Energy-Efficient Photonic Neuromorphic Computing for Telecommunication Applications

Energie-efficiënt fotonisch neuromorf rekenen voor telecommunicatietoepassingen

Andrew Katumba

Promotoren: Prof. Dr. Ir. Peter Bienstman, Prof. Dr. Ir. Joni Dambre
Proefschrift tot het bekomen van de graad van
Doctor in de Ingenieurswetenschappen: Fotonica

Vakgroep Informatietechnologie
Voorzitter: Prof. Bart Dhoedt
Faculteit Ingenieurswetenschappen en Architectuur
Academiejaar 2018-2019

GHENT UNIVERSITY

GHENT UNIVERSITY

Universiteit Gent
Faculteit Ingenieurswetenschappen en Architectuur
Vakgroep Informatietechnologie

Promotoren: Prof. Dr. Ir. Peter Bienstman
Prof. Dr. Ir. Joni Dambre

Examencommissie:		
	Prof. Dr. Ir. Filip de Turck (voorzitter)	Universiteit Gent, INTEC
	Prof. Dr. Ir. Peter Bienstman (promotor)	Universiteit Gent, INTEC
	Prof. Dr. Ir. Joni Dambre (promotor)	Universiteit Gent, ELIS
	Dr. Ir. Thomas Van Vaerenbergh	Hewlett Packard Enterprise, USA
	Prof. Dr. Ir. Guy Van der Sande	Vrije Universiteit Brussel
	Prof. Dr. Ir. Xin Yin	Universiteit Gent, INTEC
	Prof. Dr. Ir. Geert Morthier	Universiteit Gent, INTEC
	Prof. Dr. Ir. Tom Dhaene	Universiteit Gent, INTEC

Universiteit Gent
Faculteit Ingenieurswetenschappen en Architectuur

Vakgroep Informatietechnologie
Technologiepark-Zwijnaarde, 9052 Gent, België

Tel.: +32-9-331.49.00
Fax.: +32-9-331.48.99

Acknowledgements

Any Ph.D. thesis sits atop a pyramid of effort and support from many people; mine is no exception. First and foremost, I would like to thank my promoters Prof. Peter Bienstman and Prof. Joni Dambre for the generous advice and guidance from the onset of my research. I am grateful for the freedom and trust you accorded me to explore and experiment extensively. I appreciate the many opportunities you have created for me to develop my career professionally and to grow as a person. I really couldn't have wished for better guides on this journey.
To my Ph.D. defense jury members, I am grateful for your concise feedback that has helped to shape-up this thesis to its current quality.

I would like to acknowledge the steering team of the Photonics Research Group, with Prof. Roel Baets at the helm. Together with professors Wim, Geert, Bart, Gunther, Nicola and Dries, my work was positively impacted by a short conversation here and a nudge there over the years. I am also thankful to the support staff: Ilse Meersman, Ilse Van Royen, Jasper, Bert, Kristien, Mike, Michael, Peter, and Jeroen, for your aid in all matters admin and technical.

Turning to the PRC group, we have spent a lot of time in meetings poking and prodding at the various nuances and complexities of reservoir computing. My predecessors Thomas, Martin, and Bendix, your assistance when I was just striking out centered and grounded the basis upon which the rest of my research was built. Martin and Thomas, thanks for bearing with my flood of Caphe questions and feature requests, even long after you had left the group. As for the current generation, Matthias and Floris, the timing of your entry could not have been more fortuitous. Our discussions and collaboration have without any doubt born numerous fruit; all the best as you wind up. And to the "new kids on the block"Stijn, Chonghuai, and Emmanuel, you have brought with you new energy and a fresh take on matters. I am looking forward to seeing where you will take this burgeoning field of photonic reservoir computing.

I would like to appreciate all the collaborators I had the chance to work with both within and without Gent University. I am especially grateful to the members of the Horizon 2020 PHRESCO project. I have benefitted tremendously from your technical expertise and our discussions that have shaped my work, especially in the last 2 years.
I would also like to single-out Prof. Xin Yin and Prof. Johan Bauwelinck of

IDLAB for their important insights, both and the beginning and towards the end of my Ph.D., into how reservoir computing might fit into a telecom setting.

Working in a big research group comes with opportunities for a variety of interactions, both at and outside work. Sanja, Sarah, and Yanlu, we have been office mates since the Technicum days, thanks for the convivial work environment. The same goes for you Nayerra, Stijn and Yuting.
On the research side, I had the pleasure of collaborating with various people. Amin, Abdul, Kasper, Jochem and Joris, it was always exciting to push the limits of both devices and equipment. I also recall numerous thought-provoking conversations with Andreas, Jesper, Paul, Jing, Koen, Utsav, Haolan, Antonio, Sören, Thijs and Eva. Thanks for the willingness to share your time and expertise.
As for the De Brug and, later on, the 5^{th} floor lunch group, lunch was always packaged with some pretty interesting discussions on just about everything (most interesting, some bizarre and occasionally both).
Outside the office, Jesper, Paul, Joan, Tommaso, Mahmoud, Sanja, Nina, Alessio, Margherita, Savvina, Ana, Camiel, Irfan, Jing, Nuria, Fabio, Floris, Grigorij, Mattias, Clemens, Hanna, Alexandros and Alejandro, those after-work and weekend meet-ups served for quality downtime. Not forgetting Sulakshna and Saurav the impeccable culinary duo. And for those that made the trip to Uganda in 2017, my family and I look back fondly to those times.

No sooner had I arrived in Belgium than I was adopted by the Okumus, the Njugunas, the Ssimbas, the Mugishas, and the Asiyas. I am eternally thankful for your hospitality and regard, and for making sure that I was properly settled-in and, in a way, like I never left home.

Thanks, Michael, Gloria, Billy, Felly, Marvin, Ray, Agaba, Gina, Kaka, Tori, Ayor, Cathy L., Brenda, Malcolm, Esther, Eric, Cosmas and Onesmus for always making the long periods between our meetings seem inconsequential. Michael, your annual visits were inexplicably timed at the moments when I needed an interlude from the rigor of research.
The Nkrumah company FX, Michael, Kimera, Deno, Dero, Danny, Paulo, Bukenya and Mugagga, I am humbled by the solidarity and brotherhood. Gracias for always coming through when it matters.
Carlos, Paula, and David Owek, I am indebted to you for always challenging me to dream bigger and continually strive for excellence. Carlos and Paula, aren't we overdue for another "dinner for schmucks"?

A great many people have had an impact on the success of my research and personal well-being than I could enumerate here. The absence of your name does not diminish my gratitude for your generosity.

Finally, I would like to thank the GBS and Kyeyune families who have been with me at every turn. Your love, guidance, and support have been the beacon for this journey.

Gent, February 2019
Andrew Katumba

Table of Contents

List of Figures

List of Tables

List of Acronyms

A

ADC	Analog-to-Digital Converter
ANN	Artificial Neural Network

B

BP	Backpropagation
BER	Bit error rate
BPDC	Backpropagation-Decorrelation
BTO	Barium Titanate

C

CMOS	Complementary metal oxide semiconductor
CD	Chromatic Dispersion
CW	Continuous wave

D

DSP	Digital Signal Processor
DBP	Digital Backpropagation
DFE	Decision Feedback Equalizer
DFB	Distributed Feedback

E

EDFA	Erbium Doped Fiber Amplifier
EAM	Electro-absorptiion Modulator

F

FDTD	Finite-Difference Time-Domain
FFE	Feed Forward Equalizer
FFNN	Feed Forward Neural Network
FIR	Finite Impulse Response

H

HD-FEC	Hard Decision Forward Error Correction

I

IBM	International Business Machines Corporation
IM-DD	Intensity Modulation - Direct-Detection
IoT	Internet of Things

M

ML	Machine Learning
MIT	Metal-Insulator Transition

N

NRZ	Non-return-to-zero
NEP	Noise Equivalent Power

O

OOK	On-Off keying
OBPF	Optical Bandpass Filter s

P

PhRC	Photonic Reservoir Computing
PIC	Photonic Integrated Circuit
PMD	Polarization Mode Dispersion

R

RC	Reservoir Computing
RNN	Recurrent Neural Network

S

SOA	Semiconductor Optical Amplifier
SOI	silicon on insulator
SNN	Spiking Neural Network
SNR	Signal-to-noise ratio
Si	Silicon
SD-FEC	Soft Decision Forward Error Correction
SPM	Self Phase Modulation
SSMF	Standard single-mode fiber

T

TE	Transverse Electric
TM	Transverse Magnetic

V

VO_2	Vanadium Oxide

Nederlandse samenvatting –Summary in Dutch–

Telecommunicatie en computers zijn de twee meest toonaangevende voorbeelden van technologische vooruitgang in onze moderne maatschappij. Hun impact op ons leven is zo wijdverbreid dat het ondenkbaar zou zijn dat er een dag voorbij zou gaan zonder dat we gebruik maken van een slimme telefoon of een ander gelijkaardig toestel. In belangrijke mate steunen alle facetten van onze moderne maatschappij (gezondheidszorg, financies, zakenleven, onderwijs, ...) op vooruitgang in deze domeinen.

Maar hoewel onze maatschappij deze technologieën omarmd heeft, heeft de onderliggende infrastructuur last om gelijke tred te houden met de als maar groeiende behoefte aan meer informatie, meer gegevens, meer rekenkracht. Er wordt veel werk verricht om iedere laatste gram capaciteit uit de onderliggende digitale elektronische infrastructuur te persen, en men probeert nieuwe manieren uit om deze systemen te bouwen, of om belangrijke flessenhalssubsystemen te vervangen.

Als inspiratie voor oplossingen zijn sommige mensen op zoek naar antwoorden in de manier waarop ons menselijk brein werkt. Ons brein heeft vaardigheden in communicatie en informatieverwerking die voor sommige toepassingen efficiënties halen ver boven die van de meest geavanceerde supercomputers. Een mogelijke aanpak die succesvol gebleken is, is om kunstmatige neuronale structuren te bouwen en die aan te passen via algoritmes die oplossingen genereren op basis van gekende (input, output) paren, en die dan toe te passen om outputs te creëren voor nieuwe, ongeziene inputs. De manier waarop deze computationele structuren getraind worden om taken op te lossen verschilt drastisch van de manier waarop traditionele digitale elektronische informatieverwerking plaatsgrijpt, en maakt gebruik van een verzameling van technieken die collectief bestempeld worden als machinaal leren.

In dit doctoraatsonderzoek gebruiken we fotonica om deze artificiële neurale netwerken te bouwen. Het fotonicaplatform bezit een gigantische bandbreedte, voor de hand liggende paden naar parallellisatie en zou significante verbeteringen kunnen brengen in het energieverbruik om bepaalde taken op te lossen. Een andere belangrijke eigenschap van dit werk is dat het gebaseerd is op het paradigma van analoog rekenen, en het kan daarom de flessenhalzen vermijden die typisch zijn voor de standaard von Neumann architecturen. We focussen op het toepassen van de combinatie van fotonica en machinaal leren om optische hogesnelheidssignalen

te verwerken voor telecommunicatie, op een manier die uitsteekt boven bestaande elektronische implementaties, vooral op het gebied van energieëfficiëntie.

1 Fotonica

Fotonica als wetenschappelijke discipline is begaan met het ontwikkelen van componenten en systemen om licht op te wekken, te manipuleren en te detecteren voor specifieke toepassingen. Toepassingen voor de fotonica situeren zich courant in de domeinen van metingen, beeldopname, beeldschermtechnologie, telecommunicatie, niet-invasieve chirurgie, gegevensopslag, ...

Een van de redenen voor het commerciële succes van fotonische technologie is het feit dat een aantal functies op silicium chips geïntegreerd kunnen worden aan de hand van CMOS technologie. Dit leidt tot compacte systemen, schaalvoordelen wat betreft productie en verdeling, en vele andere voordelen.

2 Optische communicatie

Informatieoverdracht over lange afstanden via licht is ongetwijfeld een van de grootste successen van fotonische technologieën. Aangespoord door de uitvinding van efficiënte bronnen (zoals lasers), het terugdringen van vezelverliezen en de vooruitgang in het manipuleren van optische signalen, is het zo dat vandaag het versturen van informatie over de meeste lengteschalen hoofdzakelijk met licht gebeurt.

Maar hoewel licht de koning is voor transmissie, blijft digitale elektronica nog steeds de belangrijkste manier om informatie te verwerken, hetzij informatie in beweging (wat het geval is voor optische telecomsignalen), hetzij in rust (zoals in digitale computers). Dit impliceert dat er voor optische communicatiesystemen heen en weer geschakeld moet worden tussen fotonische en elektronische voorstellingen van de signalen, wat leidt tot snelheidsflessenhalzen die ook gepaard gaan met extra energieverbruik.

In dit doctoraatsonderzoek gebruiken we fotonische reservoircomputing als een platform om diverse functies te implementeren voor optische signaalverwerking, en stellen we verschillende wijzigingen en richtlijnen voor om dit platform toe te passen in optische telecommunicatie met maximale uitbuiting van hun energieëfficiëntievoordeel.

3 Fotonisch neuromorf rekenen

Een fotonisch neuromorf rekensysteem bouwen kan typisch gebeuren op twee verschillende manieren. Aan de ene kant hebben we systemen die gebruik maken van componenten die gepiekte signalen opwekken en verwerken. Aan de andere kant zijn er de meer analoge fotonische rekensystemen die typisch een nietlineaire transformatie uitvoeren gevolgd door een uitleesstap. In dit werk richten we

ons op de tweede aanpak met een willekeurig geïnitialiseerd netwerk genaamd het reservoir, en een getrainde uitlezing, nl. een eenvoudig model dat een lineaire combinatie neemt van de uitgangen van het reservoir.

Hoewel neuromorfe fotonica nog in de kinderschoenen staat, hebben zowel de academische als de industriële wereld het enorme potentieel ingezien van deze techniek als een alternatief met hogere bandbreedte en lager energieverbruik dan traditionele elektronische systemen. Getuige hiervan is het groeiend aantal pogingen in de laatste jaren om volledige neuromorfe systemen te bouwen om de beperkingen van elektronische gegevensverwerking te omzeilen. Er is echter nog steeds geen volledige demonstratie van een aanpak die totaal compatibel is met de vereisten voor gebruik buiten een laboratoriumcontext. In dit werk willen we stappen zetten in de richting van een geïntegreerd fotonisch neuromorf rekensysteem dat zich richt op telecomtoepassingen.

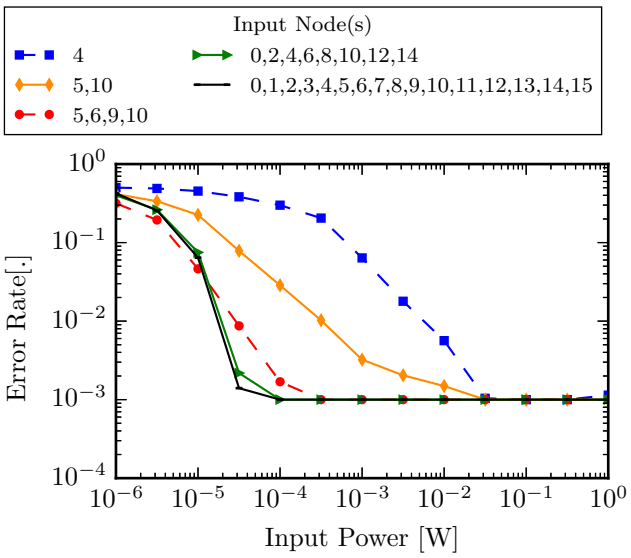
Dit doctoraatsonderzoek valt uiteen in drie thema's. Ten eerste, het verbeteren van de energieëfficiëntie van fotonische neuromorfe rekensystemen. Ten tweede, deze toepassen op telecommunicatieproblemen. Ten derde, werken in de richting van een grootschalig experimenteel prototype.

Deel I: Verbetering van de energieëfficiëntie van fotonische reservoirs

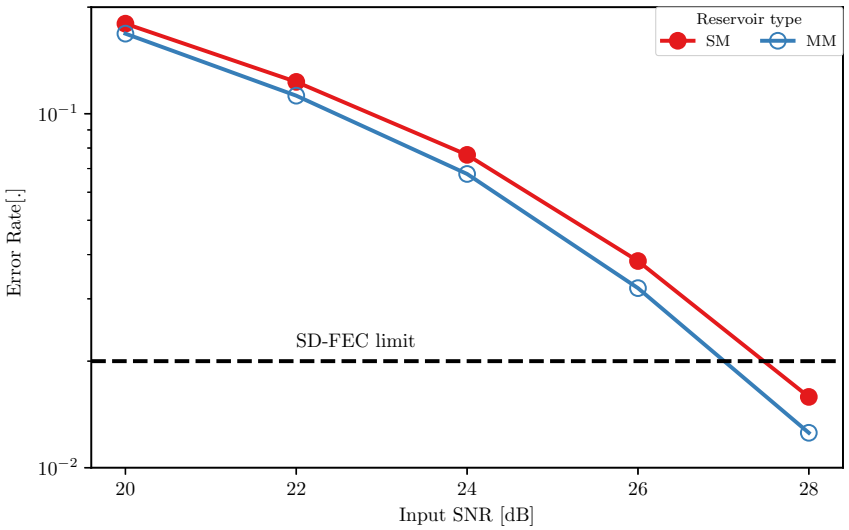
We vertrekken van het passieve reservoirsysteem, nl. een vereenvoudigd fotonisch reservoirrekensysteem waar de nietlineariteit verplaatst is naar de detector. Dit ontkoppelt de verwerkingssnelheid van de snelheid van het ingangssignaal, wat de energieëfficiëntie verbetert.

Om te beginnen, onderzoeken we de beste manier om het te verwerken signaal in het reservoir te krijgen. We komen er achter dat het veel efficiënter is om, in plaats van alle vermogen in een enkele knoop te injecteren zoals gebeurde in het verleden, het vermogen op te splitsen en het te injecteren in een verzameling van ingangsknopen. Zoals we tonen in Figuur 1, is er minder ingangsvermogen vereist voor foutloze prestaties wanneer het vermogen opgesplitst wordt over meerdere knopen, vergeleken met de situatie waarbij al het vermogen in een enkele knoop geïnjecteerd wordt.

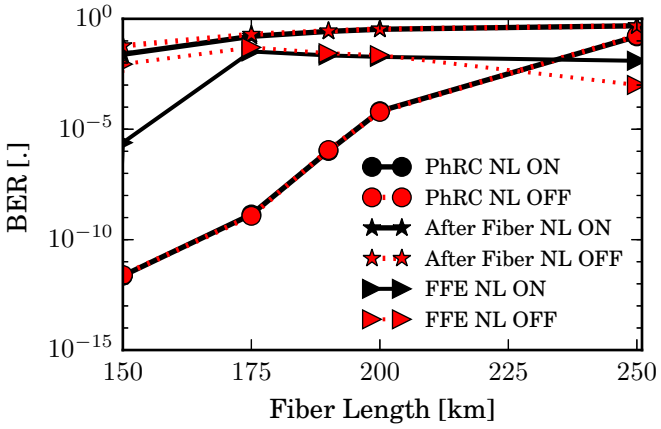
Vervolgens gebruiken we numerieke simulaties om het vermogenverbruik te bestuderen van een reservoir bestaande uit multimode eerder dan de typische monomode componenten. We tonen aan dat de extra vrijheidsgraden geïntroduceerd door het propageren van meerdere modes ons in staat stellen om een reservoir te ontwerpen dat een lagere signaalruisverhouding voor foutloze prestaties vereist dan zijn monomode evenknie. Figuur 2 is een illustratief voorbeeld voor de driebitshoofdingsherkenningstaak voor telecom.



Figuur 1: Foutdebiet vs totaal ingangsvermogen voor verschillende injectiescenarios. De kleinst meetbare fout, gegeven het aantal bits gebruikt voor testen, is 10^{-3}.



Figuur 2: Foutdebiet voor monomode en multimode reservoirs voor de driebitshoofdingsherkenningstaak voor verschillende waarden voor de signaalruisverhouding aan de ingang.



Figuur 3: Bitfoutdebiet voor de neuromorfe fotonische compensator vergeleken met die van een FIR Feed Forward Equalizer (FFE) voor verschillende vezellengtes met een lanceringsvermogen van $15\,\mathrm{mW}$. *NL ON - nietlineaire propagatie aan. NL OFF - nietlineaire propagatie af.*

Deel II: Een neuromorfe nietlineaire compensatiecomponent in siliciumfotonica

Daarna onderzoeken we aan de hand van simulaties de performantie van een neuromorfe nietlineaire compensatiecomponent in siliciumfotonica gebaseerd op fotonisch reservoirrekenen. We tonen aan dat deze compensator verschillende grootteordes beter presteert dan klassieke FFE-gebaseerde implementaties voor metroverbindingen (zie Figuur 3) en dat deze ook goed presteert in hogesnelheidkorteafstandsverbindingen. Deze component is zo gepositioneerd dat hij een valabele complementaire rol kan spelen wat betreft DSP-gebaseerde oplossingen, wat resulteert in een reductie van hun complexiteit of vermogenverbruik, of hen zelfs volledig zou kunnen vervangen.

Deel III: Naar een experimentele demonstratie van een neuromorfe fotonische component met uitlezing op de chip

We hebben vooruitgang geboekt bij de ontwikkeling van een demonstratiechip gebaseerd op fotonisch reservoirrekenen, waar elektronische en fotonische elementen samen geïntegreerd worden, en met een groter aantal knopen dan de state-of-the-art. De bedoeling van de demonstrator is om de elektro-optische eigenschappen uit te buiten van nieuwe materialen zoals vanadium dioxide (VO_2) en bariumtitanaat (BTO), voor de implementatie van een uitleessysteem op dezelfde chip als het reservoir, een uitleessysteem dat bovendien niet volatiel is en in staat om in echttijd te werken. Bedoeling is om een volledig platform te genereren dat de com-

putationele mogelijkheden van reservoirrekenen kan uitbuiten voor toepassingen in hogesnelheidstelecommunicatie op industriële schaal.

We hebben een eerste generatie van demonstratiechip ontworpen die een reservoir van 18 knopen combineert met siliciumstructuren om de verschillende geïntegreerde uitleessystemen te verkennen. Na fabricage hebben we de componenten uitgemeten en hebben we vooruitgang geboekt met de initiële experimentele karakterisatie van het volledige fotonisch reservoirrekensysteem. Ondanks het feit dat de volledige systeemkarakterisatie nog bezig is, hebben de initiële resultaten ons in staat gesteld om vooruitgang te boeken met de uitleestechnologiëen en om richtlijnen te formuleren die we hebben toegepast in het ontwerp van de tweede generatie van chip met 60 knopen.

4 Conclusie

Samengevat hebben we aangetoond dat we de energieëfficiëntie van fotonische reservoirs kunnen verbeteren door 1) een zorgvuldige keuze van de strategie waarmee het ingangssignaal geïnjecteerd wordt in het reservoir 2) de opbouw van het reservoir met multimode in plaats van monomode componenten. Bovendien hebben we aangetoond dat een passieve fotonische chip gebaseerd op reservoirrekenen kan ingezet worden als een nietlineaire compensator voor optischevezelcommunicatieverbindingen.

Deze evoluties zijn een belangrijke stap om neuromorfe fotonische rekensystemen aan te tonen voor de volgende generaties van informatieverwerking en telecommunicatie.

English summary

Telecommunication and computing are some of the most significant technological advances in modern day society. Their impact on our lives is so wide-spread to the level that it is inconceivable that a day would go by without one making use of a phone or some computing device for one purpose or another. To a major extent, all facets of our modern society, health, finance, business, education, etc., heavily rely on their advancements.

But while society has embraced these technologies, the supporting infrastructure is struggling to keep up with the ever-growing needs for more information, more data, more computational power. At the precipice of squeezing every last ounce of capacity out of the underlying digital electronic infrastructure, there is a lot of on-going work that intends to find alternative ways to build these systems, or to replace core bottleneck subsystems/components.

For inspiration for answers, some solutions seek inspiration from the way the human brain operates. The human brain has communication and computing powers operating at efficiencies that trump even the most advanced supercomputers today for certain applications. One approach that has been successful, is building artificial neuron structures and adapting them using algorithms towards producing solutions from known input-output pairings, and then apply them to create outputs for new unseen inputs. This way of training these computational structures to solve tasks deviates significantly from the way digital electronic computers and information processors work, and makes use of a body of techniques collectively known as machine learning.

In this PhD research we use photonics as the platform to build these artificial neural networks. The photonics platform has tremendous bandwidth, easy paths to parallelization and could yield significant improvements in the power required to solve tasks. Another key attribute of this work is that it is based on the analog computing paradigm and therefore can circumvent the bottlenecks of the standard von-Neumann architecture. We focus on applying the combination of photonics and machine learning techniques for the purpose of processing high speed optical telecom signals in a way that surpasses existing electronic implementations, especially regarding energy efficiency.

5 Photonics

Photonics as a scientific field concerns itself with developing devices and systems to generate, modify and detect light for specific applications. Photonics applications are routinely encountered in sensing, imaging and display technology, telecommunications, non-invasive surgery techniques, data storage, etc.

One of the reasons for the commercial success of photonics technology is that a decent number of functions can be integrated on silicon chips using CMOS technology. This leads to compact systems, economies of scale with regards to production and distribution, and numerous other benefits.

6 Optical Communications

Transmitting information over long distances using light as a medium is arguably one of the biggest successes of photonic technology. Spurred by the invention of efficient sources (such as lasers), reduction in fiber losses, and advances in techniques for manipulating optical signals, the transmission of signals for any reach today is predominantly done with light. But while light is king for transmission, digital electronics remains the prevalent way to process data; whether in motion (such as is the case on optical telecom data) or at rest (as in digital computers). This, for optical communications systems, implies going back and forth between photonic and electrical representations of signals, leading to speed bottlenecks that are also associated with extra energy requirements.

In this PhD research, we take photonic reservoir computing as a platform for implementing various functions on optical signals, and make various adaptations and guidelines towards applying it to optical telecommunications while maximizing its energy efficiency.

7 Photonic Neuromorphic Computing

Building a neuromorphic computing platform in photonics typically involves taking one of two paths. On one hand we find systems making using of components that generate and process streams of spikes. On the other hand are the more analog computing photonic systems that typically involve some nonlinear transformation followed by a readout. For this work, we focus on the latter approach with a randomly initialized network called the reservoir and a trained readout, which is a simple model taking weighting linear combinations of the reservoir outputs.

While neuromorphic photonics is still a nascent field, its tremendous promise for larger bandwidth and low energy consumption alternatives to electronic processing systems has been recognized by both research and industry. As evidence, in the past few years, there has been an increasing number of attempts at building complete neuromorphic systems to overcome the limitations of electronic processing. However, there is yet to be a full demonstration of an approach that fully

operates within the requirements suitable for deployment in a real world setting, beyond lab demonstrators. In this work, we provide steps towards an integrated neuromorphic photonic computing platform targeting telecom.

The PhD research can be summarized in three key themes, first improving the energy efficiency of photonic neuromorphic computing systems, and then applying it to a telecom problem and finally work towards a large-scale experimental prototype.

Part I: Improving the energy efficiency of photonic reservoirs

We start with the passive reservoir computing system, i.e. a simplified photonic reservoir computing system that moves the nonlinearity to the detector, which decouples the speed of processing from the speed of the input signal, and make improvements towards its power efficiency.

First, we investigate the best way to get the signal to be processed into the reservoir. We find that rather than injecting all the power into the reservoir at a single node as was done in the past, it is much more efficient to split the power and inject it at a collection of inputs nodes. As we illustrate in Figure 4, lower input power is required for error-free performance when splitting the power over multiple nodes than when inject it all into a single node.

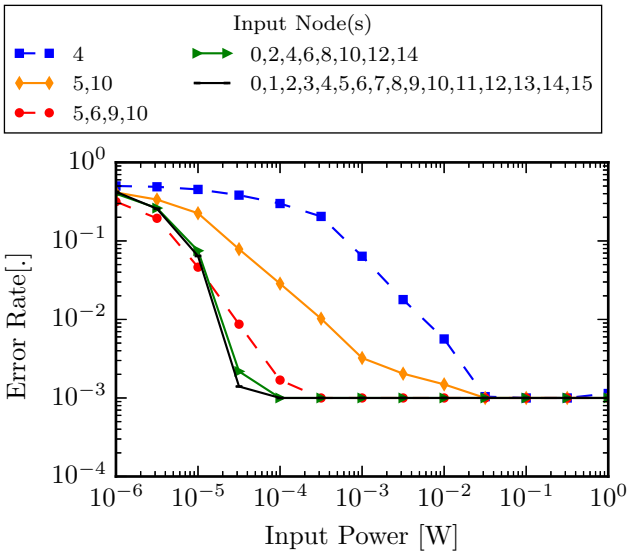


Figure 4: Error rate vs total input power for different injection scenarios. The minimum measurable error, given the number of bits used for testing, is 10^{-3}.

Next, we use numerical simulations to study the power consumption of a reservoir composed of multimode rather than the typical single-mode components. We

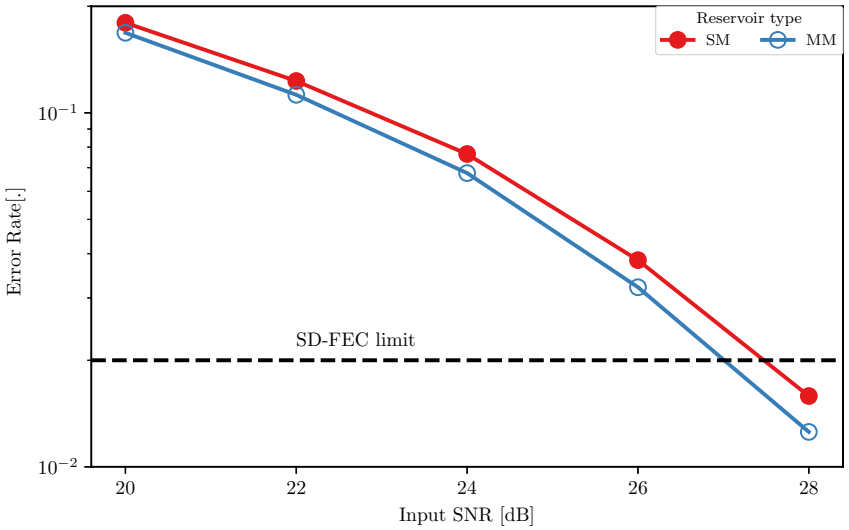


Figure 5: Error rates for single-mode and multimode on the 3 bit header recognition task for different values of the input SNR.

show that the extra degrees of freedom introduced by propagating with more than one mode enable us to design a reservoir that requires a lower input signal-to-noise ratio, than its single mode counterpart, for error-free performance. Figure 5 is an illustrative example for the 3-bit header recognition telecom task.

Part II: A neuromorphic silicon photonics nonlinear compensation device

We then investigate, through simulations, the performance of a neuromorphic silicon photonics nonlinear compensator based on photonic reservoir computing. We show that this neuromorphic equalizer can outperform FFE-based implementations by a few orders of magnitude for metro links (see Figure 6) and that it also performs well in high-speed short-reach applications. The device is positioned to be a viable complement to DSP-based solutions (resulting in a reduction in their complexity and power consumption) or as a replacement.

Part III: Towards experimental demonstration of a neuromorphic photonics device with on-chip readout

We have made progress on the development of a photonic reservoir computing demonstrator chip featuring co-integrated photonic and electronic components, with a larger number of nodes than the state-of-the-art. The demonstrator is intended to leverage the electro-optical properties of the novel materials vanadium-

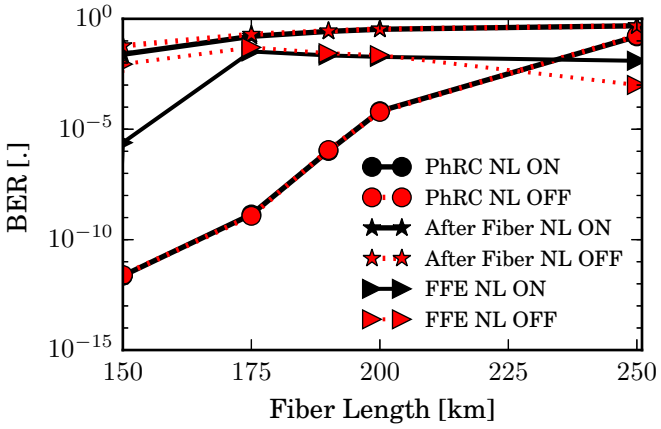


Figure 6: Bit Error Rate of the neuromorphic photonic equalizer compared to that of an FIR Feed Forward Equalizer (FFE) for different fiber lengths with a launch power 15 mW. *NL ON - nonlinear propagation. NL OFF - nonlinear propagation is deactivated.*

dioxide (VO_2) and barium-titanate (BTO) to implement a real-time capable, non-volatile readout on the same chip as the reservoir. The goal is to generate a complete platform that can exploit the computational capabilities of RC for real-world high-speed telecom applications.

We designed and taped out a first-generation demonstrator chip combining an 18 node reservoir together with silicon structures to be used to explore the integrated readouts. After fabrication, we characterized the components and made progress with the initial experimental measurements of the full photonic reservoir computing system. While the full system characterization is on-going, with the initial results we were able to make progress on development of the readout technologies and obtain guidelines that we have applied in the design of a second-generation 60-node chip.

