Acknowledgments

The journey leading to this thesis has had its ups and downs, and I am deeply grateful to the amazing people who supported me along the way.

Working with Peter was an incredible experience and this thesis owes much to his guidance. Thank you for making research easier by always dedicating time to discuss and advise, making it easy to bring up issues as they arise, and encouraging the pursuit of ambitious ideas. Just as importantly, thanks for always extending kindness and empathy when self-doubt was too high or when life was a bit more challenging. I was equally fortunate to have Joni as a second supervisor. Thank you for your insights, feedback, and for encouraging discussions with the 8th floor community, where thought-provoking conversations widened my perspective.

As supervisors they also fostered a warm and supportive atmosphere within the RC group, with colleagues who made the day-to-day experience very enjoyable. I am especially grateful to Stijn and Emmanuel for their guidance early on and for generously offering their time and expertise throughout. I'm also grateful to have worked with Ruben who brought fresh energy and perspective to our discussions and measurements. And thank you to Alessio, Chonghuai, Gouda, Federico, Ruqi, Robbe, Radwan, Alessandro, and Yuanli, whom I shared lesser time or lesser topic overlap with, but were always supportive and fun to work with.

I'm also grateful to my officemates who always filled our office with homely sweets, friendly chattering, and all sorts of advice. Thank you Shiqi, Xiujun, Cheming, Chao, Ali, Sarah, Suzzane, and Amin. And a special thanks to my immediate neighbors, Laurens, Zhongtao, Stijn and Jens who fielded far more of my questions and conversations than they probably expected but were always fun to chat with and always answered questions with patience and thoroughness.

To the wider PRG community, who shared lunches, coffees, and discussions, thanks for creating a friendly and engaging environment and a special thanks to the 5th floor colleagues. I'm also grateful to all the professors and support staff who made sure everything runs like a well-oiled machine. Naming everyone in such a large, supportive community is tough and I'm sure I'll miss many, but I can't pass up the chance to mention those who played a key role in the measurements and made

these procedures so efficient: Clemens, Ewoud, Hasan, Peter Guns, and IDLab's Joris and Nish — thank you for all your help.

I'm very fortunate to have done this work within the Postdigital project, and I'm grateful for all the opportunities this has afforded me. Thank you to the Postdigital colleagues, supervisors, and everyone who contributed to the valuable discussions and collaborations that came out of this project.

The secondment at Aston was a fantastic experience, and I am grateful to the community of researchers, support staff, and professors who made the office and lab such a welcoming place. A special thanks to Stelios, who always goes above and beyond to support his students. Working alongside you was a valuable learning experience, and I truly appreciate the dedication you put into ensuring my time at Aston was as productive and rewarding as possible. Finally, I am thankful to ISOC and MABY for their efforts in building such a strong and supportive community at Aston and Birmingham. Thank you for the great times and the amazing memories.

To my friends, thank you for listening to my endless updates, celebrating every small victory, and being there through the tough moments. To those who traveled across the world to visit — Batool, Safiya, Shahd, Gina, Mariam, Joud and (all the way from Leuven and Hasselt) Menna and Afnan — thank you for bringing a bit of home to Ghent. Batool, thank you for always being there, no matter what; nothing was ever too small or too big to talk about. And to Afnan, whose home and family have always felt like my own, thank you for the warmth, for sharing the load, and for your unwavering presence.

To my parents, thank you for always believing in me through many years of yet more school. Your love and endless support have been my safety net and the anchor I rely on when I am away. To my family and my support system that transcends all borders, thank you for creating a loving and warm home, and for always being the people I can count on, no matter what. This thesis may have my name, but it is built on your unwavering love, support, and encouragement.

Sarah Masaad
Ghent, January 2025

Contents

List of Acronyms

A

ADC	Analog to Digital Converter
AI	Artificial Intelligence
AIPT	Aston Institute of Photonic Technologies
ASIC	Application Specific Integrated Circuit
AWG	Arbitray Waveform Generator

B

BER	Bit-Error Rate
BPS	Blind Phase Search
BTO	Barium-Titanate

C

CATV	Cable TV
CDC	Chromatic Dispersion Compensation
CMA	Constant Modulus Algorithm
CMA-ES	Covariance Matrix Adaptation Evolution Strategy
CMOS	Complementary Metal-Oxide-Semiconductor
C.R	Carrier Removal
CSPR	Carrier-to-Signal Power Ratio
CW	Continuous Wave

D

DC	Data Center
DCI	Data Center Interconnect
DCN	Data Center Network
DSP	Digital Signal Processing

E

EDFA	Erbium Doped Amplifier
E.R	Extinction Ratio

F

FEC	Forward-Error Correction
FFE	Feed-Forward Equalizer
FFT	Fast-Fourier Transfrom
FIR	Finite Impulse Response
FNN	Feed-forward Neural Network
F.S	Frequency Shift
FTTC	Fiber-to-the-Curb
FTTH	Fiber-to-the-Home
FOCS	Fiber Optic Communication System

G

GS	Gerchberg-Saxton

H

HOE	Higher Order Error

I

IC	Integrated Circuit
IFFT	Inverse Fast Fourier Transform
IL	Insertion Loss
IM/DD	Intensity Modulation and Direct Detection
InP	Indium Phosphide
ISI	Intersymbol Interference

K

KK	Kramers-Kronig

L

LO	Local Oscillator
LR	Linear Regression

M

MEMS	Micro-Electro Mechanical System
MIMO	Multiple-Input Multiple-Output
ML	Machine Learning
MMI	Multi-Mode Interferometer
MPW	Multi-Project Wafer
MRR	Micro-Ring Resonator
MSE	Mean-Square Error
MZI	Mach-Zender Interferometer
MZI	Mach-Zender Modulator

N

NN	Nerual Network

O

OADM	Optical Add-Drop Multiplexer
OSNR	Optical Signal-to-Noise Ratio

P

PAM	Pulse-Amplitude Modulation
PBS	Polarizing Beam Splitter
PD	Photodiode
PIC	Photonic Integrated Circuit
PON	Passive Optical Network

Q

QAM	Quadrature-Amplitude Modulation

R

RC	Reservoir Computing
RDE	Radius Directed Equalization
RNN	Recurrent Neural Network
ROADM	Reconfigurable Add Drop Multiplexer
RRC	Root-Raised Cosine

RTO	Real-Time Oscilliscope

S

S-Matrix	Scattering Matrix
SiN	Silicon Nitride
SiP	Silicon Photonics
SNR	Signal-to-Noise Ratio
SOA	Semiconductor Optical Amplifier
SOI	Silicon on Insulator
SOP	State of Polarization
SPS	Samples per Symbol
SSB	Single Sideband Signal
SSBI	Signal-Signal Beat Interference

T

TIA	Transimpedance Amplifier
TDM	Time Division Multiplexing

V

VOA	Variable Optical Attenuator

W

WDM	Wavelength Division Multiplexing

Samenvatting

1 Inleiding

De uitrol van glasvezelnetwerken concentreert zich steeds meer op toepassingen met kortere afstanden, waarbij individuele systemen slechts een bereik hebben van enkele kilometers, maar waar hun gecombineerde schaal aanzienlijk is. Gezien de grote hoeveelheid van deze implementaties, vereisen netwerken voor korte- en middellange afstanden kosteneffectieve oplossingen, die verschillen van de systemen voor lange afstanden, waar hogere investeringen over langere afstanden en grotere gebruikersgroepen kunnen worden verdeeld.

Coherente transmissie is bijvoorbeeld effectief voor langeafstandsnetwerken, maar brengt hoge kosten voor ontvangers en aanzienlijke verwerkingsvereisten met zich mee, wat de haalbaarheid ervan beperkt voor netwerken met hoge dichtheid. Onderzoek richt zich daarom op het verminderen van de complexiteit van ontvangers door alternatieve architecturen en vereenvoudigde signaalverwerkingsmethoden, om coherente oplossingen praktischer te maken voor netwerken voor korte en middellange afstanden.

Een benadering om ontvangers te vereenvoudigen is het gebruik van zelf-coherente architecturen, waarbij een lokale oscillator meegestuurd wordt met het berichtsignaal. Er worden ook minder fotodiodes gebruikt, gevolgd door verschillende verwerkingstechnieken om complexe signalen uit detectoruitgangen te halen. Voorbeelden hiervan zijn de Kramers-Kronig (KK) ontvanger en de silicium-fotonische signaal-signaal-zwevings-interferentie onderdrukkende ontvanger (SiP SSBI). Deze systemen hebben echter vaak strenge operationele beperkingen die hun praktische inzetbaarheid kunnen beperken.

Pogingen om signaalverwerking te vereenvoudigen zijn gericht op analoge oplossingen, die de noodzaak van digitale verwerking met overbemonsterde signalen omzeilen. Dit wordt ook gedreven door zorgen over de langetermijngeschiktheid van digitale verwerking, aangezien de verwerkingscapaciteit voor baudrate van digitale signaalverwerkingschips (DSP) jaarlijks slechts met 13% groeit, wat ver achterblijft bij de 60% groei in dataverkeersvraag. Tegelijkertijd maakt de huidige matuurheid van geïntegreerde fotonica miniaturisatie en integratie van verschillende fotonische componenten op

een compacte chip mogelijk, waardoor licht efficiënt binnen een minimale ruimte kan worden gemanipuleerd. Deze integreerbaarheid, samen met eigenschappen zoals frequentieparallelisme, lage latentie en lage transmissieverliezen, positioneert optische verwerking als een veelbelovend alternatief voor traditionele elektronica.

Geïnspireerd door eerdere successen waarbij optische subsystemen traditionele elektronica vervangen in een glasvezelcommunicatienetwerk, verkent dit proefschrift optische alternatieven voor het egaliseren van vervormde communicatiesignalen. Dit werk sluit ook aan bij het "post-digitaleönderzoeksthema, dat onconventionele, niet-binaire computersystemen onderzoekt voor taken die economisch of ecologisch onpraktisch zijn om digitaal af te handelen. Dit concept vormt de basis van het PostDigital-project, een Marie Skłodowska-Curie Innovative Training Network, binnen het welke dit proefschrift werd ontwikkeld en waarvan we de term hebben geleend.

Met dit doel vervangt dit proefschrift traditionele digitale berekeningen door analoge, geïntegreerde fotonische verwerking. De verwerkingsarchitectuur is gebaseerd op reservoir computing, een machine learning-paradigma. Net als bij andere machine learning-benaderingen, past de chip zijn parameters aan op basis van blootstelling aan gelabelde datasets om het gewenste gedrag te bereiken. Het gebruik van fotonische verwerking wordt zo benut om bredere inzet van coherente modulatie in korte- en middellangeafstandscommunicatiesystemen mogelijk te maken. Dit doel wordt bereikt door twee belangrijke aspecten aan te pakken: het vereenvoudigen van verwerkingsvereisten en het mogelijk maken van zelf-coherente ontvangerarchitecturen om onder minder strenge omstandigheden te kunnen functioneren.

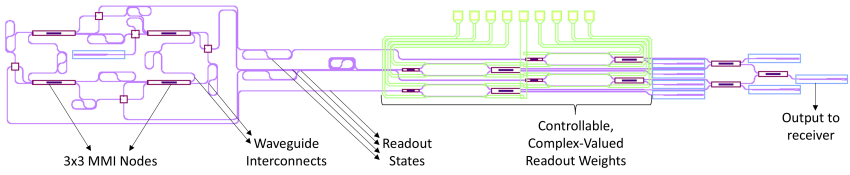
2 Fotonische Reservoircomputing

Het trainen van machine learning-netwerken kan uitdagend zijn en vereist vaak iteratieve optimalisatiemethoden die aanzienlijke computationele middelen en grote datasets gebruiken om te convergeren. Om het trainingsproces te vereenvoudigen, biedt Reservoir Computing (RC) een alternatieve benadering door de interne netwerkverbindingen (het reservoir) vast te zetten en alleen de uitgangslaag, die bekend staat als de readout, te trainen. Het reservoir fungeert als een interne projectie van de invoer(en), waarbij een voldoende rijke verscheidenheid aan kenmerken ('features') wordt gegenereerd waarop de readout-laag kan worden getraind. Ondanks de schijnbare eenvoud heeft deze aanpak succes gehad in verschillende toepassingen, waaronder tijdreeksvoorspellingen en classificatie/regressietaken.

RC is ook naar voren gekomen als een veelbelovende aanpak voor machine learning in fysieke computing. Dit wordt gedreven door zijn vermogen om efficiënt netwerken te construeren in omgevingen waarin het manipuleren

van interne gewichten uitdagend is. Bovendien heeft de capaciteit van het reservoir om diverse fysieke fenomenen in hun natuurlijke vorm te benutten, de toepassingsmogelijkheden vergroot. Als gevolg hiervan is RC onderzocht in een breed scala aan substraten, waaronder elektronica, spintronica, biologische systemen en fotonica.

Het fotonische reservoir dat in deze thesis wordt gebruikt, wordt de vierpoortarchitectuur genoemd. Het reservoir maakt gebruik van 3x3 MMI's als knooppunten die onderling verbonden zijn door spiraalgolfgeleiders, zoals weergegeven in Figuur 1. Bij elk knooppunt interfereren invoersignalen coherent om verschillende uitvoersignalen te produceren. De architectuur van dit reservoir is passief, omdat er geen externe energie nodig is om de knooppunten of verbindingen aan te drijven. Bovendien kunnen niet-vluchtige optische gewichten ook op de chip worden geïntegreerd voor de readout. Hoewel deze gewichten tijdens de training aanvankelijk energie vereisen, verbruiken ze, eenmaal ingesteld, nauwelijks energie.



Figuur 1: Schakelinglayout van de vierpoortarchitectuur met 3x3 MMI's als knooppunten die onderling zijn verbonden door spiraalgolfgeleiders. Een geïntegreerde readout met complex-waardige gewichten op de chip is ook weergegeven.

3 Resultaten

3.1 Egalisatie met Zelf-Coherente Ontvangers

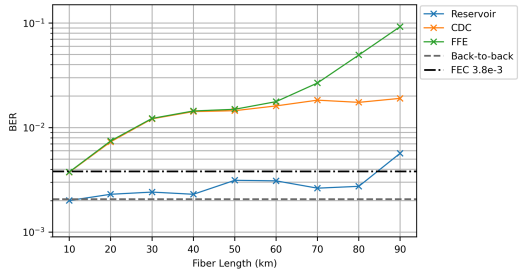
Transmissie in optische vezelcommunicatiesystemen induceert doorgaans signaalverstoringen, die voortkomen uit de respons van verschillende componenten, waaronder bronnen, optische vezels en ontvangers. Dientengevolge is signaalegalisatie essentieel voor het waarborgen van foutloze communicatie en dit wordt doorgaans uitgevoerd via digitale signaalverwerking. Dit omvat procedures zoals het schatten en inverteren van de kanaalrespons, evenals adaptieve egalisatie voor variabele verstoringen.

De prestaties van dergelijke procedures, en in het bijzonder de correctie voor de kanaalverstoringen, zijn effectiever wanneer het volledige optische veld beschikbaar is voor verwerking. Dit zorgt ervoor dat de compensatie op een eenvoudige manier kan worden geïmplementeerd met behulp van de transferfunctie van de vezel die

wordt gerealiseerd met lineaire filters. Echter, wanneer gebruik wordt gemaakt van zelf-coherente architecturen, kan de ontvangerstransferfunctie niet-lineair zijn als de operationele omstandigheden niet optimaal zijn. Dit introduceert een niet-lineaire transformatie van het optische veld, wat resulteert in suboptimale digitale egalisatie.

Echter, eenvoudigere oplossingen kunnen worden toegepast als de compensatie wordt uitgevoerd *voor* detectie. Deze benadering maakt gebruik van het optische veld voordat het de niet-lineaire ontvangertransformaties ondergaat, wat op zijn beurt een nauwkeurige egalisatie van kanaalverstoringen mogelijk maakt. Bovendien wordt deze verwerking uitgevoerd in het analoge optische domein, waardoor de noodzaak voor digitale egalisatie die op dubbele baudsnelheid werkt, wordt verminderd. Over het algemeen maakt dit het mogelijk om zelf-coherente ontvangers te gebruiken met lage operationele beperkingen, terwijl bevredigende bitfoutpercentages (BER) worden bereikt.

Met dit doel verkennen we in simulaties het gebruik van een fotonisch reservoir voor egalisatie in systemen die zelf-coherente ontvangers gebruiken. De onderzochte systemen maken gebruik van 64-QAM-transmissie, die zeer gevoelig is voor niet-lineaire transformaties, en omvatten implementaties met zowel een KK-ontvanger als een SiP SSBI-onderdrukkende ontvanger. De operationele omstandigheden van de ontvangers worden gevarieerd om de voordelen van optische egalisatie te bestuderen met variërende graden van ontvangerverstoringen. De simulatieresultaten tonen aan dat het aanpakken van kanaalverstoringen vóór een perfect coherente ontvanger DSP-prestaties evenaart, terwijl dit vóór een verstorende ontvanger de DSP-prestaties aanzienlijk overtreft. Een voorbeeld hiervan wordt getoond in Figuur 2, waar een KK-ontvanger (wiens werking wordt beïnvloed door het vermogen van de co-gepropagateerde draaggolf) wordt gebruikt. Er wordt een draaggolf met laag vermogen gebruikt, en daarom is het gedrag van de ontvanger verstorend. De resultaten tonen aan dat reservoiregalisatie (in het blauw) de DSP-oplossingen (in het oranje en groen) overtreft voor verschillende transmissielinks.



Figuur 2: BER vs vezellengte voor het reservoir (blauw), de benchmark chromatische dispersiecompensatie (CDC, oranje) en de benchmark feedforward-egalisatie (FFE, groen). De gestippelde grijze lijn geeft de back-to-back resultaten aan en de puntgestippelde zwarte lijn geeft een forward-error-correction (FEC) drempel van 3,8e-3 aan.

De verkregen resultaten tonen aan dat compensatie voor kanaalverstoringen vóór de zelf-coherente ontvangers de algehele systeemprestaties kan verbeteren. Om deze resultaten verder uit te breiden, bestuderen we vervolgens of de prestaties van de KK-ontvanger zelf kunnen worden verbeterd door fotonische reservoirverwerking.

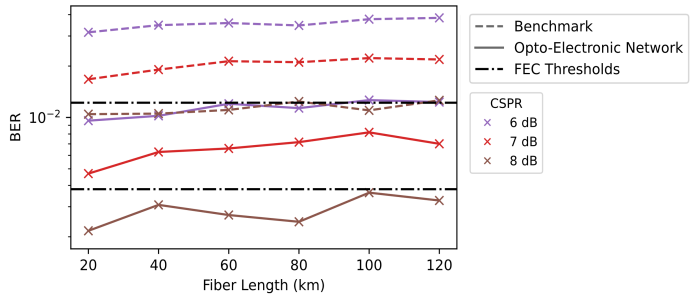
3.2 Linearizeren van de KK-ontvanger

De KK-ontvanger is een zelf-coherente ontvanger die coherente detectie ondersteunt via een zeer eenvoudige hardwarearchitectuur bestaande uit een enkele fotodiode. Het waarborgen van lineariteit binnen deze ontvanger vereist echter strikte operationele voorwaarden. Namelijk, een hoge vermogensverhouding tussen de co-gepropagateerde drager en het berichtsignaal is nodig. Dit wordt de carrier-signal-power ratio (CSPR) genoemd en is doorgaans meer dan 9 dB. Bovendien zijn er 6 samples/symbool (sps) vereist om de verwerkingsstappen van de ontvanger uit te voeren. Deze voorwaarden onderstrepen de afwegingen tussen hardwarecomplexiteit en verwerkingsvereisten van zelf-coherente ontvangers, wat hun praktische toepasbaarheid in korte netwerken beïnvloedt.

We verkennen daarom het gebruik van opto-elektronische machine learning-netwerken om de CSPR- en sps-beperkingen van KK-ontvangers te verlichten, zodat ze beter toepasbaar zijn in korteafstandssystemen. Het netwerk bestaat uit een fotonisch reservoir dat gevolgd wordt door een trainbare elektronische feedforward-egalisator. Dit trainbare opto-elektronische netwerk bevindt zich rondom de KK-ontvanger en wordt gebruikt om niet-lineaire fouten te compenseren die ontstaan bij een KK-ontvanger die onder ernstig beperkte condities opereert, waarbij 3 sps in plaats van de vereiste 6 worden gebruikt en bij CSPR's onder de 9 dB.

Numerieke resultaten tonen aan dat deze opto-elektronische architectuur een significante verbetering in de lineariteit van de ontvanger laat zien, wat betere integratie mogelijk maakt in een digitale signaalverwerkingspijplijn voor chromatische dispersiecompensatie (CDC) en blinde fasezoektocht (BPS). Belangrijk is dat het optische reservoir een banddoorlaatfilter vervangt vóór de KK-ontvanger, wat doorgaans vereist is in zelf-coherente systemen. Bovendien wordt de training uitgevoerd op een back-to-back systeem, zonder dat kennis van de verbindingsparameters vereist is, waardoor het netwerk op een "plug-and-play" manier in variabele verbindingen kan worden ingevoegd. We richten ons op korteafstandssystemen van maximaal 250 km die 64 QAM inzetten, en tonen tot 4 keer reductie in BER aan in simulaties.

Een aantal van de behaalde resultaten is weergegeven in Figuur 3, waar de BER als functie van vezellengtes tussen 20 en 120 km voor drie CSPR-waarden wordt getoond. Naarmate het opto-elektronische netwerk de lineariteit van de ontvanger verbetert, worden lagere BER's bereikt, met een verbetering van ongeveer 3-4 keer.



Figuur 3: Resultaten van het opto-elektronische netwerk worden weergegeven als doorgetrokken lijnen, terwijl resultaten van de benchmark worden weergegeven als gestippelde lijnen. Puntgestippelde lijnen komen overeen met de FEC-drempels. De kleuren verwijzen naar verschillende CSPRs variërend tussen 6 en 8 dB.

3.3 Egalisatie met Coherente Ontvangers

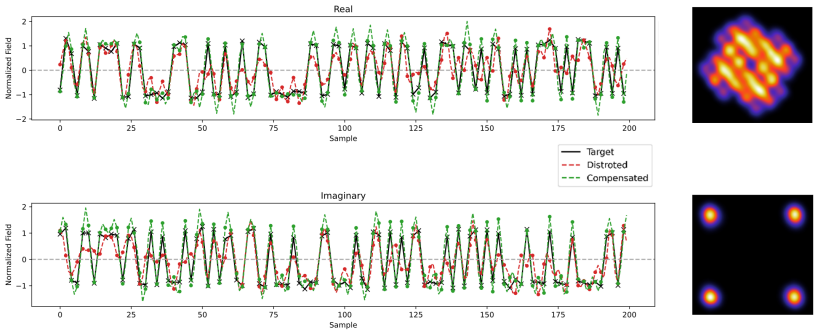
Tenslotte onderzoeken we ook het gebruik van fotonische reservoiregalisatie met standaard coherente ontvangers, waarbij we ons richten op de signaalverwerkingsvereisten die doorgaans door vermogenintensievere digitale methoden worden afgehandeld. Analoge alternatieven, die het energieverbruik kunnen verlagen en de digitale bandbreedtebeperkingen kunnen omzeilen, zijn bijzonder relevant voor functies zoals chromatische dispersiecompensatie en actieve egalisatie, die oversampling vereisen bij twee samples per symbool. Deze twee taken alleen al verbruiken een aanzienlijk deel van de DSP-kracht—ongeveer 15% voor chromatische dispersie en tot 50% voor actieve egalisatie.

Met dit doel demonstreren we experimenteel het gebruik van een analoge oplossing, gerealiseerd in een geïntegreerd fotonisch reservoir, om de elektronische verwerkingsbehoeften van coherente transceivers die 4 en 16 QAM gebruiken, te verlichten. Dit stelt ons niet alleen in staat om de grote bandbreedtes van optica te benutten voor het verwerken van signalen, maar ondersteunt ook op natuurlijke wijze de gezamelijke verwerking van beide kwadraturen. Het reservoir kan egalisatietaken uitvoeren en vervangt daardoor digitaal geïmplementeerde egalisatoren voor chromatische dispersie en ontvangerverstoringen. Het verplaatsen van deze stappen, die normaal gesproken op dubbele baudsnelheid werken, naar het optische domein vermindert aanzienlijk de rekenkundige complexiteit van de pijplijn en de bandbreedte van de elektronica.

In deze experimenten koppelen we ons netwerk aan standaard DSP-blokken die andere noodzakelijke procedures uitvoeren, waaronder frequentiecompensatie en blinde fasezoektocht, en tonen zo aan dat ons netwerk naadloos kan integreren in bestaande DSP-pijplijnen. Deze modulaire benadering maakt geleidelijke migratie naar een optische co-verwerkingsoplossing mogelijk, voordat we overstappen naar

volledige optische verwerking of optisch-analoog hybride systemen.

Om de impact van deze verwerking te visualiseren, toont Figuur 4 de tijdsdomeingolfvormen van een 4-QAM-signaal, waarbij de reële en imaginaire componenten afzonderlijk zijn uitgezet. De verstoorde en gecompenseerde signalen worden respectievelijk in rood en groen weergegeven, met het ideale signaal in zwart. Zoals te zien is, komt het gecompenseerde signaal overeen met het ideale signaal, wat wijst op succesvolle egalisatie. De effectiviteit van de compensatie wordt verder gevisualiseerd door constellatiediagrammen, waarin de verstoorde (boven) en gecompenseerde (onder) signalen worden vergeleken.



Figuur 4: Golfvormen en constellatiediagrammen van een 4-QAM-signaal vóór en na reservoircompensatie. Aan de linkerkant toont het bovenste paneel de reële component, terwijl het onderste paneel de imaginaire component toont. De zwarte, rode en groene golfvormen vertegenwoordigen respectievelijk het ideale, verstoorde en gecompenseerde signaal. De constellatiediagrammen tonen het verstoorde signaal (boven) en het reservoirgecompenseerde signaal (onder).

4 Conclusie en Vooruitzichten

De bevindingen gepresenteerd in deze thesis vormen een basis voor de verdere ontwikkeling van het gebruik van fotonische reservoircomputing om systemen gebaseerd op zowel coherente als zelf-coherente ontvangers te vereenvoudigen. Deze benadering toont veelbelovendheid in het verminderen van de complexiteit van traditionele DSP-taken door de voordelen van fotonische systemen, zoals parallelle verwerkingscapaciteiten en grote bandbreedtes, te benutten.

