement the Hopfield network [67] to solve constraint problems, for example the typical benchmark constraint problem known as the 'traveling salesman problem'. The problem seeks to find 'the shortest possible route that visits each city exactly once and returns to the original city, given a list of *N* cities' [457]. In **Fig. 7.11, a** and **b**, a comparison between all-parallel electronic Hopfield network and the equivalent photonic Hopfield network is presented. It is apparent from the figure that the photonic approach greatly simplifies the network, due to the availability of WDM to transmit data in parallel without interaction. The photonic Hopfield network is simulated to solve the traveling salesman problem for the shown coordinates in **Fig. 7.11c**, based on the steps provided in [67]. The network weights are optimised using the numpy.random.seed 0 to 3 [435]. The input initial value and step sizes are set as 0.0006, and 0.0004 respectively. After 10^3 iterations, the solution converges (**Fig. 7.11d**): all four trials indicate that the shortest route is C → D → A → B. The correct solution can be readily ascertained by testing each possibilities by brute force.

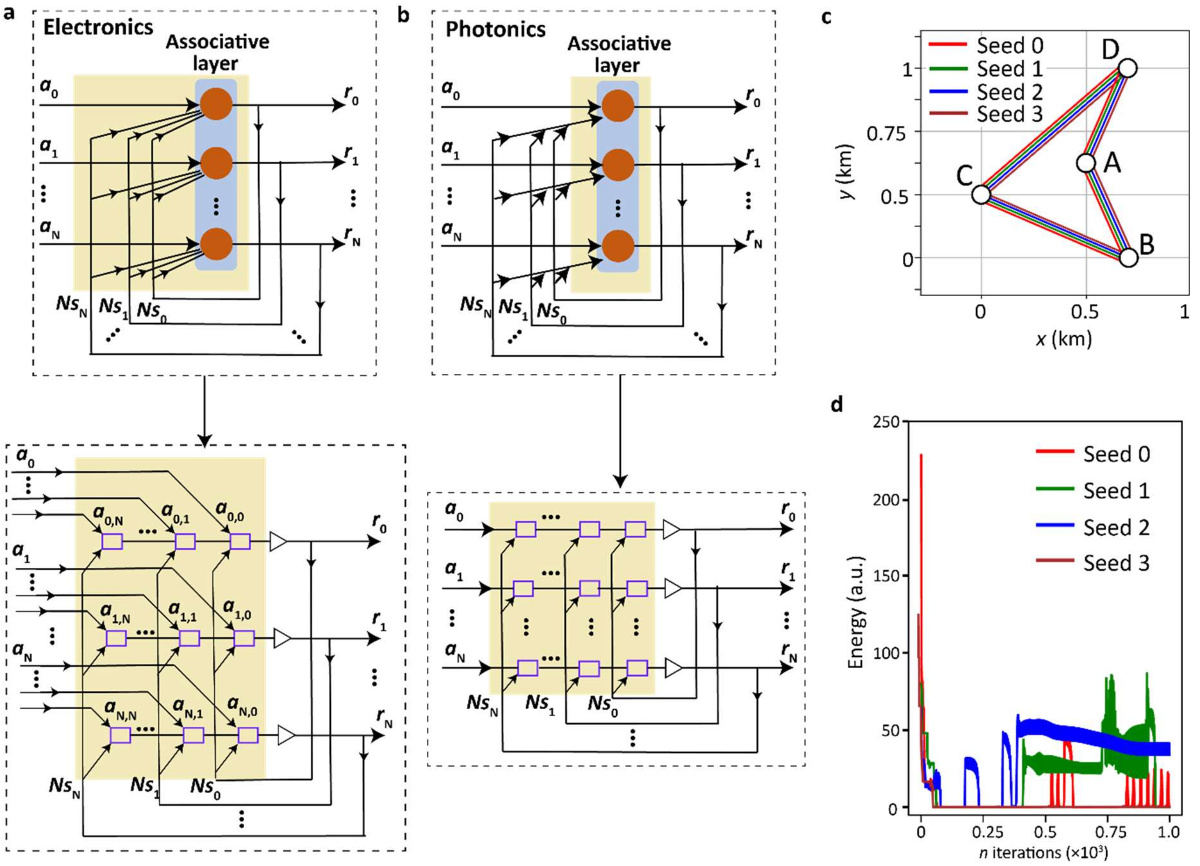

Fig. 7.11. Photonic CAM for Hopfield network: comparison and applications. a,b, Comparison between all-parallel electronic and photonic Hopfield network. All-parallel electronic network can achieve full parallelisation, but at the expense of complex wiring that scales ~ exponentially with the number of inputs. Photonic network can implement parallel processing using WDM. **c**, The photonic Hopfield network is simulated to solve the traveling salesman problem for the shown coordinates (A, B, C, and D). The shortest possible route is shown (C → D → A → B). **d**, Energy (convergence) plot. The route converges as the energy is low (within ~ 50 a.u.).

As depicted in **Fig. 7.12a**, the connections in each row of an electronic CAM is similar to those in perceptron (compare with **Fig. 1.2** in **Chapter 1**). With WDM, instead of using distinct channels, weighted inputs can share the same channel (details in **Fig. 7.12b**). The photonic CAM thus provide a cost-effective, convenient and highly-scalable alternative to the electronic CAM.

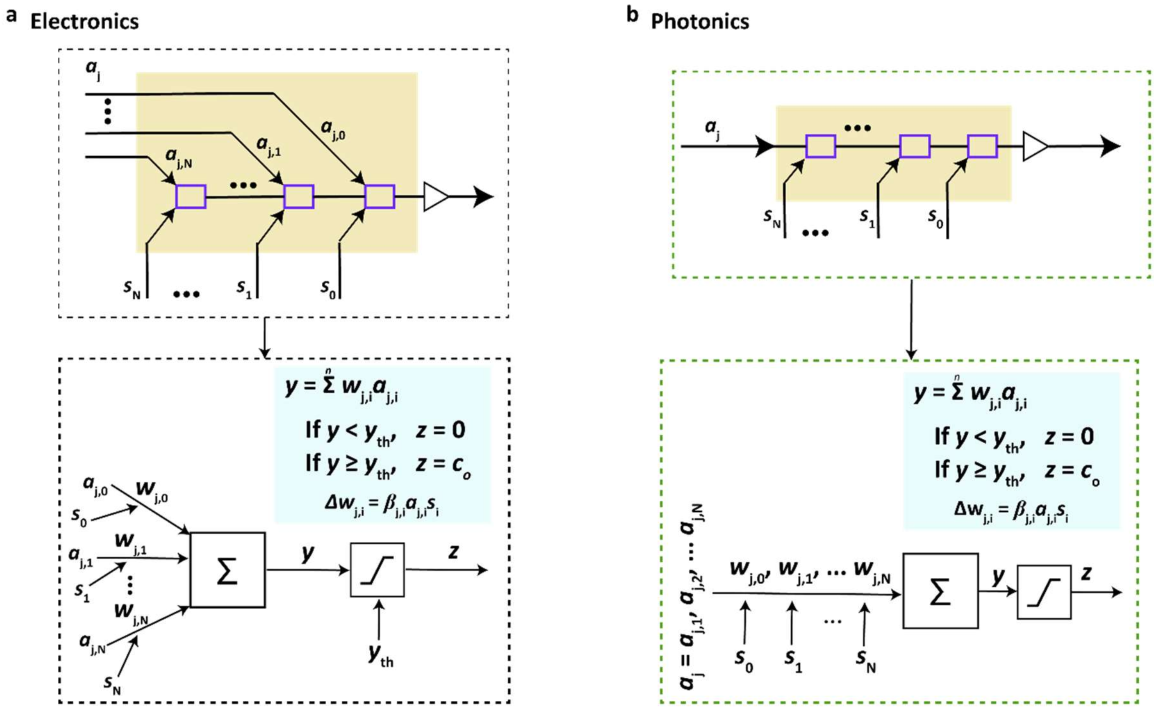

Fig. 7.12. Perceptron-like simplification in CAM. a, The row elements in CAM can be likened to a perceptron. The weights are modified when both 'search' and 'content' data is respectively sent. **b**, The weighted sums can be parallelised on a single channel when implemented photonically.

We have discussed the limitations of the architecture proposed and adopted in Chapter 6. Although the architecture is capable of association detection, inductive learning, and deductive reasoning (as summarised in Fig. 1.5 in Chapter 1), the features extracted from the presented patterns and images are apparent in contrast to conventional DNN that can handle complex feature extractions. Attempts to apply associative learning for such complex extractions – by modifying from DRTP – show that the same order of iterations to convergence as DNN can be reached. Other devices in the broader family of 'associative learning' devices such as Y-combiner based device, mutual two-way associative learning and photonic CAM are also discussed. In particular, a highly-scalable photonic CAM is designed by the author to be used for solving the practical benchmark constraint problem (travelling salesman problem).

Chapter 8

8. Conclusion

This chapter concludes the thesis by summarising key results in this thesis, determines the current limitations of the device and system, and projects future directions and implementations to overcome them.

The work described in this thesis mainly investigates photonic Pavlovian associative learning elements, and a scaled-up implementation of pattern and image recognition using these elements. A learning framework that facilitates the transition from a monadic Pavlovian *Input-Teacher* association to any arbitrary *n Input-Teacher* associations is provided. The inner workings of the hardware, that spatiotemporally correlates two initially distinct inputs (s_1 and s_2) to the same output is elucidated.

The learning approach is benchmarked with other conventional methods. Based on the comparison, it is determined that the architecture can be advantageous for problems that require quick detection and inferencing, and for problems that do not require complex feature extractions. For such higher-level features, a modification of DNN that applies associative learning is examined. The results obtained in this work demonstrate the practical use of the element for learning applications, and offer new insights for future developments in the nascent yet rapidly expanding field of hardware accelerators based on input associations.

8.1 All-optical associative learning element

An outline to realise photonic associative learning is provided in **Chapters 3 and 4**. It consists of seamlessly cascading two directional couplers, with distinct functionalities for each cascade segment. The first determines the input combinations that induce the learning process, while the second regulates the output signal response. On the first segment, optical superposition *or* interference accumulates signals at the region prior to modulating element GST (to sufficiently heat up the GST for GST structural phase change) when *two*-input, and not *one*-input is sent to the element. On the second segment, the coupling extents of waveguide modes varies asymmetrically depending on ‘before or after learning’ states. The signal regulation is allocated here due to the modulating element GST. To this end, the length of coupled waveguides, and the effective loss from GST thickness are judiciously determined.

As the signals have to constructively interfere (and not destructively interfere) for the learning process, stable control of input optical phase is essential. However, it is observed that optical phases vary erratically upon the slightest vibration and/or temperature change in the measurement environment. Two layouts are devised to eliminate this: ring- and heater-based layouts. The ring-based layout enables real-time observation of associative learning. To scale up the devices (by integrating multiple elements using wavelength-multiplexed inputs), heater-based layout is required. It consists of cascaded MZMs to modulate input amplitude and optical phases to the element. It forms the basis for subsequent learning demonstrations.

8.2 All-optical associative learning for AI acceleration

The element is placed strategically in the context of AI hardware acceleration, given the comparable concepts of Pavlovian associative learning [5] and supervised learning in AI [444]. Both involve pairing of input with correct output, in one way or another. To reinforce the relation, a scaled-up integration of the elements is implemented.

In the hardware network, the cascaded MZMs (modulated using NiCr thermo-optic integrated waveguide heaters) enable optical signal control between the layers. They provide reliable means to wavelength-multiplex on a large scale. Multimode interferometers (MMIs) split the output signals from the elements to separate paths each. The first path forms the individual outputs while the second are cumulatively summed with outputs from other elements.

AI acceleration using associative learning hardware is demonstrated to solve pattern and image recognition tasks. In the pattern recognition task, the hardware recognises specific patterns. Once bit patterns to be detected are programmed to the elements, simultaneous detections can be carried out via WDM. In the image recognition task, the hardware creates predictive model of the object-to-be-detected to generalise a set of images.

The hardware is benchmarked with other conventional approaches. From simulations on solving tasks of varying sizes and network depths (in DNN), the associative learning network requires fewer iterations ($\sim 10^1$ compared to $\sim 10^2$ to $\sim 10^4$ in other approaches), as with other symbolic methods. A numeric method [445] placed in the context of associative learning is proposed to extract higher-level features in data.

8.3 Concluding remarks

Although the results demonstrate promising use of optical associative learning element as building blocks for AI hardware acceleration, more could be done. On the learning architecture and algorithm, the following improvements can be made:

- An architecture that can detect associations, generalise both apparent and complex patterns, and deductively reason based on explicit similarities. The approach discussed in **Chapter 6** lacks complex generalisation abilities.
- A platform that supports learning process guided by automatically-created and passively-retrieved labels (or in general objective). Currently, the device supports learning process that rely on actively-retrieved labels.
- An architecture that addresses the conventional tradeoff between the ability to acquire complex features and the computational efficiency [458]. In addition to the approach discussed in **Chapter 6**, numeric approach to associative learning modified from DRTP in **Chapter 7** to acquire complex features has low computational efficiency. We may perhaps address this by finding answers from biologically-plausible models [35-37] .

The prospective research streams summarised from the points above is depicted in **Fig. 8.1**. This work represents the first implementation of associative learning as a building block in learning scaled-up hardware, and could potentially open up new avenues of research in machine learning algorithms and architectures.