8 Conclusion

In conclusion, we show that we can improve the energy efficiency of photonic reservoirs by i) carefully selecting the strategy with which the input signal is injected into the reservoir and ii) by constructing the reservoir from multimode components rather than the typical single mode components. Furthermore, we show that a passive photonic reservoir chip can be operated as a nonlinear compensation device for fiber optic communication links. These findings are key steps towards demonstrating neuromorphic photonic computing systems for the the next generation of information processing and computing systems for telecommunication.

1 Introduction

1.1 Introduction

Over the past decades, humanity's information processing needs have accelerated at a tremendous pace to overwhelm the capabilities of the underlying support infrastructure. In light of the stumbling Moore's law, gate integration densities have stagnated, and the von-Neumann architecture that has been the stalwart of the digital revolution can no longer be relied on to cope with the exponential growth of the data consumption requirements of the modern world. This stumbling block for CMOS technology has re-ignited interest in analog computing platforms that depart from the von-Neumann approach to information processing.

Analog computing, unlike conventional computing, relies on the information processing capabilities of certain physical systems. The processing usually hinges on the evolution of the internal state of the dynamical behavior of a physical system in response to an appropriately encoded input. Analog information processing platforms are typically coupled with an adaptation and readout system to guide the evolution of the system state (or a subset of it) towards a solution whose accuracy metric is usually known beforehand. This is then followed by a readout that transforms the observed physical signal of interest into a form that is suitable for interpretation by digital computers. Machine learning is a prevalent way of implementing the adaptation of the computational system's internal states to extract the desired solution .

One important class of machine learning techniques are the so-called artificial

neural networks (ANNs), that consist of a network of interconnected computational units, dubbed "artificial neurons". The layout and operation of the ANN is inspired by the structure and information processing mechanism of the human brain. Recurrent neural networks (RNNs), a subtype of neural networks, introduce memory into the network by creating directed interconnection cycles between neurons in order to tackle tasks with temporal extent. Reservoir computing (RC) [1–3] was proposed as a methodology to ease the training of these recurrent networks, which is typically rather challenging. But while software-only reservoirs have successfully been applied to solve complex tasks such as speech recognition and time series prediction, they are still just another algorithmic way of digital information processing and are inevitably still bound to the same scaling limitations of CMOS technology (digital computers). To overcome this hurdle, researchers have figured out a number of ways to implement reservoirs in physical systems to form truly analog information processing and computing platforms [4–8].

Photonics has proven itself as a leading platform for analog information processing as it comes with tremendous bandwidth, has practical advantages such as immunity to electromagnetic interference, and can be integrated with microelectronics technology. Photonic reservoir computing is one such analog computing approach. It combines a nonlinear dynamic network of photonic nodes with a machine-learning-based readout and has been shown to reach state-of-the-art performance on a number of specific tasks such as speech recognition, time series forecasting, Boolean logic operation, etc. [9–12].

The integrated photonic reservoir computing platform of this work is an instance of this approach using nanophotonic technology on the silicon photonics platform. Apart from the obvious scaling and mass-production advantages, using integrated nanophotonics allows us to compute using both phase and amplitude, which has been shown to dramatically improve the performance of the reservoir. The platform is versatile and over time, a large collection of primitive components such as optical splitters, filters as well as detectors and high-speed modulators have been made available to implement various signal processing and conditioning functions. Many non-native devices can also typically be added via hybrid or monolithic integration procedures such as bonding, and in lieu of that, external discrete components can be pipelined with the photonic chips using available input-output primitives (such as fiber-chip couplers). Furthermore, a selection of non-linearities at time scales ranging from ms to fs are available that could be leveraged for reservoir construction. Before continuing the discussion of photonic reservoir computing, we should mention that together with photonic reservoir computing, a number of analog integrated photonic computing platforms are have been studied such as in [13] and [14]. All these systems tend to be neuromorphic in nature, seeking to mimic the information processing semantics of the human brain as well as its power efficiency (the brain has about 10^{11} neurons and can

perform up to 10^{16} operations per second while consuming only $25\,\mathrm{W}$ of power, orders of magnitude lower than any other information processor [15]).

While prototypical photonic reservoirs are implemented by interconnecting non-linear optical nodes, it has been shown that it is possible to achieve state-of-the-art performance using a completely passive linear network. This network of components does not include amplification or nonlinear elements. The nonlinearity required for computation is introduced at the readout point, typically with a photodetector [10, 12]. This approach to reservoir design leads to conceptually simple designs with fewer parameters to tune and possible energy efficiency benefits as no external energy is required for the computation itself (other than the energy in the input signal which needs to overcome e.g. propagation losses). The passive photonic reservoirs in [10, 12] achieved state-of-the-art performance on Boolean logic operations, header recognition and speech recognition tasks.

However, while the majority of work on integrated photonic RC has been focused on solving specific (and usually benchmark) tasks, very little attention was paid to the energy requirements for the computation. An understanding of the energy efficiency of a chosen reservoir design will allow the engineer to appropriately trade off energy efficiency and performance when building a solution for a particular task. This is an especially serious challenge in passive integrated reservoirs where the losses increase as the signal traverses the reservoir. In this thesis, we first show how the choice of the strategy to get the input signal into the reservoir impacts the performance and energy efficiency of the reservoir. We then study the possibility of taking advantage of the additional propagation channels available in multimode photonic integrated circuits to design a reservoir with lower internal losses. Finally, we demonstrate how photonic reservoirs can be used for nonlinear compensation in optical communications links in place of, or together with Digital Signal Processors (DSPs) in a way that meets the power, bandwidth and energy requirements of the telecom sector.

The rest of this section is structured as follows: we first introduce the problems plaguing optical communications and how photonics is at the forefront of signal processing and computing innovations in optical communications systems. We then briefly introduce silicon photonics technology and why it is the platform of choice for implementing high-speed, energy-efficient signal processing functions. This is followed by a discussion of how machine learning has re-emerged as a versatile tool for general-purpose information processing and how coupled with photonics it yields a powerful neuromorphic information processing paradigm. Finally, we introduce the objectives of the thesis and the outline.

1.2 Optical Communications

Telecommunication, the conveyance of information from point to point, is integral to many aspects of our lives: social, security, finance, leisure, etc. Each evolutionary step on the journey to what is considered modern telecommunication systems has been crafted to increase the reach over which information can be relayed. Light-based techniques of communication have always been at the forefront of innovation surrounding telecommunication. The earliest forms of communication were all optical in nature, from the early smoke signals to beacon fires and semaphore lines. Today, fiber-based communications systems use light as the conduit to transmit information over 100s and 1000s of kilometers across cities, oceans and continents. Even at much smaller distances, such as in on-chip interconnects, there is an increase in demand for light-based technology solutions to move the enormous amounts of data at a lower power budget than the extant electronic alternatives.

Fiber optical communications systems can take advantage of the lightwave medium's huge bandwidths, low susceptibility to electromagnetic interference and energy efficiency benefits when compared to their electronics-based counterparts, to push the range of optical communications. Today commercial optical communication links routinely transmit 100s of gigabits of data per second, and a record of 1 petabit/s over $206\,\mathrm{km}$ fiber has been demonstrated in a laboratory setup [16]. The growth in technology surrounding optical telecommunication is due to the exponential adoption of the internet as the primary global information warehouse. The requirements of communication systems have therefore changed from transmitting simple voice and text data, to delivering large volumes of video, structured and unstructured data. And in the past decades, this has further been compounded by the similarly massive increase in the number of connected devices that need to access data as well as communicate with each other, ushering in the so-called age of "Internet of Things" (IoT). All these factors have aligned to give rise to "big data" that needs to both be transmitted and processed. For now, this trend is envisioned to continue uninhibitedly upward.

Optical communication research and industrial sectors have reacted by fervently innovating to keep up with this deluge of data. The outcome has been a reduction of fiber losses, invention of low-cost sources and amplifiers, and a slew of innovations to support the lightwave communication technology. At the same time, electronics-based implementations of communication systems, faced with fundamental scaling challenges originating from the limitations of the underlying DSP technology, have generated the need for techniques that seek to replace them with more optically rooted versions. The result is continued penetration of optics towards the end-user to provide the backbone to support the huge bandwidth appetites of the consumers.

From source to destination, optical signals need to go through various manipulations for routing, conditioning, alignment etc., during generation, transmission and reception. Except for a few specialized cases (such as linear dispersion compensation), the traditional approach to this has always been to convert the signal back to electrical domain, carry out the processing and then convert back, for transmission on the next link, or consume electrically. However, as we alluded to above, these electronic implementations are encumbered by a number of challenges that impede them from processing, in real-time, signals beyond certain transmission rates (for example electrical signals at $20\,\mathrm{Gb/s}$ are already challenging to characterize and reliably receive). Using photonic reservoir computing-based signal processors would allow the processing to occur all-optically and therefore eliminate the need for optical-electrical and electro-optical conversions that are the main source of bandwidth limitations and other inefficiencies. In this thesis, we will provide general design guidelines on how these photonic reservoirs can be applied in the most power efficient configurations.

One particular challenge in optical communication for which electronic implementations have failed to provide a complete solution is the area of non-linear compensation. The Kerr nonlinearity has very recently surfaced as a fundamental limit to the scaling up of optical communication fiber capacity. Nonlinearity in fibers limits the maximum allowable SNR in optical communication links, especially coherent ones. Nonlinearity mitigation is very much an open problem that the telecom industry has to address to avoid the looming capacity crunch. While DSPs have proven themselves indispensable at combating almost all other forms of distortions, they are not adept at reliably tackling this nonlinearity challenge. To date, a number of digital mitigation techniques have been proposed. Digital Backpropagation (DBP), typically implemented on DSP chips, combats both linear and nonlinear impairments by solving the inverse nonlinear Schrödinger equation (NLSE) to estimate the transmitted signal [17]. However, DBP is resource intensive and complex to implement for practical applications. Nonlinear Volterra series equalizers (V-NLEs) model fiber nonlinearities as a series of Volterra kernels which are then inverted to undo the distortions to the transmitted signal. V-NLEs are easier to implement and more computationally efficient than DBP, especially for intrachannel applications [18], but are still quite resource-demanding. A nonlinear compensator based on photonic reservoir computing is therefore well-poised to fill this gap in technology to provide a stand-alone solution or to work in concert with DSPs to yield a solution to this problem that is very relevant in the telecom industry.

1.3 Silicon Photonics

Silicon photonics technology is the prevalent platform for building photonic integrated circuits (PICs), especially those operating in the infrared region. It is based on silicon-on-insulator (SOI) technology. Silicon is transparent in the C and O bands, $1.55\,\mu\mathrm{m}$ and $1.3\,\mu\mathrm{m}$ respectively, that are the workhorse of optical communication networks. SOI PICs can be fabricated through various techniques, however the CMOS-compatible fabrication technology – the same technology that is used to make CMOS electronic chips, – stands ahead of the pack. It offers a cost-effective, industrially proven process for mass producing silicon photonics chips for various applications. Apart from the obvious "do not re-invent the wheel" advantages, it presents opportunities for co-integration of photonics and electronics on the same substrate which enables high-speed interfaces. PICs fabricated this way can take advantage of the enormous investment and advances in CMOS technology, and can significantly tighten the research - industry cycle.

Silicon photonics devices are at the core of all networks types: high-speed modulators, resonant filters, Si-Ge detectors, add-drop multiplexers, etc., form the backbone of modern optical communication systems. Moreover, light sources can be (in some cases easily) integrated with III-V on the same substrate (by hybrid integration techniques), to yield compact transmitter modules. The versatility of the silicon photonics platform makes it suitable for implementation of various optical signal communication functions. Additionally, silicon photonics chips come in compact footprints that make it relatively straightforward to package them in industry-grade form factors easing the path to adoption in networks as needed.

In summary, with our devices implemented in silicon photonics, we can take advantage of the existing technology that has been developed for current optical communication technology and are also in a position to quickly leverage any advancements at both device and fabrication levels.

1.4 Machine Learning

Machine learning encompasses a set of techniques to teach computer systems how to perform complex tasks on previously unseen data without explicitly programming them. Examples of tasks that are suitable for some form of machine learning are classification and regression. The collection of machine learning techniques is extensive and for every application the most appropriate technique has to be selected, depending on the application's specific demands.

Machine learning is at the forefront of a large number of breakthroughs in the past decade alone, seeking to make sense of big data and processing IoT streams. Recent advances in machine learning techniques have led to the invention of Deep Learning, a class of techniques that holds cutting edge performance on computer

vision tasks (such as image recognition) and related applications (using Convolutional Neural Networks), and Natural Language Processing tasks like translation as well as speech recognition [19–24]. These advances have spurred innovation in search, autonomous transportation, virtual assistants, and numerous other applications that are key to defining the current age of technological advancement.

With respect to the nonlinear compensation challenge in optical communications we introduced earlier, some popular machine learning-based techniques have been proposed using Artificial Neural Networks (ANNs) [25] and Support Vector Machines (SVMs) [26, 27]. [28] presents a summary of machine-learning techniques in optical communications and contributes approaches for nonlinearity distortion removal based on nonlinear state-space based Bayesian filtering and Gaussian Mixture Models (GMMs). A more recent study of a photonic machine learning implementation for signal recovery in optical communications can be found in [29].

The work in this thesis is modelled on reservoir computing, a form of Recurrent Neural Networks (RNNs), which in turn is a subtype of neural networks with memory, that is appropriate for processing sequential data such as in the case of speech recognition and written character recognition and even generation. We rely on well-known linear models to train the readout of the reservoir in order to extract the desired outcome from the complex state of the reservoir. A more detailed description of the machine learning techniques used in this work will be discussed in Chapter 2.

1.5 Optical Signal Processing

The various challenges in optical communication systems, mentioned above, have over the past decade sparked interest in techniques that do processing activities on the transmitted signals (both linear and nonlinear) optically, techniques collectively referred to as optical signal processing. The result is the adoption of optical filters, switches, multiplexers, etc., in optical communication networks. Optical signal processing is particularly attractive as it comes with a potential for the reduction of the energy per bit consumed during processing.

The failure of current DSP-based technologies, which we discussed earlier, to fully account for nonlinear distortions in fiber-optic links, yields sub-optimal detection (leading to more than allowable receiver errors), which has driven interest in all-optical approaches to this problem of nonlinearity distortions. Optical nonlinearity compensation has successfully been demonstrated using optical phase conjugation (OPC), where a single optoelectronic component is inserted in the center of the transmission line to invert the polarity of the dispersion and nonlinearity parameters from the first phase of the transmission [30], such that they are compensated for in the second part of the link. However, the placement of the

OPC node limits network design flexibility and puts restrictions on the expected dispersion and power evolution profiles. Phase-conjugated twin wave (PCTW) is a more recent technique that offers a low complexity (simpler DSP), and is an effective solution for optical nonlinearity mitigation [31]. It is however coupled to reduced spectral efficiency (SE) because of the extra overhead to transmit a phase-conjugated copy of the transmitted signal.

Because of the issues mentioned above, in this thesis we study photonic reservoirs as an alternative platform for nonlinearity mitigation. Various implementations of photonic reservoirs have proven that the technology is competitive for analog information processing [9, 12, 32] and integrated photonic reservoirs such as those in [10, 33, 34] can push processing speeds even higher for digital optical information processing, making them suitable for processing optical telecom data which is typically modulated at high speeds.

The integrated photonic reservoir computing nonlinear compensator we propose operates in the optical domain, meaning that it is very fast, compact and ostensibly energy efficient. Moreover it can easily be co-integrated with receiver electronics via monolithic photonic-electronic co-integration in, for example, a transceiver module. Equally important is that, as its operation is based on machine-learning techniques, it is adaptable to different link conditions and network types. Therefore, a single chip could be used in different links, only requiring a retraining of the weights required to generate the final result. Alternatively, the photonic reservoir computing nonlinear compensator could also be used to augment the processing capabilities of DSPs. The reservoir could start the distortion removal process on the raw optical signal and hand over a more tractable signal to the DSP.

Finally, the success of any optical signal processing system will depend on how it can meet, or even surpass, the cost, bandwidth and energy efficiency requirements that have been set (quite high) by the existing digital options. So, while possessing numerous benefits, the passive integrated photonic reservoirs that have been proposed before are plagued by the issue of loss accumulation of the signal as it propagates. This must be addressed if they are to be a viable contender for telecom applications. In this dissertation we propose a number of techniques to put integrated photonic reservoir computing signal processors on a path to meet these industrial demands if they are to be used to solve more complex tasks such as nonlinear compensation. This is a prime goal of this thesis.

1.6 Objectives

Broadly, this dissertation focuses on the construction of optical information processing kernels based on photonic reservoir computing to yield state-of-the-art performance on telecommunication-relevant tasks, while simultaneously maximally utilizing the power they consume.

First, we investigate the best way in which to present the input to the reservoir. We track the power requirements, computational capabilities and task performance of reservoirs subject to various input configurations. In so doing, we will propose reservoir input strategies that not only lead to compact designs but also require the least amount of input power for error-free performance.

Second, we study whether constructing the reservoir from multimode components rather than the typical single-mode components can yield benefits in terms of loss reduction. Particularly, we check if the extra modes can enable us to preserve the signal in the reservoir for longer times and whether this can correspond to improvements in required SNRs for error-free performance.

Then, we assess the design of a photonic reservoir computer nonlinear equalizer for fiber-optic telecommunication applications. We seek to take advantage of the versatility of photonic reservoirs to generate a design that can compensate distortion in various link conditions and for various reaches and transmitter/receiver combinations. We demonstrate that the photonic reservoir computing nonlinear equalizer can either be a competitive alternative for power-hungry DSPs, or could provide physical layer pre-processing (first stage compensation) for DSPs, which translates into a reduction in their complexity and an increase in their power efficiency.

Finally, we present results from experiments on various silicon photonics chips encompassing a number of component and system-level designs geared at the next generation of even more energy-efficient integrated photonic reservoirs.

1.7 Thesis outline

The thesis is structured as follows: Chapter 2 introduces reservoir computing focusing on numerical modelling and the machine-learning techniques used for training reservoirs.

Chapter 3 presents the architectural search for the best input strategy for passive integrated reservoirs. The focus is on finding guidelines on how the input should be provided to the reservoir without trading off power efficiency, footprint or performance.

Thereafter, Chapter 4 focuses on the design and characterization of multimode reservoirs as a potential replacement for single-mode reservoirs.

Chapter 5 deals with the application of photonic reservoir computing to nonlinear compensation in optical communication networks. We consider nonlinear compensation in Intensity-Modulation Direct-Detection few $100\,\mathrm{km}$ metro links, as well high-speed short-reach applications.

Then in Chapter 6, we present designs and results for various chips designed to characterize components for the next-generation reservoir computing devices, as well reservoirs with on-chip all-optical integrated readout.

Finally, Chapter 7 discusses the main results and outlines future perspectives.

1.8 Publications

1.8.1 Publications in international journals

1. **A. Katumba**, Matthias Freiberger, F. Laporte, A. Lugnan, S. Sackesyn, C. Ma, Joni Dambre, P. Bienstman, *Neuromorphic computing based on silicon photonics and reservoir computing*, IEEE Journal on Selected Topics in Quantum Electronics (invited), 24(6), p.8300310 (10 pages) (2018)

2. **A. Katumba**, Jelle Heyvaert, B. Schneider, S. Uvin, Joni Dambre , P. Bienstman, *Low-Loss Photonic Reservoir Computing with Multimode Photonic Integrated Circuits*, Scientific Reports, 8(1), p.2653 (2018) .

3. **A. Katumba**, Matthias Freiberger, Joni Dambre, P. Bienstman, *A Multiple-Input Strategy to Efficient Integrated Photonic Reservoir Computing*, Cognitive Computation, 9(3), p.307-314 (2017).

4. **A. Katumba**, X. Yin, J.Dambre, P. Bienstman, *A Neuromorphic Silicon Photonics Nonlinear Equalizer for Optical Communication with Intensity Modulation and Direct-Detection*, submitted to the Journal of Lightwave Technolgy, IEEE, August 2018 (under review).

5. F. Laporte, **A. Katumba**, Joni Dambre, P. Bienstman, *Numerical demonstration of neuromorphic computing with photonic crystal cavities*, Optics Express, 26(7), p.7955-7964 (2018).

6. M. Freiberger, **A. Katumba**, P. Bienstman and J. Dambre, *Training Passive Photonic Reservoirs With Integrated Optical Readout*, in IEEE Transactions on Neural Networks and Learning Systems (2018)

7. K. Van Gasse, L. Bogaert, L. Breyne, J. Van Kerrebrouck, S. Dhoore, C. Op de Beeck, **A. Katumba**, C.Y. Wu, H. Li, J. Verbist, A. Rahim, A. Abbasi, B. Moeneclaey, Z. Wang, H. Chen, J. Van Campenhout, X. Yin, B. Kuyken, G. Morthier, J. Bauwelinck, G. Torfs, G. Roelkens, *Analog radio-over-fiber transceivers based on III-V-on-silicon photonics*, IEEE Photonics Technology Letters (invited), (2018).

8. S. Radosavljevic, N. Teigell Beneitez, **A. Katumba**, M. Muneeb, M. Vanslembrouck, B. Kuyken, G. Roelkens, *A mid-infrared Vernier racetrack resonator tunable filter implemented on a Germanium on SOI waveguide platform*, Optical Materials Express (invited), 8(4), p.824-835 (2018)

9. Florian Denis-le Coarer, Marc Sciamanna, **A. Katumba**, Matthias Freiberger, Joni Dambre, P. Bienstman, Damien Rontani, *All-optical reservoir computing on a photonic chip using silicon-based ring resonators*, IEEE Journal on Selected Topics in Quantum Electronics, 24(6), (2018)

10. Y. Wu , Q. Huang, S. Keyvaninia, **A. Katumba**, J. Zhang, W. Xie, G. Morthier, J. He, G. Roelkens, *All-optical wavelength conversion of 40 Gb/s NRZ data in a heterogeneously integrated III-V-on-silicon SOA*, Optics Express, 24(18), p.20318 (2016) .

11. A. Abbasi, C. Spatharakis, G. Kanakis, N. Sequeira Andre, H. Louchet, **A. Katumba**, J. Verbist, H. Avramopoulos, P. Bienstman, X. Yin, J. Bauwelinck, G. Roelkens, G. Morthier, *High Speed Direct Modulation of a Heterogeneously Integrated InP/SOI DFB Laser*, Journal of Lightwave Technolgy (invited), 34(8), p.1683-1687 (2016).

1.8.2 Publications in international conferences

1. K. Harkhoe, G. Verschaffelt, **A. Katumba**, P. Bienstman, G. Van der Sande, *Delay Based Reservoir Computing on Integrated InP Photonic Chip*, accepted for publication in the proceedings for Cognitive Computing: Merging Concepts with Hardware, Hannover, Germany, December 2018.

2. K. Harkhoe, G. Verschaffelt, **A. Katumba**, P. Bienstman, G. Van der Sande, *Integrated Delay Based Photonic Reservoir Computing*, Proceedings of the Annual Symposium of the IEEE Photonics Society Benelux Chapter, Brussels, Belgium, November 2018.

3. **A. Katumba**, Xin Yin, Joni Dambre, P. Bienstman, *Silicon Photonics Neuromorphic Computing and its Application to Telecommunications*, (invited), European Conference on Optical Communication (ECOC), Italy, 2018

4. A. Rahim, A. Abbasi, M. Shahin, Nuno Sequeira Andre Andre Richter, Joris Van Kerrebrouck, K. Van Gasse, **A. Katumba**, Bart Moeneclaey, Xin Yin, G. Morthier, R. Baets, G. Roelkens, *50 Gb/s DMT and 120 Mb/s LTE signal transmission over 5 km of optical fiber using a silicon photonics transceiver*, Advanced Photonics Congress, Switzerland, (2018)

5. Florian Denis-le Coarer, Damien Rontani, **A. Katumba**, Matthias Freiberger, Joni Dambre, P. Bienstman, Marc Sciamanna, *Toward neuro-inspired computing using a small network of micro-ring resonators on an integrated photonic chip*, SPIE PHOTONICS EUROPE, 10689, France, p.1068908-9 (2018)

6. S. Radosavljevic, N. Teigell Beneitez, **A. Katumba**, M. Muneeb, M. Vanslembrouck, B. Kuyken, G. Roelkens, *A Widely Tunable Vernier Filter on a Ge-on-SOI Platform for Sensing Applications*, Conference on Lasers and Electro-Optics (CLEO), United States, p.STh3I.7 (2018)

7. P. Bienstman, Joni Dambre, **A. Katumba**, Matthias Freiberger, F. Laporte, A. Lugnan, *Photonic reservoir computing: a brain-inspired approach for information processing*, The Optical Fiber Communication Conference (OFC) (invited), United States, p.paper M4F.4 (3 pages) (2018)

8. P. Bienstman, Joni Dambre, **A. Katumba**, Matthias Freiberger, F. Laporte, A. Lugnan, *Silicon photonics for neuromorphic information processing* , SPIE Photonics West (invited), DL 10551, United States, (2018).

9. Damien Rontani, Florian Denis-Le Coarer, **A. Katumba**, Matthias Freiberger, Joni Dambre, P. Bienstman, Marc Sciamanna, *Reservoir Computing with Nonlinear Micro-Resonators on a Silicon Photonics Chip*, International Symposium on Nonlinear Theory and Its Applications, Mexico, (2017) .

10. Matthias Freiberger, **A. Katumba**, Joni Dambre, P. Bienstman, *On-chip passive photonic reservoir computing with integrated optical readout*, ICRC2017, the IEEE International Conference on Rebooting Computing 2017, United States, p. 255 - 258 (2017)

11. D. Rontani, F. Denis, **A. Katumba**, M. Freiberger, J. Dambre, P. Bienstman, M. Sciamanna, *Photonic reservoir computing using a small network of micro-ring resonators*, Workshop on Dynamical Systems and Brain-inspired Information Processing, Germany, (2017)

12. **A. Katumba**, Matthias Freiberger, Jelle Heyvaert, Joni Dambre, P. Bienstman, *Building Scalable Integrated Silicon Photonics Reservoirs for Signal Processing*, Frontiers in Optics, United States, (2017)

13. **A. Katumba**, F. Laporte, A. Lugnan, J. Dambre, P. Bienstman, *Integrated-photonics implementation of reservoir computing neural networks*, Machine learning workshop at ECOC (invited), Sweden, (2017)

14. F. Denis, D. Rontani, **A. Katumba**, M. Freiberger, J. Dambre, P. Bienstman, M. Sciamanna, *Reservoir computing on an active silicon photonics chip using nonlinear microring resonators*, XXXVII Dynamic Days Europe, Hungary, p.187 (2017)

15. **A. Katumba**, Matthias Freiberger, Joni Dambre, P. Bienstman, *Passive integrated photonics reservoir computing with on-chip readout for telecommunication applications*, Dynamical Systems and Brain Inspired Computing 2017 (invited), Belgium, (2017) .

16. F. Denis, D. Rontani, **A. Katumba**, M. Freiberger, J. Dambre, P. Bienstman, M. Sciamanna, *Nonlinear microring resonators on silicon photonic chip for brain inspired computing*, Workshop on Dynamical Systems and Brain-inspired Information Processing, Belgium, (2017).

17. A. Rahim, A. Abbasi, Nuno Sequeira Andre, **A. Katumba**, Hadrien Louchet, K. Van Gasse, R. Baets, G. Morthier, G. Roelkens, *69 Gb/s DMT Transmission Using Directly Modulated Heterogeneously Integrated InP-on-Si DFB* , Optical Fiber Communication Conference and Exhibition (OFC) , United States, p.paper Th1B5 (3 pages) (2017)

18. **A. Katumba**, B. Schneider, J. Dambre, P. Bienstman, *An integrated photonics reservoir computing approach to signal equalization for telecommunications*, International Conference on Transparent Optical Networks (ICTON) 2016, Italy, p.Tu.D5.1 (2016)

19. A. Abbasi, C. Spatharakis, G. Kanakis, N. S. Andre, H. Louchet, **A. Katumba**, J. Verbist, X. Yin, J. Bauwelinck, H. Avramopoulos, G. Roelkens, G. Morthier, *PAM-4 and Duobinary Direct Modulation of a Hybrid InP/SOI DFB Laser for 40 Gb/s Transmission over 2 km Single Mode Fiber*, Optical Fiber Communication Conference (OFC) 2016, United States, p.M2C.6 (2016) .

20. **A. Katumba**, P. Bienstman, J. Dambre, *Photonic reservoir computing for telecom applications*, Dynamical systems and brain-inspired information processing, France, (2015) .

21. **A. Katumba**, P. Bienstman, J. Dambre, *Photonic reservoir computing approaches to nanoscale computation*, International Conference on Nanoscale Computing and Communication (NANOCOM, United States, (2015)

1.8.3 Patents

1. Peter Bienstman, **Andrew Katumba**, and Jelle Heyvaert, Joni Dambre, Matthias Freiberger *Training of photonic reservoir computing systems*, Patent Application No. PCT/EP2018/063857, Filed 26/05/2018

2. Peter Bienstman, **Andrew Katumba**, and Jelle Heyvaert, *Multimode reservoir*, Patent Application No. PCT/EP2018/063856, Filed 26/05/2018

1.8.4 Book Chapters

1. **A. Katumba**, M. Freiberger, F. Laporte, A. Lugnan, S. Sackesyn, C. Ma, J. Dambre, P. Bienstman, *Integrated on-chip reservoirs*, (invited) accepted

for publication in Photonic Reservoir Computing: Optical Recurrent Neural Networks, De Gruyter, (to be published July 2019).

References

[1] Wolfgang Maass, T Natschläger, and Henry Markram. *Real-time computing without stable states: A new framework for neural computation based on perturbations*. Neural computation, 2560:2531–2560, 2002.

[2] Herbert Jaeger and Harald Haas. *Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless communication.* Science (New York, N.Y.), 304:78–80, 2004.

[3] D. Verstraeten, B. Schrauwen, M. D'Haene, and D. Stroobandt. *An experimental unification of reservoir computing methods*. Neural Networks, 20(3):391–403, 4 2007.

[4] Helmut Hauser, Auke Ijspeert, Rudolf Füchslin, Rolf Pfeifer, and Wolfgang Maass. *Towards a theoretical foundation for morphological computation with compliant bodies*. Biological Cybernetics, 105(2011):355–370, 12 2011.

[5] Henry O Sillin, Renato Aguilera, Hsien-Hang Shieh, Audrius V Avizienis, Masakazu Aono, Adam Z Stieg, and James K Gimzewski. *A theoretical and experimental study of neuromorphic atomic switch networks for reservoir computing.* Nanotechnology, 24:384004, 2013.

[6] Manjari S. Kulkarni and Christof Teuscher. *Memristor-based reservoir computing*. In Proceedings of the 2012 IEEE/ACM International Symposium on Nanoscale Architectures - NANOARCH '12, pages 226–232, New York, New York, USA, 2012. ACM Press.

[7] K Vandoorne. *Photonic reservoir computing with a network of coupled semiconductor optical amplifiers*. PhD thesis, 2011.

[8] Y Paquot, F Duport, A Smerieri, J Dambre, B Schrauwen, M Haelterman, and S Massar. *Optoelectronic Reservoir Computing*. Scientific Reports, 2:287, 2 2012.

[9] L Appeltant, M C Soriano, G Van der Sande, J Danckaert, S Massar, J Dambre, B Schrauwen, C R Mirasso, and I Fischer. *Information processing using a single dynamical node as complex system.* Nature communications, 2:468, 9 2011.

[10] Kristof Vandoorne, Pauline Mechet, Thomas Van Vaerenbergh, Martin Fiers, Geert Morthier, David Verstraeten, Benjamin Schrauwen, Joni Dambre, and Peter Bienstman. *Experimental demonstration of reservoir computing on a silicon photonics chip.* Nature communications, 5:3541, 1 2014.

[11] L. Larger, M. C. Soriano, D. Brunner, L. Appeltant, J. M. Gutierrez, L. Pesquera, C. R. Mirasso, and I. Fischer. *Photonic information processing beyond Turing: an optoelectronic implementation of reservoir computing*. Optics Express, 20(3):3241, 1 2012.

[12] Quentin Vinckier, Franois Duport, Anteo Smerieri, Kristof Vandoorne, Peter Bienstman, Marc Haelterman, and Serge Massar. *High-performance photonic reservoir computer based on a coherently driven passive cavity*. Optica, 2(5):438–446, 2015.

[13] Alexander N. Tait, Mitchell A. Nahmias, Bhavin J. Shastri, and Paul R. Prucnal. *Broadcast and weight: An integrated network for scalable photonic spike processing*. Journal of Lightwave Technology, 32(21):3427–3439, 2014.

[14] A. Hurtado, K. Schires, I. D. Henning, and M. J. Adams. *Investigation of vertical cavity surface emitting laser dynamics for neuromorphic photonic systems*. Applied Physics Letters, 2012.

[15] Ralph C. Merkle. *Energy Limits to the Computational Power of the Human Brain*. Foresigh Update, 6, 1989.

[16] Nippon Telegraph and Telephone Corporation. *One Petabit per Second Fiber Transmission over a Record Distance of 200 km*, 2017.

[17] Danish Rafique, Jian Zhao, and Andrew D. Ellis. *Digital back-propagation for spectrally efficient WDM 112 Gbit/s PM m-ary QAM transmission*. Optics Express, 19(6):5219, 2011.