Gezien deze veelbelovende resultaten is de migratie van de resterende DSP-procedures naar het optische domein een belangrijk doel. Dit kan de afhankelijkheid van traditionele DSP aanzienlijk verminderen, waardoor de verwerkingsoverhead en de latentie die samenhangen met elektronische componenten tot een minimum worden beperkt. Het bereiken hiervan zou een

belangrijke stap zijn naar volledig optische verwerkingsunits, die, in combinatie met vereenvoudigde ontvangerarchitecturen, superieure prestaties en verbeterde energie-efficiëntie kunnen leveren. Dergelijke vooruitgangen zouden de creatie van schaalbare en kosteneffectieve oplossingen voor een breed scala aan optische communicatieapplicaties mogelijk maken, waarmee wordt ingespeeld op de groeiende vraag naar hoogsnelheid-, hoge-capaciteitsnetwerken.

Summary

1 Introduction

The deployment of fiber networks is becoming increasingly concentrated in shorter-reach applications, where individual systems span only a few kilometers, but their combined scale is substantial. Given the sheer volume of these deployments, short- and mid-reach networks require cost-effective solutions, distinct from those used in long-haul systems, where higher investments can be distributed over longer distances and larger user bases.

For instance, coherent transmission is effective for long-haul systems but involves high receiver costs and substantial processing demands, limiting its feasibility for high-density networks. Research is therefore focused on reducing receiver complexity through alternative architectures, as well as simplified signal processing methods to make coherent solutions more practical for short- and mid-reach networks.

One approach to simplify receivers is to use self-coherent architectures, where the local oscillator is co-propagated with the message signal. Fewer photodiodes are also used, followed by different processing techniques to retrieve complex-valued signals from detector outputs. Examples include the Kramers-Kronig (KK) receiver and the silicon photonic signal-signal beat interference canceling receiver (SiP SSBI). However, these systems often face stringent operational constraints that may limit their practical deployment.

Efforts to simplify signal processing are driven towards analog solutions, which circumvent the need for digital processing using over-sampled signals. This is also driven by concerns about the long-term suitability of digital processing, since baud rate processing capacity for digital signal processing (DSP) chips is growing at just 13% annually, far behind the 60% growth in data rate demands. Simultaneously, the current maturity of integrated photonics allows for the miniaturization and integration of various photonic components onto a compact chip, enabling efficient manipulation of light within a minimal footprint. This integrability, along with attributes such as frequency parallelism, low-latency operation, and low transmission loss positions optical processing as a promising alternative to traditional electronics.

Inspired by previous successes where optical subsystems replace traditional electronics in a fiber optic communication network, this thesis explores optical alternatives for equalizing distorted communication signals. This work also aligns with the ”post-digital” research theme, which explores unconventional, non-binary computing systems for tasks that are economically or environmentally impractical to handle digitally. This concept underpins the PostDigital project, a Marie Skłodowska-Curie Innovative Training Network, within which this thesis was developed and from which we borrow the term.

To this end, this thesis replaces traditional digital computation with analog, integrated photonic processing. The processing architecture is based on reservoir computing which is a machine learning paradigm. Similar to other machine learning approaches, rather than explicit programming, the chip adjusts its parameters based on exposure to labeled datasets to achieve the desired behavior. The use of photonic processing is thus leveraged to unlock wider deployment of coherent modulation in short- and mid-reach communication systems. This goal is achieved by addressing two key aspects: simplifying processing requirements and enabling self-coherent receiver architectures to function under less stringent conditions.

2 Photonic Reservoir Computing

Training machine learning networks can be challenging and often requires iterative optimization methods that utilize substantial computational resources and large datasets to converge. To simplify the training process, Reservoir Computing (RC) offers an alternative approach by fixing the internal network connections (the reservoir) and only training the output layer, which is known as the readout. The reservoir acts as an internal projection of the input(s), generating a sufficiently rich variety of features that the readout layer can be trained on. Despite its apparent simplicity, this approach has demonstrated success in various applications, including time series forecasting and classification/regression tasks.

RC has also emerged as a promising approach for machine learning in physical computing. This is driven by its ability to efficiently construct networks in environments where manipulating internal weights is challenging. Furthermore, the reservoir’s capacity to leverage diverse physical phenomena in their natural form has expanded its application scope. As a result, RC has been explored across a wide range of substrates, including electronics, spintronics, biological systems, and photonics.

The photonic reservoir used in this thesis is called the four-port architecture. The reservoir employs 3x3 MMIs as nodes that are interconnected by spiral waveguides, as shown in Figure 5. At each node, input signals interfere coherently to produce different output signals. This reservoir’s architecture is passive, as it does not require external power to drive the nodes or connections. Furthermore, non-volatile

optical weights can also be integrated on the chip for the readout. While these weights require initial power during training, once set they hardly consume any electrical power.

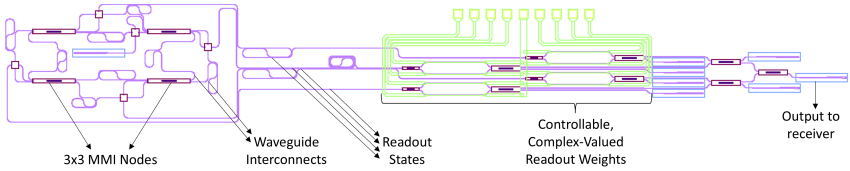


Figure 5: Circuit layout of four-port architecture employing 3x3 MMIs as nodes interconnected by spiral waveguides. An integrated readout with on-chip complex-valued weights is also shown

3 Results

3.1 Equalization with Self-Coherent Receivers

Transmission in fiber optic communication systems typically induces signal distortions, which arise from the response of various components including transmitters, optical fibers, and receivers. As such, signal equalization is essential for ensuring error-free communication and is typically performed through digital signal processing. This includes procedures like estimating and inverting the channel response, as well as adaptive equalization for variable impairments.

Performance of such procedures, and in particular correction for the channel impairments, is more effective when the full optical field is available for processing. This ensures that the compensation can be implemented in a straightforward manner using the fiber's transfer function implemented with linear filters. However, when using self-coherent architectures, the receiver transformation can be nonlinear if the operational conditions are not optimal. This introduces a nonlinear transformation on the optical field, resulting in suboptimal digital equalization.

However, simpler solutions can be employed if the compensation is performed *prior* to detection. This approach leverages the use of the optical field before it undergoes the nonlinear receiver transformations, which in turn enables accurate equalization of channel impairments. Additionally, this processing is performed in the analog optical domain, thus alleviating the need for digital equalizers that operate at twice the baudrate. Overall, this allows using self-coherent receivers with low operational constraints while achieving satisfactory bit error rates (BER).

To this end, we explore in simulations the use of a photonic reservoir for equalization in systems deploying self-coherent receivers. The systems examined use 64-QAM transmission, which is highly sensitive to nonlinear transformations, and include

implementations with both a KK receiver and a SiP SSBI-cancelling receiver. The operational conditions of the receivers are varied to study the advantages of performing optical equalization with varying degrees of receiver distortions. The simulation results show that addressing channel impairments prior to a perfectly coherent receiver will match the DSP performance, while doing so before a distorting receiver significantly surpasses the DSP performance. An example of this is shown in Figure 6, where a KK receiver (whose operation is affected by the power of the co-propagated carrier) is utilized. A low carrier power is used, and as such the behaviour of the receiver is distorting. The results show that reservoir equalization (in blue) outperforms the DSP solutions (in orange and green) for various transmission links.

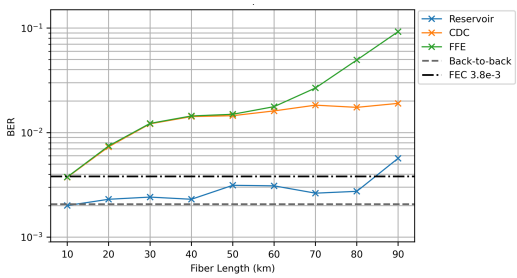


Figure 6: BER vs fiber length for the reservoir (blue), the benchmark chromatic dispersion compensation (CDC, orange), and the benchmark feed-forward equalizer (FFE, green). The dashed grey line indicates the back-to-back results and the dash-dotted black line indicates a forward-error-correction (FEC) threshold of 3.8e-3.

The obtained results show that compensation for channel impairments before the self-coherent receivers can improve the overall system performance. To further extend these results, we next study whether the performance of the KK receiver itself can be enhanced by photonic reservoir processing.

3.2 Linearizing the KK Receiver

The KK receiver is a self-coherent receiver that supports coherent detection through a very simple hardware architecture comprising a single photodiode. However, ensuring linearity within this receiver necessitates stringent operational conditions. Namely, a high power ratio between the co-propagated carrier and the message signal is needed. This is termed carrier-signal-power ratio (CSPR) and is typically over 9 dB. Furthermore, 6 samples/symbol (sps) are required to implement the receiver's processing steps. These conditions underscore the tradeoffs between hardware complexity and processing requirements of self-coherent receivers, which affects their practicality in short-haul networks.

We thus explore using opto-electronic machine learning networks to ease the CSPR and the sps constraints of KK receivers, such that they are more applicable to short-haul systems. The network comprises a photonic reservoir that is followed by a trainable electronic feed-forward equalizer. This trainable opto-electronic network sandwiches the KK receiver and is used to compensate for nonlinear errors arising from a KK receiver operated under severely limited receiver resources, where 3 sps instead of the required 6 are use and at CSPRs below 9 dB.

Numeric results demonstrate that this opto-electronic architecture shows significant improvement in receiver linearity, which allows for better integration in a digital signal post-processing pipeline for chromatic dispersion compensation (CDC) and blind phase search (BPS). Importantly, the optical reservoir replaces a bandpass filter before the KK receiver, which is typically required in self-coherent systems. Furthermore, the training is done on a back-to-back system, without requiring knowledge of the link parameters, allowing the network to be inserted in a "plug-and-play" manner in varying links. We target short-reach systems of up to 250 km deploying 64 QAM, and show up to 4 times reduction in BER in simulations.

Some of the achieved results are shown in Figure 7, where the BER against fiber lengths spanning between 20 and 120 km for three CSPR values are shown. As the opto-electronic network enhances the receiver's linearity, lower BERs are achieved, with around 3-4 times improvement seen.

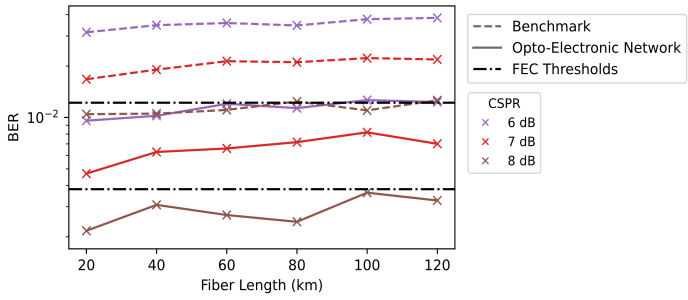


Figure 7: Results from the opto-electronic network are shown as solid lines, while results from the benchmark are shown as dashed lines. Dash-dotted lines correspond to the FEC thresholds. The colors refer to different CSPRs ranging between 6 and 8 dB.

3.3 Equalization with Coherent Receivers

Finally, we also look into using photonic reservoir equalization with standard coherent receivers, focusing on the signal processing demands typically handled by power-intensive digital methods. Analog alternatives, which may lower power consumption and bypass digital bandwidth constraints, are particularly relevant for

functions like chromatic dispersion compensation and active equalization, which require oversampling at two samples per symbol. These two tasks alone consume a significant portion of DSP power—about 15% for chromatic dispersion and up to 50% for active equalization.

To this end, we experimentally demonstrate the use of an analog solution, realized in an integrated photonic reservoir, to relax electronic processing needs of coherent transceivers deploying 4 and 16 QAM. This not only allows harnessing the large bandwidths of optics for processing signals, but also naturally supports one-shot processing of both quadratures. The reservoir can perform equalization tasks and therefore replaces digitally implemented equalizers for chromatic dispersion and transceiver impairments. Moving these steps, which normally operate at twice the baud rate, into the optical domain considerably reduces the computational complexity of the pipeline and the bandwidth of the electronics.

In these experiments we pair our network with standard DSP blocks performing other necessary procedures including frequency offset compensation and blind phase search, thus showcasing that our network can seamlessly integrate with existing DSP pipelines. This modular approach allows gradual migration to an optical co-processing solution, before moving to fully optical processing or optical-analog hybrids.

To visualize the impact of this processing, Figure 8 shows a 4-QAM signal's time-domain waveforms, with the real and imaginary components plotted separately. The distorted and compensated signals are show in red and green respectively, with the ideal signal shown in black. As seen, the compensated signal matches the ideal signal, thus indicating successful equalization. The effectiveness of the compensation is further visualized through constellation diagrams, comparing the distorted (top) and compensated (bottom) signals.

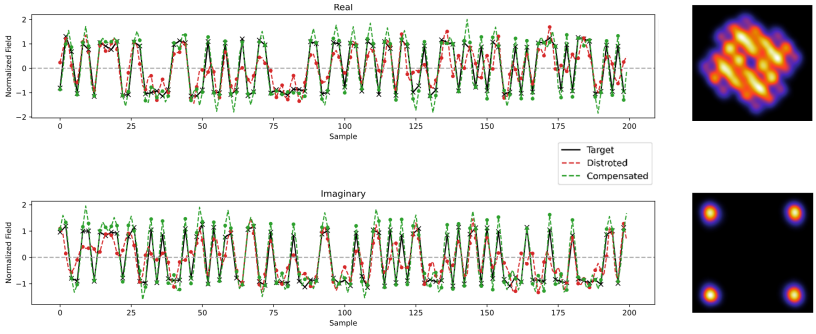


Figure 8: Waveforms and constellation diagrams of a 4-QAM signal before and after reservoir compensation. On the left, the top panel shows the real component while the bottom panel shows the imaginary component. Black, red, and green waveforms represent ideal, distorted, and compensated signals, respectively. The constellation diagrams show the distorted signal (top) and the reservoir compensated (bottom)

4 Conclusion and Outlook

The findings presented in this thesis provide a foundation for advancing the use of photonic reservoir computing to simplify systems based on both coherent and self-coherent receivers. This approach shows promise in reducing the complexity of traditional DSP tasks by leveraging the benefits of photonic systems such as parallel processing capabilities and large bandwidths.

Given these promising results, the continued migration of the remaining DSP procedures to the optical domain is an important objective. This would significantly reduce the reliance on traditional DSP, thereby minimizing the processing overhead and latency associated with electronic components. Achieving this would be a major step towards fully optical processing units, which, when combined with simplified receiver architectures, can deliver superior performance and improved energy efficiency. Such advancements would enable the creation of scalable and cost-effective solutions for a wide range of optical communication applications, addressing the growing demand for high-speed, high-capacity networks.

1

Introduction

1.1 Overview

Fiber optic communication systems (FOCSs) are the backbone of today's highly connected world, enabling high-speed data transport over lengths ranging from short distances in data centres to thousands of kilometers in submarine fibre-optic links. This proliferation can be attributed to several optical characteristics, including large bandwidths, low-loss channels, and multiplexing in amplitude, phase, space, polarization, and wavelength [1]. Indeed, our incremental but fast-paced ability to exploit these advantages have supported the growth of aggregate data rates from modest Megabytes/s at the start of commercial FOCSs around 50 years ago to today's Terabytes/s and beyond. Naturally, this scaling would not have been possible without driving advances in individual components, subsystems, and processing techniques.

A notable example is the advent of the erbium-doped amplifier (EDFA), which allowed optical amplification of signals. In its absence, the first 20 years of FOCSs were marked by electric regeneration, where signals were detected, electrically amplified, and then re-transmitted on a span-by-span basis [1]. Through supporting optical amplification, EDFAs not only saved costs and enabled longer-reach transmission, but also allowed the use of wavelength division multiplexing (WDM), which would not have been feasible if electric regeneration were necessary.

Another example is the introduction of Optical Add-Drop Multiplexers (OADMs) for optical networking. OADMs facilitate efficient wavelength routing at add/drop network nodes, especially when several wavelengths carry pass-through traffic and must be forwarded to the next node. Before OADMS, pass-through traffic was handled through opto-electro-optical conversion or a back-to-back demultiplexer-multiplexer (where wavelengths are separated at the demultiplexer, some are dropped at the node, and the remaining wavelengths are recombined at the multiplexer) [2]. These solutions were suboptimal as the first was too expensive and the latter introduced prohibitive optical losses especially in multi-hop scenarios [3]. Therefore, an efficient optical networking layer that bypassed high-speed electronics at routing interfaces was imperative, which led to the development of OADMs at first and later the reconfigurable variant known as ROADMs [4]. This allowed maximizing the use of optical bandwidth and enabled using flexible frequency grids.

These are but two of the many milestones that FOCSs have witnessed over the years. However, they were specifically selected to exemplify how significant strides can be achieved "simply" by replacing system blocks from their traditional electric

implementation to an optical one, thus extending the advantages of optics beyond mere links. Following in this theme, this thesis investigates the use of *optical* as opposed to standard *electronic* processing for optical communication signals. These processing procedures are necessary to compensate for distortions that occur at different system blocks, either due to their intrinsic distorting nature or due to an imperfect implementation. Motivating this transition are the challenges of deploying coherent transceivers in short- and medium-reach links due to their associated power-hungry and computationally-intensive signal processing requirements. Such operational prerequisites are prohibitive of exploiting technologically viable solutions for next-generation networks until their economical and environmental impact is lowered [5]. In fact, issues including greenhouse gases *already* drive such decisions on governmental and organizational levels [6].

In this regard, this work also overlaps with the research theme of a "post-digital" world which investigates more natural forms of computing for certain applications that are too expensive - economically or environmentally - to perform on digital computers [7]. This results in unconventional computing systems that are either non-binary, non-electronic, or both. Since this commonly also means bypassing logic-based computation, other forms of programability, including those based on machine learning techniques, are relied on. Additionally, nature is used for inspiration on computational concepts. For example, the overarching biological theme "the form is the function" inspires building application-specific computing devices whose architectures closely match their purposes [8]. Inspired by this terminology, in this thesis we replace digital electronic computation with analog, hardware-based, integrated photonic computation. The architecture of the computing device supports (limited) in-situ memory and computes following a machine learning technique called reservoir computing. Just like in other machine learning approaches, the chip is not explicitly programmed, but rather adjusts its programmable parameters such that it can produce a desired behaviour from exposure to labelled data sets.

In the remainder of this chapter, we situate our work at the intersection of the aforementioned fields. We first introduce the targeted optical communication systems for which our optical signal processing is implemented. We then introduce some post-digital computing concepts, as well as a brief introduction on integrated photonics as the substrate for performing these computations. The chapter then discusses in more detail the objectives of this thesis, the thesis outline, and attributions of the work. Finally, a list of publications in which parts of this work were disseminated is provided.

1.2 Short- and Mid-Reach FOCSs

Optical data traffic has evolved from predominantly traversing long-haul networks that cover extensive distances, to being increasingly transported within short-reach networks [9]. This stems from the surge in inter- and intra-data center traffic, as well as from new emergent applications where short but fast communication links are needed. Such applications include the widespread adoption of cloud services, the deployment of fiber-to-home networks, and the expanded use of optical front-hauling in mobile communication. These changes in fiber deployment trends are shown in Figure 1.1, where the accumulated lengths of fiber for different applications are shown. Interestingly, the use of fiber in access networks like Fiber to the Curb (FTTC) and Fiber to the Home (FTTH) (collectively referred to as FTTx) only started in the late 1990s but became a leading use-case for fiber networks. Furthermore, Figure 1.1 shows a constant 15 percent increase in fiber deployment since the 2000s that is largely attributed to short-reach networks [1]. This trend also translates to an explosive number of individual systems, each a few kilometers in length, and whose accumulated total is significant.

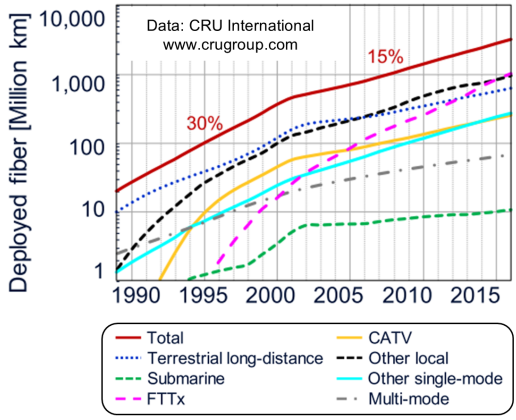


Figure 1.1: Fiber installed for different applications worldwide since 1990 shows significant increase in deployment of short and mid reach fibers as opposed to long haul and submarine fibers. FTTx: Fiber to the x (e.g., curb / building / home), CATV: Cable TV. Image from [1]

Considering the volume at which these networks are deployed, short- and mid-reach applications demand cost-effective solutions different from those used in long-haul systems, where expensive investments can be spread over vast distances and a number of users. Coherent communication, for example, which encodes data in both amplitude and phase, has been a key enabler for increasing transmission throughputs in long-haul communications without taxing the limited bandwidths of electronic

interfaces, since it allows multilevel modulation over two orthogonal quadratures (i.e. quadrature-amplitude modulation or QAM). However, the expensive hardware of a coherent receiver, as well as the power-hungry digital signal processing (DSP) procedures that are required post-detection, remain hindering for high-baudrate implementations and/or for mass deployment of coherent transceivers in short- and mid-reach systems [10].

In this introduction, two such short- and mid-reach network use-cases are presented, with the aim of placing into context the proposed solutions of this thesis. Namely, data center networks and interconnects as well as passive optical networks are introduced with their respective design constraints and future outlooks. Then, current research proposals for bridging the coherent gap in these networks are summarized.

1.2.1 Data-Center Networks and Interconnects

Modern internet-driven applications - including search services, Internet of Things, video streaming, and social networking - rely heavily on the infrastructure provided by data centers (DC) [11]. These facilities house massive server farms, interconnected by high-performance networks that facilitate communication and data exchange, both between servers in one data center as well as between servers across geographically dispersed locations [12]. These internal connections in a DC form the Data Center Networks (DCNs), while communication between DCs is performed through Data Center Interconnects (DCIs). The design and optimization of these networks are therefore paramount to ensuring the scalability, performance, and efficiency of the applications they support.

For DCNs, which are typically a few kilometers in length, intensity modulation and direct detection (IM/DD) has been the main transceiver technology. More recently, research advances supported the upgrade from standard IM/DD to 4-level pulse amplitude modulation (PAM4), thus increasing the spectral efficiency while maintaining compatibility with IM/DD systems [13]. However, to address the ever-growing data rate demands within DCNs, research is actively exploring the feasibility of coherent transceivers, which are anticipated to play a pivotal role in future high-bandwidth solutions [13–16]

DCIs span considerably longer distances, typically in the range of 100 km but may even reach up to a few hundred kilometers. While IM/DD remains the prevalent technology in DCIs, coherent transceivers are witnessing increasing adoption in longer-reach networks ($>$100 km) due to their ability in supporting higher data rates and in mitigating dispersion through signal processing [1, 11, 17–19].

1.2.2 Passive Optical Networks

Data-rate demands from user-driven applications not only rely on efficient networking in data centers, but indeed also require high-speed low-latency connections until the last mile. This constitutes large infrastructural and operational costs, since it involves connecting all users through FOCSs. Therefore, simple network architectures are of interest here. To this end, passive optical networks (PONs) offer a cost-effective solution by employing passive optical components throughout the network, with electronics only required at the transmitter and receiver ends. This is commonly represented by the architecture in Figure 1.2. Here, a fiber link extends from the service provider's central office and through passive optical splitters distributes traffic to end users. Users can then share this centralized connection through time-division multiplexing (TDM) or WDM. Connections can also be bi-directional by utilizing different wavelengths for the uplink and downlink data. As a cornerstone technology for FTTx deployments, PONs currently dominate the FTTH landscape, to the extent that the terms are often used interchangeably [20].

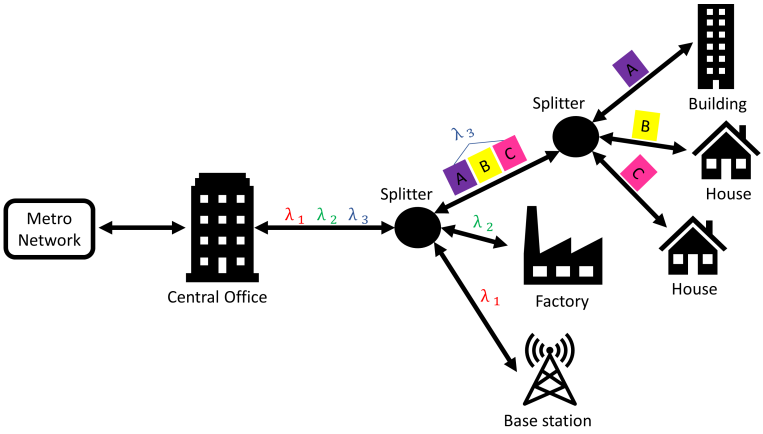


Figure 1.2: PON connecting end users to the metro network through the use of passive splitters. Users share the network infrastructure through WDM and/or TDM technologies, where λ_1, λ_2, and λ_3, are different wavelength carriers and where A, B, and C are time multiplexed packets

So far, all PON generations have relied on the use of IM/DD due to the cost sensitivity of these networks and the cost-effective nature of such a transceiver. Additionally, next generation's 50 Gbits/s High-Speed Capable PON is foreseen to continue using IM/DD technologies [21]. However, there is general consensus that scaling beyond these data rates would require transitioning to coherent systems [20].

1.2.3 Bridging the Coherent Gap

Although concepts enabling coherent detection were introduced as early as the 1980s, the fast-paced growth of WDM systems kept it on the sidelines for around two decades [22]. It was not until the mid-2000s, following the growth and popularization of the internet, that it was brought back due to the need for supporting higher data rates. It was clear then that the targeted 40 Gbits/s transmission would not be possible without dispersion compensation enabled by coherent detection. Furthermore, the spectral efficiency of the deployed IM/DD systems were sub-optimal, and thus it was important to increase it to maximize use of the limited electronic bandwidths available at the interfaces [1]. Industrial deployments started with Nortel's successful demonstration of coherent detection with real-time processing [23], marking the start of a new coherent-based era for long-haul systems.

