Design of Reconfigurable On-Chip Optical Architectures based on Phase Change Material

Parya Zolfaghari

A Thesis

In

The Department

Of

Electrical and Computer Engineering

Presented in Partial Fulfillment of the Requirements

For the Degree of

Master of Applied Science (Electrical and Computer Engineering) at

Concordia University

Montréal, Québec, Canada

November 2022

Concordia University

School of Graduate Studies

This is to certify that the thesis prepared

By :

Parya Zolfaghari

Entitled :

Design of Reconfigurable On-Chip Optical Architectures

Based on Phase Change Material

and submitted in partial fulfillment of the requirements for the degree of

Master of Applied Science (Electrical and Computer Engineering)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

	Chair/Examiner
Dr. Otmane Ait Mohamed	
	Examiner
Dr. Mohsen Gafouri (CIISE)	
	Thesis Supervisor
Dr. Sébastien Le Beux	

Approved by:

	Department Chair, Electrical and Computer Engineering
Dr. Yousef R. Shayan	
	Dean, Gina Cody School of Engineering
Dr. Mourad Debabbi	

Abstract

Design of Reconfigurable On-Chip Optical Architectures based on Phase Change Material

Parya Zolfaghari

Integrated optics is a promising technology to take the advantage of light propagation for high throughput chip-scale computing architectures and interconnects. Optical devices call for reconfigurable architectures to maximize resource utilization. Typical reconfigurable optical computing architectures involve micro-ring resonators for electro-optic modulation. However, such devices require voltage and thermal tuning to compensate for fabrication process variability and thermal sensitivity. To tackle this challenge we propose to use non-volatile Phase Change Material (PCM) to configure optical path. The non-volatility of PCM elements allows maintaining the optical path without consuming energy and the high contrast between two state of crystalline (cr) and amorphous (am) allows to route signal only through the required resonators, thus saving the calibration energy of bypassed resonators. We evaluate the efficiency of PCM based design on Reconfigurable Directed Logic (RDL) and nanophotonic interconnect. We develop a model allowing to estimate optical and electrical energy consumption. In the context of nanophotonic interconnect we evaluate the efficiency of the proposed PCM-based interconnects using system level simulations carried out with SNIPER manycore simulator. Results show that the proposed implementation allows reducing the static power by 53% on average for RDL and communication power saving up to 52% is achieved for nanophotonic interconnect.

To My Beloved Family

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Context and Motivation

Today's data intensive and computation intensive applications demand for high performance and low energy architectures. Conventional computing systems cannot fulfill the computation requirement of these applications which necessitates the emergence of more efficient computing paradigm. In the past, shrinking the size of integrated circuits and increasing the number of transistors was the most promising approach toward ever growing computation need. The approach led to the emergence of Very Large Scale Integration (VLSI) technology which allowed for accelerated computation [77]. However the technology faced with challenges involving high energy dissipation induced by transistor small size and low performance memories which could not kept pace with ever growing computation speed. To address these issues the solution was moving from single core processors with sequential execution to multi core processors with parallel execution which improved the performance of architectures[78]. Considering the large number of cores, memories and accelerators, network on chip naturally become the backbones of manycore architectures. Despite the advantages provided by manycore architectures they suffer from limitations induced by electronic interconnects. High loss, crosstalk, high energy dissipation and low bandwidth of the electrical interconnects present a bottleneck and as the density of interconnect is increased it is more difficult to achieve high performance computing. In such architectures it costs more money and energy to move data than to process it[40]. As a result to achieve energy efficient computing architectures new technologies such as photonic integrated circuits are required.

1.2 History of Optical Computing and Nanophotonic

Interconnect Architectures

Photonic integrated circuits allowing for on chip processing and interconnect was developed due to the factors involving compatibility with existing manufacturing process, integration capability, low manufacturing cost and scalability. The first demonstration of optical waveguides traces back to 1970 with realization of 2D and 3D waveguides. In 1990 the waveguides fabricated on Silicon on Insulator (SOI) platform demonstrated large propagation loss (~30dB/cm) [26]. The mature optical fiber technology in telecommunication industry paved the way for optical network on chip. Chip scale optical interconnect was the most promising candidate to overcome the limitation induced by electrical interconnects. Since 2000 there was a massive deployment of optics in chip scale interconnect. According to Beausoleil et al. [84] published in 2010, the communication bandwidth of optical interconnect per unit of dissipated power exceeds by factor of 20 from communication bandwidth of electrical interconnects. Nowadays improved fabrication technology allows to realize waveguides with loss as low as 0.1 dB/cm[26]. This shows the potential of optics in achieving high speed, low latency and high bandwidth communication between chips in manycore architectures.

While the efficiency of optics in on chip communication has been proved, there were many attempts to take the advantages of light in processing. The first use of optics in processing involves free space optical processing and dates back to 1953 with the employment of lens to obtain Fourier transform of light[24]. These free space optical computing architectures did not develop due to the bulky optical devices. Later on, the development of photonic integrated circuits and the emergence of high speed optical devices such as modulators, lasers and photodetectors provided the ground for the use of integrated optics in processing[26]. The technology of photonic integration allows to realize high speed low latency chip scale optical processors. In 2012 Directed Logic (DL) architecture dedicated to the simultaneous execution of AND and NAND was reported[29]. In DL rings are organized as an array of optical switches to control light propagation. Later, on 2014, Reconfigurable Directed Logic which allows to use one architecture for multiple applications was developed[8]. As the technology continues to mature, emerging optical computing architectures are developed to accelerate neural networks

applications [80] and microwave processing [81]. The design of optical circuits dedicated to matrix multiplications, logic functions [82] and adders [83] are also investigated. The emergence of disruptive material such as phase change material allows to realize non-volatile photonic integrated circuits. PCM based spiking neural network reported in 2018[10] and PCM based optical switch realized in 2019 [11] are among the latest achievements in photonic integrated circuits.

1.3 Micro Ring Resonator

Micro Ring Resonators (MRR) is one of the most important components in optical architectures. Compared to Mach–Zehnder Interferometer (MZI), rings are smaller and can achieve narrowband filtering and modulation. Most optical processing architectures and nanophotonic interconnects rely on rings to modulate or receive signal. For instance, in the context of RDL [8] MRRs are used as switch to control light propagation and in the context of interconnects based on WDM they are used to modulate and filter out the signal. A feature shared by such architectures is high static power consumption, which is mostly due to losses experienced by optical signals and devices calibration requirements. Indeed, optical devices are sensitive to manufacturing process and thermal variations, which call for constant calibration of resonating devices such as ring resonators. While high contrast can be achieved, the method requires voltage and thermal tuning to calibrate the rings, which leads to static power consumption overhead. Disruptive materials and architectures are thus needed to overcome the low energy efficiency of optical devices calibration.

1.4 Phase Change Material

Phase Change Material (PCM) has been widely studied to design non-volatile photonic circuits such as neural networks [10]. Indeed, the non-volatility of PCM based devices allows maintaining the configuration of optical device without consuming energy. Typical configurations involve amorphous (*am*) and crystalline (*cr*) states, which can be obtained by heating the device [56]. These configurations are highly distinctive in optical and electrical properties which provide the ground for PCM utilization in different applications. Among recently demonstrated PCM based devices, a Directional Coupler (DC) reported in [11] leads to

bar and cross under *cr* and *am* states respectively. The low attenuation and the associated high optical contrasts allow envisioning new optical architectures involving reconfigurable optical paths.

1.5 Problem Statement

Most optical architectures rely on the use of rings to carry out modulation and filtering. To insure their accurate performance constant calibration is required which leads to static power consumption, area overhead induced by control systems and latency. Therefore the question is to find a way to avoid the calibration of rings which are not used. For this purpose we focus on using PCM.

1- Can we use non-volatility property of PCM devices to bypass unused rings?
2- Can we define a generic PCM based architecture in order to bypass ring which can be used in both optical computing architectures and nanophotonic interconnects?
3- Can we define design methods allowing to configure PCM according to given application or connectivity requirement?

1.6 Contribution

In this work we propose an optical architecture allowing to bypass unused optical devices. To achieve this, PCM-based directional couplers are placed before and after rings, thus allowing either to transmit optical signals to devices for modulation (filtering) purpose or to bypass them. The use of the bypass path allows avoiding calibration of the optical devices, thus leading to significant reduction in the static power consumption. We investigate the efficiency of the proposed design on the Reconfigurable Directed Logic (RDL) architecture and nanophotonic interconnects. The architectures are configured according to the application mapping and the model allows to estimate laser power overhead as well as ring power saving. To explore design space we consider different configurations such as application mapping on different number of clusters in the context of optical interconnect.

1.6.1 Generic Architecture

In order to adopt the proposed generic design in the context of computing and interconnect we derive different implementations of it. These implementations allow to bypass one ring or multiple rings per set of DC or to bypass several groups of rings. The implementations are defined as: i) Single Ring-Single Group (SRSG), ii) Multiple Rings-Single Group (MRSG), iii) Single Rings-Multiple Groups (SRMG) and iv) Multiple Rings-Multiple Groups (MRMG). Single ring per set of DC achieves reconfiguration at a scale of one ring. Implementation involving multiple rings allows bypassing groups of rings and is used for applications based on WDM such as architectures with multiple receivers and transmitters.

1.6.2 RDL

We propose PCM based non-volatile RDL involving the use of SRMG in its architecture. To implement a function on the architecture we first configure it based on the function requirement and we only reconfigure the architecture when new function is mapped. Results show that our proposed non-volatile RDL can achieve 19% power saving on average and for functions involving the bypassing of maximum number of rings %35 saving is obtained. We also propose another implementation of non-volatile RDL with different interface. This implementation uses coupler instead of rings to merge signals and leads to 53% power saving on average. One key challenge of architecture is the slow state change of PCM. To obtain realistic evaluation of architecture efficiency taking this limitation into account, we estimate power consumption according to reconfiguration frequency. Based on results architecture is power efficient up to 14MHz.

1.6.3 Nanophotonic Interconnect

We proposed PCM based nanophotonic interconnect composed of Single Writer Multiple Reader (SWMR) channels. We use MGMR implementation of the proposed architecture in designing the channel. We develop a method to configure the interconnect according to the mapping of applications on the cores. To do this, SNIPER simulator is modified to enable the thread distribution of multi-applications from Splash2 and PARSEC benchmarks. A key limitation of PCM elements for Optical Network on Chip (ONoC) is the slow phase state changes compared to the required nanosecond scale latency communication requirements in manycore. We tackle this challenge by partitioning the manycore to execute different applications and by

reconfiguring the interconnect only when new applications are executed. We show that an average of 21% power saving is obtained compared to channels without PCM, and up 52% saving is reached for mapping involving 4 clusters. Finally, we simulate the execution of the applications in parallel by considering the proposed reconfigurable interconnect. This allows us to divide the architecture to execute two applications independently and simultaneously. On average, we obtain a 21.6% reduction in the execution time.

1.7 Thesis Organization

The thesis is organized as following: In chapter 2 we present optical computing and nanophotonic interconnects, then we introduce the state of art that have investigated the use of PCM in different architectures involving switches, memories, processors and neuromorphic computing.

In chapter 3 we present our proposed design to bypass unused MRR then we introduce four implementations of it that can be utilized for different applications and finally we introduce two use cases for PCM based DC involving RDL and nanophotonic interconnect.

In chapter 4 we develop a model which allows us to evaluate the power consumption in PCM based architectures. We consider PCM induced loss to obtain laser power and we also take into account the number of bypassed rings to achieve ring calibration saving.

Chapter 5 presents the results for each use case based on the mapped application. For RDL we study the power consumption for each function and compare it with baseline which is the architecture without PCM. For nanophotonic interconnect we evaluate the power based on application mapping on different numbers of clusters. We also study power saving according to the reconfiguration frequency of PCMs.

In chapter 6 we conclude the work and present future work directions.

Chapter 2

Related Work

Photonic Integrated Circuits (PIC) involve the use of heterogeneous devices allowing to take advantage of light to accelerate on-chip data transmission and computation. Among the demonstrated platforms for PIC, Silicon Photonic (SP) has developed due to the compatibility with CMOS manufacturing process. As a result, emerging optical computing architectures and chip scale nanophotonic interconnects are developed. While the former allows the realization of neural network applications and microwave processing, the latter delivers the bandwidth required by data intensive applications. However, these architectures suffer from high static power consumption induced by inefficient lasers and calibration requirement of optical devices. This calls for non-volatile materials to overcome the high static energy. In this section we first introduce the optical devices, then we present optical processing and nanophotonic interconnects and finally we introduce Phase Change Material (PCM) and its applications.

2.1 Optical Devices

In following we introduce devices mostly used in nanophotonic architectures.

2.1.1 Waveguide

Optical signal propagates trough waveguides, which are used to connect devices. It provides a good confinement of propagated light. One of the most important challenges in designing waveguides is to minimize loss. Waveguides with loss as low as 0.11 dB/cm has been reported in [1] which allow designing passive PICs involving splitters, filters and combiners.

2.1.2 Micro Ring Resonator

Micro ring resonators (MRR) are one of the key components of the SP. It is an optical waveguide with round shape which allows the coupling of optical signal if its resonance wavelength matches the signal wavelength. The coupling occurs when optical path is an integer coefficient

of wavelength. Therefore, a MRR shows multiple resonances and the distance between these resonances is known as Free Spectral Range (FSR). MRR can be coupled to one or two waveguides defined as all-pass ring and pass/drop rings respectively. Figure 2.1 represents both types of rings. In all-pass ring the output transmission spectra shows dips corresponding to the wavelengths where coupling to ring occurs. In pass/drop ring light is coupled to two waveguides. Dips in transmission spectra of pass port occur due the coupling to the ring which results in peaks in the transmission spectra of drop port [2]. MRR finds application as filters, modulators, sensors and laser cavity. MRR tuned to a resonant wavelength of signal are used to filter the signal in optical interconnects based on WDM. It is also used as buffers and delay lines in PICs [3]. One of the main applications involves its use as modulators. For this purpose, rings are tuned and their resonant wavelengths are changed through applying electronic signals which leads either to transmission or blocking of optical power. The electronic signal is applied thorough embedded pin junctions which work in carrier injection or depletion mode. The modulation of free carriers leads to the change in refractive index of the medium and results in resonant wavelength shift [4]. MRRs are highly sensitive to any variation involving temperature, shape, size which provides the possibility of using them as sensor [5][6]. However, this property requires them to be calibrated constantly to work accurately which leads to high static power consumption. MRRs are used in designing directed logics (DL) [7] or reconfigurable DLs [8]. These architectures allow the implementation of functions such as XOR through modulating the optical signal. An optical look up table relying on MRR is reported in [9] in which WDM leads to the implementation of multiple functions. MRRs deposited with PCM finds application in neural network [10] and switching [11].
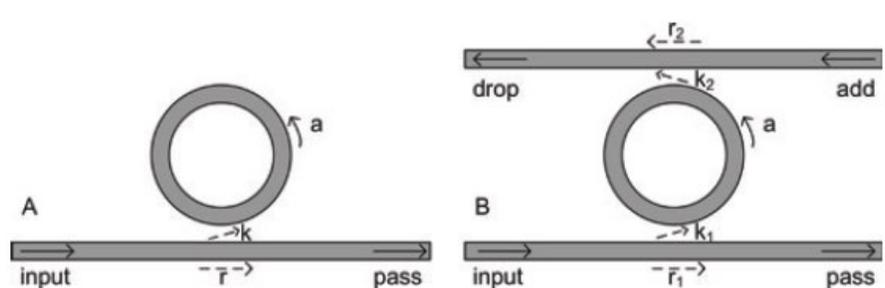
Figure 2.1: a) All-pass ring, b) Add/drop ring

2.1.3 Directional Coupler

Directional Coupler typically consists of two parallel waveguides, which are close enough to allow signal coupling. The coupling can be controlled either at design time (e.g. by changing the distance between the waveguides or the waveguide length) or at run-time (e.g. by changing the reflecting index of one of the waveguide). Numerous variations of coupler exist; for instance, a directional coupler enabling the implementation of optical pass-gate (OPG) has been designed using photonic crystal in [12]. From the pass-gate, an optical full adder and subsequently an optical multiplier demonstrating a x3 speedup over a CMOS Wallace tree multiplier were reported in [13].

2.1.4 Mach-Zehnder Interferometer

Mach-Zehnder Interferometer (MZI) is commonly used as modulator and switch. It is a key component of the photonic multipurpose processor proposed in [14] and all optical neural networks [15]. In the latter, matrix multiplication is implemented using reconfigurable MZI. In [16], up to 56 MZI were used to implement the optical interference unit of NN, which led to x10 speed up compared to electronic devices. The application of MZI as switch has been reported in [17] which achieves switching with High extinction ratio and fast response time.

2.1.5 Photodetector

Photodetectors are used to convert optical signals into electronic signals. Received signal in photodetector is amplified by transimpedance amplifier (TIA) and finally is distinguished as '0' or '1' by a comparator [18]. The efficiency of photodetector depends on numerous parameters such as i) operating frequency (Gb/s) [19], ii) dark current (nA), iii) Bit Error Rate (BER) and iv) responsivity [20][21].

2.1.6 Laser

Laser input power depends on different technological parameters involving i) lasing efficiency ii) loss induced by optical devices, iii) photodetector sensitivity, and IV) crosstalk. At the system level, taking into account the physical characteristics and fabrication process related parameters in laser's model is not suitable due to the long simulation time it would involve. Instead, laser model considered for system level design include the lasing efficiency. It corresponds to the ratio of output optical power to input electrical power [22]. Photodetector sensitivity is the produced

output electrical power per input optical power [23]. Crosstalk noise is the result of coupling light between channels or between devices. This calls for efficient lasers to compensate for the noise and ensure the required signal to noise ratio at output [23].

2.2 Optical Computing

The idea of taking the advantages of high speed and parallelism of light triggered the development of optical processing. Early optical processors consisted of three planes of input, processing and output. Spatial light modulators were used as input plane to convert signal to optics [24]. Liquid crystals drove the research in this field which was later replaced by micro electromechanical mirrors. They provided the possibility of modulating arrays of light. Attempts in exploiting light to process information dates back to 1940 with the invention of holography. Later on the invention of laser made it possible to obtain the 3D holograms of fast moving objects [25]. The Fourier transforms property of lens led to applications like spatial filtering and optical correlation. To realize optical computing an optical memory with parallel access was needed. The works in this field led to the invention of holographic memory and refractive crystal with angular multiplexing [24]. These free space optical computing architectures did not develop due to the bulk devices like mirrors and lens which restrict the scaling of free space applications. On the other hand PIC have attracted attention due to many factors involving the compatibility with existing manufacturing process, integration capability, low manufacturing cost and scalability [26]. Emerging computing paradigm based on PIC can be categorized as digital and analogue. While the former involves the implementation of logic functions, adder and multiplier the latter includes the matrix multiplication and neural network implementation.
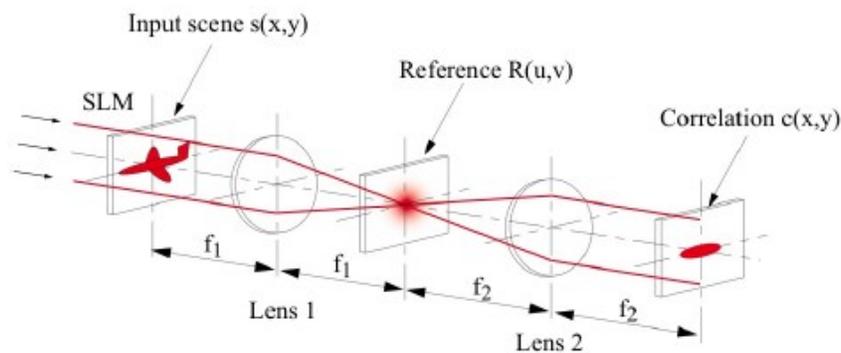
Figure 2.2: Basic optical correlator

10

2.2.1 Digital Architectures

Numerous optical accelerators have been designed to execute both arithmetic and logic operation. They involve key optical devices such as MRRs, micro-disks, photonic crystal cavities and waveguides. A common objective is to reduce the critical path delay, which can be obtained by simultaneously applying multiple electro-optic modulation on optical signals propagating along a waveguide. By doing so, an 8-bit ripple carry adder with a 20ps critical path delay has been demonstrated in [27]. The same approach has been used in [28] for the design of an n-bit multiplier. One of the emerging optical computing architectures is Directed logic (DL). The key device of such architectures is MRR. In DL the rings are organized as an array of optical switches to control light propagation. Directed logic architectures have been proposed to efficiently utilize optical devices by simultaneously executing AND and NAND [29], the outputs being available on through port and drop port of a ring resonator. The approach has then been extended to XOR and XNOR operations [30]. A key issue with the above-mentioned architectures is the limited number of operations that can be executed, which is solved by the Reconfigurable Directed Logic (RDL) [8]. Reconfigurable optical architectures allow to efficiently use bulky, optical devices for multiple operation, thus allowing to reduce the cost overhead induced by the technology. The RDL involves parallel waveguides on which modulators are serially placed, thus allowing to map sum-of-product functions. A feature shared by such architecture is the need to calibrate ring resonators in order to control optical signal transmissions. To do so, the architecture relies on modes (named pass/pass, pass/block, block/pass and block/block) which are configured by calibrating the modulator using thermal tuning. Hence, the main drawback of the architecture is the need to constantly thermally tuning ring resonators, even if no modulation is carried out, which is power consuming. To realize large scale Boolean functions, logic gate synthesis methods have been investigated. [31] Proposes to use virtual gate (VG) as the building block of the logic synthesis. VG is a crossbar switch which provides two paths defined as bar and cross for signal propagation based on the input electronic signal. The VGs can be combined to obtain larger functions. The drawback of this binary decision diagram (BDD) based design is large number of splitters which attenuate signal to large extent. Another optical implementation of BDD based synthesis has been proposed in [32], in which the low power optical output problem is solved through removing splitters and replacing them with combiners. It prevents the attenuation of signal through its split. Although combiner

prevents the splitting of signal, however splitters and combiners attenuate signal by 3dB leading to a low power signal at output. [33] Proposes to use directional coupler by appropriately assigning the coupling ratio and to remove the combiner. This way 3dB loss imposed by splitters and combiners are avoided.

2.2.2 Analogue Architectures

One of the emerging optical computing paradigms is Optical neural network (ONN). The analog computation of NN can use light to accelerate computation. Artificial neural network (ANN) based on electronics has drawbacks which has motivated the emerging optical computing paradigm. High density connection is required to implement NN which leads to the interference of electronic signals with each other. It also demands high energy which limits the real time processing of such networks. Digital computers handle the computation with sequential clock rates which results in long processing time. Such drawbacks have created attention toward using light in ANN. A fast improvement in machine learning has created a need for architectures with high speed low latency processing capabilities. Recently demonstration of ONN achieved 10 trillion of operation per second [35]. The progress in the field has led to the invention of application specific photonic neural networks. The first photonic integrated silicon neural network was introduced in 2017 which was approved to solve differential equation efficiently [36]. Later, a programmable nanophotonic processor (PNP) with MZI as processing unit was introduced as ANN. The architecture involves phase shifter, optical splitter and beam attenuator. While architecture is able to implement any size of input, directional couplers and phase modulators limits the size of the architecture to be scaled for more than 1000 neurons [37]. In 2019 PCM based optical neural network was introduced in which the weight is adjusted by setting the state of the PCM [38]. The proposed architecture is able to handle the image processing with high accuracy. The field of ONN is evolving every day; however there are challenges such as appropriate nonlinear optical materials that should be addressed to efficiently exploit the proficiency of light in realization of ONN.