[18] Ling Liu, Liangchuan Li, Yuanda Huang, Kai Cui, Qianjin Xiong, Fabian N. Hauske, Changsong Xie, and Yi Cai. *Intrachannel nonlinearity compensation by inverse Volterra series transfer function*. Journal of Lightwave Technology, 30(3):310–316, 2012.

[19] T U Munich. *Framewise Phoneme Classification with Bidirectional LSTM and Other Neural Network Architectures*. Neural Networks, 18(5):602–610, 2004.

[20] Sepp Hochreiter and Jurgen Jürgen Schmidhuber. *Long short-term memory*. Neural Computation, 9(8):1–32, 1997.

[21] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. *Gradient-based learning applied to document recognition*. Proceedings of the IEEE, 1998.

[22] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. *ImageNet Classification with Deep Convolutional Neural Networks*. Advances In Neural Information Processing Systems, pages 1–9, 2012.

[23] Alex Graves. *Generating Sequences With Recurrent Neural Networks*. pages 1–43, 2013.

[24] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. *Domain Adaptation for Large-Scale Sentiment Classification: A Deep Learning Approach*. Proceedings of the 28th International Conference on Machine Learning, (1):513–520, 2011.

[25] Mutsam Jarajreh, Elias Giacoumidis, Ivan Aldaya, Son Le, Athanasios Tsokanos, Zabih Ghassemlooy, and Nick Doran. *Artificial Neural Network Nonlinear Equalizer for Coherent Optical OFDM*. IEEE Photonics Technology Letters, 27(4):1–1, 2015.

[26] Yi Han, Song Yu, Minliang Li, Jie Yang, and Wanyi Gu. *An SVM-based detection for coherent optical APSK systems with nonlinear phase noise*. IEEE Photonics Journal, 6(5):1–10, oct 2014.

[27] Elias Giacoumidis, Sofien Mhatli, Tu Nguyen, Son T. Le, Ivan Aldaya, Mary E. McCarthy, Andrew D. Ellis, and Benjamin J. Eggleton. *Comparison of DSP-based nonlinear equalizers for intra-channel nonlinearity compensation in coherent optical OFDM*. Optics Letters, 41(11):2509, 2016.

[28] Darko Zibar, Molly Piels, Rasmus Jones, and Christian G. Schaeffer. *Machine Learning Techniques in Optical Communication*. Journal of Lightwave Technology, 34(6):1442–1452, mar 2016.

[29] Apostolos Argyris, Julián Bueno, and Ingo Fischer. *Photonic machine learning implementation for signal recovery in optical communications*. Scientific reports, 8(1):8487, 2018.

[30] Danish Rafique. *Fiber Nonlinearity Compensation: Commercial Applications and Complexity Analysis*. Journal of Lightwave Technology, 34(2):544–553, 2016.

[31] Xiang Liu, A. R. Chraplyvy, P. J. Winzer, R. W. Tkach, and S. Chandrasekhar. *Phase-conjugated twin waves for communication beyond the Kerr nonlinearity limit*. Nature Photonics, 7(7):560–568, 2013.

[32] Daniel Brunner, Miguel C. Soriano, Claudio R. Mirasso, and Ingo Fischer. *Parallel photonic information processing at gigabyte per second data rates using transient states.* Nature communications, 4:1364, 1 2013.

[33] Hong Zhang, Xue Feng, Boxun Li, Yu Wang, Kaiyu Cui, Fang Liu, and Weibei Dou. *Integrated photonic reservoir computing based on hierarchical time-multiplexing structure*. Opt. Express, 22(25):31356–31370, 12 2014.

[34] Charis Mesaritakis, Alexandros Kapsalis, and Dimitris Syvridis. *All-optical reservoir computing system based on InGaAsP ring resonators for high-speed identification and optical routing in optical networks*. volume 9370, page 937033, 2 2015.

2 Reservoir Computing

2.1 Introduction

In the introduction chapter, we pointed out that a number of modern information processing approaches are analog in nature, typically employing the dynamics of a physical system and a readout system that extracts a meaningful signal from the complex representation created by the dynamical system. Photonic reservoir computing takes advantage of the photonics platform for the dynamics and also heavily relies on machine-learning techniques. Photonics and machine learning are very broad subjects, whose depth cannot be exhausted in a single chapter or even a single book. Therefore, in this section we will only briefly introduce the key theory and techniques from both areas required to develop integrated photonic reservoirs and use them to solve tasks.

This chapter is structured as follows: in section 2.2 we first give a general introduction the field of machine learning including a discussion on Artificial Neural Networks (ANNs), Feedforward Neural Networks (FFNNs) and Recurrent Neural Networks (RNNs). Section 2.3 focuses on classical software-based reservoir computing, first from a historical perspective motivating the purpose and path to their invention, and then discussing the nuances of reservoir computing system implementations. Section 2.4 then discusses various approaches by which photonics has been combined with reservoir computing with the aim of processing optical signals. Section 2.4.1 discusses how the processing can be achieved with a passive photonic network coupled to a nonlinear layer resulting in simpler designs for

photonic reservoirs. Section 2.5 presents a mathematical model for the reservoir, together with a discussion of the procedure for training it.

2.2 Machine Learning

Machine learning embodies the collection of techniques to teach computers to solve tasks without being explicitly programmed for this purpose [1]. If for example you would like to use a computer to perform a cancer diagnosis, you would collect the historical data available on past cancer diagnoses and use this to train a machine learning model based on some suitably chosen algorithm. If the model has successfully learned from the data, it should do well at predicting the correctness of cancer diagnoses for new cases that were not part of the data that was used in the training phase.

Based on the requirements of the task to be solved, machine learning techniques can broadly be categorized into 3 different classes:

Supervised Learning

The task here is to learn the relationship between the data ('observations') and a target variable ('labels'). The goal is to use certain attributes of the input data ('feature vectors') to influence the model to yield good predictions on the labels. Learning is considered successful (generalization) if the model can then make accurate predictions when fed input data that was unseen during the training phase. Supervised learning usually involves either classification, where the labels are discrete, or regression, where the label to be predicted is continuous in nature. Supervised learning is by far the most commonly applied form of machine learning. Common algorithms for supervised learning are Nearest Neighbor [2], Naive Bayes [3], Decision Trees [4], Support Vector Machines (SVMs) [5] and Neural Networks [6, 7].

Unsupervised learning

In some cases, we wish to discover some latent (hidden) structure within a dataset. In this case, we are solving an unsupervised learning problem since the data points are not explicitly labelled. Two of the most popular algorithms for unsupervised learning are used for clustering and dimensionality reduction. A canonical example of a clustering application is an online retailer who is interested in finding out the different market segments to which their clients may belong, from historical sales data, so that they can better target their advertisement campaigns. On the other hand, dimensionality reduction algorithms are popularly used as preprocessors for supervised methods for feature selection and compression of input data to speed up model training or to provide a means to visualize higher dimensional data.

Some of the most popular algorithms for unsupervised learning are k-means [8] and Gaussian mixture models (GMMs) [9] for clustering, and Pricipal Component Analysis (PCA) [10], t-SNE [11] and Auto Encoders [12] for dimensionality reduction.

Reinforcement Learning

In reinforcement learning, an 'agent' has to figure out what the best action to perform is in a certain environment given its current configuration (state). The agent learns what to do in each situation through a trial-and-error procedure driven by the goal of reward maximization. Reinforcement learning differs from other (previously mentioned) forms of machine learning as it reaches its final goal through this search mechanism. Its decisions are driven by a delayed reward mechanism: the agent takes the best action in its current state to maximize its future reward [13]. Similar to the way children and animals learn, agents learn from examples when they get "rewarded" for taking good decision and getting "punished" when they make the wrong choices. These kinds of algorithms excel in gaming environments [14] and control engineering such as in robotics [15]. In summary, reinforcement learning agents build their own understanding of their environment and figure out a way achieve their goal without any historical data or explicit instruction. Some common algorithms in reinforcement learning are Q-Learning [16], Temporal Difference (TD) learning [17], Deep Adversarial Networks [18, 19] etc.

2.2.1 ANNs

An Artificial Neural Network is a simplified model of the brain that is capable of some form of computation without being explicitly programmed. It learns the structure of the problem from a collection of examples from which it derives the rule to project any unseen input to a solution. The neural network is composed of artificial neurons interconnected together (analogous to the neurons and synapses for example in the human brain). The neurons carry out processing while the synaptic connections allow information to flow between the different neurons.

A number of neuron types have been used in ANNs. There are the threshold type neurons also known as perceptrons that produce a digital output and are computationally universal [20]. Another set of neurons that are more extensively used are the analog neurons, also collectively termed *activation functions*. Activation functions provide the nonlinear transformation between inputs and outputs at the nodes and enable ANNs to model complex relationships between inputs and outputs. Activation functions are usually differentiable in nature for compatibility with typically used optimization techniques such as backpropagation. Popularly used activation functions are the *sigmoid* neuron (Figure 2.1 (b)) and the *tanh* neuron (Figure 2.1 (b)). Another type of neuron that has gained significant popularity

in the newer deep neural networks is the Rectified Linear Unit (ReLU) (Figure 2.1 (b)) neuron which thresholds the input at 0. Despite being non-differentiable at zero, the ReLU has been show to be more biologically plausible and to have better or equal performance than the other two analog neuron types on various tasks (its derivatives have particularly proven reliable at tackling the issue of vanishing gradients that plagues other activation functions). In practice *sigmoid* and *tanh* activation functions are used in all layers of the network (input, output and hidden) while *ReLU* neurons are usually only found in the hidden layers.

These 3 activation functions can be represented by the following equations:

$$sigmoid(x) = \frac{1}{1 + e^{-x}} \tag{2.1}$$

$$tanh(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \tag{2.2}$$

$$relu(x) = max(0, x) \tag{2.3}$$

The final type of neuron encountered is the spiking neuron. Spiking neurons are even more biological plausible than all their other counterparts and are the basis for a class of ANNs known as the Spiking Neural Networks (SNNs). In SNNs, communication between the neurons happens via discrete spikes (instead of passing of continuous values as is the case in analog neuron-based ANNs) and information is encoded in the precise timing and the neuron firing rate. The biggest challenge of SNNs is the difficulty of representing the real world analog nature of signals in the discrete form required by the network.

This work focuses on ANNs based on analog neurons. As already mentioned, in an ANN, each neuron receives input from the neighbouring neurons or from the input layer, and generates as output the weighted sum of these inputs transformed through an activation function . There are two main types of ANN, Feed Forward Neural Networks (FFNNs) and Recurrent Neural Networks (RNNs) which we will discuss next.

2.2.2 Feed Forward Neural Networks

FFNNs are by far the most well known and studied type of neural network. They consist of an input layer, an output layer and one or more intermediate layers (hidden layers). Their name comes from the fact that information flows strictly from the input layer towards the output layer in contrast to the recurrent neural networks of the next section that allow for both forward and backward propagation of information. An example of a simple single hidden layer feed forward neural network is illustrated in Figure 2.2.

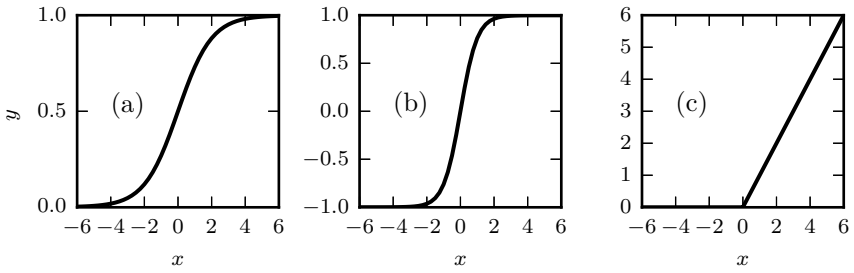


Figure 2.1: Activation functions commonly used in neural networks: (a) sigmoid*, (b)* tanh*, and (c)* ReLU.

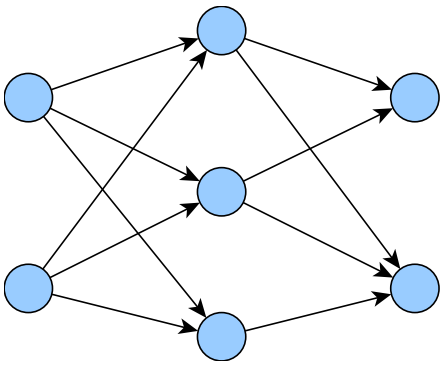


Figure 2.2: A feed forward neural Network

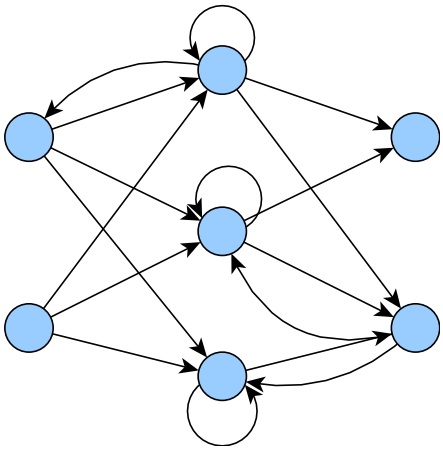


Figure 2.3: A recurrent neural network.

These networks can model highly nonlinear and complex relations and generalize well. The most common approach to training FFNNs is by error backpropagation where the final error between the expected and predicted outputs is used to make changes to the weights of the network to do better on the training set on a per epoch basis. FFNNs have successfully been applied extensively in areas such as image processing, character recognition as well as forecasting in finance, medicine, security, agriculture etc.

2.2.3 Recurrent Neural Networks

Without cyclic connections, the the feed forward neural nets of the previous section cannot handle tasks involving sequential, temporal data in a natural way. To tackle this, RNNs introduce loops within the network allowing them to process signals that are associated with a time sequence. Figure 2.3 is an illustration of a RNN in which the feedback loops are visible. Having these extra connections that are not available in their FFNNs counterparts, makes RNNs more suitable for tasks such as speech recognition, time series prediction etc. which require some form of memory. With memory, RNNs can model dependencies between inputs and/or between inputs and outputs. For example, when predicting what character comes next in a written text you require information about the current character and whichever characters came before (usually a finite window of characters is used in each prediction or classification step). So, rather than the nonlinear function approximators of FFNNs, RNNs form a nonlinear dynamical system capable of memorizing and processing past sequential inputs [21]. Recently RNNs (and particularly their variants such as Long Short Term Memory (LSTMs) networks) have led to significant advances in the field of Natural Language Processing (NLP). RNNs have shown state-of-the-art performance for language modelling and text generation [21–23].

While RNNs have multiple applications and address the challenges of FFNNs in relation to sequential data processing, they come with their own challenges. RNNs are usually trained using techniques based on gradient updates (mainly backpropagation), but since the process is repeated over and over, this leads to a situation where the changes in network parameters either become too small or too large over a number of timesteps. These problems are known respectively as the vanishing and exploding gradients problem in artificial learning parlance and this makes it difficult to learn long-term dependencies. A number of techniques have been proposed to address this issue, such as techniques based on higher-order gradients (see for example [24]), but these tend to be quite complicated to setup and tune. As a result, in the past few decades there has been interest in techniques for handling the challenge of reliably training RNNs within treasonable time and computational constraints. One approach is using novel architectures that extend RNNs by keeping extra state (typically called cell state) and making use of gate

units in the network that modulate how this new state is updated, to what degree new inputs are taken into account and how much the internal state of the RNNs is converted into the final outputs of the RNNs. These approaches are used in LSTM and Gated Recurrent Unit (GRU) RNNs. These types of networks are very interesting for various software-based implementations of RNNs but do not easily translate to hardware implementations. In this work, we focus on the other approach which involves not training the internal RNN parameters but only setting them up to be in the correct dynamical regime and only training the readout. This approach, known as reservoir computing, is conceptually easier to implement and is amenable to hardware implementation. In the next section we will provide details on the origin, theory and practice of reservoir computing.

2.3 Reservoir Computing

Reservoir Computing (RC) [25–28] initially emerged as a way around the intricacies associated with correctly training recurrent neural networks for supervised learning. Particularly, the training of RNNs required learning all network weights for input, internal connectivity and output which was challenging for the algorithms of the time (1990s). As indicated above, most of the algorithms of the time used some kind of gradient descent that was inherently lengthy and could not provide convergence guarantees. However, as of today, advances in machine learning have mostly solved the challenges of gradient descent training of RNNs. Today's RNNs and their variants (notably LSTMs that make it possible to learn long term dependencies) can reliably be trained to solve very complex sequence tasks such as in natural language processing.

While RNNs trained with gradient descent have become the go-to technology toolkit for handling many of the sequential processing tasks today, reservoir computing still remains a competitive paradigm in two specific cases. First, when solving tasks with low levels of complexity and when quick convergence is required [29]. However, the most important second application is in the realm of analog computation and information processing hardware platforms to which photonic reservoir computing belongs.

Strictly speaking, reservoir computing was independently developed by Herbert Jaeger as Echo State Networks (ESNs) [26] in 2001 and by Wolfgang Maass as Liquid State Machines (LSMs) [25] in 2002 [30].

Jaeger's ESN is composed of a RNN initialized with a random topology and random weights and is engineered to asymptotically forget its initial state when driven by an external signal – the so-called echo state property which loosely translates into fading memory of previous inputs. A key design guideline for ESNs is that the dynamics of the network should allow for sufficiently rich mixing of inputs and the reservoir state (due to previous inputs), but not so rich as to hinder

the fading memory property. Jaeger called this operating the reservoir at the *edge of stability*. To extract the results of the computation from the complex reservoir state, the RNN is coupled with a simple linear readout to extract the desired result [26]. As the ESN conception was from an engineering perspective, it turns out to be the most practical choice and it is what is popularly used in many analog implementations of RC.

Taking a different approach from Jaeger's engineering practicality, Maass approached the problem from a robotics and neuroscience angle, driven by the biology plausibility of the resulting construct [26]. He defined the LSM as consisting of two parts: a high-dimensional map that projects the current and past inputs into a new state space and a linear readout that projects this state into the output. The map should be highly discriminative between similar inputs, and the readout should be powerful enough to approximate any function on a closed and bounded domain (universal approximation property). In practice, LSMs typically have a spiking RNN as the map and have been used in a number of practical applications such as in [31, 32].

A final approach to reservoir computing worth mentioning, an extension to the other two, is Backpropagation-Decorrelation (BPDC) introduced by Steil in 2004 [33]. BPDC implements an online learning rule that seeks to improve the performance of the linear readout layer by temporally decorrelating the network activations with respect to the inputs and the one-step backpropagated errors. Again here the RNN is fixed (to within a global scaling parameter) and only the weights of the readout are modified during training.

As a summary, software RC involves setting up a large randomly initialized nonlinear dynamical system (*the reservoir*) that is tuned to a specific dynamical regime to allow for the following three conditions: separability of the inputs, generation of similar outputs for similar inputs and some form of finite memory of the previous inputs. Under these circumstances, the states of the reservoir can be linearly combined, following task-imposed optimization criteria, to extract the desired outputs for the specified inputs.

Software reservoirs are typically simulated on general-purpose computers by setting up a model of the reservoir and executing it in a sequential manner to generate outputs from the inputs. While implementing reservoir computing this way is most straight-forward, it is by no means optimal and is subject to several limitations. The sequential nature of the execution leads to throttling of the reservoir speeds as it discards the inherent parallelism of the reservoir. As a consequence, RC research has strived to invent ways to enable computing with physical nonlinear dynamical systems that could take full advantage of the reservoir computing platform starting with parallel information processing. Examples of reservoir demonstrations in mechanical systems, memristive systems, atomic switch networks, boolean logic elements and photonic systems can be found in [34–38].

In all implementations, the RC system consists of three basic parts: an input layer which couples the input signal into a non-linear dynamical system, "the reservoir" (i.e. the recurrent neural network, which is kept untrained) and finally the output layer that typically linearly combines the states of the reservoir to provide the time-dependent output signal. An illustration of this reservoir computing architecture is given in Fig. 2.4 .

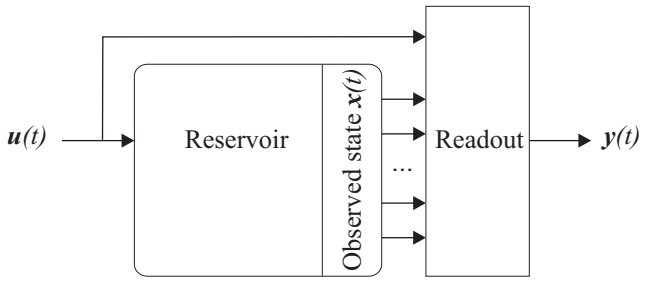


Figure 2.4: Schematic representation of a reservoir computing system. The input signal $u(t)$ is fed into the reservoir and the resulting reservoir states $x(t)$ are used to learn a linear readout that is then used to generate the output signal $y(t)$.

2.4 Photonic Reservoir Computing

Of the hardware adaptations of reservoir computing, photonic RC presents a number of benefits as it offers a large bandwidth and is inherently massively parallel. Photonic reservoirs can tackle various temporal tasks (often with state-of-the-art performance) such as speech recognition, nonlinear channel equalization, robot control, time series forecasting, handwriting recognition etc. In the previous section we introduced the concepts underpinning reservoir computing as a field. In the rest of the sections we will continue the discussion but focusing on photonic implementations.

Two main paradigms for designing photonic reservoir computing systems have emerged as the most promising, especially considering practical implementations.

The first set of approaches are based on a single nonlinear node with delayed feedback architecture, and can be found in [39–47]. These approaches typically make use of the internal dynamics of the nonlinear node to map the inputs to the higher dimensional space. They make use of *virtual nodes* defined within the delay line that are defined in a time multiplexing sense. Examples of nodes that have been studied include lasers and Mach-Zehnder modulators. These approaches will also typically require masking of the input (i.e. multiplication with a fixed, higher-speed periodic signal) to set up the virtual nodes.

The other class of techniques is implemented on an integrated photonics platform [37, 48–53]. These techniques typically take a selection of nonlinear photonic components and compose them into a network of nodes to create the reservoir. Some examples of nodes that have been studied are semiconductor optical amplifiers (SOAs) [48], photonic crystal cavities [50] and microring resonators [52]. To keep with the limitations of the platform, integrated reservoirs tend to be planar in nature and crossings have to be dealt with or avoided. Unlike the former approaches, these reservoirs typically have physical nodes corresponding to the points where the outputs of the nodes are read out. The work of this thesis belongs to this class of techniques. Integrated reservoirs tend to be more compact than their counterparts and are amenable to the scaling advantages of their platforms. This makes them a more attractive alternative in some fields, such as telecom, where mass production is required.

Generally, photonic RC implementations are advantageous when the signals to be processed are already in the optical domain. This is true in the case of tasks oriented towards fiber-optic based telecommunication systems, like bit-sequence processing tasks (such as logical temporal XOR, AND, OR), header recognition and equalization. The reservoir manipulates the light signals directly without the need for any extra electrical-optical and/or optical-electrical conversions. This leads to processing speed-ups and overall reduction in system complexity and energy requirements, which is of huge importance for telecommunication applications.

2.4.1 Passive Photonic Reservoir Computing

Most photonic reservoir computing systems were developed by broadly following the structure of their software counterparts. Reservoirs were built by interconnecting the required number of nodes (real or virtual), where in the case of integrated photonics, the node was a photonic component with functional behaviour mimicking in some way that of a nonlinear activation function. Passive photonic reservoir computing is a more recent invention that deviates from this approach. Here the input signal is propagated though a passive linear photonic network, i.e., one without amplification or nonlinear elements, and the required nonlinearity is introduced through the readout process [53]. Since the initial demonstration in integrated photonics, the concept has been adapted to the single node with delayed feedback architecture in the form of a coherently driven passive cavity [39].

In the case of integrated photonics, the reservoir consists of a linear network of passive photonic integrated circuit (PIC) components, specifically combiners/splitters and long waveguides (spirals). The required nonlinear transformation is provided by the readout system (an optical nonlinearity would also a viable alternative). In current passive photonic RC implementations, the photodetector which converts the complex-valued reservoir states to real-valued intensities,

suitably serves this purpose [53].

Apart from simplicity from a fabrication point-of-view, passive reservoir architectures tend to operate with reduced power requirements, as the computation itself does not require external energy. As an example, the SOA active reservoir of [37] turned out to be quite challenging in practice as it featured a number of amplifiers that needed to be biased and operated under strict conditions to avoid introducing instability. The other challenge here was that, with all the energy being injected into the system, temperatures could rise to very high values where thermal management becomes non-trivial.

With respect to interconnection topology, in [53], the reservoir nodes are laid out in a swirl topology as a way to minimize crossings that manifest as a source of signal cross-talk and extra losses, and to satisfy planarity constraints of the CMOS SOI platform while allowing for a reasonable mixing of the input signals. A 16-node photonic swirl reservoir is shown in Fig. 2.5. Unless stated otherwise, the work in this thesis is based on this architecture.

With regards to their applicability to problem solving, in general for tasks that are not strongly nonlinear, it is possible to achieve state-of-the-art performance using completely passive reservoirs. Nevertheless, even with the nonlinearity only occurring at readout time, with the richness originating from the mixing of the complex-valued electromagnetic field of the optical signals in the reservoir, it is still possible to solve a number of non-trivial tasks with significant industrial relevance.

In the next section we will provide a high-level model of the passive integrated photonics reservoir as was used in the numerical analysis of various reservoir architectures and tasks in this thesis. We will also discuss in more detail how the training and validation proceeds.

2.5 Reservoir model, training and validation

2.5.1 Model

In discretized time, the passive reservoir state update equation can be generalized as:

$$\vec{x}[k+1] = \boldsymbol{W}_{res}\vec{x}[k] + \vec{w}_{in}\vec{u}[k+1] + u_{bias} \tag{2.4}$$

where $\vec{u}$ is the input to the reservoir and u_{bias} is a fixed scalar bias applied to the inputs of the reservoir. For an N-node reservoir, $\boldsymbol{W_{res}}$ is an $N \times N$ matrix representing the interconnections between reservoir components taking into account splitting ratios and losses, with phases drawn from a random uniform distribution on $[-\pi, \pi]$, $U(-\pi, \pi)$. $\vec{w}_{in}$ is an N-dimensional column vector whose elements are nonzero for each active input node. These input weights are similarly chosen

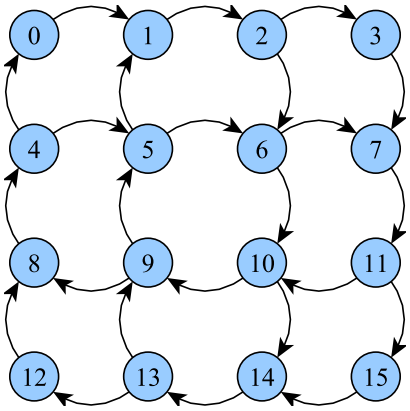


Figure 2.5: Signal flow in a 16-node swirl reservoir architecture. The time-dependent output at each numbered node is linearly combined to result in the answer of the computation. As for inputs, depending on the application the input can be inserted in one or more of the numbered nodes.

from $U(-\pi, \pi)$. Equation 2.4 is re-cast graphically to show the evolution of the reservoir states unrolled over a few timesteps in Figure 2.6.

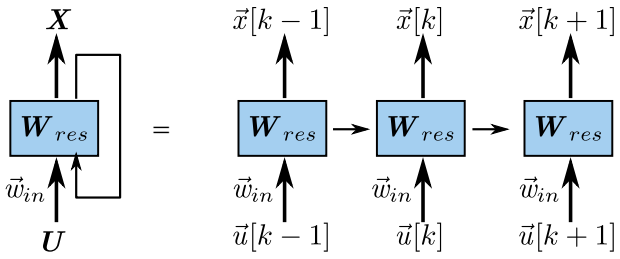


Figure 2.6: Unfolding the passive reservoir computing system for 3 time-steps. In the left figure, $\boldsymbol{U}$ is a concatenation of all the input vectors including the bias and $\boldsymbol{X}$ is the concatenation of the reservoir states collected at each timestep, these will then be used to train the readout of the reservoir.

2.5.2 Reservoir training and validation

The reservoir readout can be trained in either an offline or online manner. In this work, we mainly focus on offline training. However, we will briefly touch on the subject of online training in a later section, as it is pertinent to some of the newer reservoir designs currently under investigation, such as those with coherently combined states that we will introduce in a later chapter.

Training the reservoir readout proceeds in two main phases: a supervised train-

ing phase and a testing (validation) phase. In the training phase, a set of known inputs with known outputs are fed into the reservoir, the resulting states are collected, concatenated into a matrix $\boldsymbol{A}$ ($m \times N$) where N is the number of nodes readout from the reservoir and m is the number of examples in the training set. Similarly, the desired outputs are also concatenated into a matrix $\boldsymbol{B}$ ($m \times c$) where c is the length of each desired output element (for example in the case of binary classification $c = 1$ with $B_i = \{0, 1\}$). $\boldsymbol{A}$ and $\boldsymbol{B}$ are then used in the training phase to obtain a set of weights $\vec{w_{out}}$ (the model) such that

$$\boldsymbol{B} = \boldsymbol{A}\vec{w_{out}}. \tag{2.5}$$

In most practical cases, $m \gg N$ meaning that equation 2.5 is overdetermined so we proceed to solve the system in the least squares linear regression sense. We solve the quadratic program

$$\vec{w} = \arg\min_{\vec{w}} J(\vec{w}; \boldsymbol{A}, \boldsymbol{B}) = \arg\min_{\vec{w}} \frac{1}{m} \sum_{i=1}^{m} (\vec{w}^{\intercal} \vec{a_i} - \vec{b}_i)^2 \tag{2.6}$$

where $\vec{a_i}$ is the i-th example in $\boldsymbol{A}$ and $\vec{b}_i$ is the i-th element of $\boldsymbol{B}$. $J(\vec{w}; \boldsymbol{A}, \boldsymbol{B})$ is generally referred to as the objective function.

The solution to equation 2.6 could be obtained using the same gradient descent approach that is popular with ANN users, but here we take the vectorized version of the equation and use an approach based on the Moore-Penrose pseudoinverse [54] to obtain the weights in one go as follows

$$\vec{w_{out}} = \boldsymbol{A}^{\dagger} \boldsymbol{Y} = (\boldsymbol{A}^H \boldsymbol{A})^{-1} \boldsymbol{A}^H \boldsymbol{B}. \tag{2.7}$$

The hermitian conjugate of $\boldsymbol{A}^H$ of $\boldsymbol{A}$ to account for cases of training on complex values. In the real-valued case, $\boldsymbol{A}^H$ degenerates to $\boldsymbol{A}^{\intercal}$.

In the testing phase, the learned weights $\vec{w_{out}}$ are used together with the testing data to make predictions and evaluate the overall performance of the model on unseen data. We can generally claim that the model has learned how to solve the task encoded in the input-output pairs (learned to generalize) if we can get comparable performance on the test dataset. By feeding the test data into the reservoir we obtain states $\boldsymbol{A}'$ and compute the output of the model as

$$\boldsymbol{Y} = \boldsymbol{A}' \vec{w_{out}} \tag{2.8}$$

The empirical performance of the model is subsequently determined using some measure of error of the desired performance on the known test set $\boldsymbol{B}'$ together with the predicted (model output) output $\boldsymbol{Y}$. The test error is known as the generalization error and represents the expected performance of the model on data that was not in the training set. In this thesis, the majority of the tasks solved are encoded as classification problems and therefore the error measure used, for

the binary case, will take the form $e(\vec{w},(a,y)) = I(f(a\vec{w}) \neq y)$; where I is the indicator function that gives 0 when the class y does not correspond to the prediction of the trained model $\vec{w}$ from reservoir states a. Since we typically train on intensities rather than electric fields, $a\vec{w} \subset \mathbb{R}^N$ so it needs to go through extra processing by a function f to turn it into a integer class that is compatible with the error measure. This could for example be done by squashing it with the sigmoid or tanh activation function followed by thresholding (in some cases just thresholding suffices).

An issue that may arise with our current training procedure is that we could end up with a model that performs well on the training data but poorly on the test dataset, a situation referred to as overfitting. This implies that the model is too complex and fits to the noise in the data rather than discovering the underlying trends (distribution) in the data. To avoid this, we need to control the complexity of the model so that it generalizes well rather than memorize the training data. One approach is to add restrictions to the model weights in a way that favors simpler models, a process called regularization.

A second issue is how to best structure the model training procedure, particularly regarding the splits for training and testing data, and the selection of global model parameters. We shall adopt cross-validation techniques to this end.

These two issues will be described in detail in the next two sections.

2.5.3 Regularization

If during training the weights of the readout model are left unbounded, there is a risk that they can take on arbitrarily large values that lead to an overly complex model that struggles to yield good performance outside the training dataset. It is therefore common for an extra term to be added to the loss function to keep the weights within reasonable values.