Currently, short- and mid-reach networks face the same set of challenges that led to the adoption of coherent detection in long-haul systems, however with more stringent requirements. Firstly, the cost of the transceiver itself must be low due to the large volume of deployed transceivers. This is a concern since the hardware of a coherent receiver is much more complex and traditionally consists of two pairs of balanced photodiodes, a local oscillator and a 90-degree hybrid. Secondly, the power consumption of the associated DSP must be kept to a minimum due to its environmental impact and economical costs. To bridge this coherent gap, common research directions investigate both reduction in receiver complexity through alternative architectures, as well as more efficient methods for the required signal processing procedures post detection [13].

Efforts in reducing the receiver's complexity include self-coherent receiver architectures, which simplify the receiver's hardware by co-propagating the local oscillator with the modulated message signal. In addition, fewer photodiodes are used, followed by different processing techniques to retrieve the complex signal from the detectors' outputs. Notable examples include the Kramers-Kronig receiver [24], the Gerchberg-Saxton receiver [25], and a silicon photonics signal-signal beat interference (SSBI) cancelling receiver [26]. While these receivers have gained traction recently, they often have stringent operational constraints that make their operational feasibility a concern, a topic we discuss more in Chapter 3. Other solutions, including simplified but fully coherent receivers based on homodyne detection, have also been investigated [27–29].

To reduce the environmental and operational costs for signal processing, common research directions investigate replacing digital implementations with analog ones. This can circumvent the need for high-speed analog-to-digital converters (ADCs)

as well as reduce the operational bandwidth, since digitized solutions typically operate at twice the baud rate. To this end, analog electronic solutions coupled with simplified signal processing procedures can be promising candidates [28, 30]. Additionally, optical solutions are also investigated, since they offer processing bandwidths that match the transmission bandwidths and naturally support complex (i.e., real and imaginary) operations [31–38].

Finally, we remark that the transition to coherent detection in the 2000s was not just *need* driven, but also *opportunity* driven [1]. At the time, the Complementary Metal-Oxide-Semiconductor (CMOS) technology had scaled enough to allow developing signal processing chips for such high-speed transmission. This progress aligned with Moore's Law, which predicted the doubling of transistors on a chip every two years, leading to faster, more cost-effective, and compact computing solutions [39]. However, arising concerns regarding the long-term sustainability of Moore's Law are becoming increasingly prevalent [40]. Furthermore, within Application-Specific Integrated Circuits (ASICs) used for DSP chips, baudrate processing capacity is increasing by a mere 13% annually, which forms a significant bottleneck as it falls short of the growth observed in other transceiver components [41]. This also falls short of the ever-increasing demand for higher data rates, scaling at a constant 60% per year [1]. More ominously, this disparity is expected to widen since the current 13% growth rate for ASICs is already slowing down and may saturate [41]. Current maturity in photonic technologies may thus present as a timely opportunity to leverage their high bandwidth characteristics for signal processing.

In the next sections, we show that the search for alternatives to traditional digital computing is not unique to this field and that this pursuit is gaining traction across the wider computing community. Furthermore, we discuss advantages of optical computing as well as the current maturity in integrated photonics as propellers for this evolving computing paradigm.

1.3 Post-Digital Computing

The remarkable achievements of digital computers, rooted in symbolic logic and the Turing machine, are undeniable. However, these machines are limited by their reliance on electronic hardware elements operating in a binary on/off state [42]. This constraint forces discrete operation on continuous hardware and imposes a discrete framework on the natural continuity of the physical world [7]. On the other hand, computing that exploits the natural dynamics of physical substrates to perform computations in harmony with their inherent physical properties may be better suited for certain applications. This formed the basis of the PostDigital

project [43], a Marie Skłodowska-Curie Innovative Training Network in the context of which this thesis was performed and from which we borrow the term.

This domain of alternative computing paradigms is large and encompasses a vast landscape, including terms like neuromorphic computing, physical computing, material computing, unconventional computing, and even analog computing. This introduction focuses specifically on two areas relevant to this thesis, namely neuromorphic computing and optical computing.

1.3.1 Neuromorphic Computing

The field of Artificial Intelligence (AI) is booming, transforming healthcare, finance, language processing, and beyond. Despite impressive computational power, these artificial models struggle to match the biological brain's efficiency in flexibility and performance for many cognitive tasks. Remarkably, the brain achieves this feat with minimal power consumption – roughly equivalent to a light bulb (around 20 watts) – while supercomputers require significantly more, often by orders of magnitude [44]. This power consumption is not only relevant during training AI models, but according to recent estimates from industry leaders like Nvidia and AWS, total power consumption during repeated inference surpasses training [45,46].

Bridging this efficiency gap requires reconsidering the von Neumann architecture, as it is ill-suited for performing AI-required computations efficiently [47]. To this end, neuromorphic computing attempts to replace such conventional computing architectures with ones that more closely emulate the brain's underlying structure and processing methods. This ambitious pursuit hinges on advancements in three crucial areas: materials, circuits, and algorithms [48]. Novel materials are investigated for creating devices that emulate various neuromorphic functions like neuron behaviours and synapse behaviours. Examples include synaptic transistors [49], memristors [50], and phase change memory [51]. These devices can be integrated into engineered circuits whose architectures are efficient in supporting the necessary computing, for example due to inherent parallelism. The circuits, in turn, may be biologically-plausible like spiking networks or just biologically-inspired, including various topologies of non-spiking neural networks [52]. Finally, algorithms or frameworks for learning in these networks are developed or borrowed and must be suited for the network's architecture. A prominent framework in this regard is reservoir computing (RC) which localizes the learning process within specific circuit components instead of distributing it across the entire network [53]. RC holds potential to significantly reduce energy consumption and simplify hardware while maintaining good performance, making it well-suited for neuromorphic computing.

This multi-faceted approach opens up avenues for exploring neuromorphic computing. By combining various aspects of materials, circuits, and algorithms, numerous implementations have been proposed. Notably, optical implementations have gained significant traction due to inherent advantages of optics. In the next section, these attributes will be discussed to show how they enable not only optical neuromorphic systems, but also optical computing in general.

1.3.2 Optical Computing

Optical computing has a rich, albeit intermittent history, marked by periods of interest and subsequent setbacks. Early explorations encompassed both analog photonic processing, including holography and Fourier Transforms, as well as digital photonic processing through optical transistors [54]. However, associated challenges and the early success of electronics often tempered enthusiasm. Currently, there is a strong resurgence that is driven by multiple challenges in conventional computing, including the impending limitations in transistor scaling, thermal effects from higher clock frequencies, the von-Neumann bottleneck, and the insatiable computational demands of AI applications [55]. Simultaneously, advancements in photonics have reached a maturity level that enables practical implementation on integrated circuits (IC). This integrability, as well as other optical attributes including massive frequency parallelism, low-latency operation, low transmission loss and negligible dissipation are some of the reasons optical computing may offer superiority to electronic computing [56]

To capitalize on this, industry giants including Nvidia and Intel, alongside emerging startups like Lightmatter, are investing substantial resources to harness optical advantages in computing. Despite this, general-purpose fully-optical computers are still challenging to realize and may be long-term aspirations [57, 58]. As such, the immediate focus is on integrating optical components into existing electronic systems, primarily for increasing the efficiency of interconnects and speeding up matrix operations [59–61]. These hybrid approaches can therefore offer a promising pathway to near-term benefits.

Given the challenges (and skepticism [61–63]) associated with developing general-purpose digital optical computers, significant efforts are targeted towards high-performance, application-specific, analog photonic computers. Inherent strengths of analog optics can thus be harnessed to perform computations like filtering, routing, convolution, and Fourier transforms [56]. This approach involves engineering optical circuits tailored for specific tasks, which may require long wait times and hamper advancements in this field. To shorten wait times and support the design process of such circuits, efforts in creating programmable photonic circuits

are also investigated [64]. One way this can be implemented is by realizing a mesh of optical basic units and altering the light path between them to achieve a desired behaviour [65].

Concurrently, there is a prominent rise in neuromorphic optical implementations as novel computing paradigms that can circumvent electronic bottlenecks. This approach seeks to map the inherently parallel nature of these models onto hardware that can efficiently support their parallel and distributed computational demands [66–68]. RC based-models are particularly prominent because they support physical implementations that are more hardware efficient and more tolerant to fabrication variability [69]. Furthermore, a growing emphasis on *domain-tailored* neuromorphic designs is evident. By strategically selecting node functions and optimizing network connectivity, researchers are developing efficient, yet intelligent, signal processors without the complexities of deep, highly nonlinear architectures [31, 70–72]. Note that this approach is not strictly application-specific but rather domain-tailored, a methodology increasingly relevant even in traditional computing [73].

Instrumental to all these endeavours in realizing optical computing is that there have been enormous advances in the scale, speed, and energy efficiency of photonic devices. This period has also seen the development and commercialization of photonic integrated circuits, giving a miniaturized alternative to bulk optics. In the next section, we briefly introduce integrated photonics, paying special attention to silicon photonics as a prime platform for realizing these circuits.

1.4 Integrated Photonics

Integrated photonics offers the capability of miniaturizing and integrating various photonic components onto a compact chip. This facilitates the generation, modulation, processing, and detection of light with minimal footprint [74]. This concept emerged in the 1960s to explore the possibility of replacing bulk optics with optical components on a single chip, similar to the ICs revolutionizing electronics at that time [75]. Early efforts focused on materials like Indium Phosphide (InP) that showed great potential for integrating lasers and detectors, but the 1980s saw a turning point with the realization that silicon, the dominant material in microelectronics, could also be used for photonic integration. [74]. This offered two primary advantages. First, the high index contrast of silicon photonic waveguides supported compact structures which allowed scaling to complex circuits on a small footprint. Second, the existing CMOS fabrication technology facilitated cost-effective, large-volume manufacturing for silicon photonic chips without substantial overhead investments to develop these fabrication facilities [76].

Silicon photonic (SiP) platforms are commonly realized through Silicon on Insulator (SOI), which uses silicon as the core material, and silicon nitride (SiN), which uses SiN for the core [76]. Both platforms can employ air or silicon dioxide as cladding materials to confine light within waveguides, thus achieving high refractive index contrasts that enable compact structures. These waveguides can also be engineered to perform functions such as optical interferometry and filtering. However, plain SiP platforms are sub-optimal for other functionalities including detection in the telecommunication O-band and C-band (due to transparency in the 1.3 to 1.55 μm region) as well as generation or amplification of light (due to an indirect bandgap). Thus, to support a wider scope of devices, extra layers can be incorporated. For example, germanium can absorb light efficiently in the O-band and C- band and can also be epitaxially grown on silicon, and thus is commonly used for integrating photodiodes on these platforms [77]. Other layers can also be added during fabrication as part of the integration flow to allow implementation of devices like phase shifters and heaters. Additionally, III-V semiconductors can be added for realizing lasers and amplifiers [78].

Despite the possible functionalities and advantages, photonic integrated circuits (PICs) have yet to achieve the widespread adoption of their electronic counterparts. While large-scale foundries support high-volume PIC production for applications like optical communications, smaller-scale startups or research and development efforts face significant cost barriers due to limited foundry capacity. To mitigate these costs, multi-project wafer (MPW) services enable cost-sharing among multiple users. [79]. Moreover, the lack of robust design software that supports component variability modeling and prediction often necessitates multiple fabrication iterations for PICs, leading to increased development time and costs [80]. Since accurate design software may take years to develop, programmable photonic circuits offer a potential solution by enabling rapid design prototyping on generic and tunable PICs, accelerating the development process.

1.5 Objectives

This thesis investigates the use of photonic processing, realized through an integrated reservoir computing architecture, to unlock wider deployment of coherent modulation in short- and mid-reach communication systems. This goal is achieved by addressing two key aspects: simplifying processing requirements and enabling self-coherent receiver architectures to function under less stringent conditions.

Simplified signal processing is achieved by utilizing a simple and integrated photonic architecture that performs equalization at the speed of transmission and without the need for oversampling. This approach is validated experimentally

using coherent receivers with 4- and 16-QAM formats. This is further extended through simulations to evaluate performance with 64-QAM self-coherent receivers, demonstrating the reservoir's versatility in signal equalization across different transmission scenarios and receiver architectures.

Additionally, the applicability of two self-coherent receivers—the Kramers-Kronig receiver and the silicon photonics SSBI-cancelling receiver—is enhanced through relaxing their operational conditions. The simulation results show how reservoir-based equalization can achieve low error rates both by addressing impairments before the receiver and by enhancing the receiver's overall behavior.

1.6 Thesis Outline

The remainder of this thesis is organized as follows:

- **Chapter 2** begins by introducing reservoir computing and some key machine learning concepts. This chapter also details the photonic reservoir system used in this thesis, covering both its simulated model and the physical chip implementation.
- **Chapter 3** covers the theory behind the coherent and self-coherent receivers employed in the thesis. It then presents and discusses the corresponding simulation models and the receivers' simulated behaviors.
- **Chapter 4** focuses on the application of the photonic reservoir in systems deploying self-coherent receivers. The KK receiver and SiP SSBI-cancelling receivers are simulated under a range of operational conditions, and equalization of the system impairments is performed using the photonic reservoir.
- **Chapter 5** explores the creation of advanced network architectures by combining the photonic reservoir with electronic feed-forward equalizers. Simulations demonstrate the enhancement of a Kramers-Kronig receiver operating under the limitations of restricted sampling rates and carrier powers.
- **Chapter 6** discusses the experimental results for compensating transmission impairments using the photonic reservoir in systems deploying a coherent receivers .
- **Chapter 7** concludes this thesis, summarizing the key findings and potential future directions.

1.7 Attribution of Work

The work presented in this thesis builds upon previous efforts that resulted in the simulated and physical models of the reservoir used. Namely, the simulation package Photontorch was developed by Floris Laporte [81] and was used as a photonic simulation tool for simulating the reservoirs used in Chapters 4 and 5. The experiments in Chapter 6 were performed using a reservoir chip designed by Stijn Sackesyn [82].

Additionally, contributions from colleagues and collaborators were integral to achieving some of the results presented in this thesis. The simulations pertaining to the SiP SSBI-cancelling receiver in Chapter 4 were part of a Master's thesis I supervised that was conducted by Aimen Zelaci. The experiments in Chapter 6 were performed with essential guidance and expertise from Dr. Stylianos Sygletos of the Aston Institute of Photonic Technologies (AIPT). Efforts of AIPT members that facilitated measurements within their labs are also acknowledged.

1.8 List of Publications

This dissertation has led to the following publications.

1.8.1 International Journals

1. **S. Masaad**, S. Sackesyn, S. Sygletos, P. Bienstman, "Experimental Demonstration of 4-Port Photonic Reservoir Computing for Equalization of 4 and 16 QAM Signals,"*Journal of Lightwave Technology,* (2024)

2. Aimen Zelaci, **S. Masaad**, P. Bienstman, "Reservoir Computing for Equalization in a Self-Coherent Receiver Scheme," *Submitted: Optics Express* (2024)

3. **S. Masaad**, P. Bienstman, "Opto-Electronic Machine Learning Network for Kramers-Kronig Receiver Linearization", *Optics Express* 32, pp 23561-23574 (2024)

4. S. Abreu, I. Boikov, M. Goldmann, T. Jonuzi, A. Lupo, **S. Masaad**, L. Nguyen, E. Picco, G. Pourcel, A. Skalli, L. Talandier, B. Vettelschoss, E.A. Vlieg, A. Argyris, P. Bienstman, D. Brunner, J. Dambre, L. Daudet, J.D. Domenech, I. Fischer, F. Horst, S. Massar, C.R. Mirasso, B.J. Offrein, A. Rossi, M.C. Soriano, S. Sygletos, S.K. Turitsyn, "A Photonics Perspective on Computing with Physical Substrates," *Reviews in Physics*, vol.12 (2024)

5. **S. Masaad**, E.J.C. Gooskens, S. Sackesyn, J. Dambre, P. Bienstman, "Photonic Reservoir Computing for Nonlinear equalization of 64-QAM Signals with a Kramers-Kronig Receiver," *Nanophotonics*, vol. 12, no. 5, pp. 925-935 (2022)

1.8.2 International Conferences

1. **S. Masaad**, S. Sackesyn, S. Sygletos, P. Bienstman, "Experimental Demonstration of 4-Port Photonic Reservoir Computing for Equalization of 4 and 16 QAM Signals," *European Conference on Optical Communication (ECOC),* Germany (Sept. 2024)

2. **S. Masaad**, P. Bienstman, "Photonic Reservoir Computing for Kramers-Kronig Receiver Linearization," *Telecommunications, Optics, and Photonics Conference (TOP),* United Kingdom (Feb. 2024)

3. P. Bienstman, A. Lugnan, S. Aggarwal, F. Bruckerhoff-Pluckelmann, W. H. P. Pernice, H. Bhaskaran, C. Ma, S. Sackesyn, E.J.C. Gooskens, **S. Masaad**, M. Gouda, R. De Prins, "Optical Computing in Silicon Photonics: Self-Adapting Ring Networks and Quantum Recurrent Neural Networks," *Natural and Physical Computing*, Germany (Oct. 2023)

4. Tigers Jonuzi, **S. Masaad**, Alessandro Lupo, J. David Domenech Gomez, P. Bienstman, Serge Massar, "Numerical Analysis of a Self-Calibrating Time-Spatial Interleaving Photonic Convolutional Accelerator," *Photonics in Switching and Computing (PSC)*, Italy (Nov. 2023)

5. **S. Masaad**, E.J.C. Gooskens, M. Gouda, R. De Prins, F. Marchesin, R. Shi, A. Lugnan, S. Sackesyn, C. Ma, Joni Dambre, P. Bienstman, "Integrated Photonic Reservoir Computing for Telecommunication Applications," *Neuromorphic photonics and applications* (**invited**), Greece, (July 2023)

6. E.J.C. Gooskens, S. Sackesyn, **S. Masaad**, J. Dambre, P. Bienstman, "Photonic Reservoir Computing for Wavelength Multiplexed Nonlinear Fiber Distortion Mitigation," *IEEE SiPhotonics*, United States, (April 2023)

7. P. Bienstman, A. Lugnan, C. Ma, S. Sackesyn, E.J.C. Gooskens, **S. Masaad**, M. Gouda, R. De Prins, "Coherent optical computing in silicon photonics," *Coherent Network Computing (CNC)*, United States (2022)

8. **S. Masaad**, E.J.C. Gooskens, S. Sackesyn, J. Dambre, P. Bienstman, Photonic Reservoir Computing for Nonlinear Equalization of 64-QAM Signals with a Kramers-Kronig Receiver, *European Conference on Optical Communication (ECOC),* Switzerland, (Sept. 2022)

9. S. Sackesyn, **S. Masaad**, C. Ma, P. Bienstman, "Reservoir Computing with Optical Weights," *European Conference on Optical Communication (ECOC)* (**invited**), Switzerland, (Sept. 2022)

10. G. von Hunefeld, G. Ronninger, P. Safari, I. Sackey, R. Thomas, P. Cegielski, S. Suckow, E. Seker, D. Stahl, **S. Masaad**, E.J.C. Gooskens, P. Bienstman, C. Schubert, J. Fischer, R. Freund, "Enabling optical modulation format identification using an integrated photonic reservoir and a digital multiclass classifier," *European Conference on Optical Communication (ECOC*, Switzerland, (Sept. 2022)

11. D. Chatzitheocharis, T. Chrysostomidis, I. Roumpos, M. Kohli , U. Koch , J. Leuthold , F. Eltes , S. Sackesyn, **S. Masaad**, P. Bienstman, T. Buriakova, M. Zervas, C. Caillaud, H. Bertin, C. Besançon, D. K. Sharma, A. Bouhelier, V. Grimaldi, C. De Vita, F. Zanetto, A. Ivan Martinez Rojas, G. Ferrari, F. Morichetti, M. Sampietro, E. Kyriazi, D. Apostolopoulos, H. Avramopoulos, D. Petousi, B. Wohlfeil, S. Alnairat, N. Argyris, P. Bakopoulos13, B. Atias, E. Mentovich, J. Declercq, J. Lambrecht, J. Van Kerrebrouck, X. Yin, J. Van Campenhout, B. Offrein, F. Hermann and K. Vyrsokinos, "NEBULA project: combining BTO based plasmonic modulators with neuromorphic augmented receivers for higher than 100 Gbaud intra- and inter-data center interconnects," *19th International Conference on Nanosciences & Nanotechnologies*, Greece (July 2022)

12. P. Bienstman, J. Dambre, A. Lugnan, S. Sackesyn, C. Ma, E.J.C. Gooskens, M. Gouda, **S. Masaad**, "Photonic Neuromorphic Computing Using Silicon Chips", *Huawei STW* (**invited**), (2021).

13. A. Lugnan, S. Sackesyn, C. Ma, E.J.C. Gooskens, M. Gouda, **S. Masaad**, J. Dambre, P. Bienstman, "Reservoir computing for high-speed photonic information processing," *Photonics in Switching and Computing (PSC)* (**invited**), (2021)

14. A. Lugnan, S. Sackesyn, C. Ma, E.J.C. Gooskens, M. Gouda, **S. Masaad**, Joni Dambre, P. Bienstman, "Photonic reservoir computing for high-speed neuromorphic computing applications", *2021 IEEE Summer Topicals Meeting Series* (**invited**), Mexico, (2021)

15. P. Bienstman, J. Dambre, A. Lugnan, S. Sackesyn, C. Ma, E.J.C. Gooskens, M. Gouda, **S. Masaad**, "Silicon photonics for brain-inspired neuromorphic information processing," *1st Workshop on Neuromorphic Photonics* (**invited**), (2020)

1.8.3 National Conferences

1. P. Bienstman, J. Dambre, A. Lugnan, S. Sackesyn, C. Ma, E.J.C. Gooskens, M. Gouda, **S. Masaad**, "Photonic Neuromorphic Computing Using Silicon Chips", *BePOM* (**invited**), (2021)

1.8.4 Awards

1. Best Poster Award for: S. Masaad, P. Bienstman, "Photonic Reservoir Computing for Kramers-Kronig Receiver Linearization," *Telecommunications, Optics, and Photonics Conference (TOP),* United Kingdom (Feb. 2024)

2

Reservoir Computing

2.1 Overview

Machine learning (ML) is a field of computer science that gives computers the ability to learn without being explicitly programmed. This capability has revolutionized industries and applications, from medical diagnosis to autonomous vehicles, by enabling systems to identify patterns in data and make predictions or decisions. Depending on whether this data is labeled or unlabeled, ML models fall into two primary categories: supervised and unsupervised learning. Given the diverse

spectrum of ML problems, a multitude of models can be used, including clustering, regression, kernel methods, and neural networks (NNs) [83].

NNs represent a prominent class of ML models renowned for their ability to extract obscured patterns and to approximate complex functions, allowing them to excel at tasks ranging from image recognition to natural language processing. Inspired by biological neural systems, these networks comprise interconnected nodes organized in various topologies suited to different tasks [84]. Figure 2.1 illustrates two common architectures: feedforward neural networks (FNNs), primarily used for tasks without sequential dependencies, and recurrent neural networks (RNNs), designed to process sequential data through recurrent connections.

To understand the fundamental operation of NNs, we exemplify the flow of information in an FNN since it has a straightforward structure. Data traverses the network layer by layer. At each node, input values are multiplied by corresponding weights, summed, and then processed through a nonlinear transformation. The result is then broadcast to the next layer with different weights for different connections. Figure 2.1 also shows the input/output connections of a particular node in the FNN, where $w_{l,i}$ is the i^{th} weight in the l^{th} layer and $x_{l,j}[n]$ is the j^{th} signal in the l^{th} layer. A biasing constant $b_{l,k}$ may also be added at each node to shift the transformation function, introducing a degree of flexibility in the network's response.

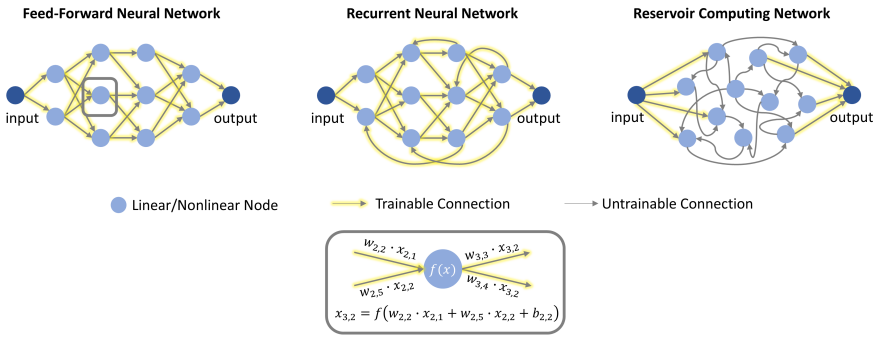


Figure 2.1: Topologies of Feed Forward Neural Networks (FNN), Recurrent Neural Networks (RNN), and Reservoir Computing (RC) Networks. The input/output connections of a node in the FNN network are shown to exemplify the flow of information in a network. The first subscript in our notation identifies the layer, while the second indexes the specific signal or parameter within that layer, such that $w_{l,i}$ is the i^{th} weight, $x_{l,j}$ is the j^{th} signal and $b_{l,k}$ is the k^{th} bias in the l^{th} layer. Figure based on [85] and [86]

Training NNs involves optimizing these weights and biases to minimize a loss function that quantifies the model's performance, i.e. the discrepancy from the ground truth. Given the number and depth of these parameters, the training task is

challenging and often requires iterative optimization methods that utilize substantial computational resources and large datasets to converge. Moreover, recurrent connections in RNNs introduce the challenges of dynamical systems, increasing training difficulty [87].

To simplify the training process, Reservoir Computing (RC), shown in Figure 2.1, offers an alternative approach by fixing the internal network connections (the reservoir) and only training the output layer, which is known as the readout [53, 88]. The reservoir works as an internal projection of the input(s), generating a sufficiently rich variety of features that the readout layer can be trained on. While some implementations also train input connections, RC still significantly reduces the number of trainable parameters compared to traditional NNs. Furthermore, RC often involves convex optimization, thus avoiding computationally intensive iterative training methods. Despite its apparent simplicity, this approach has demonstrated success in various applications, including time series forecasting [89] and classification/regression tasks [90].