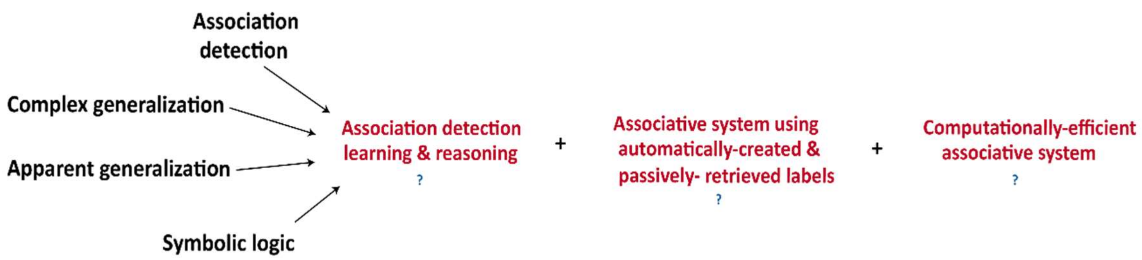

Fig. 8.1. Prospective research streams on associative learning AI architecture. Architecture proposed and adopted in **Chapter 6** enables associative detection, apparent (but not complex) generalisation, and deductive reasoning (or symbolic logic) on actively retrieved labels. It is desirable to embed the ability to infer/generalise complex features from datasets that are automatically-created or passively-retrieved labels. Attempts to link associative learning for complex generalisation in **Chapter 7** are based on simplifications from DNN, and are thus computationally-intensive. It is advantageous to have an algorithm that offers higher computational efficiency.

More can also be done on the hardware device. Currently, the associative learning devices are limited by the factors outlined in **Table 8.1**.

Table 8.1. Limitations of AMLE device.

Parameters	**Pumping** (switching or training)	**Probing** (testing, detection or inferencing)
1. Speed	Limited by **EOM modulation speed** (currently ~ 10 to 100 GHz) and **GST switching speed** (~ 10 to 100 MHz)	Limited by **signal-to-noise ratio** of probing signal (to resolve the pulse signals). Signal power is conventionally limited to ~ 1 mW to ensure **stability of GST structural state** [459] - which limits the probing speed at ~ 50 GHz
2. Energy consumption	~ several tens of pJ to nJ to **switch GST structural state**	~ fJ per bit
3. Throughput	Not applicable	Number of wavelengths for signal multiplexing is limited by the **maximum signal power that can be supported by GST-deposited waveguide without affecting stability of GST structural state**, which is ~ 1 mW [459]. This is projected to be ~ 50 wavelengths.
4. Computational density	Not applicable	Limited by **signal-to-noise ratio** of probing signal and spatial density of optical waveguides. This is projected to be approximately ~ 10^5 TOPS/mm^2, considering 50 wavelengths modulated at 50 GHz) for device spatial area of 30 μm^2
5. Spatial density	~ 30 μm^2	

On <u>learning speed (from pump pulses)</u>, the speed is limited by the EOM modulated speed – which in turn is limited by RF signal modulation speed (~ several hundred of GHz). If the modulation from EOM is to be ruled out as the limiting factor, then other phase change materials (for example metal antimony [460]) with lower glass transition temperature over melting temperature ratio can be considered [111,152]. On the <u>probing speed (from probe pulses)</u>, the signal power is limited by the instability of the structural state of phase change material beyond ~ 1 mW. It would be necessary to study in further detail the cause for such an observation, as this limits the <u>signal throughput</u> and <u>computational density</u> as well (details in **Table 8.1**).

On <u>energy efficiency</u>, the energy consumption mainly lies in switching the phase change material. This energy could be further reduced by using shorter pulse lengths, for example ~ 15 pJ switching pulse energy at ~ 2 ps pulse length [461]. The switching energy efficiency improves with the reduction of programming pulse width, as thermal confinement can be achieved by limiting the heat diffusion time for short pulses [462].

Currently, the element relies on near-field coupling to switch the GST on a waveguide, which is not spatially concentrated and is thus energy inefficient. Moreover, such near-field coupling approach for GST switching could also lead to overall switching latency: as the GST structural phase switching relies on optical pulse heating on the GST material, the heat from the optical pulses would have to be transferred over a relatively considerable distance to the GST material for the switching to occur. Further well thought out approaches to couple light efficiently to the phase change material could mitigate these. On the spatial density of the device, both the large optical waveguide mode size and optical waveguide coupling length impede integration densities. On reducing mode size, although plasmonic waveguides have been proposed [463], its lossy nature renders it impractical. Ways to approach these may include device optimisation [464].

All in all, these limitations have to be placed in the 'co-processing' context to consider bottlenecks in the system as a whole. On the computation speed, although multiple signals at different wavelengths can be simultaneously sent to increase detection or inferencing speed, the overall speed bottleneck mainly lies in ADC/DAC conversion and CPU computations. As AI computations tend to involve multiple types of calculations, hardware with co-processing capabilities is key in the effort to truly accelerate the computations.

Other future work that could follow on the work presented in this thesis include

- Removing the optical phase-dependence of the device. The element, in its present form, relies on specific optical phase difference between inputs to induce learning, to compensate the lack of light-light interaction in linear optical medium [128,130]. However, it fluctuates with the slightest disturbance to the measurement environment. Although different layouts have been conceived to address this, they come at the expense of consuming large physical space prior to the inputs on the on-chip layout, and tighter parameter restrictions (for example, wavelength range, modulation speed, cascadability, ability for real-time measurements). Solutions may include optical phase-insensitive signal propagation.

- Solving the 'fan-in and fan-out problem' of optical interconnections. Optical interconnections - conventionally realised via MMIs and DiCs – either rely on optical phase difference between the input signals (when their optical wavelengths are the same) or standard MMI/DiC structures meant to split signals equally. For example, using a 1 × *N* MMI optical splitter for a *N* × 1 MMI optical combiner would introduce *N*-fold loss [465]. Other more specialised optical interconnections e.g., asymmetric optical splitter (such as those applied in photonic CAM – discussed in **Section 6.2.3**) likewise introduce loss as well. Unless interferometer-based structures are used for the optical interconnections (which requires electrical/RF-modulation), such optical loss can be prohibitively high when the optical circuit is significantly scaled-up.

- Identifying memory materials that addresses speed, energy efficiency and structural stability issues (highlighted earlier). Current lack of other viable 'optical memories' (discussed in **Chapter 2**) implies that these may remain relevant for a foreseeable time. Although GST phase change material is commonly used, combination of Ge , Sb [466], Te [110], Al, In [467], V [468], Nb, Ti, S [469] Mo [470], and C [471] may be considered.

- Identifying and applying other associative learning elements into new learning architectures. In **Chapter 7**, a two-way associative learning process (distinct from the AMLE, which is based on Pavlovian associative learning) is proposed. Such 'associative learning elements', applied as a building blocks in a hardware network, could accelerate certain forms of associations.

In this concluding chapter, a summary of the author's research work, and limitations of the devices implemented by the author are described. Future projections from the work and prospective research streams are laid out, which consist of two main points:

- **Firstly, a computational architecture and algorithm that addresses all constituents of a truly intelligent construct depicted in Fig. 8.1 (association detection, inductive learning - both apparent and complex generalisation - and deductive reasoning).**
- **Secondly, a more exhaustive co-processing approach to AI acceleration that is unbounded by electronic parameters (for example, electrical or RF modulation speed on the EOM) –that can be drawn upon from the limitations presented in Table 8.1.**

These shall be reserved for a future time.

Appendix A: Fabrication and characterisation equipment

The equipment used to realised the associative learning hardware is shown in **Fig. A**.

Fig. A. Fabrication and characterisation equipment. **a**, Electron beam lithography tool (JEOL JBX-5500ZC). **b**, Thermal evaporator (Edwards 306). **c**, Sputtering tool chamber (**i**) and control system (**ii**) (Nordiko). **d**, Reactive-ion etching (Plasmalab 80Plus). **e**, Mask aligner (Karl Suss MJB 3). **f**, Plasma cleaning tool (Henniker Plasma HPT-200). **g**, Profilometer (DektakXT Bruker). **h**, Spincoater. **i**, Wire bonder (K&S 4524 Manual Ball Bonder) **j**, Scanning electron microscope (Hitachi S-4300). **k**, Optical microscope (Nikon Eclipse LV100 ND).

Appendix B: Electron-beam proximity effect correction

During electron beam lithography, the electron beams to draw custom shapes on the resist experience diffuse backscattering as the beam hits on the resist and/or substrate. This causes the electron beam exposure to not be as uniform as intended – in certain cases with apparent under-exposed regions. Such underexposure is more pronounced with closely-packed regions and sharp corners; and less pronounced when the pattern is isolated. The dimension and magnitude of the backscattered electrons are described by the point spread function of the focused beam incidence.

To compensate such 'proximity effect', the Beamer software (GenISys GmbH) [472] is used. The software simulates the point spread function of the focused beam incidence to determine the optimal exposure doses to increase uniformity and contrast. Using the optimised exposure dose values (at specified discrete points in the pattern), the dose profile is specified in the electron beam control software. As an example, the proximity effect correction for the waveguide layer for the hardware shown in **Fig. 7.9** is shown in **Fig. B**.

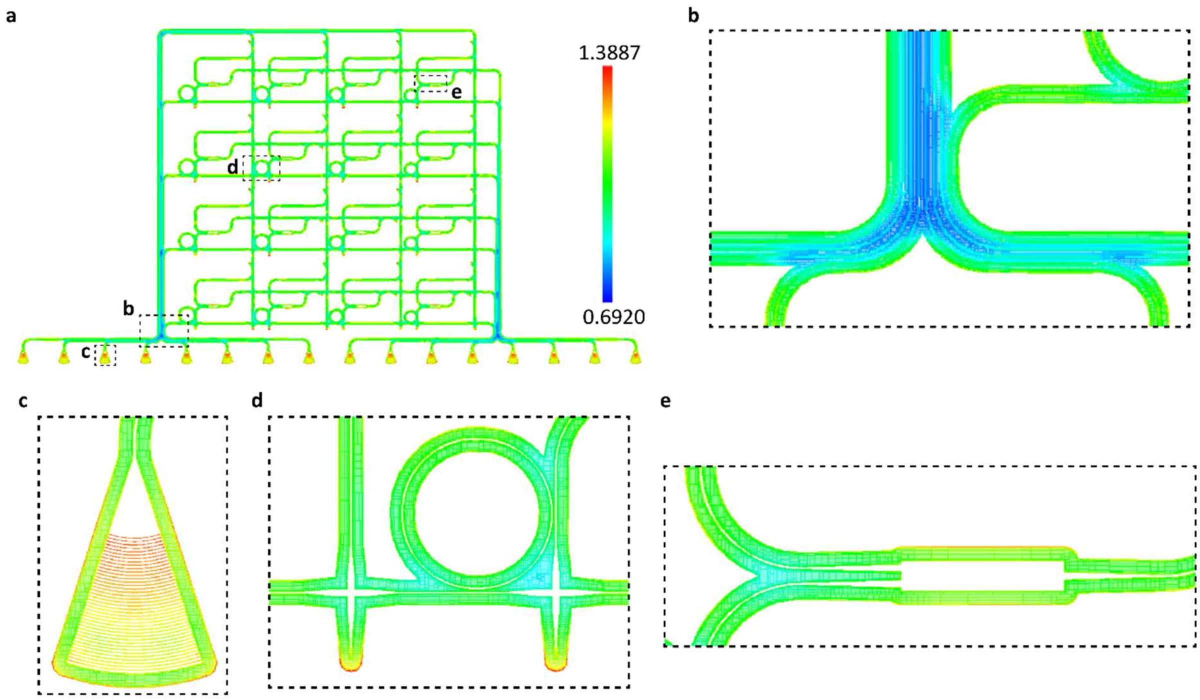

Fig. B. Proximity effect correction. a, Corrected dose profile as simulated using Beamer software. Correction is significant at closely-packed regions with sharp corners (boxed). **b-e**, Enlarged view of the boxed regions in (**a**). Colour scale indicates relative dose proportional to the degree of underexposure.

Appendix C: Steps to fabricate photonic CAM on SOI platform

The fabrication of CAM on SOI platform follows the following sequence:

a. Fabrication of Si structure. To define the Si_3N_4 waveguide, markers and frame structures, the following steps are carried out:

i. Si_3N_4 substrate (220nm Si/2µm SiO_2) is cleaned.

ii. The substrate is spincoated and pre-baked with positive resist CSAR 62 (Allresist GmbH) at 150°C for 3 minutes.

iii. The substrate is patterned using electron beam lithography (JEOL 5500FS) at 50kV.

iv. The patterned resist is developed in AR 600-546 and MIBK for 30 seconds and 15 second, respectively; and stopped in IPA for 15 seconds.

vii. Reactive ion etching (PlasmaPro 80, Oxford Instruments) is performed in $CHF_3/O_2/Ar$ to partially etch (100 nm) the exposed Si.

v. The remaining resist is removed by oxygen plasma in the RIE.

b. Deposition of GST/ITO patches. To define the GST cell with a thin ITO layer, the following steps are carried out:

i. The substrate is cleaned.

ii. The substrate is spincoated and pre-baked with positive resists 495 PMMA at 4% and 950 PMMA at 2% (MicroChem Corp.) at 150°C for 5 and 15 minutes respectively.

iii. Similar process as in a.iii. is performed to define GST/ITO patch patterns.

iv. The patterned resist is developed in IPA/MIBK:EMK (15:5:1; MicroChem Corp).

v. The substrate is sputtered with 10nm GST/10nm ITO using sputtering tool (Nordiko RF Sputter Deposition tool, Nordiko Technical Services Ltd.).

vi. The remaining resist is removed by Remover PG (MicroChem Corp.).

Appendix D: Comment on extracting 'less implicit' features

In line with the 'no free lunch theorem' [473] (placed in context in [474]), there is no one universally-best AI algorithm. The choice of learning method depends on the learning applications. The associative learning method described in **Chapter 6** is best applied for quick detection of less implicit/complex features.