2.3 Optical Network on Chip

Data intensive and computation intensive applications demand for high speed processing capabilities which call for fast, low power processing units. According to the Moore's law the

number of transistors in an integrated circuit double every 2 years. As the size of transistors shrink high speed computing chip can be achieved which leads to increased power consumption. On the other hand, shrinking the size of transistor is limited by quantum tunneling effect which can degrade its performance. This motivates the move from one bigger core to multiple smaller cores with lower power consumption. Applications demanding high speed computation and technology scaling has called for integration of large number of cores in a single chip [39]. This requires high speed communication among cores which cannot be fulfilled by conventional electrical interconnects. Capacitive and inductive coupling, parasitic resistance, interconnect noise and propagation delay are among the associated problems with electrical interconnects which prevents their scaling. Long distance electrical interconnects require repeaters to obtain a good quality signal at outputs [40]. These problems motivate the move toward emerging technologies such as SP. The compatibility of CMOS manufacturing process with silicon photonics allow to achieve low cost large scale manufacturing capacity [41]. The technology allows for the integration of other materials such as silicon nitride and III-V compound semiconductors which can enhance the performance of interconnects. Silicon photonic can fulfill the high speed, high bandwidth and low latency requirement of optical channels [26]. WDM allows transmitting different wavelength at the same time in a channel, hence increasing the bandwidth of the network [42].

Figure 2.3 shows an optical link in Optical Network on Chip (ONoC). At transmitter side the optical signal is modulated through MRRs which are tuned to resonate at the same wavelength of optical signal. Data is first serialized, then the driver applies signal to the MRRs leading to match or mismatch of their resonant wavelength with that of optical signal. This leads to modulation of 0 or 1 which is known as on/off keying (OOK) modulation. At receiver side each wavelength is filtered out through an MRR which is calibrated to that wavelength. The filtered signal is first converted to electrical signal by a photodetector and then is passed to transimpedance amplifier (TIA) and finally is distinguished as 0 or 1 by a comparator [18].
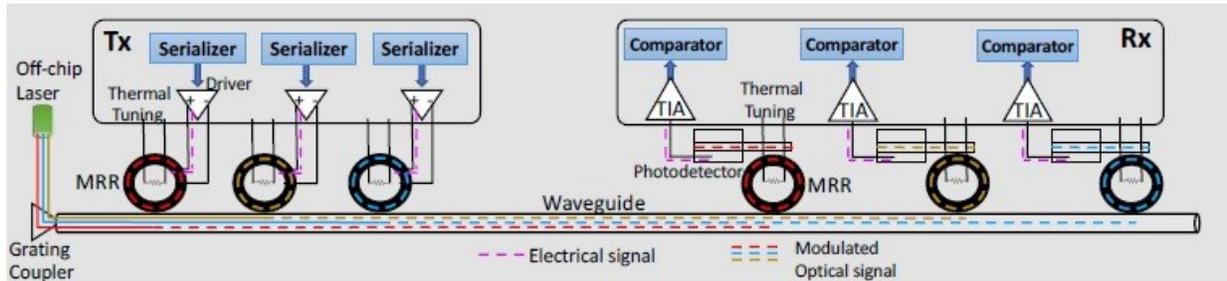
Figure 2.3: An example of optical channel

2.3.1 Optical Buses

There are different configuration of optical links allowing for sharing one link between writers and readers. In following we introduce these configurations.

- **Single Writer Single Reader (SWSR):** SWSR provides an optical link through which one sender communicates with one receiver as shown in Figure 2.4.a.

- **Multiple Writer Single Reader (MWSR):** One optical link is shared by all writers which send WDM signals to one reader. Crossbar architectures based on MWSR buses involves one waveguide per reader [4].

- **Single Writer Multiple Reader:** SWMR link allows the transmission of data from one sender to all readers which allows for implementation of broadcast network. However in order to feed the photodetectors of all readers laser with sufficient power is required [4]. Reservation assisted SWMR (R-SWMR) addresses this issue by supplementing SWMR with extra low bandwidth channel used to send the reservation packet. Therefore, before communication starts the reservation packet which involves only the destination address and number of bits, is sent. This allows for other readers to detune their MRRs and only the receiver tunes its MRRs. Supplementing SWMR with extra channel imposes latency and area overhead however it leads to laser power saving [43].

- **Multiple Writers Multiple Readers (MWMR)**: In MWMR all senders and receivers are connected to one waveguide. To avoid the corruption of data due to the simultaneous transmission of it by writers, arbitration is performed. This link provides low power consumption due to the shared number of wavelengths but it imposes latency [43].

14
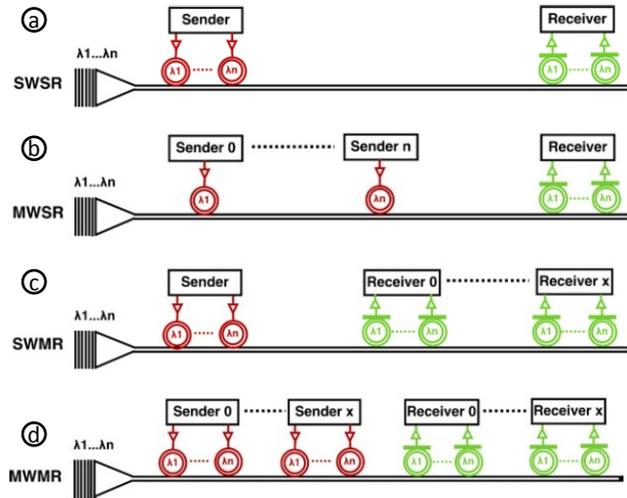
Figure 2.4: a) Single writer single reader channel, b) Multiple writer single reader channel, c) Single writer multiple reader channel, d) Multiple writer multiple reader channel

2.3.2 Classification of ONoC

The design approaches targeting efficient networks for inter chip communications can be classified as All-optical Network on Chip (NoC) based on Wavelength Routed Optical NoC (WRONoC) and Hybrid NoC [4].

2.3.2.1 Wavelength Routed Optical NoCs

WRONoC routs optical signals according to their wavelengths. The network is an optical crossbar which implies an all to all communication. The crossbar can be implemented using optical buses such as SWSR, SWMR and MWSR. However, this leads to area overhead which is impractical as the number of nodes are increased. To tackle with this issue, efficient topologies relying on MRR switches are proposed. To name a few of these topologies we can refer to snake [45], λ-router [46] and folded-crossbar [45] as shown in Figure 2.5 [4].
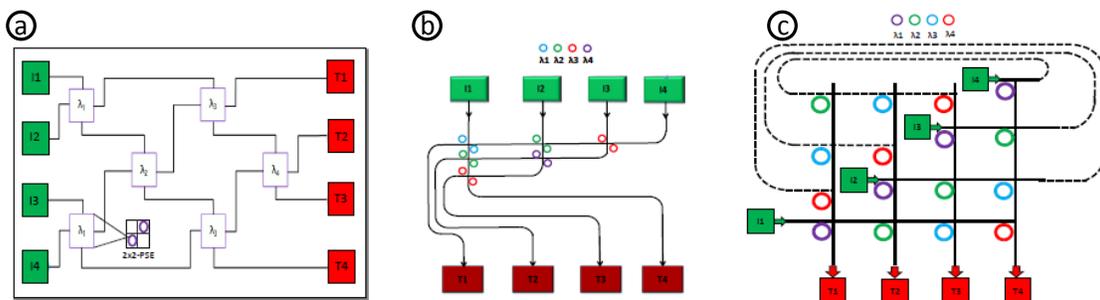
Figure 2.5: a) 4x4 Lambda-Router , b) 4x4 Snake, c) 4x4 folded crossbar

2.3.2.2 Hybrid NOC

Both electrical and optical links are used to implement hybrid NoC. This approach leads to electrical connection among cores in short distance and optical connection in long distance. **Firefly** is a hybrid network consisting of 64 cores, 16 of which are grouped as cluster. Intra cluster communication is implemented through 2D electrical link. Each cluster involves 4 routers which are used to communicate with routers of other clusters. Every router in a cluster has a counterpart in other clusters which implement the optical communication among the cores. This is referred as optical assembly and forms a crossbar topology. Firefly uses R-SWMR to implement the optical link [47].

2.3.3 Design Challenges

The attempts in addressing high power consumption of optical networks can be classified as resonating device calibration and laser power management approaches [4].

2.3.3.1 Resonating Device Management

MRR as a main building block of optical channels are very narrow band with high extinction ratio allowing for high precision modulation and filtering. However, they are highly sensitive to process imperfection and thermal variation. These environmental and process uncertainties lead to resonant wavelength shift of this device which in turn degrade its performance and consequently reduces the signal to noise (SNR) ratio. Circuit level approaches focuses on reducing the impact of environmental variation through calibrating MRR. For this purpose, output signal is monitored and based on that controller applies required signal to compensate for the variation. The tuning of MRRs is classified based on the type of feedback signal, tuning method and controller design. The feedback signal can be direct or indirect. The direct feedback signal is the optical power detected by a photodetector placed in the through or drop port of a switch. Temperature gradient is an indirect feedback signal which can be measured by a sensor. By appropriately modeling the channel it gives an estimation of output optical power. 0/1 statistics is another indirect feedback signal. This is due to the fact that degradation of MRR affects the optical power which alters the 0/1 statistics. The most used techniques for MRR tuning are classified as thermal tuning and electrical tuning. Electrical tuning works based on injecting or depleting carries from the medium causing the change of refractive index and consequently shifting the resonant wavelength. An advantage of electrical tuning is being fast;

however, it can only be used to compensate for low temperature gradients. Thermal tuning is the most used technique. It works based on the joule heating or resistive heating in which the current passes through resistor and produces heat. Although the process is slow it can fulfill the tuning requirement of high temperature gradients. Among the methods used for control we can refer to Finite-State-Machine (FSM), proportional-integral-derivative (PID) control and model predictive control (MPC) [48].
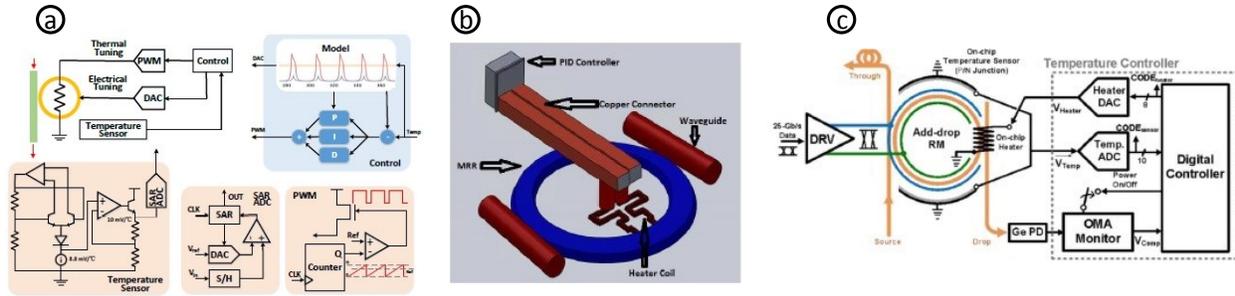
Figure 2.6: a) MRR tuning based on indirect feedback signal (temperature) utilizing MPC [48], Tuning of MRR with heater using PID controller [49], c) MRR tuning based on temperature singnal [50].

2.3.3.2 Laser Power Management

High power consumption of ONoC calls for efficient lasers to alleviate the issue. In addition to efficient laser development methods, there are techniques used to manage laser power in efficient way. In following we present related techniques**.**

- **Approximate Computing:** Approximate computing is an emerging method which relies on relaxing the accuracy of data representation at an acceptable level for error tolerable applications. The technique allows to improve the energy efficiency and execution time of applications. Floating point numbers are good candidate for approximation through which the quality of the output is not affected. Works utilizing approximate computing consider different laser power level for MSB and LSB. While Most Significant Bits (MSB) are sent with high laser power, Least Significant Bits (LSB) are sent with low laser power [51][52].

- **PROWAVES: Proactive Runtime Wavelength Selection for Energy-Efficient Photonic NoCs:** Bandwidth requirement of an application during execution is variable. Therefore, selecting fixed number of wavelength for execution can increase the power consumption. To tackle with this issue, an online wavelength selection based on bandwidth need is employed. For this purpose, the total execution of application is divided into intervals and for each interval the average latency of packets is extracted. The proposed method improves the

power consumption of ONoC by only activating the required lasers for each interval. Due to the online prediction strategy of active lasers the execution time is affected. Results show 18% and 33% power saving against 1% and 5% performance loss [18].

- **Suor: Sectioned Undirectional Optical Ring for Chip Multiprocessor:** Suor is a ring waveguide with bidirectional data transmission capability which is divided into non-overlapping sections, each of which can support communication independently. Each node is assigned an agent and the communication between node and agent is implemented optically. The transceiver involves the use of two MRRs defined as bridge ring and switch ring. They are used to provide the connection between channel and transceiver and to change the direction of light. Sour shows improvement in power consumption compared with MWMR and MWSR channels [53].
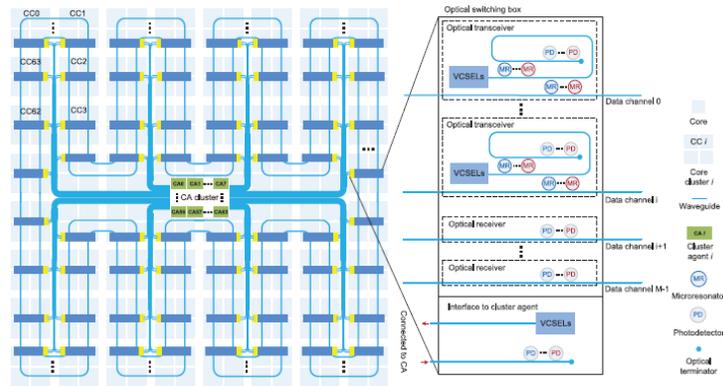
Figure 2.7: Suor architecture

- **Crosstalk Aware Channel Optimization:** Crosstalk is an intrinsic property of optical channels. It happens when two waveguides cross therefore signal propagating in one of them induce crosstalk noise to the other one. Channels based on WDM are highly vulnerable to crosstalk noise. Distribution of wavelengths inside a channel results in small spacing between them resulting in crosstalk noise. In order to exploit the high bandwidth property of WDM channels it is necessary to tackle with the problem. Optimal mapping of application in a multi core system can reduce the impact induced by crosstalk noise. Figure 2.8.a shows an example of task mapping. In b mapping of application on cores results in two waveguides crossing which induces crosstalk noise, however for the same task graph optimal mapping is represented in c which avoids crossing [54].

18

- **CHAMELEON (CHANNEL Efficient Optical Network-on-Chip):** CHAMELEON allows to runtime implementation of multiple point to point communication at the same time without waveguide crossing. Each wavelength can be reused by optical network interface (ONI) to realize another communication. The ONIs can adopt the number of wavelengths based on bandwidth requirement allowing for turning of the laser when low bandwidth is required. The channel also allows for both clockwise and counter-clockwise communications. This leads to energy efficient and scalable network [55].
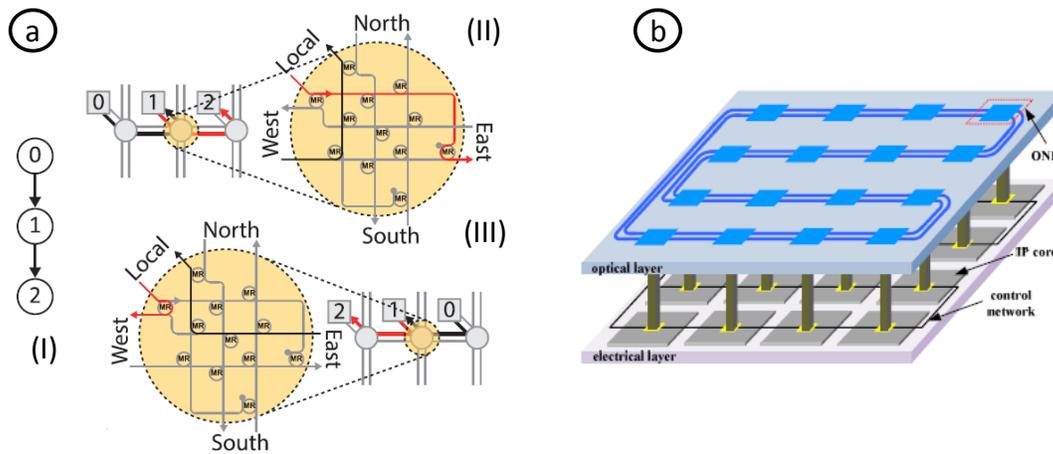
Figure 2.8: a) Impact of different mapping on crosstalk, I) Communication of an application, II) One mapping solution leading to crosstalk noise, III) Different mapping solution without induced crosstalk noise, b) CHAMELEON is implemented on optical layer

2.4 Phase Chang Material (PCM)

Mach-zehnder interferometer and micro ring resonator are the key optical devices to realize switching and modulation. Mach-zehnder interferometer offers a wide range of modulation and switching bandwidth and micro ring resonator achieves narrow band modulation with high extinction ratio. The modulation in these devices is achieved through plasma dispersion, thermo optic and electro optic effects. Optical resonating devices are highly sensitive to process imperfection and temperature gradient of environment which calls for constant calibration to insure accurate performance [18]. The calibration requirement of these devices leads to increased static power consumption which limits the scalability of PICs. To obtain efficient scalable PICs the above mentioned switches should have low static and dynamic power consumption and achieve fast and high contrast switching. Therefore, non-volatile optical devices with low static power consumption are needed to overcome the low energy efficiency. The use of PCM in

photonic platforms has been widely studied in recent years. Indeed, non-volatility (zero static power consumption), sub-nanosecond phase transition, femtojoule-scale phase transition energy consumption, 10^{15} switching cycle endurance and years long state retention have provided the ground for the massive deployment of PCM in numerous applications. (Ge)-antimony (Sb)-telluride (Te) (or shortly GST) is a well-known PCM material which has the abovementioned characteristics. The two phases of PCM are known as amorphous (*am*) and crystalline (*cr*) which have distinctive optical and electrical properties and lay the foundation for its deployment in different applications. To achieve *cr* state long and moderate heating energy is required which leads to the transition of material to its glass state. *Am* state is obtained by applying high energy short pulses to melt the material and then cool it fast. Phase transition of PCM is obtained by thermal annealing using external heaters, optical pulses (photothermal effect) or electrical pulses (electrothermal effect) [56][57].

Thermal annealing achieves phase state transition through using oven or rapid thermal processing (RTP) system. This is a slow process which limits its application in fast PICs and only provides one-way switching (from *am* to *cr*), which is due to the fact that re-amorphization cannot be handled by heaters. Therefore, in order to achieve fast and reversible state changes of PCM optical and electrical approaches are considered. Using optical signals to change the state of PCM is defined as photothermal approach. The process can be perfumed using in plane or out plane laser beams. The out-plane optical excitation is a slow process which suffers from low efficiency due to the diffraction of light beams. It also limits the implementation of fully optical integrated circuits. For on-chip optical excitation the transmission of optical signal to the PCM and state change of large material are challenging. This calls for other methods which can target large PCM integrated circuits. We can refer to electro-thermal and mixed electrical/optical methods as a good substitute to above mentioned methods[56].

PCM have been deployed in many areas involving implementation of optical switches and modulators, optical memories, optical processing units and optical neural network. In following we introduce works that have investigated the use of PCM in each area.

2.4.1 PCM based Optical Switches

Optical switches and modulators are the key components of optical processing and communication. PCM based optical switches can be implemented using different configurations.

In following we introduce configurations and the control used to achieve the state change of the PCM.

2.4.1.1   PCM based Resonator and Bus Waveguides

This configuration consists of a MRR or racetrack resonator covered with a layer of PCM and coupled to, i) one bus waveguide as shown in Figure 2.9.a and Figure 2.9.b or ii) two bus waveguides with through and drop outputs as represented in Figure 2.9.c and Figure 2.9.d. *Am* state leads to the coupling of signal to resonator. This implies no signal transmission to output for one bus waveguide configuration. For two bus waveguide configuration drop output is achieved. Upon the phase transition of PCM to *cr*, refractive index and extinction coefficient changes. The change in the refractive index results in the red shift of resonance wavelength which detunes resonator from signal wavelength. As a result, no coupling occurs. This implies that optical signal transmits to the output of former configuration and through port of latter one [57].
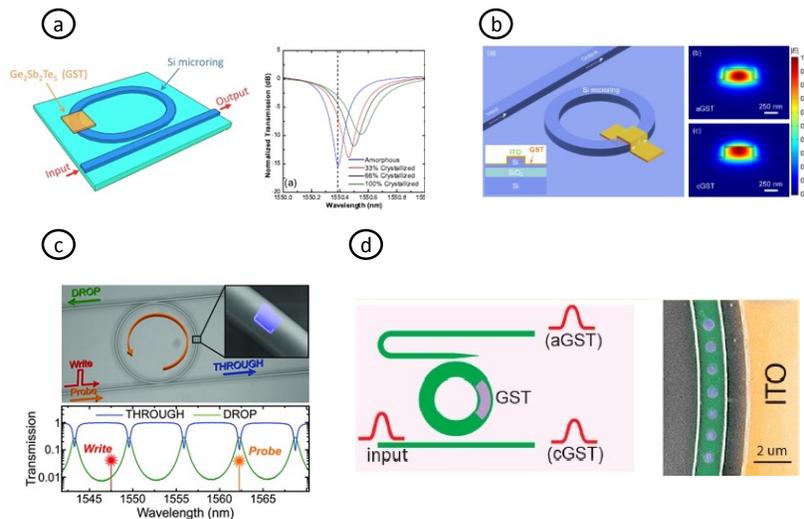
Figure 2.9: a, b) PCM based resonator coupled to a bus waveguide, state transition is achieved utilizing out-plane optical signal[58][57], c,d) PCM based resonator coupled to two bus waveguide [59][60], c) State transition is achieved with optical pulses, d) Crystallization is achieved using optical pulses, amorphization is obtained electrically

2.4.1.2   Waveguide Covered with PCM

This configuration of switch relies on the integrating of PCM on the surface of bus waveguide. *Am* state is low loss which allows the transmission of large ratio of signal to the output. State

conversion to *cr*, red shifts the imaginary part of refractive index which in turn increases the absorption. As a result large ratio of signal is absorbed which leads to '0' at output[64].
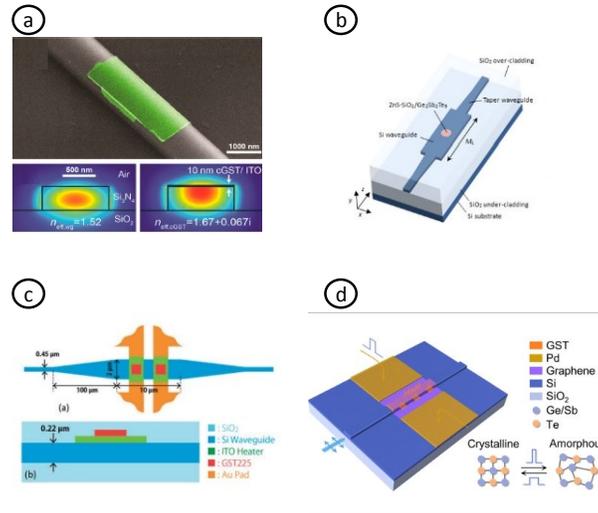
Figure 2.10: Optical switch incorporating PCM on top of bus waveguide a, b) Optical pulses are used to change GST state[62][61], c, d) Electrically driven heaters change GST state[63][64].