In this thesis we make use the L^2-strategy which penalizes the norm of the model weights. L^2 is also commonly referred to in literature as *weight decay* to point to the fact that it forces the values of the weights closer to the origin. It is also known as *ridge regression* or *Tikhonov Regularization*. The regularization is achieved by appending the term $\frac{\lambda}{2}\|\vec{w}\|_2^2$ to the objective function $J(\vec{w};\boldsymbol{A},\boldsymbol{B})$ in equation 2.6, where λ is the regularization parameter which is a model parameter that has to be selected (details of which will be discussed in the next section). With regularization, the model equation 2.7 becomes

$$\vec{w_{out}} = (\boldsymbol{A}^H\boldsymbol{A} + \lambda\boldsymbol{I})^{-1}\boldsymbol{A}^H\boldsymbol{B}. \tag{2.9}$$

Weight decay in general forces the model to have weights with small values. This is in effect forcing the model to choose simpler functions over more complicated ones. Another interpretation as to why this approach works is that if the

weights are large they tend to amplify the noise in the training data. Another popular approach to regularization is L^1 regularization which results in sparse solutions (here the term $\lambda \|\vec{w}\|$ is added to $J(\vec{w}; \boldsymbol{A}, \boldsymbol{B})$). It is also common to combine both approaches.

Another approach that has worked in other domains is the addition of noise with a very small variance to the to the reservoir states or to the weights themselves. This again will also result in a norm penalty on the weights. In fact, the two approaches can be shown to be similar.

2.5.4 Cross-validation

Now that we have a description of the model, we turn to the specifics of how to get the best generalization error. Two specific issues need to be addressed: one is how to select the hyper-parameters of the model and second one is how to structure the training and evaluation phases of our procedure to get the best estimates of the performance of the model for out-of-sample data.

Up until now, our description of the training process involved using a so-called held-out test set where we split the total available data set in a single way into a train and test set. The test set was used to estimate the generalization error of the model after training. However, as highlighted above, we also need a way to select the best value for the regularization hyperparameter λ (see equation 2.9). If we however use the test set to determine λ, the generalization error will be underestimated [55]. The solution is to split the training data further into two disjoint sets: one set that will be used to train the model and a validation set that will be used to select the best value for the regularization parameter. The candidate regularization parameters themselves are generated by constructing a logarithmic grid that spans a large range of values. Once the best value for the regularization parameter is determined, the generalization error can then be evaluated as the performance on the test set.

However, this hold-out test approach (with a single split of the data in a train, validation and test set) is only applicable when we work with a large enough dataset. For smaller datasets, the same approach is likely to generate a high-variance model. This is because the testing performance will strongly depend on which particular samples ended up in the testing set versus the training set. For most of our work, we used small datasets to keep the training time to a minimum. Cross-validation provides a solution to this challenge by allowing us to use all our data for both train and testing.

The idea is to create a number of test/train splits and average the test results. One of the most popular ways of going about this is a technique known as k-fold cross-validation. The data is randomized and split into k disjoint subsets ('folds') and the training proceeds by taking each fold in turn and using it as the testing set

and using all the other folds as the training set. The performance of the model is then the average test error across all the folds and is a much better estimation of the performance of the model. Pseudocode for the k-fold cross-validation algorithm is given in Algorithm 1. As mentioned above, an advantage of this approach is that it allows us to use the available data efficiently (as each sample will be used once in the training and once testing set).

In this work, we made use of an extension of the cross-validation technique that simultaneously optimizes the hyperparameters (in our case the regularization parameter λ). Within each of the k trials, the procedure for selection of λ is the same as was described for the hold-out test approach above (see Algorithm 2 for details). In order to also estimate the error, we construct a nested cross-validation strategy where the inner cross-validation is used for model selection, while the outer cross-validation process is used to estimate the generalization error. The complete pseudocode for nested k-fold cross-validation is given in Algorithm 3. Of note is that, while any number can be picked for k, in practice the number of folds used is typically either 5 or 10. In this work we use $k = 5$

2.5.5 Training a readout with coherently combined states

One important extension to the training of photonic reservoirs is that we can take advantage of the complex-valued nature of the reservoir states, not only for the computing phase, but also during the training phase as we shall now discuss. One concern that should be apparent when it comes to applying reservoir computing to telecom tasks, as we have described in the sections above, is that we would require one high-speed receiver (photodetector, high-speed Analog-to-Digital Converter (ADC), slicer, ...) per output node (the extreme case being when we read out all the reservoir nodes). For real-time applications, this would imply an impractical and very costly system.

We have therefore been looking at a way to reduce the number of high-speed readouts required for a reservoir computing system in order to operate with reasonable accuracy. In order to achieve this, we perform the summation of the states in the analogue optical domain instead of in the digital electrical domain, so that we only need a single high-speed photodiode at the output (Figure 2.7).

A key difference between this approach of coherently combining the reservoir states and the offline training procedure described in section 2.5.2 is that the readout weights are applied to the states before they are processed by the detector. This means that the weights are always complex and in the optical domain. In terms of implementation, the weights could for example be implemented using Mach-Zehnder modulators. In theory, any other optical component (or combination of optical components) that can carry out amplitude and phase modulation could be used.

Algorithm 1 $CV(\mathbb{D}, \mathcal{A}, \mathcal{L}, \lambda)$
The k-fold cross-validation algorithm to generate an accurate estimate of the generalization error. It especially comes in handy for small datasets. The algorithm takes a dataset $\mathbb{D}$ with training examples, makes use of a learning algorithm $\mathcal{A}$ to learn a model f whose performance is measured with a loss function of $\mathcal{L}$ [55]. The final output is an average error across the folds.

Require:
A given dataset $\mathbb{D} = \{(x^{(i)}, y^{(i)}), \quad i = 1, ..., m\}$
Require:
A regularization parameter, λ.
Require:
A learning algorithm, $\mathcal{A}(\geq)$, a function that takes a dataset and returns a learned function f and makes use of λ to avoid overfitting.
Require:
A loss function, $\mathcal{L}$, that gives the scalar error between the estimate of the learned function on the example $(x^{(i)}, y^{(i)}) \in \mathbb{D}$

Split $\mathbb{D}$ into k mutually exclusive subsets $\mathbb{D}_i$ (i.e. $\mathbb{D} = \bigcup_{i=1}^{k} \mathbb{D}_i$ and $\mathbb{D}_i \cap \mathbb{D}_j = 0$)
for i from 1 to k **do**
  $f_i = \mathcal{A}(\mathbb{D} \setminus \mathbb{D}_i)$
  for $(x^{(j)}, y^{(j)})$ in $\mathbb{D}_i$ **do**
    $e_j = \mathcal{L}(f_i(x^{(j)}), y^{(j)})$
  end for
  Calculate the average error for the current fold
  $e_i = \langle e_j \rangle$
end for
Calculate the average error across folds
$e = \langle e_i \rangle$
return e

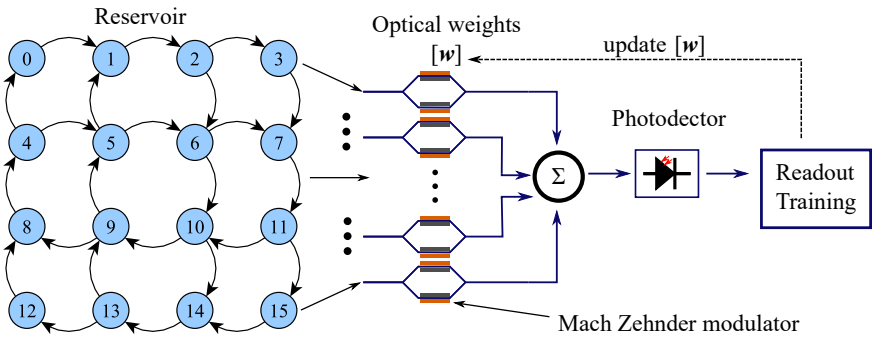


Figure 2.7: Schematic for training a 16 node photonic swirl reservoir with coherently combined reservoir states.

Algorithm 2 $CVWithOptimalLambda(\mathbb{D}, \mathcal{A}, \mathcal{L}, \Lambda)$
Implements k-fold cross-validation with optimization of the regularization parameter. While algorithm 1 uses a fixed regularization parameter (λ), it provides no leads on how to go about selecting its value to yield the best generalization error. To overcome that hurdle, this algorithm provides a solution by taking in a set of regularization parameters, usually constructed as a grid on a logarithmic scale, performs cross-validation on each of them and uses the resulting errors to determine the optimal value. The final output is the model generated by the learning algorithm using all the data and the found best value of the regularization parameter.

Require:
A given dataset $\mathbb{D} = \{(x^{(i)}, y^{(i)}), \quad i = 1, ..., m\}$
Require:
A learning algorithm, $\mathcal{A}$, a function that takes a dataset and returns a learned function f
Require:
A loss function, $\mathcal{L}$, that gives the scalar error between the estimate of the learned function on the example $(x^{(i)}, y^{(i)}) \in \mathbb{D}$
Require:
A set of regularization parameters, $\Lambda \subset \mathbb{R}$, usually drawn on a logarithmic scale.

for λ in Λ **do**
Determine
$e_\lambda = CV(\mathbb{D}, \mathcal{A}, \mathcal{L}, \lambda)$
end for
Find the best regularization parameter
$\lambda^* = \arg\min_\lambda e_\lambda$

Train a model with λ^*
$f^* = \mathcal{A}(\mathbb{D}, \lambda^*)$
return f^*

Algorithm 3 $NestedCV(\mathbb{D}, \mathcal{A}, \mathcal{L}, \Lambda)$
The k-fold nested cross-validation that makes use of algorithms 1 and 2 to provide an estimate of the generalization error while optimizing for the regularization parameter.

Require:
A given dataset $\mathbb{D} = \{(x^{(i)}, y^{(i)}), \quad i = 1, ..., m\}$
Require:
A set of regularization parameters, $\Lambda \subset \mathbb{R}$, usually drawn on a logarithmic scale.
Require:
A learning algorithm, $\mathcal{A}(\geq)$, a function that takes a dataset and returns a learned function f and makes use of λ to avoid overfitting.
Require:
A loss function, $\mathcal{L}$, that gives the scalar error between the estimate of the learned function on the example $(x^{(i)}, y^{(i)}) \in \mathbb{D}$

Split $\mathbb{D}$ into k mutually exclusive subsets $\mathbb{D}_i$ (i.e. $\mathbb{D} = \bigcup_{i=1}^{k} \mathbb{D}_i$ and $\mathbb{D}_i \cap \mathbb{D}_j = 0$)
for i from 1 to k **do**
    Learn
    $f_i^* = CVWithOptimalLambda(\mathbb{D} \setminus \mathbb{D}_i, \ldots)$

    Calculate the error on f_i^* for the current fold
    for $(x^{(j)}, y^{(j)})$ in $\mathbb{D}_i$ **do**
        $e_j = \mathcal{L}(f_i^*(x^{(j)}), y^{(j)})$
    end for
    Calculate the average error for the current fold
    $e_i = \langle e_j \rangle$
end for
Calculate the average error across folds
$e = \langle e_i \rangle$
return e

One challenge that arises is that we can no longer easily apply the model of equation 2.6 as it requires that we have access to all the states for all timesteps before calculating the readout weights. Indeed, using only a single photodiode loses the direct observability of all the states.

We considered two options for training this type of readout. The first option involves an iterative reconstitution of the amplitude and phase of the reservoir states [56, 57], based on presenting the input multiple times, but with only 1 or 2 weights turned on at any time. More details on this and related techniques will be detailed as part of Matthias Freiberger's doctoral dissertation.

2.6 Conclusion

In this chapter, we have introduced the theoretical underpinnings and techniques required to implement an integrated optical neuromophic information processing platform based on combining machine learning and integrated photonics.

In section 2.2 we introduced the field of machine learning including a common taxonomy of machine learning techniques into supervised, unsupervised and reinforcement learning approaches. We then focused on ANNs as a simplified model of a computation unit popularly used in machine learning. Sections 2.2.2 and 2.2.3 introduced FFNNs and RNNs, the two most popularly encountered ANN architectures.

We then introduced Reservoir Computing in section 2.3. RC started out as technique to train RNNs quickly and efficiently, but has grown to represent one of the most successful approaches to analog computing. Next, in section 2.4 we introduce photonic RC that takes advantage of light as a medium and photonics platforms to implement information processors for optical signals. Here we alo introduced integrated photonic reservoirs, particularly implemented in silicon photonics, as a way of achieving this processing on a CMOS compatible platform, that comes with an attractive path to commercialization. The we discussed a more recent approach to photonic RC that separates the nonlinear reservoir into a linear part and nonlinear part, presenting further advantages for faster processing and potential energy consumption benefits.

In section 2.5 we discussed the circuit model used in this work for discrete time simulations of the passive photonic reservoir.

Finally in section 2.5.2 we discussed the general procedure for training the reservoir to solve a specific task. We spent some time discussing techniques such as regularization and cross-validation that can be used to ensure that we get the best generalization performance for on the task we are attempting to solve.

References

[1] A. L. Samuel. *Some Studies in Machine Learning Using the Game of Checkers*. IBM Journal of Research and Development, 3(3):210–229, jul 1959.

[2] Thomas Cover and Peter Hart. *Nearest neighbor pattern classification*. IEEE transactions on information theory, 13(1):21–27, 1967.

[3] Thomas Bayes, Richard Price, and John Canton. *An essay towards solving a problem in the doctrine of chances*. 1763.

[4] J. Ross Quinlan. *Induction of decision trees*. Machine learning, 1(1):81–106, 1986.

[5] Corinna Cortes and Vladimir Vapnik. *Support-vector networks*. Machine learning, 20(3):273–297, 1995.

[6] Warren S McCulloch and Walter Pitts. *A logical calculus of the ideas immanent in nervous activity*. The bulletin of mathematical biophysics, 5(4):115–133, 1943.

[7] Bernard Widrow and Marcian E Hoff. *Adaptive switching circuits*. Technical report, Stanford Univ Ca Stanford Electronics Labs, 1960.

[8] Stuart Lloyd. *Least squares quantization in PCM*. IEEE transactions on information theory, 28(2):129–137, 1982.

[9] Jeffrey D Banfield and Adrian E Raftery. *Model-based Gaussian and non-Gaussian clustering*. Biometrics, pages 803–821, 1993.

[10] Karl Pearson. *LIII. On lines and planes of closest fit to systems of points in space*. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 2(11):559–572, 1901.

[11] Laurens van der Maaten and Geoffrey Hinton. *Visualizing data using t-SNE*. Journal of machine learning research, 9(Nov):2579–2605, 2008.

[12] Dana H Ballard. *Modular Learning in Neural Networks.* In AAAI, pages 279–284, 1987.

[13] R.S. Sutton and A.G. Barto. *Reinforcement Learning: An Introduction*. IEEE Transactions on Neural Networks, 9(5):1054–1054, sep 1998.

[14] David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, Yutian Chen, Timothy Lillicrap, Fan Hui, Laurent Sifre, George Van Den Driessche, Thore Graepel, and Demis Hassabis. *Mastering the game of Go without human knowledge*. Nature, 550(7676):354–359, 2017.

[15] J. Andrew Bagnell. *Reinforcement Learning in Robotics: A Survey*. Springer Tracts in Advanced Robotics, 97:9–67, 2014.

[16] Christopher JCH Watkins and Peter Dayan. *Q-learning*. Machine learning, 8(3-4):279–292, 1992.

[17] Gerald Tesauro. *Temporal difference learning and TD-Gammon*. Communications of the ACM, 38(3):58–68, 1995.

[18] Alec Radford, Luke Metz, and Soumith Chintala. *Unsupervised representation learning with deep convolutional generative adversarial networks*. arXiv preprint arXiv:1511.06434, 2015.

[19] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. *Generative adversarial nets*. In Advances in neural information processing systems, pages 2672–2680, 2014.

[20] F. Rosenblatt. *The perceptron: A probabilistic model for information storage and organization in the brain*. Psychological Review, 65(6):386–408, 1958.

[21] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. *Deep sparse rectifier neural networks*. AISTATS '11: Proceedings of the 14th International Conference on Artificial Intelligence and Statistics, 2011.

[22] Toma Mikolov, Stefan Kombrink, Luka Burget, Jan Cernocky, and Sanjeev Khudanpur. *Extensions of recurrent neural network language model*. ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, pages 5528–5531, 2011.

[23] Tomáš Mikolov, M Karafiat, Lukáš Burget, J Cernocky, and Sanjeev Khudanpur. *Recurrent Neural Network based Language Model*. Interspeech, (September):1045–1048, 2010.

[24] James Martens and Ilya@cs Utoronto Ca. *Learning Recurrent Neural Networks with Hessian-Free Optimization Ilya Sutskever*. Technical report, 2011.

[25] Wolfgang Maass, T Natschläger, and Henry Markram. *Real-time computing without stable states: A new framework for neural computation based on perturbations*. Neural computation, 2560:2531–2560, 2002.

[26] Herbert Jaeger. *The echo state approach to analysing and training recurrent neural networkswith an Erratum note 1*. Technical Re-port, GMD148. Bonn, Germany: German ..., pages 1–47, 2001.

[27] Herbert Jaeger and Harald Haas. *Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless communication.* Science (New York, N.Y.), 304:78–80, 2004.

[28] D. Verstraeten, B. Schrauwen, M. D'Haene, and D. Stroobandt. *An experimental unification of reservoir computing methods*. Neural Networks, 20(3):391–403, 4 2007.

[29] Herbert Jaeger. *Echo state network*. Scholarpedia, 2(9):2330, 2007.

[30] Mantas Lukoeviius and Herbert Jaeger. *Reservoir computing approaches to recurrent neural network training*. Computer Science Review, 3(3):127 – 149, 2009.

[31] Prashant Joshi and Wolfgang Maass. *Movement generation with circuits of spiking neurons*. Neural Computation, 2005.

[32] D. Verstraeten, B. Schrauwen, D. Stroobandt, and J. Van Campenhout. *Isolated word recognition with the Liquid State Machine: A case study*. Information Processing Letters, 2005.

[33] Jochen J. Steil. *Backpropagation-Decorrelation: Online recurrent learning with O(N) complexity*. IEEE International Conference on Neural Networks - Conference Proceedings, 2:843–848, 2004.

[34] Helmut Hauser, Auke Ijspeert, Rudolf Füchslin, Rolf Pfeifer, and Wolfgang Maass. *Towards a theoretical foundation for morphological computation with compliant bodies*. Biological Cybernetics, 105(2011):355–370, 12 2011.

[35] Henry O Sillin, Renato Aguilera, Hsien-Hang Shieh, Audrius V Avizienis, Masakazu Aono, Adam Z Stieg, and James K Gimzewski. *A theoretical and experimental study of neuromorphic atomic switch networks for reservoir computing.* Nanotechnology, 24:384004, 2013.

[36] Manjari S. Kulkarni and Christof Teuscher. *Memristor-based reservoir computing*. In Proceedings of the 2012 IEEE/ACM International Symposium on Nanoscale Architectures - NANOARCH '12, pages 226–232, New York, New York, USA, 2012. ACM Press.

[37] K Vandoorne. *Photonic reservoir computing with a network of coupled semiconductor optical amplifiers*. PhD thesis, 2011.

[38] Y Paquot, F Duport, A Smerieri, J Dambre, B Schrauwen, M Haelterman, and S Massar. *Optoelectronic Reservoir Computing*. Scientific Reports, 2:287, 2 2012.

[39] Quentin Vinckier, Franois Duport, Anteo Smerieri, Kristof Vandoorne, Peter Bienstman, Marc Haelterman, and Serge Massar. *High-performance photonic reservoir computer based on a coherently driven passive cavity*. Optica, 2(5):438–446, 2015.

[40] Daniel Brunner, Miguel C. Soriano, Claudio R. Mirasso, and Ingo Fischer. *Parallel photonic information processing at gigabyte per second data rates using transient states.* Nature communications, 4:1364, 1 2013.

[41] L Appeltant, M C Soriano, G Van der Sande, J Danckaert, S Massar, J Dambre, B Schrauwen, C R Mirasso, and I Fischer. *Information processing using a single dynamical node as complex system.* Nature communications, 2:468, 9 2011.

[42] L. Larger, M. C. Soriano, D. Brunner, L. Appeltant, J. M. Gutierrez, L. Pesquera, C. R. Mirasso, and I. Fischer. *Photonic information processing beyond Turing: an optoelectronic implementation of reservoir computing*. Optics Express, 20(3):3241, 1 2012.

[43] Franois Duport, Bendix Schneider, Anteo Smerieri, Marc Haelterman, and Serge Massar. *All-optical reservoir computing*. Optics Express, 20(20):22783, 9 2012.

[44] Antoine Dejonckheere, Franois Duport, Anteo Smerieri, Li Fang, Jean-Louis Oudar, Marc Haelterman, and Serge Massar. *All-optical reservoir computer based on saturation of absorption*. Optics Express, 22(9):10868, 5 2014.

[45] M. C. Soriano, S. Ortín, D. Brunner, L. Larger, C. R. Mirasso, I. Fischer, and L. Pesquera. *Optoelectronic reservoir computing: tackling noise-induced performance degradation*. Optics Express, 21(1):12, 1 2013.

[46] Romain Modeste Nguimdo, Guy Verschaffelt, Jan Danckaert, and Guy Van der Sande. *Fast photonic information processing using semiconductor lasers with delayed optical feedback: Role of phase dynamics*. Optics Express, 22(7):8672, 4 2014.

[47] Konstantin Hicke, M. Escalona-Morán, Daniel Brunner, Miguel Cornelles Soriano, Ingo Fischer, and Claudio R. Mirasso. *Information Processing Using Transient Dynamics of Semiconductor Lasers Subject to Delayed Feedback*. IEEE Journal of Selected Topics in Quantum Electronics, 19(4):1501610–1501610, 7 2013.

[48] Kristof Vandoorne, Joni Dambre, David Verstraeten, Benjamin Schrauwen, and Peter Bienstman. *Parallel reservoir computing using optical amplifiers.* IEEE transactions on neural networks, 22(9):1469–81, 9 2011.

[49] C Mesaritakis, V Papataxiarhis, and D Syvridis. *Micro ring resonators as building blocks for an all-optical high-speed reservoir-computing bit-pattern-recognition system*. JOSA B, (October), 2013.

[50] Martin Andre Agnes Fiers, Thomas Van Vaerenbergh, Francis Wyffels, David Verstraeten, Benjamin Schrauwen, Joni Dambre, and Peter Bienstman. *Nanophotonic reservoir computing with photonic crystal cavities to generate periodic patterns*. IEEE Transactions on Neural Networks and Learning Systems, 25(2):344–355, 2014.

[51] Hong Zhang, Xue Feng, Boxun Li, Yu Wang, Kaiyu Cui, Fang Liu, and Weibei Dou. *Integrated photonic reservoir computing based on hierarchical time-multiplexing structure*. Opt. Express, 22(25):31356–31370, 12 2014.

[52] Charis Mesaritakis, Alexandros Kapsalis, and Dimitris Syvridis. *All-optical reservoir computing system based on InGaAsP ring resonators for high-speed identification and optical routing in optical networks*. volume 9370, page 937033, 2 2015.

[53] Kristof Vandoorne, Pauline Mechet, Thomas Van Vaerenbergh, Martin Fiers, Geert Morthier, David Verstraeten, Benjamin Schrauwen, Joni Dambre, and Peter Bienstman. *Experimental demonstration of reservoir computing on a silicon photonics chip.* Nature communications, 5:3541, 1 2014.

[54] R. Penrose and J. A. Todd. *A generalized inverse for matrices*. Mathematical Proceedings of the Cambridge Philosophical Society, 51(03):406, jul 1955.

[55] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. `http://www.deeplearningbook.org`.

[56] Matthias Freiberger, Andrew Katumba, Peter Bienstman, and Joni Dambre. *Training Passive Photonic Reservoirs with Integrated Optical Readout*. oct 2018.

[57] Matthias Freiberger, Andrew Katumba, Peter Bienstman, and Joni Dambre. *On-chip Passive Photonic Reservoir Computing with Integrated Optical Readout*. In Rebooting Computing (ICRC), 2017 IEEE International Conference on, 2017.

3

Reservoir Computing with Multiple-Input Injection Strategy

The most obvious way to get the signal to be processed into a planar integrated photonic reservoir is through a single node, for example with a fiber grating coupler, after which it can propagate throughout the network to reach other nodes. This method is straightforward to implement and typically yields designs that do not require crossings which would create cross-talk. Examples of integrated reservoirs employing the single-input technique for tackling speech signal processing and bit-sequence processing tasks can be found in [1, 2]. Despite this success, single-input reservoirs suffer from major drawbacks due to the inherent limitations of the integrated photonics platform. Particularly, the losses increase with the size of the reservoir [3]. We therefore need to consider how loss reduction in such reservoirs could be achieved without sacrificing performance.

In line with this, the work in [1] introduced the idea that it may be beneficial to inject multiple copies of the input signal into the reservoir. However, only a very specific case of presenting the input to all nodes with different random phases was discussed. Here, we carry out a detailed investigation of the impact of the choice of the number and configuration of the input nodes on the robustness of the reservoir. Additionally, to improve the realism of our numerical simulations, we introduce a photodetector model at each readout node that takes into account bandwidth limitations, as well as optical and electrical noise properties encountered in real-world detectors [3]. With this model in place, we are able to examine, for the first time, the impact of the input power level on the performance and make conclusions

about the energy efficiency of various reservoir designs.

The details of the numerical model used for the simulation, together with the task setup will be discussed in section 3.1. In section 3.2, we will then study the performance for the *delayed* XOR task of reservoirs driven in various configurations across multiple data rates in section 3.2.1 and determine the power consumption associated with each configuration in section 3.2.2. Following this, we will choose the best input configuration that achieves error-free performance using the minimum input signal level in section 3.2.3.

The results in this chapter are an extended version of the paper we published in [3].

3.1 Methods

3.1.1 Numerical Model

The reservoir state update equation is as in equation 2.4. All the previous numerical work in our research group on integrated photonic reservoir computing has assumed perfect reconstruction of the states at the readout nodes. The absolute square value of the reservoir states (electric field values) was used as the input for the machine learning model. We now introduce a detector model that takes into account the responsivity, as well as various noise contributions and the response-time limitation encountered in real photodetectors [3]. The total noise σ_n^2 of the photodetector has shot noise and thermal noise contributions as follows:

$$\sigma_n^2 = 2qB(\langle I \rangle + \langle I_d \rangle) + 4k_B T B / R_L \tag{3.1}$$

where B is the bandwidth of the detector, $\langle I \rangle$ is the photocurrent, I_d is the dark current, q is the elementary particle charge, k_B is Boltzmann's constant, R_L is the load impedance and T is the temperature (in K).

The first part of equation 3.1 represents shot noise terms due to the input signal and the dark current, while the last part is the thermal noise contribution due to the detector load resistor. The bandwidth limitation of the detector is approximated by a low-pass filter with 3dB cutoff corresponding to the detector bandwidth. The filter operates on the states after adding the noise.

The output from the reservoir is then given as:

$$\vec{y}_{out} = \boldsymbol{W_{out}} \vec{x_{pd}} \tag{3.2}$$

where $\boldsymbol{W}_{out}$ are the linear output (readout) weights to be determined through training with ridge regression, and $\vec{x_{pd}}$ are the reservoir states after the photodetector.

Introducing this model for the detector dictates that we pay extra attention to the receiver power levels and in general the overall power budget of our systems,

to prefer designs that not only yield acceptable performance, but are also energy efficient [3].

For the input patterns (input configurations), we consider cases where input signal is injected into combinations of 2, 4, and 8 nodes, or into all 16 nodes. In each case we compare the resulting performance and energy requirements, for the temporal XOR task, to the case of input to a single node.

3.1.2 Simulation Setup and Parameters

The reservoir states are obtained as per equation 2.4 by propagating the inputs through a photonic reservoir model implemented in the photonics circuit simulator Caphe [4]. The photodetector used in the simulations is modeled based on the Alphalas UPD-15-IR2-FC photodector [5] that is available in our lab. The specific parameters used are a bandwidth of 25 GHz, a responsivity of 0.5 A/W (a pessimistic value as the datasheet value is 0.75 A/W) and a dark current of 0.1 nA. This NEP corresponds to an average signal power of 1.6 nW at an SNR of 10. It should be mentioned that the ultimate minimum power at the reservoir input will be set by the requirements of the downstream processing electronics.

Each combination of reservoir initialization (described in section 2.5.2) and input injection strategy was then tasked to solve the delayed XOR task. The current output bit for this task is the XOR of the current input bit with one n_{delay} bits in the past. Here we express it as:

$$y[n] = x[n] \oplus x[n - n_{delay}], \tag{3.3}$$

where $x[n]$ is the bit-level representation of the input data stream and $y[n]$ is the bit-level representation of the output. Before injection into the reservoir, the inputs $x[n]$ are converted from logical levels to discrete sampled data by upsampling and pulse shaping steps. We consider the delayed XOR task as it is the most difficult of all delayed binary tasks involving two bits. This is the case because, in machine learning terms, XOR is not linearly separable (see for example [6]).

For all considered input cases, the 4x4 (16 node) reservoir architecture was used to generate the states. This number of nodes was chosen as it is a design that is both cost-effective to produce with multi-project wafer runs, but also has a good performance on a number of tasks.

Once the states are obtained and transformed with the detector model, the readout is trained with a combination of the Oger machine learning toolbox [7] and the scikit-learn library [8].

We feed 10000 randomly chosen bits into the reservoir and use the resulting states for training with 5-fold cross validation to optimise the design parameters and yet another 10000 for testing. We use regularized ridge regression to train the linear readout. Testing is done on the best set of design parameters resulting from

the cross-validation. All reported error rates are related to the test data. With 10000 bits for testing, error rates are reported at a confidence level of about 90% [9].

We compare the performance of the multiple input reservoirs with the single-input reservoir architecture of [1], with the same total input power injected into the reservoir but distributed over the different input nodes to have a fair comparison. Note that in all cases, for the input and output coupling, we assumed a grating of $5\,\mathrm{dB}$ per grating.

3.2 Results and Analysis

3.2.1 Data Rate Studies

For the cases of single-input and multiple-input reservoirs, we studied the error rate of the reservoir across multiple data rates. To match the limitations of currently available measurement equipment in our lab, we restrict the maximal data rate to 32 Gbps. The data stream is a NRZ OOK modulated signal, which for simulation purposes is over-sampled 24 times to achieve sufficient simulation accuracy.

For a fair comparison between the different cases, the same aggregate input power of $100\,\mathrm{mW}$ across all input nodes was used. Where the input was fed into more than one node, the power was equally divided between the nodes. Results are reported as averages across 30 different random initializations of the input weights and reservoir waveguide phases (each using different randomly generated bit streams).

For plotting and interpreting the results, we make use of the the reservoir interdelay parameter r_{id}, which is defined as:

$$r_{id} = \frac{\tau_{bit}}{\tau_{id}}, \tag{3.4}$$

where τ_{bit} is the bit duration for the given data rate and τ_{id} is the interconnection delay time, corresponding to the the time it takes signals to propagate between two adjacent reservoir nodes. The reservoir interdelay parameter can be directly interpreted as the number of times the bit duration fits into the reservoir interconnection delay and can be used to identify under which regime the current computation is being carried out.

For the single-input simulations, we chose the input node from a representative sample of the available nodes as dictated by the symmetry of the swirl architecture relative to the central loop. The error rates for different reservoir interdelays are given in Figure 3.1 for input to nodes 0, 1, 2, 4 and 5 where the node labelling is the same as in Figure 2.5. The results show the typical single sharp minimum that translates into the reservoir only being able to process signals at a single data rate. The origin of these sharp optima is potentially due to the fact that as in single input reservoirs the signal is strongest around the input node (and quickly dies

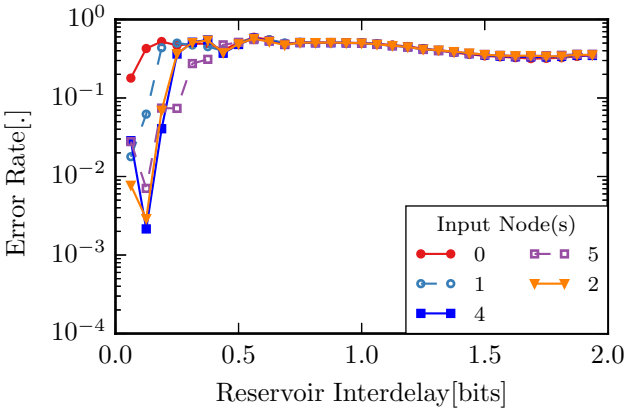


Figure 3.1: Error rate vs. reservoir interdelay for various nodes for the input to single node case. The minimum acceptable error rate is 1.0×10^{-3}

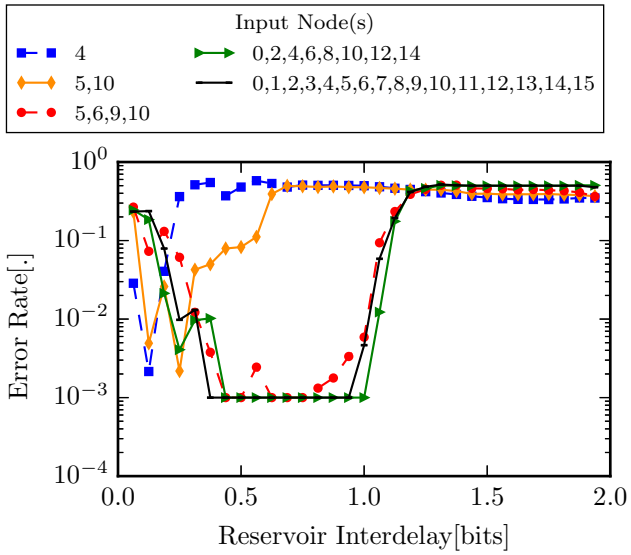


Figure 3.2: Error rate vs. reservoir interdelay for the different injection strategies. Minimum acceptable error rate is $10^{-}3$

away after that), the strongest contribution to the computation will be local and via single interference path corresponding to a single delay.