RC has also emerged as a promising approach for machine learning in physical neuromorphic computing. This is driven by its ability to efficiently construct networks in environments where manipulating internal weights is challenging. Furthermore, the reservoir's capacity to leverage diverse physical phenomena in their natural form has expanded its application scope. As a result, RC has been explored across a wide range of substrates, including electronics, spintronics, biological systems, and many more [90, 91]. Amongst these, photonic implementations demonstrate particular promise due to their potential for high speed and low power consumption [69].

The remainder of this chapter provides an overview of prominent photonic reservoir computing architectures from the literature to establish a relevant context. Then, the four-port reservoir employed in this thesis is discussed, covering both the physical system and its corresponding simulation model. Finally, the training algorithms deployed in this thesis are introduced.

2.2 Photonic Reservoir Computing

Photonic reservoir computing leverages physical optical systems to implement the reservoir, which allows using the inherent parallelism and high speed of optics. Various optical components and architectures can be used to realize reservoirs, including free space, discrete optics, and integrated platforms. Common partitioning of these photonic RC systems separates them as time-delayed reservoirs or space-multiplexed reservoirs [69].

Time-delayed RC typically uses a single nonlinear node with delayed feedback as a means of minimizing the expected hardware complexity in photonic systems [92]. This approach divides the delay loop duration (τ) into N intervals, effectively creating N virtual reservoir nodes. Thus the node is driven both by the fed input and the delayed feedback from the node's output. To enhance system dynamics, a random mask is applied to the input, transforming every input sample into piecewise constant signals. Different implementations have been realized using various masks and nonlinear nodes, including semiconductor optical amplifiers (SOAs) [93], lasers [94, 95], or fiber cavities [96].

In contrast, space-multiplexed RC leverages different parts of the photonic system to correspond to the different nodes in a network. This can be realized using various components including diffractive optical elements [97], cameras [98], scattering media [99], or integrated optical networks with e.g., SOA nodes [100].

SOAs are interesting for photonic RC and have been used in both time-delayed and spatial multiplexed systems because they exhibit a power-dependant nonlinear behaviour. This behaviour closely resembles the (positive) tanh function commonly used in traditional NN nodes. Indeed, the network of SOA nodes in [100] showed comparable performance to standard tanh networks in tasks like speech recognition [101]. However, this was only shown in simulations because of the high power consumption in SOAs, which hindered practical implementation. Instead, the first hardware implementation ridded the network of the nonlinear node transformations, and replaced them with linear multi-mode interferometers (MMIs). This exchange made the network fully linear, however nonlinearity was introduced at the output since the state at each reservoir node was read out through a photodiode, converting complex-valued amplitudes to real-valued intensities [31]. Note that nonlinearity is essential for a network, as it allows solving nonlinear problems and approximating complex functions. While this is traditionally integrated within the network at every node, decoupling the recurrent connections from the nonlinear transformation was shown to perform well on many nonlinear tasks in this linear photonic network [31]. Currently, the same concept is gaining momentum and is explored within software-based NN and RC architectures, demonstrating the potential to rival or surpass the performance of fully nonlinear systems [102–104].

Since then, several enhancements were made to the architecture introduced in [31]. Notably, the input scheme was refined by distributing the input signal over multiple nodes to optimize power distribution within the network. Modifications to node connectivity were also introduced, such that every node had the same number of inputs and outputs in order to avoid radiation losses. This resulted in the four-port architecture employed in this thesis [82]. In the next section, the physical realization and simulation model of this network are presented in more detail.

2.3 Four-Port Architecture

The four-port architecture employs 3x3 MMIs as nodes that are interconnected by spiral waveguides as seen in Figure 2.2. Each node has two input ports connected to other nodes and two output ports leading to subsequent nodes. The third input port allows external signal injection, but can also be left unused. The third output port connects to the readout where controllable weights can be implemented. At each node, input signals interfere coherently to produce different output signals. This reservoir's architecture is passive, as it does not require external power to drive the nodes or connections. Furthermore, non-volatile optical weights can also be integrated on the chip for the readout. While these weights require initial power during training, once set they hardly consume any electrical power.

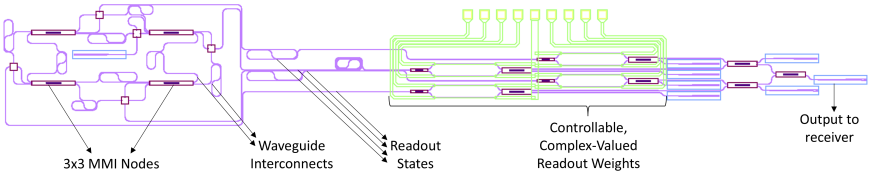


Figure 2.2: Circuit layout of four-port architecture employing 3x3 MMIs as nodes interconnected by spiral waveguides. An integrated readout with on-chip complex-valued weights is also shown

To clarify the system's operation, we recall that the photonic reservoir processes data (or bits) modulated on an optical carrier. This means that there are two relevant timescales to be observed: the rapid oscillations of the optical carrier itself, measured in THz, and the slower rate at which the data is carried in the envelope, typically in the GHz range. These timescales roughly correspond to the group and phase velocity characteristics of the wave, leading to two types of interactions within the reservoir.

On the one hand, effective data processing, particularly for memory-intensive tasks, necessitates the interaction of data from different time instances. For example, calculating a weighted sum of the current and previous bits requires a recurrent connection that delays the signal sufficiently before it is fed back to the MMI. This is done by controlling the length l of the waveguides linking the different MMI nodes, which results in a group delay Δt as per the following equation [105]

$$\Delta t = \frac{l}{c}\, n_g(\lambda), \tag{2.1}$$

where c is the speed of light and $n_g(\lambda)$ is the wavelength-dependent group index. Given that the data rate is in the GHz range, standard bit periods are fractions of nanoseconds. As such, Δt is typically in that range.

On the other hand, the phase of the optical carrier controls the extent to which an interference at an MMI is destructive or constructive. This phase change due to prorogation is governed by the following equation [22]

$$\phi = \frac{2\pi}{\lambda}\, l\, n_{\text{eff}}(\lambda) \tag{2.2}$$

where $n_{\text{eff}}(\lambda)$ is the wavelength-dependent effective index.

To illustrate the effect of these two quantities, the top panels of Figure 2.3 show a message signal with 5 bits modulating an optical carrier. For clarity, the envelope of the carrier is indicated in a dashed black line showing the bit stream 1 1 0 1 0. If this signal propagates in a short waveguide, a group delay is induced which is seen in the second panels of Figure 2.3. The delay in the first 2 columns corresponds to half the bit period, whereas the last column shows a delay of a full bit period. The interference of the signals and their time-delayed copies are shown in the last panel. Here, the relative phases of the two carriers control the resulting output. As seen in the first two columns, despite the envelope having the same delay, the phase shift of the carrier creates different outputs.

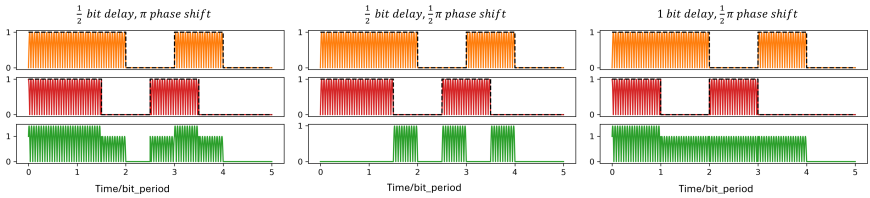


Figure 2.3: The impact of delays and carrier phases on signal interactions in the reservoir is illustrated. The top row shows a modulated optical carrier with its corresponding data envelope. The middle row depicts the signal after experiencing different time delays in a waveguide. The bottom row demonstrates the interference between the original and delayed signals, showing the impact of both data-induced delays and carrier phase differences on the resulting output.

As mentioned earlier, enhancing the system's memory and facilitating signal interactions is achieved by controlling the path length light traverses in the interconnecting waveguides, which introduces the nanosecond time delays seen above. Rather than specifying absolute waveguide lengths, we express these as a ratio of induced delay to the the bit duration. As such, higher data rates necessitate shorter waveguides for equivalent effective delays. On the other hand, as will be discussed in the next subsection, the carrier phase shifts are a random parameter, and therefore result in the random weight projections of the reservoir. Note that the example shown in Figure 2.3 is lossless to emphasize the effect of the delay and phase shifts, but in practice insertion losses from the MMIs, as well as splitting and propagation losses in the waveguides are also incurred.

These are the fundamental properties of the reservoir used in this thesis. In simulations, several aspects of this architecture can easily be adjusted and are often varied in different tasks to investigate their influence. This includes the input strategy, the length of the connections and the connection topology. In the physical system, however, these aspects are fixed, given that the chip was designed and fabricated in advance. Details of the physical system and simulation model are further discussed in the following subsections.

2.3.1 Physical System

The four-port architecture employed in this thesis was designed by Stijn Sackesyn and fabricated by Ligentec on a silicon nitride (SiN) platform [106]. As the chip remained unpackaged, grating couplers at the input and output of the chip facilitated coupling of light from the vertically supported optical fibers to the planar chip and vice versa. While grating couplers are less sensitive to misalignment of the optical fiber, they incur significantly higher insertion losses (IL) compared to edge couplers. This is shown in Figure 2.4, where the the measured IL of a straight test structure using either edge or grating couplers is plotted as a function of the optical wavelength.

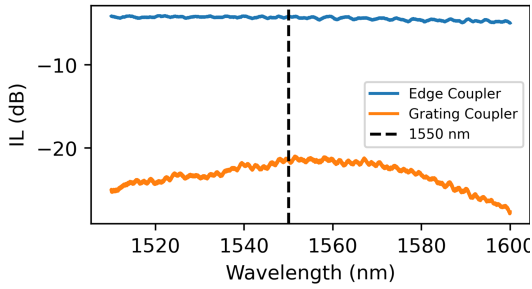


Figure 2.4: Insertion loss (IL) in dB for two coupling mechanisms. The loss at 1550 nm is indicated with a dashed line

The MMIs within the network were designed to produce symmetric output distributions. This design choice was done in contrast with the preceding generation of chips, though its superiority over asymmetric configurations remains uncertain. Nevertheless, the intended symmetric behavior was achieved and confirmed as depicted in Figure 2.5. The observed wavelength-dependent variations in the MMI output are due to the grating couplers.

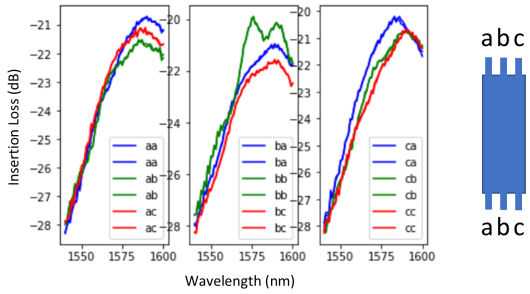


Figure 2.5: Power distribution of the 3x3 MMI. The notation in the legend indicates which port was used as an input (first letter) and which port was used as an output (second letter). Due to the symmetric nature of the MMI, reversing the input and output results in the same behaviour. Figure from Stijn Sackesyn's measurements in [107].

Waveguides serve as interconnects between MMIs, providing the necessary delay for the required memory. The SiN platform's high index contrast enables highly compact waveguide structures, measuring approximately $1\,\mu\mathrm{m}$ in width and $0.8\,\mu\mathrm{m}$ in thickness. The length is determined by the data rate to be processed and the delay ratio intended. This chip was designed for 28 Gbaud signals and a 0.5 delay ratio for most connections, making the waveguides about $1\ \mathrm{mm}$ long. The platform's high index contrast also allows for tight bend radii, resulting in a total chip footprint of less than $1.5\ \mathrm{cm}^2$ for a 16-node configuration.

However, this high index contrast makes n_{eff} highly sensitive even to nanometer-scale fluctuations in waveguide width [76]. Given the inherent nanometer-scale roughness of waveguide sidewalls, as depicted in Figure 2.6 (where air cladding is shown for clarity), n_{eff} exhibits significant variability. This roughness-induced fluctuation in n_{eff} directly translates to phase variations along the waveguide. For approximately 1 mm-long waveguides, these accumulated phase differences easily exceed 2π [108]. Consequently, uncontrollable phase shifts arise, leading to the aforementioned random MMI outputs. Furthermore, due to fabrication imperfections, each reservoir possesses a unique, yet fixed, set of internal parameters.

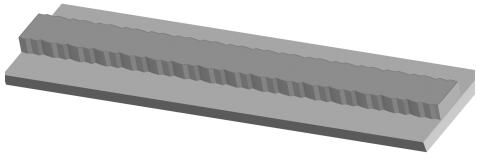


Figure 2.6: Integrated waveguide with air cladding showing sidewall roughness

For the readout, optical weights are commonly implemented using controllable phase shifters and Mach-Zender Interferometers (MZIs). Through shifting the phase of one signal with respect to the other, constructive or destructive summation occurs, which allows controlling the amplitude of the signal. Another phase shifter can also be used to control the resulting signal's phase, thus allowing the implementation of complex-valued weights. This controllable optical phase shift can be implemented in several ways, including tungsten heaters, micro-electro mechanical systems (MEMS), or phase-change materials like barium-titanate (BTO) [109].

It is possible, however, to use the reservoir architecture with or without the integrated optical weights. This is depicted in Figure 2.7, which shows two methods for implementing the readout. In the parallel readout, shown in brown, the readout signals from the reservoir are directed to the integrated and optically implemented weights, followed by an integrated summation tree. Finally, the summation tree output is led to a receiver.

If a parallel readout cannot be used, a serial readout, shown in purple, is an alternative for proof-of-concept lab experiments. In this scheme, each readout channel is measured separately and sequentially and is then stored offline. Through offline processing, appropriate weights can be computed and then tested on the collected data. However, a serial readout remains different from a delay-based reservoir, since sampling is done over multiple spatially separate nodes as opposed to the single node used in delay-based reservoirs.

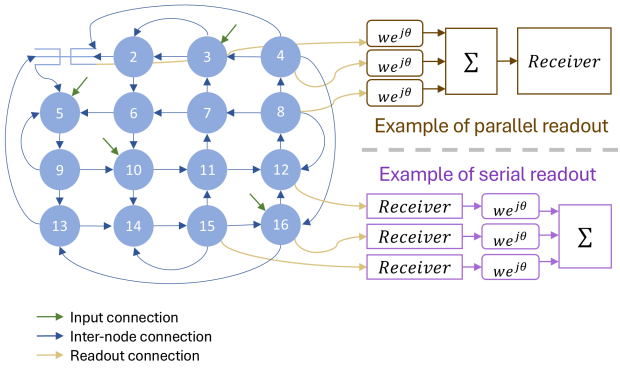


Figure 2.7: 16-node reservoir with 2 possible readout schemes. Nodes are shown as numbered vertices, with the exception of node 1, which is shown as an abstracted 3x3 MMI. Input, internode, and readout connections are shown as colored arrows. A parallel readout scheme is shown in red which includes on-chip weights and one detected signal. A serial readout scheme is shown in blue, where electronic off-chip weights are implemented after individual detection of readout states

Note that the entire reservoir would follow a single readout scheme, however they are both depicted on the same figure for conciseness. Furthermore, various implementations for receivers can be used, including photodiodes, coherent receivers, or self-coherent receivers. Naturally, this choice influences the network's behaviour as the output transformation varies.

Serial readout is generally undesirable due to the increased processing complexity (as will be detailed in Chapter 6) and its incompatibility with real-time reservoir operation. However, it can serve as a temporary workaround to mitigate excessive optical losses, as is the case with the particular generation of the four-port reservoir used this thesis. Here, significant optical losses primarily attributed to grating couplers were incurred resulting in an IL that ranged between -35 to -50 dB for individual nodes of a 16-node reservoir. These cumulative losses precluded the use of a parallel readout, as additional combining losses within the integrated readout would have placed the detected signal's power below the receiver's noise floor.

A subsequent chip, characterized during this thesis but not employed in experiments, addresses the coupling losses by using edge couplers instead of grating couplers and as such has much better power characteristics. For comparison, the loss for each node for both chip generations is shown in Figure 2.8.

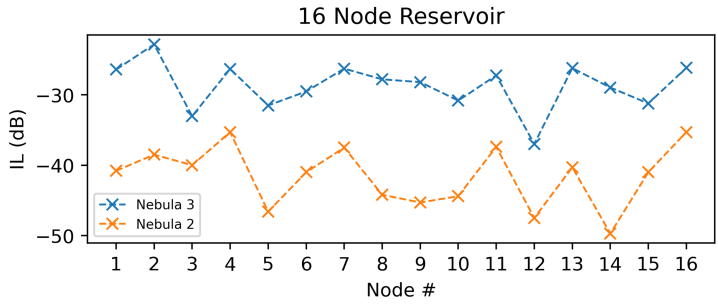


Figure 2.8: Node losses for two generations of chips. Nebula 2, experimentally used in this thesis, incurred higher losses due to grating coupling. The newer Nebula 3 chip circumvents this by using edge couplers and as such has lower losses per node.

2.3.2 Simulation Model

The reservoir simulation model is based on the Photontorch Python library developed by Floris Laporte [81, 110]. This library enables the extrapolation of circuit-level behavior from measured or simulated single-component properties.

It utilises the scattering matrix (S-matrix) formalism, which accurately relates the input and output fields of linear components. Complex-valued optical state vectors are employed to represent the phase and amplitude of the optical fields at each component port. By defining S-matrices for individual components and their interconnections, a network is formed. The network's input ports can then be specified, and simulations are conducted in either time or frequency domains.

To model the four-port architecture, S-matrices capturing MMI splitting ratios and insertion losses were incorporated. Additionally, waveguide losses of 25 dB/m and waveguide delays were modeled. Random phases and slight waveguide length variations were introduced to simulate fabrication imperfections. However, dispersion is not accounted for in the Photontorch simulator. While this may be prominent for certain circuits, including ring resonators, it is insignificant for the non-resonant reservoir system. This can be shown through simulating the dispersive effect of a waveguide. Depending on their geometry, SiN waveguides may induce dispersion as high as $-150\ \mathrm{ps/nm \cdot km}$. While this estimate is on the higher end, we can use this value to model a dispersive element and show the expected pulse broadening. This is seen in Figure 2.9, where a raised cosine pulse with a roll-off factor of 0.2 and a symbol rate of 28 Gbaud is shown in blue and the dispersed output of a waveguide is shown in orange. Various (and unrealistically long) waveguide lengths are modelled. Lengths up to 100 m show negligible broadening, and it is only at the 1 km mark that pulse broadening is evident. Since the total accumulative length of waveguides on-chip is barely a few centimeters, these results show that a photonic simulator that does not model dispersion can still accurately capture the behaviour of this system (at least in the absence of strong resonances).

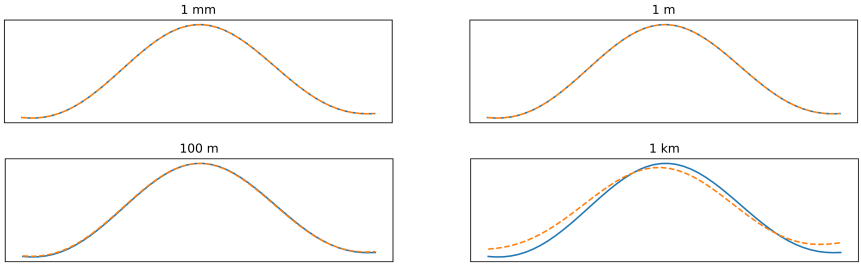


Figure 2.9: Simulation of dispersion effects in various (and unrealistically long) SiN waveguides. A raised cosine pulse with a roll-off factor of 0.2 and a symbol rate of 28 Gbauds (blue) is propagated through a waveguide and the dispersed output pulse (orange) is shown. As seen, negligible pulse broadening is incurred in waveguides as long as 100 meters

2.4 Training

To train the reservoir, readout weights are optimized to minimize the discrepancy between the generated output and a target signal. This optimization process is quantified through a loss function, which serves as a performance metric for the network's accuracy in approximating the ground truth. This section discusses loss functions frequently used in the thesis and algorithms for minimizing them.

Note that software-based and hardware-based training algorithms can vary. Software-based methods leverage the flexibility of simulation environments that offer full state observability and component modelling. This enables implementing complex algorithms using techniques like gradient-based optimization. In contrast, hardware-based methods relying on direct physical implementations are constrained by issues such as limited observability and real-time processing. This necessitates using alternative, and often less compute-intensive, algorithms (e.g, [111–116]). Given the focus of this thesis on offline processing of serial readout data and the use of a simulated model, the presented algorithms are tailored for software-based training. Thus, hardware-based training is outside the scope of this work.

2.4.1 Loss functions

Bit Error Rate

Bit Error Rate (BER) is a critical metric in digital communication systems where data is transmitted as discrete symbols. In the context of QAM modulation, each symbol represents multiple bits, with an M-QAM constellation employing M symbols to carry $log_2(M)$ bits. To determine bits at the end of the link, the received analog signal is sampled and compared to decision regions corresponding to each QAM symbol. If a sample falls within a decision region of a symbol, it is subsequently mapped to that symbol's corresponding bit pattern based on the QAM constellation. By comparing this reconstructed bit pattern with the original transmitted bit pattern, erroneous bits can be identified. BER is then calculated as the ratio of the erroneously detected bits to the total number of transmitted bits.

Employing Gray coding, wherein adjacent symbols differ by only one bit, can mitigate the BER in the event of transmission errors. A visual representation of a Gray-coded 16-QAM constellation with color-coded decision regions and corresponding bit patterns is shown in Figure 2.10. Note that QAM is a complex modulation format that utilizes both the amplitude and phase of a carrier to encode data. As such, the constellation diagram captures this through representing the imaginary component (Q) on the y-axis and the real component (I) on the x-axis.

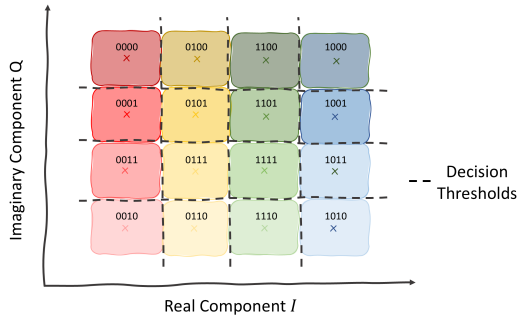


Figure 2.10: Gray-coded 16-QAM constellation with color-coded decision regions and respective bit patterns

While BER is a standard performance measure in communication systems, its application in machine learning training is limited due to its non-continuous nature. This lack of continuity renders BER non-differentiable, which is a crucial property for gradient-based optimization algorithms commonly employed in machine learning, as will be discussed later. Additionally, small parameter updates do not often result in immediate BER changes, leading to flat regions in the optimization landscape. This sparsity in feedback makes it difficult for the optimizer to guide improvements.

Mean Square Error

Mean Squared Error (MSE) is a commonly employed loss function that quantifies the average squared difference between the predicted and actual values. Mathematically this is expressed as

$$MSE = \frac{1}{N}\sum_{n=1}^{N} |y_n - \hat{y}_n|^2 \tag{2.3}$$

where N is the total number of data points, y_i denotes the true value, $\hat{y}_n$ represents the predicted value. The absolute value is an adjustment from the standard MSE to allow capturing the nature of the complex-valued data used. When substituting $I_n + jQ_n$ for y_n, the MSE equation equates to

$$MSE = \frac{1}{N}\sum_{n=1}^{N} |\Delta I_n + j\Delta Q_n|^2 = \frac{1}{N}\sum_{n=1}^{N} \left(\Delta I_n^2 + \Delta Q_n^2\right) \tag{2.4}$$

where ΔI_n and ΔQ_n are a sample's deviation from the true symbol on the real and imaginary axes, respectively. This also equals the square of the Euclidean distance.

MSE's widespread adoption as a loss function is attributed to several advantages. It is computationally efficient, effectively penalizes larger errors, and is differentiable when split up into real and imaginary parts, enabling its use with gradient-based optimization algorithms. These properties make MSE a versatile choice across various domains. When applied to communication systems, MSE can be computed on either individual samples or the continuous analog signal. In the former case, the error is calculated between corresponding symbol values, while in the latter, a continuous target signal derived from the symbols is used.

Higher-Order Error

Despite its advantages, minimizing MSE does not always correlate with a reduction in BER, since prioritizing overall MSE reduction can potentially introduce erroneous bits. Furthermore, unlike BER, which focuses exclusively on errors, MSE penalizes all data points even when they are not erroneous. This may hinder its effectiveness in discriminating between correct and erroneous symbols.

These issues are exacerbated when a network is already trained but requires fine-tuning, and in these cases erroneous bits tend to cluster near decision boundaries. Note that the decision regions for each symbol in Figure 2.10 are typically 2 units wide on each axes, and deviations are measured from the center of the region. When fine-tuning a model, erroneous bits will typically lie close to the decision thresholds with values ΔI and/or ΔQ just over 1. In these cases, MSE lacks a clear distinction between erroneous and correct bits due to the gradual nature of the squared error function at these boundaries. This behaviour is illustrated in Figure 2.11 which shows the computed error for the MSE function on a heat map. The x and y-axes capture ΔI and ΔQ, respectively.

To address these limitations, we modified the MSE function by replacing the squared term with a higher exponent such that

$$\text{HOE} = \frac{1}{N}\sum_{n=1}^{N} \Delta I_n^H + \Delta Q_n^H \tag{2.5}$$

where H is an even integer > 2. This adjustment renders the function more sensitive to errors while maintaining neutrality towards correct samples and preserving differentiability. This can be seen in Figure 2.11 where the heat map shows low error values when ΔI and ΔQ are low, but steeply increasing error values as ΔI and ΔQ approach 1.

This model should be used with care, as it can provide very high error values for outliers or strongly distorted signals. As such, it is typically used on problems that require fine-tuning and attention is paid to discard the contribution of outliers.

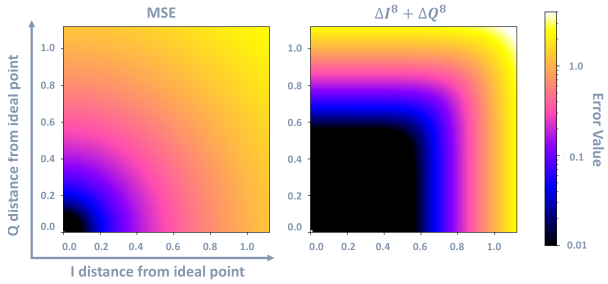


Figure 2.11: Heatmap illustrating the MSE and the Higher-Order Loss function as a function of ΔI and ΔQ, representing deviations from ideal symbol centers. The heatmap shows ability of the High-Order Error function to sharply penalize bit errors, while maintaining a relatively neutral response to small deviations. This addresses the limitations of MSE in differentiating between correct and erroneous symbols near decision boundaries

The selection of an appropriate loss function is crucial for effective training. The characteristics of the loss function, such as differentiability, sensitivity to outliers, and computational efficiency, also influence the choice of optimization algorithm. In the subsequent sections, we will explore a range of training algorithms and discuss their suitability for different loss functions and problem domains.