2.4.1.3   PCM based Directional Coupler

This configuration of PCM based optical switch employs directional coupler. While resonator based devices allows for narrow band modulation and switching, coupler based design leads to broadband switching. Figure 2.11 shows a PCM based directional coupler (DC) which consists of a Si waveguide and hybrid waveguide (PCM deposited on the waveguide). This configuration implements 1x2 and 2x2 switches. In this case, a PCM deposited waveguide is inserted between the branches of two waveguides. *Am* leads to the phase match between the modes in Silicon waveguide and hybrid waveguide, thus coupling occurs which results in cross transmission. Rapid thermal annealing is performed in 200 ˚C for 10 minutes. The obtained *cr* state results in the phase mismatch between the modes in the waveguides which prevents the coupling and results in bar transmission[65].
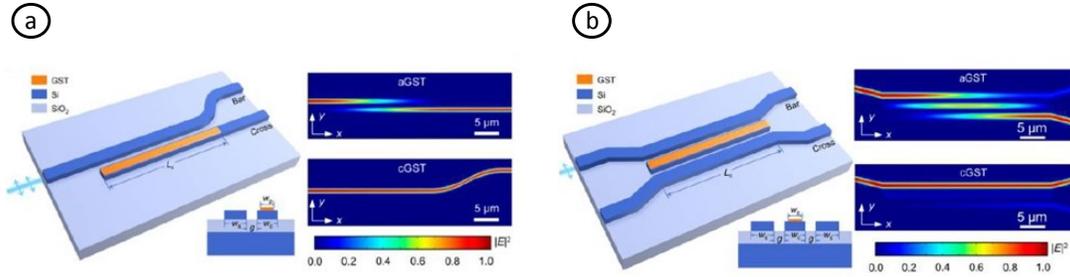
Figure 2.11: PCM based directional coupler , a) 1x2 optical switch, b) 2x2 optical switch

2.4.2 PCM based Memory

The realization of photonic memories leads to, i) replacement of high latency electronic memories with photonic ones, ii) removing the need for optoelectronic conversion and iii) the possibility of the simultaneous realization of processing and storage. Writing and erasing (phase state transition of PCM) of memory is achieved through high power optical signal. Low power optical signals are used to readout the memory. By adjusting the power of applied signal to the PCM, intermediate states are achieved which allows to realize multi-level memories. The architectures implementing PCM based memories can be classified as following: i) resonator covered with PCM and coupled to one or two bus waveguides as represented in Figure 2.12.a-b. The approach allows to implement multi wavelength access memory as represented in Figure 2.12.c and ii) bus waveguide partially covered with PCM which allows to achieve multi-level memories by controlling the degree of crystallization as represented in Figure 2.12.d[66].
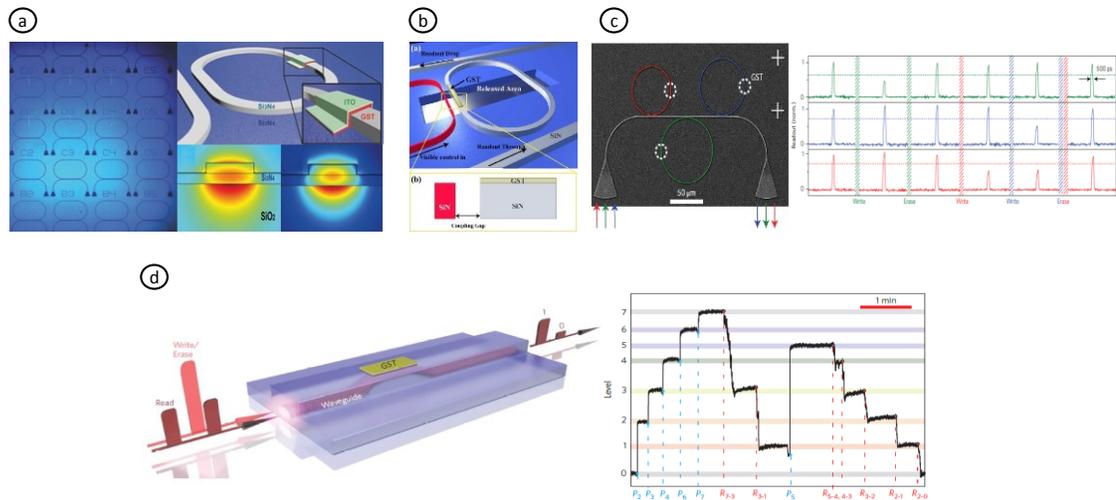
Figure 2.12: a) Optical memory composed of Si_3N_4 racetrack resonator covered with GST and coupled to a bus waveguide[66], b) Optical memory based on a recetrack resonator covered with PCM and coupled to two parallel waveguides. The state of the GST on the resonator is controlled using red waveguide[67]. c) Multi wavelength access memory based on wavelength division multiplexing[68], d) Optical memory composed of bus waveguide partially covered with PCM[68]

23

2.4.3 PCM based Processing:

The intermediate level (not fully *cr*, not fully *am*) provides the ground for PCM utilization in arithmetic processing. The approaches rely on i) encoding numbers with the crystallization level and the mathematical operations target changing the phase state of PCM as represented in Figure 2.13.a-b and ii) encoding one number in the transmission of PCM based waveguide and encoding the other number in the power of input signal as shown in Figure 2.13.c[69][70].
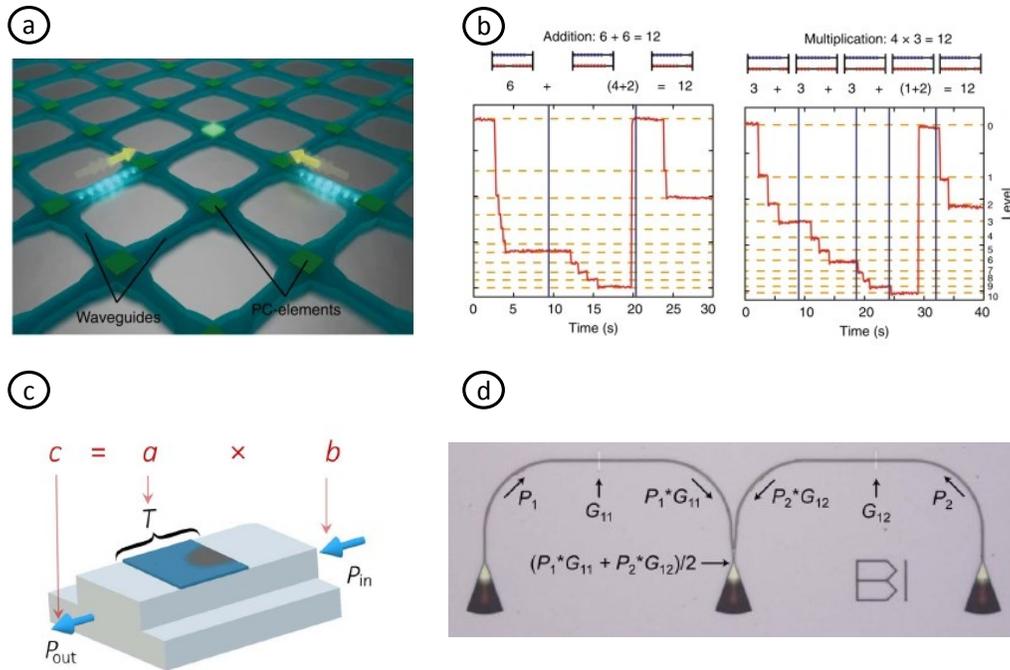
Figure 2.13: a) PCM based optical processor implementing sum, multiplication, subtraction and division[69], b) Implementation of 6+6 and 4x3, c) PCM based optical multiplier[70], d) Matrix vector multiplication based on PCM based optical multiplier

2.4.4 PCM based Neuromorphic Computing

Brain inspired neuromorphic computing can benefit data intensive applications in efficiency and speed. The paradigm has found application in image processing and deep learning. PCM allows realizing all optical neuromorphic computing. A PCM based spiking neural network capable of implementing supervised and unsupervised learning is represented in Figure 2.14. To realize synapses, waveguide covered with PCM is utilized. The degree of crystallization (amorphization) allows to control the ratio of transmitted signal, representing weights in spiking neural network. Transmitted signal from each waveguide is coupled to a MRR and is propagated through upper straight waveguide. The accumulated signal from all waveguides constitutes the

post-synaptic spike which is fed to a PCM integrated on the top of waveguide crossing. The MRR is tuned to be resonant when PCM is in *cr*. When the input signal (post-synaptic spike) lacks sufficient power, the state of PCM remains unchanged. This results in the coupling of probe signal to the MRR and being absorbed by PCM. Therefore, no signal is transmitted to the output. When the input signal has sufficient power, the state of PCM is changed into *am*. This shifts the resonant wavelength of MRR which causes the probe signal to be off resonance. Therefore coupling is not achieved and the probe signal is directed to the output [71].
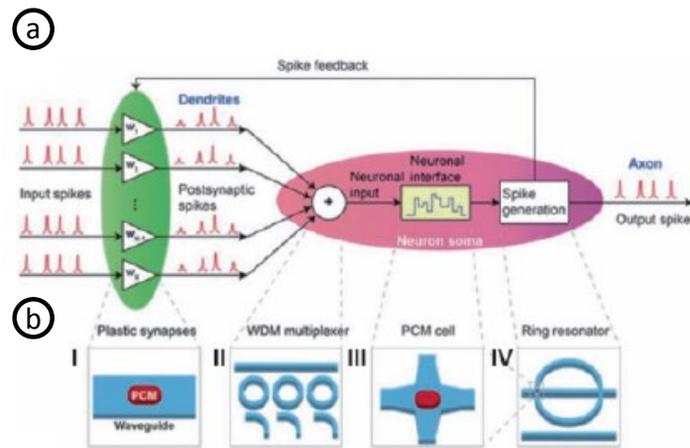
Figure 2.14: a) PCM based neurosynaptic system, b) Realization of synaptic weights using PCM covered waveguide (I), WDM multiplexer (II), Combined signals are transmitted to a PCM cell placed on top of ring resonator (III, IV)

2.5 Summary

MRRs are key components for the design of optical architectures based on WDM. The architectures rely on arrays of rings to carry out modulation and filtering. These rings require calibration even if they are not used. This leads to high static power consumption which calls for the use of disruptive materials. PCM based directional coupler is a non-volatile optical switch which provides reconfigurable optical path. We propose to use this device to rout signal through rings if modulation or filtering is carried out or to bypass them if they are not used. To the best of our knowledge this has never been investigated so far. In next section we will present our proposed design in more details and then we will introduce its use cases in the context of optical computing and nanophotonic interconnect.

Chapter 3

PCM Based RDL and Nanophotonic Interconnect

Chip scale optical computing and interconnect devices call for reconfigurable architectures to maximize resource utilization. These architectures allow to use optical devices for multiple purposes, thus allowing to reduce the cost overhead induced by the technology. Typical optical architectures involve the use of micro ring resonator which achieves narrow band modulation and filtering with high extinction coefficient. However, it is highly sensitive to any process and temperature variation. This calls for power hungry calibration which induces significant static power overhead, thus limiting the scalability of optical architectures. To tackle this challenge disruptive materials and architectures are required to overcome the low energy efficiency of optical devices. Phase Change Material (PCM) is non-volatile material allowing to maintain the configuration of optical device without consuming energy. Typical configuration involves crystalline (*cr*) and amorphous (*am*) which are highly distinctive in optical properties allowing to obtain reconfigurable optical paths.

In this thesis, we propose to use non-volatile PCM elements to route optical signals only through the required resonators, hence saving calibration energy of bypassed resonators. We investigate the efficiency of proposed design on RDL and nanophotonic interconnect. In following we first present our proposed design allowing to bypass unused devices then we introduce its use cases for RDL and nanophotonic interconnect.

3.1 Bypass of Unused Resonating Devices

To address the static power consumption of optical architectures induced by calibration requirement of resonating devices, we propose to bypass unused micro ring resonators. For this purpose we use PCM based directional coupler (DC) which is demonstrated in [65]. The device

involves the use of hybrid waveguide (PCM deposited on top of waveguide) between the branches of two parallel waveguides. The DC is configured according to the state of the embedded PCM element as shown in Figure 3.1. Based on the state of PCM two optical paths defined as bar and cross are obtained. *Am* and *cr* lead to Cross and bar respectively. The PCM is non-volatile, thus allowing maintaining the DC configuration without consuming static energy.
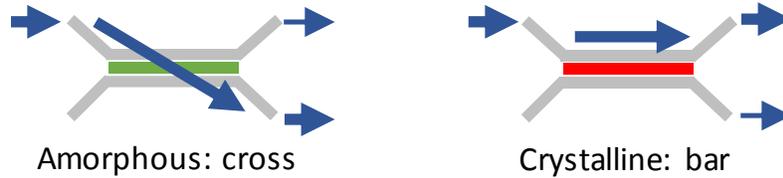
Figure 3.1: PCM based directional coupler

Our proposed cell is composed of two DCs placed before and after two parallel waveguides. At one of the waveguides one or groups of resonating devices such as MRR are placed as represented in Figure 3.2. The MRR carries out the filtering or modulation of signal. Each DC allows configuring the optical path to transmit signal to the resonating device or to the bypass. Therefore, two signal path defined as modulation (filtering) and bypass are obtained. In the modulation (filtering) path, the optical signal propagates through resonating devices where modulation (filtering) is carried out. In bypass path, the optical signal directly propagates toward DC_2. The couplers are large band, thus allowing transmitting all modulated signals to the same output waveguide. The state of the PCM in DC is electrically configured using a dedicated control signal.
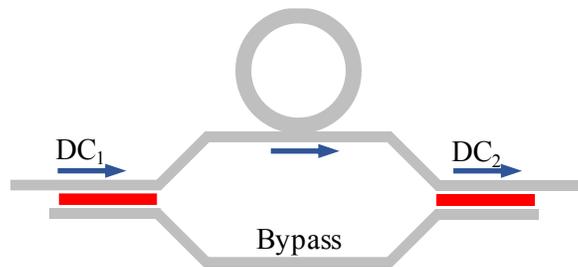
Figure 3.2: Proposed cell based on micro ring resonator and phase change directional coupler

Different implementations involving numbers of MRRs per cell or numbers of cascaded cells are considered which allows realizing various applications. In following we introduce these architectures and their working principle.

27

Single Ring-Single Group (SRSG): This architecture involves the use of one micro ring in the modulation (filtering) path as represented in Figure 3.2. This is utilized for application requiring reconfiguration at a scale of one ring. Therefore, one MRR per set of DCs are used. DC_1 configured to *cr* and *am* allows the transmission of signal to the modulation and bypass paths respectively.

Multiple Rings-Single Group (MRSG): The architecture involves the use of multiple rings in the modulation (filtering) path. This allows to bypass groups of MRRs leading to saving the calibration power of them. The proposed configuration can be used for applications based on WDM, thus requiring configuration at a scale of groups of MRRs.
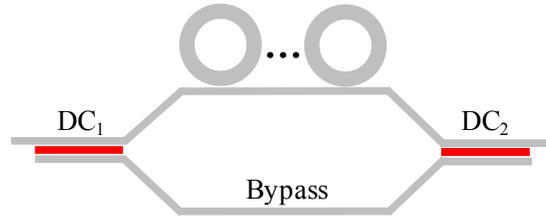
Figure 3.3: Multiple Rings-Single Group (MRSG)

Single Ring-Multiple Groups (SRMG): The architecture is realized by cascading two cells involving one MRR per set of DC as represented in Figure 3.4. To reduce the design complexity, DC_2 from the first cell is merged with DC_1 from the second cell. To connect the MRR in second cell DC_2 is configured to *cr* if MRR in first cell is connected (signal coming from modulation path) or to *am* if the MRR in first DC is disconnected. The transmitted signal through the architecture is modulated by one or two MRRs.
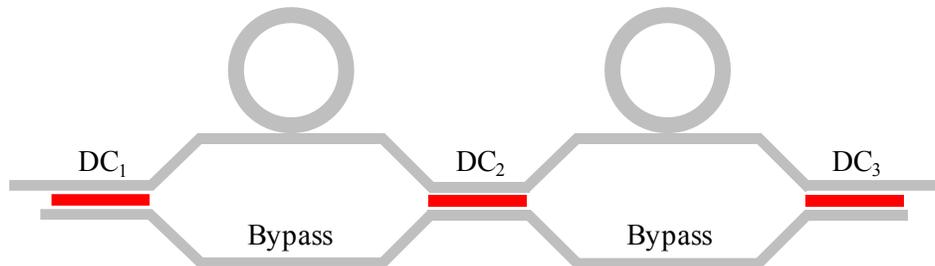
Figure 3.4: Single Ring-Multiple Groups (SRMG)

Multiple Rings-Multiple Groups (MRMG): It corresponds to the most general use case scenario. This architecture is achieved through cascading one cell involving multiple rings per

set of DC. Similarly to SRMG, DCs are merged. This allows to implement architectures in which WDM signals are modulated or received by groups of writers or readers.
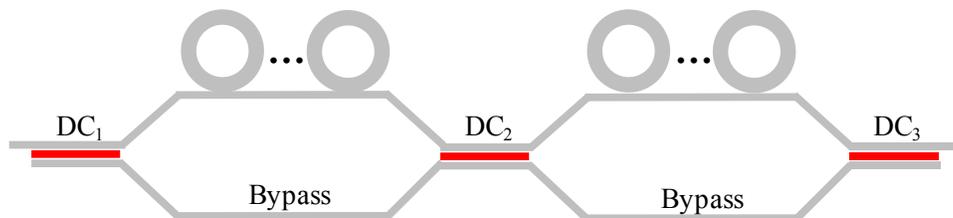
Figure 3.5: Multiple Rings-Multiple Groups (MRMG)

PCM allows to avoid tuning unused devices thus resulting in calibration power reduction. However, it induces loss on optical signal resulting in its attenuation which leads to laser power overhead. Therefore, taking into account the power overhead of PCM is required to have an actual assumption of total power saving. We investigate PCM based optical architectures considering two use case scenarios involving RDL and nanophotonic interconnect. In RDL reconfiguration is at the scale of one ring per set of DC which demands the use of SRMG architecture. PCM based nanophotonic based on WDM relies on groups of MRRs to realize transmitters and receivers. This is achieved through using MRMG architecture. Power consumption in PCM based non-volatile RDL and nanophotonic interconnect is obtained taking into account technological parameters involving laser efficiency, number of bypassed rings and number of crossed PCMs.

3.2 PCM based Realization of Logic Functions

In this section, we study the application of the proposed cell in RDL. As mentioned earlier, RDL involves the use of reconfigurable rings to realize different logic functions. The reconfiguration in RDL is achieved at the scale of each ring; therefore, to implement a non-volatile RDL, we use SRSG configuration. The resulting cell used is represented in Figure 3.6. Calibration is achieved through using heater and electrical signal (Data) realizes the modulation of optical signal. In the modulation path, data is modulated on optical signal, before reaching DC_2. In the bypass path, the optical signal directly propagates towards DC_2. Depending on the state of DC_2, signals are transmitted either to the output of the cell or to a terminator.
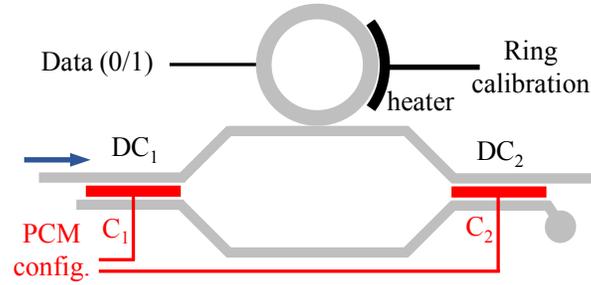
Figure 3.6: RDL architecture based on SRSG architecture

3.2.1 Cell Configuration

The cell is configured according to i) the state of the PCM elements in the DCs and ii) the tuning of the ring. By combining the states of the PCM and ring tuning, the following cell configurations are defined:

- Pass/Pass: Both DC_1 and DC_2 are in the *am* state as shown in Figure 3.7.a. The input signal propagates through the bypass path and is transmitted to the output. Since the signal does not propagate through modulation path, no thermal calibration of the ring is needed.

- Block/Block: Figure 3.7.b represents the block/block mode. Similarly to pass/pass mode, the signal propagates through the bypass path since DC_1 is set to the *am* state. However, instead of transmitting the signal to the output, DC_2 is configured to the *cr* state, which leads to a transmission of the signal to the terminator. Hence, the optical signal is strongly attenuated on the output.

- Pass/block: The input signal is transmitted to the modulation path, which is achieved with DC_1 is configured in the *cr* state as shown in Figure 3.7.c. The signal is first modulated by the input data and is then transmitted to the output (DC in *cr* state). Since a modulation occurs, the ring is thermally calibrated to the signal wavelength (λ_s). Therefore, data input '0' leads to the coupling of the signal, which results in a strong attenuation, while data input '1' detunes the resonance of the ring, which leads to a high transmission of the signal.