We can also conclude that proximity of the node to the central loop (nodes 5, 6, 9 and 10) is important for realizing low error rates on the task. Nodes 0 and 1, which are furthest away from the central loop, have the worst error performance while 4, 2, and 5, which inject either directly into the central loop or are only one hop away, yield the best performance.

For the multiple-input reservoir case, we consider input configurations involving simultaneous injection of the input bit stream into 2, 4, 8 or all 16 nodes of the reservoir. The input node combinations with best error rates in each of the groupings are plotted together in Figure 3.2. From the plot, we observe that in general the multiple-input reservoirs perform better than their single-input counterparts. As more reservoir nodes are driven, we discern the emergence of an increasingly wider basin in which the error is at or below the measurable minimum (10^{-3} in this case). The all-input case provides the widest basin. A wide basin implies more flexible architectures that can operate at multiple data rates. To change the data rate of operation, one simply has to re-train the reservoir readout for that data rate.

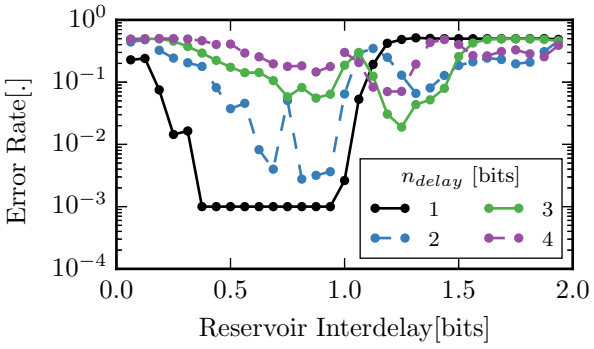


Figure 3.3: Error rate vs reservoir interdelay for the input to all nodes case. n_{delay} specifies the separation,in number of bits, of the two bits used for the XOR computation.

We further checked the influence of using multiple-input reservoir configurations on the computational power of the reservoir, more specifically its memory. Figure 3.3 depicts the error rates corresponding to the single-input versus the all-input case for various values of the delay in the temporal XOR task, n_{delay}. A larger n_{delay} corresponds to a task that requires more memory, because the current output bit is the XOR of the current input bit with a bit much further back in time.

For the single-input case, no error rates below 0.1 can be obtained for $n_{delay} > 1$. Even though for multiple-input reservoirs the performance similarly deteriorates

with increasing n_{delay}, it is clear that they can be operated for larger values of n_{delay}. This is because the useful signal (with a level significantly above the noise floor) remains present in the reservoir for a longer time.

3.2.2 Power level analysis

A key design guideline for signal processing systems for fiber-optic telecommunications systems is to keep the energy consumption as low as possible. In all our previous works, simulations assumed idealized detection of the reservoir states at each detection point for the readout nodes. In this work, on top of the search for the lowest error rate and robust reservoir designs, we now also look at how power efficiency maps to the different choices.

The data rates for the power sweeps were chosen at the minima of the error rate versus reservoir interdelay sweep curves (like the ones in Figure 3.2). The simulations were repeated 10 times for each reservoir design with different initializations

Figure 3.4 shows averaged error rates plotted against total input power.

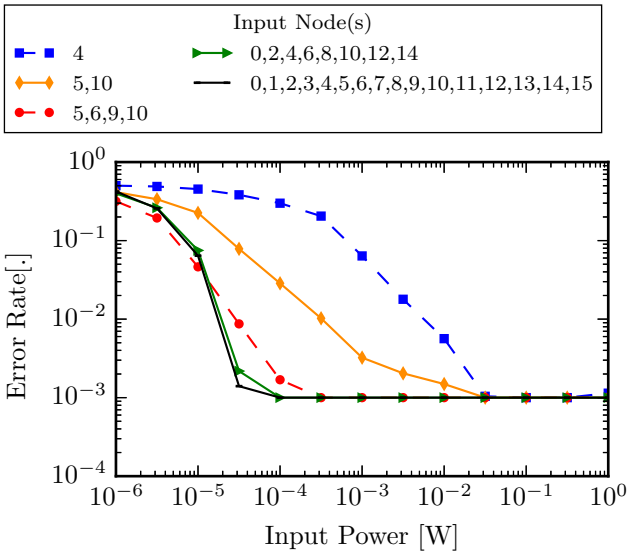


Figure 3.4: Error rate vs total input power for different injection scenarios. The minimum measurable error, given the number of bits used for testing, is 10^{-3}.

We observe that as we increase the number of the input nodes, the minimum power requirements for error-free performance also go down. The most significant jump in power efficiency is an approximately 2 orders of magnitude decrease for the best 4-input node combination as compared to the 1 or 2 node input combinations. This can be attributed to the fact that the [5, 6, 9, 10] combination is the

central loop in the swirl architecture which allows for significant signal distribution for a small number of inputs. We also observe that increasing the number of input nodes beyond 4 does not significantly impact the power efficiency. Since each input that needs to be driven incurs an additional hardware cost, we can conclude that driving the central four nodes is the most cost and power-efficient solution for the 16 node swirl architecture.

Looking in more detail at what happens inside the reservoir, Figures 3.5, 3.6 and 3.7 show the average power intensity in all reservoir nodes for the cases of single-node input, input to the central loop and input to all nodes respectively. For the single-node input case, the power decreases significantly within a few hops from the driving node. As an example, node 8, which is just below node 4, has more than 10 dB less power than node 4. When all nodes are driven, the power is most evenly distributed across all the nodes. This scenario also corresponds to the best power efficiency (3 orders of magnitude higher than the best single input case) obtained in our simulations. With the power injected in the central loop nodes only, the power efficiency lies between the two extreme cases. In this instance, there is still a significant subset of the reservoir nodes with similar power levels and only the furthest nodes exhibit a power drop of more than 5dB compared to the input nodes.

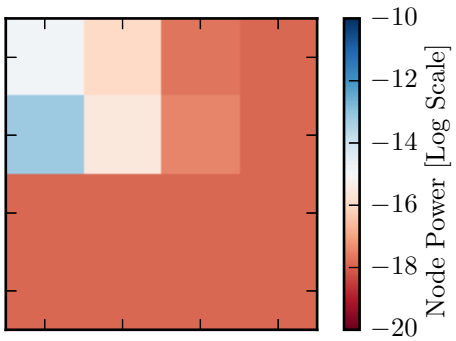


Figure 3.5: Average power distribution over the reservoir nodes for input to node 4.

3.2.3 Optimal design

Simulation results from sections 3.2.1 and 3.2.2 above indicate that injection of power into the central nodes of the reservoir, [5, 6, 9, 10] (see Figure 3.8), provides the best combination of performance and energy efficiency.

Figures 3.9 and 3.10 illustrate the bounds of the errors for the results within 1 standard deviation of average over the repetitions for error rate studies and power

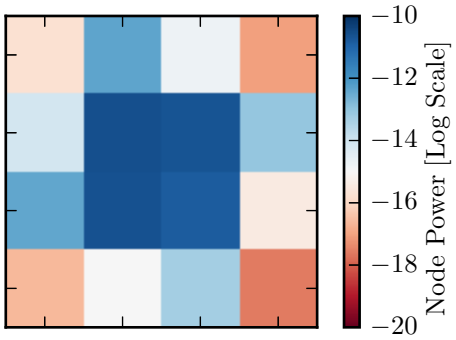


Figure 3.6: Average power distribution over the reservoir nodes for input to the central loop .

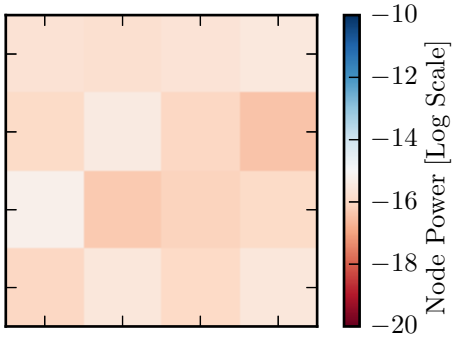


Figure 3.7: Average power distribution over the reservoir nodes for input to to all nodes.

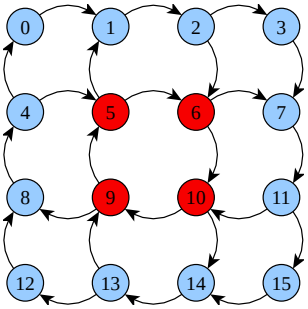


Figure 3.8: Schematic of the 16-node swirl reservoir architecture highlighting the central loop nodes that yield the optimal input configuration. This injection strategy results in the best performance - energy efficiency operation mode for this reservoir architecture.

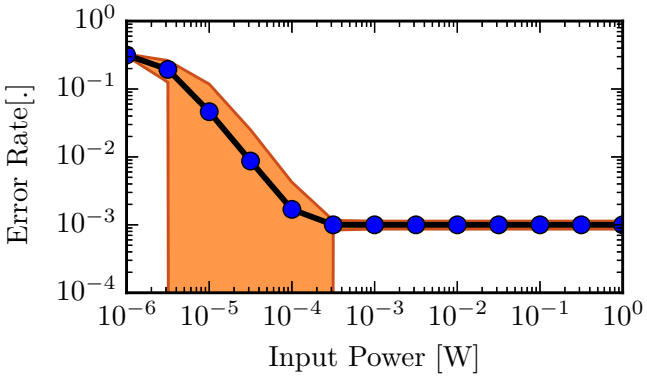


Figure 3.9: Error rate vs total input power for input to the central swirl loop (nodes [5, 6, 9, 10]). The solid line indicates the mean value over all repetitions while the shaded areas indicate the error bounds within 1 standard deviation of the mean.

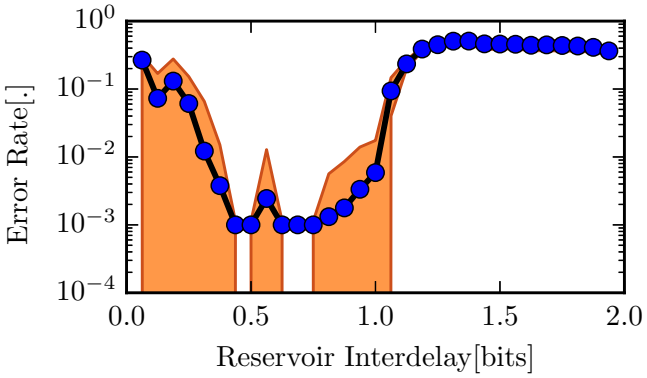


Figure 3.10: Error rate vs reservoir interdelay for input to the central swirl loop (nodes [5, 6, 9, 10]). The solid line indicates the mean value over all repetitions while the shaded areas indicate the error bounds within 1 standard deviation of the mean.

level studies respectively. Unsurprisingly, the transition regions between the zones of best performance and those with the highest error rates have the highest uncertainty. This uncertainty can be shrunk by, for example, considering a larger number of bits in the test dataset. Concerning the minimum input power for this design, and since the voltage required for the subsequent machine learning electronics is on the order of a few mV, the equivalent power at the input of the reservoir is on the order of a few mW. In comparison, the most complete energy efficiency calculation for an optical reservoir can be found in [10] for a fully nonlinear reservoir based on a laser with feedback. The authors reported a power consumption of 10mJ per bit for the speech processing task. In [11], a minimum input power of 0.57mW at the input is reported for the coherently driven passive cavity reservoir with a fiber loop. The optimal reservoir design's energy efficiency is therefore similar, and coupled together with the advantages of integration, makes for a state-of-the-art signal processing platform.

3.3 Conclusion

In this chapter we have presented an exploration of the input injection configurations available in a planar, passive integrated photonic reservoir computing system. Error rates obtained from circuit simulations establish that multiple-input injection reservoirs perform better than single-input reservoirs for a larger number of data rates. The key benefit of using multiple inputs is that they tend to result in a more even power distribution landscape across the nodes of the reservoir. This creates a larger usable richness in the reservoir since more signals with roughly similar amplitudes are mixed. Moreover, multiple input-designs present a better power efficiency and so present better odds for correct extraction of all reservoir states, since there are more nodes that have power that is higher than the noise floor. Moreover, with more input points, the signal tends to stick around longer in the reservoir which increases the reservoir memory allowing for processing of more challenging signals.

Another reason for these performance improvements is that the varied mixing between the multiple copies of the input signals with different phases translates into increased computational power of the reservoir. This is because there is a richer signal from which the machine learning algorithm can extract useful features.

However, driving more nodes comes at an additional hardware cost, because the optical signals need to be distributed to all nodes. Since most of the improvement in robustness and power efficiency is obtained by driving the four central nodes instead of just one, we consider this to be the most promising and cost-effective solution for 4×4 reservoirs. In its current state, this optimal design requires a few mW of input power. The next step would be research on how bring-

ing this value further down, for example, by reducing the the internal losses in the reservoir, or by using more compact architectures in which losses do not scale directly with reservoir sizes. I should mention here that while in this work we focused on crystalline silicon waveguides, a number of ultra-low loss integrated photonics waveguide platforms exist that could also be used as platforms for photonic reservoir computing (see e.g. [12]). The loss reductions techniques presented in this thesis can be directly applied on those platforms. Morever at the data rates considered in this work, and the correspoding delay lines, the linear propagation loss in the waveguides is not the critical source of loss but rather the splitter/combiner loss across a given path in the reservoir.

In the next chapter, we will show an approach to reducing these internal losses in the reservoir that relies on multimode photonics.

References

[1] Kristof Vandoorne, Pauline Mechet, Thomas Van Vaerenbergh, Martin Fiers, Geert Morthier, David Verstraeten, Benjamin Schrauwen, Joni Dambre, and Peter Bienstman. *Experimental demonstration of reservoir computing on a silicon photonics chip.* Nature communications, 5:3541, 1 2014.

[2] K Vandoorne. *Photonic reservoir computing with a network of coupled semiconductor optical amplifiers*. PhD thesis, 2011.

[3] Andrew Katumba, Matthias Freiberger, Peter Bienstman, and Joni Dambre. *A Multiple-Input Strategy to Efficient Integrated Photonic Reservoir Computing*. Cognitive Computation, pages 1–8, 4 2017.

[4] Martin Fiers, Thomas Van Vaerenbergh, Ken Caluwaerts, Dries Vande Ginste, Benjamin Schrauwen, Joni Dambre, and Peter Bienstman. *Time-domain and frequency-domain modeling of nonlinear optical components at the circuit-level using a node-based approach*. J. Opt. Soc. Am. B, 29(5):896–900, 5 2012.

[5] *Ultrafast InGaAs PIN photodetector UPD-15-IR2-FC*, 2012.

[6] Sergios Theodoridis and Konstantinos Koutroumbas. *04 Nonlinear Classifiers*. Pattern Recognition (Fourth Edition), pages 151–260, 2009.

[7] David Verstraeten, Benjamin Schrauwen, Sander Dieleman, Philemon Brakel, Pieter Buteneers, and Dejan Pecevski. *Oger: Modular Learning Architectures For Large-Scale Sequential Processing*. Journal of Machine Learning Research.

[8] F Pedregosa, G Varoquaux, A Gramfort, V Michel, B Thirion, O Grisel, M Blondel, P Prettenhofer, R Weiss, V Dubourg, J Vanderplas, A Passos, D Cournapeau, M Brucher, M Perrot, and E Duchesnay. *Scikit-learn: Machine Learning in {P}ython*. Journal of Machine Learning Research, 12:2825–2830, 2011.

[9] M Jeruchim. *Techniques for Estimating the Bit Error Rate in the Simulation of Digital Communication Systems*. IEEE Journal on Selected Areas in Communications, 2(1):153–170, 1 1984.

[10] Daniel Brunner, Miguel C. Soriano, Claudio R. Mirasso, and Ingo Fischer. *Parallel photonic information processing at gigabyte per second data rates using transient states.* Nature communications, 4:1364, 1 2013.

[11] Quentin Vinckier, Franois Duport, Anteo Smerieri, Kristof Vandoorne, Peter Bienstman, Marc Haelterman, and Serge Massar. *High-performance photonic reservoir computer based on a coherently driven passive cavity*. Optica, 2(5):438–446, 2015.

[12] Jared F. Bauters, Martijn J. R. Heck, Demis John, Daoxin Dai, Ming-Chun Tien, Jonathon S. Barton, Arne Leinse, René G. Heideman, Daniel J. Blumenthal, and John E. Bowers. *Ultra-low-loss high-aspect-ratio Si_3N_4 waveguides*. Optics Express, 19(4):3163, feb 2011.

4

Multimode Reservoir Computing

Silicon photonics reservoirs are composed of nodes that are interconnected together in a planar topology such as that of Figure 2.5. Since the interconnections between nodes are made up of spirals of a few centimeters, the material loss – around $2\,\mathrm{dB/cm}$ for single-mode $220\,\mathrm{nm}$ Si waveguides – is important. In passive reservoirs, in the absence of an intervening on-chip amplification mechanism, the signal is exponentially attenuated. An equally (often more) significant source of signal loss of the architecture used in our group until now, is the use of Y-junctions which results in a loss at the combiner points (a combiner/splitter component is needed at each point where light paths merge or split). Indeed, based on supermode theory, combining single-mode waveguides in a Y-junction only has 100% transmission if the two inputs are exactly in phase. For anti-phase inputs, the transmission is 0%. Therefore, averaged over all possible phase differences, there is only a 50% transmission for each combiner traversed. An alternative way of expressing the same fact is by saying that if we only excite a single input of the combiner, we will have only 50% transmission. Later on we will use this single-side excitation as a quicker way to model the average transmission for different phase differences. The accumulation of subsequent 50% losses quickly leads to substantial overall loss values for large reservoirs with a lot of combiners. Both this combination loss and the propagation loss constrain the performance gain that can be achieved by scaling up reservoirs to more nodes. This in turn limits the complexity of the tasks a single reservoir can tackle [1].

In this chapter we focus on lowering the combination loss in passive integrated photonic reservoirs by using multimode waveguides. An alternative ap-

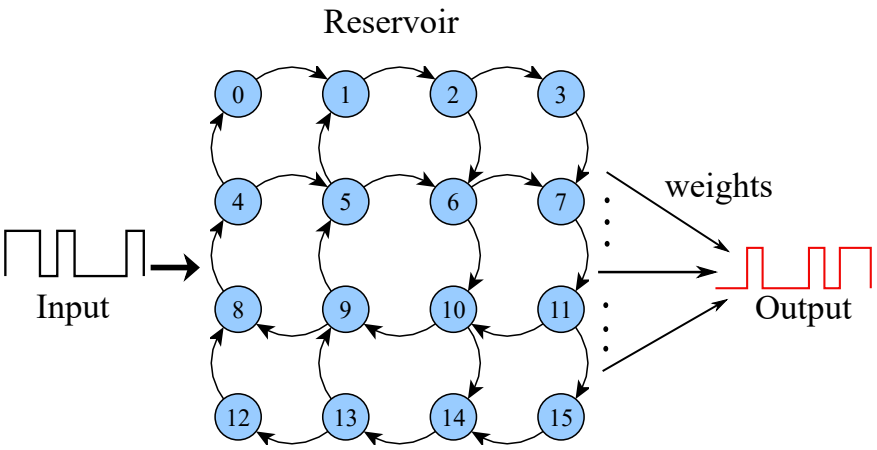


Figure 4.1: Schematic of a photonic reservoir computing setup for handling tasks involving digital optical signals. The input is a non-return-to-zero on-off-keying (NRZ-OOK) digital optical signal, the reservoir is composed of 16 nodes arranged in a swirl topology and the output is the bitstream that results from training the readout to solve a particular task.

proach based on 3×3 MMI devices will be described in section 6.3. With the loss build-up curtailed, the number of nodes in the reservoir, and hence its computational power, can be increased significantly. We show that even a relatively modest improvement in trans-nodal transmission will yield a substantial overall gain, since splitting and combining of signals occurs a multitude of times before the signal is read out. The improvement hinges on using broader waveguides which support multiple modes.

One way to take advantage of multiple modes in reservoirs is by considering them as distinct channels of computation and reading them out separately through a demultiplexing structure such as a cascade of asymmetrical directional couplers [2]. The total number of modes supported by the network represents the factor of increase of the number of observables in the system and is an indicator for the performance improvement. A multimode reservoir also directly implies richer dynamics due to multiple mixing avenues that could happen between the modes. However, this approach is outside the scope of this thesis.

A second way to take advantage of the multimodal character is the one that we pursue here, where we focus on how it can be leveraged for loss reduction. A critical component in this design is a novel multimode Y-junction structure that is used at the combiner/splitter points. The junction employs a taper section that is deliberately designed to be not perfectly adiabatic, ultimately resulting in energy efficiency benefits. Note that multimodal photonics is something which is typically not considered for other applications, because it leads to several complications (modal dispersion, more complex design, difficulty to selectively excite and main-

tain a specific number of modes throughout the whole circuit, ...). However, in the context of the RC computing paradigm, none of these are of any consequence.

The key advantage of using the multimode Y-junction comes from the fact that a portion of the light that was previously scattered into radiation modes of the single-mode structure can now be captured into the higher-order guided modes. Indeed, in a case where e.g. light is only sent into one of the two input arms of a single-mode Y-combiner, when entering the output waveguide 50% of the light radiates away, as mentioned previously. However, if the output waveguide of the combiner supports multiple modes, this transmission increases to 100% for a perfectly adiabatic taper. Still, since part of this combined light is now carried by a higher-order mode as opposed to the fundamental mode, it is still a fact that this higher-order light will radiate away at the next combiner. This is where the next part of the design comes into play: by deliberately designing the Y-junction to be non-adiabatic to a certain extent, we can hope to get a degree of conversion from these higher-order modes back to the lower-order modes which can propagate unhindered. Still, this beneficial process competes with the harmful reverse process of conversion from lower-order to higher-order modes. Therefore, it is not a-priori obvious that there will be a design that offers a bigger than 50% transmission averaged over different modal compositions of input excitation. We present exactly such a design .

The rest of this chapter is structured as follows. In section 4.1, we discuss the techniques used to design a low-loss multimode Y-junction combiner and the setup of the circuit simulations to be used to characterize its performance when used in a photonic reservoir. Section 4.2.1 explains in detail the steps taken to reach the final design parameters for the Y-junction, including waveguide width, taper optimizations and wavelength dependence. Finally, section 4.2.2 discusses the performance and energy efficiency of the identified optimal Y-junction design on the 3-bit header recognition task as compared to a single-mode reservoir of the same size using a symmetric Y-junction combiner.

The results in this chapter are an extended version of the paper we published in [1]. The work was performed together with Jelle Heyvaert, a master student we supervised.

4.1 Methods

4.1.1 Electromagnetic simulations

For calculating the eigenmodes of the waveguides, Lumerical© Mode Solutions, a commercial mode solver package, was used. The Film Mode Matching Method-based solver in Fimmwave© was used to double-check the results. For the Lumerical© varFDTD method selected for the propagation simulations and optimisation of the

parameters of the Y-junctions, an auto non-uniform meshing strategy was chosen at level 5 accuracy and a conformal variant 1 mesh refinement was applied. The center wavelength was set to $1300\,\mathrm{nm}$ and a silicon-oxide cladding of refractive index 1.4469 surrounding the $220\,\mathrm{nm}$ high silicon waveguide structure with material parameters according to Palik's model in the Lumerical material database was used. The geometries of the structures were designed with Luceda Ipkiss© Photonic Design Environment, exported into the GDSII format and then imported into Lumerical©. The varFDTD method was selected as it enables fast simulations (2D FDTD simulation speeds) with an accuracy approaching that of 3D-FDTD simulations for devices with omnidirectional in-plane propagation such as the Y-junction considered in this work.

4.1.2 Circuit simulations and task setup

We use custom time-domain circuit simulation scripts based on the Caphe software [3] and scikit-learn library [4] for the machine learning. Unless stated otherwise, a 4×4 (16 node) reservoir architecture was used to generate the states. As in the previous chapter, this number of nodes was chosen as it is a design that is both cost-effective to produce with multi-project wafer runs, but also has a good performance on a number of tasks. In all cases, the length of the interconnections between the reservoir translates to a propagation time of $62.5\,\mathrm{ps}$, matching the current generation of experimentally available chips [5].

The methodology is very similar to the one in the previous chapter. Once the states were obtained and transformed with an appropriate photodetector model taking into account various noise contributions and bandwidth limitations, the readout was trained using the scikit-learn library [4]. For the training, 10,000 randomly chosen bits were fed into the reservoir and the resulting states were used for training with 5-fold cross-validation to optimise the model hyperparameters and yet another 10,000 for testing. We used a regularised ridge regression algorithm to train the linear readout. Testing is done on the best case resulting from the cross-validation. All reported error rates are related to the test data. With 10,000 bits for testing, error rates are reported at a confidence level of about 90% [6].

4.2 Results

4.2.1 Design of a multimode Y-junction with low combiner loss

4.2.1.1 Waveguide cutoff conditions

To effectively design multimode waveguides and components, we must first determine the cutoff conditions for different modes. This has to be done numerically, as the well-established closed form 1D solution cannot easily be extended to practical

2D cases. Lumerical© Mode Solutions was used to determine the modes of the waveguides. The fully vectorial FFM solver in Fimmwave©, which is based on the Film Mode Matching Method, was used to check the results for consistency.

We assume oxide-clad $220\,\mathrm{nm}$ silicon at $1300\,\mathrm{nm}$ on the SOI material platform. The reason for this wavelength is a requirement set by the materials used for modulation and amplification in the H2020 PHRESCO project. The project goal is to build large scale photonic reservoirs that can tackle more real-world telecom tasks (see chapter 6 for details). From here on, we will only focus on the TE-like modes but a similar argument applies for TM-like modes (we also confirmed that there was negligible interaction between the two mode groups). From the mode simulations we obtain the cut-off points for each mode and can therefore choose the waveguide width to support a given number of modes (Figure 4.2).

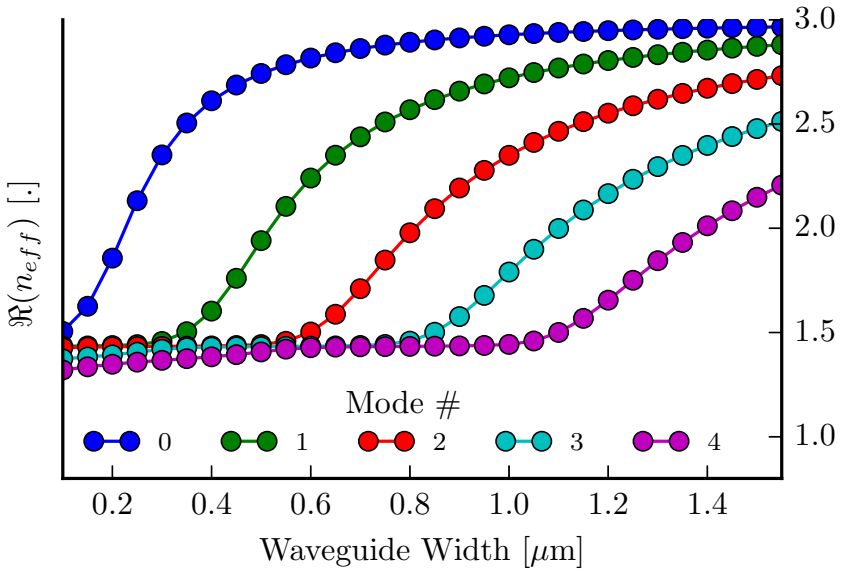


Figure 4.2: Dispersion diagram for $220\,\mathrm{nm}$ *SOI waveguide for the TE polarization and a center wavelength of* $1300\,\mathrm{nm}$. n_{eff} *is the complex effective index of the mode. [1]*

4.2.1.2 Relevant geometrical parameters of Y-junction

The Y-junction design for the simulations was composed of three sections as seen in Figure 4.3. The input waveguide (section 1) of width w_1 is followed by a linear taper of length t (section 2) leading into a wider waveguide with width w_2. The third section starts off right after the taper and terminates into a split into the two output waveguides. The output waveguides (also referred to as arms) are all the

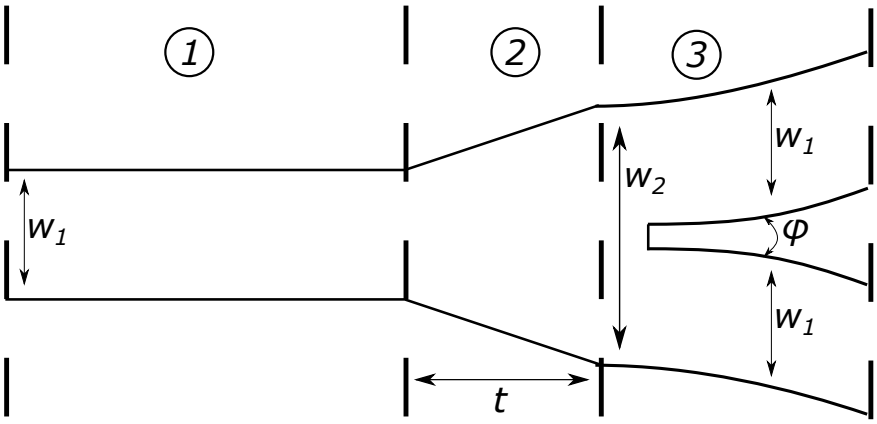


Figure 4.3: Sketch of the Y-junction indicating the sections critical to its performance [1].

same width w_1 as the input waveguide of section 1. ϕ is the angle between the two output waveguides and is determined by the bend radius of the output waveguides. The smaller the bend radius, the larger the value of ϕ and vice versa.

We shall refer to the splitter configuration as the case when the input is on the left side of Figure 4.3 in section 1. For the combiner configuration, the junction is operated in the reverse sense.

In contrast to the single objective criterion of maximizing adiabaticity for the single-mode Y-junction, designing a low loss multimode Y-junction combiner is a more intricate procedure, as the interactions and evolution of the modes through the junction drive the choice of the best geometrical parameters. Since sweeping all the geometric parameters would be very time-consuming, the simulations proceeded as follows. For all designs considered, $w_2 = 2w_1$, and the bend radius for the two arms is set to $5\,\mu\mathrm{m}$. Subsequently, the taper length t was fixed and the width of the input and output waveguides w_1 was scanned in the range $600\,\mathrm{nm}$ - $1200\,\mathrm{nm}$. Finally, the taper length t was optimized to achieve maximum transmission for the Y-junction with the selected width.

4.2.1.3 Waveguide Width Optimisation

A first relevant parameter is the input and output waveguide width w_1 for the Y-junction, for which we will compare four values, inspired by different regimes in the dispersion diagram of Figure 4.2. At $600\,\mathrm{nm}$ width, the first two modes (i.e. fundamental and first-order mode) are strongly guided. A third mode exists, but is guided less strongly. At a width of $800\,\mathrm{nm}$, the first three modes are well-guided, but now a fourth mode (i.e. third-order mode with effective index of 1.46 with a cladding index of 1.4469) is weakly guided. The situation for a $1000\,\mathrm{nm}$ wide waveguide has evolved further towards four well-guided modes and a fifth (i.e. the

fourth-order mode) that is barely guided with an effective index 1.4495. The last design, with a width of $1200\,\mathrm{nm}$, guides this fourth-order mode better and does not support the fifth-order mode yet.

The transmission was then simulated for these Y-junction designs through propagation simulations. As mentioned, for this purpose Lumerical© FDTD solutions, a commercial FDTD software package, was used.

If we consider the transmission results from the simulations for the combiner configuration with input to the upper arm only as presented in Table 4.1, the gains of moving to wider waveguides are evident.

Each column in the table represents transmission values for excitation with a given input mode to a single arm only. The final column is an average transmission over all the previous columns for guided modes. Note that in a real reservoir, the modal composition at the input will take many different forms, and therefore this average transmission is used as a straightforward way to derive an approximate but relevant single figure-of-merit for the device, without having to deal with fully coherent multimodal simulations of an entire RC network. From the results, it is evident that for wider waveguides and junctions, the losses are substantially smaller than for smaller waveguides and junctions by virtue of the power ending up in higher order guided modes rather than being radiated away. For the waveguide widths considered, the highest average transmission is obtained in the $1000\,\mathrm{nm}$ - $1200\,\mathrm{nm}$ width range (e.g. for a width of $1000\,\mathrm{nm}$ the average transmission is 53.18%, to be compared with the 50% transmission for the single-mode combiner). One can argue that maximum transmission occurs in this region because it constitutes the widths at which power conversion between the different modes is most efficient for the given taper length t = $0.1\,\mu\mathrm{m}$ (t is chosen adhoc for this section). However, to reach the maximum of transmission for both the waveguide widths w_1 and taper lengths t, we need to probe the influence of the taper length t on the transmission. This will be the focus of the *Taper Optimization* section 4.2.1.5 below.

4.2.1.4 Modal Power Distribution

Before optimizing the taper length t, we first make a detour to check the output modal power composition of the multimode Y-junction with respect to input excitation of a particular mode. We focus on the case of a $1\,\mu\mathrm{m}$ wide Y-junction with a $0.1\,\mu\mathrm{m}$ taper. Table 4.2 represents the transmission from the upper arm into the input in the combiner configuration, while table 4.3 shows the transmission from the input into upper arm in the splitter configuration.