2.4.2 Linear Regression

In a typical reservoir computing system, the network generates a sequence of readout states, denoted as X. These states are linearly combined using a readout weight matrix, W_{readout}, to produce the output signal, $y[n]$, according to the equation:

$$y = W_{\text{readout}} \cdot X \tag{2.6}$$

To establish the optimal mapping between reservoir states and the desired output, linear regression can be employed. This involves finding the hyperplane that best fits the relationship between the reservoir states, X, and the target signal, $\hat{y}$, minimizing the error between the predicted and actual outputs. A closed-form solution for the readout weights, W, can be derived using the ordinary least squares method, incorporating regularization to prevent overfitting:

$$W = (X^H X + \alpha I)^{-1} X^H \hat{y} \tag{2.7}$$

where W is the weight matrix, α is a regularization parameter to limit overfitting, I is the identity matrix. The superscript H refers to the conjugate-inverse of the matrix, which is needed for complex-valued data.

Linear regression is an appealing choice for initial modeling due to its computational efficiency and interpretability. However, this method assumes a linear relationship between the states and the target signal, which may lead to suboptimal model performance when the actual relationship is nonlinear. Moreover, in the presence of a nonlinear output transformation (as is the case when the receiver in the readout is nonlinear), the target signal exists after the receiver and thus must pass through an inversion of the receiver transformation to generate the required $\hat{y}$. This inversion is non-exact as it often involves a noisy receiver and can even be multi-valued (e.g., the inverse of a photodetected signal is non-deterministic since it involves inverting the squared magnitude of a complex-valued signal). Furthermore, depending on the nonlinear transformation of the receiver, optimizing weights based on an inverted target does not necessarily lead to an optimal solution for the post-receiver signal. As such, other training methods may be better-suited.

2.4.3 Backpropagation and Gradient Descent

Backpropagation is a cornerstone algorithm in training neural networks, as it provides a method for adjusting parameters in deep and/or nonlinear networks [117]. This process happens in two stages: the forward pass and the backward pass. During the forward pass, input data propagates through the network, generating an output. Then, a backward pass is performed by calculating the loss, taking into account the difference between the predicted and actual outputs. The gradient of this loss is then computed and propagated backward through the network, layer by layer, using the chain rule. This process yields gradients for all network parameters. Note that due to the gradient computations, the loss function employed must be differentiable.

After the gradients are computed, a method for optimizing the weights of the network is needed. This is achieved through gradient descent, an optimization algorithm that iteratively moves the network parameters in the direction of the negative gradient of the loss function [118]. The update rule for one weight is then:

$$w_{t+1} = w_t + \eta \nabla w \tag{2.8}$$

where w_t is the current value of the weight, w_{t+1} the new one, ∇w is the gradient, and η is the learning rate that controls the step size of the changes. This process is repeated for all the network's weights and biases over several iterations. This helps the algorithm approach minimum values in the loss function as seen in Figure 2.12.

In the context of reservoir computing, gradient descent can be used to optimize the weights of the readout layer when the readout layer involves a nonlinear transformation. However, one of the main disadvantages of gradient descent is that it is prone to getting stuck in local minima.

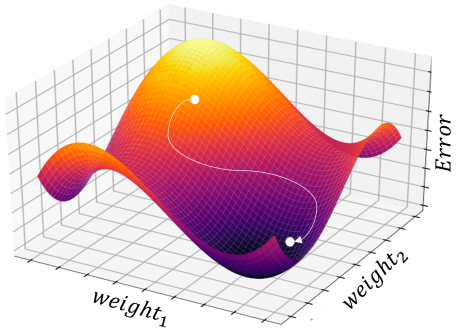


Figure 2.12: Gradient descent algorithm tracking local minima in a loss function with two weights. Figure adapted from [119]

2.4.4 CMA-ES

Covariance Matrix Adaptation Evolution Strategy (CMA-ES) is an optimization algorithm inspired by the principles of natural evolution [120]. It is particularly effective for non-linear, non-convex optimization problems with continuous variables, making it well-suited for optimizing communication system losses based on the BER loss function.

CMA-ES generates a population of candidate weights and tests them to generate an error value. It does this by sampling candidate solutions from a multivariate normal distribution. The mean and covariance matrix of this distribution are iteratively updated based on the fitness of the sampled solutions, guiding the search towards promising regions of the solution space.

Unlike backpropagation, CMA-ES does not require gradient information, making it robust to noisy and discontinuous loss functions. This property is advantageous in scenarios where the loss function is complex or difficult to differentiate. In the context of reservoir computing, CMA-ES can be used to optimize the weights of the readout layer when it involves a nonlinear transformation or when the loss function has multiple local minima. By exploring the solution space more globally, CMA-ES can potentially find better solutions than gradient-based methods.

As previously discussed, receivers at the readout can be implemented in multiple ways. This choice impacts not only the transmission system's performance but also the reservoir's behavior and the effectiveness of different training methods. The following chapter elaborates on the receiver implementations explored in this thesis, including their functionalities, system performance, and the potential for nonlinear transformations.

3

Coherent and Self-Coherent Receivers

3.1 Overview

Coherent communication systems are indispensable for high-capacity optical networks. However, complexities arise with coherent detection since photodiodes are inherently sensitive to the optical *intensity* rather than the optical *field*. This necessitates engineered receiver architectures that are designed to enable fully coherent or self-coherent operation.

This chapter covers the working principles of coherent receivers in addition to two prominent self-coherent architectures. Furthermore, these receivers are constructed

as simulation models to allow exploration of their system-level trade-offs. For completeness, the chapter begins with a brief introduction on coherent transmitters and some information theory concepts.

3.2 Coherent Transmitters

Coherent transmission exploits both the amplitude and phase of an optical carrier to carry information, such that a modulated optical field is expressed as

$$E_x(t) = A(t)\cos(2\pi f_c t + \phi(t)) \tag{3.1}$$

where $A(t)$ and $\phi(t)$ are the amplitude and phase modulation components, and f_c is the frequency of the optical carrier. This can also be rewritten as [22]

$$E_x(t) = x_I(t)\cos(2\pi f_c t) - x_Q(t)\sin(2\pi f_c t) \tag{3.2}$$

where $x_I(t) = A(t)\cos(\phi(t))$ and $x_Q(t) = A(t)\sin(\phi(t))$ are the in-phase and quadrature components of the signal.

To realize this physically, an IQ modulator is required whose architecture will be discussed in the next subsection. Furthermore, spectrally efficient pulse-shapes to represent $x_I(t)$ and $x_Q(t)$ are introduced in the subsequent subsection.

3.2.1 IQ Modulator

A coherent IQ modulator, as illustrated in Figure 3.1a, consists of nested Mach-Zehnder modulators (MZMs). To analyze its operation, the behaviour of a single MZM is first considered. The input signal is evenly split between the two arms of the MZM and a differential voltage, V_t, is applied to induce a phase difference between the two arms. This causes the interference at the MZM's output to result in varying amplitudes and polarities depending on the induced phase difference. This relationship between the input voltage, V_t, and the output optical field, E_{out} is sinusoidal, and a portion of this is shown in Figure 3.1b. By appropriately selecting the operational points on this transfer function, the desired number of modulation levels can be achieved.

To realize the full IQ behaviour, an optical carrier is split between two such MZMs that are independently driven by electric signals $x_I(t)$ and $x_Q(t)$. The resulting optical outputs are then combined with a $\pi/2$ shift induced on the lower arm, as seen in Figure 3.1a. This achieves modulation like the 16 QAM in Figure 3.1c.

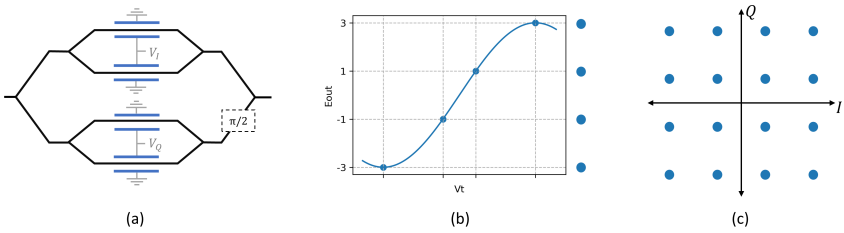


Figure 3.1: Architecture and behaviour of the coherent transmitter. (a) IQ modulator formed by nested MZMs. (b) Portion of an MZM's transfer function driven to generate 4 distinct amplitude/polarity levels. (c) 16 QAM complex constellation generated from combining the outputs of two MZMs as is done in the IQ modulator shown in (a). Based on [22]

This ideal behaviour of the IQ modulator is not straightforward to realize due to several transmitter imperfections. Most importantly, the generated constellations seen in Figure 3.1b and c rely on correct biasing of the MZMs. However, under the influence of drifts in the biasing points, a nonlinear signal distortion will arise due to the sinusoidal form of the transfer function. This is seen in Figure 3.2a, where the ideal constellation points and the constellations resulting from slight drifts in the IQ MZMs are shown. Furthermore, the $\pi/2$ shift in the IQ modulator is also implemented through biasing of a phase shifter. Imperfections in its biasing or drifts will result in a loss of orthogonality in the constellation, as seen in Figure 3.2b. Finally, the non-zero linewidth of the optical carrier creates a phase difference between two samples collected at different points in time. This phase noise translates to constellation rotations as seen in Figure 3.2c.

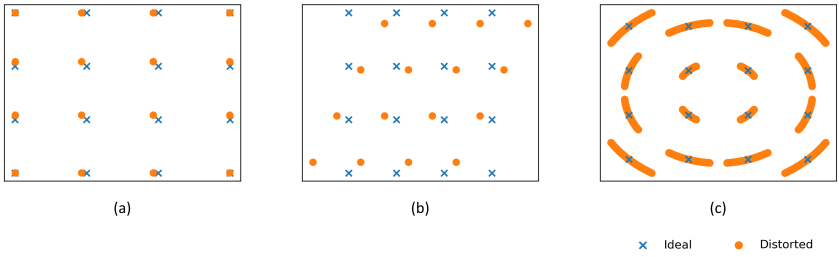


Figure 3.2: Distortions caused by transmitter imperfections. (a) Nonlinear constellation distortion due to imperfect IQ MZM biases. (b) Loss of constellation orthogonality due to imperfect biasing of the $\pi/2$ phase shifter. (c) Phase rotations due to non-zero linewidth of optical carrier.

The constellation diagrams in Figure 3.2 assume noise-free conditions. This, naturally, is unrealistic since most system components will add some noise and under its influence I and Q spreads will be observed. Noise also exacerbates the influence of the system's imperfections and causes higher BERs. Therefore,

resolving these issues is necessary and is normally done through dedicated hardware solutions or signal pre/post processing.

Choices pertaining to the driving signals $x_I(t)$ and $x_Q(t)$ are also crucial, since they influence the achievable BER in a noisy system. For example, using higher-level modulations not only lowers the signal's noise tolerance, but also adds transmitter-side noise due to the finite resolution of the digital-to-analog converters. Furthermore, the frequency response of the bandwidth-limited transceiver electronics may be distorting. This necessitates choosing bandwidth-limited waveforms for the signals $x_I(t)$ and $x_Q(t)$, as discussed in the following subsection.

3.2.2 Pulse Shaping

The constellation diagrams in Figure 3.1 are formed by periodically sampling a time-continuous signal whose waveform must be bandlimited. This condition ensures that both transmission and sampling can happen without intersymbol interference (ISI). Furthermore, due to the limited available bandwidth, spectrally-efficient waveforms are needed. Here, spectral efficiency can be defined as the ratio of the achieved symbol rate R to the spectral bandwidth W required to support this rate [22]. Therefore, for a given rate, it is of interest to minimize the occupied bandwidth to achieve higher spectral efficiency.

Spectral efficiency is directly influenced by the shape of the time-continuous pulses representing the discretized symbols. For instance, using square pulses leads to the spread of significant energy over a large bandwidth, because their spectrum is given by a sinc function. Apart from being physically unimplementable (due to the zero transition time), they are also spectrally inefficient. In contrast, employing sinc pulses will result in the lowest possible occupied bandwidth, where $W = R$ for a passband signal. Despite being spectrally efficient, this too is physically impractical due to the infinite duration of these pulses.

Alternatively, raised cosine pulses are a practically implementable and bandwidth-limited family of pulses. The pulse shape, and therefore the occupied bandwidth, can be adjusted by the so-called roll-off factor, α, of the pulse shaping filter. This parameter controls the smoothness of the transition between the passband and stopband, with $\alpha = 0$ corresponding to a sinc pulse. As the roll-off factor increases, the occupied bandwidth expands, following the relation $W = (1 + \alpha)\, R$ [121]. The frequency and impulse responses of three roll-off factors are shown in Figure 3.3. The main lobe of the pulse has a time period $T_s = 1/R$, and the pulse tails exhibit zero-crossings at integer multiples of T_s.

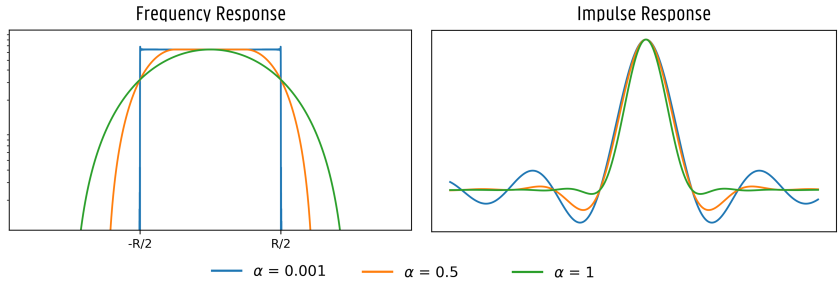


Figure 3.3: Behaviour of raised cosine pulses, where the frequency response is shown on the left and the impulse response is shown on the right for three roll-off factors α. Higher α values result in larger bandwidths and faster decaying pulses.

To enhance ISI mitigation, the pulse shaping filter can also be split between the transmitter and the receiver, thus realizing matched filtering. This results in the root-raised cosine (RRC) filter, which is identically applied on both the transmitter's side and the receiver's side.

One drawback of using raised cosine and RRC pulses is that accurate sampling is required at the receiver's end to select the pulses' centers. If sampling jitters are encountered, the neighboring pulses will have a non-zero contribution to the sampled pulse due to their tails, resulting in signal distortions. This is further discussed in the next section in addition to other receiver-related designs and issues.

3.3 Coherent Receivers

Despite the photodiode's (PD) insensitivity to the optical phase, coherent detection of optical signals can be realized by relying on the following squaring property

$$|E_x(t) + E_{LO}(t)|^2 = |x(t)|^2 + |LO(t)|^2 + 2\mathbb{R}\{x(t)\, LO(t)^*\} \tag{3.3}$$

where $E_x(t)$ is the optical field of a modulated optical carrier and $E_{LO}(t)$ is the field of an unmodulated continuous wave laser, known as the local oscillator (LO). The signals $x(t)$ and $LO(t)$ are the baseband representations of their respective optical fields.

The last term of Equation 3.3 indicates that the real part of the field $E_x(t)$ can be retrieved in the presence of this LO. By shifting the phase of $E_x(t)$ or $E_{LO}(t)$ by $\pi/2$, the same term also becomes proportional to the imaginary part of $E_x(t)$, thus enabling detection of the full complex field. However, a challenging aspect of this detection is $|x(t)|^2$, which is known as the signal-signal beat interference (SSBI).

The receiver must be designed to suppress this term, since its frequency contents overlap with the desired $2\mathbb{R}\{x(t)\ LO(t)^*\}$ and thus are distorting to the signal.

To realize this behaviour, coherent receivers are formed by cascaded couplers and two pairs of balanced PDs, as shown in Figure 3.4. The complex optical field $E_x(t)$ and a narrow-linewidth continuous wave laser forming the LO, $E_{LO}(t)$, are sent into the receiver at different branches. By propagating through the couplers, these optical fields are phase shifted by multiples of $\pi/2$ as indicated on the various branches of Figure 3.4. As such, the structure formed by these cascaded couplers is also known as a 90-degree hybrid.

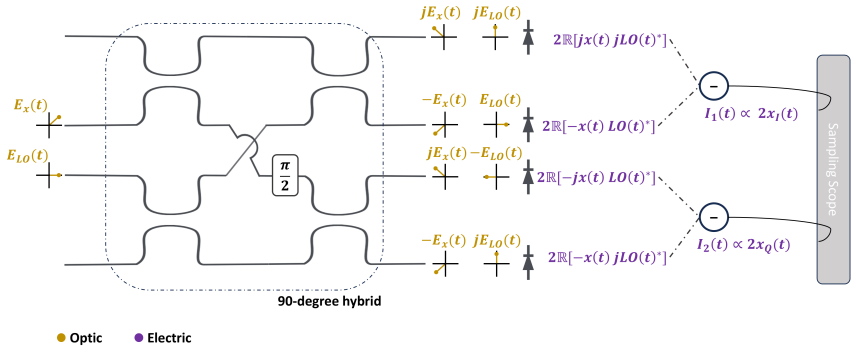


Figure 3.4: Structure of a coherent receiver composed of a 90-degree hybrid and two pairs of balanced photodiodes. The couplers of the 90-degree hybrid create phase shifts for the input signal $E_x(t)$ and the local oscillator $E_{LO}(t)$. These shifts are indicated at the output of the 90-degree hybrid. When mixed coherently, the output of the top pair of photodiodes is proportional to $x_I(t)$ (i.e, the baseband representation of $\mathbb{R}[E_x(t)]$) and the output of the lower pair is proportional to $x_Q(t)$ (i.e. the baseband representation of $Im[E_x(t)]$). The subtraction discards the other terms arising from the squaring, namely $|x(t)|^2 + |LO(t)|^2$, which are present at each branch but were omitted from the figure for conciseness.

As seen in Figure 3.4, the detected outputs from the top two branches are proportional to the real component of $E_x(t)$, while the outputs from the lower two branches are proportional to the imaginary component of $E_x(t)$. Furthermore, due to the use of balanced PDs, the difference between every pair of branches removes the contribution of $|x(t)|^2$ and $|LO(t)|^2$, which achieves the SSBI suppression required. These terms are omitted from the detected outputs in the figure for conciseness.

In addition to their ability to detect phase information, coherent receivers are attractive because they have superior noise tolerance. This arises from subtraction in the balanced branches, which reduces the common-mode noise, including intensity fluctuations of the LO. Furthermore, the detected signals are amplified since they are multiplied by E_{LO}, which results in improved receiver sensitivity.

However, similar to the transmitter, the receiver's complex architecture makes it sensitive to imperfections, which generate system errors. For example, the performance of the receiver is affected by attributes of the PDs used in the receiver, including limited bandwidth, shot noise, thermal noise, and responsivity. Additionally, the path lengths traversed by the real and imaginary components may be unequal due to a mismatch in the optical or electrical paths. This causes a so-called IQ skew, an example of which is seen in Figure 3.5a. Furthermore, as discussed in the previous section, an ADC is required to digitize the analog signals. The sampling rate and sampling phase must be matched to the transmitter, otherwise sampling errors occur causing distortions, as seen in Figure 3.5b. Finally, the equations in this section assume a matched frequency between the optical carrier and the LO. However, this is not often the case since both frequencies drift, resulting in an ambiguous frequency difference. This results in the signal rotations seen in Figure 3.5c. These issues are also addressed using dedicated hardware and/or signal processing as will be discussed in the following chapters.

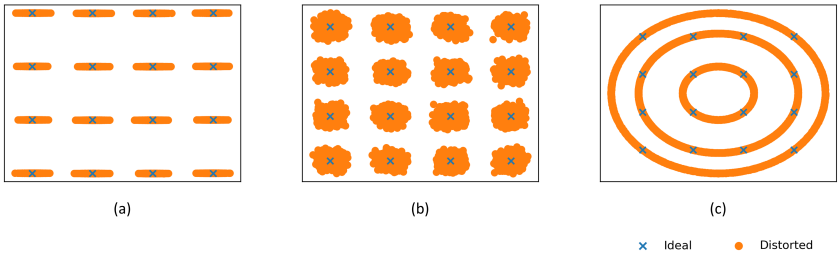


Figure 3.5: Disortions due to receiver imperfections . (a) IQ skew due to mismatched electric or optic path lengths. (b) Distortions caused by sampling errors (c) Phase rotations due to frequency drift.

3.4 Self-Coherent Receivers

Given the complex architecture of coherent receivers, alternative systems are often investigated to reduce the cost and complexity of these receivers. Self-coherent receivers offer this advantage by ridding the receiver from the need for a LO, and instead explore methods for invoking self-coherence. A common approach for achieving this involves co-propagating an unmodulated optical carrier (sometimes referred to as a reference carrier) along with the modulated message signal. This reference carrier is placed at a frequency just outside the passband of the message signal, resulting in a single sideband (SSB) signal, where the entire modulated signal is located on one side of the reference carrier [15]. An example of this SSB frequency spectrum is shown in Figure 3.6.

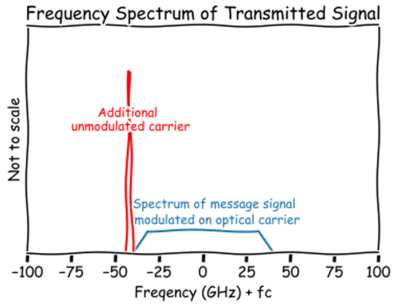


Figure 3.6: Frequency spectrum of a single-sideband signal (SSB). A message signal is modulated on an optical carrier with frequency f_c and is co-propagated with a reference carrier at the left edge of the message signal

The presence of the reference carrier allows a PD response similar to that of Equation 3.3. Furthermore, instead of devising a receiver architecture that constructs the $x_I(t)$ and $x_Q(t)$ components separately, self-coherent receivers devise methods to reconstruct the complex-valued $x(t)$ from the measured signal attributes. This often reduces the need for two pairs of balanced PDs, however at the expense of some signal processing requirements.

Other forms of self-coherence that do not rely on SSB signals are also possible. For example, the Gerchberg-Saxton (GS) receiver achieves this through the well-known GS iterative algorithm [25]. Here a pair of PDs are used, with an optical dispersive element preceding one of them. This dispersive propagation can convert phase modulation to amplitude modulation, and is therefore used to recover the message signal's phase information.

In this thesis however, the focus is on SSB self-coherent receivers, and two such architectures are discussed in this chapter. The Kramers-Kronig (KK) receiver is first explored, which utilizes the infamous KK relations to reconstruct the complex signal from the *amplitude* measured by a single PD. In the subsequent subsection, a receiver architecture known as a SiP SSBI cancelling receiver is explored. Here, the focus is on reconstructing the complex signal from the *real* component measured by a pair of PDs.

3.4.1 Kramers-Kronig Receiver

The Kramers-Kronig receiver is a phase-retrieval based self-coherent receiver that employs a single PD to detect a complex signal's power and then reconstructs its complex nature through a series of processing steps. This is possible when a

complex signal's phase is uniquely related to its amplitude, which is a property of minimum-phase signals. These signal can be constructed provided that the following two necessary and sufficient conditions are met [122].

First, the signal must be a single side-band (SSB) signal. Thus, an additional optical carrier should be added to either edge of the signal's spectrum. Second, the reference carrier must have sufficiently higher power with respect to the signal; this is termed the carrier-signal power ratio (CSPR). This creates the optical field $E_s(t)$ such that

$$E_s(t) = \left(x(t)\, e^{j2\pi f_z t} + c(t)\right)\, e^{j2\pi f_r t} \tag{3.4}$$

where $c(t)$ is the reference carrier, $x(t)$ is the complex-valued message signal, f_r is the reference carrier frequency, and f_z is the difference between the f_r and the main carrier's frequency f_c.

The goal of the reconstruction is to retrieve the complex-valued, down-converted signal $s(t)$ from the measured quantity $|s(t)|^2$. This is achieved through relating the signal's phase information $\phi(t)$ to the magnitude of the signal $|s(t)|$. In [24] it is shown that

$$\phi = \frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{\ln(|s(t)|)}{t-\tau}\, d\tau \tag{3.5}$$

where P is the Cauchy principal value and $\ln$ is the natural logarithm.

This equation is based on the well-known KK relations that relate the *real* and *imaginary* parts of a signal. However, when taking the logarithm of the signal, its *amplitude* and *phase* become separated into the sum of real and imaginary parts, allowing the use of these relations. Indeed, consider the following baseband representation of the signal $s(t)$

$$s(t) = |s(t)|e^{j\phi(t)} \tag{3.6}$$

and by taking its natural logarithm becomes

$$\ln(s(t)) = \ln\left(|s(t)|\right) + j\phi(t) \tag{3.7}$$

This is the reason the natural logarithm is seen in Equation 3.5. A complexity arising from the use of this logarithm however, is that the frequency spectrum of the signal is broadened and therefore requires at least 3-fold upsampling (i.e., 6 samples/symbol) for correct operation. In addition, the equation calls for the use of $|s(t)|$, which is the square root of the measured $|s(t)|^2$. This nonlinear operation is another reason for the required upsampling. When $\ln(s(t))$ from Equation 3.7 is retrieved, exponentiating will allow reconstructing $s(t)$, and in turn arrives to the down-converted form of Equation 3.4, which is $x(t)e^{j2\pi f_z t} + c(t)$.

Practical implementations of the KK receiver may choose to circumvent performing the time integral of Equation 3.5, as it can be quite involved. Alternatively, going through the frequency domain is simpler, where the KK relationship of Equation 3.5 assumes the form of a sign function [123,124]. This implementation is shown in the pipeline of the KK receiver in Figure 3.7 with insets of the frequency components of the signals at different stages. Initially, the optical SSB signal is photo-detected and then sampled by an ADC. The signal is then up-sampled, such that the square root and logarithm operations can be implemented on over-sampled signals. The KK relations are then applied in the frequency domain through a Hilbert transform. The complex signal produced after the KK relations is then exponentiated to reverse the effect of the logarithm and can subsequently be down-sampled. This reconstructed signal is a down-converted $s(t)$ and therefore includes the message signal and the reference carrier, similar to the optical signal impinging on the photodiode. Further processing is thus required to retrieve the message signal, which involves removing the carrier and frequency shifting the signal by f_z.