- Block/Pass: Similarly, to Pass/block, the signal propagates through the modulation path, as illustrated in Figure 3.7.d. However, the ring is tuned to $\lambda_s - \Delta\lambda$, i.e. the ring is off

signal resonance for data input '0'. Data input '1' leads to a red shift of the ring and hence a strong attenuation of the optical signal.
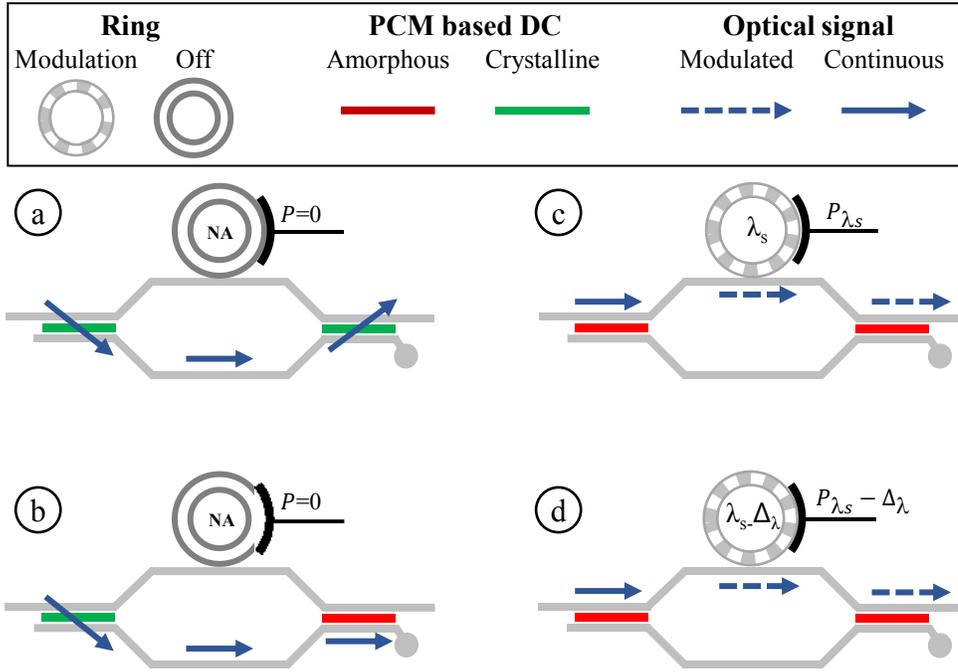
Figure 3.7: Non-volatile implementation of a) Pass/pass. b) Block/block. c) Pass/block d) Block/pass modes from RDL[8] using PCM-based directional couplers

3.2.2 Implementation of AND Function

In order to implement the product of two operands, we use SRMG configuration as shown in Figure 3.8. Since DC_2 from first cell is merged with DC_1 from second cell, the configuration of block/block mode is only available in the second cell, which implies to configure the first cell in the pass/pass mode. The design allows to implement functions such as A, B, AB, AB'. Figure 3.8 illustrates the implementation of AB'. For this purpose, first and second cells are configured in pass/block and block/pass modes respectively. This is obtained by configuring DC_1, DC_2 and DC_3 in *cr* state and tuning the first and second cell to λ_0 and $\lambda_0 - \Delta_\lambda$ respectively.
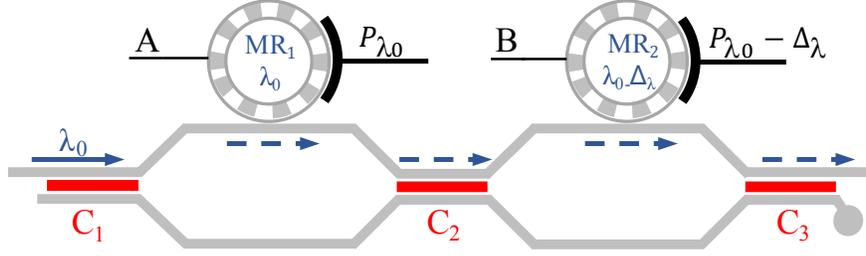
Figure 3.8: Implementation of AB' by configuring PCMs to *cr* and tuning rings to λ_s and λ_s-Δ_λ

3.2.3 Non-Volatile RDL Architecture

In order to implement the function OR, the SRMG architecture is duplicated on two parallel waveguides as represented in Figure 3.9. Signals propagating from waveguides are transmitted to multiband photodetector which results in the sum of products. Proposed PCM based RDL architecture feature the implementation of XOR function, i.e. AB'+BA', with AB' being implemented in the upper waveguide. It is obtained by configuring first and second cell in pass/block and block/pass modes respectively. This is achieved by configuring DC_1, DC_2 and DC_3 in *cr* states and tuning the first and second rings to λ_0 and $\lambda_0-\Delta_\lambda$ respectively. Therefore, signal at λ_0 is transmitted to the output when rings are off resonance, which requires A=1 and B=0. BA' is implemented on the lower waveguide by configuring first and second cells in block/pass and pass/block modes respectively.
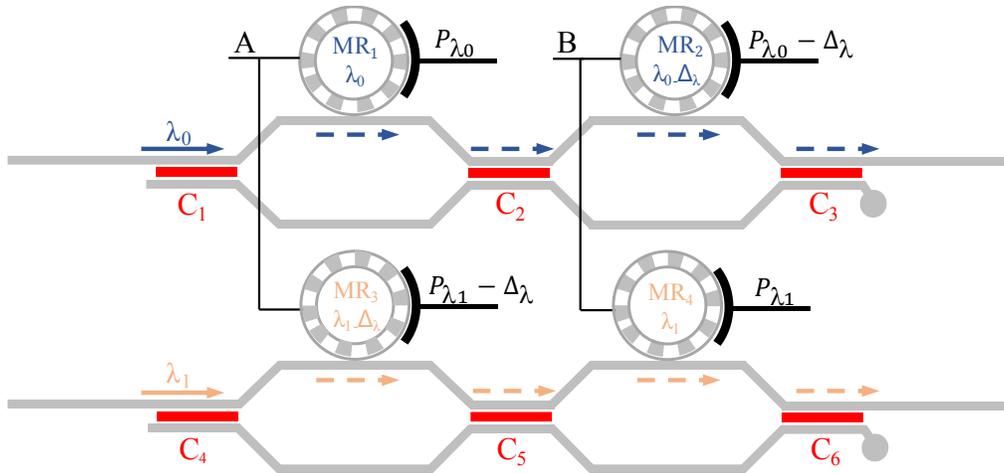
Figure 3.9: Ring tuning and PCM configuration of RDL for XOR

3.2.4 Interfaces

In the following, we consider two implementations for the interface defined as ring filter based RDL and coupler based RDL. The interfaces allow to i) transmit signal from laser to the architecture and ii) to transmit modulated signal to the photodetector. Ring filter based RDL involves the use of add/drop rings to drop signal matching with their resonant wavelength. In coupler based RDL, lasers are inserted on each waveguide, thus requiring a coupler to merge the signals transmitting to the photodetector. In following we present PCM configuration and ring tuning for each implementation for different functions as provided in Table 3.1 and Table 3.2**.**

3.2.4.1 Ring Filter based RDL

Figure 3.10 represents ring filter based RDL. The MRR filters on the left-hand side are used to couple signal from lasers to the horizontal waveguides. The modulated signals are transmitted to a photodetector through MRRs located on the right-hand side. The filter MRRs call for constant calibration to ensure the coupling.
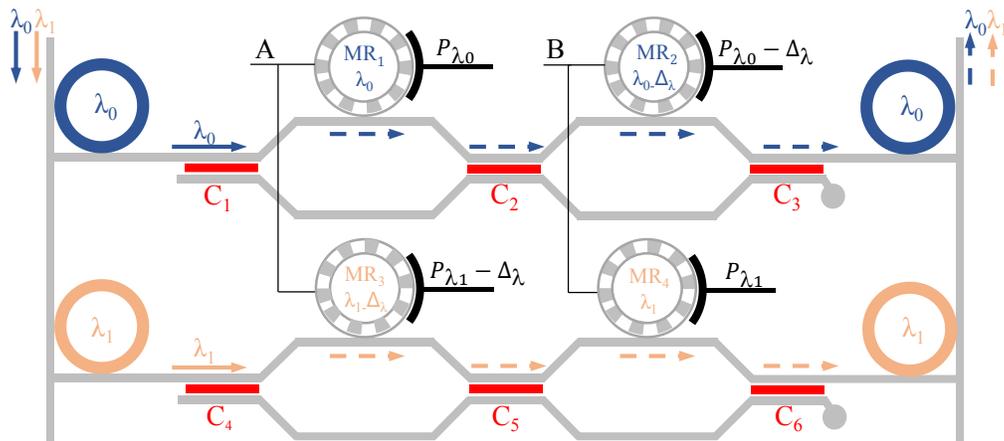
Figure 3.10: Ring filter based RDL configured for XOR

Table 3.1 summarizes the configurations of PCMs and the rings according to the logic function for ring filter based RDL architecture. Functions involving a single product induce block/block mode for the lower waveguide which leads to bypassing of signal. XOR and XNOR functions involve modulation on all the rings, which requires to configure all the DCs in the *cr* state. MR_3 modulates data when functions involving a second product include operand 'A' (e.g. XOR and XNOR). Since all functions can be executed without reconfiguring DC_6, the device could be removed for reduced hardware complexity purpose. However, since keeping DC_6 offers the opportunity to map single produce function on the lower waveguide, we didn't consider this optimization.

Table 3.1: Device state according to the function for RDL with PCM and ring filter

Device	Function						
	A	B	AB	AB'	$A + B$	$A + B'$	$AB + A'B'$
DC_1	*cr*	*am*	*cr*	*cr*	*cr*	*cr*	*cr*
DC_2	*am*	*am*	*cr*	*cr*	*am*	*am*	*cr*
DC_3	*am*	*cr*	*cr*	*cr*	*am*	*am*	*cr*
MR_1	λ_0	off	λ_0	λ_0	λ_0	λ_0	λ_0
MR_2	off	λ_0	λ_0	$\lambda_0 - \Delta_\lambda$	off	off	λ_0
DC_4	*am*	*am*	*am*	*am*	*am*	*am*	*cr*
DC_5	*cr*	*cr*	*cr*	*cr*	*am*	*am*	*cr*
DC_6	*cr*	*cr*	*cr*	*cr*	*cr*	*cr*	*cr*
MR_3	off	off	off	off	off	off	$\lambda_1 - \Delta_\lambda$
MR_4	off	off	off	off	λ_1	$\lambda_1 - \Delta_\lambda$	$\lambda_1 - \Delta_\lambda$

3.2.4.2 Coupler based RDL

Figure 3.11 represents coupler based RDL. Lasers are placed on each waveguide allowing turning them off when signal is not used. Therefore, block/block mode is not needed which allows to remove the terminator. Signals propagating from two waveguides are merged through coupler.
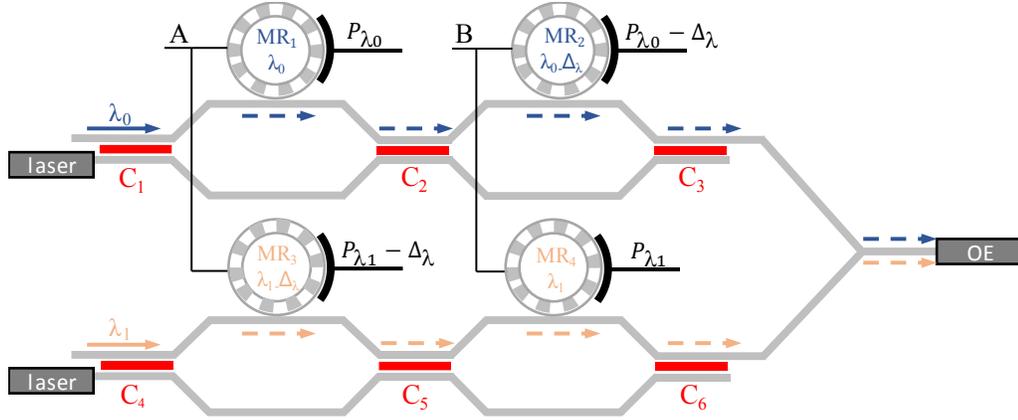
Figure 3.11: Coupler based RDL configured for XOR

Table 3.2 summarizes the PCM configuration and ring tuning for coupler based RDL. Laser is turned off on lower waveguide for functions involving the use of one waveguide such as A, AB. This allows to avoid the configuration of PCMs which are shown with don't care (i.e. x) in the table.

Table 3.2: Device state according to the function for RDL with PCM and coupler

Device	Function						
	A	B	AB	AB'	$A + B$	$A + B'$	$AB + A'B'$
DC$_1$	cr	am	cr	cr	cr	cr	cr
DC$_2$	am	am	cr	cr	am	am	cr
DC$_3$	am	cr	cr	cr	am	am	cr
MR$_1$	λ_0	off	λ_0	λ_0	λ_0	λ_0	λ_0
MR$_2$	off	λ_0	λ_0	$\lambda_0-\Delta_\lambda$	off	off	λ_0
DC$_4$	x	x	x	x	am	am	cr
DC$_5$	x	x	x	x	am	am	cr
DC$_6$	x	x	x	x	cr	cr	cr
MR$_3$	off	off	off	off	off	off	$\lambda_1-\Delta_\lambda$
MR$_4$	off	off	off	off	λ_1	$\lambda_1-\Delta_\lambda$	$\lambda_1-\Delta_\lambda$

The common characteristic of both non-volatile RDLs are as following: i) to allow bypassing unused MRRs, thus saving the calibration power and ii) laser power overhead induced by insertion loss of PCMs. Ring filter based RDL suffers from several limitations. For instance, two lasers are considered to be active and PCMs on all waveguide require configuration for functions

involving the use of one waveguide. Another challenge is the calibration requirement of input and output rings either to drop signal or to ensure off resonance condition. Coupler based RDL involves the use of lasers on each waveguide which leads to improvements as followings: i) laser is deactivated for functions involving the use of one waveguide, ii) For these functions PCM configuration on unused waveguide is avoided and iii) ring filter calibration power is saved. However, the limitation involved with the use of coupler based RDL is the coupler induced loss which would results in overhead in laser power. Therefore, taking into account all parameters is necessary to obtain an actual comparison between the architectures.

3.2.5 Toward Large Scale Architectures

In this section, we study the usage of the architecture to enable the processing of multi operand functions. From the coupler based architecture we define architecture illustrated in Figure 3.12. It is essentially composed of two cascaded RDL cores which are interconnected using waveguides linking upper and lower branches of DC_3 and DC_9 to DC_4 and DC_{10}. The other two branches are used to sum the output signals and transmit the results to a photodetector. Therefore, different optical connections between the cores are obtained depending on the PCM configurations. In following we show how sum of products for four operands can be achieved using proposed architecture. We also discuss the limits of the architecture.
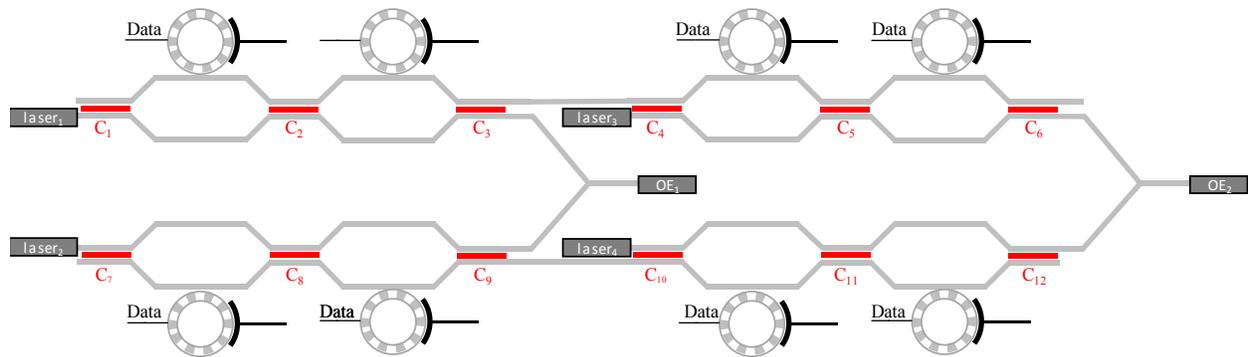
Figure 3.12: Architecture for processing multi operand functions

3.2.5.1 Sum of Product Implementation of Four Operands

Figure 3.13 illustrates the implementation of ABCD+EFGH. Both ABCD and EFGH are implemented through configuring all cells in pass/block mode. In order to transmit the signal from first core to the second one, both DC_3 and DC_9 are configured in *cr* state. This allows to avoid turning on the laser$_3$ and laser$_4$. Signals transmitting from lower and upper waveguides

propagate to the OE of second core through configuring DC_6 and DC_{12} in *am* state. Therefore, the sum of products is obtained.
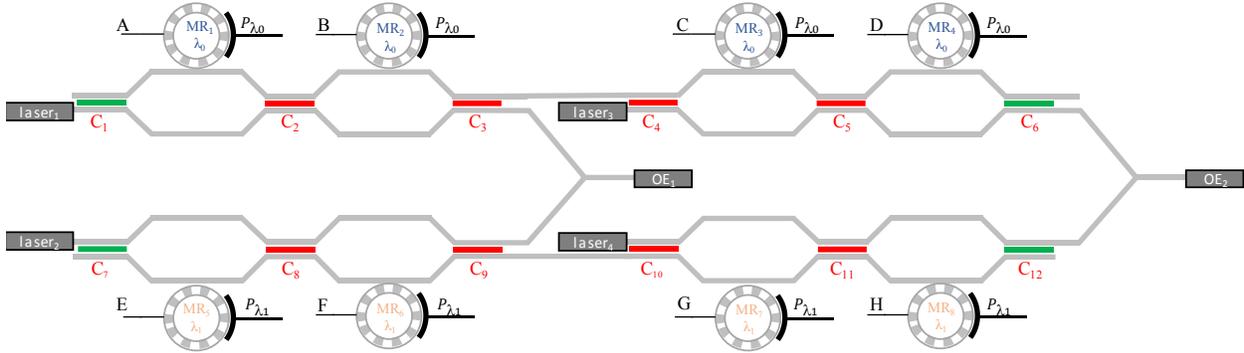
Figure 3.13: Implementation of ABCD+EFGH

3.2.5.2 Limitation Associated with Large Scale Architecture

Although cascading of RDL cores allows implementing multiple operand functions, the architectures suffer from several limitations. For instance, the implementation of multi-input XOR cannot be achieved since it requires crossing of data between the cores. While using electro-optical solutions would solve the issue, the use of electronics would also considerably limit the advantages of such architectures. Hence, topologies involving heterogeneous and specialized cores [7][30] are probably needed. Another challenge remains the losses induces by PCM material. This will call for synthesis tools enabling the mapping of functions to minimize the crossing of PCMs [72]. Finally, as already previously discussed, PCM suffers from a limited endurance and high reconfiguration time. This will call for optimized application mapping solutions while taking into account the current PCM state to minimize changes of states.

3.3 PCM based Nanophotonic Interconnects

In this section, we investigate the use of proposed cell in the context of nanophotonic interconnects. ONoC -or nanophotonic interconnects- uses WDM to ensure high bandwidth required by data intensive applications. To realize WDM, the architectures rely on group of MRRs to filter out signals on the receiver side. To investigate the efficiency of design we focus on SWMR since it is most commonly used approach. To reduce the calibration requirement, we use (MRSG) configuration for each transmitter or receiver. Since each channel is composed of groups of readers and one group of writers MRMG architecture is used.

We consider the 3D architecture illustrated in Figure 3.14: bottom layer implements processing units and memories, while top layer implements the optical interconnect. It is composed of 16 clusters of 4 cores. Each cluster includes a shared last level cache (L3). Each core has a private L1 data (L1d), instructions (L1i) caches, of 32KB each, and a private L2 cache of 512KB. MESI protocol ensures coherency between the distributed caches. All clusters are connected with ONIs through Through-Silicon-Via (TSVs). Each Optical Network Interface (ONI) is connected to one transmitting and 15 receiving waveguides which are featured with SWMR channel and WDM, and include PCM based DCs, as depicted in Figure 3.14. The channel is reconfigurable and allows to bypass unused readers using PCM-based DCs. Each waveguide transmits 8 optical signals at different wavelengths.
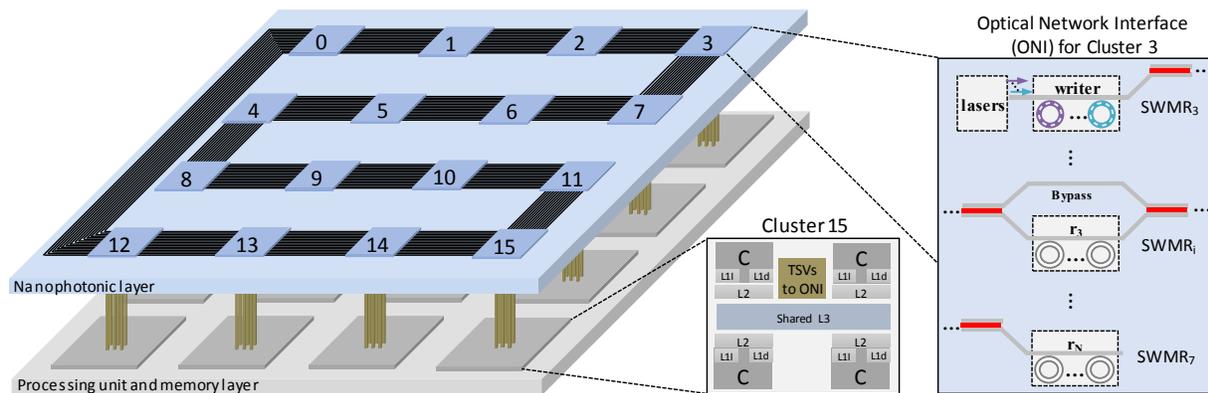
Figure 3.14: Considered 3D hardware architecture.

Figure 3.15 represents the proposed configurable SWMR channel assuming one waveguide and N wavelengths. As in conventional SWMR channel, the writer modulates the optical signal, which propagates towards the destination receiver where Opto-Electronic (OE) conversions are carried out. However, unlike conventional SWMR channels where all intermediate receivers must be crossed before reaching the destination, the configuration ability of the proposed SWMR channel allows to connect only a selected set of readers. By disconnecting readers from the optical path, we aim at reducing the power consumption as follows: i) MRRs in disconnected receivers do not require, power hungry, calibration and ii) optical losses can be reduced since signals do not propagate through all the MRRs.

To achieve this, we use MRMG configuration by placing readers in modulation path of the proposed cell as represented in Figure 3.15. For each DC, the path can be configured to transmit signals towards the bypass waveguide or towards the receiver.
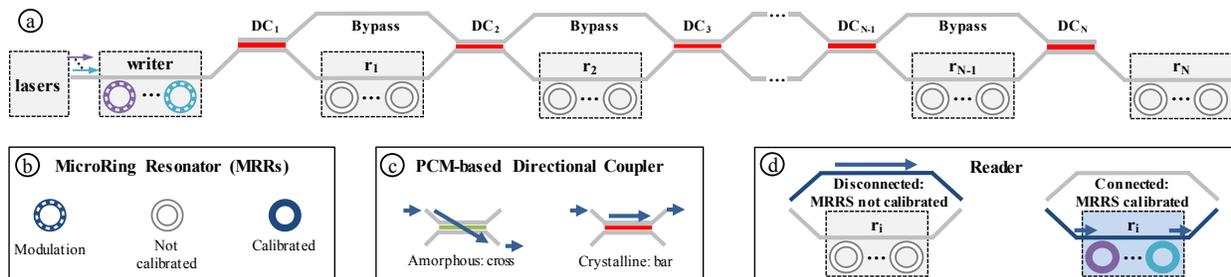
Figure 3.15: a) Proposed SWMR channel with PCM-based directional couplers to configure the optical path through readers and bypass waveguides, b) MRRs states, c) Signal transmission (for all wavelengths) through the directional coupler according to the state of the PCM and d) Signal transmission to connected reader or use bypass path for disconnected reader.