It is tempting to raise the question on the practical use of extracting such explicit (or less implicit) features, when these features are afterall self-evident. To put things into perspective, it is due to the very fact that such features are apparent, that we can delineate reason from a particular observed data. The reasoning arises from symbolic manipulations (for example, element wise AND operations) that form the requisite to solving many well-defined AI problems including:

- <u>Logical inferencing</u>. AI systems that symbolically form associations enable capturing compositional and causal structure in the dataset to provide reasoning to the conclusions they arrive at. An example of practical AI capabilities that make use of such logical operations to infer new knowledge is the system 'Adam' [475]. Due to its ability to reason and interpret results, it is able to methodically hypothesise and test the hypothesis it come up with. In the same fashion, the associative learning system built by the author can explicitly interpret from the 'cat representation' results in **Fig. 6.14** in **Chapter 6** to logically conjure up the proposition that an image must have the shown clear, distinctive, and comprehensible physical attributes to constitute a cat image. This ability is absent in conventional AI architectures. The representations from conventional AI are typically abstract and uninterpretable at best.
- <u>Expert system tasks</u>. To emulate the decision-making ability of a human expert, an AI expert system has to understand the problem at hand. This is achieved by using symbolic rules and decision trees to reach a verifiable and explainable conclusion. An example expert system is the Stanford HIV Drug Resistance Database (HIVDB) GRT interpretation system (GRT-IS) which predicts virological response by discovering rules from associations between the many drug-resistance mutations [476]. The associative learning hardware applied in the experiment summarised in **Fig. 6.9** in **Chapter 6** provides the foundation for subsequent optical hardware to solve such associative mining tasks.
- <u>Constraint satisfaction</u>. It is common to apply associative mining to solve constraint satisfaction problems [477], similar to that is summarized in **Fig. 7.11** in **Chapter 7**.
- <u>Natural language processing</u>. In language processing today, conventional AI is used to predict what text comes after some given text after being trained on a large amount of uncategorised texts. However, the system fails when given a simple test (Turing Test) that interrogates its understanding

of a subject matter [478]. In contrast, systems with basic association capabilities [479] thrives at such tasks. They are capable of figuring out contextual nuances. The associative memory in the work is comparable to our associative hardware shown in **Fig. 6.5a** in **Chapter 6**.

Overall, conventional AI architectures (and algorithms) 'struggle' to solve well-defined problems, such as those listed above. These architectures are essentially black-boxes that handle NP-hard [480] optimisation problems; with no guarantee that a right solution can be reached within a reasonable (polynomial) time, but relies instead on a 'good enough' gradient descent minima to indicate that a reliable solution is possibly reached typically by running many iterations using many examples. In contrast, in solving these problems, it is straightforward to determine whether or not a solution from symbolic manipulations is reliable, simply with very few iterations and examples. Thus, so long as conventional AI architectures are not capable of bestowing reasoning to their solutions, symbolic manipulations will continue to find relevance to solving many practical AI problems. When these architectures are explicitly implemented on a purpose-built hardware – such as the associative learning hardware shown in **Fig. 6.5a** in **Chapter 6** – they accelerate practical AI problems more efficiently than is possible on a general-purpose computer, with significant increase in signal throughput due to parallelism attributed to optical wavelength division multiplexing.

It should also be noted that the term 'less complex features' does not necessarily translate to 'simple problems'. Problems involving extracting 'less complex features' can be complex, for example those processed by IBM Watson machine that won the quiz show 'Jeopardy!' in 2011 [481]. Using conventional AI architectures to solve such problems involving extracting 'less complex features' would prove to be highly inefficient, if not insurmountable.

References

1 Vazza, F. & Feletti, A. in *Nautilus Quarterly* (New York, U.S.A., 2017).

2 Krizhevsky, A. Learning Multiple Layers of Features from Tiny Images. (2009).

3 Silver, D. *et al.* Mastering the game of Go with deep neural networks and tree search. *Nature* **529**, 484–489 (2016).

4 Senior, A. W. *et al.* Improved protein structure prediction using potentials from deep learning. *Nature* **577**, 706–710 (2020).

5 Pavlov, I. P. Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex. *Ann. Neurosci.* **17**, 136–141 (1927).

6 The simplicity of sea slugs. *Lab Anim* **42**, 195 (2013).

7 Hebb, D. O. The organization of behavior *New York: Wiley & Sons* (1949).

8 Morford, M. P. O., Lenardon, R. J. & Sham, M. *Classical mythology*. (Oxford University Press, 2018).

9 Rhodios, A. & Green, P. *The Argonautika*. (Berkeley : University of California Press, 2007).

10 Petersen, L. I. R. Sultans of science: An opportunity lost. *Technology and Culture* **55**, 973-987 (2014).

11 Brett, G. The Automata in the Byzantine "Throne of Solomon". *Speculum* **29**, 477-487 (1954).

12 Riley, J. *Chinese theatre and the actor in performance*. (Cambridge University Press, 1997).

13 Moran, M. E. The da Vinci Robot. *Journal of Endourology* **20** (2006).

14 Grillo, F. *Hero of Alexandria's Automata: a critical edition and translation, including a commentary on Book One* Ph.D. thesis, University of Glasgow, (2019).

15 Price, D. J. d. S. Automata and the Origins of Mechanism and Mechanistic Philosophy. *Technology and Culture* **5**, 9-23 (1964).

16 Babbage, C. *Chapter VIII – Of the Analytical Engine*. (Longman, Green, Longman, Roberts & Green, 1864).

17 Lovelace, A. Notes to a Sketch of the Analytical Engine invented by Charles Babbage, by L.F. Menabrea. *Scientific Memoirs* **3** (1843).

18 Cajal, S. R. y. *Histology of the nervous system of man and vertebrates (Translated by Neely Swanson and Larry W. Swanson in 1995)* (Oxford University Press, 1909).

19 Sherrington. C. S. to Schafer, E.A., Letter 27th Nov 1897, in the Sharpey-Schafer papers *Contemporary Medical Archives Centre, the Wellcome Institute for the History of Medicine* (1897).

20 Lapicque, L. Quantitative investigations of electrical nerve excitation treated as polarization (translated by Brunel, M. and M. C. W. Rossum). *Biol Cybern.* **97** (1907).

21 McCulloch, W. S. & Pitts, W. A logical calculus of the ideas immanent in nervous activity. *Bull Math Biophys* **5** (1943).

22 Rosenblatt, F. The perceptron - a perceiving and recognizing automaton (Cornell Aeronautical Laboratory, New York, U.S.A., 1957).

23 Chollet, F. in *Lex Fridman Podcast* (ed Lex Fridman) (YouTube, 2020).

24 Descartes, R. *Rules for the direction of the natural intelligence (translated by George Heffernan in 1998)* (Brill Rodopi, 1619-1628).

25 Galilei, G. *The assayer (translated by Stillman Drake in 1960)*. (University of Pennsylvania Press, 1623).

26 Hobbes, T. *Leviathan (edited by Edwin Curley in 1994)*. (Hackett Publishing Co, Inc, 1651).

27 Pólya, G. *How to solve it* (Princeton University Press, 1945).

28 Courant, R. Variational methods for the solution of problems of equilibrium and vibrations. *Bull. Amer. Math. Soc* **49** (1943).

29 Cauchy, A. Methodes generales pour la resolution des syst'emes dequations simultanees (translated by Richard J. Pulskamp in https://cs.uwaterloo.ca/~y328yu/classics/cauchy-en.pdf). *C.R. Acad. Sci. Par* **25**, 536–538 (1847).
30 Rumelhart, D. E., Hinton, G. E. & Williams, R. J. Learning representations by back-propagating errors. *Nature* **323**, 533-536 (1986).
31 LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
32 Minsky, M. & Papert, S. A. *Perceptrons: An Introduction to Computational Geometry*. Vol. (2nd edition with corrections, first edition 1969) (The MIT Press, 1972).
33 Krizhevsky, A., Sutskever, I. & Hinton, G. E. ImageNet classification with deep convolutional neural networks. *Communications of the ACM.* **60** (2017).
34 Ivakhnenko, A. G. e. & Lapa, V. G. v. *Cybernetic predicting devices*. (CCM Information Corp., 1965).
35 Crick, F. The recent excitement about neural networks *Nature* **337**, 129–132 (1989).
36 Marblestone, A. H., Wayne, G. & Kording, K. P. Toward an integration of deep learning and neuroscience *Front. Comput. Neurosci.* **10**, 2016 (2016).
37 Whittington, J. C. R. & Bogacz, R. An approximation of the error backpropagation algorithm in a predictive coding network with local Hebbian synaptic plasticity. *Neural Comput.* **29** (2017).
38 Schmidhuber, J. Deep learning in neural networks: An overview. *Neural Netw.* **61**, 85-117 (2015).
39 Cassidy, A. S., Georgioub, J. & Andreou, A. G. Design of silicon brains in the nano-CMOS era: Spiking neurons, learning synapses and neural architecture optimization. *Neural Netw.* **45**, 4-26 (2013).
40 Maass, W. Networks of spiking neurons: the third generation of neural network models. *Neural Netw.* **10** (1996).
41 Neumann, J. v. *The computer and the brain*. (Yale University Press, 1958).
42 Liu, L. in *Nautilus* (Nautilus Quarterly, New York, U.S.A., 2020).
43 Moravec, H. *Mind children: the future of robot and human intelligence*. (Harvard University Press, 1988).
44 Kandel, E. R., Schwartz, J. H. & Jessell, T. M. *Principles of Neural Science*. (McGraw-Hill, 1981).
45 Ho, T. K.Random decision forests in *Proceedings of the 3rd International Conference on Document Analysis and Recognition.* (IEEE, 1995).
46 Oguchi, T., Higuchi, M., Uno, T., Kamaya, M. & Suzuki, M. in *International Solid State Circuit Conference* (IEEE San Francisco, 1981).
47 Walker, R. & Tersini, N. *Silicon destiny: The story of Application Specific Integrated Circuits and .LSI Logic Corporation*. (C.M.C Publications, 1992).
48 Katz, R. H. & Borriello, G. *Contemporary logic design*. (Pearson Prentice Hall, 1993).
49 Furze, T. A. & Bennett, B.Using the principles of classical conditioning to learn event sequences in *Proceedings of the Artificial Intelligence and the Simulation of Behaviour (AISB) Conference 2011.* (eds D. Kazakov & G. Tsoulas) 40-47 (Society for the Study of Artificial Intelligence and the Simulation of Behaviour, York, U.K., 2011).
50 Furze, T. A. *The application of classical conditioning to the machine learning of a commonsense knowledge of visual events* PhD thesis, University of Leeds, (2013).
51 Andreae, J. H. *Associative learning for a robot intelligence* (Imperial College Press, 1998).
52 Kohonen, T. *Content addressable memories*. (Springer-Verlag, 1980).
53 Sebastian, A. *et al.* Temporal correlation detection using computational phase-change memory. *Nat. Commun.* **8** (2017).
54 Thorndike, E. L. The elements of psychology. *A. G. Seiler* (1905).
55 Skinner, B. F. *The behavior of organisms: an experimental analysis*. (Appleton-Century-Crofts, 1938).
56 Watson, J. B. & Rayner, R. Conditioned emotional reactions. *J. Exp. Psychol. Learn. Mem. Cogn* **3**, 1–14 (1920).

57 Budampati, R. & Kolavennu, S. *Industrial wireless sensor networks*. (Elsevier Ltd. , 2016).
58 Turkcan, E., Ciftcioglu, O. & Hagen, T. H. J. J. v. in *Diagnostic systems in nuclear power plants, Proceedings of a Technical Committee Meeting* (International Atomic Energy Agency, Istanbul, Turkey, 1998).
59 Pao, Y.-H. in *Final report* (Office of Scientific and Technical Information, U.S. Department of Energy U.S.A., 1982).
60 Sobajic, D. J.Neural network computing for the electric power industry in *Inns Summer Workshop.* (Lawrence Erlbaum Associates Publishers 1992).
61 Ruz-Hernandez, J. A., Suarez, D. A., Garcia-Hernandez, R., Sanchez, E. N. & Suarez-Duran, M. U.Optimal training algorithm application to design fossil electric power plant in *8th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes (SAFEPROCESS).* (International Federation of Automatic Control (IFAC), 2012).
62 Minsky, M. Logical versus analogical or symbolic versus connectionist or neat versus scruffy. *AI Magazine* **12**, 34-51 (1991).
63 Garcez, A. d. A. & Lamb, L. Neurosymbolic AI: The 3rd Wave. *arXiv:2012.05876* (2020).
64 Tan, J. Y. S. *et al.* Monadic Pavlovian associative learning in a backpropagation-free photonic network. *arXiv:2011.14709* (2020).
65 Barbour, G., Blackwell, K., Busse, T., Alkon, D. & Vogl, T. Dystal: a Self-organizing ANN with Pattern Independent Training Time *Intern. Join. Conf. Neur. Net.* **4**, 814-819 (1992).
66 Hopfield, J. J. Neural networks and physical systems with emergent collective computational abilities. *Proc. NatL Acad. Sci. USA* **79**, 2554-2558 (1982).
67 Hopfield, J. J. & Tank, D. W. Computing with neural circuits: a model *Science* **233** (1986).
68 Andraea, J. H. & Cleary, J. G. A new mechanism for the brain *Int. J. Man-Machine Studies* **8**, 89-119 (1976).
69 Tan, J. Y. S., Feldmann, J. & Bhaskaran, H. 'Photonic content addressable memory', Patent No. UK Patent Application No. 2109302.6 (filed on 28 June 2021), (2021).
70 Minzioni, P. *et al.* Roadmap on all-optical processing. *J. Opt.* **21** (2019).
71 Berggren, K. *et al.* Roadmap on emerging hardware and technology for machine learning. *Nanotechnology* **32** (2020).
72 Christensen, D. V. *et al.* 2021 roadmap on neuromorphic computing and engineering. *arXiv:2105.05956* (2021).
73 Willner, A. E., Khaleghi, S., Chitgarha, M. R. & Yilmaz, O. F. All-optical signal processing. *J. Light. Technol.* **32** (2014).
74 MenloSystems. 'C-fiber 780 femtosecond Erbium laser'. Retrieved from https://www.menlosystems.com/products/femtosecond-lasers-and-amplifiers/c-fiber-780/. Accessed on 27 August 2021.
75 Washburn, B. R. *et al.* Phase-locked, erbium-fiber-laser-based frequency comb in the near infrared. *Opt. Lett* **29**, 250-252 (2004).
76 Kim, Y., Kim, S., Kim, Y.-J., Hussein, H. & Kim, S.-W. Er-doped fiber frequency comb with mHz relative linewidth. *Opt. Express* **17**, 11972-11977 (2009).
77 Kim, Y.-J., Chun, B. J., Kim, Y., Hyun, S. & Kim, S.-W. Generation of optical frequencies out of the frequency comb of a femtosecond laser for DWDM telecommunication. *Laser Phys. Lett* **7** (2010).
78 Rudman, P. S. *How Mathematics Happened: The First 50,000 Years*. (Prometheus Books, 2007).
79 Brattain, W. H. & Bardeen, J. 'Three-electrode circuit element utilizing semiconductive materials', Patent No. US 2524035, (1948).
80 Bardeen, J. & Brattain, W. H. The transistor, a semi-conductor triode *Phys. Rev.* **74** (1948).
81 Heilmeier, G. H. Transient behavior of domains in liquid crystals. *J. Chem. Phys.* **44** (1966).
82 Hoff, M. E., Mazor, S. & Faggin, F. 'Memory system for a multi-chip digital computer', Patent No. US3821715, (1973).