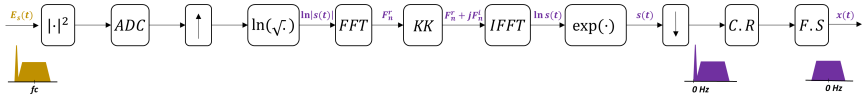


Figure 3.7: KK receiver pipeline, showing receiver blocks, several signals along the pipeline, and their frequency spectrum. Receiver blocks: $|\cdot|^2$: Photodiode – ADC: Analog-to-Digital Converter – $\uparrow$: upsampling - ln: natural logarithm – FFT: Fast Fourier Transform – KK: KK relations – $IFFT$: Inverse Fast Fourier Transform exp: exponential – $\downarrow$: down-sampling – $C.R$: carrier removal (through filtering) – $F.S$: frequency shifting. Signals: $s(t)$: single sideband signal (includes unmodulated carrier and message signal) – F_n^r: frequency components of real signal – $F_n^r + jF_n^i$: frequency components of complex signal – $x(t)$: complex message signal

This pipeline will accurately regenerate the complex-valued nature of the message signal, provided that the minimum phase conditions are satisfied and the oversampling requirements in the nonlinear steps are respected. When violating these conditions however, signal distortions will arise. This can be seen in Figure 3.8, where an undistorted 64 QAM signal is detected at CSPR levels ranging between 4 and 9 dB and using 4-fold upsampling in the procedures requiring oversampling. It is only around 9 dB CSPR that accurate reconstruction is seen.

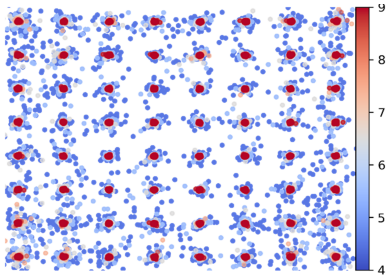


Figure 3.8: Influence of CSPR on reconstruction of an undistorted 64 QAM signal shown as a heatmap.

Figure 3.9 also illustrates the combined impact of altering the CSPR and samples-per-symbol (sps) on the receiver's performance in a realistic system that has some additive noise. It showcases BER vs CSPR in a back-to-back KK transceiver configuration utilizing 64 QAM signals. Here, the signal impinging on the PD has an Optical Signal-to-Noise Ratio (OSNR) of 27 dB. The trendlines depict the influence of employing signals with 6, 4, and 3 sps in the nonlinear steps of the pipeline in Figure 3.7. The figure shows that the bit error rates are very high for low sps, but drop as the CSPR increases irrespective of the oversampling. Since the BER plateaus at around 9 dB and 6 sps, this can be considered near ideal reconstruction for the system parameters under consideration.

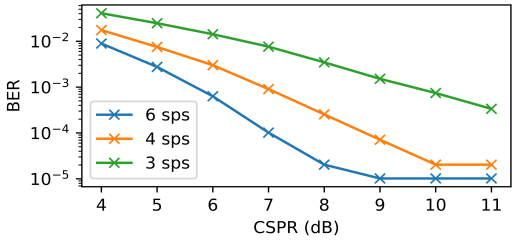


Figure 3.9: BER vs CSPR of a back-to-back 64 QAM KK transceiver system operated at 6,4 and 3 samples per symbol

The errors incurred in the back-to-back scenario stem exclusively from the receiver's nonlinearity. Some of these error values appear relatively low, but since they are the result of nonlinear transformations, they are expected to significantly impact subsequent digital signal post-processing stages, which typically require linear receiver operation. The impact of such distortions will further be discussed in the following chapters. We will also investigate how such conditions can be relaxed using photonic processing and how this distortive impact can be mitigated.

3.4.2 Signal-Signal Beat Interference Cancelling Receiver

The high CSPR and oversampling requirements of the KK receiver both arise from attempting to utilize the KK equations in an amplitude-to-phase relationship. However, when using them in their original real-to-imaginary form, the required CSPR is significantly reduced and the need for oversampling is circumvented since nonlinear processing operations are no longer needed. This approach is used in the SiP signal-signal beat interference (SSBI) canceling receiver.

Recall that the squared signal produced by the photodiode in the presence of a LO already contains the real component of the signal and the conjugate of the LO, as in Equation 3.3. With an SSB signal, the reference carrier $c(t)$ replaces the LO and can therefore be used to ensure the term $2\mathbb{R}\{x(t)\ c(t)^*\}$ is produced after the PD. However, the presence of the SSBI term $|x(t)|^2$ obscures immediate use of it. Furthermore, the overlapping frequency components of the desired and interfering signals, as illustrated in Figure 3.10, make its suppression in post-signal processing challenging.

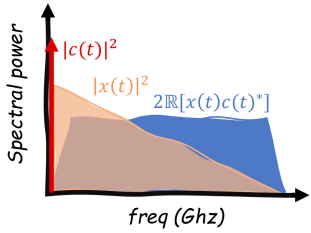


Figure 3.10: Output terms from detection of an SSB signal in the frequency domain. The SSBI term $|x(t)|^2$ overlaps with the desired signal $2\mathbb{R}\{x(t)\ c(t)^*\}$ causing distortion. The term $|c(t)|^2$ is the unmodulated reference carrier which can be removed by filtering

Alternatively, hardware suppression of the SSBI term can be implemented in various forms. Indeed, an example of such a behaviour was already demonstrated in the coherent receiver's architecture. A simpler hardware implementation (proposed in [26]) that can also be realized in integrated photonics, is shown in Figure 3.11. Here, the SSB signal is split between two branches, where one is immediately detected while the other passes through a notch filter before detection to suppress the reference carrier. Therefore, detection of the top branch generates the desired term, $2\mathbb{R}\{x(t)c(t)^*\}$, in addition to the SSBI term $|x(t)|^2$. Detection from the lower branch, on the other hand, only yields the SSBI term. This term can subsequently be subtracted from the detected output of the upper branch, which achieves SSBI mitigation. This allows utilizing the real component $2\mathbb{R}\{x(t)c(t)^*\}$ to reconstruct the complex-valued $x(t)$ using the KK relations. Finally, the signal is frequency shifted by f_z

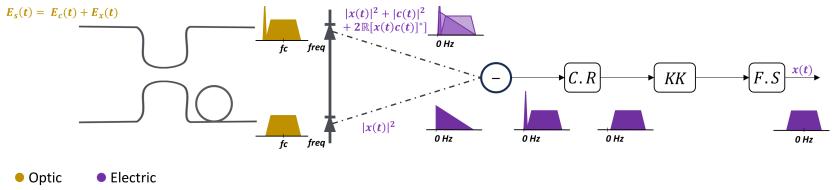


Figure 3.11: Architecture of SiP SSBI canceling receiver. The upper branch leads to a PD where the desired and SSBI terms are generated post detection. On the lower arm, an MRR behaving as a notch filter suppresses the reference carrier, which generates only the SSBI term post detection. Subtracting the outputs of the two arms yields the desired $2\mathbb{R}\{x(t)c(t)^*\}$. This is subsequently sent for carrier removal (C.R), after which the KK relations can be used to reconstruct the complex valued $x(t)$. Finally a frequency shift (F.S) adjusts for the shift in frequency components post-detection

This technique is made possible by the use of SiP integration technologies, which facilitate having two short paths with minimal and stable path length difference. The required notch filter can be realized by a micro-ring resonator (MRR), as shown in Figure 3.12 [125]. The resonator is formed by connecting the top ports of a directional coupler together by a waveguide, thus creating a ring. The portion of the amplitude of the signal that cross-couples to the ring is denoted by κ, while the portion that self-couples is given by r. The round-trip losses in the ring are given by a, which include the waveguide bend losses, the waveguide propagation losses, and the coupling losses [126].

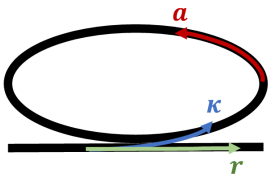


Figure 3.12: Micro-ring resonator in an all pass filter form realized by feeding one of the directional coupler's output back to itself, thus creating a ring. κ is the cross-coupling coefficient, r is the self-coupling coefficient, and the ring's round-trip losses are given by a,.

Note that a is computed in the Nepper units, and relates to the Decibel α as follows

$$a = e^{\frac{\alpha}{8.686}} \tag{3.8}$$

The circumference of this ring L is designed to introduce a wavelength-dependent phase delay. When this incurred phase delay for a particular wavelength is an integer multiple of 2π, constructive interference occurs for that wavelength. This is known as the resonance wavelength of the MRR and is given by

$$\lambda_{\text{res}} = \frac{n_{\text{eff}}(\lambda)\, L}{m} \tag{3.9}$$

where m is an integer. The quality factor Q of the ring is a measure of its sharpness at λ_{res}, and is given by [126]

$$Q = \frac{\pi n_g L \sqrt{ra}}{\lambda_{res}(1-ra)} \tag{3.10}$$

High Q-factors are desirable since they allow suppression of the reference carrier without affecting the neighboring frequency components of the message signal. For a given length, a higher Q factor is achieved by minimizing the cross-coupling factor κ (or similarly, increasing the self-coupling r).

The extinction ratio $E.R.$ of the filter is a measure of how well the carrier signal is suppressed in reference to the pass portion of the spectrum. This is given by

$$E.R. = \frac{(r+a)^2\,(1-ra)^2}{(r-a)^2\,(1+ra)^2} \tag{3.11}$$

To maximize the extinction ratio, the self-coupling coefficient must be as close as possible to the total losses of the MRR (a condition called critical coupling). Therefore, higher ring losses require a higher r to ensure a sufficiently large E.R. This relationship indicates that, depending on the losses in the ring, a high Q-factor may not yield a high extinction ratio. This trade-off is seen in the Figure 3.13, which shows the frequency responses from varying coupling coefficients of a ring whose combined losses are 0.1 dB. The ring's circumference is designed for a resonance frequency of -19 GHz away from the carrier's frequency. The responses of these MRRs show that the highest simulated Q-factor corresponds to the lowest extinction ratio.

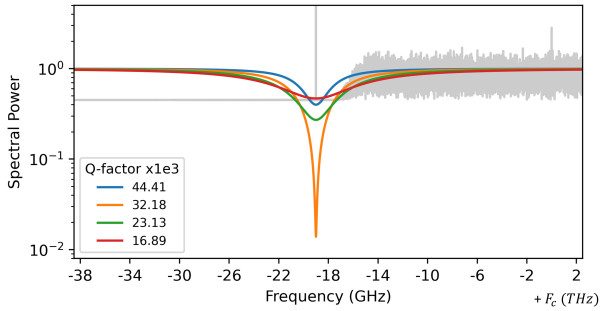


Figure 3.13: Frequency response of an MRR with 0.1 dB losses designed for resonance at $f_c - 19Ghz$. Various coupling coefficients are simulated and the figure shows how both the Q-factor and extinction ratio of the ring are affected. The ring spectrum is overlaid on the frequency response of an SSB signal for which f_z is -19 GHz. The message signal is a 32 Gbaud signal, pulse-shaped by a raised cosine filter with a 0.1 roll-off. Note that the spectrum's power is adjusted to enhance the figure's readability.

Given the challenges of designing an ideal notch filter with a high Q-factor and a high E.R., we need to solve an optimization problem that balances the suppression of the reference carrier against the flatness of the frequency response at the message signal's frequencies. Indeed, on the one hand, an unsuppressed carrier results in the output of the receiver's lower branch having terms other than the SSBI term. Thus, subtraction of the two branches will not yield full SSBI mitigation, which results in nonlinear distortions. On the other hand, the finite roll-off of the filter distorts the message signal, producing an SSBI term on the lower branch different from that on the upper branch, which also hinders full mitigation. To visualize the relevance of this roll-off, Figure 3.13 shows overlaid on the ring spectra the frequency response of a 32 Gbaud signal pulse shaped using a raised cosine filter with a 0.1 excess bandwidth factor. Despite the spacing, or guard-band, between the reference carrier and the message signal, the filter roll-offs overlap with a substantial portion of the message signal, indicating potential distortions.

To mitigate the influence of these effects, a trainable parameter, β, is introduced to the digital output of the receiver's lower branch. This multiplicative factor adjusts the magnitude of the SSBI term, thereby improving SSBI suppression. However, β also scales all other outputs of the lower receiver branch, which in the case of carrier residuals will include the desired term as well.

Additional SSBI mitigation can also be achieved by increasing the CSPR of the signal, which boosts the power of the desired term in the upper receiver branch. This improves the ratio of the desired signal to any remaining SSBI, leading to enhanced BER performance.

The influence of the CSPR and β as well as the ring design parameters can be seen in Figure 3.14. Here the BER for a back-to-back system transmitting 64 QAM signals with a 27 dB OSNR is shown against the Q-factor of the ring. The message signal is a 32 Gbaud signal pulse shaped using a raised cosine filter with a 0.1 roll-off and the reference carrier is placed 19 GHz to the left of the carrier frequency. The ring's length is thus set for a resonance at the reference carrier's frequency. The E.R. values are displayed on the top of the figure for reference. Solid trend lines are used for unoptimized β, while dashed lines indicate an optimized β parameter. Finally, the various trendline colors reflect different CSPR values.

As seen, both training β and increasing the CSPR improve the receiver's SSBI mitigation, especially in high-error scenarios. Figure 3.14 further shows how the Q-factor and E.R influence the receiver performance. The results indicate a consistent local minimum for systems with optimized β regardless of the CSPR which lies around a Q-factor of 34,500 with an E.R. of 20.67 dB. When unoptimized, the local minimum is found at a higher Q-factor, which comes at the cost of a lower E.R. This optimal point shifts towards even higher Q-factors as the CSPR reduces.

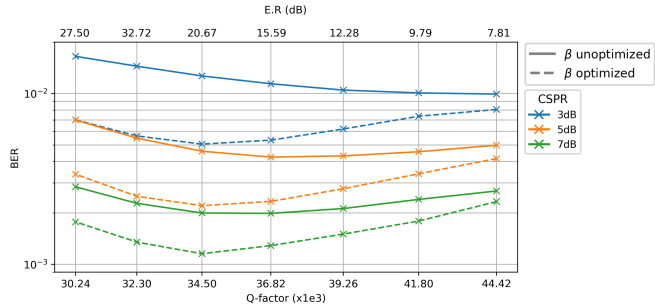


Figure 3.14: BER for a back-to-back system transmitting 64 QAM signals with a 27 dB OSNR vs. the MRR's Q-factor. The corresponding E.R. values are displayed at the top of the figure. Solid trend lines indicate unoptimized β, while dashed lines represent optimized β. Different colors reflect various CSPR values.

Finally, we note that an MRR's resonance frequency is very sensitive to its length and effective refractive index. This indicates that variability in these parameters, for example due to the waveguide's sidewall roughness, will cause offsetting of the resonance frequency. The impact of this offset depends on whether it is positive (i.e., closer to the signal) or negative (i.e., further away from the signal) as illustrated in Figure 3.15

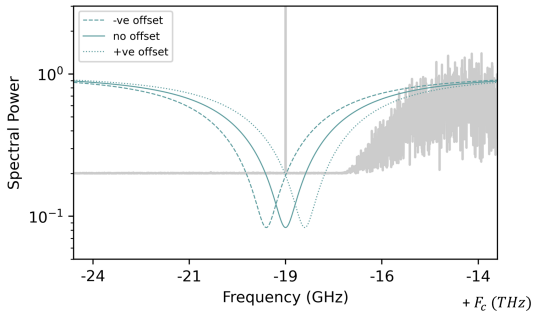


Figure 3.15: Illustration of positive and negative offsets for an MRR's resonance frequency. The frequency responses are shown and are overlaid with the spectrum of the same SSB signal in Figure 3.13. Note that the spectrum's power is adjusted to enhance the figure's readability.

In both cases, the offset results in a reduction of the effective E.R that the reference carrier sees. Additionally, in the case of a positive offset, a more pronounced distortion of the message signal is incurred due to the filter's roll-off. However, for small negative offsets the receiver's performance may potentially improve since the impact of the filter's roll-off is lowered. At larger negative offsets, this advantage is lost and the BER worsens, but at a lesser rate compared to the positive offsets. All

of this is shown in Figure 3.16, where the BER of the same back-to-back system detailed above is plotted against the frequency offset. The CSPR is set to 5 dB and several Q-factor values are simulated that are depicted on the figure as trendlines. Again, the impact of an optimized vs an unoptimized β is shown through the use of solid and dashed lines respectively.

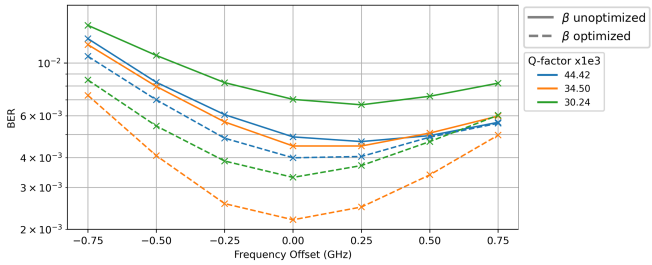


Figure 3.16: Impact of offsets in the resonance frequency of the MRR on the performance of the receiver. The BER vs offset is shown for various Q-factors that are depicted as trendlines. The solid lines indicate an unoptimized β, while the dashed lines indicate an optimized β. The same system parameters as those used for Figure 3.14 are used and CSPR is set to 5 dB

Similar to the errors of the KK receiver, the errors incurred in this back-to-back scenario stem from the receiver's nonlinearity, given that they arise from the nonlinear impairments term $|x(t)^2|$. As such, they are expected to impact subsequent digital signal post-processing stages, which typically require linear receiver operation. In the following chapter, further improvements on this receiver's behaviour due to the photonic reservoir will be shown. Additionally, the advantages of performing optical signal equalization for channel impairments, as opposed to an electronic equalization, is shown in the presence of imperfect detection from both self-coherent receiver architectures.

4

Simulated Reservoir Equalization with Self-Coherent Detection

Parts of this chapter were published in [127], [128], and included in the MSc thesis of Aimen Zelaci [129]

4.1 Overview

Transmission in fiber optic communication systems typically induces signal distortions, which arise from the response of various components including

transmitters, optical fibers, and receivers. As such, signal equalization is essential for ensuring error-free communication and is typically performed through digital signal processing. This includes procedures like estimating and inverting the channel response, as well as adaptive equalization for variable impairments.

Performance of such procedures, and in particular correction for the channel impairments, is more effective when the full optical field is available for processing. This ensures that the compensation can be implemented in a straightforward manner using the fiber's transfer function implemented with linear filters. However, when using self-coherent architectures, the receiver transformation can be nonlinear if the operational conditions are not optimal. This introduces a nonlinear transformation on the optical field, resulting in suboptimal digital equalization.

Generally, when the receiver introduces such nonlinear transformations, advanced digital equalizers including those that leverage ML techniques are required. This can also be seen in the use of ML-based networks for compensation of direct-detected signals, where the nonlinear transformation is the squaring response of the photodiode [130–132].

However, simpler solutions can be employed if the compensation is performed *prior* to detection. This approach leverages the use of the optical field before it undergoes the nonlinear receiver transformations, which in turn enables accurate equalization of channel impairments. Additionally, this processing is performed in the optical analog domain, thus alleviating the need for digital equalizers that operate at twice the baudrate. Several such implementations for optical equalization were demonstrated [36, 133–137], albeit primarily in IMDD systems.

In this chapter, we explore the use of a photonic reservoir for equalization in systems deploying self-coherent receivers. The systems examined use 64-QAM transmission, which is highly sensitive to nonlinear transformations, and include implementations with both a KK receiver and a SiP SSBI-cancelling receiver. The operational conditions of the receivers are varied to study the advantages of performing optical equalization with varying degrees of receiver distortions. The simulation results show that addressing channel impairments prior to a perfectly coherent receiver will match the DSP performance, while doing so before a distorting receiver significantly surpasses the DSP performance. Additionally, given the trainable nature of the photonic reservoir, we show how this architecture can simultaneously perform equalization and optical bandpass filtering. This behaviour allows it to replace both the DSP equalizer and the optical bandpass filter required for out-of-band noise rejection.

4.2 Simulation Setup

4.2.1 Reservoir Pipeline

The simulation setup for equalization with the photonic reservoir is shown in Figure 4.1. The system blocks in the red box are implemented in VPI, of which a detailed schematic is shown in Figure 4.3. The rest of the system blocks are implemented in Python using the libraries Photontorch and Pytorch for the reservoir and receiver respectively.

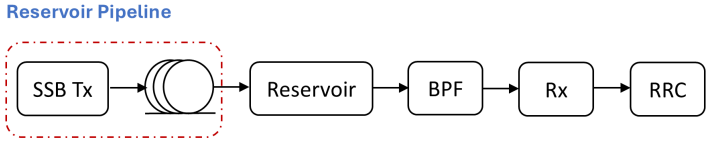


Figure 4.1: Simulation setup for equalization with the photonic reservoir. The system blocks in the red box are implemented in VPI, of which a detailed schematic is shown in Figure 4.3. The rest of the system blocks are implemented in Python using the libraries Photontorch and Pytorch for the reservoir and receiver respectively

At the transmitter, a 64 QAM signal is generated at 32 Gbauds with a modulator that incurs slight nonlinearity due to imperfect biasing. The effect of this nonlinearity on a noise-free 64 QAM signal is shown in Figure 4.2.

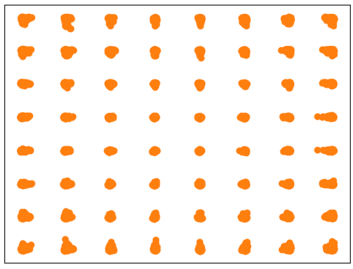


Figure 4.2: The effect of modulator nonlinearity on a noise-free 64 QAM signal. This is controlled by toggling the "electric pre-distortion" parameter in VPI which pre-compensates for the modulator nonlinearity. Without this pre-compensation, the constellation in the figure is generated

The signals are then noise loaded to realize a 27 dB OSNR before being coupled with a reference carrier to generate an SSB signal. The carrier is placed -19 GHz away from the main carrier (i.e., $f_z = -19$ GHz), and various CSPR values are simulated. Note that Figure 4.1 abstracts the transmitter behaviour into a single block, but this is further elaborated upon in Figure 4.3a. This SSB signal is sent

through lengths of SSMF ranging between 10 km and 90 km. The distorted signal is then injected into a parallel-readout 16-node reservoir whose architecture was shown in Figure 2.7, with all 16 nodes connected to the readout. The signals are then passed to an optical bandpass filter, followed by detection with either a KK receiver or a SiP SSBI-cancelling receiver. Finally, the signals are passed to an RRC filter with a 0.1 rolloff for matched filtering. The transmission parameters are also tabulated in Table 4.1.

The transmitter and fiber blocks, indicated in the red box in Figure 4.1 , are implemented using the VPI Transmission Maker simulation software [138], using the VPI schematic in 4.3a. This setup enables the export of optical fields from VPI as Python arrays, which are subsequently used in the remaining system blocks within the Python environment. The photonic library Photontorch [110] is used for implementing the reservoir, while the Pytorch [139] library is used for implementing the receiver.

The receivers are simulated as per their pipelines shown in Figure 3.7 for the KK receiver and in Figure 3.11 for the SiP SSBI-cancelling receiver. An ADC is used at the output of both receivers with a sampling rate of 2.5$\times$baudrate. This signal is then up-sampled to either 4 or 6 samples/symbol for the KK receiver, while the SSBI-cancelling receiver does not require up-sampling.

Parameter	Value
Symbol rate	32 Gbaud
OSNR	27 dB
Signal power	3 dBm
Pulse shape	RRC, 0.1 rolloff
Carrier linewidth	0 kHz
CSPR	3 - 12 dB
f_z	-19 GHz
Fiber length	10 - 90 km
Fiber loops	1
Fiber attenuation	0.2 dB / km
Fiber nonlinear index	$2.610^{-20}\mathrm{m}^2/\mathrm{W}$
Fiber dispersion parameter	16 ps / (nm km)

Table 4.1: Simulation parameters. Note that signal power is measured without the reference carrier's power

Optimization of the reservoir weights, as will be detailed in the following section, requires the use of a target signal. This is a noise-free and undistorted signal that did not incur the transmitter and fiber impairments. To generate this signal, the same VPI simulation parameters shown in Table 4.1 are used, however, without the noise

loading and fiber blocks. The transmitter's nonlinearity is also pre-compensated. This is implemented through the VPI schematic shown in Figure 4.3b. The resulting optical fields are then also exported as Python arrays.

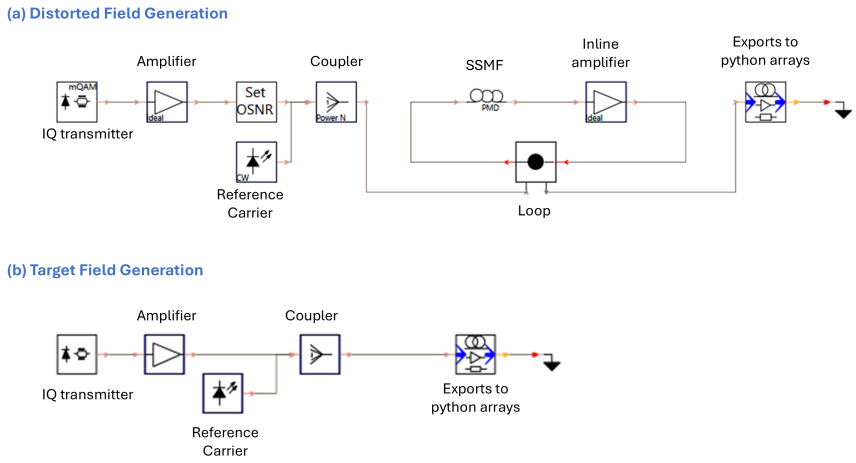


Figure 4.3: VPI schematic to generate the SSB optical fields and export them as Python arrays. (a) Pipeline to generate distorted signals, including noise loading, a nonlinear modulator, an SSMF, and a noise-free inline amplifier to compensate fiber losses (b) Pipeline to generate undistorted target signals with ideal modulator and without noise loading.