3.3.1 Configuration Method and Use Case Scenarios

A key limitation of PCM elements is the slow phase state changes (around 100ns [73]) compared to the required nanosecond scale latency communication requirements in manycore. We tackle this limitation by partitioning the manycore to execute different applications and by reconfiguring the interconnect only when new applications are executed. For each application configuration is obtained according to the mapping of application on the cores. The execution times of the targeted benchmarks applications, which typically range from 100ms to 10s, will ensure a low reconfiguration frequency of the SWMR channel (<<1Hz). The connectivity requirements depend on the number of writers (i.e. the number of SWMR channels) and number of readers per channel. In the context of manycore architectures where multiple SWMR channels are used, each channel is configured according to the list of readers to be reached by the writer. To connect a reader, the preceding DC is configured either to cross if the previous reader is disconnected (i.e. signals coming from bypass) or to bar if the previous reader is also connected. Since we assume that a SWMR configuration does not change during the execution of an application, the MRRs of disconnected readers are not calibrated. On the contrary, the MRRs of connected readers are calibrated as follows: MRRs of the destination receiver drop the signals from the waveguides, while the others MRRs let the signals propagating (through). Maintaining the calibration of all connected MRRs allows minimizing the communication latency overhead [48]. To further save energy, the injected laser power is adapted to losses experienced by the optical signal for each configuration.

Figure 3.16 illustrates various connectivity scenarios for a 4-readers SWMR channel. We assume three wavelengths, thus leading to three MRRs per reader. The scenarios are:

- **All readers connected** (Figure 3.16.a) leads to a regular SWMR channel for which all the MRRs require calibration. To achieve the channel configurations, all the PCM elements are set to *cr* state, thus leading to bar transmission for the directional couplers. As previously mentioned, any of the connected receivers can be reached by the writer by calibrating the corresponding MRRs to the drop transmission (the others MRRs are calibrated to through transmission). Regarding the optical losses, this configuration leads to the highest losses, and hence highest laser power requirements, since the optical signal can propagate through all MRRs.

- **r_2, r_3 and r_4 connected** (Figure 3.16.b) require bypassing r_1, which is achieved by configuring DC_1 to the *am* state. Since r_2 is part of the connected readers, DC_2 is also configured to *am*. The rest of the PCMs are set to *cr* state, thus allowing to transmit the signals to r_3 and r_4. Since one reader is bypassed, the MRR through losses are reduced, thus leading to laser power saving. Furthermore, reduction of calibration power is also achieved since signals do not propagate through r_1.

- **r_1 and r_3 connected** (Figure 3.16.c) involves bypassing r_2. PCM element in DC_1 is configured to *am*, thus allowing the signal propagation through r_1. To bypass r_2, PCM elements in DC_2 and DC_3 are also configured to *am*. Since r_4 is also disconnected and is beyond the last connected reader, DC_4 does not require any specific configuration. In term of power, further reduction can be achieved compared to previous scenario due to lower MRRs losses, waveguide propagation losses and MRR calibration.

- **r_2 connected** (Figure 3.16.d) leads to SWSR channel. To achieve this, DC_1 and DC_2 are set to cross state while DC_3 and DC_4 can be in any state (i.e. no reconfiguration needed).

Based on scenarios we conclude that PCM based interconnect leads to an overhead in laser power due to the loss induced by DCs. The worst scenario corresponds to a case where all readers are connected. Therefore, the less the connected readers are the higher the ring calibration power saving can be achieved. Scenarios involving disconnected readers lead to saving in laser power and ring calibration power. For these readers while the calibration of unused rings is avoided, laser power according to the decrease in total loss is also reduced.
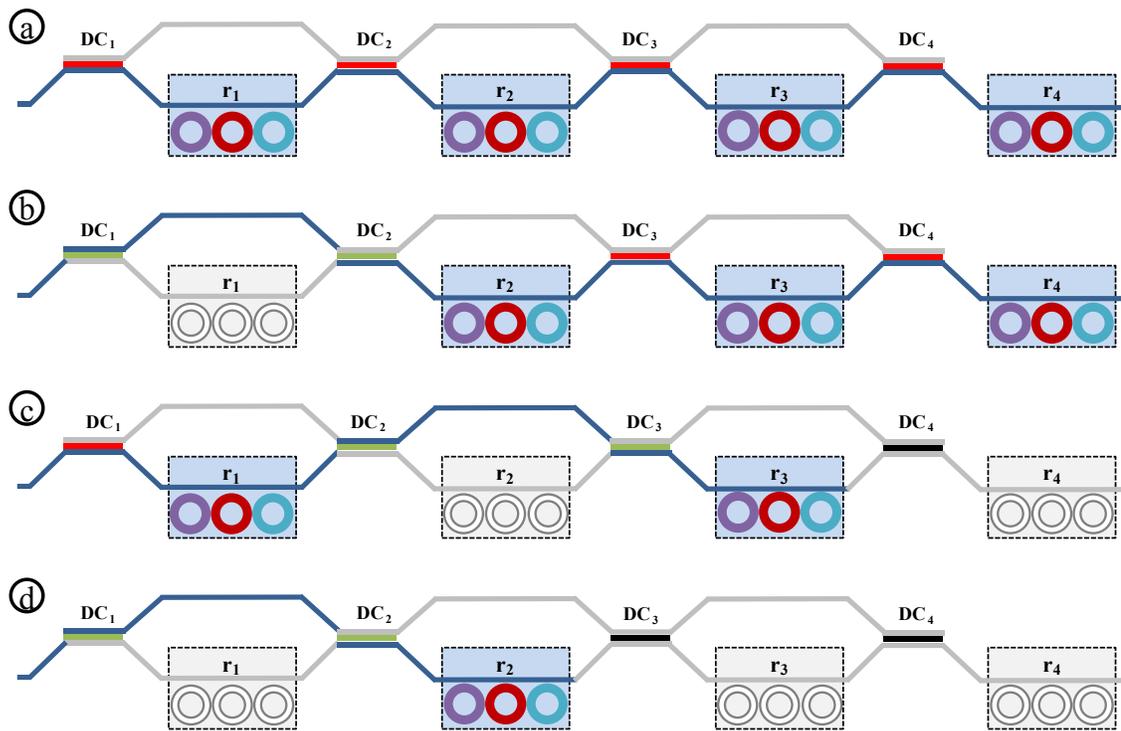
Figure 3.16: PCM elements configuration and ring calibration for various scenario: a) All interfaces connected, leading to a regular SWMR channel, b) r_1 disconnected, which allows to not calibrate its MRRs, c) r_2 and r_4 disconnected d) r_2 only connected leading to SWSR channel.

3.3.2 Application Mapping

To efficiently use the non-volatile capability of PCM based interconnect, mapping application on dedicated number of clusters is considered. Figure 3.17.a represents mapping on four numbers of clusters. It leads to use of four SWMR channels which implies that each used channel is configured to bypass disconnected readers. Figure 3.17.b-e illustrates mappings on 6, 8, 9 and 12 clusters.
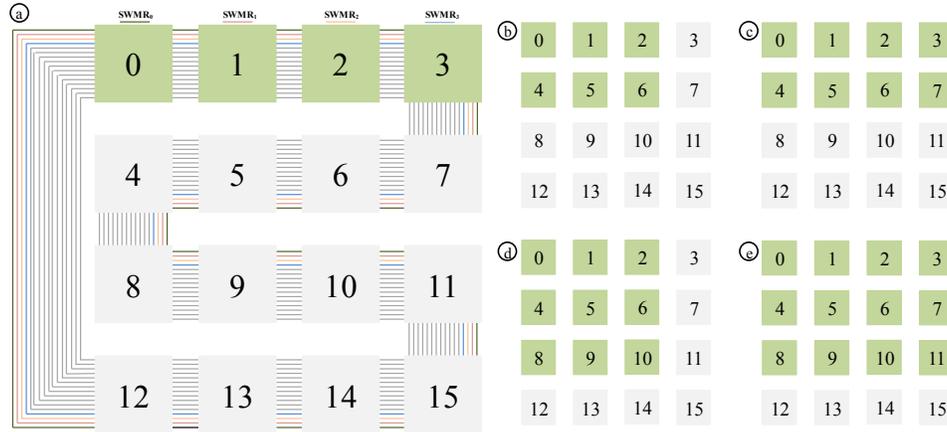
Figure 3.17: Mapping example: a) Application mapped on Cluster 0 to 3, which leads to the use of Channel 0 to 3 only ; b-e) Mapping of one application on 6, 8, 9 and 12 clusters

To address the latency induced by reduced number of dedicated clusters, we partition the network to execute applications in parallel. PCMs allow partitioning the architecture in order to allocate a dedicated number of clusters per application. This is compared with sequential execution of two applications. Since readers which are not part of a SWMR channel partition are disconnected, the application tasks can be executed in parallel without any resources sharing. Figure 3.18.a-b illustrates cluster partitioning for the parallel execution of two applications. It is possible to distribute more than two applications on any number of clusters as represented in Figure 3.18.c, which is suitable to explore the design space. In result section we will show how partitioning interconnect affects the execution time and energy consumption.
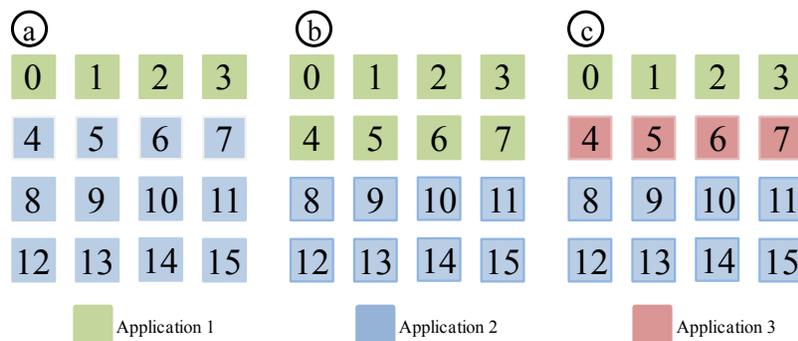
Figure 3.18: a) Mapping of two applications on 4 and 12 clusters, b) Mapping of two applications on equal number of clusters, c) Mapping of three applications on 4, 4 and 8 clusters

42

3.4 Summary

In this chapter we presented a generic method of using PCM based DC to bypass rings. We then proposed four implementations involving different number of rings per cell or different number of cascaded cells. The use case of proposed architectures were investigated in the context of optical computing and optical interconnect. In next chapter we propose a model allowing to evaluate laser power overhead as well as saving achieved through bypassing unused rings.

Chapter 4

Modeling PCM based Optical Architectures

In this section we first present a model used to evaluate power consumption in optical architectures. The model takes into account the optical power as well as electrical power. It also allows to estimate the laser power overhead and the reduced ring calibration power consumption. Based on that we derive the models for PCM based RDL and interconnect.

4.1 Proposed Power Model

In each optical architecture, power consumption takes into account the laser power, electrical to optical to electrical conversion power, ring power and PCM configuration power as defined by:

$$P_{total} = P_{laser} + P_{EOE} + P_{ring} + P_{reconfig} \qquad (1)$$

Where P_{laser} is the laser power needed to reach the targeted optical power based on photodetector sensitivity, P_{EOE} is the conversion power from electrical to optical at source and optical to electrical at receiver side, P_{ring} is the power required to calibrate ring and $P_{reconfig}$ is the required power to change the state of PCM in directional coupler when architecture is reconfigured.

4.1.1 Laser Power

Optical signal propagating through the architecture experiences loss induced by optical devices. Laser power is obtained according to the laser efficiency and photodetector sensitivity considering worst case loss as defined by Equation (2). The loss experienced by signal depends on MRR through loss, MRR drop loss, waveguide propagation loss, insertion loss induced by the DCs, coupler loss and crosstalk power penalty according to Equation (3).

$$P_{laser} = (P_{received} + IL_{wc})/eff \qquad (2)$$

$$Loss_{wc} = L_{MR-t} + L_{MR-d} + L_{wg} + \sum_{a=0}^{A-1} IL_{cr}^{bar} + \sum_{b=0}^{B-1} IL_{am}^{cross} + IL_{coupler} + X_{talk} \tag{3}$$

L_{MR-t} is the loss induced by rings calibrated to through (letting signal propagate), L_{MR-d} is the loss induced by MRRs calibrated to drop (same resonance wavelength of signal). *A* is the number of bar transmission for PCM configured in *cr* state and *B* is the number of cross transmissions for PCM in *am* state. As previously explained, *am* state leads to the cross transmission of most signal power (IL_{am}^{cross}) while only small fraction of the power is transmitted to bar (IL_{am}^{bar}) as shown in Figure 4.1. The opposite occurs for *cr* state: most of the signal power is bar transmitted while a small fraction of the signal power is cross transmitted ($IL_{cr}^{cross} \ll IL_{cr}^{bar}$).
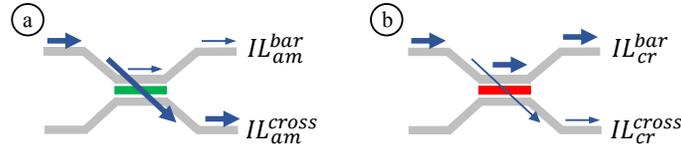
Figure 4.1: IL for DC according to the state of PCM and output port, a) *am*: cross transmission of most signal power, b) *cr*: bar transmission of most signal power

4.1.2 MRR Calibration Power

The MRR in optical architectures carry out modulation or filtering. For modulation, MRR are either calibrated to signal wavelength or to a wavelength with small deviation from signal. Filter MRRs are either calibrated to through or drop. Since MRRs are highly sensitive to any process or environmental imperfection, the resonant wavelength shift is monitored and compensated through thermal, electro-thermal or opto-thermal processes. In next sections we will investigate the MRR calibration power at the architecture level in more details.

4.1.3 PCM Configuration Power

The reconfiguration of the architecture involves changing the state of PCMs. This requires phase state conversion from *cr* to *am* or from *am* to *cr*. While static power consumption depends on loss induced by optical devices and ring calibration power, the dynamic power, $P_{reconfig}$ depends on PCM state conversion energy and the architecture reconfiguration frequency *f* as defined by:

$$P_{config} = f \times \left(\sum_{a=0}^{N_{PCM}^{cr}} E_{cr_i \to am_i} + \sum_{b=0}^{N_{PCM}^{am}} E_{am_j \to cr_j} \right) \tag{4}$$

4.2 Modeling for PCM based RDL

In this section we first present a framework describing the designing of non-volatile RDL then we evaluate the power consumption for the architecture by instantiating generic model proposed in section 4.1.

4.2.1 Design Flow

Figure 4.2 represents the flow for obtaining the power consumption for RDL architecture. To estimate the power consumption for non-volatile RDL we first consider function mapping on the version of RDL implementation. Power consumption in both implementations of RDL is evaluated taking into account the parameters as following: i) number of rings calibrated as modulator (pass/block or block/pass), ii) number of PCMs configured to *cr/am*. Ring filter based RDL involves the calibration of ring filters to drop signal from/to vertical waveguide. Therefore, for this implementation of RDL we also consider number of rings filters tuned to resonance. Coupler base RDL does not involve the use of ring filters; instead it uses coupler to merge signals. For this version of RDL laser is turned off on unused waveguide which implies that there is no need to configure PCMs.
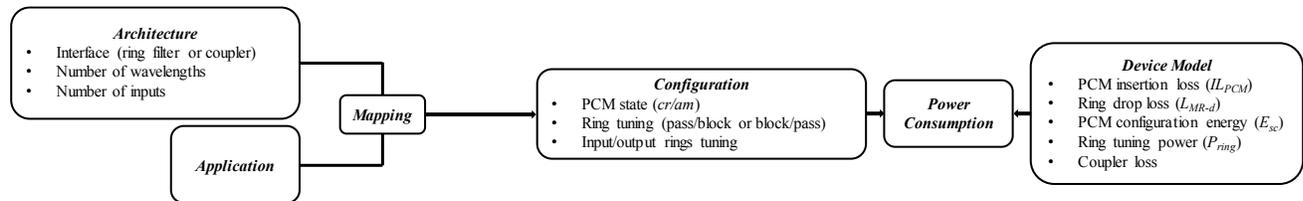
Figure 4.2: Modeling framework for PCM based non-volatile RDL

In order to estimate the required laser power, we obtain worst case loss at the architecture level. To evaluate loss for non-volatile RDL we consider the model proposed in [8]. Based on that, we do not take into account MRR drop loss and waveguide propagation loss. However we consider MRR through loss as represented in Table 4.1. It summarizes the ring transmission parameters according to the selected tuning resonance wavelength and the modulated data. Tuning ring to λ_s (resp. λ_s-Δ_λ) leads to IL_λ (resp. $IL_{\lambda -\Delta\lambda}+ER_{\lambda -\Delta\lambda}$) and $IL_\lambda +ER_\lambda$ (resp. $IL_\lambda -\Delta_\lambda$) for logic inputs of '1' and '0' respectively. When the ring is tuned to λ_s+Δ_λ, the loss is independent from the data.

Table 4.1: Ring loss according to the tuning and modulated data

Tuning	Data	
	0	1
λ_s	$(IL)_{\lambda s} + ER_{\lambda s}$	$IL_{\lambda s}$
$\lambda_s - \Delta_\lambda$	$IL_{\lambda s - \Delta \lambda}$	$(IL)_{\lambda s - \Delta \lambda} + (ER)_{\lambda s - \Delta \lambda}$
$\lambda_s + \Delta_\lambda$	$IL_{\lambda s + \Delta \lambda}$	

PCM induced loss is obtained as represented in Figure 4.1. In our model, we do not consider the crosstalk induced by bar and cross transmission through IL_{cr}^{cross} and IL_{am}^{bar} respectively. However, in block/block mode where signal mostly propagates toward the terminator, we consider IL_{cr}^{cross} for the last DC to obtain the ratio of signal propagation to the output. In addition to abovementioned losses, coupler based RDL involves the loss induced by coupler to merge signals.

The worst-case loss occurs for functions in which signal is propagating through two modulating rings such as XOR and AB. XOR involves the use of four modulating rings. To propagate signal through these rings, DC$_{1-3}$ are configured to *cr* which lead to $3IL_{cr}^{bar}$ loss for PCM. Each waveguide includes one ring tuned to λ_s (pass/block) and one ring tuned to λ_s-Δ_λ (block/pass), which results in MRR through losses of $L_{\lambda s}$ and $L_{\lambda s-\Delta\lambda}$. Therefore the worst-case loss is obtained as $3IL_{cr}^{bar} + L_{\lambda s} + L_{\lambda s-\Delta\lambda}$.

To obtain ring power for coupler based RDL we only consider the number of rings tuned to λ_s or λ_s-Δ_λ. For ring filter based RDL in addition to the tuning of modulating rings we consider the number of filter rings tuned to resonance (λ_s). For instance, AND involves the tuning of three filter rings while XOR requires four number of ring filters to be tuned.

To obtain the reconfiguration power we consider two scenarios. We first consider the worst-case scenario in which i) we assume that all PCM elements change state when a new function is configured and ii) we use the largest of $E_{(cr \rightarrow am)}$ and $E_{(am \rightarrow cr)}$ for the state conversion. We then consider a more realistic scenario in which we take into account the actual number of PCMs that change state for each possible reconfiguration. For this purpose, we assume that architecture is initially configured for a function then we consider its reconfiguration to other function and obtain the number of PCMs that must change state. For instance, if we consider architecture is initially configured for XOR which requires all PCMs to be in *cr* state, its reconfiguration for AB

requires four PCM state conversions. Table 4.2 presents required number of PCM state changes for each pair of functions before and after reconfiguration.

Table 4.2: Number of PCMs state changes for each reconfiguration

		Function after reconfiguration							
		A	B	AB	AB$'$	A+B	A+B$'$	XNOR	XOR
Function before reconfiguration	A	-	2	2	2	1	1	3	3
	B	2	-	2	2	3	3	3	3
	AB	2	2	-	0	3	3	1	1
	AB$'$	2	2	0	-	3	3	1	1
	A+B	1	2	3	3	-	0	4	4
	A+B$'$	1	2	3	3	0	-	4	4
	XNOR	2	2	1	1	4	4	-	0
	XOR	2	2	1	1	4	4	0	-

4.2.2 RDL Power Model

In this section, we detail the proposed power model for non-volatile RDL. Total power consumption at the architecture level is obtained using Equation (1). For this architecture we do not consider the power consumption induced by electrical to optical conversion at laser and optical to electrical at photodetector. P_{EOE} only involves the power consumed to modulate optical signal with arrays of rings. Therefore, by substituting P_{EOE} with P_M we define the total power consumption for non-volatile RDL as following:

$$P_{total}=P_{laser}+P_M+P_{ring}+P_{reconfig} \tag{5}$$

- Laser Power

The Laser power for non-volatile RDL is obtained taking into account the worst-case loss as defined by Equation (2). As mentioned earlier for this architecture we do not consider MRR drop loss and waveguide propagation loss. MRR through loss involves the loss induced by modulating rings tuned to λ_s and λ_s-Δ_λ as defined by Equation (6). Finally by replacing MRR through loss with Equation (6) we obtain worst-case loss for the architecture according to Equation (7):

$$L_{MR-d}=L_{MR-t/\lambda s}+L_{MR-t/\lambda s-\Delta\lambda} \tag{6}$$

$$Loss_{wc/RDl} = \sum_{m=0}^{M} L_{MR\text{-}t/\lambda s} + \sum_{n=0}^{N} L_{MR\text{-}t/\lambda s\text{-}\Delta\lambda} + \sum_{a=0}^{A\text{-}1} IL_{cr}^{bar} + \sum_{b=0}^{B\text{-}1} IL_{am}^{cross} + IL_{coupler} \qquad (7)$$

Where M and N are the number of rings tuned to λ_s and $\lambda_s-\Delta\lambda$ respectively.

- Ring Power

Non-volatile RDL involves the use of modulating and filter rings. Modulating rings allow to realize functions by modulation of data into optical signal and ring filters drop signal from/to vertical waveguide. Considering the propagation of signal with wavelength (λ_s) through the architecture, the modulating rings are tuned to λ_s or $\lambda_s-\Delta\lambda$. Ring filters are tuned to λ_s. Therefore for PCM based RDL total ring power is defined by i) the calibration power of modulating rings (i.e. rings which are not bypassed using the directional couplers) and ii) the calibration power of ring filters, as defined by:

$$P_{ring} = \sum_{i=1}^{I} P_{\lambda s} + \sum_{j=1}^{J} P_{\lambda s\text{-}\Delta\lambda} \qquad (8)$$

Where I and J represent the number of rings calibrated at λ_s and $\lambda_s-\Delta\lambda$ respectively. The calibration power involves the compensation of resonant wavelength shift induced by environmental changes.

- Reconfiguration Power

The states of PCMs are configured based on the implemented function. To realize another function reconfiguration of PCMs are carried out. As previously mentioned we consider two scenarios for PCM reconfiguration in non-volatile RDL. For realistic scenario we obtain the reconfiguration power as defined by Equation (4). For worst case scenario the reconfiguration power is obtained considering state change for all PCMs as defined by Equation (10).

$$E_{sc} = max\,(\,E_{cr\rightarrow am}, E_{am\rightarrow cr}\,) \qquad (9)$$

$$P_{reconfig} = f \times \sum_{i=0}^{\substack{number\ of \\ PCMs}} E_{sc} \qquad (10)$$

4.3 PCM based Nanophotonic Interconnect

In this section we first present how an interconnect is configured based on connectivity requirement and what parameters are considered to estimate the power consumption, then we instantiate the generic model proposed in section 4.1 for PCM based interconnect.