First, we observe that indeed several modes participate in the guiding of power through the junction, and that the distribution of power over the different modes is very different for different input configurations. We can also see some modal conversion taking place from input in higher-order modes to output in lower-order,

	Combiner transmission (input to one arm only)					
width w_1 (nm)	fundamental	1^{st} order	2^{nd} order	3^{rd} order	4^{th} order	average transmission
600	89.14%	28.41%	0.29%	-	-	39.28%
800	76.84%	70.55%	17.36%	0.45%	-	41.30%
1000	89.99%	55.25%	55.02%	12.47%	-	53.18%
1200	93.47%	75.95%	17.49%	41.27%	3.97%	46.43%

Table 4.1: Transmission for Y-junction combiners of different input and output waveguide widths w_1 when the input is a particular modal excitation. The results here correspond to the case of excitation to the upper arm only. Missing values indicate that that particular mode is not guided for that waveguide width.

better guided modes.

Second, the results allow us to verify the reciprocity of the device (to within simulation tolerances). For example, the transmission from the fundamental mode to the first-order mode in the combiner configuration in Table 4.2 is 41.94%, whereas the transmission from the first-order mode to the fundamental mode in the splitter configuration is 41.90% in Table 4.3.

	Transmission from upper arm to input (combiner)				
input source	fundamental	1^{st} order	2^{nd} order	3^{rd} order	4^{th} order
fundamental	34.94%	41.94%	4.67%	5.01%	3.44%
1^{st} order	0.46%	6.38%	16.12%	20.72%	11.56%
2^{nd} order	5.07%	1.30%	26.43%	17.70%	4.52%
3^{rd} order	6.61%	0.05%	0.39%	2.90%	2.51%

Table 4.2: Modal decomposition of Y-junction combiner with $1\,\mu\mathrm{m}$ *wide waveguides and* $0.1\,\mu\mathrm{m}$ *long taper for single excitation from upper arm.*

	Transmission from input to upper arm (splitter)				
input source	fundamental	1^{st} order	2^{nd} order	3^{rd} order	4^{th} order
fundamental	34.95%	0.56%	5.05%	6.52%	0.36%
1^{st} order	41.90%	6.52%	1.36%	0.05%	0.01%
2^{nd} order	4.67%	16.10%	26.72%	0.39%	0.01%
3^{rd} order	5.06%	20.77%	17.88%	2.94%	0.48%

Table 4.3: Modal decomposition of Y-junction splitter with $1\,\mu\mathrm{m}$ *wide waveguides and a* $0.1\,\mu\mathrm{m}$ *long taper.*

4.2.1.5 Taper Length Optimisation

In section 4.2.1.3 above, we demonstrated that by going to multimode we can gain a small improvement in the transmission of the Y-junction combiner when averaged over many possible input modal excitations. However, to continue to boost the transmission, we need to further tailor the adiabaticity of the evolution of the input modeset by making appropriate changes to the geometry. We choose to do this by altering the taper making up section 2 of the design. The simulations for the choice of the waveguide width were done with a taper length t of $0.1\,\mu\mathrm{m}$. We now select the $w_1 = 1.0\,\mu\mathrm{m}$ design, which we found to be in the region of highest transmission in the first phase of our simulations, and track the transmission for taper lengths $t = 0.1\,\mu\mathrm{m}$, $1\,\mu\mathrm{m}$, $2\,\mu\mathrm{m}$ and $2.5\,\mu\mathrm{m}$.

The resulting transmission values for the multimode Y-junction combiner are plotted in Figure 4.4 against taper length t for the case of excitation in the upper arm only. As input excitations, we use the fundamental, 1st, 2nd and 3rd order

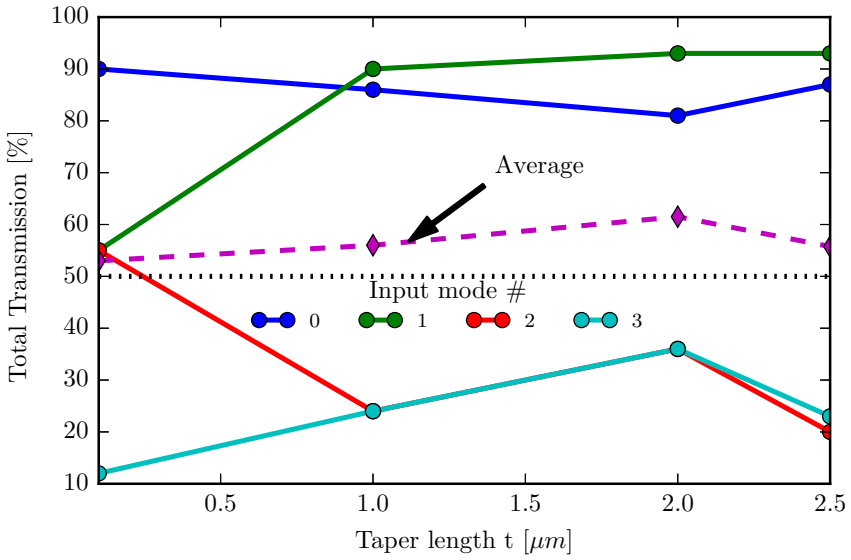


Figure 4.4: Total transmission (summed over all output modes) in the Y-junction combiner for different taper lengths. Results are shown for input to the upper arm of the junction consisting of the fundamental (0), 1st , 2nd and 3rd order modes as well as the average transmission across all input modes. The baseline transmission of 50%, for the adiabatic single-mode Y-junction, is also indicated.

mode. Each solid line corresponds to the denoted input mode and at the output a sum of the transmission across all output modes is plotted as the total transmission. The average transmission across all input modes per taper length t is also plotted and is the same figure-of-merit as used in the last column of Table 4.1. As anticipated, the optimization has resulted in higher transmission values as can be seen by comparing with the results of the $t = 0.1\,\mu\mathrm{m}$ case (which was used in a previous section to determine the best waveguide width for the multimode Y-junction) to the longer taper lengths.

As we alluded to earlier, there are two conflicting trends at work which result in an optimum average transmission occurring at taper lengths around $2.0\,\mu\mathrm{m}$. On one hand, increasing the taper length increases the adiabaticity which decreases the scattering losses for the guided supermodes and increases the radiation loss of the unguided supermodes. Note that also for this multimode junction, in the limit of a perfectly adiabatic taper we expect 50% average transmission: 100% transmission for modes 0 and 1, and 0% transmission for modes 2 and 3, which will couple to unguided supermodes. On the other hand, having a certain degree of non-adiabaticity is beneficial to convert some of the higher-order modes into better guided lower-order modes, as shown earlier. This is the mechanism that allows us to get above a 50% average transmission. When looking at the transmission of e.g. modes 2 and 3 as a function of taper length, there is no clear trend, indicating a complicated modal mixing and conversion taking place.

At this point, although we could further optimize the design, simulation results already show an average transmission of 61% for the combiner, which is much better than the standard 50% average loss in single-mode junctions.

As this is the design we will be using in the reservoir, we carry out further simulations to check its performance when operated in the splitter configuration and obtain 42% average transmission per output arm. While the transmission in the splitter is lower than the 50% of the fully adiabatic single-mode case, we will confirm later that the improvement in the transmission of the combiner will yield net gains in the power efficiency of the reservoir.

4.2.1.6 Wavelength dependence

We additionally evaluated the impact the wavelength of operation could have on the performance of a passive reservoir system that uses this multimode Y-junction design. We track the transmission to the fundamental mode of the output Y-junction combiner from input excitation with the fundamental mode in a $40\,\mathrm{nm}$ bandwidth around $1300\,\mathrm{nm}$. From the results for the Y-junction of width w_1 set to $1\,\mu\mathrm{m}$ and taper length t set to $2\,\mu\mathrm{m}$ we can conclude that variations in the wavelength will not impact the performance of the reservoir. The transmission remains at $\approx 32\%$ for the entire wavelength range. Similar behaviour is observed for other input-output mode pairs.

4.2.2 Performance of the improved Y-junctions in photonic reservoirs

We focused our numerical simulations on a model of a 16-node passive integrated photonics reservoir with the nodes arranged in a swirl topology [5] (see Figure 4.1) with the input power injected into all nodes.

As mentioned, in the design of the multimode Y-junction, we used as a figure-of-merit the average transmission across all considered excitations with different modes. Therefore, we need to setup our simulations in a way that matches with this scenario. Specifically, at all points where a combiner is needed we use 61% transmission, and similarly for all splitting locations we use splitters with 42% efficiency. Additionally, as we are taking average powers across all modes, our reservoir simulations will not be coherent (as was the case in most of our previous works) but will rather work with the intensity of the signals.

The reservoir was tasked to solve the 3-bit header recognition task. The setup for the header recognition task is similar to that of the XOR task described in section 3.1.2 with the exception of the readout. For 3-bit header recognition there are $2^3 = 8$ unique headers meaning that 8 classifiers need to be trained, one per header. Each classifier is trained to give a +1 output at the last bit of the header and -1 if the header is not detected. As the readout generates real-valued outputs, the winner-takes-all (WTA) mechanism is used to select which header is present [5]. The input data stream is an NRZ OOK signal with an oversampling factor of 24 and the maximum considered data rate is 32 Gbps. A total input power of $100\,\mathrm{mW}$ was used and for each data rate errors were obtained for 10 different random initialisations of the input bit stream and input weights.

4.2.2.1 Error rates for different data rates

We evaluated and compared the error rates of the reservoir on the 3-bit header recognition task for multiple data rates for single-mode and multimode reservoirs. Figures 4.5 and 4.6 show the results from the error rate vs reservoir interdelay [1] experiments for the single-mode and the multimode reservoir cases respectively, for the case of input to node 0. For the most part there is no significant difference in performance when going from single to multimode reservoirs, which means there is no performance hit associated to moving to multimodal Y-junctions. In both cases we have regions of performance below the Soft Decision Forward Error Correction (SD-FEC) limit (corresponding to a BER of 2×10^{-2}) which means we are able to reach error-free performance by applying FEC codes. To see the real benefit of this work, however, we need to look at energy efficiency.

[1] Recall that this is a normalization of the data rate to the interconnection delay between two adjacent nodes.

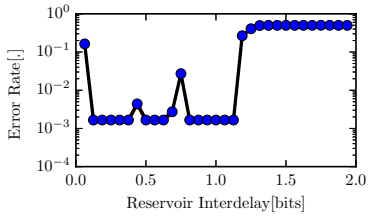


Figure 4.5: Error rate vs reservoir interdelay for the single-mode reservoir for the 3-bit header recognition task.

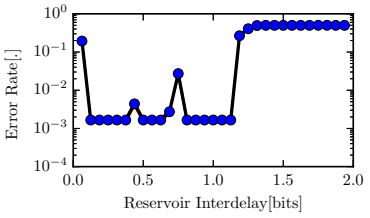


Figure 4.6: Error rate vs reservoir interdelay for the multimode reservoir for the 3-bit header recognition task.

4.2.2.2 Energy efficiency considerations

The key goal of this work is to demonstrate that we can achieve energy efficiency benefits by replacing a single-mode reservoir with a multimode version. First, if we track the error rate for the single-mode and multimode reservoirs on the 3-bit header recognition task against the input signal-to-noise ratio as plotted in Figure 4.7, we observe that errors for the multimode reservoir are always lower than those of the single-mode case. The trend also tends to show higher divergence for higher signal-to-noise ratios. While at the level of the SD-FEC limit the difference seems small, it should be noted that we anticipate that this diverging trend will persist as we characterize for even lower BERs.

Next, we compare the overall loss of a single-mode reservoir with that of a multimode reservoir. Figure 4.8 indicates the per-node power ratios for the 16 node passive reservoir for the multimode versus single-mode case for the case of input to node 0. This power ratio is obtained by dividing the average power of the states at each node in the multimode reservoir by the power of the same node in the single-mode reservoir and is shown for the 3-bit header recognition as we considered for the performance evaluation. Because node 0 is used as the input in both cases, its power did not change and consequently its power ratio is equal to one. All other nodes have a power ratio above the dashed red line that demarcates where the ratio is equal to one. This means that more power is measured at all those nodes in the multimode reservoir. We observe that we can get up to 20% more power in nodes such as 4, 8 and 12 (they also happen to be the furthest from the input node in terms of optical path length).

We carried out one more experiment to further investigate how the improvement in the component energy efficiency affects the power distribution in the reservoir as we move to larger reservoirs. We re-simulated the 3-bit header recognition task but this time with a 6×6 (36 node) swirl reservoir. We observe, as shown in Figure 4.9, that we now have nodes that have more than 30% improvement in their power level (see e.g. nodes 6, 12, ...). One can expect for the gains to increase even

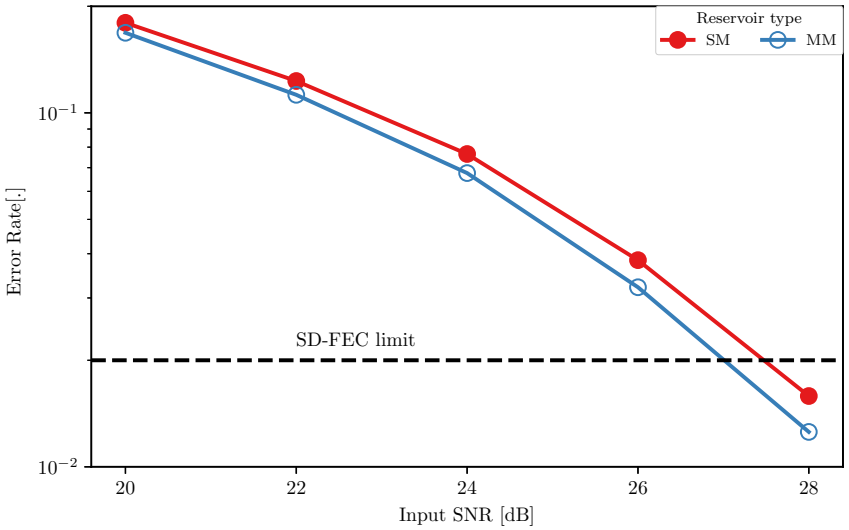


Figure 4.7: Error rates for single-mode and multimode on the 3 bit header recognition task for different values of the input SNR.

further as we move to larger networks. The additional power boost obtained for multimode reservoirs could be the difference between being below or above the noise floor. This could have significant impact on the performance and scalability of a real-world photonic reservoir computing setup.

4.3 Conclusion

We can improve the overall energy efficiency of silicon photonics integrated circuit reservoirs by replacing the typical single-mode components and waveguides with multimode versions. FDTD electromagnetic simulations show that we can design a low-loss multimode Y-junction with widths ranging from $500\,\mathrm{nm}$ onwards but with a preference for $1000\,\mathrm{nm}$ to keep the circuit size down while keeping the advantageous aspects of mode evolution at such a junction.

The best designs can be implemented with taper lengths around $2.0\,\mu\mathrm{m}$ at an optimum level of adiabaticity. The best design can yield a combiner with an average transmission of 61% as opposed to the 50% in the best case single-mode junction combiner. The same device operating as a splitter gives 42% transmission, less than the 50% of the single-mode case. However, when the device is used in a reservoir, the improvements in the combiner transmission make up for this reduction in the splitter performance. Further parameter optimization can be used to get even better improvements on the loss. One approach that could par-

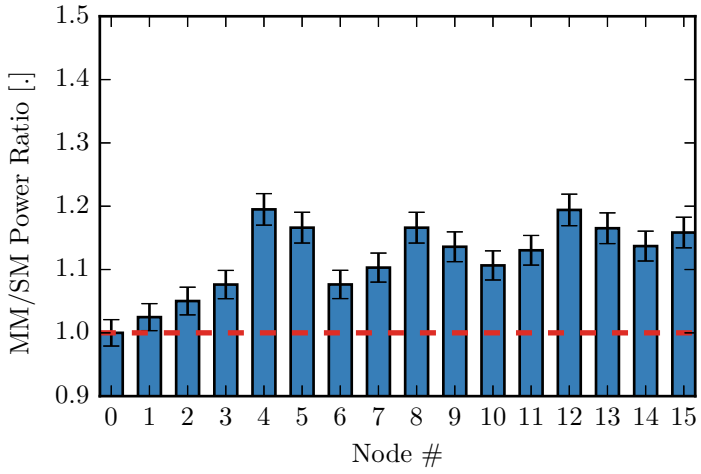


Figure 4.8: Comparison of single-mode and multimode 16 node reservoir average power per node for input to node 0.

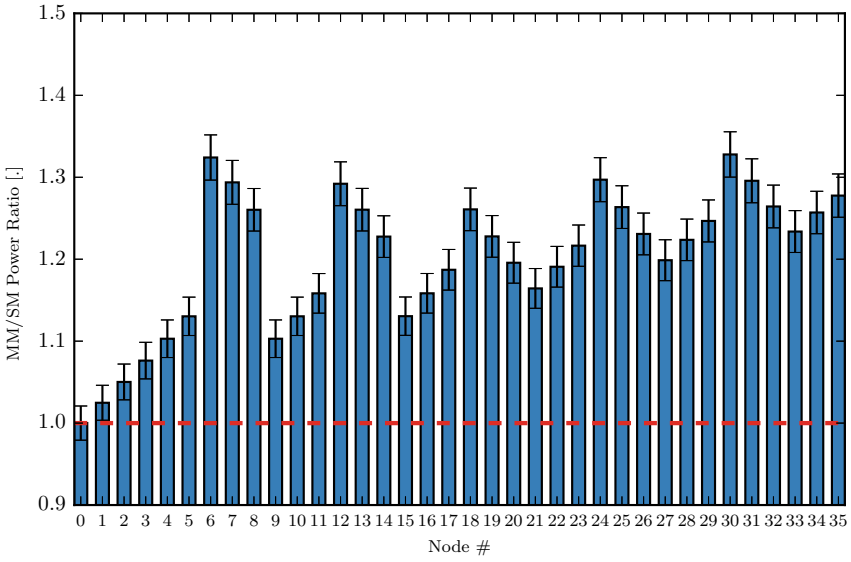


Figure 4.9: Comparison of single-mode and multimode 36 node reservoir average power per node for input to node 0.

ticularly come in handy is that of inverse design [7, 8]. The figure of merit would have to be defined to simultaneously minimize losses while improving combiner transmission.

Constructing a reservoir based on the average transmission characteristics for our optimal Y-junction combiner/splitter design yields the same task performance but gains in energy efficiency, when computing with intensities of the input signal. For single-node input to the reservoir, energy at some nodes could be enhanced up to as much as 30% in the case of a 36-node reservoir. Overall, for the 36-node multimode reservoir we gain approximately 15 to 20% in effective reservoir size over the 36-node single mode reservoir. This highlights how the small improvements in a single component design yield benefits that scale with the size of the reservoir. Such an improvement paves the way for building reservoirs with a larger number of nodes than is currently possible and that can therefore solve more complex tasks.

References

[1] Andrew Katumba, Jelle Heyvaert, Bendix Schneider, Sarah Uvin, Joni Dambre, and Peter Bienstman. *Low-Loss Photonic Reservoir Computing with Multimode Photonic Integrated Circuits*. Scientific Reports, 8(1), 2018.

[2] Daoxin Dai, Jian Wang, and Yaocheng Shi. *Silicon mode (de)multiplexer enabling high capacity photonic networks-on-chip with a single-wavelength-carrier light*. Optics Letters, 38(9):1422, 5 2013.

[3] Martin Fiers, Thomas Van Vaerenbergh, Ken Caluwaerts, Dries Vande Ginste, Benjamin Schrauwen, Joni Dambre, and Peter Bienstman. *Time-domain and frequency-domain modeling of nonlinear optical components at the circuit-level using a node-based approach*. J. Opt. Soc. Am. B, 29(5):896–900, 5 2012.

[4] F Pedregosa, G Varoquaux, A Gramfort, V Michel, B Thirion, O Grisel, M Blondel, P Prettenhofer, R Weiss, V Dubourg, J Vanderplas, A Passos, D Cournapeau, M Brucher, M Perrot, and E Duchesnay. *Scikit-learn: Machine Learning in {P}ython*. Journal of Machine Learning Research, 12:2825–2830, 2011.

[5] Kristof Vandoorne, Joni Dambre, David Verstraeten, Benjamin Schrauwen, and Peter Bienstman. *Parallel reservoir computing using optical amplifiers.* IEEE transactions on neural networks, 22(9):1469–81, 9 2011.

[6] M Jeruchim. *Techniques for Estimating the Bit Error Rate in the Simulation of Digital Communication Systems*. IEEE Journal on Selected Areas in Communications, 2(1):153–170, 1 1984.

[7] J.S. Jensen and O. Sigmund. *Topology optimization for nano-photonics*. Laser & Photonics Reviews, 5(2):308–321, mar 2011.

[8] Alexander Y. Piggott, Jesse Lu, Konstantinos G. Lagoudakis, Jan Petykiewicz, Thomas M. Babinec, and Jelena Vučković. *Inverse design and demonstration of a compact and broadband on-chip wavelength demultiplexer*. Nature Photonics, 9(6):374–377, jun 2015.

5

Nonlinear Compensation in optical communication links using passive photonic reservoirs

There is currently significant industrial effort towards extending the reach and increasing the capacity of optical communication networks to meet user demands. This is however, no trivial undertaking, as the various elements constituting fiber-based lightwave networks also contribute to the degradation of optical signals during generation, transmission and reception phases [1]. In this chapter, we propose an approach that can be applied to combat these unwanted signal modifications using passive integrated photonic reservoir computing chips.

The taxonomy of fiber optic communication impairments consists of linear and nonlinear forms [2]. Linear impairments include Chromatic Dispersion (CD), Polarization Mode Dispersion (PMD), symbol timing offset and optical filtering. Typically encountered nonlinear impairments are laser phase noise, Self Phase Modulation (SPM), Cross Phase Modulation (XPM), Four Wave Mixing (FWM) and nonlinear phase noise. In this thesis, as we focus on single-wavelength communication, XPM and FWM do not apply. Similarly, the inelastic nonlinearities of Stimulated Brillouin Scattering (SBS) and Stimulated Raman Scattering (SRS) are not considered in this work as they occur at power levels that are not of interest to the telecom links studied here.

Traditionally, linear imperfections in optical communication networks have been addressed with well-known techniques: optically, using dispersion compen-

sation fibers and dispersion-shifted fibers, and electrically, using tapped delay line based Feedforward Equalizers (FFEs) or Decision Feedback Equalizers (DFEs), or a combination of both. The advent of coherent optical communications for 100G and beyond has generated an increase in reliance on DSP technology as these networks almost exclusively rely on the DSP for their linear and nonlinear impairment mitigation, as well as a plethora of other transmitter and receiver-side processing tasks. However, the non-linear Shannon limit, which limits the use of higher powers to increase capacity as nonlinear effects set in, is evidence that these electronic (DSP-based) solutions cannot reliably tackle the issue of non-linear compensation. Similarly, proposed all-optical solutions come with numerous disadvantages that hinder their large-scale adoption. Even so, the same DSPs (typically with some reduction in complexity) have been used to implement various functions in Intensity Modulated-Direct Detection (IM-DD) links for high-speed short-reach applications involving PAM-4 and DMT such as in [3] and [4]. As a consequence, there is a lot of interest in solutions that seek to either entirely replace DSP functions (DSP-less) or simplify the complexity of the task that has to be solved by the DSP chip (and consequently bring their wall plug efficiency within manageable limits) for both coherent and IM-DD optical communications systems.

In this chapter, we present our proposal for a novel signal equalizer to undo both impairments in optical communication networks based on reservoir computing and study its performance in IM-DD links. Our approach operates in the optical domain, meaning that it is very fast, compact (only 16 nodes in a footprint of $\approx 16\,\mathrm{mm}^2$) and energy efficient. Additionally, being CMOS-compatible means that we can take advantages of CMOS mass production benefits for scaling. Moreover the device can be packaged together with receiver electronics in, for example, a transceiver module.

The equalizer is implemented using an integrated photonic reservoir computing (PhRC) system and is the same size and design as that in [5]. A high level representation of RC used for nonlinear compensation is shown in Figure 5.1 where the procedure for obtaining the distorted input signal will be explained in the next section. Conceptually, the PhRC equalizer is a high-dimensional complex nonlinear filter that acts in the optical domain to undo distortions. As the the reservoir has memory, it combines delayed versions of the input signal (similar to a tapped filter) and additionally performs a nonlinear transformation of the input signal (which is necessary for creating rich nonlinear combinations of delayed versions of the input signal) that can be used to combat distortions such as those induced by SPM and CD. The integrated PhRC equalizer will naturally benefit from the advantages of the CMOS platform in terms of scalability and energy efficiency. The design is flexible and can be applied to virtually all sorts of links without limit in terms of required bandwidth. Designing a $40\,\mathrm{Gb/s}$ or a $400\,\mathrm{Gb/s}$ equalizer is a matter of changing the internal timescale of the reservoir (set by the length of the delay

lines); the ultimate limit lies in the readout electronics, making the PhRC equalizer a future-proof option to signal equalization. Once deployed, the equalizer can easily be retrained in the field (on-demand or on a schedule) to adapt to the link conditions. The PhRC equalizer could also be deployed just before a DSP to reduce the computational load associated with cleaning up severely distorted signals.

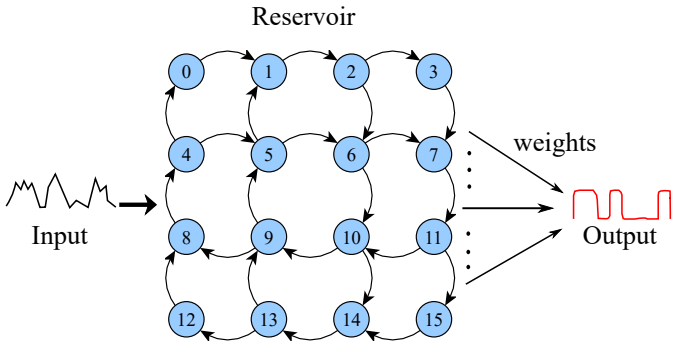


Figure 5.1: Schematic representation of the photonic reservoir computing setup as is used in this dissertation for impairment removal. The input signal is a distorted non-return-to-zero on-off-keying (NRZ-OOK) signal that has traversed a fiber optic link, the integrated photonics reservoir is composed of 16 nodes arranged in a swirl topology and the reservoir states are recorded at each node with aid of a photodetector. These states are then recorded and used to train a set of weights that represent the readout function that generates the final output signal. Note that in the reservoir, the numbered lightlight blue circles (n_i) are the nodes in the reservoir and in this architecture are the locations where states are combined and split, and also serve as the input and detection points.

The rest of this chapter is structured as follows. In section 5.1, we present the general simulation setup, training details and machine learning techniques used in the study. This will be followed by a presentation and discussion of the results in section 5.2, for a variety of different links, ranging from metro over short-reach to long-haul. We then provide a summary of results and a conclusion in section 5.3

The work in this chapter is an extended version of a paper submitted to the Journal of Lightwave Technology [6].

5.1 Simulation Setup and Methodology

The optical communication links were set up in VPI Transmission Maker v9.8 software. The setup used for the metro link is given in Figure 5.2. For the high-speed short-reach links the schematic in Figure 5.7 was used, while for the long-haul link we used the setup in Figure 5.10. The VPI software incorporates a wealth of validated models of transmitter and receiver components as well as implementations of

all signal degradation mechanisms encountered as signals traverse optical telecom links. Standard single-mode fiber (SSMF) was used in all cases with parameters given in Table 5.1.

Parameter	Value	Units
Attenuation	0.2	$\mathrm{dB/km}$
Dispersion Coefficient	16×10^{-6}	$\mathrm{s/m^2}$
Dispersion Slope	0.08×10^{3}	$\mathrm{s/m^3}$
Nonlinear Index	2.6×10^{-20}	$\mathrm{m^2/W}$
Core Area	80.0×10^{-12}	$\mathrm{m^2}$

Table 5.1: Key parameters for SSMF used in transmission simulations.

After traversing the link, the electric field of the signal was recorded and saved offline, to be processed by the numerical circuit level model of the passive integrated photonics reservoir. In an experimental setup, this coupling from the fiber into the reservoir chip could for example be achieved using fiber-to-chip grating couplers. We should mention that, as we operate in a single polarization (Transverse Electric (TE) in this case), a polarization controller component is inserted just before the reservoir chip to ensure maximum transmission. The simulation model was built with custom time-domain circuit simulation scripts based on the Caphe software [7]. In all cases a 4×4 (16-node) reservoir architecture was used to generate the states. This number of nodes was chosen as it yields a design that is cost-effective to produce with multi-project wafer runs, and has achieved good performance on a number of tasks [5], as well as the impact of other model and training parameters, on the performance of the PhRC equalizer will further be discussed in subsequent sections.

The states recorded from each reservoir node were transformed by a receiver model to take into account various noise contributions and bandwidth limitations . The specific model used is for a PIN photodetector with responsivity $1.0\,\mathrm{A/W}$ and thermal noise of $3.0 \times 10^{-12}\,\mathrm{A}/\sqrt{\mathrm{Hz}}$. It has a 3$^{\mathrm{rd}}$ order low pass Bessel electrical bandpass filter (ELBF) with cut-off frequency corresponding to 0.7 times the bit rate of the input signal. Shot noise in the PIN photodetector was accounted for as well.

After the receiver model, the states were collated and used to train the readout weights using the scikit-learn library [8]. For each run, 10,000 randomly chosen input PRBS7 bits were used and the resulting states (after the receiver) were used for training with 5-fold cross-validation to optimize the model hyperparameters. Note that for each 1,000 bits a different random seed was used for the PRBS signal. The regularized ridge regression algorithm was used to train a linear readout. The reported bit error rate (BER) is the average validation error across folds. This process was repeated 8 times for different random initializations of the link condi-

tions and reservoir phases. For each training run, the internal reservoir phases and the input phases are randomly generated and fixed.

The BER was calculated using VPI based on Gaussian statistics and optimization of the sampling instant. To calculate the BER error measure used in the readout training, the received signal is first sampled at a sampling instant (the optimal sampling instant is determined by simply checking at which of the samples the minimum error occurs). The resulting sampled signal has the average current value for the ones fluctuating around I_1 with standard deviation σ_1 while the zeros have an average value of I_0 with standard deviation σ_0. To make a decision on the current sample, a threshold current value I_D has to be determined such that whenever $I > I_D$ the decision is for a "1" bit, and a "0" for when $I < I_D$ [9]. The bit error rate can be defined as

$$BER = p_1 P_{0/1} + p_0 P_{1/0} \tag{5.1}$$

where p_1 and p_0 are the probabilities of a "1" and a "0", respectively in the original bitstream, and $P_{1/0}$ and $P_{0/1}$ represent the probability of making the wrong decision. Specifically, $P_{0/1}$ is the probability of receiving a "1" and deciding the bit is a "0", and $P_{1/0}$ is the probability of deciding the current bit is a "0" when actually a "1" was received. Assuming gaussian statistics for both the thermal noise and shot noise currents, the BER can be rewritten as [9]

$$BER = \frac{1}{2}\Big[p_1 erfc(\frac{I_1 - I_D}{\sigma_1\sqrt{2}}) + p_0 erfc(\frac{I_D - I_0}{\sigma_0\sqrt{2}})\Big] \tag{5.2}$$

where $erfc(.)$ is the complementary error function. The decision threshold I_D is determined through minimizing the BER by, for example, setting up an optimization problem and solving it with e.g. the Newton-Raphson method (there are also a number of simplifications to this approach).

To check that the results were not subject to the pitfalls of applying Neural Networks in Optical Communications outline in [10] , the training procedure was repeated using the much longer PRBS31 sequence. Results showed that their was no BER estimation bias. From here-on, we will therefore only discuss results related to the PRBS7 signals.

5.2 Results and Discussion

5.2.1 Metro links

In this section, we investigate the performance of the PhRC equalizer on unrepeated fiber optic communications links in the metro regime for fiber lengths ranging from $100\,\mathrm{km}$ to $250\,\mathrm{km}$ for a $10\,\mathrm{Gb/s}$ NRZ OOK link. The setup is shown in Figure 5.2. The amplifier, with a noise figure (NF) of 4.0 (the same amplifier is

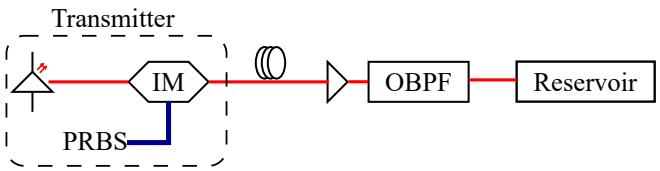


Figure 5.2: Schematic representation of the simulation setup to generate data for the signal equalization task. The input pseudo-random bit sequence (PRBS) signal is modulated onto a laser signal, transmitted over a fiber link, amplified and filtered, after which the field of the optical signal is saved to file to be used as input to the photonic reservoir simulation model. OBPF - Optical band pass filter.

used throughout the work), is set to entirely undo the link attenuation. The filter gets rid of out-of-band noise and is a 3rd-order Bessel filter with bandwidth equal to 4 times the data rate. Note that this data rate is chosen here to keep it compatible with the measurement capability in our characterization lab, but the same procedure can be followed for higher-speed links.