4.2.2 Benchmark Pipeline

To benchmark the results from the reservoir's equalization, a DSP pipeline is set up that replaces the pre-receiver reservoir with post-receiver processing implemented either through a frequency-based chromatic dispersion compensating (CDC) filter or a time-domain feed-forward equalizing (FFE) filter. All of the other system blocks have the same behaviour and parameters as the reservoir pipeline.

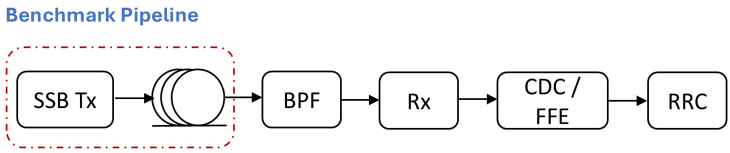


Figure 4.4: Simulation setup for equalization with DSP blocks. The system blocks in a red box are implemented in VPI, where a detailed schematic is shown in Figure 4.3. The rest of the system blocks are implemented in Python using the libraries Photontorch and Pytorch for the reservoir and receiver respectively.

The CDC block is implemented in the frequency domain by applying the inverse frequency response of the fiber to the signal's spectrum. This process adjusts the phases of the different frequency components, since chromatic dispersion arises due to different spectral components of a guided mode propagating with different group velocities. The frequency response can be written as

$$H_{CDC}(\omega) = e^{j\frac{\beta_2 L\omega^2}{2}} \tag{4.1}$$

where L is the fiber length, ω is the angular frequency, and β_2 quantifies the spreading of pulses in $s/m/[rad/s]$. It relates to the fiber dispersion parameter D, given in $ps/nm/km$, as

$$D = -\frac{2\pi c}{\lambda^2}\beta_2 \tag{4.2}$$

For practical digital implementations, the fiber response is truncated such that it can be implemented with a finite impulse response (FIR) filter. Such a filter has a non-recursive structure and is typically implemented using a tapped delay line [140, 141], as shown in Fig. 4.5. Here, a sequence of delay elements are cascaded, each introducing a time delay corresponding to half the symbol period. A portion of the signal is tapped off after every delay, and weighted by a programmable weight. The weighted signals are then summed to generate the equalized output signals. Given the truncated response of the FFE, it has limited memory compared to the $H_{CDC}(\omega)$ implementation, with the number of taps controlling the amount of dispersion that can be compensated. In this case, we set the number of taps to 16 to match the number of reservoir nodes.

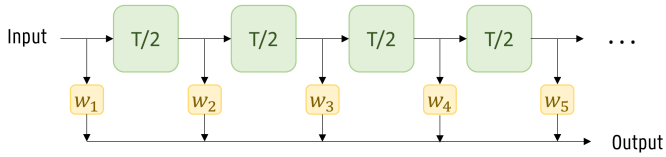


Figure 4.5: Feed-forward equalizer composed of a sequence of delay elements introducing a time delay $T/2$, corresponding to half the symbol period. A portion of the signal is tapped off after every delay, and weighted by a weight w_i, where i is the index of the i^{th} weight. The weighted signals are then summed to generate the equalized output signal

Despite its limited memory, FFE offers the advantage of being a trainable filter, allowing its coefficients to be optimized to compensate for dispersion and other residual impairments. This makes it closer to the reservoir's behavior, since such flexibility is not achievable with the $H_{CDC}(\omega)$ implementation, which relies on fixed, non-trainable parameters. In the following section, the training procedure followed for both the reservoir and FFE is detailed.

4.3 Training Procedure

The training data set is composed of over 16,000 symbols that are generated using a Wichmann-Hill generator [142]. Training first starts by computing weights using linear regression as in Equation 2.7. Here, $\hat{y}$ is the target signal generated by the pipeline in Figure 4.3b and the input matrix X is composed of the 16 states from each reservoir node. The resulting computed weight vector appropriately compensates for dispersion, which is the biggest error contributor in these systems. However, the effect of the receiver's distortions - if any - are not factored in when performing linear regression. This is because the input matrix on which the weights are computed consists of the reservoir states before they are passed into the receiver. To incorporate the receiver's behaviour and thus achieve further BER improvement, training by backpropagation is used. This method attempts to minimize the difference between the downsampled output of the receiver and the transmitted symbols. To quantify this error, the higherorder error function, as in equation 2.5, is used to generate the error term $\mathrm{error}_{\mathrm{HOE}}$.

However, since the reservoir behaves as a trainable filter, a possible solution that backpropagation can arrive at is one where the signal power is attenuated with respect to the carrier, thus increasing CSPR. To maintain a fair comparison with the benchmark, we ensure that the CSPR does not grow by introducing in the error function a second term, $\mathrm{error}_{\mathrm{CSPR}}$, that heavily penalizes growth of CSPR. This term should also be neutral to the reduction of the CSPR, as the optimizer would otherwise be stuck. Additionally, CSPR reduction is already inherently accounted for in the calculation of the other error term $\mathrm{error}_{\mathrm{HOE}}$, since CSPR reduction typically leads to lower performance. A suitable function for implementing this $\mathrm{error}_{\mathrm{CSPR}}$ is a softplus [143] function, which has the desired behaviour, in addition to being differentiable and continuous. The expression for $\mathrm{error}_{\mathrm{CSPR}}$ is thus given by

$$\mathrm{error}_{\mathrm{CSPR}} = \frac{1}{\beta} \ln\left(1 + e^{\beta \Delta \mathrm{CSPR}}\right) \tag{4.3}$$

$$\Delta\mathrm{CSPR} = \mathrm{CSPR}_{\mathrm{reservoir}} - (\mathrm{CSPR}_{\mathrm{transmitted}} + \alpha) \tag{4.4}$$

where $\mathrm{CSPR}_{\mathrm{transmitted}}$ is the desired CSPR to maintain (which corresponds to the CSPR the signal was transmitted at) and $\mathrm{CSPR}_{\mathrm{reservoir}}$ is the CSPR measured at the output of the readout. This loss function should make sure that at the end of the training, $\mathrm{CSPR}_{\mathrm{reservoir}}$ is less than or equal to $\mathrm{CSPR}_{\mathrm{transmitted}}$. The hyperparameters β and α allow optimizing the behavior of the softplus function to appropriately penalize $\Delta\mathrm{CSPR}$ with respect to $\mathrm{error}_{\mathrm{HOE}}$. This ensures that the the latter's contribution is still significant and optimized for during training. The training error is then a sum of the two errors as described by Equation 4.5.

$$\mathrm{error}_{\mathrm{training}} = \mathrm{error}_{\mathrm{CSPR}} + \mathrm{error}_{\mathrm{HOE}} \tag{4.5}$$

A sweep of fiber lengths and CSPRs is performed to generate data under various operational points. For each combination of length and CSPR, a different random seed was used in the VPI simulator to generate the data stream and noise profile. Furthermore, each of these resulting signals were passed to a different and randomly initialized reservoir. This means the internal phases of the reservoir varied in each simulation. However, if the same CSPR-length combination is used for different receivers, the same reservoir is also used to ease comparability. For testing sets, over 32000 symbols were generated using a different VPI seed for the bit and noise generators compared to the one used for generating the training set. However, the same reservoir seed must be used in both the training and testing sets since the trained weights are reservoir-specific.

For training the FFE benchmark, linear regression is also used to compute the weight vector. This is done on the same training set used for the reservoir. However, the target signal consists of the transmitted symbols since the FFE computation happens after the receiver. The testing results reported are also on the same testing sets described above.

4.4 Results and Discussion

4.4.1 Performance with KK Receivers

Figure 4.6 shows the BER as a function of fiber length for the reservoir (blue), the benchmark CDC (orange), and the benchmark FFE (green). The results are presented for two variations of the KK receiver, both operating under distorting conditions. The left subplot depicts performance at a low samples-per-symbol rate (4 sps) with a higher CSPR of 7 dB, while the right subplot shows results for a low CSPR (5 dB) with a higher sampler-per-symbol (6 sps). In both cases, resolving dispersion before the receiver significantly improves the BER compared to performing it after the receiver. For comparison, we also show the back-to-back results achieved by these receivers (i.e., with no fiber). With 5 dB CSPR, the reservoir's behaviour approaches the receiver's back-to-back performance, indicating that the majority of channel-induced errors are resolved. With the higher CSPR, the results show residual uncompensated errors where the performance penalty compared to a back-to-back scenario is visible. In both cases however, the performance of the equalized system does not exceed the performance of an unequalized back-to-back system.

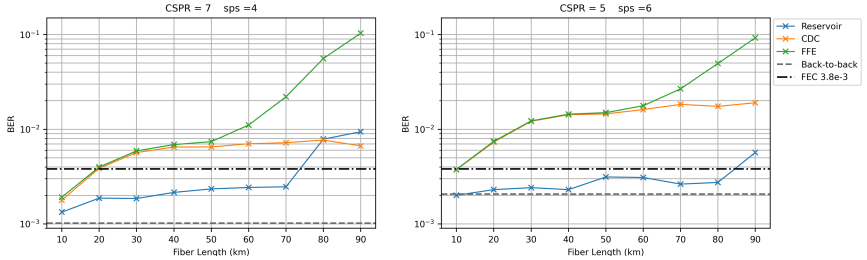


Figure 4.6: BER vs fiber length for the reservoir (blue), the benchmark CDC (orange), and the benchmark FFE (green). The results are shown for two variations of the KK receiver, both operated under distorting conditions. The left subplot show low samples/symbol operation (4 sps) at a higher CSPR (7 dB), while the right shows low CSPR (5 dB) with higher samples/symbol (6 sps). In both cases, resolving dispersion before the receiver significantly enhances the BER compared to after the receiver. Data points at and above 80 km indicate that the reservoir is out of memory. This happens sooner for the FFE, which is seen to perform worse than the CDC at around 60 km. The dashed grey line indicates the back-to-back results and the dash-dotted black line indicates a FEC threshold of 3.8e-3

Data points at and beyond 60 km indicate that the FFE reaches its memory limit leading to poorer performance compared to CDC. This is also seen to affect the performance of the reservoir, but at around 80 km on the left subplot and 90 km on the right. Despite both reservoirs having the same size, the different reservoir simulations used different internal phases leading to a slight difference in the memory capacity. This is expected, since the variations in phase affect how the signal interferences happen within the reservoir, which in turn leads to variations in how long the signals echo in the network.

The effect of equalization is also demonstrated in the constellation diagrams of Figure 4.7. Constellation diagrams for the three solutions obtained after 70 km, detected by a KK receiver deploying 6 sps and with a CSPR of 5 dB (as in the right subplot of Figure 4.6). The constellation from the reservoir processing (left) indicates better equalization compared to CDC (center) and FFE (right).

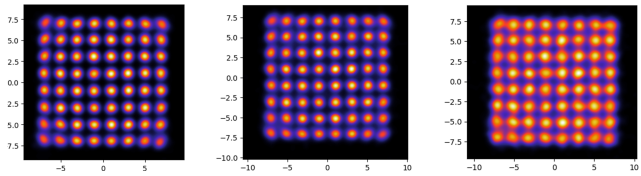


Figure 4.7: Constellation diagrams for the three solutions obtained after 70 km, detected by a KK reciever deploying 6 sps and with a CSPR of 5 dB. The constellation from the reservoir processing (left) indicated better equalization compared to CDC (center) and FFE (right).

Figure 4.8 shows the effect of varying the CSPR on the performance of the reservoir and benchmarks. The results are shown for equalization after a 50 km link using a KK receiver with 4 sps (left subplot) and 6 sps (right subplot). The results show that reservoir-based equalization allows for the use of lower CSPRs, achieving a reduction of over 2 dB compared to digital-based equalization while maintaining the same BER. Given the reliance of the receiver on the CSPR, the CSPR sweep shows at which points the receiver operation becomes linear. This happens at the values where the reservoir performance matches the DSP performance, i.e. there are no added gains from compensating before the receiver.

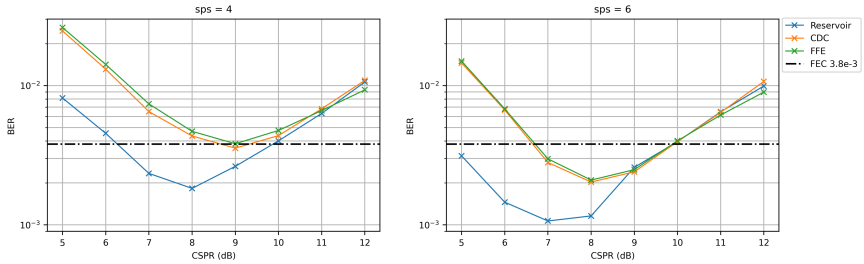


Figure 4.8: BER vs CSPR for the reservoir (blue), the benchmark CDC (orange), and the benchmark FFE (green). The results are shown for equalization after a 50 km link using a KK receiver with 4 sps (left subplot) and 6 sps (right subplot). The dash-dotted black line indicates the FEC limit of 3.8e-3

In addition to equalizing signals, the reservoir is also capable of performing bandpass filtering. This is shown in Figure 4.9, where the reservoir is trained without a bandpass filter in the system, thus forcing its behaviour to adapt to both equalization and filtering. While this penalizes the performance of the reservoir compared to the case where a filter is present, it still shows BER improvements compared to the benchmark CDC pipeline deploying a filter. In this case, the reservoir's dispersion compensation is limited and is seen to reach capacity at 40 km. Better results that extend to higher lengths could probably be achieved with a larger reservoir, which would allow it to have better noise rejection while simultaneously performing dispersion compensation. However, with the available 16 degrees of freedom, a solution for partial dispersion compensation and partial bandpass filtering that achieves the lowest BER is found. For comparison, the effect of no bandpass filtering on the CDC benchmark is also shown, which indicates the necessity of the filter for a good system performance.

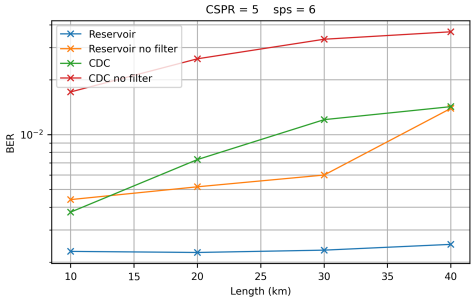


Figure 4.9: BER vs Length for a KK receiver with CSPR of 5 dB and operating at 6 samples per symbol. The results of equalization without a bandpass filter in the system are shown, to study the effect of training a reservoir for equalization and filtering. This is shown in orange, contrasted with the reservoir trained with a filter (blue), as well as the CDC benchmark with (green) and without (red) a filter.

4.4.2 Performance with SiP SSBI-Cancelling Receiver

Finally, we look at the performance of the reservoir with the SiP SSBI-cancelling receiver. The ring is designed with a Q-factor of 44,420 and SSB signals with a CSPR of 5 dB are transmitted in 30 km fibers. As a form of distorting behaviour, we study the influence of a resonance shift in the MRR, where positive and negative frequency shifts are simulated. Figure 4.10 shows the achieved BER performance for the reservoir, CDC, and FFE solutions, in addition to the back-to-back performance of the receiver. All trendlines have an optimized β.

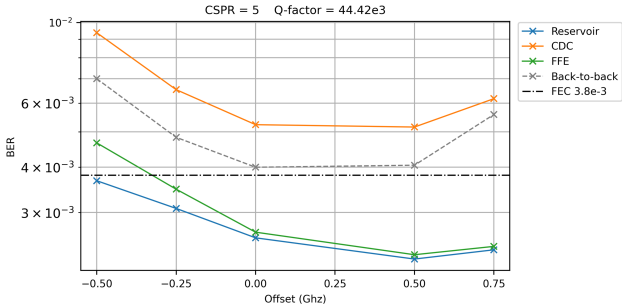


Figure 4.10: BER vs frequency offset for a SiP SSBI-cancelling receiver. The ring is designed with a Q-factor of 44,420 and SSB signals with a CSPR of 5 dB are transmitted in 30 km fibers. Results for the reservoir (blue), the benchmark CDC (orange), and the benchmark FFE (green) are shown, in addition to the back-to-back performance (grey) and the FEC threshold (black).

The results indicate that performing standard CDC after the receiver shows the worst performance. However, trainable equalization, before or after the receiver, offer significant improvements, with the advantage of using the reservoir being more evident at higher distorting scenarios. This allows over 1 GHz of safety margin for the frequency offsets This equalization addresses not only the channel impairments, but also some of the receiver impairments, which results in the equalized system's performance being better than that of the unequalized back-to-back system. This is in contrast to what was shown with a KK receiver, where all solutions were still worse than the receiver's back-to-back. In the following chapter, we show how the KK receiver's imperfections can be addressed with the use of an opto-electric network.

4.4.3 Comparison of Equalization Techniques

Table 4.2 compares the different equalization techniques employed in this chapter across several key features, highlighting their strengths, limitations, and application-specific suitability.

The photonic reservoir supports equalization with low latency and potentially low power consumption, due to its passive photonic implementation. However, its scalability is constrained when compensating for high accumulated dispersion due to the increased complexity of adding more nodes. Similarly, the digital FFE faces scalability challenges as a time-domain solution, becoming inefficient with larger accumulated dispersion. By contrast, the frequency-domain CDC effectively handles high dispersion, making it well-suited for long-haul transmission. However, this advantage comes at the cost of increased latency and power consumption.

The results in this chapter highlight the benefits of the photonic reservoir in scenarios involving receiver distortion or nonlinearity. By compensating for dispersion prior to a distorting receiver, the reservoir improves overall system performance. Its trainability further enhances this advantage by enabling equalization to target multiple error sources, including transceiver imperfections. While digital FFEs also offer this trainability, receiver distortions can complicate the equalization process, requiring optimization techniques and hyper-parameter tuning to determine appropriate weights. Frequency-domain CDC, on the other hand, eliminates the need for such optimization but lacks the adaptability of trainable solutions. This rigidity can be advantageous in highly distorted scenarios, as it avoids the challenges associated with finding weights in environments with significant nonlinearity or distortion.

Feature	Reservoir	Digital FFE	Frequency CDC
Latency	Very low	Low-moderate	High
Power Consumption	(Potentially) low	Moderate (2 samples/symbol required)	Moderate-high (FFT/IFFT procedures required)
Complexity for shorter lengths	Low	Low	Moderate
Complexity for Longer lengths	High	High	Low
Receiver Requirements for optimal behavior	Irrelevant	Fully Coherent or limited receiver nonlinearity	Fully Coherent
Adaptability	Trainable	Trainable	Fixed

Table 4.2: Comparison of the different equalization techniques employed

4.5 Conclusion

In this chapter, we demonstrated the effectiveness of photonic-based reservoir equalization through 64 QAM simulations, employing two self-coherent receivers under varying operational conditions. Specifically, the KK receiver was evaluated under scenarios with low CSPR and low oversampling rates, while the SiP SSBI-cancelling receiver was tested with a shifted frequency response. Given the stringent requirements for achieving full linearity with these receivers, we showed that optical compensation significantly enhances the overall system performance up to transmission lengths of 80 km.This approach enables the receivers to operate under relaxed condition, with over 2 dB of CSPR reduction for the KK receiver and over 1 GHz offset margin for the SiP SSBI-cancelling receiver, while still maintaining BERs below the FEC threshold.

5

Opto-Electronic Network for KK Receiver Equalization

Parts of this chapter were published in [144].

5.1 Overview

KK receivers support coherent detection through a very simple hardware architecture comprising a single PD. However, ensuring linearity within this receiver necessitates stringent operational conditions. Namely, a high CSPR value is required alongside 3-fold oversampling (6 sps). These conditions underscore the tradeoffs between hardware complexity and processing requirements of self-coherent receivers, which affects their practicality in short-haul networks.

Furthermore, essential DSP operations, including equalization and forward-error correction, are operational prerequisites that contribute to the overall complexity and cost.

In this chapter, we explore using opto-electronic machine learning networks to ease the CSPR and the sps constraints of KK receivers, such that they are more applicable to short-haul systems. The network comprises a photonic reservoir that is followed by a trainable electronic feed-forward equalizer. This trainable opto-electronic network sandwiches the KK receiver and is used to compensate for nonlinear errors arising from a KK receiver operated under severely limited receiver resources. Namely, the receiver is operated using 3 sps instead of the required 6 and at CSPRs below 9 dB. These conditions, as was shown in Chapter 3, contribute to significantly nonlinear receiver behaviour. Numeric results demonstrate that this opto-electronic architecture shows significant improvement in receiver linearity, which allows for better integration in a digital signal post-processing pipeline for chromatic dispersion compensation (CDC) and blind phase search (BPS). Importantly, the optical reservoir replaces a bandpass filter before the KK receiver, which is typically required in self-coherent systems. Furthermore, the training is done on a back-to-back system, without requiring knowledge of the link parameters, allowing the network to be inserted in a "plug-and-play" manner in varying links. We target short-reach systems of up to 250 km deploying 64 QAM, and show up to 4 times reduction in BER in simulations.

Efforts in utilizing machine learning networks to aid self-coherent receivers have also been reported in other works, albeit utilizing fully digital neural networks. For example, authors in [145] report the use of a convolutional neural network that can emulate the behavior of the KK algorithm while lowering its CSPR requirements. Deep neural networks have also been shown to lower the complexity and the required CSPR of the Gerchberg-Saxton receiver [146, 147].

Compared to previous chapter, this chapter utilizes an *opto-electronic* network for equalizing *receiver-based* distortions, as opposed to a *fully optical* network for equalizing *channel-based* distortions. Furthermore, as mentioned above, the opto-electronic network is trained on a back-to-back transceiver system such that it can be inserted in various links and seamlessly cascaded with a DSP pipeline for channel impairments. This arrangement splits the responsibilities and allows the network to be utilized for transceiver impairments, while channel impairments like dispersion are deferred to the DSP.

5.2 Opto-Electronic Network Architecture

The opto-electronic network architecture is composed of a 24-node photonic reservoir, with all nodes connected to the readout, co-trained with a 16-tap electronic feed-forward equalizer (FFE). The photonic reservoir replaces the optical bandpass filter normally required before a KK receiver and hence is not an additional component in the system. Both the reservoir and FFE are composed of generic delay and weighting elements realized either in the optical or electrical domain, and are therefore simple to implement.

Naturally, the photonic part of the network is placed before the detector, while the FFE part comes after the detector. The algorithmic portion of the KK receiver can then be performed before or after the FFE block. As such, two possible architectures can be constructed and these are shown in Figure 5.1. Note that the algorithmic portion of the KK receiver denotes all the blocks in the KK pipeline of Figure 3.7 that follow the detector, including the carrier removal and the frequency shifting.

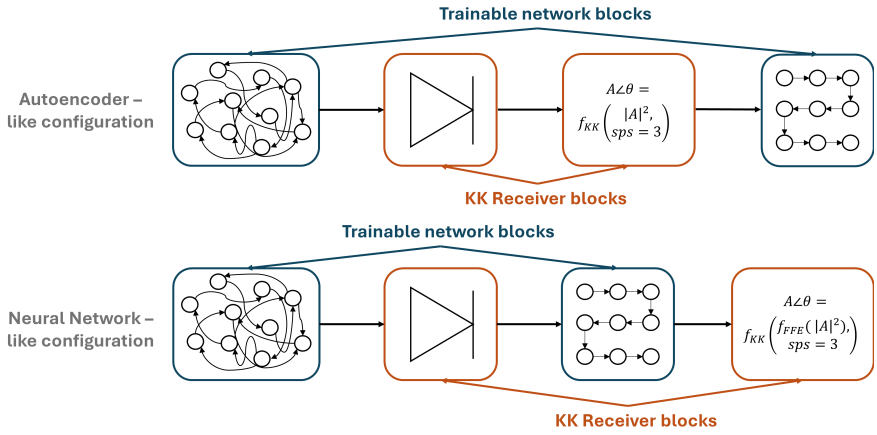


Figure 5.1: Two configurations for the opto-electronic network with the KK receiver. The network is composed of two trainable blocks (blue), namely the photonic reservoir (always preceding the photodiode) and the electronic FFE. The KK receiver consists of two blocks (orange), namely the photodiode and the algorithmic steps shown in Figure 1. The complex-valued output of the receiver in the top configuration (autoencoder-like) is a direct function of the raw measured amplitude and the sps. In the lower configuration (neural network-like), the measured amplitude undergoes an FFE transformation before it continues to the algorithmic portion of the receiver.

In the top configuration, the FFE portion of the trainable network comes at the end of the full KK receiver. The idea is to create a network where a photonic pre-distorting filter imposes meaningful distortions on the signal, making it more

robust to the KK receiver's nonlinearity. Then, an FFE post-filter is used to negate these distortion effects and reconstruct the message signal. To better enable a symmetric behavior between the reservoir and FFE, we use a single input node in the reservoir. However, the input node is chosen in the middle of the reservoir to allow for richer reservoir dynamics. Note that the reservoir can use the KK receiver's behavior as a nonlinear kernel. Both the FFE and reservoir have interconnections that introduce half symbol-period delays. This architecture bears similarity to equalization autoencoders used in end-to-end learning of fiber optic systems [148].

In the lower configuration, the KK receiver pipeline is broken up by the FFE block, where the FFE is performed after the detector, but before the KK receiver algorithm. This configuration separates the KK receiver nonlinearities such that every linear trainable part of the network is followed by a nonlinear KK receiver transformation. The reservoir's delay lines are designed to introduce half symbol-period delays, while the FFE effectively delays the signal by 40 percent of the symbol period in every tap. This is because the ADC samples the signal at 2.5 samples/symbol, and the FFE granularity is set to delay 1 sample per tap. Because we aim to enhance the reservoir's dynamics, three input nodes are used: one from the top left corner, one in the center, and one from the bottom right nodes. All nodes receive a copy of the same input signal. This configuration bears resemblance to a two-layer neural network where linear layers are followed by nonlinear activation functions.

5.3 Training and Testing Systems

For training the weights in both configurations, a back-to-back system is used to ensure that only the KK receiver errors are compensated for. This facilitates arriving at a generic solution that linearizes the KK receiver without being trained on a specific channel length. Then, the resulting architecture can be inserted in a plug-and-play manner into any channel length, provided that the transceiver operational conditions remain fixed and the channel impairments are compensated for using DSP modules. In addition to reducing the number of times the system needs to be trained, this method also simplifies the training process, because the DSP pipeline is not traversed in every training step.

The system schematic is shown in Figure 5.2, where the training pipeline is seen in (a) and the testing pipeline in (b). VPI transmission suite [138] is used to simulate the behavior of the transmitter and the fiber, while the receiver and trainable network are simulated using PyTorch and the photonic library based on it, Photontorch [110]. The figure shows that three different configurations are deployed for both the training and testing pipelines. These are the two network configurations shown in Figure 5.1 in addition to a benchmark configuration.