83 Boone, G. W. & Cochran, M. J. 'Variable function programmed calculator', Patent No. US4074351, (1977).

84 Nagarajan, R. *et al.* Large-scale photonic integrated circuits. *IEEE J. Sel. Top. Quantum Electron* **11** (2005).

85 Coombs, A. W. M. The making of Colossus. *IEEE Ann. Hist. Comput* **5** (1983).

86 McCartney, S. *ENIAC: The Triumphs and Tragedies of the World's First Computer*. (Berkley Books, 2001).

87 Turing, A. M. Computing machinery and intelligence *Mind* **49**, 433-460 (1950).

88 Neumann, J. v. First Draft of a Report on the EDVAC. (1945).

89 Williams, F. C., Kilburn, T. & Tootill, G. Electronic digital computers. *Nature* **162**, 487 (1948).

90 Jalali, B. & Fathpour, S. Silicon photonics *J. Light. Technol* **24** (2006).

91 Ambrogio, S. *et al.* Equivalent-accuracy accelerated neuralnetwork training using analogue memory. *Nature* **558** (2018).

92 Li, C. *et al.* Analogue signal and image processing with large memristor crossbars. *Nat. Electron* **1**, 52-59 (2018).

93 Nvidia. 'Inference Platforms for HPC Data Centers. Retrieved from https://www.nvidia.com/en-sg/deep-learning-ai/inference-platform/hpc/. Accessed on 17 July 2021. '. Retrieved from Accessed

94 Hanlon, A. G. Content-addressable and associative memory systems: a survey. *IEEE Trans. Comput.* **EC-15** (1966).

95 Jouppi, N. P. *et al.*In-datacenter performance analysis of a Tensor Processing Unit in *ISCA '17: Proceedings of the 44th Annual International Symposium on Computer Architecture.* (2017).

96 Sefat, M. S., Aslan, S., Kellington, J. W. & Qasem, A.Accelerating HotSpots in Deep Neural Networks on a CAPI-based FPGA in *2019 IEEE 21st International Conference on High Performance Computing and Communications; IEEE 17th International Conference on Smart City; IEEE 5th International Conference on Data Science and Systems (HPCC/SmartCity/DSS).* 248–256 (2019).

97 Merolla, P. A. *et al.* A million spiking-neuron integrated circuit with a scalable communication network and interface. *Science* **345** (2014).

98 Sze, D. K. a. S. M. A floating gate and its application to memory devices. *Bell Syst. Tech. J* **46** (1967).

99 Pavan, P., Bez, R., Olivo, P. & Zanoni, E. Flash memory cells—An overview. *Proc. IEEE* **85**, 1248-1271 (1997).

100 Aunet, S. & Hartmann, M. Real-time reconfigurable linear threshold elements and some applications to neural hardware. *ICES 2003: Evolvable Systems: From Biology to Hardware*, 365-376 (2003).

101 Koo, J. *et al.* Nonvolatile electric double-layer transistor memory devices embedded with Au nanoparticles. *ACS Appl. Mater. Interfaces* **10**, 9563−9570 (2018).

102 Leong, W. L. *et al.* Towards printable organic thin film transistor based flash memory devices. *J. Mater. Chem.* **21**, 5203 (2011).

103 Lim, E. W. & Ismail, R. Conduction mechanism of valence change resistive switching memory: A survey. *Electronics* **4**, 586-613 (2015).

104 Gale, E. TiO2-based memristors and ReRAM: materials, mechanisms and models (a review). *Semicond. Sci. Technol.* , 104004 (2014).

105 Strukov, D. B., Snider, G. S., Stewart, D. R. & Williams, R. S. The missing memristor found. *Nature* **453**, 80–83 (2008).

106 Williams, R. S. How we found the missing memristor *IEEE Spec.* **45**, 28 - 35 (2008).

107 Chen, J.-Y., Huang, C.-W., Chiu, C.-H., Huang, Y.-T. & Wu, W.-W. Switching kinetic of VCM-based memristor: Evolution and positioning of nanofilament. *Adv. Mat.* **27** (2015).

108 Kawahara, T., Ito, K., Takemura, R. & Ohno, H. Spin-transfer torque RAM technology: Review and prospect. *Microelectron. Reliab.* **52**, 613-627 (2012).

109 Chumak, A. V., Vasyuchka, V. I., Serga, A. A. & Hillebrands, B. Magnon spintronics. *Nat. Phys.* **11**, 453–461 (2015).
110 Ovshinsky, S. R. Reversible electrical switching phenomena in disordered structures. *Phys. Rev. Lett.* **21**, 1450 (1968).
111 Yamada, N., Ohno, E., Nishiuchi, K., Akahira, N. & Takao, M. Rapid-phase transitions of GeTe-Sb2Te3 pseudobinary amorphous thin films for an optical disk memory. *J. Appl. Phys* **69**, 2849 (1991).
112 Raoux, S., Wełnic, W. & Ielmini, D. Phase change materials and their application to nonvolatile memories. *Chem. Rev.* **110** (2010).
113 Siegel, J., Schropp, A., Solis, J., Afonso, C. N. & Wuttig, M. Rewritable phase-change optical recording in Ge2Sb2Te5 films induced by picosecond laser pulses. *Appl. Phys. Lett.* **84**, 2250 (2004).
114 Agarwal, S. C. Role of potential fluctuations in phase-change GST memory devices. *Phys. Status Solidi B* **249**, 1956–1961 (2012).
115 Pirovano, A. *et al.*Scaling analysis of phase-change memory technology in *IEEE International Electron Devices Meeting.* 699-702 (Institute of Electrical and Electronics Engineers (IEEE), 2003).
116 Lencer, D., Salinga, M. & Wuttig, M. Design rules for phase-change materials in data storage applications. *Adv. Mat.* **23**, 2030-2058 (2011).
117 Raoux, S., Xiong, F., Wuttig, M. & Pop, E. Phase change materials and phase change memory. *MRS Bulletin* **39**, 703-710 (2014).
118 Jacobi, D. P. W. 'Halbleiterverstärker', Patent No. DE 833366C, (1949).
119 Kilby, J. S. C. 'Miniaturized electronic circuits', Patent No. US 3138743, (1959).
120 Noyce, R. N. 'Semiconductor device and lead structure', Patent No. US 2981877, (1959).
121 Lilienfeld, J. E. 'Method and apparatus for controlling electric currents', Patent No. US 1745175, (1926).
122 Lilienfeld, J. E. 'Specification of electric current control mechanism patent application', Patent No. CA 272437, (1927).
123 Lilienfeld, J. E. 'Amplifier for electric currents', Patent No. US 1877140, (1932).
124 Lilienfeld, J. E. 'Device for controlling electric current ', Patent No. US 1900018, (1933).
125 Heil, O. 'Improvements in or relating to electrical amplifiers and other control arrangements and devices', Patent No. GB 439457, (1935).
126 Shockley, W. 'Circuit element utilizing semiconductor material ', Patent No. US 2569347, (1948).
127 Agrell, E. *et al.* Roadmap of optical communications. *J. Opt.* **18** (2016).
128 Maxwell, J. C. *A treatise on electricity and magnetism*. (Clarendon Press, 1873).
129 Huygens, C. *Traité de la Lumière (translated by Silvanus P. Thompson as Treatise on Light in 1912)*. (Macmillan 1690).
130 Born, M. & Wolf, E. *Principles of Optics*. (Pergamon Press, 1959).
131 Wu, Y.-D. Nonlinear all-optical switching device by using the spatial soliton collision. *Fiber and Integrated Optics* **23** (2004).
132 Pramono, Y. H. & Endarko. Nonlinear waveguides for optical logic and computation. *J. Nonlinear Opt. Phys. Mater.* **10** (2001).
133 Sauter, F. On the behavior of an electron in a homogeneous electric field in Dirac's relativistic theory (Translated by D. H. Delphenich). *Zeit. f. Phys.* **69** (1931).
134 Almeida, V. R., Barrios, C. A., Panepucci, R. R. & Lipson, M. All-optical control of light on a silicon chip *Nature* **431**, 1081–1084 (2004).
135 Kinsey, N., DeVault, C., Boltasseva, A. & Shalaev, V. M. Near-zero-index materials for photonics. *Nat. Rev. Mater* **4**, 742–760 (2019).
136 Reshef, O., Leon, I. D., Alam, M. Z. & Boyd, R. W. Nonlinear optical effects in epsilon-near-zero media. *Nat. Rev. Mater* **4** (2019).

137 Zhang, J., MacDonald, K. F. & Zheludev, N. I. Controlling light-with-light without nonlinearity *Light Sci. Appl* **1** (2012).
138 Kitayama, K.-i. *et al.* Novel frontier of photonics for data processing - Photonic accelerator. *APL Photonics* **4** (2019).
139 Tu, X., Song, C., Huang, T., Chen, Z. & Fu, H. State of the art and perspectives on silicon photonic switches. *Micromachines* **10** (2019).
140 Nozaki, K. *et al.* Femtofarad optoelectronic integration demonstrating energy-saving signal conversion and nonlinear functions. *Nat. Photon.* **13** (2019).
141 Baranov, D. G., Krasnok, A., Shegai, T., Alù, A. & Chong, Y. Coherent perfect absorbers: linear control of light with light. *Nat. Rev. Mater* **2** (2017).
142 Wang, X., Weigei, P. O., Zhao, J., Ruesing, M. & Mookherjea, S. Achieving beyond-100-GHz large-signal modulation bandwidth in hybrid silicon photonics Mach Zehnder modulators using thin film lithium niobate. *APL Photon* **4** (2019).
143 Harris, N. C. *et al.* Efficient, compact and low loss thermo-optic phase shifter in silicon. *Opt. Express* **22**, 10487-10493 (2014).
144 Smith, P. W. On the physical limits of digital optical switching and logic elements. *Bell Labs Tech. J.* **61** (1982).
145 Qin, F., Liu, Y. & Li, Z.-Y. Optical switching in hybrid semiconductor nonlinear photonic crystal slabs with Kerr materials. *J. Opt* **12** (2010).
146 Dani, K. M. *et al.* Subpicosecond optical switching with a negative index metamaterial. *Nano Lett.* **9** (2009).
147 Ríos, C. *et al.* Integrated all-photonic non-volatile multi-level memory. *Nat. Photon.* **9**, 725-733 (2015).
148 Rudé, M., Simpson, R. E., Quidant, R., Pruneri, V. & Renger, J. Active control of surface plasmon waveguides with a phase change material. *ACS Photonics* **2** (2015).
149 Bach, H.-G. *et al.* InP-based waveguide-integrated photodetector with 100-GHz bandwidth. *IEEE J. Sel. Top. Quantum Electron* **10** (2004).
150 Luque, A. & Hegedus, S. *Handbook of photovoltaic science and engineering*. (Wiley, 2011).
151 Zhang, W., Mazzarello, R., Wuttig, M. & Ma, E. Designing crystallization in phase-change materials for universal memory and neuro-inspired computing. *Nat. Rev. Mater* **4** (2019).
152 Sharma, Y. D. & Bhatnagar, P. K.Characterization of Ag-Sb-Te alloy and their films for phase change optical memories in *Proceedings Volume 5188, Advanced Characterization Techniques for Optics, Semiconductors, and Nanotechnologies.* (SPIE, 2003).
153 Vahala, K. Optical microcavities *Nature* **424**, 839-846 (2003).
154 Perot, A. & Fabry, C. On the application of interference phenomena to the solution of various problems of spectroscopy and metrology. *Astrophysical Journal* **9**, 87 (1899).
155 Strutt, J. W. *The theory of sound*. (MacMillan, 1878).
156 Mie, G. Contributions on the optics of turbid media, particularly colloidal metal solutions (translated by Patricia Newman in 1978). *Ann. Phys.* **330**, 377-445 (1908).
157 Debye, P. The light pressure on spheres of arbitrary material (Der lichtdruck auf kugeln von beliebigem material). *Ann. Phys.* **30**, 57-136 (1909).
158 Yang, K. Y. *et al.* Bridging ultrahigh-Q devices and photonic circuits *Nat. Photon.* **12**, 297–302 (2018).
159 Abbe, E. Contributions to the theory of the microscope and of microscopic perception (translated by Barry R. Masters in 'Superresolution optical microscropy' Springer (2021)) *Archiv für Mikroskopische Anatomie* **9**, 413–468 (1873).
160 Alexoudi, T., Kanellos, G. T. & Pleros, N. Optical RAM and integrated optical memories: a survey. *Light Sci. Appl* **9** (2020).
161 Hunter, D. K., Chia, M. C. & Andonovic, I. Buffering in optical packet switches. *J. Light. Technol* **16** (1998).