First, we study the influence of the node interconnection delay time on the BER of the PhRC equalizer. The node interconnection delay is the most important parameter for passive planar integrated photonics reservoirs as it determines the timescale at which signals interfere within the reservoir. It is set by changing the length of the on-chip delay lines at design time. The transmitter (consisting of the pattern generator, continuous wave (CW) laser and external Lithium Niobate modulator) parameters are given in Table 5.2. The symmetry factor is an indicator of the chirp behaviour of the modulator and is a measure of the imbalance of the drive-induced phase shifts in its two arms. A value of -1 means ideal intensity modulation with no chirp, meaning that the two drive voltages are equal in magnitude but out of phase.

Parameter	Value	Units
	Laser	
Emmision Wavelength	1550	nm
	Modulator	
Symmetry factor	-1	-
Extinction Ratio	20	dB
Rise time	0.4 $\times$ bit period	-

Table 5.2: Transmitter parameters for metro link setup.

As seen in Figure 5.3, the regime of best operation of the PhRC equalizer corresponds to an interconnection length that corresponds to a delay of half the bit duration. This result demonstrates that for these link parameters, the condition of

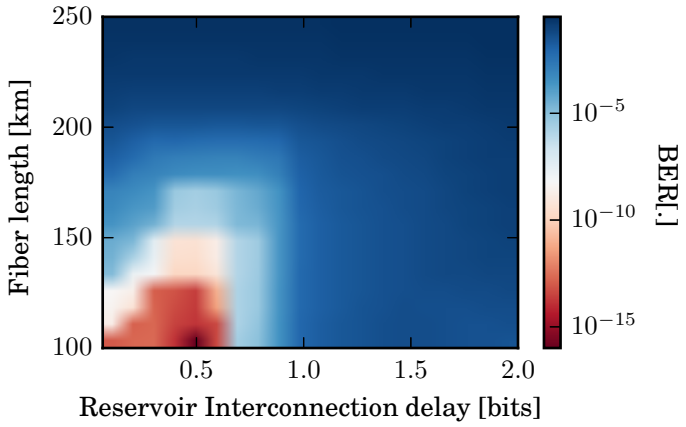


Figure 5.3: Error rates vs reservoir interconnection delay and fiber length for 16 node PhRC equalizer for an NRZ-OOK link at $10\,\mathrm{Gb/s}$.

the optimal interconnection delay coincides with those for the header recognition and bit level tasks of [5]. For the rest of the discussions in this section we therefore fix the interconnection length of the reservoir to this value.

As in any causal system, the reservoir output requires a finite time before responding to the input. We therefore need to find the delay between the input and the corresponding desired output for which the reservoir can best solve the task. This delay is termed latency here (we assume here that there are no external, more rigid latency constraints to take into account). In the interconnection delay studies above (as illustrated in Figure 5.3), we arbitrarily set the value of the latency to 1 bit duration. This turns out to be a good choice: Figure 5.4 shows that the latency has a significant impact on the performance of the reservoir (more than 10dB difference in BER for most fiber lengths). The best latency depends on the fiber length, but is usually close to one bit period.

For all subsequent simulations in this section, we therefore set the reservoir inter-node delay time to half the data rate and the latency to 1 bit duration.

We then investigate the performance of this reservoir design for equalization in links running at different data rates. The results, given in Figure 5.5, show that for link lengths less than $200\,\mathrm{km}$, the PhRC equalizer can operate at data rates higher than the design data rate of $10\,\mathrm{Gb/s}$. For example for the $100\,\mathrm{km}$ link, we get error rates below the FEC limit for data rates close to $20\,\mathrm{Gb/s}$. This result gives a measure of robustness for this particular PhRC equalizer. When designing equalizers for links with even higher bit rates, the analysis would need to be repeated to find the best operating parameters (most notably the interconnection delay) for the equalizer.

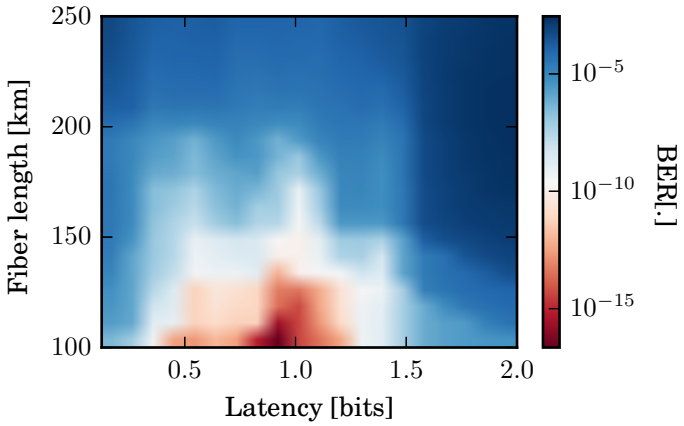


Figure 5.4: Error rate vs latency and fiber length for a reservoir with interconnection delay time equal to half the bit duration.

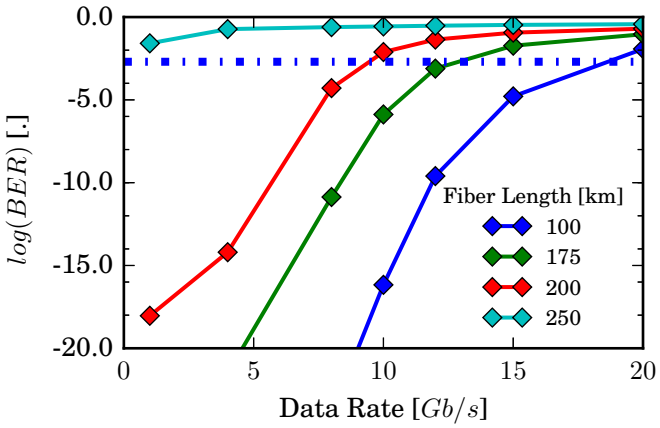


Figure 5.5: Error rate vs data rate for a reservoir with interconnection delay time equal to half the bit duration and latency 1 bit. A Soft Decision Forward Error Correction limit (SD-FEC limit) of 0.2×10^{-2} is also shown. Error free operation is possible for all error rate values below this limit.

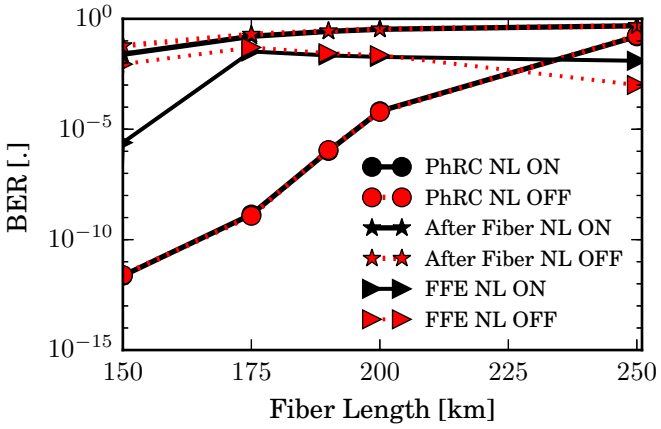


Figure 5.6: BER of the PhRC equalizer as compared to that of an FIR Feed Forward Equalizer (FFE) trained on the same amount of data for different fiber lengths. The launch power is set to $15\,\mathrm{mW}$. *NL ON- nonlinear propagation. NL OFF - nonlinear propagation is deactivated (Nonlinear Index of the fiber in Table 5.1 is set to 0).*

Finally, we compare the performance of the PhRC equalizer for a link with the launch power changed to $15\,\mathrm{mW}$ (compared to the $5\,\mathrm{mW}$ of the previous simulations) to an FIR FFE filter trained on the same amount of data. $15\,\mathrm{mW}$ is significantly higher power for these kinds of links. Launch powers are typically kept below $10\,\mathrm{mW}$ to avoid distortions. An adaptive FFE filter with 31 taps is used (The filter goes over the training data four times to allow for convergence). The results are shown in Figure 5.6. The PhRC equalizer outperforms the FFE equalizer with BERs over 5 orders of magnitude lower at, for example, $150\,\mathrm{km}$ and an order or two of magnitude lower at $200\,\mathrm{km}$. The difference in performance originates in the fact that the PhRC equalizer is a nonlinear compensation device; it takes advantage of the nonlinear transformation in the reservoir to better model the distortion and outstrip the performance of the FFE filter. In Figure 5.6 we also plot the cases with and without the fiber nonlinearity. We observe that at the distances under consideration, the fiber nonlinearities are not yet deleterious. We do however observe that the reservoir is able to make use of its nonlinear nature to outperform the FFE equalizer for these links.

5.2.2 High-speed short-reach links

We now investigate the applicability of the PhRC equalizer to high-speed short-reach IM/DD applications. EAMs are particularly suitable for high-speed short-reach communication links as they are low-cost and compact compared to external $LiNbO_3$ modulators, can be designed to have relatively low chirp (for example

Parameter	Value	Units
	Laser	
Output Power	10	mW
Line width	20	MHz
Emmision Wavelength	1550	nm
Driver Voltage	2.6	V
Driver Bias	-3	V
	EAM	
Transmission characteristics (1550 nm)	see Figure 5.8	-

Table 5.3: Transmitter parameters for high-speed short-reach link setup.

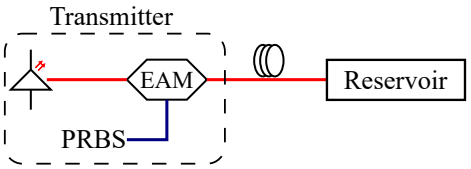


Figure 5.7: Schematic representation of the setup used for short reach simulations. The Mach-Zehnder Modulator of Figure 5.2 is replaced by an Electro-Absorptive Modulator. The EDFA at the end of the link has also been removed.

when compared to directly modulated DFB lasers) and can be driven with lower driving voltage amplitudes. They can additionally be monolithically integrated on the same chip as a DFB continuous wave (CW) laser to yield a compact transmitter module - so called electroabsorption modulator-integrated distributed-feedback (EADFB) lasers. EADFB lasers with speeds over $50\,\mathrm{Gb/s}$ were demonstrated in [11] and are competitive candidates for the next beyond $400\,\mathrm{Gb/s}$ optical communication links (multilane).

We set up a $40\,\mathrm{Gb/s}$ link using a CW DFB laser and an EAM with parameters given in Table 5.3. The input PRBS signal is then propagated for distances ranging from $1\,\mathrm{km}$ to $25\,\mathrm{km}$. Just as in the previous section, we perform a search for the optimum reservoir interconnection length and find that it lies somewhere between 20% and 50% of the bit duration. We assume a latency of 1 bit duration for the readout training phase as before.

The results for the BER obtained with the PhRC equalizer for the different fiber lengths are given in Figure 5.9 and show that we can triple the reach of links from 5km to 15km by using the PhRC equalizer compared to the case of no equalization.

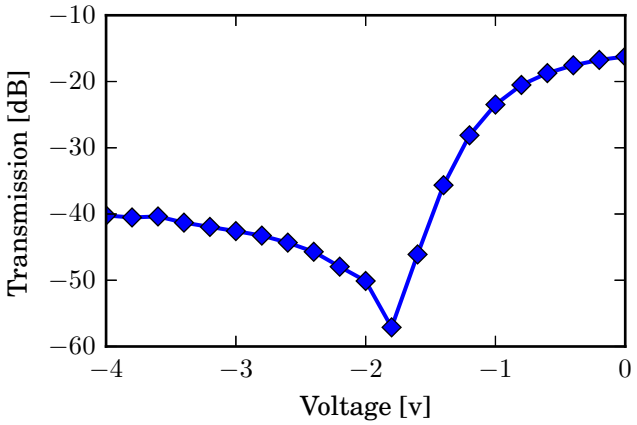


Figure 5.8: Voltage-dependent transmission of the EAM used for the high speed short reach link at 1550nm.

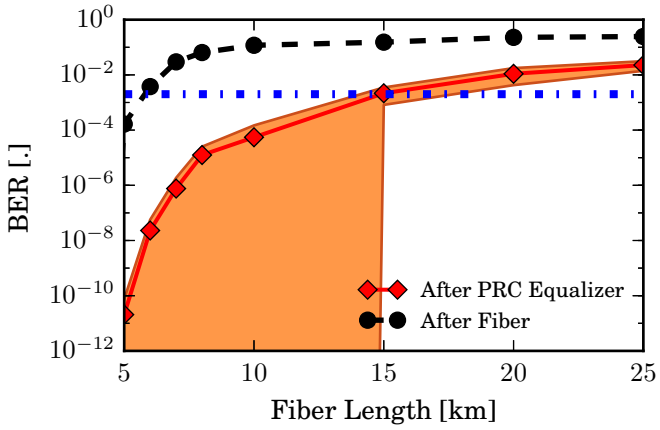


Figure 5.9: BER vs fiber length for $40\,\mathrm{Gb/s}$ *short reach access link with a DFB and an EAM. The blue dashed line indicates the HD-FEC limit The filled section demarcates the error margin.*

5.2.3 Long-Haul Communications link

We now briefly consider how the PhRC equalizer design process could be adapted to dispersion-unmanaged long-haul links. The set up is similar to Figure 5.2 but with the fiber replaced by a recirculating loop to reach large propagation distances. The span length is $100\,\mathrm{km}$ as in Figure 5.10. In each span the fiber is followed by an EDFA (noise figure N_f = 4) to undo losses, and finally an OBPF (bandwidth $80\,\mathrm{GHz}$) is added to the end of the link to get rid of out-of-band ASE.

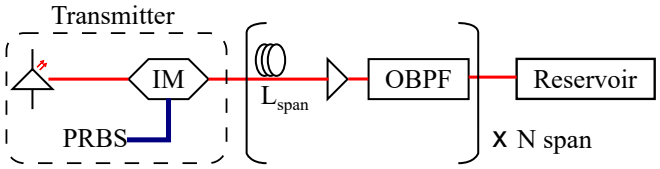


Figure 5.10: Schematic representation of the setup used for the long haul simulations. The setup is a modification of Figure 5.2 with a recirculating fiber loop used to reach large propagation distances. Each span was fixed to a length L_{span} of 100km and the amplifier gain was set to undo the attenuation in each span.

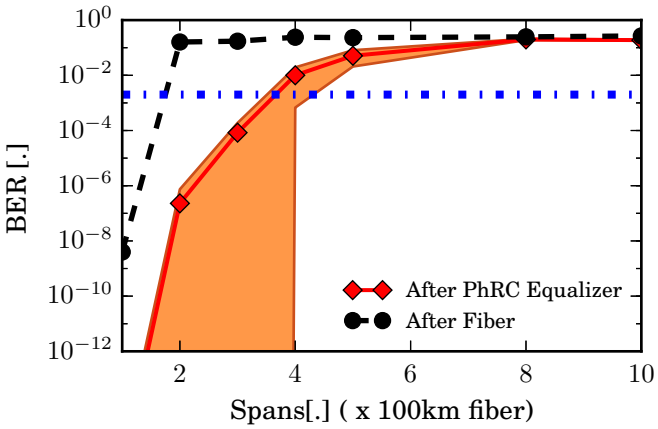


Figure 5.11: BER vs number of spans for a long-haul communications link. Blue dashed line indicates the HD-FEC limit.

If we plot the results of the BER against the number of spans (see Figure 5.11), we see that even for this preliminary study, we can achieve below FEC error rates for distances up to about $400\,\mathrm{km}$. However, to completely characterize the performance (and the corresponding limits) of the PhRC equalizer for these kinds of links, it will necessary to do a further systematic study involving various parameter settings such as span lengths, EDFA noise figures and placement, optical filters.

This is beyond the scope of this thesis.

5.3 Conclusion

In this chapter we have presented a design for an integrated photonics reservoir equalizer that can undo imperfections in optical communication links. We have numerically studied the performance of this PhRC equalizer for three different types of IM/DD links: metro, high-speed short-reach and long-haul. We can reach below the SD-FEC limit for metro links up to $200\,\mathrm{km}$ and we can extend the reach for a high-speed short-reach link running at $40\,\mathrm{Gb/s}$ with an EAM from $5\,\mathrm{km}$ to $15\,\mathrm{km}$. Preliminary simulations on long-haul links show below FEC error rates up to 400 km. We expect these tasks will serve as important experimental benchmarks for our integrated photonics reservoirs targeting optical communications tasks.

References

[1] Ivan Djordjevic, William Ryan, and Bane Vasic. *Coding for Optical Channels*. Springer US, Boston, MA, 2010.

[2] Ezra Ip and Joseph M. Kahn. *Compensation of dispersion and nonlinear impairments using digital backpropagation*. Journal of Lightwave Technology, 26(20):3416–3425, 2008.

[3] Zhou Xian, Huo Jiahao, Zhong Kangping, Wang Liang, Yuan Jinhui, Haiquan Cheng, Keping Long, Alan Pak Tao Lau, and Chao Lu. *PDM PAM-4 with IM-DD using a simple MIMO DSP-based receiver for short reach communications*. In Asia Communications and Photonics Conference 2015, page AM3E.3. Optical Society of America, 2015.

[4] X. Zhou, K. Zhong, J. Huo, J. Yuan, F. Li, L. Wang, K. Long, A. P. T. Lau, and C. Lu. *Polarization-Multiplexed DMT With IM-DD Using 2 x 2 MIMO Processing Based on SOP Estimation and MPBI Elimination*. IEEE Photonics Journal, 7(6):1–12, Dec 2015.

[5] Kristof Vandoorne, Pauline Mechet, Thomas Van Vaerenbergh, Martin Fiers, Geert Morthier, David Verstraeten, Benjamin Schrauwen, Joni Dambre, and Peter Bienstman. *Experimental demonstration of reservoir computing on a silicon photonics chip.* Nature communications, 5:3541, 1 2014.

[6] Joni Dambre Andrew Katumba, Xin Yin and Peter Bienstman. *A Neuromorphic Silicon Photonics Nonlinear Equalizer for Optical Communications with Intensity Modulation and Direct Detection*. Unpublished, Submitted for review to the IEEE/OSA Journal of Lightwave Technology.

[7] Martin Fiers, Thomas Van Vaerenbergh, Ken Caluwaerts, Dries Vande Ginste, Benjamin Schrauwen, Joni Dambre, and Peter Bienstman. *Time-domain and frequency-domain modeling of nonlinear optical components at the circuit-level using a node-based approach*. J. Opt. Soc. Am. B, 29(5):896–900, 5 2012.

[8] F Pedregosa, G Varoquaux, A Gramfort, V Michel, B Thirion, O Grisel, M Blondel, P Prettenhofer, R Weiss, V Dubourg, J Vanderplas, A Passos, D Cournapeau, M Brucher, M Perrot, and E Duchesnay. *Scikit-learn: Machine Learning in {P}ython*. Journal of Machine Learning Research, 12:2825–2830, 2011.

[9] Govind P. Agrawal. *Fiber-Optic Communication Systems*. John Wiley & Sons, 2012.

[10] Tobias A. Eriksson, Henning Bulow, and Andreas Leven. *Applying Neural Networks in Optical Communication Systems: Possible Pitfalls*. IEEE Photonics Technology Letters, 29(23):2091–2094, 2017.

[11] T. Fujisawa, S. Kanazawa, K. Takahata, W. Kobayashi, T. Tadokoro, R. Iga, and H. Ishii. *Ultra-high-speed electroabsorption modulator integrated lasers for 100GbE and beyond*. In Technical Digest - 2012 17th Opto-Electronics and Communications Conference, OECC 2012, pages 75–76, 2012.

6

Towards a photonic reservoir computing hardware implementation with integrated readout

This dissertation set out to demonstrate that integrated photonic reservoirs could be designed and exploited in a way that conforms to the energy efficiency and performance requirements of their target applications domains, especially telecom. The previous chapters have presented various studies to demonstrate both architectural and task-specific advances towards achieving this goal. This chapter builds on these advances and reports progress towards setting up a hardware platform, in the form of silicon photonics chips, to be used to experimentally demonstrate these concepts and provide a tool for further experimentation in the field of integrated photonic reservoir computing.

The activities discussed here are broadly encapsulated under the umbrella of the European Horizon 2020 project PHRESCO that aims to build the next generation of large scalable integrated photonic reservoirs that can tackle more complex and industrially relevant applications. The project aims to build an integrated photonic RC demonstrator chip featuring co-integrated photonic and electronic components, with a larger number of nodes than the state-of-the-art (see [1]). The system leverages the electro-optical properties of new materials and uses novel learning algorithms to exploit the computational capabilities of RC for high-speed telecom applications.

The expected outcome is a state-of-the-art all-optical cognitive computing plat-

form capable of processing large amounts of data with very low power requirements. Beyond scaling in terms of speed and task complexity, real-time operation and energy efficiency are also of paramount importance.

The project brings together Gent University (Belgium), IBM Zurich (Switzerland), the Katholieke Universiteit Leuven (Belgium), Innovations for High Performance Microelectronics (IHP) (Germany) and Centrale Supélec (France), with the goal of leveraging their various extensive experience in modelling and building state-of-the-art photonic devices and systems for the design of a functionally scaled, photonic RC platform. The project intends to use a larger (with more nodes) version of the passive integrated reservoir of [1] at its core and add various extensions by integrating optically active components such as nonlinear elements and weighting elements, amplifiers, and detectors to yield a more telecom compatible RC implementation.

While the activities of the project are very broad, here we focus on some of the work related to the first generation of chips intended to provide early proof-of-concept feedback for a number of critical conceptual and technical challenges regarding the reservoir design and integration of materials for implementing an optical readout. Issues such as inclusion of nonlinear elements and optical amplifiers as well as full system characterization are still a work-in-progress and will therefore not be discussed in detail in this dissertation.

The rest of this chapter is organized as follows: in section 6.1 we discuss the different optical weighting elements considered for the integrated readout for the photonic reservoir. In section 6.2 we then present the design and preliminary characterization of an 18-node silicon photonics reservoir chip with integrated readouts. Section 6.3 briefly introduces a new reservoir architecture that improves signal retention in photonic reservoirs. Finally, in section 6.5 we give concluding remarks for the chapter.

6.1 Optical weighting elements

As introduced in section 2.5.5, operating an integrated photonic reservoir computing system with reduced observables and in a real-time manner requires some key changes to the way the readout of the reservoir is designed. Reducing the number of required high-speed ADCs could take the form of somehow combining the reservoir states. In this work the approach taken is to coherently combine the states in the optical domain. Since the states are coherently combined, this necessitates that the readout weights are also implemented optically and applied before the combination of the states to avoid catastrophic loss of information, both in terms of phase and amplitude. This necessarily means that the weights need to be integrated on the same chip as the reservoir (see Figure 2.7). Physically, the weights need to implement both phase and amplitude modulation.

Optically, the weights could be implemented using silicon photonic components, for example MZM modulators, ring resonators, thermal phase shifters, etc. A more interesting approach, however, is to investigate the possibility of using novel advanced photonic devices that can implement these weights in the most energy efficient way possible. In the PHRESCO project, two material platforms are considered: Vanadium Oxide (VO_2) which exhibits a Metal-Insulator Transition (MIT) and Barium Titanate (BTO), a ferroelectric material. These materials are considered as they present the possibility of implementing the optical weights in a non-volatile manner. This presents an opportunity for improving the energy efficiency of the RC system, as energy would only be expended to set or change the weights as opposed to other implementations that would require constant bias to maintain the weights.

The PHRESCO project therefore focuses on studying the implementation of optical weighting elements based on VO_2 (in collaboration with mainly KU Leuven) and BTO (in collaboration with mainly IBM). For the chip design, silicon-based optical weights are also considered as a baseline. The first demonstrator chip for the project, discussed in section 6.2, includes provisions for all three weighting element technologies.

6.1.1 Vanadium Oxide weights

Vanadium Oxide has a metal-insulator transition that could be used to create non-volatile optical weighting elements. We rely on the fact that the optical properties (refractive index and absorption coefficient) of VO_2 change across this Insulator-to-Metal transition. This means that, when combined with photonic devices such as straight waveguides or resonant structures, we can influence the phase and amplitude of the light propagating through the device. The switching of Vanadium Oxide could be done electrically [2, 3], optically [4] or thermally [5]. In the project, the focus is on triggering the MIT with voltage pulses. However, the process of achieving reliable and controllable electrical switching in VO_2 thin films at the level required for photonic device applications is challenging [6]. The first step towards VO_2-based optical weights is therefore to show MIT switching in VO_2 and to study the transmission behaviour of VO_2-deposited waveguides and race-track resonators. These can then be used as weighting elements in the readout for a PhRC system. An additional goal, for this part of the work, was to investigate if it is possible to lower the temperature at which the MIT transition occurs. Results have shown that the transition occcurs at approximately $70°$C. A more desirable outcome would be to switch at room temperature. A hybrid silicon photonics VO_2 chip designed with the intention of exploring these goals will be discussed in the next section.

6.1.1.1 Design and characterization of a silicon photonics chip for VO_2 optical weighting elements

To serve as a test-platform for optimizing the processing steps for fabricating VO_2 thin films for optical weighting, I designed a chip including various test structures to be used in the characterization and evaluation process. This was based on requirements set by the partners involved in the fabrication of the device. The end-goal was to establish the device parameters that result in the most efficient electro-optical switching behaviour.

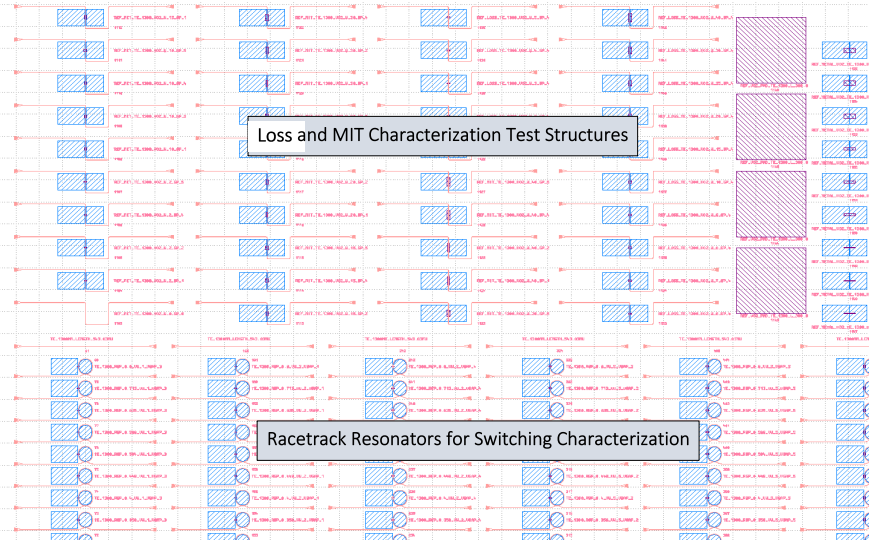


Figure 6.1: A section of the VO2 test chip built for the development of VO_2 based non-volatile optical elements. The layout fit in an area of $4 \times 4\ \mathrm{mm}^2$. Enlarged versions of some of the key components in the design are given in Figure 6.2 and Figure 6.3

.

The specific goals for the design were :

(i) to develop the deposition process for the VO_2 with focus on obtaining strong MIT characteristics.

(ii) to develop the steps for fabrication of photonic devices based on VO_2 (lift-off, etching process, etc.)

(iii) to characterize the electrical and optical properties of VO_2 films focusing on the change of real and imaginary parts of the refractive indices which are pertinent to developing optical weghting elements.

The desired outcome was to generate recommendations for device designs that would then be integrated into a hybrid silicon-VO_2 with these devices serving as weighting elements for the photonic reservoir.

Four different designs were considered: TE and TM polarizations at 1300 nm and 1550 nm. The structures on each of the chips can be grouped into three classes: those to characterize the loss of the Vanadium Oxide, those to characterize the Metal-Insulator Transition of Vanadium Oxide, and finally those to study the switching characteristics of racetrack resonators on which Vanadium Oxide was deposited. A part of the design for the VO_2 test structures chip is given in Figure 6.1.

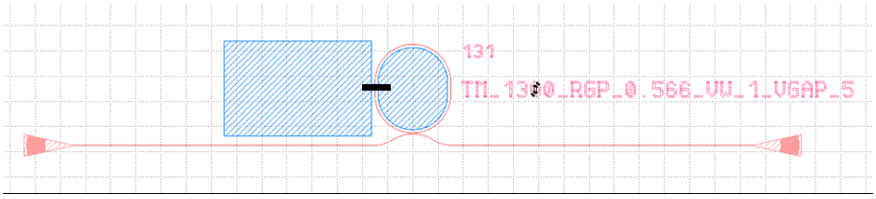


Figure 6.2: Example of a race-track resonator device with VO_2 and metal pads. Red areas correspond to the silicon photonic structures while the blue ones correspond to the metallic contacts for the VO2 stripes (marked in black).

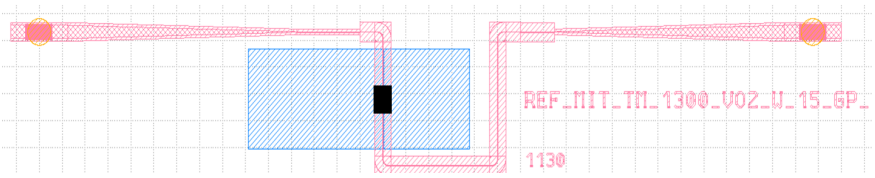


Figure 6.3: An example of a straight waveguide structure used to characterize the loss and characteristics of the metal-insulator transition. Red areas correspond to the silicon photonic structures while the blue ones correspond to the metallic contacts for the VO_2 stripes (marked in black).

The chip layout was done with Luceda Photonics' IPKISS design platform and fabricated using e-Beam lithography at IBM, Zurich. After fabrication at IBM, the chips were sent to Leuven for the VO_2 deposition and then back to IBM for metal pad deposition and static characterization.

For the characterization procedure, first the passive properties of the structures were investigated, in particular the propagation losses in the insulating phase. This was done by analyzing the transmission of the hybrid VO_2-Si waveguides with varying widths of the VO_2 strips. The propagation losses of the hybrid waveguides were determined at IBM to be $10^4\ \mathrm{dB\,cm^{-1}}$. The value of the imaginary part of VO_2 was also extracted at Leuven and found to be $n'' = 0.17$ and in good

agreement with values reported in literature. This was a good indication for the success of the VO_2 deposition process.

For the weighting elements targeted in the PHRESCO reservoir system, we intend to use resonant structures covered with VO_2 patches, as shown in Figure 6.2. In order to obtain a large contrast between the high and low transmission states, operation of the devices at a resonant wavelength with large extinction ratio is desired. Due to the initially unknown material properties of VO_2, it was necessary to experimentally determine well-performing ring resonators by varying the coupling parameter between the bus waveguide and the ring resonator. By analyzing all the race-track devices on the chip, it was found that with the 5 μm wide VO_2 patches, critical coupling with an extinction ratio of about 18 dB is reached for a waveguide-to-ring gap of 160 nm (see Figure 6.4). This value was selected to serve as the design point for future iterations of VO_2 - Si weighting devices.

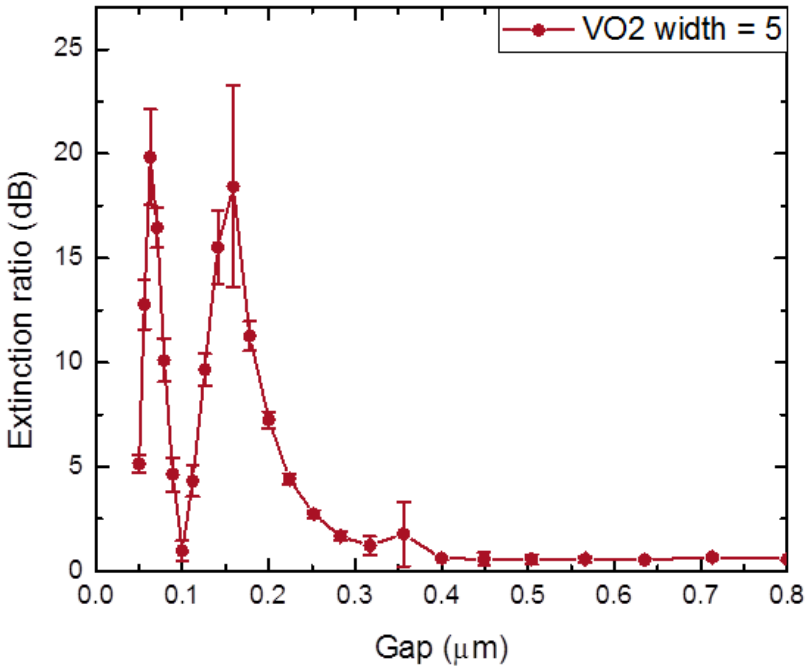


Figure 6.4: Extinction ratio for various gaps between the bus waveguide and the ring resonator. Racetrack resonators are deposited with VO_2 strips with width 5 µm.

Subsequently, electrical characterization of the VO_2 films on electrical structures deposited on the same chip revealed a metal-insulator transition occuring at input voltage of 3V. This result confirmed the presence of the metal-insulator

transition that is required before proceeding with electro-optical measurements.

Next was the electro-optic analysis of the VO_2-Si devices. However, the results revealed limited contrast in optical transmission when changing the applied devices, meaning that it was not possible to demonstrate stable non-volatile VO_2 transmission states. Demonstrating non-volatile electro-optic properties of VO_2 thin films is an on-going endeavor within the PHRESCO project by the Functional Nanosystems group of prof Jean-Pierre Locquet at KU Leuven.