4.3.1 Design Flow

In following we evaluate the power consumption in PCM based nanophotonic interconnect as represented in Figure 4.3. This part is done in collaboration with Cedric Killian and Joel Ortiz from University of Rennes 1 and all simulations using SNIPER were carried out by them. We evaluate the performance of architecture considering application mapping from PARSEC and SPLASH benchmark suites which are commonly used to study the performance of manycore architectures. They contain applications from different domains, which lead to various processing and communication requirements. Six representative applications have been selected according to their parallelization level: i) FFT and Raytrace from Splash2 and ii) x264, Blackscholes, Barnes and Dedup from PARSEC. The execution of these applications on the architecture is simulated using the SNIPER manycore simulator. Splash2 applications were modified to allow their parallel executions and SNIPER was modified to handle the distribution of threads. It is possible to distribute applications on any number of clusters, which is suitable to explore the design space. The number of bits obtained from simulation allows to estimate the energy per bit. I would like to emphasize that the simulations using SNIPER were carried out by Joel Ortiz and Cedric Killian from University of Rennes 1.

To estimate power consumption in PCM based interconnect we first define the architecture parameters which are as following: i) number of SWMR channels, ii) number of wavelength and iii) distance between network interfaces. The power consumption in the architecture depends on required connectivity which allows to estimate parameters such as number of connected readers and number of crossed PCMs. To obtain power usage we first configure the architecture for a given connectivity. This implies to allocate required number of channels for the communication and configure PCMs in the used channels. PCMs are configured to allow signal propagation through connected readers and to bypass disconnected ones. For instance, for a connectivity requiring four connected interfaces only four SWMR channels are used. This needs the PCM

before and after the readers to be in *am* state. The PCM before connected readers allows to propagate signal from bypass to the readers and the PCM after readers leads to signal transmission to the bypass. The rest of PCMs are configured to *cr* to maintain signal propagating in the corresponding path. If we consider the number of connected readers to be N, MRRs in N-1 interfaces are tuned to through and only one reader is tuned to drop.
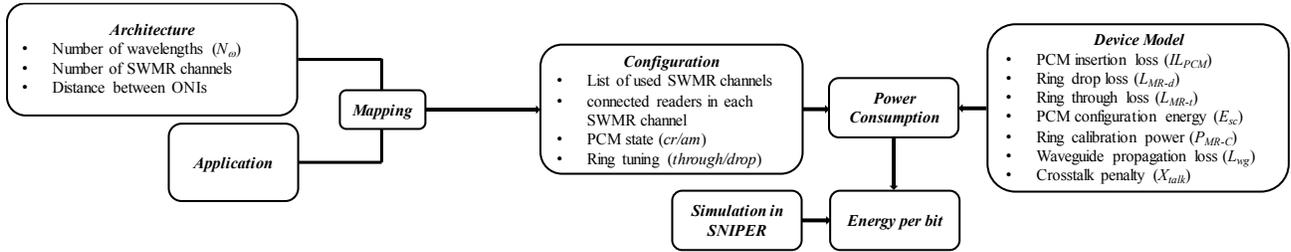
Figure 4.3: Modeling framework for PCM based nanophotonic interconnect

To obtain laser power for each channel we consider the worst case loss. To evaluate loss we consider the number of crossed PCMs, number of connected readers, number of wavelengths, position of last connected reader and crosstalk. The PCM insertion loss is obtained as defined in Figure 4.1. Numbers of connected readers along with number of wavelengths allow to estimate MRR through loss. Position of last connected reader is used to obtain waveguide propagation loss.

To evaluate ring calibration power, we only consider the rings located in connected readers in the channels which are used. Transmitter MRRs are tuned to the same wavelength of signals. Readers are either calibrated to through or drop. MRR calibration targets compensating for resonant wavelength shift induced by temperature or process variations. For this purpose we use the model proposed in [18] which we investigate in more details in next section.

Reconfiguration power takes into account the state conversion required for execution of a new set of applications. For this purpose, interface connectivity might change, thus leading to a different set of connected readers. This requires changing the state of PCMs from *cr* to *am* or from *am* to *cr*, which requires energy.

4.3.2 Interconnect Power Model

In this section, we detail the proposed power model for non-volatile nanophotonic interconnect. We estimate the total network power consumption by considering the contribution of each

SWMR channel, as defined in Equation (11). P_{SWMR_i} corresponds to the power consumption of channel i, which depends on its configuration and is obtained using Equation (1). For this, P_{ring} is the power required to calibrate the rings in connected readers and P_{config_i} is the power needed to configure the PCM elements in the directional couplers. P_{EOE} involves the transmitter (P_T) and receiver power consumptions (P_R) for each SWMR channel. P_T is the power required to serialize and modulate optical signals and (P_R) is the receiver power consumption, required to photo detection, amplification, comparison and deserialization. Therefore we can rewrite Equation (1) for each SWMR channel according to Equation (12).

$$P_{channels} = \sum_{i=0}^{N-1} P_{SWMR_i} \tag{11}$$

$$P_{SWMR_i} = P_{laser_i} + P_T + P_R + P_{config_i} + P_{ring_i} \tag{12}$$

- Laser Power

For each channel, we define the required laser power according to the Equation (2). The worst-case losses in a channel is defined by Equation (13).

$$Loss_{wc_i} = L_{MR\text{-}t} \times N_\omega \times n + L_w \times p_{r_i} \times d + L_{MR\text{-}d} + \sum_{a=0}^{A-1} IL_{cr}^{bar} + \sum_{b=0}^{B-1} IL_{am}^{cross} + X_{talk} \tag{13}$$

The propagation of the signals towards the last connected reader involves crossing all rings in the n connected interfaces. L_{MR-t} is the through loss per ring and N_ω is the number of wavelengths, which also corresponds to the number of rings per reader. The waveguide propagation loss experienced by the signals depends on the position of the last connected reader, the distance between interfaces (d) and the waveguide loss (L_w in dB/cm). L_{MR-d} is the drop loss in the receiver. X_{talk} is crosstalk power penalty which depends on the number of wavelengths and the ring position in the reader interface [74].

- Ring Calibration

For nanophotonic interconnect we obtain required thermal power compensating resonant wavelength shift induced by temperature. To evaluate the calibration power, we consider the Thermal Tuning Efficiency (TT_e) of the closed loop feedback system proposed in [18]. Only the

rings in connected readers require calibration, which leads to Equation (14). In the equation, $\Delta\lambda_{MR-C_{kj}}$ corresponds to the required wavelength shift of ring at position j in reader k. Assuming a homogeneous distribution of the signal wavelengths among the FSR, the maximum shift for each ring is FSR/N_λ. Therefore, the required shift for each MRR is defined by Equation (15), where $\Delta\lambda_{shift_{kj}}$ is the wavelength shift of MRR j located at reader k. It is obtained with Equation (16) where ΔT is the temperature variation and $\frac{d\lambda}{dT}$ is the MRR thermal sensitivity.

$$P_{MR-C} = \sum_{k=1}^{n} \sum_{j=1}^{N_\lambda} \Delta\lambda_{MR-C_{kj}} . TT_e \tag{14}$$

$$\Delta\lambda_{MR-C_{kj}} = \frac{FSR}{N_\lambda} - (\Delta\lambda_{shift_{kj}} \bmod \frac{FSR}{N_\lambda}) \tag{15}$$

$$\Delta\lambda_{shift} = \frac{d\lambda}{dT} \Delta T \tag{16}$$

- Channel Configuration

To realize new application SWMR channels are reconfigured, thus leading to different sets of connected readers. By assuming PCM already configured for a given connectivity requirement, we obtain reconfiguration power using Equation (4). It is worth mentioning that the reconfiguration time of the PCMs is not critical in our system since SWMR channels are reconfigured only when new applications are executed.

4.4 Summary

In this section we proposed a model which allows to estimate the power consumption in PCM based architectures taking into account the optical and electrical power. The optical power involves laser power, PCM reconfiguration power and ring calibration power. The model evaluates laser power overhead induced by DCs. It also estimates the ring calibration power and allows investigating the impact of bypassing unused rings. The model evaluates the PCM reconfiguration power according to the state conversion frequency. In result section we will consider architecture configuration for different applications and for each we will study the power consumption.

Chapter 5

Results

In this section we investigate the potential of PCM in saving static power consumption in the context of optical computing and optical network. The first part is dedicated to PCM based RDL and in second part we will evaluate the PCM based nanophotonic interconnect.

5.1 RDL

In this section, we evaluate the power consumption of the proposed RDL architectures. We first estimate the laser power overhead needed to compensate for losses induced by PCM elements and coupler. We then estimate the impact of the reconfiguration frequency on the cell power efficiency. Table 5.1 summarizes the considered parameters for micro ring resonator and DC at 1521.5nm wavelength. We assume 0.9mW modulation power (P_M) [8].

Table 5.1: Cell parameters

Device	Parameter type	Parameter	
MR	Tuning power (mw)	$P_{\lambda s}$	9.9[8]
		$P_{\lambda s - \Delta \lambda}$	9.7[8]
		$P_{\lambda s + \Delta \lambda}$	12.9[8]
	Loss (dB)	$IL_{\lambda s}$	-1.25[8]
		$ER_{\lambda s}$	-12.25[8]
		$IL_{\lambda s - \Delta \lambda}$	-1.25[8]
		$ER_{\lambda s - \Delta \lambda}$	-8.75[8]
		$IL_{\lambda s + \Delta \lambda}$	0[8]
DC	Phase transition energy (nJ)	E_{sc}	2[11]
	Loss (dB)	IL_{cr}^{bar}	-0.16[11]
		IL_{cr}^{cross}	-13.7[11]
		IL_{am}^{bar}	-22.9[11]
		IL_{am}^{cross}	-0.72[11]

5.1.1 Cell Insertion Loss

We evaluate the cell insertion loss for each configuration, as reported in Table 5.2. Pass/pass leads to the lowest loss since the signal propagating from input to the output cross two DCs in the *am* states. Assuming IL_{am}^{cross}=0.72dB, this leads to 1.44dB total loss. Block/block leads to the 14.42dB loss, i.e. the highest attenuation, by configuring DC_1 and DC_2 in *am* and *cr* states respectively. Pass/block involves using the modulation path, i.e. DC_1 and DC_2 are in *cr* state and ring is tuned to λ_s. Depending on the modulated data, the ring involves an attenuation of $IL_{\lambda s} = 1.25dB$ (data '1') and $IL_{\lambda s}+ER_{\lambda s}$=13.5dB (data '0'). The only difference for block/pass is the ring detuning, which is set to λ_0-$\Delta\lambda$. This leads to 1.57dB and 10.32dB loss for data '0' and '1' respectively, thus resulting in high extinction ratio for both modulation modes. Since comparable insertion losses are obtained for all the modes, data '1' on the cell output will be represented by similar power levels. We thus conclude that a same laser power can be used for all the configurations and that no laser power tuning is needed.

Table 5.2: Cell insertion loss wrt cell configuration

Mode	Device configuration			IL (dB)	
	DC_1	MR	DC		
`pass/pass	*am*	NA	*am*	$2 \times IL_{am}^{cross}$	1.44
block/block	*am*	NA	*cr*	$IL_{am}^{cross} + IL_{cr}^{cross}$	14.4
pass/block	*cr*	λ_s	*cr*	$2 \times IL_{cr}^{bar} + IL_{\lambda s}$	1.57
block/pass	*cr*	λ_s-$\Delta\lambda$	*cr*	$2 \times IL_{cr}^{bar} + IL_{\lambda s-\Delta\lambda}$	1.57

5.1.2 Laser Power

In order to estimate the required laser power, we estimate the worst-case loss at the architecture level for each implementation of the non-volatile RDL architecture. Figure 5.1 illustrates the loss breakdown for each RDL. The worst-case loss occurs for functions in which signal is propagating through two modulating rings such as AB and XOR, which involves $3IL_{cr}^{bar}$ and $2IL_{\lambda s/\lambda s-\Delta\lambda}$ and results in 2.98dB. For same functions RDL in [8] leads to 2.5dB loss.
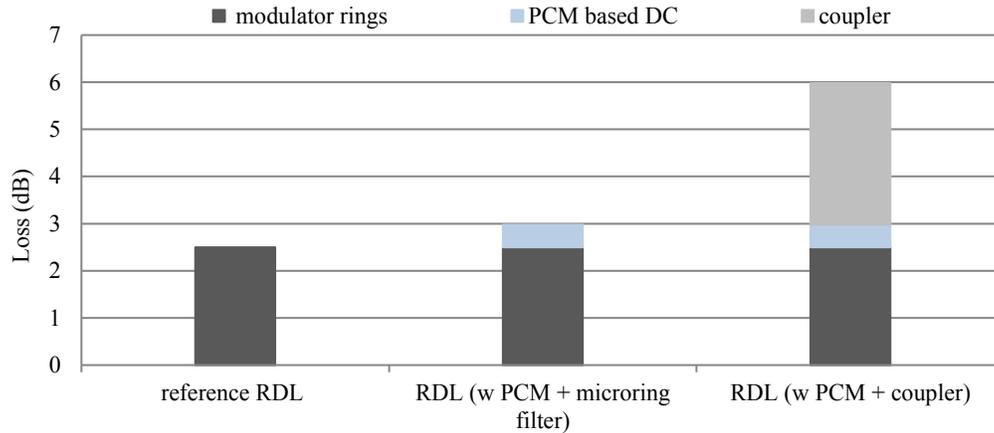
Figure 5.1: Loss breakdown for RDL architecture

To compensate the 0.48dB and 3.48dB additional loss for RDLs with ring filters and coupler, the injected optical power are set to 2.25mW and 4.5mW respectively. Assuming a 25% lasing efficiency [75], this leads to 1mW and 10mW laser power overhead respectively. In the following, we discuss how energy saving can be achieved for RDL with PCM and ring filters thanks to i) the use of the bypass path, which allows to avoid tuning unused rings. For coupler based RDL extra saving is achieved thanks to the ii) removal of ring filters which reduces MRR calibration power and iii) turning off laser for functions which involves the use of one waveguide such as A and AB.

5.1.3 Power Saving Analysis

In the following we investigate the power saving of the two implementations of non-volatile architecture wrt RDL in [8] as reported in Figure 5.2. For functions A and B, three rings out of four are bypassed thanks to the PCM based DC. This results in 35% saving for RDL with ring filters. For coupler based RDL in addition to bypassing rings, turning off the laser on the lower horizontal waveguide and saving the calibration power of ring filters lead to 72% power saving. This is achieved despite of 10mW laser power overhead needed to compensate for the loss induced by DCs and coupler. Functions involving two operands (A+B, AB, AB', A+B') allow bypassing two rings, thus leading to 22% power saving for RDL with ring filters. For coupler based RDL, functions AB and AB' lead to 61% power saving, while functions A+B and A+B' result in 50% power saving. While in all the above mentioned functions two rings are bypassed and calibration power of ring filters are saved, however turning off laser for functions of AB and

56

AB' leads to extra saving. XOR and XNOR involve the use of all rings. Therefore, due to the higher laser power needed to compensate loss induced by PCM, RDL with ring filters leads to slight power increase of (+0.2%). For coupler based RDL calibration power saving of ring filters outperform the laser power overhead and results in 29% power saving. Therefore, while RDL with ring filters leads to 19% average power saving, 53% is obtained for coupler based RDL.

The results demonstrate that using PCM to bypass ring resonators not needed to modulate data lead to significant improvement in the power efficiency. While PCM leads to saving in both implementations of proposed RDLs, keeping laser off for some functions and saving the calibration power of ring filters result in extra saving for coupler based RDL.
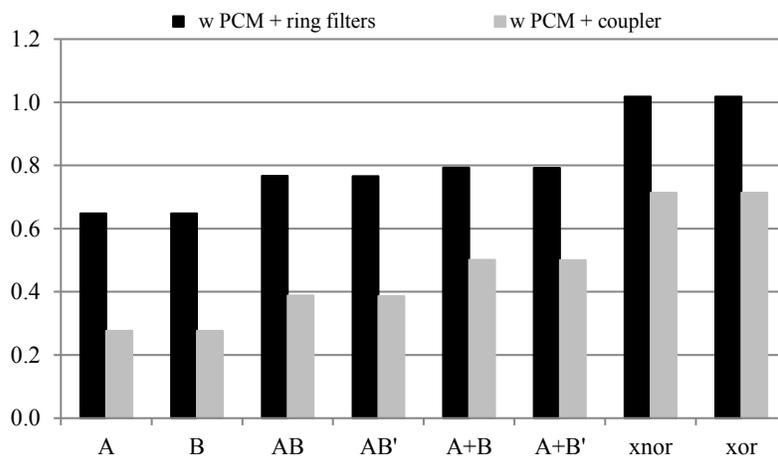
Figure 5.2: Normalized power of ring filter based and coupler based RDLs wrt RDL in [8] Power analysis based on Laser Efficiency and MRR Calibration Power

5.1.4 Power Analysis According to the Lasing Efficiency

In this section we investigate the impact of MRR calibration power and laser efficiency on power saving of coupler based RDL. For this purpose, we consider laser efficiencies of 10% and 25% and we focus the study on functions A+B and XOR, as illustrated in Figure 5.3 and Figure 5.4. We assume MRR calibration power ranging from 1mW to 10mW for pass/block mode, which corresponds to the power consumption needed to detune the rings from the signal wavelength. We also consider MRR calibration power for block/pass and pass/pass modes to be respectively 0.2mw below and 3mw above the calibration power for pass/block mode.

The implementation of A+B with coupler based RDL considering 25% laser efficiency is more power efficient for all considered MRR calibration power, as shown on Figure 5.3. However, for

10% laser efficiency, the proposed implementation is power efficient from 6mW. Implementation of A+B on coupler based RDL involves saving the calibration power of four filter rings and two modulating rings. However, the PCM also induces laser power overhead. Therefore, the architecture is more power efficient when laser efficiency is 25% or MRR calibration power is greater than 6mW.
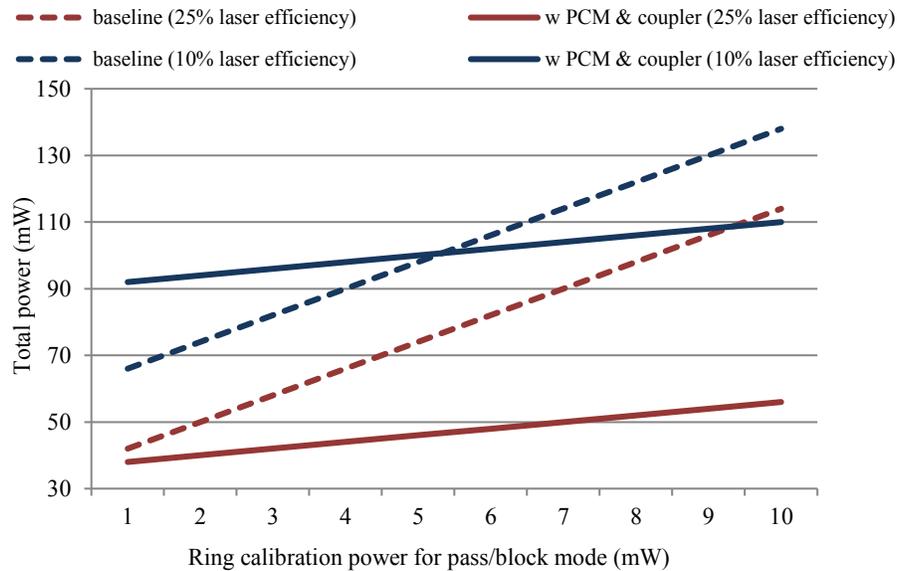
Figure 5.3: Total power consumption for A+B considering laser efficiencies of 10% and 25%

Figure 5.4 shows the implementation of XOR on both RDLs. Coupler based RDL is more power efficient from 2mW and 9mW for laser efficiencies of 25% and 10% respectively. Since XOR involves the use of all modulating rings, the coupler based RDL is less efficient for implementation of this function compared with A+B. However, considering laser efficiency of 25% makes coupler based RDL a more power efficient candidate for implementation of functions for calibration power ranging from 2mW to 10mW.
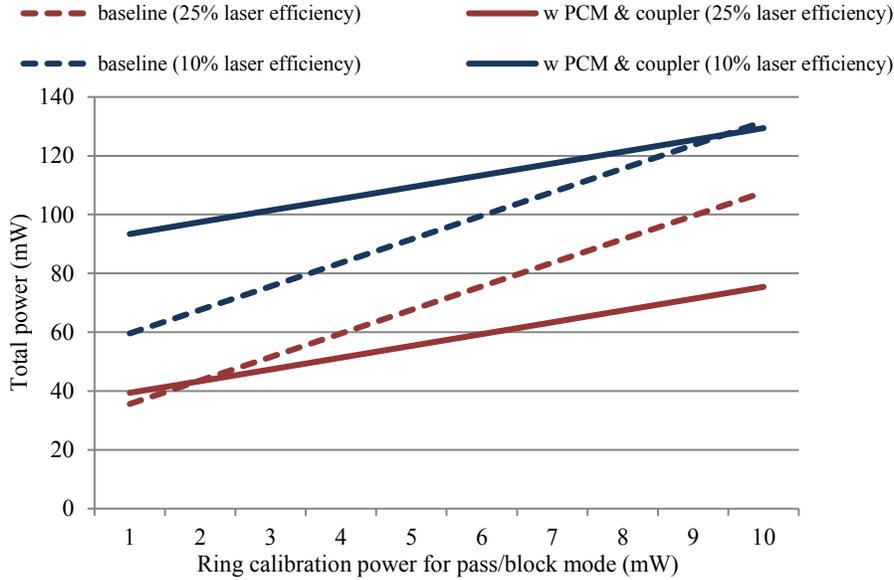
Figure 5.4: Total power consumption for XOR considering laser efficiencies of 10% and 25%

5.1.5 Reconfiguration Power

We evaluate the impact of state change of PCM elements according to the architecture reconfiguration frequency. For this purpose, we assume a 2nj [11] energy consumption to change the state of a PCM element. We assume a minimum reconfiguration period of 100ns since, according to [57], the amorphization and crystallization times are in the range of ps to ns. To obtain reconfiguration power two scenarios are considered. First we assume all PCMs are reset between each reconfiguration which leads to the worst case. In second scenario we take into account the actual number of PCMs that change state between each possible reconfiguration. For this purpose we assume that architecture is initially configured for a function then we consider its reconfiguration to all other functions and obtain the number of PCMs that must change state for each reconfiguration as summarized in Table 4.2 in section 4.2.

Figure 5.5 illustrates an example in which the initial function is A+B. Reconfiguring the architecture for A+B' does not require any PCM state change. Only ring tuning on lower waveguide changes from λ to λ-Δλ. However, implementing XOR requires four of PCMs to be reset and all rings to be tuned. To obtain reconfiguration power we consider the average of all PCM reconfiguration listed in Table 4.2.