162 Burmeister, E. F., Blumenthal, D. J. & Bowers, J. E. A comparison of optical buffering technologies *Opt. Switch. Netw.* **5**, 10-18 (2008).
163 Khurgin, J. B. & Tucker, R. S. *Slow light: Science and applications*. (CRC Press 2018).
164 Corning SMF-28 Ultra 200 optical fiber: Product information (2019).
165 Biot, M. A. General theorems on the equivalence of group velocity and energy transport *Phys. Rev.* **105** (1957).
166 Lighthill, M. J. Group velocity. *J. Inst. Maths Applics.* **1**, 1-8 (1964).
167 Yariv, A., Xu, Y., Lee, R. K. & Scherer, A. Coupled-resonator optical waveguide: a proposal and analysis *Opt. Lett* **24** (1999).
168 Lee, M., Gehm, M. E. & Neifeld, M. A. Systematic design study of an all-optical delay line based on Brillouin scattering enhanced cascade coupled ring resonators. *J. Opt* **12** (2010).
169 Sedgwick, F. G., Chang-Hasnain, C. J., Ku, P. C. & Tucker, R. S. Storage-bit-rate product in slow-light optical buffers. *Electron. Lett* **41** (2005).
170 Tucker, R. S., Ku, P.-C. & Chang-Hasnain, C. J. Slow-light optical buffers: Capabilities and fundamental limitations. *J. Light. Technol* **23** (2005).
171 Lu, W. T., Savo, S., Casse, D. F. & Sridhar, S. Slow microwave waveguide made of negative permeability metamaterials. *Microw. Opt. Technol. Lett.* **51** (2009).
172 Tsakmakidis, K. L., Boardman, A. D. & Hess, O. 'Trapped rainbow' storage of light in metamaterials *Nature* **450**, 397–401 (2007).
173 Savo, S., Casse, B. D. F., Lu, W. & Sridhar, S. Observation of slow-light in a metamaterials waveguide at microwave frequencies. *Appl. Phys. Lett.* **98** (2011).
174 Ginsberg, N. S., Garner, S. R. & Hau, L. V. Coherent control of optical information with matter wave dynamics *Nature* **445** 623–626 (2007).
175 Gibbs, H. M., McCall, S. L. & Venkatesan, T. N. C. Differential gain and bistability using a sodium-filled Fabry-Perot interferometer. *Phys. Rev. Lett.* **36** (1976).
176 Parthenopouslos, D. A. & Rentzepis, P. M. Three-dimensional optical storage memory. *Science* **245** (1989).
177 Kawaguchi, H. Bistable laser diodes and their applications: state of the art. *IEEE J. Sel. Top. Quantum Electron* **3** (1997).
178 Bakopoulos, P. *et al.* All-optical T-flip-flop using a single SOA-MZI-based latching element. *IEEE Photonics Technol. Lett.* **24** (2012).
179 Ueki, H., Kawata, Y. & Kawata, S. Three-dimensional optical bit-memory recording and reading with a photorefractive crystal: analysis and experiment. *Appl. Opt.* **35** (1996).
180 Sasaki, H., Fainman, Y., Ford, J. E., Taketomi, Y. & Lee, S. H. Dynamic photorefractive optical memory. *Opt. Lett* **16** (1991).
181 Chen, F. S. Optically induced change of refractive indices in LiNbO3 and LiTaO3. *J. Appl. Phys* **40** (1969).
182 Hill, M. T., Waardt, H. d., Khoe, G. D. & Dorren, H. J. S. All-optical flip-flop based on coupled laser diodes. *IEEE J. Sel. Top. Quantum Electron* **37** (2001).
183 Nozaki, K. *et al.* Ultralow-power all-optical RAM based on nanocavities. *Nat. Photon.* **6** (2012).
184 Pleros, N., Apostolopoulos, D., Petrantonakis, D., Stamatiadis, C. & Avramopoulos, H. Optical static RAM cell *IEEE Photonics Technol. Lett.* **21** (2009).
185 Hill, M. T. *et al.* A fast low-power optical memory based on coupled micro-ring lasers *Nature* **432**, 206–209 (2004).
186 Liu, L. *et al.* An ultra-small, low-power, all-optical flip-flop memory on a silicon chip. *Nat. Photon.* **4** (2010).
187 Naito, Y., Shimizu, S., Kato, T., Kobayashi, K. & Uenohara, H. Investigation of all-optical latching operation of a monolithically integrated SOA-MZI with a feedback loop. *Opt. Express* **20** (2012).
188 Bruck, R. *et al.* All-optical spatial light modulator for reconfigurable silicon photonic circuits. *Optica* **3** (2016).

189 Kuhr, S. *et al.* An ultrahigh Finesse Fabry-Perot superconducting resonator as a photon box for cavity-QED experiments. *arXiv:quant-ph/0612138* (2006).
190 Noh, W., Dupré, M., Ndao, A., Kodigala, A. & Kanté, B. Suspended single mode microdisk lasers. *arXiv:1812.03805* (2018).
191 Lee, H., Chen, T., Li, J., Painter, O. & Vahala, K. J. Ultra-low-loss optical delay line on a silicon chip. *Nat. Commun.* **3** (2012).
192 Goldzak, T., Mailybaev, A. A. & Moiseyev, N. Light stops at exceptional points. *Phys. Rev. Lett.* **120** (2018).
193 Miller, S. E. Integrated optics: an introduction. *Bell Labs Tech. J.* **48** (1969).
194 Mercatili, E. A. J. Dielectric rectangular waveguide and directional coupler for integrated optics. *Bell Labs Tech. J.* **48** (1969).
195 Mercatili, E. A. J. Bends in optical dielectric guides. *Bell Labs Tech. J.* **48** (1969).
196 Kao, K. C. & Hockham, G. A. Dielectric-fibre surface waveguides for optical frequencies. *Proc. IEE.* **113** (1966).
197 Maiman, T. H. Stimulated optical radiation in Ruby. *Nature* **187** (1960).
198 Kakitsuka, T. *et al.* Lateral current-injection membrane lasers fabricated on a silicon substrate. *NTT Tech. Rev.* **14** (2016).
199 Arafin, S., Simsek, A., Lu, M., Rodwell, M. J. & Coldren, L. A. Heterodyne locking of a fully integrated optical phase-locked loop with on-chip modulators. *Opt. Lett* **42** (2017).
200 Feng, N.-N. *et al.* Vertical p-i-n germanium photodetector with high external responsivity integrated with large core Si waveguides. *Opt. Express* **18** (2009).
201 Ribeiro, A. & Bogaerts, W. Digitally controlled multiplexed silicon photonics phase shifter using heaters with integrated diodes. *Opt. Express* **25** (2017).
202 Feng, N.-N. *et al.* High speed carrier-depletion modulators with 1.4V-cm VπL integrated on 0.25µm silicon-on insulator waveguides. *Opt. Express* **18** (2010).
203 Augustin, L. M., Hanfoug, R., Tol, J. J. G. M. v. d., Laat, W. J. M. d. & Smit, M. K. A compact integrated polarization splitter/converter in InGaAsP–InP. *IEEE Photon. Technol. Lett* **19** (2007).
204 Vlasov, Y. A. & McNab, S. J. Losses in single-mode silicon-on-insulator strip waveguides and bends. *Opt. Express* **12** (2004).
205 Yang, L. *et al.* Compact 2x2 tapered multimode interference couplers based on SU-8 polymer rectangular waveguides. *Appl. Phys. Lett.* **93** (2008).
206 Dai, D., Wang, J., Chen, S., Wang, S. & He, S. Monolithically integrated 64-channel silicon hybrid demultiplexer enabling simultaneous wavelength and mode-division-multiplexing. *Laser Photonics Rev.* **9** (2015).
207 Alexander, K. *et al.* Nanophotonic Pockels modulators on a silicon nitride platform. *Nat. Commun.* **9** (2018).
208 Iyer, S. S. & Xie, Y.-H. Light emission from silicon. *Science* **260** (1993).
209 Levinshtein, M., Rumyantsev, S. & Shur, M. *Handbook series on semiconductor parameters*. (World Scientific Publishing, 1996).
210 Feng, D. *et al.* High-speed Ge photodetector monolithically integrated with large crosssection silicon-on-insulator waveguide. *Appl. Phys. Lett.* **95**, 261105 (2009).
211 Colace, L. *et al.* Metal–semiconductor–metal near-infrared light detector based on epitaxial Ge/Si. *Appl. Phys. Lett.* **72** (1998).
212 Pintus, P., Pasquale, F. D. & Bowers, J. E. Integrated TE and TM optical circulators on ultra-low-loss silicon nitride platform. *Opt. Express* **21** (2013).
213 Luke, K., Okawachi, Y., Lamont, M. R. E., Gaeta, A. L. & Lipson, M. Broadband mid-infrared frequency comb generation in a Si3N4 microresonator. *Opt. Lett* **40** (2015).
214 Fan, Y. *et al.* Optically integrated InP–Si3N4 hybrid laser. *IEEE Photon. Technol. Lett* **8** (2016).
215 Theurer, M. *et al.* Flip-chip integration of InP and SiN. *IEEE Photon. Technol. Lett* **31** (2019).

216 Munakata, F. *et al.* Optical properties of b-Si3N4 single crystals grown from a Si melt in N2 *Appl. Phys. Lett.* **74** (1999).
217 Piels, M., Bauters, J. F., Davenport, M. L., Heck, M. J. R. & Bowers, J. E. Low-loss silicon nitride AWG demultiplexer heterogeneously integrated with hybrid III–V/silicon photodetectors. *J. Light. Technol* **32** (2014).
218 Kitao, A., Imakita, K., Kawamura, I. & Fujii, M. An investigation into second harmonic generation by Si-rich SiNx thin films deposited by RF sputtering over a wide range of Si concentrations. *J. Phys. D: Appl. Phys.* **47**, 215101 (2014).
219 Izhaky, N. *et al.* Development of CMOS-compatible integrated silicon photonics devices. *IEEE J. Sel. Top. Quantum Electron* **12** (2006).
220 Bucio, T. D., Tarazona, A., Khokhar, A. Z., Mashanovich, G. Z. & Gardes, F. Y. Low temperature silicon nitride waveguides for multilayer platforms. *Proc. SPIE, Silicon photonics and photonic integrated circuits V* **9891** (2016).
221 Iler, R. K. *The Chemistry of Silica: Solubility, Polymerization, Colloid and Surface Properties, and Biochemistry*. (Wiley, 1979).
222 Kruckel, C. J., Fulop, A., Ye, Z., Andrekson, P. A. & Torres-Company, V. Optical bandgap engineering in nonlinear silicon nitride waveguides. *Opt. Express* **25**, 15370-15380 (2017).
223 Crystran. The Crystran handbook of infra-red and ultra-violet optical materials. (Crystran Ltd., Poole, U.K., 2021).
224 Park, H., Fang, A. W., Kodamaa, S. & Bowers, J. E. Hybrid silicon evanescent laser fabricated with a silicon waveguide and III-V offset quantum wells. *Opt. Express* **13** (2005).
225 Komljenovic, T. *et al.* Heterogeneous silicon photonic integrated circuits. *J. Light. Technol* **34** (2016).
226 Liu, J., Kimerling, L. C. & Michel, J. Monolithic Ge-on-Si lasers for large-scale electronic–photonic integration. *Semicond. Sci. Technol.* **27**, 094006 (2012).
227 Liu, A. Y., Srinivasan, S., Norman, J., Gossard, A. C. & Bowers, J. E. Quantum dot lasers for silicon photonics. *Photon. Res.* **3** (2015).
228 Matsumoto, T. *et al.* Hybrid-integration of SOA on silicon photonics platform based on flip-chip bonding. *J. Light. Technol* **37** (2019).
229 Davenport, M. L. *et al.* Heterogeneous silicon/III–V semiconductor optical amplifiers. *IEEE J. Sel. Top. Quantum Electron* **22** (2016).
230 Chang, H.-H., Kuo, Y.-h., Jones, R., Barkai, A. & Bowers, J. E. Integrated hybrid silicon triplexer. *Opt. Express* **18** (2010).
231 Masini, G., Colace, L. & Assanto, G. 2.5 Gbit/s polycrystalline germanium-on silicon photodetector operating from 1.3 to 1.55 µm. *Appl. Phys. Lett.* **82** (2003).
232 Yu, H. *et al.* in *2012 14th International Conference on Transparent Optical Networks (ICTON)* (IEEE, Coventry, U.K., 2012).
233 Adachi, S. Optical dispersion relations for GaP, GaAs, GaSb, InP, InAs, InSb, AlxGa1−xAs, and In1−xGaxAsyP1−y. *J. Appl. Phys* **66** (1989).
234 Juodawlkis, P. W., Plant, J. J., Huang, R. K., Missaggia, L. J. & Donnelly, J. P. High-power 1.5-µm InGaAsP–InP slab-coupled optical waveguide amplifier. *IEEE Photon. Technol. Lett* **17** (2005).
235 Jelen, C. L., Slivken, S., David, T., Brown, G. J. & Razeghi, M. Responsivity and noise performance of InGaAs/InP quantum well infrared photodetectors. *Proc. SPIE* **3287** (1998).
236 Liang, D., Huang, X., Kurczveil, G., Fiorentino, M. & Beausoleil, R. G. Integrated finely tunable microring laser on silicon. *Nat. Photon.* **10**, 719-722 (2016).
237 Pavesi, I., Piazza, F., A. Rudra, Carlin, J. F. & Ilegems, M. Temperature dependence of the InP band gap from a photoluminescence study. *Phys. Rev. B* **44** (1991).
238 Zelmon, D. E., Small, D. L. & Jundt, D. Infrared corrected Sellmeier coefficients for congruently grown lithium niobate and 5 mol. % magnesium oxide–doped lithium niobate. *J. Opt. Soc. Am. B* **14** (1997).