Meanwhile, a new chip has been designed and fabricated that builds on the outcomes of the first chip for improvements, especially in the deposition process of the VO_2. This chip is expected to yield better results than the first iteration.

6.1.2 BTO weights

For these weights, the intention is to take advantage of Barium-Titanate-based photonic devices, developed at IBM, to build optical weights for the integrated photonics reservoir readout. The development of optical weights in BTO is more further along than its VO_2 counterpart. IBM has demonstrated BTO weights with over 10 resolvable non-volatile transmission states that could be used to implement the amplitude modulation [7] (see Figure 6.5). Moreover, work is still ongoing to pack more levels into each device (where more levels implies improvement in the resolution of the weights); the target is to reach 32 levels or more. The devices themselves are either ring resonators or MZMs (for phase modulation), built on hybrid Si-BTO waveguides using the process in [8].

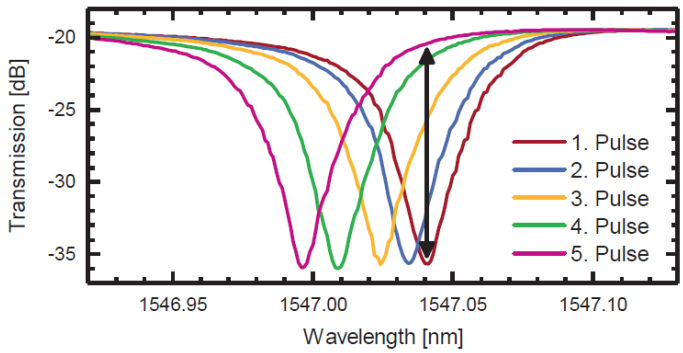


Figure 6.5: Demonstration of non-volatile tuning (change in transmission spectrum) of a hybrid Si-BTO race-track device using voltage pulses (100 ns *and between 10 and 20* V*). The black line indicates the range over which the transmission state can be changed (non-volatile memory window) when operating at a fixed wavelength. With this window, 10 unique levels of transmission can be resolved. Details in and Figure from [7]*

The optical weighting of BTO is due to the possibility to switch fractions of its

ferroelectric domains by injecting electrical pulses. In so doing, it is possible to reach more than an order of magnitude difference in the optical transmission in devices when in different domain configurations. More importantly, the BTO optical weighting elements can be non-volatile such that once set (typically iteratively), the BTO can be held in a state where it does not draw any current [7]. This has made it the preferred option for implementation of the weights in the PHRESCO project, as it means a more power-efficient implementation of the readout that does not consume power beyond what is used to set the weights.

6.2 Photonic reservoir computing chip with multiple input injection and integrated readout

A demonstrator chip based on the photonic swirl architecture with components optimized for the IHP SOI CMOS process together with various integrated readouts was then designed. For this first chip, 18 nodes were considered for the reservoir to keep its footprint small and allow for exploration of different readout technologies. The key purpose was the verification of various passive SOI devices and the development of processes to combine the optical weighting elements described in section 6.1 above with the SOI elements on the chip. The outcomes of these design and characterization activities would then guide the design of the final demonstrator chip featuring a much larger reservoir and a high-performance on-chip readout.

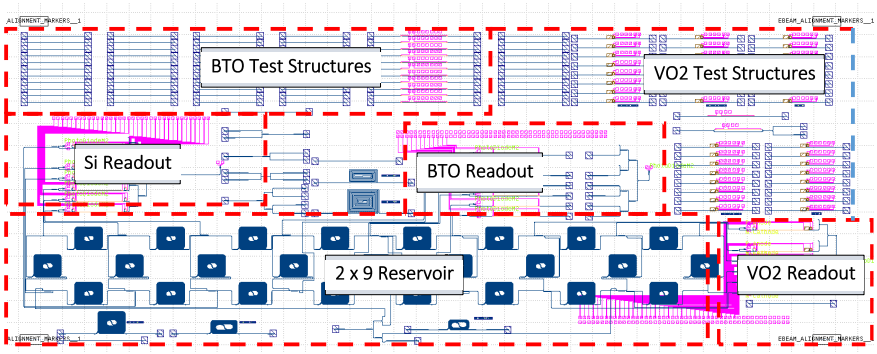


Figure 6.6: Mask plan for the silicon photonics chip consisting of a 2×9 *passive swirl photonic reservoir, and three readouts : a silicon readout, a* VO_2 *readout and a BTO readout as demarcated in the figure. The rest of the chip has different test structures to test passives components used in the reservoir and components to develop the BTO and* VO_2 *optical weighting elements.*

The chip was designed to follow the architecture described in section 2.5.5 and the mask floorplan is given in Figure 6.6. The reservoir has 3 subsections, illustrated in Figure 6.7, such that a subset of 6 nodes was connected to each of the

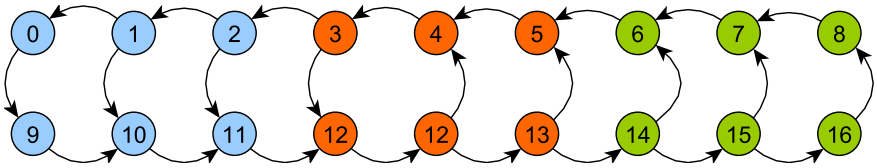


Figure 6.7: The three subsections of the reservoir each connected to one of the three different readouts: blue - Si readout, red - BTO readout and green - VO_2 readout

3 readout technologies: a silicon readout, a VO_2 readout and a BTO readout. The reservoir uses the multiple-input strategy described in chapter 2, where the input power was split using a 2-stage combiner tree and injected into nodes 9, 16 and 4. This input strategy was chosen as it yielded the best BER in simulations on the 3-bit header recognition task for all the subreservoirs.

The design targeted the IHP full-platform BiCMOS Photonic Integrated Circuit technology. This platform is particularly handy for including high-speed photodetectors and doped-silicon heaters (used to create thermo-optic phase-shifters) on the same chip as the photonics components. The chip was designed for the O-band to make the results compatible with the follow-up demonstrator that will potentially make use of SOAs designed in the O-band, at IBM, to provide extra non-linearity at the readout and/or linearly amplify reservoir states before the readout. In terms of design primitives, IHP provides a decent Photonic Design Kit (PDK) featuring low-loss fiber-chip couplers, splitters/combiners, photodetectors, etc., in the O-band.

Next, we will describe the details of the three different readouts as referenced by Figure 6.6.

6.2.1 Design of the integrated readouts

A single VO_2 readout unit consists of three parts (see Figure 6.8 (a)). The first is an SOI ring resonator that acts as the nonlinear component. This is followed by a phase shifter waveguide and a VO_2 deposited racetrack resonator that together form the weighting component. The VO_2 racetrack resonator is designed with some guidance of the results obtained from the test structures chip outline in the section 6.1.1 above. The phase shifter between the two rings is necessary to add an additional degree of freedom for adjusting the readout weights since in the VO_2 racetrack resonator, one cannot independently adjust the amplitude and phase of the input light. The full 6 units of the VO_2 readout can be found in Figure 6.8 (b).

The silicon readout is a pure SOI readout that is included for benchmark and fallback purposes. The nonlinearity is provided by a silicon ring resonator, while the weighting is done with the combination of a phase shifter and a ring resonator. The actual design of the silicon readout is the same as that of the VO_2 in Figure

6.8, minus the deposition of the VO_2.

Just like the other two readouts, the BTO readout is also assigned 6 out of 18 available nodes. In this case, the light from the reservoir node first propagates through a phase shifter to be used for extra tuning, and is then coupled upwards to the BTO layer (to be bonded at IBM) using grating couplers, where it will be channeled through the BTO devices. Specifically, the light first goes through a BTO racetrack resonator acting as a nonlinear element and then goes through a BTO Mach-Zehnder modulator device which will provide the amplitude and phase modulation to implement the weights. The full BTO readout is depicted in Figure 6.9.

Apart from the reservoir and the three different readouts, a significant amount of space on the chip was allocated for test structures. BTO test structures are to be used to determine the optimal placement for grating couplers to coupler light between the silicon photonics and BTO devices, and to ascertain the performance of the BTO racetrack resonator and modulator devices. VO_2 test structures are to be used to determine the geometry and deposition parameters to use for the final readout, as well as to determine the best placement strategy for the metal pads that will be used to drive the device.

6.2.2 Characterization

Once the chips were received from IHP, we carried out a number of transmission measurements followed by high-speed characterization of the components and the integrated reservoir. From the transmission characterization, done with the test structures included on the prototype, grating coupler losses were determined to be $\approx 5\,\mathrm{dB}$ and the spiral waveguide loss to be $2.5\,\mathrm{dB\,cm^{-1}}$. Other components, such as 1×2, 2×2 MMIs, phase shifters, photodetectors (see Figure 6.10), etc., were also characterized and found to be within acceptable ranges for bandwidth and loss.

After the component characterization, an attempt was made at an end-to-end measurement of the reservoir states using the split-off states to the high-speed photodetectors before the nonlinearity. First, simple transmission sanity checks ensured that the signal was propagating through the reservoir as expected. The setup depicted in Figure 6.11 was then used for the rest of the measurements to record the reservoir states at different nodes. For these, a pseudo-random bit sequence was generated by the pattern generator and used to modulate the CW Laser output. The modulated signal is then injected into the reservoir through a fiber-chip grating coupler on a vertical coupling stage.

Specifically, we measured node 16 of the VO_2 readout (one of the nodes with direct input, see Figure 6.6). While we could successfully record the signal (see Figure 6.12), the SNR was too small for further processing. In the end we obtained

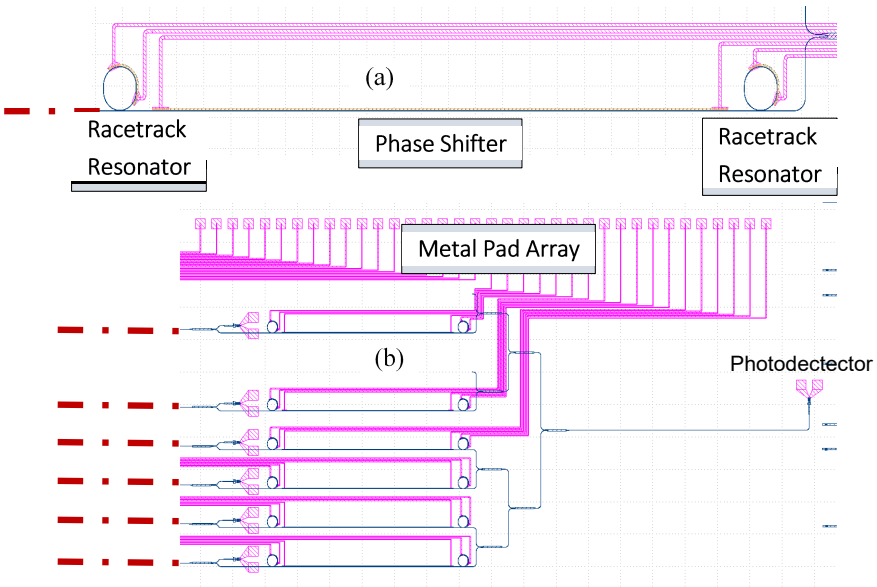


Figure 6.8: Layout of the VO_2 readout. Dashed red lines indicate connections from reservoir nodes. (a) is the zoom-in of the pipeline for a single readout unit which consists of a ring as the nonlinear element and a phase shifter followed by a VO_2 deposited racetrack resonator that together constitute a weighting component. (b) shows the full readout consisting of all 6 readout units. Wires routing to metal pads for tuning the rings and driving the phase shifter are indicated. Note that the implementation of the Si readout is the same but without the deposition of the VO_2.

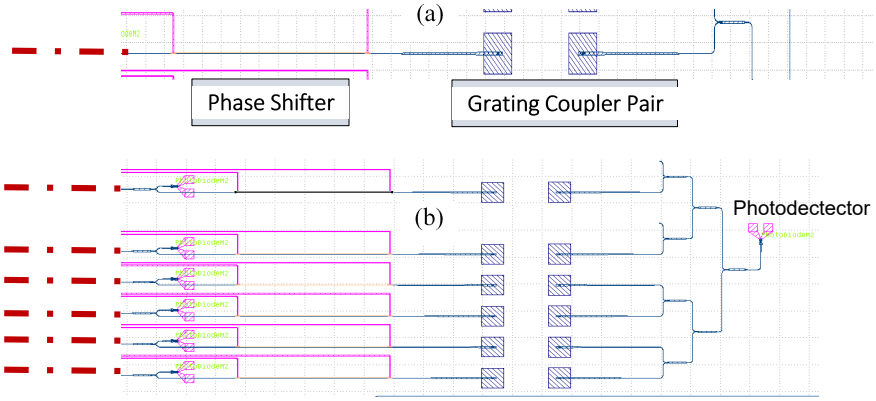


Figure 6.9: Layout of the BTO Readout. Dashed red lines indicate connections from reservoir nodes. (a) represents the pipeline for a single readout unit which consists of a a phase shifter followed by a pair of grating couplers. The grating couplers will be used to couple light off the SOI chip into the BTO layer bonded on top of the chip. The nonlinear element will be provided by BTO racetrack resonator and the weighting element will be a BTO MZI. (b) shows the full readout consisting of all 6 readout units. Wires routing to metal pads for tuning the rings and driving the phase shifter are indicated.

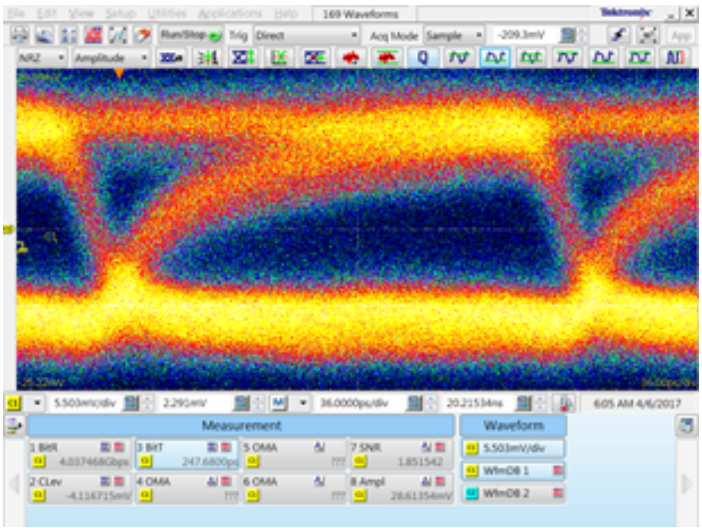


Figure 6.10: Eye diagram of integrated high speed photodetector on demonstrator chip measured at $4\,\mathrm{Gb/s}$*. An open eye (an indicator of error free operation) was obtained for input signals upto* $4\,\mathrm{Gb/s}$*. For the measurements a G-S-G RF probe was used to contact the PD and the resulting signal amplified with an RF amplifier before capture with a real-time oscilloscope.*

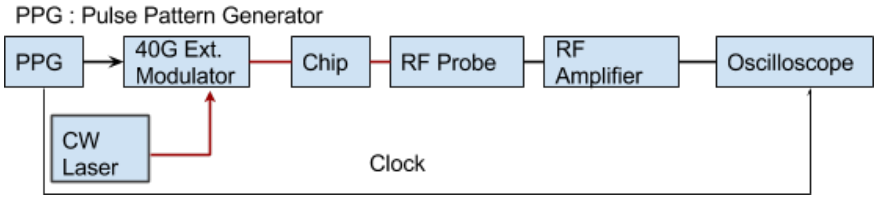


Figure 6.11: Schematic of the setup used for system level characterization of the photonic RC chip. CW - Continuous Wave, RF - Radio Frequency.

similar results for all the nodes with direct input (0, 4, 16) while for the other nodes the SNR level was too low to obtain reliable results. Boosting of the signal using a Semiconductor Optical Amplifier (SOA) could not yield sufficient amplification before saturation. This was mainly because unlike the C-Band that can make of Erbium Doped Fiber Amplifiers (EDFAs) to reach gain values of more than $40\,\mathrm{dB}$, it is not straight forward to get this level of amplification in the O-Band.

I then looked at alternatives for amplification in the O-band and settled on trying to boost the input to the reservoir using Fiber Raman amplification. I visited prof Chigo Okonkwo's Electro-Optical Communication group at the Institute for Photonics Integration, at the Eindhoven University of Technology (TU/e), that has the required equipment and prior experience with similar setups. Over a few weeks, we built a Raman amplification setup similar to that detailed in [9]. However, while we could get more than $10\,\mathrm{dB}$ of Raman gain, it was not sufficient to surmount the loss of the standard single-mode fiber (SSMF) used as the gain medium. And while as an alternative to the SSMF, highly nonlinear fiber (HNLF) was available, it was optimized for C-Band rather than O-Band operation and could therefore not yield the same gain.

Without sufficient input SNR to the reservoir, the characterization could not proceed. However, more recently one of the project partners is in the process of acquiring a Praseodymium-Doped Fiber Amplifier (PDFA) and we plan on finalizing the characterization of this chip in the next few months.

We should mention that while this initial generation of reservoir presented a number of challenges, especially regarding power budget, it has provided crucial feedback to drive the design of the next generation of chips. One example is that in the meantime, the power budget limitations gave us an opportunity to re-think the reservoir architecture with a view of bringing down the internal losses in the reservoir. In the next section we will discuss a new architecture that has been one of the outcomes of this exercise.

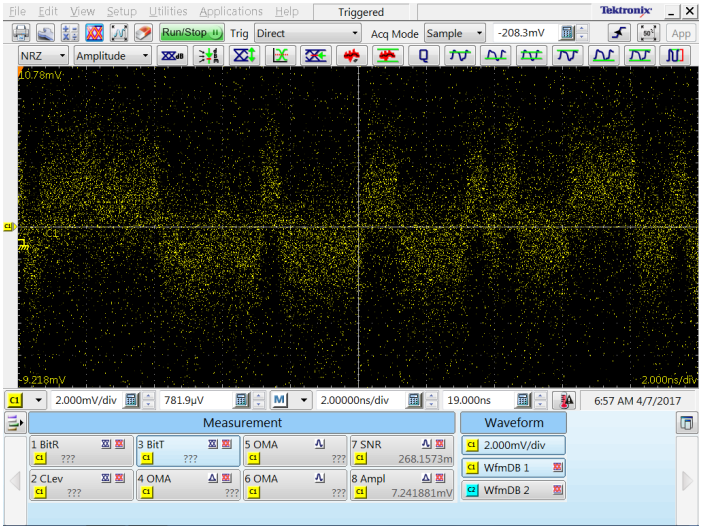


Figure 6.12: A slice of states from the chip measured at 4Gb/s at node 16 in Figure 6.7.

6.3 An improved architecture for passive photonic reservoir computing

In section 6.2, we showed that while the separate components of the fabricated reservoir computing chip worked as expected, the loss due to the reservoir was too high to enable any significant outcome from the measurements without significant boosting of the input signals. This shortcoming has led to a re-visit of the architecture of the reservoir to ensure that less power is lost in the reservoir. Indeed, in the previous swirl architecture significant loss at combination points occurs due to the presence of Y-branch combiners, as was discussed in Chapter 4. Eliminating these 3-port devices and making sure there is always more than 1 output port in each node, closes off this radiation loss channel, which improves energy-efficiency drastically.

Taking this into account, one of the leading new architectures that has been conceived is illustrated in Figure 6.13. This design is conceptually similar to the swirl topology reservoir used for all the work in this thesis, but the removal of 3-port devices results in additional skip connections between nodes that allow for the input signal at a particular node to propagate further in a single hop than in the classical swirl. For example, while in case of the swirl topology the most direct path for light to propagate from node 0 to node 3 would be to go through nodes 1 and 2, in this new architecture it can get there in a single hop directly from node 0 to node 3. This kind of behaviour also improves the efficiency of signal mixing in the reservoir as there are higher chances of the signals combining to be of similar

power levels.

Additionally, the internal implementation of the reservoir combiner/splitter components has been changed from combinations of 1×2 and 2×2 MMIs to 3×3 MMIs. This overall avoids losses due to cascading splitter components and ensures a more homogeneous reservoir architecture which is, again, good for mixing and hence computation.

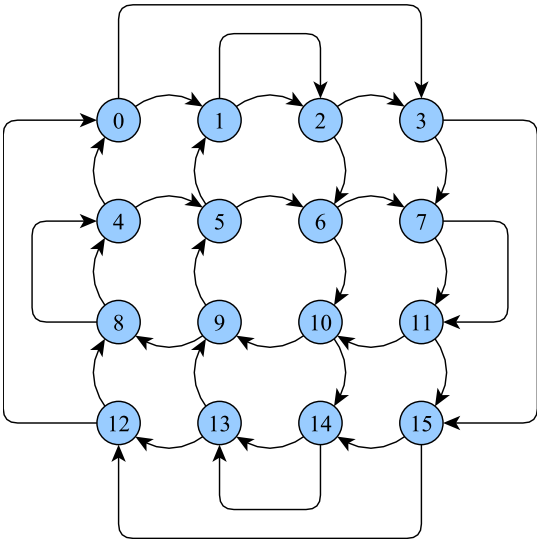


Figure 6.13: Schematic of the new photonic RC architecture. In data flow terms, it is the swirl architecture augmented with loop connections.

This new architecture is used in the reservoir design of the second generation demonstrator that has been designed and taped out to IHP. The 3×3 MMI used for this design was part of a separate chip I designed with the purpose of characterizing a number of passive components. The full characterization of the MMI structure is part of the work of the master thesis of Ashutosh Ashokkumar Vyas, while the detailed discussion and full characterization of this new reservoir architecture is subject of the doctoral thesis work of Stijn Sackesyn.

I should also mention that the new demonstrator chip is composed of more nodes (60) than the initial reservoir chip described in section 6.2 above, but targets a data rate of 32 $\mathrm{Gb/s}$ meaning that it will automatically have much lower internal losses as it will have delay lengths that are much shorter (by almost an order of magnitude).

6.4 Acknowledgements

I would like to thank all the PHRESCO project partners for their contributions and guidance throughout this work. I especially thank Mariela Menghini (Leuven) and Stefan Abel (IBM), who did most of the VO_2 deposition and characterization, and generally provided crucial guidance surrounding the optical weighting technology discussed in this section.

I would also like to thank Prof. Chigo Okonkwo and Prof. Huug de Waardt for hosting me at the Electro-Optical Communications Group at TU Eindhoven.

6.5 Conclusion

Our ultimate goal is to design a passive photonic reservoir that can be used to solve tasks in an online manner as may, for example, be required for telecom applications. Such a chip could integrate non-volatile optical weighting elements to perform the weighting of the reservoir states optically rather than electrically, which has the potential for significant improvements in processing speeds and energy efficiency.

A first generation demonstrator chip combining an 18 node reservoir chip together with silicon structures to be used to explore three integrated readouts was designed and taped out. Initial characterization of this chip showed expected performance for various components. The integration of the optical readout weights is still on-going work.

The high-speed characterization of the reservoir has shown promising results but is to be continued at a later date when the SNR of the input signal can be significantly improved to a level compatible with the electronics used to record and analyze the reservoir states.

The process parameters extracted from, and the lessons learnt during the characterization of the first chip have led to various design improvements, including a new architecture, that have guided the development of a new chip with more reservoir nodes that is expected to tackle complex telecom-related tasks at much higher data rates.

References

[1] Kristof Vandoorne, Pauline Mechet, Thomas Van Vaerenbergh, Martin Fiers, Geert Morthier, David Verstraeten, Benjamin Schrauwen, Joni Dambre, and Peter Bienstman. *Experimental demonstration of reservoir computing on a silicon photonics chip.* Nature communications, 5:3541, 1 2014.

[2] Dmitry Ruzmetov, Gokul Gopalakrishnan, Jiangdong Deng, Venkatesh Narayanamurti, and Shriram Ramanathan. *Electrical triggering of metal-insulator transition in nanoscale vanadium oxide junctions*. Journal of Applied Physics, 106(8):083702, oct 2009.

[3] Byung Gyu Chae, Hyun Tak Kim, Doo Hyeb Youn, and Kwang Yong Kang. *Abrupt metal-insulator transition observed in VO2thin films induced by a switching voltage pulse*. Physica B: Condensed Matter, 369(1-4):76–80, 2005.

[4] Judson D. Ryckman, Kent A. Hallman, Robert E. Marvel, Richard F. Haglund, and Sharon M. Weiss. *Ultra-compact silicon photonic devices reconfigured by an optically induced semiconductor-to-metal transition*. Optics Express, 21(9):10753, 2013.

[5] A. Zimmers, L. Aigouy, M. Mortier, A. Sharoni, Siming Wang, K. G. West, J. G. Ramirez, and Ivan K. Schuller. *Role of thermal heating on the voltage induced insulator-metal transition in VO2*. Physical Review Letters, 110(5):1–5, 2013.

[6] L. D. Sánchez, A. Rosa, T. Angelova, J. Hurtado, A. Griol, P. Sanchis, M. Menghini, P. Homm, B. Van Bilzen, J. P. Locquet, and L. Zimmermann. *Electrical switching in hybrid VO2/Si photonic structures*. International Conference on Transparent Optical Networks, 2016-August:3–6, 2016.

[7] Stefan Abel, David J. Stark, Felix Eltes, Daniele Caimi, Jean Fompeyrine, and John E. Ortmann. *Multi-level optical weights in integrated circuits*. 2017 IEEE International Conference on Rebooting Computing, ICRC 2017 - Proceedings, 2017-January:1–3, 2017.

[8] Stefan Abel, Thilo Stöferle, Chiara Marchiori, Daniele Caimi, Lukas Czornomaz, Michael Stuckelberger, Marilyne Sousa, Bert J. Offrein, and Jean Fompeyrine. *A hybrid barium titanate-silicon photonics platform for ultraefficient electro-optic tuning*. Journal of Lightwave Technology, 34(8):1688–1693, 2016.

[9] Pawel Mazurek, Pawel Czyzak, Huug de Waardt, and Jaroslaw Piotr Turkiewicz. *Semiconductor optical amplifiers and Raman amplification for*

1310-nm dense wavelength division multiplexed transmission. Optical Engineering, 54(11):116104, 2015.

7 Conclusion

7.1 Summary

This dissertation has focused on various energy efficiency considerations for passive integrated reservoirs, if they are to be successfully exploited in the real world for telecommunications applications. We have shown various approaches that can be used to improve the energy efficiency of integrated reservoirs, and highlighted some application scenarios for which these reservoirs could be competitive to existing digital electronic approaches with regards to cost and performance .

In chapter 3 we have shown that by distributing the input signal over a number of input nodes, rather than injecting it all in one node, yields a reservoir that performs better for a broad range of data rates. Additionally, the reservoir with this multiple-input strategy achieves error-free performance, below the HD-FEC limit check that the text does not reference SD-FEC, at orders of magnitude less power than with the single-input strategy. We attribute this improved performance to a more even distribution of power levels in the reservoir leading to an overall increase in the reservoir richness, which in turn means increased computational power. We have also found an optimal strategy for the 16-node reservoir, whereby it is best to inject the input light into the central loop of the reservoir. In this situation, the reservoir requires about 4 orders of magnitude less power than the single-input case for the same total input power to the reservoir.

We then studied another approach to improving the energy efficiency of passive integrated photonic reservoirs in Chapter 4. The success of this approach required

us to re-design the whole reservoir to operate with multiple modes rather than the typical single mode. The success of the approach hinged on demonstrating a novel Y-junction combiner to capture in higher order modes the light that is typically radiated away at the junction. We achieved this using a geometry that takes advantage of a non-adiabatic taper section in the junction to yield a re-distribution of the light into higher-order guided modes whereby after the junction it is converted back into the first-order mode, rather than radiated away as for the single mode Y-junction. When considering performance on the 3-bit header recognition task at $10\,\mathrm{Gb/s}$, the multimode reservoir has error rates similar to those of the single-mode reservoir. However, its real benefits lie in the improvement in the signal-to-noise ratio for error-free performance. This becomes particularly important when building larger reservoirs for which the signal traverses much longer paths before reaching nodes far away from input points.

Then in chapter 5, we discussed the numerical simulations of a scenario where passive integrated reservoir computing system can be used as a viable solution for high-speed distortion compensation in optical communications systems. As a non-linear compensator, the RC-based nonlinear compensator acts as a complex nonlinear filter that can model and undo imperfections introduced in the generation, transmission and reception phases for optical signals traversing optical networks. We took examples of distortion inversion in metro links up to about $200\,\mathrm{km}$ running at $10\,\mathrm{Gb/s}$, and high-speed short-reach links at $10\,\mathrm{Gb/s}$ employing Non-Return-to-zero On-Off Keying (NRZ OOK) modulation with Intensity-Modulation Direct-Detection. We reported improvements in BER over 5 orders of magnitude when compared to FFE-based solutions for metro links, and that we can extend the reach of the short-reach applications by about $10\,\mathrm{km}$.

Finally, in chapter 6, we presented our progress on developing a platform that integrates the readout on the same chip as the reservoir itself (rather than the typical approach of taking the linear combination off-chip electronically). Having an integrated all-optical readout means makes the RC suitable for online learning and real-time application. Together with the integrated readout, we introduced coherent combination as a technique to reduce the number of observables from the reservoir computing system, relaxing the output requirements to a single high-speed electronic output signal rather than the usual one per reservoir node. This improvement directly implies overall cost reduction for deployments. We also discussed progress towards implementing readouts that incorporate novel optical weighting elements based on vanadium dioxide and barium titanate, materials that allow for non-volatile setting of their optical properties that could also leading to additional savings in energy consumption of the systems.

7.2 Perspectives

Passive integrated photonic reservoirs could be leveraged in a way that enables them to tackle various information processing and computation tasks, relevant to modern telecommunication systems at very large bandwidths and in a power efficient way. This dissertation has laid the ground-work for various energy consumption reduction approaches to both building and applying passive RC systems. However, there is still a lot to be done before a complete RC-based solution can be deployed at the scale required for today's applications.

7.2.1 Improvements in reservoir design

In chapter 3, we showed how multiple copies of the input signal applied to a selection of the reservoir nodes can lead to significant reduction in the power required for error-free operation of the reservoir. However, this work targeted a fixed reservoir size and a single task. The next step would be to study the generality of the guidelines generated by this study across different tasks, and possibly across different reservoir architectures. Another extension would be to study how the 16-node reservoir we studied, driven under the optimal strategy, can be used as a tile to create larger reservoirs. This could lead to some insight into alternative approaches to building larger reservoirs, rather than just increasing the number of nodes, which may not always lead to the best case, especially regarding losses in the reservoir.

As an approach to combating internal reservoir losses, we numerically studied the multimode reservoir in chapter 4. To simplify the analysis, we assumed average behaviour across the modes by simply summing up intensities; the next step would be the full coherent exploration of the multimode reservoir. Among other modalities, such a study would require significant modifications to the in-house circuit simulation tools to support multiple modes. Subsequent to this, a chip could be designed and taped-out to experimentally verify the results of this work. Another way that the multimode reservoir could be exploited is to apply it in such a way that the different modes are treated as different processing channels. Here, each supported mode would encode a different signal corresponding to a different task (it should even be possible to simultaneously process analog and digital signals) to be solved. Designed properly, this essentially unleashes another level of parallelism for RC systems effectively increasing their effective processing power.

7.2.2 Further progress in telecommunication applications

As has been a theme for this work, photonic reservoir computing presents numerous advantages that make it a future-proof platform for building information processing systems and computational kernels for optical communications systems. This fact coupled with increased interest and investment in optical approaches to

implementing functionality in optical communications networks, presents a unique opportunity for researchers to take a closer look at what challenges could benefit from this development. In this dissertation, we identified and studied a nonlinear compensation device for 10G and beyond IM-DD NRZ OOK optical communications links. We showed that the nonlinearity of the photonic reservoir equalizer allows it to achieve very large gains in signal quality over standard FIR filter based approaches. The next step would be to systematically study the performance of the reservoir as a compensation device in presence of significant nonlinear distortions in the input signals.

An even more interesting challenge is to study the behaviour of the RC nonlinear compensator in coherent optical communications systems that are the main stay for many cuttingedge deployments of optical communications systems for all sorts of reaches. Coherent optical communications, even more than IM-DD links, rely heavily on usually power-hungry DSPs for all sorts of signal conditioning and impairment removal. This makes this attractive as a target platform for RC nonlinear compensation and any other RC-based applications as any gain over DSP technology would be a welcome development for the telecom sector for which "all-optical" implementations are heavily sought after.

7.2.3 Experimental demonstration for reservoir with integrated readouts

Finally, in chapter 6 we discussed the progress with on-going work on an experimental demonstrator chip that integrates a passive photonic reservoir with a novel readout implementation. We mentioned that by taking advantage of hybrid silicon-VO_2 and silicon-BTO platforms, we have the potential to compose all-optical weights into a non-volatile readout that is compatible with telecom applications. While we reported preliminary results from the characterization of a first-generation chip featuring a passive reservoir and optical weight development structures, the full integration of the weights is yet to be finalized, and so are the full end-to-end measurements of the system. The next generation, improved version of this chip with more nodes and improvements to the first-generation components has been designed and taped-out and will be characterized in the coming months.