59
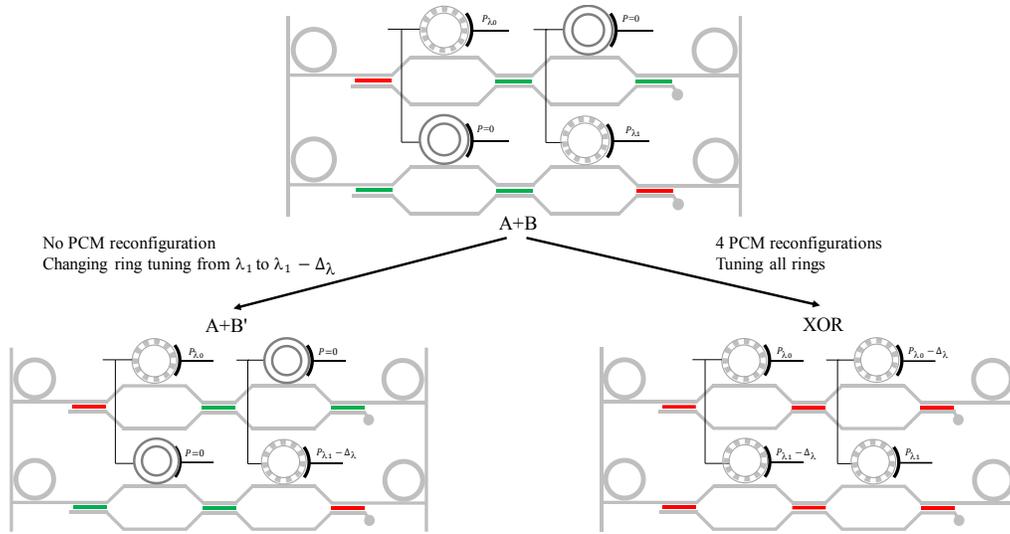
Figure 5.5: Reconfiguration of architecture to A+B' and XOR considering the initial function of A+B

While average power consumption of all functions for RDL in [8] is 107 mW, the power consumption for PCM based RDL with ring filters and with coupler is 87.3mW and 51mW respectively. Here we investigate the impact of the reconfiguration frequency on total power consumption. Figure 5.6 illustrates the power consumption for each reconfiguration scenario for two implementation of PCM based RDLs. Both coupler based RDL and ring filter based RDLs are most power efficient when no reconfiguration is required. The higher the reconfiguration frequency is the higher the power consumption is. Ring filter based RDL is power efficient up to 1.7MHz and 5MHz for worst case and actual scenarios respectively and coupler based RDL is power efficient up to 4.7 MHz and 14 MHz for corresponding scenarios. This demonstrates that taking into account the current state of PCMs is needed to efficiently reconfigure the architecture. This is especially important when the architecture is extended to process large numbers of inputs, as discussed in section 3.2.5.
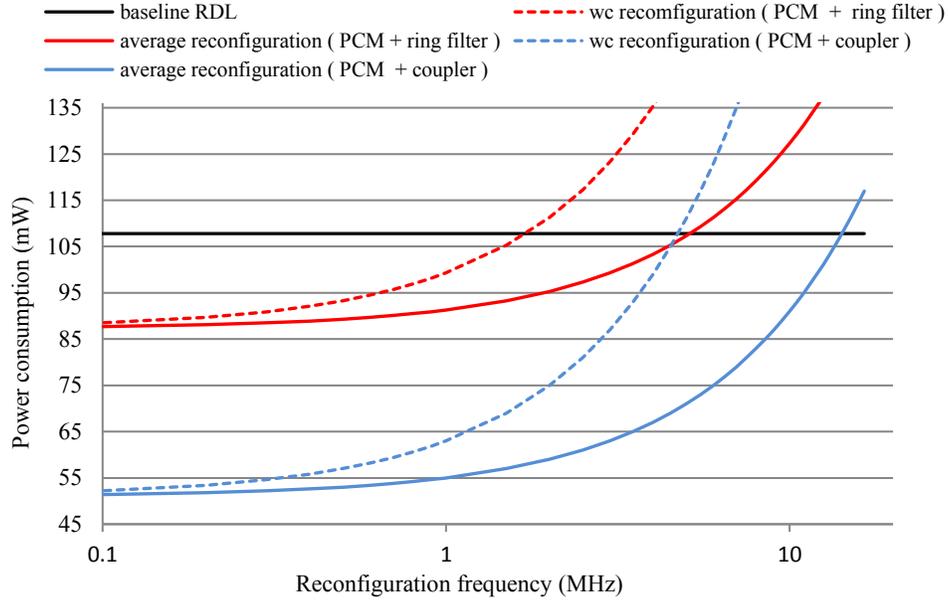
Figure 5.6: Power consumption according to reconfiguration frequency for PCM based RDLs

5.2 Nanophotonic Interconnect

In this section, we evaluate the power consumption of the proposed PCM based nanophotonic interconnect considering application mapping on different number of clusters. We then simulate the execution of benchmark application on architecture presented in section 3.3 using SNIPER environment and finally we consider the execution of pair of applications on the architecture. To obtain the traffic on the SWMR channels, we extract the communication traces between L2 caches and the distributed L3. Table 5.3 summarizes the architectural and technological parameters.

Table 5.3: Hardware and Technological Parameters

Parameter	*Definition*	*Value*
	Architecture	
	# clusters	16
	# cores per cluster	4
	L1I/D cache	32 KB each
	L2 cache	512 KB
	Cache protocol	MESI
	ONI	
d	Distance between interfaces	0.376cm [76]
N_ω	Number of wavelengths	8
B_R	Bit-rate	10Gb/s
P_T	Transmitter power	24mw [18]
P_R	Receiver power	24mw [18]
L_w	Waveguide loss in db/cm	0.25 [76]
L_{MR-d}	MRR drop loss	0.7 [76]
L_{MR-t}	MRR through loss	0.02 [76]
Q	MRR Quality Factor	20,000 [74]
TT_e	Thermal tuning efficiency	120pm/mw [18]
X_{talk}	crosstalk power penalty	0.0494 dB
	Directional Coupler	
IL_{cr}^{bar}	Bar output Insertion loss for PCM in cr state	0.16dB [11]
IL_{am}^{cross}	cross output Insertion loss for PCM in am state	0.72dB [11]
$E_{cr->am}$	Phase transition energy from cr to am	2nj [11]
$E_{am->cr}$	Phase transition energy from am to cr	2nj [11]

5.2.1 Losses and Power Analysis

We investigate the power consumption of the proposed interconnect and we compare it with a baseline interconnect without PCM elements. For this purpose, we first evaluate the loss of each channel used for 1x4 SWMR configuration and we report results in Figure 5.7. For $SWMR_0$, the waveguide loss is as small as 0.28 dB due to the short distance between the writer (cluster 0) and the last connected reader (cluster 3). Since readers located after the last connected reader are not used, the MR through loss for ONoC with and without PCM is the same (1.41dB). The need to cross three directional couplers for the PCM-based channel leads to 0.48dB overhead compared to the channel without PCM. Channels 1 to 3 demonstrate higher waveguide propagation losses since the signals propagate through the entire waveguide. Since only three readers are connected, MR through loss for channels with PCM is the same as for $SWMR_0$ (i.e. 1.41dB) while they reach 2.4 dB for without PCM due to the need to cross all the MRRs in the intermediate readers. Finally, the losses induced by the directional couplers (3.56dB) lead to 1.4dB additional losses

for the PCM based channel compared to channel without PCMs. To summarize, although PCMs allow reducing the MRR through loss when bypassing disconnected readers, the rather high insertion loss they involve lead to higher total losses. While PCM-based interconnect requires higher laser power, we will show in the following that significant power saving can be achieved thanks to significant reduction in the ring calibration power.
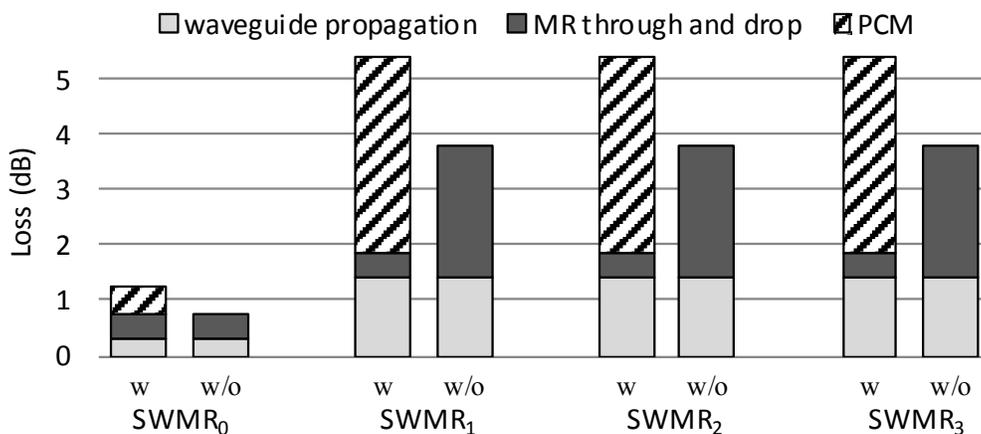
Figure 5.7: Loss breakdown on each channel for 1x4 configuration

Figure 5.8.a shows the power breakdown for previously discussed configuration and channels. For $SWMR_0$, the power consumption is 1mW higher with PCM due to laser power overhead. Calibration power is the same with and without PCM since no reader is bypassed. For channels 1, 2 and 3, we obtain x5 reduction in the calibration power since only the connected readers require calibration. Overall, despite the 8mW increase in laser power, our approach leads to up to 52% power saving. As shown in Figure 5.8.b, we obtain an average of 45% power saving per used channel power for 1x4 configuration. As the number of connected readers increases, lower power saving is obtained due to ring calibration requirements. Although configuration 4x4 leads to 6% power increase, our approach demonstrates a 21% power reduction on average. The higher energy efficiency of our approach for lower connectivity scenario ideally complements with the mapping of independent applications on the clusters, which we study in the following.
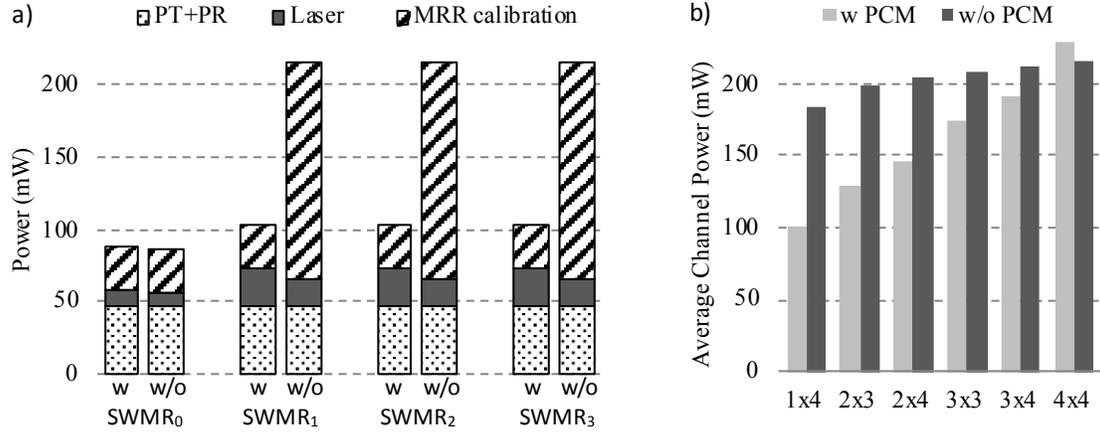
Figure 5.8: Power consumption results: a) power breakdown per channel for 1x4 configuration and b) average power consumption per channel according to the network configurations.

5.2.2 Benchmark Analysis

We simulate the execution of benchmark applications using the SNIPER environment, as reported in Figure 5.9. The mapping of Blackscholes (Figure 5.9.d) on 4 clusters (resp. 16 clusters) leads to 1376ms (resp. 1188ms) execution time and 28W (resp. 120W) power consumption. By using all clusters in the architecture, a 15% speedup is achieved at the cost of x4.2 power increase. Configuration involving 4 clusters and 8 clusters lead to intermediate speedups and powers, which we will use to optimize the multi-applications mapping and the ONoC configuration. While all tested applications follow a similar trend, the actual speedup and power consumption highly depends on the parallelization level and the communication patterns. Using 4 clusters for FFT is preferable since 16 clusters lead to 3% speedup and 3.5x power overhead (Figure 5.9.e). On the other hand, the mapping of x264 on 16 clusters allows 60% speedup. To conclude, clusters partitioning, which is achieved by reconfiguring SWMR channels, can be optimized depending on the executed applications.
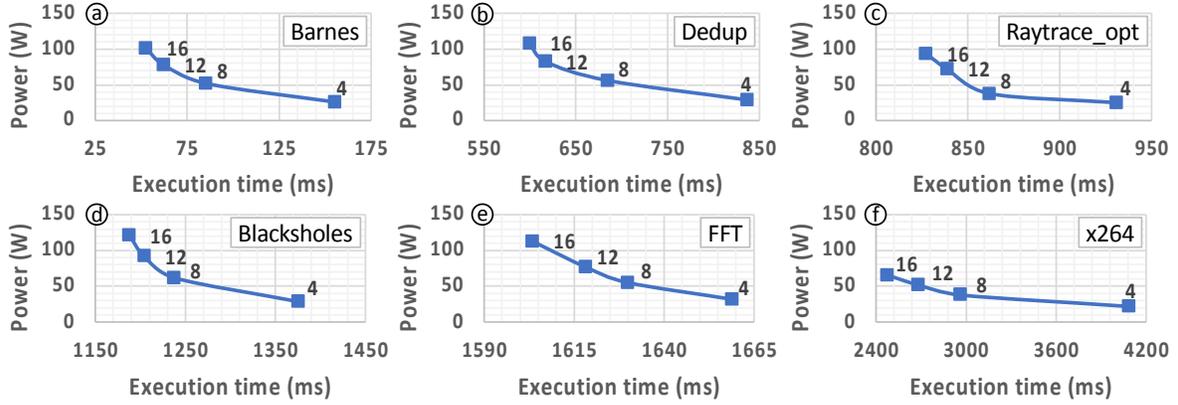
Figure 5.9: Execution time and power wrt. number of used clusters

5.2.3 Multi-Application Mapping and ONoC Configuration

We consider the execution of pairs of applications on the architecture. The baseline is a sequential execution of the two applications on all the clusters (denoted 16→16 in Figure 5.10) connected using an ONoC without PCM. As reported in Figure 5.10.b, the baseline execution time for x264 and Blacksholes is 3663ms. We now consider the proposed PCM-based ONoC. PCMs allow to partition the architecture in order to allocate a dedicated number of clusters per application (see Figure 3.18 for mapping details). Since readers which are not part of a SWMR channel partition are disconnected, the application tasks can be executed in parallel without any resources sharing, which contribute to reduce the execution time. As illustrated in Figure 5.10.b (case denoted 8/8 in the figure), the execution time is significantly reduced and reaches 2962ms. We then consider an uneven mapping of the application on the clusters (illustrated in Figure 3.18.a): 4/12 (resp. 12/4) allocates 4 (resp. 12) clusters to Blacksholes (resp. x264) and 12 (resp. 4) clusters to x264 (resp. Blacksholes). Mapping 12/4 leads to the best improvement in the execution time (26.8%). For the interconnect, SWMR channels are configured to connect only readers of interface executing the same application as the writer. Considering the 1.2Tb/s aggregated bandwidth of the optical interconnect (16 channels with 8 wavelengths at 10Gbps each), the baseline scenario would lead to a 2.7pJ/bit energy consumption under the assumption that all channels permanently transmit data. By considering the actual number of bits transmitted, we obtain a 47.7% energy per-bit reduction for the proposed ONoC compared to a network without PCM. The significant improvement in energy is due to i) the reduced static power induced by bypassed readers and ii) the higher data rate induced by shortened execution time. We carried out the same study for Blacksholes/Raytrace application pair, as reported in Figure

5.10.c. Interestingly, the best configuration is obtained by allocating 4 clusters to Blacksholes (against 12 in previous case), which is due to a lower computation load of Raytrace. Mapping 4/12 leads to 40.3% execution time reduction and 42.4% energy per bit reduction.

(a) Appli. 1 / Appli. 2	Blacksholes			x264			Raytrace_opt			Dedup			FFT		
Barnes	4/12	3.0	55.9	16→16	0	-6	4/12	4.7	36.5	4/12	5.4	33.2	4/12	2.3	27.5
FFT	12/4	42.0	68.6	4/12	34.2	47.9	12/4	33.4	41.7	12/4	26.6	31.2			
Dedup	4/12	32.6	31.8	4/12	12.8	33	4/12	41.2	40.3						
Raytrace_opt	4/12	**40.3**	42.4	4/12	18.8	39.6									
x264	12/4	**26.8**	47.7												

Best mapping wrt exec. time   *Exec. time reductions (%)*   *Energy per bit reduction (%)*
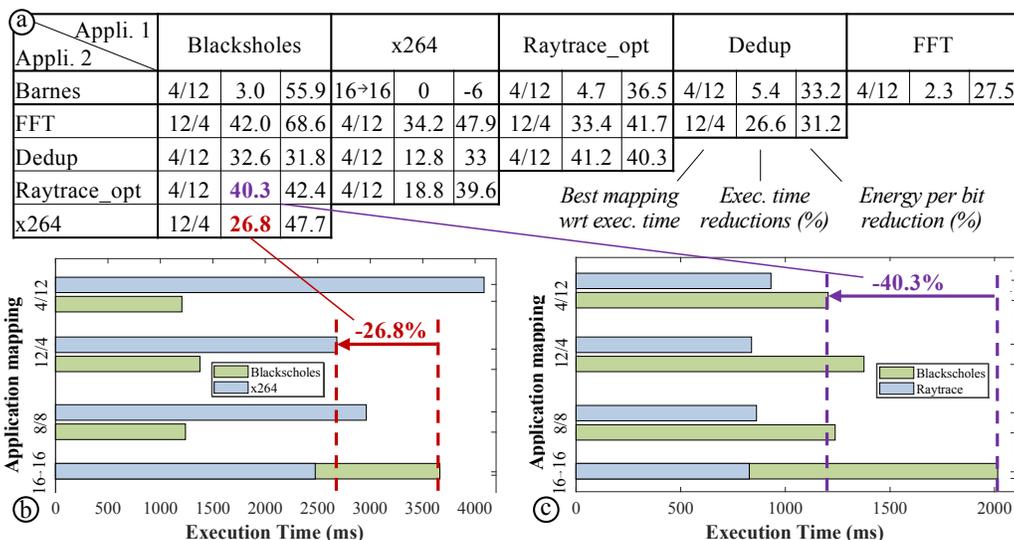
Figure 5.10: a) Best mapping results and improvements compared to execution on 16 clusters, b-c) execution time for different mappings and ONoC configurations for Blacksholes/x264 and Blacksholes/Raytraces.

Figure 5.10.a summarizes the best mapping we obtain for each application pair. Reductions in execution time and energy per bit reach up to 42% and 68.8% respectively for FFT/Blacksholes. Mappings involving Barnes lead to the lowest improvement since it is the least computation intensive application. No improvement is obtained when Barnes is combined with x264, which is the most data intensive application and calls for finer grain mapping. On average, we obtain 21.6% execution time reduction and 41.2% per-bit energy reduction for the PCM-based ONoC. This validates the potential for PCMs to improve the on-chip optical communication energy efficiency and to facilitate partitioning of manycore.

In our last study, we evaluate the power required to change the state of PCMs during network reconfigurations. For this purpose, we first consider a pessimistic scenario in which the two applications with the smallest execution time, i.e. Barnes and Dedup, require full reconfiguration of the PCMs after each execution. The resulting 1.3Hz reconfiguration frequency would lead to 5μW, which is negligible considering the total system power consumption (100W). Nevertheless, the PCMs reconfiguration power and endurance can be improved by optimizing the remapping to take into account the current state of PCMs.

5.3 Summary

We conclude that power consumption in PCM based architectures highly depends on application. If power saving through bypassing unused MRRs outperforms laser power overhead induced by DCs, power efficiency is achieved. This requires synthesis tools for optimized mapping to reduce the number of crossed PCMs and increase the number of bypassed rings. The mapping is especially important when there are large numbers of wavelengths which allows to avoid unnecessary ring calibration.

Chapter 6

Conclusion and Future Works

6.1 Conclusion

In this thesis we proposed a generic architecture based on Phase Change Material (PCM) which allows to bypass unused Micro Ring Resonators (MRR). The architecture involves the use of DCs placed before and after resonating device. Each DC allows configuring the optical path to transmit the signal to the MRR for modulation (filtering) purpose or to the bypass path when no modulation is carried out. We derived different implementations of design to adopt it to use cases. To achieve this one or groups of rings are involved per set of DC and design is cascaded. We also develop a model which allows to evaluate optical and electrical energy consumption.

6.1.1 RDL

Non-volatile PCM based RDL involve the use of Single Ring Multiple Groups (SRMG) since reconfiguration is required at the scale of one ring. The architecture is configured according to the application mapping which implies to configure PCM to transmit signal to the modulating rings and to bypass unused ones. We consider same laser power level for all functions and estimate it according to the worst case scenario. Results show that our proposed RDL leads to average power saving of 19% and up to 35% power saving is obtained if we bypass three out of four rings.

Based on results, functions involving ring bypassing allow to save calibration power of MRRs, thus saving is achieved. XOR/XNOR requires to use all rings which leads to power consumption overhead. To tackle this challenge we can decompose Boolean algebra into subset of gates or we can use a dedicated architecture for XOR.

Since we noticed that power saving obtained through ring bypassing has significant impact on static power reduction, we proposed a version of non-volatile PCM based RDL with different

interface. This implementation does not involve rings to couple signal from/to the horizontal waveguides, but instead it uses coupler to merge the modulated signals. Result show that coupler based non-volatile PCM based RDL can achieve up to 72% of power saving and on average 53% power saving is obtained. This confirms the necessity to design architectures with as little rings as possible.

6.1.2 Nanophotonic Interconnect

For nanophotonic interconnect we evaluate the proposed design on SWMR channel involving the use of Multiple Rings Multiple groups (MRMG). Due to the slow phase state change compared to the required nanosecond scale latency communication requirement in manycore we propose to reconfigure the architecture only when new applications are executed. We consider application mapping on different number of clusters and for each we study the power saving obtained through bypassing unused readers as well as laser power overhead induced by crossing PCMs. The channel is configured according to the mapping of application on dedicated clusters. It implies that PCMs are configured to allow signal propagation through connected readers and to bypass the unused ones. The model allows to obtain worst-case loss for each channel and estimates laser power according to that. To better evaluate the efficiency of design we simulate the execution of applications from Splash2 and PARSEC benchmarks on SNIPER environment. We obtain different execution time according to the number of clusters that are used. While on average 21% power saving is obtained compared with architecture without PCM, we can reach up to 52% power saving when application is mapped on 4 numbers of clusters. Based on the results, while for some applications mapping on 16 clusters lead to small or no improvements in terms of execution speed up, large power consumption is obtained. This justifies the investigation of the partial connectivity using only number of clusters. Hence, to take the advantages of PCM based nanophotonic interconnect, it is important to map application on only subset of clusters.