239 Dhar, A. & Mansingh, A. Optical properties of reduced lithium niobate single crystals. *J. Appl. Phys* **68** (1990).
240 Schmidt, W. G., Albrecht, M., Wippermann, S., Blankenburg, S. & Rauls, E. LiNbO3 ground- and excited-state properties from first-principles calculations. *Phys. Rev. B* **77** (2008).
241 Desiatov, B. & Lončar, M. Silicon photodetector for integrated lithium niobate photonics. *Appl. Phys. Lett.* **115**, 121108 (2019).
242 Kaminow, I. P. & Turner, E. H. *Linear electrooptical materials, in "Handbook of Lasers" (R. J. Pressley, ed.)*. 447-459 (Chemical Rubber Co., 1971).
243 Volk, T., Wöhlecke, M. & Rubinina, N. *Optical damage resistance in lithium niobate (in Photorefractive Materials and Their Applications 2)*. (Springer, 2007).
244 Zhang, X., Qiu, J., Li, X., Zhao, J. & Liu, L. Complex refractive indices measurements of polymers in visible and near-infrared bands. *Appl. Opt.* **59** (2020).
245 Lee, J. H. *et al.* Tunable external cavity laser based on polymer waveguide platform for WDM access network. *IEEE Photon. Technol. Lett* **17** (2005).
246 Samuel, I. D. W. & A.Turnbull, G. Polymer lasers: recent advances. *Mater. Today* **7** (1998).
247 Fan, R. S. & Hooker, R. B. Hybrid optical switch using passive polymer waveguides and semiconductor optical amplifiers. *J. Light. Technol* **18** (2000).
248 Zhang, Z. *et al.* in *ECIO 2010* (European Conference on Integrated Optics, Cambridge, U.K., 2010).
249 Wu, W., Qin, J. & Li, Z. New design strategies for second-order nonlinear optical polymersand dendrimers. *Polymer* **54**, 4351-4382 (2013).
250 Razeghi, M. High-power laser diodes based on InGaAsP alloys. *Nature* **369**, 631–633 (1994).
251 Garbuzov, D. Z. *et al.* High-power 0.8 pm InGaAsP-GaAs SCH SQW lasers. *IEEE J. Quantum Electron* **27** (1991).
252 Stabile, R., A.Rohit & Williams, K. A. Monolithically integrated 8 × 8 space and wavelength selective cross-connect. *J. Light. Technol* **32** (2014).
253 Weimann, C., Lauermann, M., Hoeller, F., Freude, W. & Koos, C. Silicon photonic integrated circuit for fast and precise dual-comb distance metrology. *Opt. Express* **25** (2017).
254 Perez-Lopez, D., Gutierrez, A. & Capmany, J. Silicon nitride programmable photonic processor with folded heaters. *Opt. Express* **29** (2021).
255 Zhang, Z. & Keil, N. Thermo-optic devices on polymer platform. *Opt. Commun.* **362**, 101-114 (2016).
256 Wang, C. *et al.* Monolithic lithium niobate photonic circuits for Kerr frequency comb generation and modulation. *Nat. Commun.* **10** (2019).
257 Howell, B. J. Optical analog computers. *J. Opt. Soc. Am.* **49** (1959).
258 Solli, D. R. & Jalali, B. Analog optical computing. *Nat. Photon.* **9** (2015).
259 Slavík, R., Park, Y., Kulishov, M., Morandotti, R. & Azaña, J. Ultrafast all-optical differentiators. *Opt. Express* **14** (2006).
260 Slavík, R. *et al.* Photonic temporal integrator for all-optical computing. *Opt. Express* **16** (2008).
261 Golovastikov, N. V., Bykov, D. A., Doskolovich, L. L. & Bezus, E. A. Spatial optical integrator based on phase-shifted Bragg gratings. *Opt. Commun.* **338** (2015).
262 Park, Y. & Azaña, J. Optical signal processors based on a time-spectrum convolution. *Opt. Lett.* **35** (2010).
263 Wilson, R. *Fourier series and optical transform techniques in contemporary optics.*, (John Wiley & Sons Inc, 1995).
264 Casasent, D. & Ghosh, A. Optical linear algebra processors: noise and error-source modeling *Opt. Lett* **10** (1985).
265 Miller, D. A. B. Self-configuring universal linear optical component. *Photon. Res.* **1** (2013).
266 Estakhri, N. M., Edwards, B. & Engheta, N. Inverse-designed metastructures that solve equations *Science* **363** (2019).
267 Silva, A. *et al.* Performing mathematical operations with metamaterials. *Science* **343** (2014).

268 Lipworth, G. *et al.* Metamaterial apertures for coherent computational imaging on the physical layer. *J. Opt. Soc. Am. A* **30** (2013).
269 Reuther, A. *et al.*Survey of machine learning accelerators in *2020 IEEE High Performance Extreme Computing Conference (HPEC).* (2020).
270 Price, J. E. 'Programmable circuit', Patent No. US 3191151A, (1962).
271 Hartmann, R. F., Chan, Y.-F., Frankovich, R. & Ou, J.-H. 'Programmable logic array device using EPROM technology ', Patent No. US 4609986A, (1986).
272 Barr, K. *ASIC Design in the Silicon Sandbox: A Complete Guide to Building Mixed-Signal Integrated Circuits*. (McGraw Hill Professional, 2007).
273 Bogaerts, W. *et al.* Programmable photonic circuits. *Nature* **586**, 207–216 (2020).
274 Chen, X. *et al.* Towards an optical FPGA - Programmable silicon photonic circuits. *arXiv:1807.01656* (2018).
275 Paine, T., Jin, H., Yang, J., Lin, Z. & Huang, T.GPU asynchronous stochastic gradient descent to speed up neural network training in *Proceedings of the 2nd International Conference on Learning Representations (ICLR).* (2014).
276 Feldmann, J. *et al.* Parallel convolutional processing using an integrated photonic tensor core *Nature* **589**, 52-58 (2021).
277 Miscuglio, M. & Sorger, V. J. Photonic tensor cores for machine learning. *Appl. Phys. Lett.* **7** (2020).
278 Bohr, M. T.Interconnect scaling-the real limiter to high performance ULSI in *Proc. IEDM.* (1995).
279 Tait, A. N., Nahmias, M. A., Shastri, B. J. & Prucnal, P. R. Broadcast and weight: An integrated network for scalable photonic spike processing. *J. Light. Technol* **32**, 4029-4041 (2014).
280 Tait, A. N. *et al.* Neuromorphic photonic networks using silicon photonic weight banks. *Sci. Rep.* **7** (2017).
281 Psaltis, D., Brady, D., Gu, X.-G. & Lin, S. Holograpy in artificial neural networks *Nature* **343**, 325-330 (1990).
282 Li, H.-Y. S., Qiao, Y. & Psaltis, D. Optical network for real-time face recognition. *Appl. Opt.* **32**, 5026-5035 (1993).
283 Farhat, N. H., Psaltis, D., Prata, A. & Paek, E. Optical implementation of the Hopfield model *Appl. Opt.* **24**, 1469-1475 (1985).
284 Miller, D. A. B. Novel analog self-electrooptic-effect devices. *IEEE J. Quantum Electron* **29**, 678-698 (1993).
285 Shen, Y. *et al.* Deep learning with coherent nanophotonic circuits. *Nat. Photon.* **11**, 441–446 (2017).
286 Lin, X. *et al.* All-optical machine learning using diffractive deep neural networks. *Science* **361**, 1004–1008 (2018).
287 Bueno, J. *et al.* Reinforcement learning in a large-scale photonic recurrent neural network. *Optica* **5** (2018).
288 Saade, A. *et al.*Random projections through multiple optical scattering: Approximating Kernels at the speed of light in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).* (IEEE, 2016).
289 Buckley, S. *et al.* All-silicon light-emitting diodes waveguideintegrated with superconducting singlephoton detectors. *Appl. Phys. Lett.* **111**, 141101 (2017).
290 Kelley, H. J. Gradient theory of optimal flight paths. *Ars Journal* **30** (1960).
291 Dreyfus, S. The numerical solution of variational problems. *J. Math. Anal. Appl* **5**, 30-45 (1962).
292 Valiant, L. Three Problems in computer science. *J. ACM* **50** (2003).
293 Garcez, A. S. d. A., Lamb, L. C. & Gabbay, D. M. *Neural-symbolic cognitive reasoning*. (Springer, 2009).
294 Chen, B. *et al.* Smart factory of Industry 4.0: key technologies, application case, and challenges. *IEEE Access* **6**, 6505 - 6519 (2017).

295 Yin, Y., Stecke, K. E. & Li, D. The evolution of production systems from Industry 2.0 through Industry 4.0. *Int. J. Prod. Res.* **56** (2018).
296 Savery, T. 'Machinery for raising water, giving motion to mills, &e.', Patent No. No. 356, (1698).
297 Jenkins, R. *Savery, Newcomen and the Early History of the Steam Engine in The Collected Papers of Rhys Jenkins.*, 48–93 (Newcomen Society, 1936).
298 Watt, J. 'Steam engines &c. ', Patent No. 913, (1769).
299 The Lenoir Engine *Scientific American* 193 (1860).
300 Otto, N. A. 'Improvement in gas-motor engines', Patent No. US 194047A, (1876).
301 Crawford, K. *The atlas of AI* (Yale University Press, 2021).
302 Schreurs, B. G. Classical conditioning of model systems: A behavioral review. *Psychobiology* **17**, 145–155 (1989).
303 Lederhendler, I. I., Gart, S. & Alkon, D. Classical conditioning of Hermissenda: Origin of a new response. *J. Neurosci.* **6**, 1325-1331 (1986).
304 Commins, S. in *Behavioral neuroscience* Ch. 9, 100-117 (Cambridge University Press, 2018).
305 Cheng, D. K. *Field and wave electromagnetics* (Addison Wesley, 1979).
306 Snyder, A. W. & Love, J. D. *Optical waveguide theory*. (Chapman and Hall 1983).
307 Jackson, J. D. *Classical electrodynamics* (John Wiley & Sons, 1962).
308 Pozar, D. M. *Microwave Engineering*. (John Wiley & Sons 1998).
309 Moss, D. J., Morandotti, R., Gaeta, A. L. & Lipson, M. New CMOS-compatible platforms based on silicon nitride and Hydex for nonlinear optics. *Nat. Photon.* **7** (2013).
310 Bauer, J. Optical properties, band gap, and surface roughness of Si3N4. *Phys. Stat. Sol. A* **39** (1977).
311 Melati, D., Morichetti, F. & Melloni, A. A unified approach for radiative losses and backscattering in optical waveguides. *J. Opt.* **16**, 055502 (2014).
312 Stoffel, A., Kovacs, A., Kronast, W. & Muller, B. LPCVD against PECVD for micromechanical applications. *J. Micromech. Microeng.* **6** (1996).
313 Gessner, T., Otto, T., Wiemer, M. & Frömel, J. *Wafer bonding in micro mechanics and microelectronics - an overview*. (Ddp Goldenbogen, 2005).
314 Li, X. *et al.* Experimental investigation of silicon and silicon nitride platforms for phase-change photonic in-memory computing. *Optica* **7** (2020).
315 Yariv, A. *Optical electronics in modern communications* (Oxford University Press 1997).
316 Xie, Y. *et al.* Self-healing of a confined phase change memory device with a metallic surfactant layer. *Advanced Materials* **30**, 1705587 (2018).
317 Ziegler, M. *et al.* An electronic version of Pavlov's dog. *Adv. Function. Mater* **22**, 2744–2749 (2012).
318 Li, Y. *et al.* Associative learning with temporal contiguity in a memristive circuit for large-scale neuromorphic networks. *Adv. Electron. Mater* **1**, 1500125 (2015).
319 Moon, K. *et al.* Hardware implementation of associative memory characteristics with analogue-type resistiveswitching device. *Nanotechnol.* **25**, 495204 (2014).
320 Tan, Z.-H. *et al.* Pavlovian conditioning demonstrated with neuromorphic memristive devices. *Sci. Rep.* **7**, 713 (2017).
321 Hu, S. G. *et al.* Synaptic long-term potentiation realized in Pavlov's dog model based on a NiOx-based memristor. *J. Appl. Phys.* **116**, 214502 (2014).
322 Ha, S. D., Shi, J., Meroz, Y., Mahadevan, L. & Ramanathan, S. Neuromimetic circuits with synaptic devices based on strongly correlated electron systems. *Phys. Rev. App.* **2**, 064003 (2014).
323 Wu, C. *et al.* Mimicking classical conditioning based on a single flexible memristor. *Adv. Mat.* **29**, 1602890 (2017).
324 Bichler, O. *et al.* Pavlov's dog associative learning demonstrated on synaptic-like organic transistors. *Neural Comput.* **25**, 549-566 (2013).