Since we observed that PCM based nanophotonic interconnect is only efficient when application is mapped on subset of clusters, we partition the architecture to investigate the parallel execution of two applications on the network. Configurability property of PCM allows partitioning the architecture for parallel execution of applications. This implies that we can isolate applications to run simultaneously and independently through configuring PCMs. The goal is to evaluate the

saving in power consumption as well as reduction in execution time. The baseline is the sequential execution of two applications on all 16 numbers of clusters. Based on the results the uneven partitioning (4/12) leads to best improvement in terms of execution time and energy and can achieve up to 42% reduction in execution time and up to 68.8% saving in power consumption for some applications. If applications are always mapped on these numbers of clusters there is no need to have PCM based DC between every pair of readers. This implies to use PCM only between the partitions and connect other readers directly to each other. The design can lead to improvements in PCM reconfiguration power due to the small number of PCMs that change state and enhanced laser power due to the reduced PCM induced loss.

6.2 Future Works

In this thesis we proposed a generic design which was applied to both optical computing and nanophotonic interconnect contexts. Although the design leads to power overhead induced by crossed DCs, the gain obtained through bypassing unused rings is significant. Therefore we conclude that our methodology in bypassing unused rings to reduce static power consumption has potential and can lead to significant gain under specific scenarios.

While the use cases for the proposed design focused on architectures with small size, to adopt the design to application with additional number of inputs, scalability should be investigated. To obtain scalable architectures it is important to tackle the limitations such as high optical power loss induced by large number of PCMs.

In PCM based architectures reconfiguration is carried out when new applications are mapped. However reconfiguration frequency is limited by endurance, high reconfiguration power and slow state change of PCM. Therefore, this is also one of the important challenges that call for methods to improve the performance of architectures.

Directional Coupler (DC) which we based our design on that, rely on complete amorphization and crystallization of PCM and leads to bar and cross. Intermediate state of PCM allows for adjusting the ratio of signal propagation to each of the bar and cross which can lead to various applications.

In following we will detail into each challenge and present future perspectives for them.

- *Scalability:* To obtain scalability it is significant to have a regular architecture. If we define each unit of the architecture as a cell, scaled architecture is achieved by cascading the cells to form connected arrays. Due to the high PCM loss induced by crossing large number of PCMs, it is not realistic to consider DC between each pair of cells. To address this challenge we can group the cells into clusters and place PCM based DC between clusters. This is defined as group bypassing and implies that cells in each clusters are connected directly to each other and cells of different clusters are connected through PCM based DC. It can lead to simple design and would reduce the loss induced by PCMs.

- *PCM Endurance:* Based on our results, we showed that computing architecture is power efficient up to 14MHz and above this frequency the gain is lost and power consumption is increased. To tackle this challenge, PCM reconfiguration power and endurance can be improved by optimizing remapping to take into account the current state of PCMs.

  It can also be improved through optimizing the sequence of applications to be executed so that minimum number of PCM reconfiguration is required. In the context of computing architecture, assuming that RDL is used as a coprocessor, the list of instructions to be executed on the database could be scheduled in a sequence which would need least number of PCM reconfiguration. For instance if it is assumed that we have to execute three functions of XOR, XNOR and 'A+B', execution sequence of XOR, XNOR and 'A+B' would require total number of 4 PCM state change with no reconfiguration between XOR and XNOR, while execution sequence of XOR, 'A+B' and XNOR would need 8 numbers of PCM reconfigurations.

  In the context of interconnect it means that if we have to schedule several applications and we map each on different number of clusters, the sequence of applications to be executed can be scheduled based on the required number of clusters in ascending order.

- *Use Case for Analogue behavior of PCM:* PCM has intermediate state (not fully crystalline, not fully amorphous) which allow for analog behavior of PCM. By controlling the degree of crystallization different ratio of signal propagation through bar and cross are obtained. In the context of RDL it can be used to split signal between two waveguides with equal or different split ratio according to the loss on each waveguide. In the context of interconnect, it can allow for multicasting communication. The device can

also allow for simultaneous communication of one writer with two readers in Single Writer Multiple Reader (SWMR) and Multiple Writer Multiple Reader (MWMR) channels.

Bibliography

[1] M. JR. Heck, J. F. Bauters, M. L. Davenport, D. T. Spencer, and J. E. Bowers. Ultra-low Loss Waveguide Platform and its Integration with Silicon Photonics. *Laser & Photonics Reviews*, 8(5):667–686, 2014.

[2] W. Bogaerts, P. D. Heyn, T. V. Vaerenbergh, K. DeVos, S. K. Selvaraja, T. Claes, P. Dumon, P. Bienstman, D.V. Thourhout and Roel Baets. Silicon microring resonators. *Laser & Photonics Reviews*, 6(1), 47–73, 2012

[3] F. Xia, L. Sekaric, and Y. Vlasov, *Nature Photonics* 1(1), 65–71 ,2007.

[4] S. Werner, J. Navaridas & M. Luján, A Survey on Optical Network-on-Chip Architectures, *ACM Computing Surveys*, 50, 1-37, 2017

[5] J. Teng, P. Dumon, W. Bogaerts, H. Zhang, X. Jian, X. Han, M. Zhao, G. Morthier, and R. Baets, *Opt. Express* 17(17), 14627–14633, 2009

[6] D. Taillaert, W. Van Paepegem, J. Vlekken, and R. Baets*, in: Proc. SPIE*, 6619, 661914, 2007

[7] L. Zhang, R. Ji, Y. Tian, L. Yang, P. Zhou, Y. Lu, W. Zhu, Y. Liu, L. Jia, Q. Fang and M. Yu, Simultaneous implementation of XOR and XNOR operations using a directed logic circuit based on two microring resonators. *Optics express*, 19, 6524-40, 2011

[8] C. Qiu, W. Gao, R. Soref, J. Robinson, and Q. Xu, Reconfigurable electro-optical directed-logic circuit using carrier-depletion micro-ring resonators, Opt. Lett. 39, 6767-6770, 2014.

[9] Z. Li, S. Beux, C. Monat, X. Letartre, I. O'Connor, Optical Look Up Table, 2013

[10] I. Chakraborty, G. Saha, A. Sengupta, et al. Toward Fast Neural Computing using All-Photonic Phase Change Spiking Neurons, *Sci Rep*, 8, 12980, 2018.

[11] P. Xu, J. Zheng, J. Doylend, A. Majumdar, Low-Loss and Broadband Nonvolatile Phase-Change Directional Coupler Switches, *ACS Photonics* ,6(2), 553 - 557, 2019.

[12] T. Ishihara, A. Shinya, K. Inoue, K. Nozaki, and M. Notomi. An Integrated Optical Parallel Adder as a First Step Towards Light Speed Data Processing. In *International SoC Design Conference. IEEE*, 123–124, 2016.

[13] Y. Imai, T. Ishihara, H. Onodera, A. Shinya, S. Kita, K. Nozaki, K. Takata, and M. Notomi. An Optical Parallel Multiplier using Nanophotonic Analog Adders and Optoelectronic Analog-to-Digital Converters. In *Conference on Lasers and Electro-Optics: Science and Innovations*, pages JW2A–50. Optical Society of America, 2018.

[14] D. Perez, I. Gasulla, L. Crudgington, D. Thomson, A. Khokhar, K. Li, W. Cao, G. ovich, J. Capmany, Multipurpose silicon photonics signal processor core, *Nature Communications*. 8, 2017.

[15] T. Ishihara, J. Shiomi, N. Hattori, Y. Masuda, A. Shinya and M. Notomi, An Optical Neural Network Architecture based on Highly Parallelized WDM-Multiplier-Accumulator, *IEEE/ACM Workshop on Photonics-Optics Technology Oriented Networking, Information and Computing Systems (PHOTONICS)*, pp. 15-21, 2019

[16] Y. Shen, N. C. Harris, S. Skirlo, M. Prabhu, T. Baehr-Jones, M. Hochberg, X. Sun, S. Zhao, H. Larochelle, D. Englund, and M. Solja£i¢, Deep learning with coherent nanophotonic circuits, *Nature Photon.*, 11(7), 441_446, 2017.

[17] K. V. Kumar, R. Rajasekar, N. Ayyanar and G. T. Raja, Photonic Crystal Mach-Zehnder Optical Switch Based on Phase Change Material, IEEE *20th International Conference on Nanotechnology (IEEE-NANO)*, 378-381, 2020.

[18] A.Narayan et *al*., PROWAVES: Proactive Runtime Wavelength Selection for Energy-efficient Photonic NoCs, *IEEE TCAD*, 2020

[19] H. T. Chen, J. Verbist, P. Verheyen, P. De Heyn, G. Lepage, J. De Coster, P. Absil, X. Yin, J. Bauwelinck, J. Van Campenhout, et al. High Sensitivity 10Gb/s Si Photonic Receiver Based on a Low-Voltage Waveguide-Coupled Ge Avalanche Photodetector. *Optics Express*, 23(2), 815–822, 2015.

[20] L. Chen, K. Preston, S. Manipatruni, and M. Lipson. Integrated GHz Silicon Photonic Interconnect with Micrometer-Scale Modulators and Detectors. *Optics Express*, 17(17), 15248–15256, 2009.

[21] C. Gunn. CMOS Photonics for High-Speed Interconnects. *IEEE Micro*,26(2), 58–66, 2006.

[22] Ian O'Connor and Gabriela Nicolescu. Integrated Optical Interconnect Architectures for Embedded Systems, *Springer Science & Business Media*, 2012.

[23] Yu-Hsiang Kao and H. Jonathan Chao, a bufferless photonic clos network-on-chip architecture. In *Fifth ACM/IEEE International Symposium on Networks on Chip (NoCS'11)*. *IEEE*, 81–88, 2011

[24] P. Ambs, A short history of optical computing: Rise, decline, and evolution. *Proceedings of SPIE - The International Society for Optical Engineering*, 7388, 2009.

[25] J. Touch, A.H. Badawy, V. Sorger, Optical computing. *Nanophotonics*, 6, 2017

[26] X. Chen *et al.*, The Emergence of Silicon Photonics as a Flexible Technology Platform, in *Proceedings of the IEEE*, 106(12), pp. 2101-2116, Dec. 2018,

[27] T. Ishihara, A. Shinya, K. Inoue, K. Nozaki and M. Notomi, An integrated optical parallel adder as a first step towards light speed data processing, *International SoC Design Conference (ISOCC), Jeju, Korea*, 123-124, 2016.

[28] J. Shiomi, T. Ishihara, H. Onodera, A. Shinya and M. Notomi, An Integrated Optical Parallel Multiplier Exploiting Approximate Binary Logarithms Towards Light Speed Data Processing, *IEEE International Conference on Rebooting Computing (ICRC), McLean, VA, USA*, 1-6, 2018

[29] Y. Tian, L. Zhang, and L. Yang, Electro-optic directed AND/NAND logic circuit based on two parallel microring resonators, *Opt. Express 20*, 16794-16800, 2012.

[30] L. Zhang, J. Ding, Y. Tian, R. Ji, L. Yang, H. Chen, P. Zhou, Y. Lu, W. Zhu, and R. Min, Electro-optic directed logic circuit based on microring resonators for XOR/XNOR operations, *Opt. Express 20*, 11605-11614, 2012.

[31] C.Condrat, P. Kalla, S. Blair. Logic synthesis for integrated optics. *GLSVLSI '11*, 2011.

[32] R. Wille, O. Keszocze, C. Hopfmuller and R. Drechsler, Reverse BDD-based synthesis for splitter-free optical circuits, *The 20th Asia and South Pacific Design Automation Conference*, 172-177, 2015.

[33] Z. Zhao, Z. Wang, Z. Ying, S. Dhar, R. T. Chen, and D. Z. Pan, Logic Synthesis for Energy-Efficient Photonic Integrated Circuits, In *Design Automation Conference, 22th Asia and South Paci_c. IEEE*, 2018.

[34] X. Sui, Q. Wu, J. Liu, Q. Chen and G. Gu, A Review of Optical Neural Networks, in *IEEE Access*, 8, pp. 70773-70783, 2020

[35] X. Y. Xu, M. X. Tan, B. Corcoran, J. Y. Wu, A. Boes, T. G. Nguyen, S. T. Chu, B. E. Little, D. G. Hicks, R. Morandotti, A. Mitchell, and D. J. Moss, 11 TOPS photonic convolutional accelerator for optical neural networks, *Nature*, 589, 44-51, 2021

[36] A. N. Tait, T. F. de Lima, E. Zhou, A. X.Wu, M. A. Nahmias, B. J. Shastri, and P. R. Prucnal, Neuromorphic photonic networks using silicon photonic weight banks, *Sci. Rep.*, 7(1), 2017

[37] Y. Shen, N. C. Harris, S. Skirlo, M. Prabhu, T. Baehr-Jones, M. Hochberg, X. Sun, S. Zhao, H. Larochelle, D. Englund, and M. Solja£i¢, Deep learning with coherent nanophotonic circuits, *Nature Photon.*, 11(7), 441_446, 2017.

[38] J. Feldmann, N. Youngblood, C. D. Wright, H. Bhaskaran, and W. H. P. Pernice, All-optical spiking neurosynaptic networks with self-learning capabilities, *Nature*, 569(7755), 208-214, 2019.

[39] A. Vajda, Multi-core and Many-core Processor Architectures, In *Book Programming Many-Core Chips*, 2011.

[40] R. Ho, K.W. Mai, and M. Horowitz, The Future of Wires, In *Proceedings of the IEEE*, 89(4), 2001.

[41] W. Bogaerts, S. K. Selvaraja, H. Yu, and et al., A Silicon Photonics Platform with Heterogeneous III-V Integration, In *Proceedings of Integrated Photonics Research, Silicon and Nano-Photonics*, 2011.

[42] C. A. Brackett, Dense wavelength division multiplexing networks: principles and applications, In *IEEE Journal on Selected Areas in Communications*, 8(6), pp. 948-964, 1990.

[43] D. Vantrease, N. Binkert, R. Schreiber and M. H. Lipasti, Light speed arbitration and flow control for nanophotonic interconnects, *42nd Annual IEEE/ACM International Symposium on Microarchitecture (MICRO)*, pp. 304-315, 2009.

[44] C. Li, M. Browning, P.V. Gratz, and S. Palermo, LumiNOC: A power-efficient, high performance, photonic network-on-chip. *IEEE Transactions on Computer Aided Design of Integrated Circuits & Systems*, 33(6), 826–838, 2014.

[45] L. Ramini, P. Grani, S. Bartolini, and D. Bertozzi, fcontrasting wavelength-routed optical NoC topologies for power-efficient 3D-stacked multicore processors using physical-layer analysis. In *Proceedings of the Conference on Design, Automation and Test in Europe*. EDA Consortium, 1589–1594, 2013.

[46] H. Li, A. Fourmigue, S. Le Beux, I. O'Connor and G. Nicolescu, Towards Maximum Energy Efficiency in Nanophotonic Interconnects with Thermal-Aware On-Chip Laser Tuning, in *IEEE Transactions on Emerging Topics in Computing*, 6(3), 343-356, 2018.

[47] Y. Pan, P. Kumar, J. Kim, G. Memik, Y. Zhang, A. Choudhary, Firefly: Illuminating Future Network-on-Chip with Nanophotonics, ACM SIGARCH Computer Architecture News, 37(3), 429–440, 2009.

[48] X. Chen, J. Feng, J. Xu, J. Zhang and S. Chen, Simultaneously Tolerate Thermal and Process Variations Through Indirect Feedback Tuning for Silicon Photonic Networks, in *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 40(7), 1409-1422, 2021,

[49] D. Dang, R. Mahapatra and E. J. Kim, PID controlled thermal management in photonic network-on-chip, *33rd IEEE International Conference on Computer Design (ICCD)*, 17-23, 2015.

[50] M. Kim, M.H. Kim, Y. Jo, H.K. Kim, S. Lischke, C. Mai, L. Zimmermann, W.Y. Choi, A Silicon Electronic-Photonic Integrated 25-Gb/s Ring Modulator Transmitter with a Built-in Temperature Controller. *Photonics Research.* 9, 2021.

[51] F. P. Sunny, A. Mirza, I. Thakkar, M. Nikdast and S. Pasricha, ARXON: A Framework for Approximate Communication Over Photonic Networks-on-Chip, in *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 29(6), 1206-1219, 2021.

[52] J. Lee, C. Killian, S. L. Beux, D. Chillet, Distance-aware Approximate Nanophotonic Interconnect, *ACM Transactions on Design Automation of Electronic Systems*, 27(2), 1-30, 2021.

[53] X. Wu et al., SUOR: Sectioned Undirectional Optical Ring for Chip Multiprocessor, *JETC*, 10, 1–25, 2014

[54] E. Fusella, A. Cilardo, Crosstalk-Aware Automated Mapping for Optical Networks-on-Chip, *ACM Transactions on Embedded Computing Systems*. 16, 1-26, 2016.

[55] S. Le Beux *et al.*, Chameleon: Channel efficient Optical Network-on-Chip, *Design, Automation & Test in Europe Conference & Exhibition (DATE)*, pp. 1-6, 2014.

[56] S. Abdollahramezani, O. Hemmatyar, H. Taghinejad, A. Krasnok, Y. Kiarashinejad, M. Zandehshahvar, A. Alù, A. Adibi, Tunable nanophotonics enabled by chalcogenide phase-change material, *Nanophotonics*, 9(5), 1189-1241, 2020.

[57] J. Zheng, A. Khanolkar, P. Xu, S. Colburn, S. Deshmukh, J. Myers, J. Frantz, E. Pop, J. Hendrickson, J. Doylend, N. Boechler, and A. Majumdar, GST-on-silicon hybrid nanophotonic integrated circuits: a non-volatile quasi-continuously reprogrammable platform, *Opt. Mater. Express* 8, 1551-1561, 2018.

[58] M. Rudé, J. Pello, R. E. Simpson, et al., Optical switching at 1.55 µ m in silicon racetrack resonators using phase change materials, *Appl. Phys. Lett.,* 103(14), 141119, 2013.

[59] M. Stegmaier, C. Ríos, H. Bhaskaran, C. D. Wright, and W. H. Pernice, Nonvolatile all-optical 1×2 switch for chipscale photonic networks, *Adv. Opt. Mater.,* 5(1), 1600346, 2017.

[60] C. Wu, H. Yu, H. Li, X. Zhang, I. Takeuchi, and M. Li, Low-loss integrated photonic switch using subwavelength patterned phase change material, *ACS Photonics,* 6(1), pp. 87–92, 2018.

[61] D. Tanaka, Y. Shoji, M. Kuwahara, et al., Ultra-small, selfholding, optical gate switch using $Ge_2Sb_2Te_5$ with a multimode Si waveguide, *Optics Express*, 20(9), pp. 10283–10294, 2012.

[62] M. Stegmaier, C. Rıos, H. Bhaskaran, and W. H. Pernice, Thermo-optical effect in phase-change nanophotonics, *ACS Photonics*, 3(5), 828–835, 2016.

[63] C.K. Kato, M. Kuwahara, H. Kawashima, T. Tsuruoka, and H. Tsuda, Current-driven phase-change optical gate switch using indium–tin-oxide heater, *Appl. Phys. Express,* 10(7), 072201, 2017.

[64] J. Zheng, S. Zhu, P. Xu, S. Dunham, and A. Majumdar, Modeling electrical switching of nonvolatile phase-change integrated nanophotonic structures with graphene heaters, *ACS Appl. Mater. Interfaces*, 12(19), 2020.

[65] P. Xu, J. Zheng, J. Doylend, A. Majumdar, Low-Loss and Broadband Nonvolatile Phase-Change Directional Coupler Switches, *ACS Photonics* ,6(2), 553 - 557, 2019.

[66] C. Rios, P. Hosseini, C. D. Wright, H. Bhaskaran, and W. H. Pernice, On-chip photonic memory elements employing phase-change materials, *Adv. Mater.,* 26(9), 1372–1377, 2014.

[67] W. H. Pernice and H. Bhaskaran, Photonic non-volatile memories using phase change materials, *Appl. Phys. Lett.,* 101(17), 171101, 2012.

[68] C. Ríos, M. Stegmaier, P. Hosseini, et al. Integrated all-photonic non-volatile multi-level memory, *Nature Photon 9*, 725–732, 2015.

[69] J. Feldmann, M. Stegmaier, N. Gruhler, et al., Calculating with light using a chip-scale all-optical abacus, *Nat. Commun.,* 8(1), p. 1256, 2017.

[70] C. Ríos, N. Youngblood, Z. Cheng, et al., In-memory computing on a photonic platform, *Sci. Adv.,* 5(2), 5759, 2019.

[71] J. Feldmann, N. Youngblood, C. Wright, H. Bhaskaran, and W. Pernice, All-optical spiking neurosynaptic networks with selflearning capabilities, *Nature,* 569(7755), 208, 2019.

[72] Z. Zhao, Z. Wang, Z. Ying, et al. Logic synthesis for energy-efficient photonic integrated circuits, *23rd Asia and South Pacific Design Automation Conference (ASP-DAC)*, pp. 355-360, 2018.

[73] N. Yamada et al., Rapid, phase transitions of GeTe-Sb$_2$Te$_3$ pseudobinary amorphous thin films for an optical disk memory. *Journal Applied Physics*, 1991.

[74] M. Bahadori et al., Crosstalk penalty in microring-based silicon photonic interconnect systems, *J. Lightwave Technol.*, 2016

[75] C. Sun et al., DSENT - A Tool Connecting Emerging Photonics with Electronics for Opto-Electronic Networks-on-Chip Modeling*, IEEE/ACM Sixth International Symposium on Networks-on-Chip*, Lyngby, Denmark, pp. 201-210, 2012.

[76] J. Lee et al., Approximate nanophotonic interconnects. *IEEE/ACM NOCS*, 2019.

[77] Goswami, Debabrata. (2003). Optical computing. Resonance. 8. 56-71

[78] P. Vivet et al., IntAct: A 96-Core Processor With Six Chiplets 3D-Stacked on an Active Interposer With Distributed Interconnects and Integrated Power Management, *IEEE Journal of Solid-State Circuits*, 2021

[79] A.H. Atabaki, S. Moazeni, F. Pavanello, et al. Integrating photonics with silicon nanoelectronics for the next generation of systems on a chip, *Nature 556*, pp. 349–354, 2018.

[80] S. Xu, J. Wang, R. Wang, J. Chen, and W. Zou, High-accuracy optical convolution unit architecture for convolutional neural networks by cascaded acousto-optical modulator arrays, *Opt. Express 27*, 19778-19787, 2019.

[81] R. A. Minasian, Photonic signal processing of microwave signals, *IEEE Transactions on Microwave Theory and Techniques,* 54(2), pp. 832-846, 2006.

[82] Z. Zhao, Z. Wang, Z. Ying, S. Dhar, R. T. Chen and D. Z. Pan, Optical computing on silicon-on-insulator-based photonic integrated circuits, *IEEE 12th International Conference on ASIC (ASICON)*, Guiyang, China, pp. 472-475, 2017.

[83] Z. Ying, Z. Wang, Z. Zhao, S. Dhar, D. Pan, R. Soref, and R. Chen, Silicon microdisk-based full adders for optical computing, *Opt. Lett. 43*, 983-986 , 2018.

[84] Beausoleil, Ray, et al. A nanophotonic interconnect for high-performance many-core computation, Integrated Photonics and Nanophotonics Research and Applications. Optical Society of America, 2008