325 Wan, C., Zhou, J., Shi, Y. & Wan, Q. Classical conditioning mimicked in junctionless IZO electric-double-layer thin-film transistors. *IEEE Electron. Dev. Lett.* **35** (2014).
326 Yu, F., Zhu, L. Q., Xiao, H., Gao, W. T. & Guo, Y. B. Restickable oxide neuromorphic transistors with spike-timing-dependent plasticity and Pavlovian associative learning activities. *Adv. Funct. Mater* **28**, 1804025 (2018).
327 Wang, Z. *et al.* Capacitive neural network with neuro-transistors. *Nat. Commun.* **9**, 3208 (2018).
328 Kumar, M., Abbas, S., Lee, J.-H. & Kim, J. Controllable digital resistive switching for artificial synapses and Pavlovian learning algorithm. *Nanoscale* **11** (2019).
329 Eckel, C., Lenz, J., Melianas, A., Salleo, A. & Weitz, R. T. Nanoscopic electrolyte-gated vertical organic transistors with low operation voltage and five orders of magnitude switching range for neuromorphic systems. *Nano Lett.* **22** (2022).
330 John, R. A. *et al.* Synergistic gating of electro-iono-photoactive 2D chalcogenide neuristors: Coexistence of Hebbian and homeostatic synaptic metaplasticity. *Adv. Mat.* **30**, 1800220 (2018).
331 Zhang, H. *et al.* Programming a Pavlovian-like conditioning circuit in Escherichia coli. *Nat. Commun.* **5**, 3102 (2014).
332 Soldano, L. B. & Pennings, E. C. M. Optical multi-mode interference devices based on self-imaging. *J. Light. Technol.* **13** (1995).
333 Besse, P. A., Bachmann, M., Melchior, H., Soldano, L. B. & Smit, M. K. Optical bandwidth and fabrication tolerances of multimode interference couplers. *J. Light. Technol* **12** (1994).
334 Darmawan, S., Lee, S.-Y., Lee, C.-W. & Chin, M.-K. A rigorous comparative analysis of directional couplers and multimode interferometers based on ridge waveguides. *IEEE J. Sel. Top. Quantum Electron* **11** (2005).
335 Haus, H. A. & Huang, W. Coupled-mode theory. *Proc. IEEE* **79**, 1505-1518 (1991).
336 Lu, Z. *et al.* Broadband silicon photonic directional coupler using asymmetric-waveguide based phase control. *Opt. Express* **23** (2015).
337 Cheng, Z., Ríos, C., Pernice, W. H. P., Wright, C. D. & Bhaskaran, H. On-chip photonic synapse. *Sci. Adv.* **3** (2017).
338 Guo, A. *et al.* Observation of PT-symmetry breaking in complex optical potentials. *Phys. Rev. Lett.* **103**, 093902 (2009).
339 Okamoto, K. *Fundamentals of optical waveguides*. (Academic Press, 2000).
340 Hunsperger, R. G., Yariv, A. & Lee, A. Parallel end-butt coupling for optical integrated circuits *Appl.Opt.* **16**, 1026-1032 (1977).
341 Burns, W. K. & Hocker, G. B. End fire coupling between optical fibers and diffused channel waveguides. *Appl. Opt.* **16**, 2048-2050 (1977).
342 Gehring, H., Eich, A., Schuck, C. & Pernice, W. H. P. Broadband out-of-plane coupling at visible wavelengths. *Opt. Lett.* **44** (2019).
343 Barwicz, T. & Taira, Y. Low-cost interfacing of fibers to nanophotonic waveguides: design for fabrication and assembly tolerances. *IEEE Photonics J.* **6** (2014).
344 Doany, F. E. *et al.* Multichannel high-bandwidth coupling of ultradense silicon photonic waveguide array to standard-pitch fiber array. *J. Light. Technol* **29** (2011).
345 Lindenmann, N. *et al.* Photonic wire bonding: a novel concept for chipscale interconnects. *Opt. Express* **20** (2012).
346 Dietrich, P.-I. *et al.* In situ 3D nanoprinting of free-form coupling elements for hybrid photonic integration. *Nat. Photon.* **12**, pages241–247 (2018).
347 Sun, P. & Reano, R. M. Cantilever couplers for intra-chip coupling to silicon photonic integrated circuits. *Opt. Express* **17** (2009).
348 Shoji, T., Tsuchizawa, T., Watanabe, T., Yamada, K. & Morita, H.Spot-size converter for low-loss coupling between 0.3-pm-square Si wire waveguides and single-mode fibers in *The 15th Annual Meeting of the IEEE Lasers and Electro-Optics Society.* (IEEE, 2002).

349 Pu, M., Liu, L., Ou, H., Yvind, K. & Hvam, J. M. Ultra-low-loss inverted taper coupler for silicon-on-insulator ridge waveguide. *Opt. Commun.* **283**, 3678–3682 (2010).
350 Park, H., Kim, S., Park, J., Joo, J. & Kim, G. A fiber-to-chip coupler based on Si/SiON cascaded tapers for Si photonic chips. *A fiber-to-chip coupler based on Si/SiON cascaded tapers for Si photonic chips* **Opt. Express**, 24 (2013).
351 Djogo, G. *et al.* Femtosecond laser additive and subtractive micro-processing: enabling a high-channeldensity silica interposer for multicore fibre to silicon-photonic packaging. *Int. J. Extrem. Manuf.* **1**, 045002 (2019).
352 Kasaya, K., Mitomi, O., Naganuma, M., Kondo, Y. & Noguchi, Y. A simple laterally tapered waveguide for low-loss coupling to single-mode fibers. *IEEE Photon. Technol. Lett* **5** (1993).
353 Shoji, T., Tsuchizawa, T., Watanabe, T., Yamada, K. & Morita, H. Low loss mode size converter from 0.3pm square Si wire waveguides to singlemode fibres. *Electron. Lett* **38** (2002).
354 Tsuchizawa, T. *et al.* Microphotonics devices based on silicon microfabrication technology. *IEEE J. Sel. Top. Quantum Electron* **11** (2005).
355 Bakir, B. B. *et al.* Low-loss (<1 dB) and polarization-insensitive edge fiber couplers fabricated on 200-mm silicon-on-insulator wafers. *IEEE Photon. Technol. Lett* **22** (2010).
356 Hatori, N. *et al.* Hybrid integrated light source on a silicon platform using a trident spot-size converter. *J. Light. Technol* **32** (2014).
357 Billah, M. R. *et al.* Hybrid integration of silicon photonics circuits and InP lasers by photonic wire bonding. *Optica* **5**, 876-883 (2018).
358 Paatzsch, T. *et al.* Very low-loss passive fiber-to-chip coupling with tapered fibers. *Appl. Opt.* **36** (1997).
359 Barclay, P. E., Srinivasan, K., Borselli, M. & Painter, O. Probing the dispersive and spatial properties of photonic crystal waveguides via highly efficient coupling from fiber tapers. *Appl. Phys. Lett.* **85** (2004).
360 Son, G. & Yu, K.High-efficiency power transfer for silicon-based photonic devices in *Proc. SPIE 10526, Physics and Simulation of Optoelectronic Devices XXVI.* (SPIE, 2018).
361 Love, J. D. & Henry, W. M. Quantifying loss minimization in single-mode fibre tapers. *Electron. Lett* **22** (1986).
362 Nelson, A. R. Coupling optical waveguides by tapers. *Appl. Opt.* **14** (1975).
363 Tamir, T. & Peng, S. T. Analysis and design of grating couplers *Appl. Phys.* **14**, 235-254 (1977).
364 Vivien, L. *et al.* Light injection in SOI microwaveguides using high-efficiency grating couplers. *J. Light. Technol* **24** (2006).
365 Vermeulen, D. *et al.* High-efficiency fiber-to-chip grating couplers realized using an advanced CMOS-compatible silicon-on-insulator platform. *Opt. Express* **18** (2010).
366 Benedikovic, D. *et al.* High-directionality fiber-chip grating coupler with interleaved trenches and subwavelength index-matching structure. *Opt. Lett* **40** (2015).
367 Fan, M., Popović, M. A. & Kärtner, F. X.High directivity, vertical fiber-to-chip coupler with anisotropically radiating grating teeth in *Conference on Lasers and Electro-Optics/Quantum Electronics and Laser Science Conference and Photonic Applications Systems Technologies.* (Optical Society of America, 2007).
368 Mekis, A. *et al.*A CMOS photonics platform for high-speed optical interconnects in *IEEE photonics conference 2012* (IEEE, 2012).
369 Taillaert, D., Bienstman, P. & Baets, R. Compact efficient broadband grating coupler for silicon-on-insulator waveguides. *Opt. Lett* **29** (2004).
370 Song, J. H. *et al.* Polarization-independent nonuniform grating couplers on silicon-on-insulator. *Opt. Lett* **40** (2015).
371 Zaoui, W. S., Kunze, A., Vogel, W. & Berroth, M. CMOS-compatible polarization splitting grating couplers with a backside metal mirror. *IEEE Photon. Technol. Lett* **25** (2013).
372 Taillaert, D. *et al.* A compact two-dimensional grating coupler used as a polarization splitter. *IEEE Photon. Technol. Lett* **15** (2003).

373 Harris, J. H. & Shubert, R.Optimum power transfer from a beam to a surface wave in *URSI Spr. Mtg.* (1969).
374 Iogansen, L. V. Theory of resonant electromagnetic systems with total internal reflection III. *Sov. Phys. Tech. Phys.* **11**, 1529–1534 (196).
375 Sarid, D., Cressman, P. J. & Holman, R. L. High-efficiency prism coupler for optical waveguides. *Appl. Phys. Lett.* **33** (1978).
376 Pasmooij, W. A., Mandersloot, P. A. & Smit, M. K. Prism-coupling of light into narrow planar optical waveguides. *J. Light. Technol* **7** (1989).
377 Ulrich, R. Optimum excitation of optical surface waves. *J. Opt. Soc. Am.* **61** (1971).
378 Cheben, P., Xu, D.-X., Janz, S. & Densmore, A. Subwavelength waveguide grating for mode conversion and light coupling in integrated optics. *Opt. Express* **14** (2006).
379 Tang, Y., Wang, Z., Wosinski, L., Westergren, U. & He, S. Highly efficient nonuniform grating coupler for silicon-on-insulator nanophotonic circuits. *Opt. Lett* **35** (2010).
380 Chen, X., Li, C., Fung, C. K. Y., Lo, S. M. G. & Tsang, H. K. Apodized waveguide grating couplers for efficient coupling to optical fibers *IEEE Photonics Technol. Lett.* **22** (2010).
381 Suh, W., Wang, Z. & Fan, S. Temporal coupled-mode theory and the presence of non-orthogonal modes in lossless multimode cavities. *IEEE J. Quantum Electron* **40** (2004).
382 Quan, Q., Deotare, P. B. & Loncar, M. Photonic crystal nanobeam cavity strongly coupled to the feeding waveguide. *Appl. Phys. Lett.* **96** (2010).
383 Klimov, N. N., Purdy, T. & Ahmed, Z. Fabrication and characterization of on-chip integrated silicon photonic Bragg grating and photonic crystal cavity thermometers. *arXiv:1508.01419* (2015).
384 Oraevsky, A. N. Whispering-gallery waves. *Quantum Electron.* **32** (2002).
385 Yi, X. *et al.* Multiple-Rayleigh-scatterer-induced mode splitting in a high-Q whispering-gallery-mode microresonator. *Phys. Rev. A* **83**, 023803 (2011).
386 Delfino, M. & Reifsteck, T. A. Laser activated flow of phosphosilicate glass in integrated circuit devices. *IEEE Electron. Dev. Lett.* **3** (1982).
387 Maine, S. *et al.*Design, fabrication and characterization of high Q-factor SOI microring resonators in *2007 4th IEEE International Conference on Group IV Photonics.* (IEEE, 2007).
388 Kazmierczak, A. *et al.*High quality optical microring resonators in Si3N4/SiO2 in *European Conference on Integrated Optics (ECIO).* (ECIO, 2008).
389 Grutter, K. *Optical whispering-gallery mode resonators for applications in optical communication and frequency control* PhD thesis, University of California, Berkeley, (2013).
390 Elshaari, A. W., Zadeh, I. E., Jöns, K. D. & Zwiller, V. Thermo-optic characterization of silicon nitride resonators for cryogenic photonic circuits. *IEEE Photonics J.* **8** (2016).
391 AZoM. 'Silicon nitride (Si3N4) properties and applications'. Retrieved from https://www.azom.com/article.aspx?ArticleID=53. Accessed on 17 June 2021.
392 Edwards, T. C. & Steer, M. B. *Foundations for Microstrip Circuit Design*. (Wiley Interscience, 1981).
393 Delan, A., Rennau, M., Schulz, S. E. & Gessner, T. Thermal conductivity of ultra low-k dielectrics. *Microelectronic Engineering* **70** (2003).
394 Autodesk. 'AutoCAD'. Retrieved from https://www.autodesk.com/products/autocad/overview. Accessed on 24 June 2021.
395 Gehring, H., Blaicher, M., Hartmann, W. & Pernice, W. 'gdsHelpers 1.2.0'. Retrieved from https://pypi.org/project/gdsHelpers/. Accessed on 24 June 2021.
396 Wold, S. & Tauber, R. *Silicon processing for the VLSI era: Volume 1 - Process technology*. (Lattice Press, 1986).
397 Abe, H., Nishioka, K., Tamura, S. & Nishimoto, A. Microfabrication of anti-reflective Chromium mask by gas plasma. *Jpn. J. Appl. Phys* **15** (1976).
398 Rahim, A. *et al.* Open-access silicon photonics platforms in Europe *IEEE J. Sel. Top. Quantum Electron* **25** (2019).

399 Boyd, R. W. *Nonlinear Optics*. (Elsevier Science, 2003).
400 Desurvire, E. *Erbium-doped fiber amplifiers: Principles and applications* (Wiley, 1994).
401 Desurvire, E., Sulhoff, J. W., Zyskind, J. L. & Simpson, J. R. Study of spectral dependence of gain saturation and effect of inhomogeneous broadening in erbium-doped aluminosilicate fiber amplifiers. *IEEE Photonics Technol. Lett.* **2** (1990).
402 Kanayama, K., Ando, Y., Nagase, R., Iwano, S. i. & Matsunaga, K. Advanced physical contact technology for optical connectors *IEEE Photon. Technol. Lett* **4** (1992).
403 Union, I. T. (International Telecommunication Union Geneva, Switzerland 2015).
404 Eickhoff, W. Stress-induced single-polarization single-mode fiber. *Opt. Lett* **7** (1982).
405 Birch, R. D., Varnham, M. P., Payne, D. N. & Okamoto, K. Fabrication of a stress-guiding optical fibre *Electron. Lett* **19** (1983).
406 Huang, H.-c. Fiber-optic analogs of bulk-optic wave plates. *Appl. Opt.* **36** (1997).
407 VanWiggeren, G. D. & Roy, R. Transmission of linearly polarized light through a single-mode fiber with random fluctuations of birefringence. *Appl. Opt.* **38** (1999).
408 Varnham, M. P., Payne, D. N., Birch, R. D. & Tarbox, E. J. Single-polarization operation in highly birefringent bow-tie optical fibers *Electron. Lett* **19** (1983).
409 Okamoto, K. Single-polarization operation in highly birefringent optical fibers. *Appl. Opt.* **23** (1984).
410 Pearsall, T. P. & R.W. Hopson, J. The Ga0.47In0.53As homojunction photodiode —A new avalanche photodetector in the near infrared between 1.0 and 1.6 µm. *J. Electron. Mat.* **7**, 133-146 (1978).
411 Graeme, J. G. *Photodiode Amplifiers: OP AMP Solutions*. (McGraw Hill 1995).
412 Hornbeck, L.Deformable-mirror spatial light modulators in *33rd Annual Technical Symposium.* (SPIE, 1989).
413 LinearTechnology. 'LTC2668: 16-Channel 16-/12-Bit ±10 V VOUT SoftSpan DACs with 10 ppm/°C Max Reference '. Retrieved from https://www.analog.com/media/en/technical-documentation/data-sheets/2668fa.pdf. Accessed on 29 July 2021.
414 Alfano, R. R. & Shapiro, S. L. Emission in the region 4000 to 7000 Å via four-photon coupling in glass. *Phys. Rev. Lett.* **24** (1970).
415 Alfano, R. R. & Shapiro, S. L. Observation of self-phase modulation and small-scale filaments in crystals and glasses. *Phys. Rev. Lett.* **24** (1970).
416 Kodama, Y. & Hasegawa, A. Nonlinear pulse propagation in a monomode dielectric guide. *IEEE J. Quantum Electron* **QE-23** (1987).
417 Blow, K. J. & Wood, D. Theoretical description of transient stimulated Raman scattering in optical fibers *IEEE J. Quantum Electron* **25** (1989).
418 Mamyshev, P. V. & Chernikov, S. V. Ultrashort-pulse propagation in optical fibers. *Opt. Lett* **15** (1990).
419 Dudley, J. M., Genty, G. & Coen, S. Supercontinuum generation in photonic crystal fiber. *Rev. Mod. Phys.* **78** (2006).
420 Koos, C., Jacome, L., Poulton, C., Leuthold, J. & Freude, W. Nonlinear silicon-on-insulator waveguides for all-optical signal processing. *Opt. Express* **15** (2007).
421 Avdokhin, A. V., Popov, S. V. & Taylor, J. R. Continuous-wave, high-power, Raman continuum generation in holey fibers. *Opt. Lett* **28** (2003).
422 Travers, J. C., Rulkov, A. B., Cumberland, B. A., Popov, S. V. & Taylor, J. R. Visible supercontinuum generation in photonic crystal fibers with a 400 W continuous wave fiber laser. *Opt. Express* **16** (2008).
423 NKT_Photonics. Supercontinuum generation in photonics crystal fibers. (Birkerød, Denmark, 2009).
424 Nishizawa, N. & Takayanagi, J. Octave spanning high-quality supercontinuum generation in all-fiber system. *J. Opt. Soc. Am. B* **24** (2007).

425 Heidt, A. M. Pulse preserving flat-top supercontinuum generation in all-normal dispersion photonic crystal fibers. *J. Opt. Soc. Am. B* **27** (2010).
426 Rampur, A. *et al.* Ultra low-noise coherent supercontinuum amplification and compression below 100 fs in an all-fiber polarization-maintaining thulium fiber amplifier. *Opt. Express* **27** (2019).
427 Genty, G., Lehtonen, M., Ludvigsen, H. & Kaivola, M. Enhanced bandwidth of supercontinuum generated in microstructured fibers. *Opt. Express* **12** (2004).
428 Chen, K. K. *et al.* Picosecond fiber MOPA pumped supercontinuum source with 39 W output power. *Opt. Express* **18** (2010).
429 Genty, G., Lehtonen, M., Ludvigsen, H., Broeng, J. & Kaivola, M. Spectral broadening of femtosecond pulses into continuum radiation in microstructured fibers. *Opt. Express* **10**, 1083-1098 (2002).
430 Hilligsøe, K. M. *et al.* Supercontinuum generation in a photonic crystal fiber with two zero dispersion wavelengths. *Opt. Express* **12** (2004).
431 Birks, T. A., Wadsworth, W. J. & Russell, P. S. J. Supercontinuum generation in tapered fibers. *Opt. Lett* **25** (2000).
432 Shi, J. *et al.* 1.06 µm picosecond pulsed, normal dispersion pumping for generating efficient broadband infrared supercontinuum in meter-length single-mode Tellurite holey fiber with high Raman gain coefficient. *J. Light. Technol* **29** (2011).
433 Horak, P., Chen, K. K., Alam, S.-u., Dasgupta, S. & Richardson, D. J.High-power supercontinuum generation with picosecond pulses in *2010 12th International Conference on Transparent Optical Networks.* (IEEE, 2010).
434 Yuksek, N. S. *et al.*Real time amplitude noise and jitter comparison of supercontinua generated at different dispersion regimes in *2008 Conference on Lasers and Electro-Optics and 2008 Conference on Quantum Electronics and Laser Science.* (IEEE, 2008).
435 SciPy.org. 'numpy.random.seed '. Retrieved from https://docs.scipy.org/doc/numpy-1.15.0/reference/generated/numpy.random.seed.html. Accessed on 5 May 2021.
436 Alkon, D. L., Blackwell, K. T., Barbour, G. S., Rigler, A. K. & Vogl, T. P. Pattern-recognition by an artificial network derived from biologic neuronal systems. *Biol. Cybern* **62**, 363-376 (1990).
437 Retrieved from https://course.fast.ai/datasets. Accessed on 27 August 2021.
438 Li Fei-Fei, R. F., Pietro Perona.Learning generative visual models from few training examples: An incremental Bayesian approach tested on 101 object categories in *2004 Conference on Computer Vision and Pattern Recognition Workshop* (IEEE, 2004).
439 Omkar M Parkhi, A. V., Andrew Zisserman, C. V. Jawahar.Cats and dogs in *IEEE Conference on Computer Vision and Pattern Recognition.* (IEEE, 2012).
440 Jia Deng, W. D., Richard Socher, Li-Jia Li, Kai Li, Li Fei-Fei.ImageNet: A large-scale hierarchical image database in *2009 IEEE Conference on Computer Vision and Pattern Recognition.* (IEEE, 2009).
441 VanElzakker, M. B., Dahlgren, M. K., Davis, C., Dubois, S. & Shin, L. M. From Pavlov to PTSD: The extinction of conditioned fear in rodents, humans, and anxiety disorders. *Neurobiol. Learn. Mem.* **113**, 3-18 (2014).
442 Stojanovic, V., Markovic, D., Nikolic, B., Horowitz, M. A. & Brodersen, R. W.Energy-delay tradeoffs in combinational logic using gate sizing and supply voltage optimization in *European Solid-state Devices and Circuits Conference (ESSCIRC) 2002 : Proceedings of the 28th European Solid-State Circuits Conference.* (IEEE, 2002).
443 Stanley, K. O. *Efficient evolution of neural networks through complexification* Ph. D thesis, University of Texas at Austin, (2004).
444 Goodfellow, I. J., Bengio, Y. & Courville, A. *Deep Learning*. (MIT Press, 2016).
445 Lillicrap, T. P., Cownden, D., Tweed, D. B. & Akerman, C. J. Random feedback weights support learning in deep neural networks. *arXiv:1411.0247* (2014).

446 Launay, J., Poli, I., Boniface, F. & Krzakala, F. Direct feedback alignment scales to modern deep learning tasks and architectures. *arXiv:2006.12878v2* (2020).
447 Launay, J. *et al.* Light-in-the-loop: using a photonics co-processor for scalable training of neural networks. *arXiv:2006.01475v2* (2020).
448 Neumaier, A. Solving ill-conditioned and singular linear systems: a tutorial on regularization. *SIAM Rev.* **40** (1998).
449 Ha, B. & Li, Y. Parallel modified signed-digit arithmetic using an optoelectronic shared content-addressable-memory processor. *Appl. Opt.* **33** (1994).
450 Krishnamoorthy, A. V., Marchand, P. J., Yayla, G. & Esener, S. C. Photonic content-addressable memory system that uses a parallel-readout optical disk. *Appl. Opt.* **34** (1995).
451 Alkabani, Y., Miscuglio, M., Sorger, V. J. & El-Ghazawi, T. OE-CAM: A hybrid opto-electronic content addressable memory. *arXiv:1912.02220v3* (2019).
452 Vagionas, C., Maniotis, P., Pitris, S., Miliou, A. & Pleros, N. Integrated optical content addressable memories (CAM) and optical random access memories (RAM) for ultra-fast address look-up operations. *Appl. Sci.* **7** (2017).
453 Chen, H. & Poon, A. W. Low-loss multimode-interference-based crossings for silicon wire waveguides. *IEEE Photon. Technol. Lett* **18** (2006).
454 Bachmann, M., Besse, P. A. & Melchior, H. General self-imaging properties in N x N multimode interference couplers including phase relations. *Appl. Opt.* **33** (1994).
455 Little, B. E., Chu, S. T., Haus, H. A., Foresi, J. & Laine, J.-P. Microring resonator channel dropping filters *J. Light. Technol.* **15** (1997).
456 Mehrabi, K. & Zarifkar, A. Ultracompact and broadband asymmetric directional-coupler-based mode division (de)multiplexer. *J. Opt. Soc. Am. B* **36** (2019).
457 Biggs, N. L., Lloyd, E. K. & Robin J. Wilson. *Graph Theory, 1736–1936*. (Clarendon Press, 1976).
458 Wolpert, D. H. *The supervised learning no free-lunch theorems*. (Springer, 2001).
459 Youngblood, N. *et al.* Tunable volatility of Ge2Sb2Te5 in integrated photonics. *Adv. Funct. Mater* **29**, 1807571 (2019).
460 Cheng, Z. *et al.* Antimony thin films demonstrate programmable optical nonlinearity. *Sci. Adv.* **7** (2021).
461 Feldmann, J. *et al.* Calculating with light using a chip-scale all-optical abacus. *Nat. Commun.* **8**, 1256 (2017).
462 Stern, K. *et al.* Uncovering phase change memory energy limits by sub-nanosecond probing of power dissipation dynamics. *Adv. Electron. Mater.* **7** (2021).
463 Oulton, R. F., Sorger, V. J., Genov, D. A., Pile, D. F. P. & Zhang, X. A hybrid plasmonic waveguide for subwavelength confinement and long-range propagation. *Nat. Photon.* **2** (2008).
464 Molesky, S. *et al.* Inverse design in nanophotonics. *Nat. Photon.* **12** (2018).
465 Goodman, J. W. Fan-in and fan-out with optical interconnections. *Optica Acta: Int. J. Opt.* **32** (1985).
466 Salinga, M. *et al.* Monoatomic phase change memory. *Nat. Mat.* **17**, 681-685 (2018).
467 Chang, Y.-Y. & Chou, L.-H. Laser-induced crystallization in AgInSbTe phase change optical disk. *Jpn. J. Appl. Phys.* **44** (2005).
468 Imada, M., Fujimori, A. & Tokura, Y. Metal-insulator transitions. *Rev. Mod. Phys.* **70** (1998).
469 Wu, K. *et al.* Phase transition across anisotropic NbS3 and direct gap semiconductor TiS3 at nominal Titanium alloying limit. *Adv. Mat.* **32** (2020).
470 Hou, W. *et al.* Strain-based room-temperature non-volatile MoTe2 ferroelectric phase change transistor. *Nat. Nanotechnol.* **14** (2019).
471 Santini, C. A. *et al.* Oxygenated amorphous carbon for resistive memory applications. *Nat. Commun.* **6** (2015).
472 GmbH, G. 'Beamer: Electron- and Laser-Beam Lithography Software'. Retrieved from https://www.genisys-gmbh.com/beamer.html. Accessed on 25 July 2021.

473 Wolpert, D. H. & Macready, W. G. No free lunch theorems for optimization. *IEEE Trans. Evol. Comput.* **1** (1997).
474 WikiWikiWeb. 'Symbolic processing'. Retrieved from https://wiki.c2.com/?SymbolicProcessing. Accessed on 27 August 2021.
475 King, R. D. *et al.* The automation of science. *Science* **324** (2009).
476 Paredes, R. *et al.* Collaborative update of a rule-based expert system for HIV-1 genotypic resistance test interpretation. *PLOS One* **12** (2017).
477 Bayardo, R. J., Agrawal, R. & Gunopulos, D. Constraint-based rule mining in large, dense databases. *Data Min. Knowl. Discov.* **4**, 217-240 (2000).
478 Floridi, L. & Chiriatti, M. GPT-3: Its nature, scope, limits, and consequences. *Minds Mach.* **30**, 681-694 (2020).
479 Martin, A. E. & McElree, B. A content-addressable pointer mechanism underlies comprehension of verb-phrase ellipsis. *J. Mem. Lang.* **58** (2007).
480 Blum, A. L. & Rivest, R. L. Training a 3-node neural network is NP-complete. *Neural Netw.* **5** (1992).
481 Ferrucci, D. A. Introduction to 'This is Watson'. *IBM J. Res. & Dev.* **56** (2012).

Publications, patents and conferences

Contributions directly related to this thesis are as follows:

Publications	• **Tan, J. Y. S.**, Cheng, Z., Li, X., Youngblood, N., Ali, U. E., Wright, C.D., Pernice, W. H.P., Bhaskaran, H. 'Monadic Pavlovian associative learning in a backpropagation-free photonic network' (under review). Preprint in https://arxiv.org/abs/2011.14709.	2020
Patent filings	• Bhaskaran, H., Feldmann, J. and **Tan, J. Y. S.** Photonic Associative Matrix. UK Patent Application No. 2109302.6.	2021
	• Bhaskaran, H., Cheng, Z. and **Tan, J. Y. S.** Associative Learning Element. International Patent Application No: PCT/GB2020/051358. Publication No: WO 2020/254781	2021
Conference/ workshop	• 'Associative learning on phase change photonics,' *SPIE Optics and Photonics, Optical Engineering + Applications*, San Diego, California, U.S.A. (presented).	2021
	• 'All-optical associative learning,' PhD Workshop on Next-Generation Cloud Infrastructure, Microsoft Research Cambridge lab, Cambridge, U.K. (presented).	2019

Contributions not directly related to this thesis are as follows:

Publications	• Farmakidis, N., Youngblood, N., Li, X., **Tan, J. Y. S.**, Swett, J.L., Cheng, Z., Wright, D., Pernice, W.H.P., Bhaskaran, H. 'Plasmonic nanogap enhanced phase-change devices with dual electrical-optical functionality' *Science Advances* 5 11 (2019). doi: 10.1126/sciadv.aaw2687.	2019
Conference/ workshop	• 'Phase change photonics for brain-inspired computing' *Micro & Nanotechnology Sensors, Systems and Applications*	2019