For the benchmark, an ideal bandpass filter before the KK receiver is used to filter out the noise pre-detection taking over the role of the trainable photonic reservoir. The benchmark pipeline also includes a trainable 16-tap FFE equalizer. This is done to showcase that the added benefits of the proposed solutions do not come from linear compensations like those achievable by a linear equalizer, despite our trainable network elements being linear. Rather, the gains from our solution are achieved through leveraging the entire pipeline to arrive to the autoencoder-like or neural network-like behaviors described in Figure 5.1.

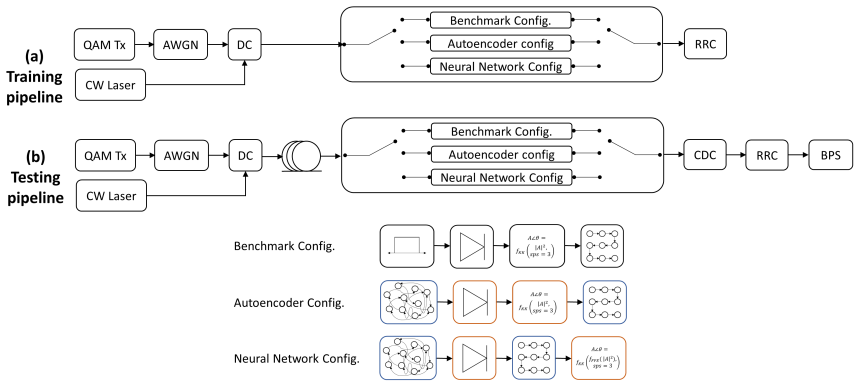


Figure 5.2: Schematic of (a) training system and (b) testing system. Systems are simulated three times: with the autoencoder-like configuration, the neural network-like configuration, and a benchmark configuration. The autoencoder-like and the neural network-like configurations follow the description detailed in Figure 5.1; the benchmark configuration utilizes an optical bandpass filter before the photodiode and an FFE block after the full KK receiver. ***Tx****: Transmitter –* ***CW*** *laser: Continuous wave laser –* ***AWGN****: Additive white gaussian noise* ***DC****: Directional coupler* ***CDC****: Chromatic dispersion compensation –* ***RRC****: Root-raised cosine filter –* ***BPS****: Blind phase search*

The transmitter simulation parameters are detailed in Table 5.1. The setup employs a 64-QAM modulator transmitting at a rate of 64 Gbaud. This data stream modulates an optical carrier with a linewidth of 200 kHz. The signal's average power remains fixed at 3 dBm, and an OSNR of 27 dB is achieved through gaussian noise loading. Subsequently, an unmodulated CW laser, positioned approximately 2 GHz away from the message signal's left edge, is optically generated and combined with the signal. The CW laser's power level is varied to achieve CSPRs within the range of 5 dB to 8 dB. The KK receiver algorithm is operated at 3 samples per symbol, which is achieved by upsampling from the ADC's 2.5 samples.

Parameter	Value
Symbol rate	64 Gbaud
OSNR	27 dB
Signal power	3 dBm
Pulse shape	RRC, 0.01 rolloff
Carrier linewidth	200 kHz
CSPR	5 - 8 dB
Guard band	2 GHz

Table 5.1: Transmitter simulation parameters. Note that signal power is measured without the reference carrier's power

In the testing pipeline, standard single-mode fiber spanning distances from 20 km to 250 km is used (Table 5.2). Standard DSP blocks are implemented to execute CDC (as described in Chapter 4) and BPS [149]. Note that the training pipeline excludes a BPS block as the laser linewidth was set to 0 Hz during training. This expedites the training process by eliminating the need for blind phase searches in each training epoch. Training details are provided in the next section.

Parameter	Value
Length	20 - 50 km
Attenuation	0.2 dB / km
Nonlinear index	$2.610^{-20}\mathrm{m}^2/\mathrm{W}$
Dispersion parameter	16 ps / (nm km)

Table 5.2: Fiber simulation parameters

5.4 Training the Networks

The training data set is composed of over 16,000 symbols that are generated using a Wichmann-Hill generator [142]. In the first training stage, the network parameters are initialized in preparation for backpropagation. Initialization of the network parameters is typically an important aspect when backpropagation is used, since the presence of multiple local minima can prevent the optimizer from finding the optimal solution and instead get stuck at suboptimal ones. For example, in our networks, a random initialization followed by backpropagation continuously underperformed the benchmark. As such, Linear Regression (LR) is used to find an initial set of weights for the reservoir weights as in Equation 2.7.

It turns out that this first linear regression step conditions the reservoir to behave as a bandpass filter, which is necessary to avoid signal-noise beating in the detector.

Indeed, it has a similar performance as the benchmark configuration, where such a bandpass filter is explicitly implemented. Initializing the reservoir for this behavior allows starting at the benchmark performance, where noise is filtered from the signal, before attempting to optimize this solution further.

Next, we address the initialization of the FFE weights. The straightforward initialization for the FFE weights is setting all the weights to 0, except the center tap, which is set to 1. This initialization is used for the autoencoder-like configuration. For the neural network-like configuration, we were concerned that the first network layer (i.e., the reservoir) would have a more advanced initialization than the second layer. Since the FFE in this case is also followed by a nonlinear receiver transformation, it might not be able to take advantage of this if the initialization is near a suboptimal local minimum. As such, we used the genetic algorithm CMA-ES as described in Chapter 2 to find a new FFE initialization point. To allow flexibility, the algorithm was also able to vary the LR initialization of the reservoir weights. However, we did not exhaustively investigate the influence of this approach.

After initialization, backpropagation is used to alter the optical reservoir weights as well as the electric FFE weights. The same procedure is followed for both configurations. A gradient descent optimizer with momentum [150] is used to minimize the MSE loss between the signal and the target symbols, where MSE is given by Equation 2.3.

However, since the reservoir behaves as a trainable filter, a solution that backpropagation can arrive at is one where the signal power is attenuated with respect to the carrier. While this would lead to a higher CSPR and thus to MSE reduction, this is an undesired effect, since it achieves performance improvements without leveraging the network's behavior and at the expense of signal attenuation. To prevent this from happening, we again introduce $\mathrm{err}_{\mathrm{CSPR}}$ (as in Equation 4.3) in the error function. This term heavily penalizes growth of CSPR, while also being neutral to its reduction. The training error is then a sum of the CSPR and MSE errors as described by Equation 5.1.

$$\mathrm{error}_{\mathrm{training}} = \mathrm{error}_{\mathrm{CSPR}} + \mathrm{MSE} \tag{5.1}$$

As in the previous chapter, this loss function should make sure that at the end of the training, $\mathrm{CSPR}_{\mathrm{reservoir}}$ is less than or equal to $\mathrm{CSPR}_{\mathrm{transmitted}}$. The hyperparameters are also optimized such that the behavior of the softplus function is adjusted to appropriately penalize $\Delta\mathrm{CSPR}$ with respect to MSE. This ensures that the MSE contribution is still significant and optimized for during training. Note that due to the highly nonlinear behaviour of the receiver, the resulting error is very sensitive to adjustments in the reservoir weights. As such, the errors of the signal are computed using MSE as opposed to the higher order error function in the previous chapter which would otherwise make the training highly unstable.

To escape potential local minima, the optimizer's learning rate is chosen such that it is a bit on the high side, causing $\mathrm{error}_{\mathrm{training}}$ to temporarily and sporadically increase during training. This is because there are two error terms, and allowing the optimization algorithm to explore options where the CSPR is temporarily increased allows arriving at better overall solutions. Naturally, the learning rate should not be too high, as the error can then spiral out of control, reaching large values and getting stuck at a local minimum.

To showcase the training dynamics, a segment of the training progress is shown in the Figure 5.3. The plot shows $\mathrm{error}_{\mathrm{training}}$ increasing momentarily during the training before decreasing again. Note that overfitting is not commonly a problem in these networks, since the network's nonlinearity is limited. As such, we rely on the training error and do not use a validation set.

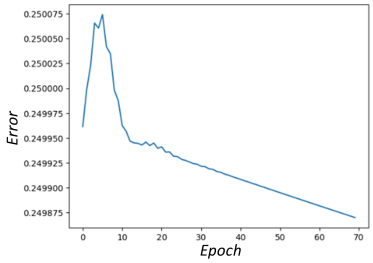


Figure 5.3: A segment of the training error briefly increasing before decreasing again.

For the benchmark pipeline, the only trainable block is the FFE, since the optical bandpass filter has fixed parameters. The FFE weights are first initialized with zeroes, except for the center tap, which is initialized with 1. Then, backpropagation is tried to optimize the weights using the MSE error function of Equation 2.3. (Note that monitoring the CSPR is irrelevant in this configuration, since the optical bandpass filter is not trainable and will not change the CSPR.) However, this backpropagation contributed to no improvements compared to the performance of the initialized weights. On the contrary, it worsened the overall BER, since optimizing for MSE does not necessarily lead to better BER (recall that backpropagation cannot use BER as an error function, since it is a non-continuous function). To rule out the possibility of being stuck in a local minimum due to our initialization, we switched to finding FFE weights through CMA-ES using BER as a metric instead of MSE. However, this training resulted in less than 3 percent improvement compared to the initialization. This shows that there are little to no gains achievable through standard linear equalization. The benchmark results reported in the next section are based on the CMA-ES solution for the FFE block.

5.5 Results and Discussion

The training BER achieved in back-to-back systems is shown in Figure 5.4 for CSPR values between 5 and 8 dB. For each CSPR, a different reservoir with random input and internal phase variations is simulated and a different dataset is also used. Both proposed network configurations had substantially lower BERs compared to the benchmark, with the autoencoder-like configuration outperforming the neural network-like configuration.

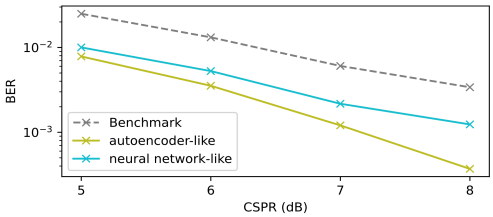


Figure 5.4: Back-to-back BER measured during the training phase for CSPRs ranging between 5 and 8 dB and at 3 sps. The two proposed network configurations are compared against the benchmark pipeline.

For the testing dataset, over 32000 symbols were generated using a Whichmann-Hill generator, seeded at random and using different seeds than during training. A different test set was also used for every CSPR value. Figure 5.5 shows the achieved test BER against fiber lengths spanning between 20 and 120 km for three CSPR values. Results for the autoencoder-like configuration are shown on the left subplot and results for the neural network-like configuration are shown on the right subplot. Results from the proposed networks are shown in solid lines, contrasted against results from the benchmark configuration, which are shown as dashed trend lines. Each color corresponds to a different CSPR.

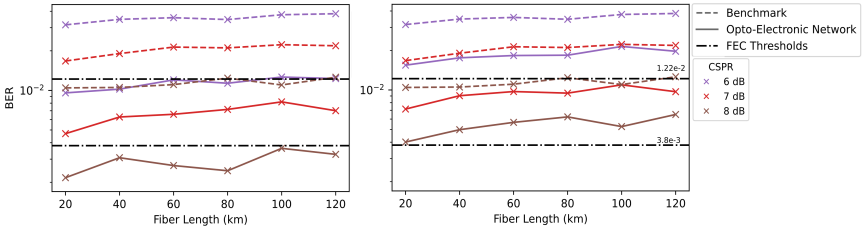


Figure 5.5: Results for the testing set showing BER vs fiber length for the autoencoder-like configuration (left) and the neural network-like configuration (right). Results from the opto-electronic networks are shown as solid lines, while results from the benchmark are shown as dashed lines. Dash-dotted lines correspond to the FEC thresholds. The colors refer to different CSPRs ranging between 6 and 8 dB.

When compared to a back-to-back system, the fiber channel in the testing pipeline adds linear errors that are nonlinearly transformed due to the receiver and therefore are incompletely compensated using standard CDC. As such, the errors of the testing pipeline are substantially higher than those observed in the training configurations (which we recall used a back-to-back configuration). Note that the nonlinear Kerr effect in the fiber is negligible at the power levels investigated in this chapter. As the opto-electronic networks enhance the receiver's linearity, lower BERs are achieved, with around 3-4 times improvement seen from the autoencoder-like configurations and around 2-3 times improvement from the neural network-like configuration. Furthermore, Figure 5.5 indicates that similar BER results were achieved by the autoencoder-like configuration at 2 dB lower CSPRs compared to the benchmark, while a 1 dB improvement was contributed by the neural network-like configuration. These improvements are consistent over the range of lengths investigated.

Compensation is also evident from the constellation diagrams in Figure 5.6, which show signals with a CSPR of 7 dB after traversing 120 km for the benchmark pipeline (left), the autoencoder-like configuration (center), and the neural network -like configuration (right).

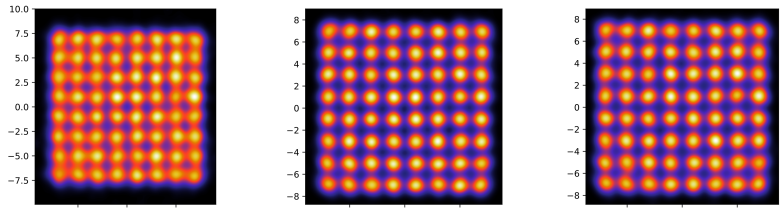


Figure 5.6: Constellation diagrams for CSPR 7 dB after 120 km when detected by the benchmark pipeline (left), the auto-encoder like configuration (center), and the neural network-like configuration (right).

Figure 5.7 shows the impact of evaluating the system at a fiber length of 250 km, i.e. roughly twice as long as the maximum of Figure 5.5, so that dispersion accumulation becomes more pronounced. A notable increase in errors is observed at 5 dB and 6 dB, attributed to the failure of the BPS in accurately compensating for the phase noise. While in theory the severity of the phase noise should remain unaffected by the length of the fiber, in practice however, when the residual errors caused by dispersion are significant, the BPS algorithm will be suboptimal and achieve a poor phase compensation. To potentially address this issue, a larger averaging window might improve performance, but this enhancement would demand increased computational resources and memory allocation.

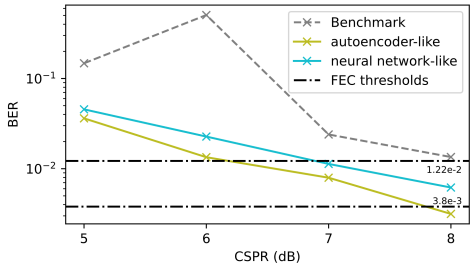


Figure 5.7: BER vs CSPR after 250 km transmission for the opto-electronic networks and the benchmark. A BER spike is seen in the benchmark pipeline due to residual phase error.

We then explore the connection between our networks' performances and the receiver's nonlinearity, aiming to discern the extent to which system performance relies on this nonlinearity. The receiver's nonlinear behavior is exhibited at the detector when CSPR levels are low, as well as at the algorithmic portion when the samples per symbol are low. While the CSPR value is an aspect of the signal itself, the sps is easily adapted on the receiver's end. As such, we study the effect of changing the sampling rate of the KK receiver after the networks have been trained for operation with 3 sps, and do this for two CSPR levels.

Figure 5.8 shows the BER vs sps of the testing set at 250 km for both configurations. The sps is varied between more nonlinear (i.e., 2.5 sps) and more linear (i.e., 4 and 6 sps) receiver operation. Note that the lower sps system can exhibit highly nonlinear behavior which, as seen in Figure 5.7, may lead to residual phase noise and thus a high BER. As such, to ensure the results of Figure 5.8 reflect the effects of changing the nonlinearity of the receiver, we set the linewidth of the laser to zero. The left subplot of Figure 11 shows the performance of the systems when the CSPR is 6 dB, and thus the overall receiver is more nonlinear compared to the right subplot with a CSPR of 8 dB.

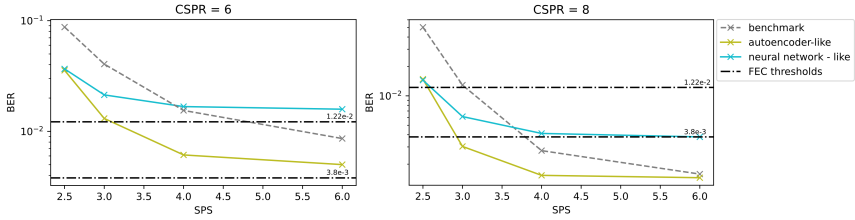


Figure 5.8: BER vs SPS for testing pipelines at 250 km. Networks were trained to operate at 3 sps but were tested at sps values between 2.5 (more nonlinear) and 6 (more linear). Both 6 dB CSPR (left) and 8 dB CSPR (right) are shown.

Both network configurations at both CSPRs continue to have good performance at 2.5 sps, resulting in errors 2 to 5 times lower than the benchmark. However, as the sps increases above 3, the neural-network like configuration becomes distorting to the signal and thus performs worse than the benchmark. This is more nuanced when the CSPR is higher, and thus the overall receiver behavior is more linear. In contrast, the autoencoder-like configuration continues to perform better than the benchmark as the sps increases. However, similar to the neural network-like configuration it too deteriorates in performance as the CSPR increasing. These results indicate that the neural network-like configuration is very dependent on the nonlinear behavior stemming from the receiver's sampling rate. The autoencoder-like configuration on the other hand continues to achieve better performance than the benchmark, as long as a portion of the KK receiver (i.e., the CSPR) remains nonlinear.

When constructing the two different configurations, several aspects were varied, including the placement of the FFE block, the input nodes of the reservoir, the delay line lengths of the FFE block, and the initialization before backpropagation. These factors make it hard discern which aspect or combination of aspects led to the difference in achievable BERs. Nonetheless, the results of both networks indicate that significant BER gains are achievable through leveraging the autoencoder-like and neural network-like behaviors.

Moreover, the opto-electronic networks reduce the complexity of the KK receiver, especially at higher baud rates, by halving the system's sampling rate, thereby streamlining computationally intensive operations. This complexity reduction is notable even with the incorporation of the additional FFE block, since it effectively functions at low sampling rates and can also be realized in an analog form.

The reservoir and FFE node count were chosen to offer increased degrees of freedom and higher programmability, although further investigation could focus on minimizing the network costs. These nodes, alongside their programmability, introduce memory that could potentially compensate—at least partially—for dispersion. Although this chapter does not explore this aspect, training the reservoir to perform several functions has been explored in previous chapters.

Finally, we address two limitations of the opto-electronic network. First, since the photonic reservoir replaces an optical bandpass filter, the network training optimizes the reservoir's behavior to achieve suitable cutoff and roll-off characteristics. Consequently, changes in the system's noise profile may necessitate new training. Second, the performance remains limited when compared to a fully coherent receiver. Since the opto-electronic neural networks have linear architectures, they are not comparable to a fully capable machine learning network. Rather, our approach offers a simple alternative designed to boost the performance of a less capable system by harnessing the nonlinear functions inherent in the receiver.

5.6 Conclusion

This chapter discussed the use of an opto-electronic network composed of an integrated photonic reservoir and an electronic feed-forward equalizer for the linearization of a KK receiver operating at 3 sps and CSPRs below 9 dB. Simulation results showed up to a 4x reduction in BER on links of 250 km and shorter, without explicitly training the reservoir on these channels and with seamless DSP integration. The opto-electronic network components are fully linear, which makes them easily realizable, and they rely on the receiver's nonlinearity to achieve the desired behavior.

6

Experimental Reservoir Equalization with Coherent Detection

Parts of this chapter were published in [151] and [152].

6.1 Overview

The growing demand for higher data rates has led to significant interest in new technologies that can enhance the capacity and efficiency of short- and mid-reach optical communication systems. Meeting these demands depends on the system constraints, however, maximizing the channel capacity per wavelength, rather than expanding the wavelength dimension, is generally a desired approach. Such efforts

are not only beneficial for the very short-term evolution, but could also be interesting for the still uncertain but upcoming generations [21]. Besides increasing capacity, coherent communication also offers increased receiver sensitivity, making it a particularly attractive solution. As such, there is substantial interest in developing a new generation of energy-efficient coherent links designed specifically for systems like data center applications and access networks [30]. However, a key issue to address remains the signal processing requirements of these receivers, which are typically realized through power-intensive digital implementations. Therefore, alternative methods that may reduce the power consumption and bypass the bandwidth limitations of digital processing are investigated, with special focus on analog-based solutions.

Some of these signal processing blocks are traditionally performed using two samples/symbol, and are therefore prime candidates for analog replacements, since there is an elevated power consumption associated with their oversampling requirements. Indeed, two such blocks are chromatic dispersion compensation and active equalization (which addresses residual transceiver impairments and performs polarization demulitplexing). The former consumes an estimated 15% of the total DSP power, while the latter can account for around 50% of the power consumption [21].

To this end, in this chapter, we experimentally demonstrate the use of an analog solution, realized in an optical signal processing PIC, to relax electronic processing needs of coherent transceivers deploying 4 and 16 QAM. This not only allows harnessing the large bandwidths of optics for processing signals, but also naturally supports one-shot processing of both quadratures. The processor, a photonic integrated reservoir, can perform equalization tasks and therefore replaces digitally implemented equalizers for chromatic dispersion and transceiver impairments. Moving these steps, which normally operate at twice the baud rate, into the optical domain considerably reduces the computational complexity of the pipeline and the bandwidth of the electronics.

In these experiments we pair our network with standard DSP blocks performing other necessary procedures including frequency offset compensation and blind phase search, thus showcasing that our network can seamlessly integrate with existing DSP pipelines. This modular approach allows gradual migration into an optical co-processing solution, before moving to fully optical processing or optical-analog hybrids.

6.2 Experimental Setup

The experimental setup is shown in Figure 6.1 and the system parameters are summarized in Table 6.1. On the transmitter's side, a dual-polarization IQ modulator was used to modulate a continuous-wave (CW) laser emitting at 1550 nm by a randomly generated data stream. The data stream was composed of either 4 or 16 QAM symbols and was pulse shaped using an RRC filter with a roll-off of 0.2. The data transmission rate was set to 28 Gbaud, which was imposed by the length of the reservoir's delay lines as discussed in Chapter 2. The output of the transmitter was amplified by Amplifier 1 to 23 dBm, and then attenuated by a variable optical attenuator that controlled the desired input power into the fiber. This setup allowed sweeping the fiber's input power without changing the noise profile of the amplifier, thus ensuring a fairer comparison between different situations. The signals then traversed 20 km of standard single-mode fiber.

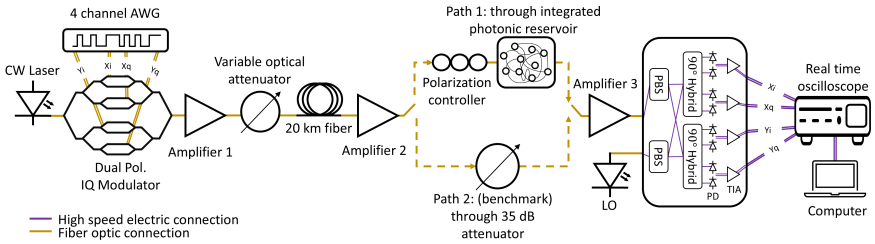


Figure 6.1: Experimental setup with two possible paths: path 1 goes through the reservoir and a polarization controller, which is used to collect readout states on a node-by-node basis. Path 2 is a benchmark which replaces the reservoir with a 35 dB attenuator. All other components shown are common to both paths. CW: continuous-wave laser, AWG: arbitrary waveform generator, VOA: variable optical attenuator, LO: local oscillator, PBS: polarizing beam splitter, PD: photodiode, TIA: transimpedance amplifier

Parameter	Value
Carrier Wavelength	1550 nm
Modulation Format	4 / 16 QAM
Pulse Shape	Root Raised Cosine
Roll-Off Factor	0.2
Fiber Input Power	2 – 15 dBm
Fiber Length	20 km
Receiver Input Power	0 dBm
Sampling Rate	64 Gsamples/s

Table 6.1: Experimental Parameters

As discussed in Chapter 2, significant losses were incurred in this 4-port reservoir generation, ranging between 35 dB and 50 dB per node. Thus, further amplification before detection was needed, and this was provided by EDFAs. This amplification was divided between Amplifier 2 (before the reservoir) and Amplifier 3 (after the reservoir) such that the input power to the receiver was maintained at 0 dBm. Furthermore, the serial readout scheme, as shown in Figure 2.7, was used due to the low power budget available. This means that signals from the output of each reservoir node were sequentially detected and stored for offline processing.

The receiver used was a dual-polarization coherent receiver with transimpedance amplifiers (TIA). Its architecture is thus similar to that of Figure 3.4, duplicated however for the second polarization. Polarizing beam splitters for the input signal and the LO are also integrated as part of the receiver input. The detected signals were then sampled by a real-time oscilloscope (RTO) at 64 Gsamples/s and stored on a computer for offline processing. Despite the dual-polarization setup, only one polarization was used for transmission. This is due to the polarization selectivity of the SiN chip, where the grating couplers and waveguides on chip are highly lossy for TM waves. Therefore, the data stream that drove the IQ arms of the transmitter's Y (or TM) polarization was loaded with zeros. The Y polarization modulators were also biased at the null point to suppress the carrier, thus leaving the output of the Y polarization arm on the transmitter's side negligible.

Nonetheless, polarization rotations that occur throughout the transmission system cause the signal to arrive at the chip's input with a random state of polarization (SOP). As such, a polarization controller is used to align the input signal with the TE polarization by maximizing the output power from the reservoir. If the signal is not polarization rotated before coupling into the chip, power losses from the loss of the TM polarization may be prohibitive. At the receiver's end, the signal also arrives with a random SOP due to rotations, but the signal is detected on both polarization branches of the receiver.

To assess the equalization of the reservoir, its performance was compared to a benchmark system that relied solely on a DSP pipeline. This benchmark setup used a 35 dB attenuator instead of the reservoir, such that roughly the same amplifier signal-to-noise ratio (SNR) behaviour was maintained for both systems. As shown in Figure 6.1, measurements on the reservoir and the benchmark were done consecutively, by connecting either the attenuator or the reservoir to the setup. The benchmark data collected was also processed offline. The DSP details of these processing steps are shown in Figure 6.2 and are discussed in the following sections.

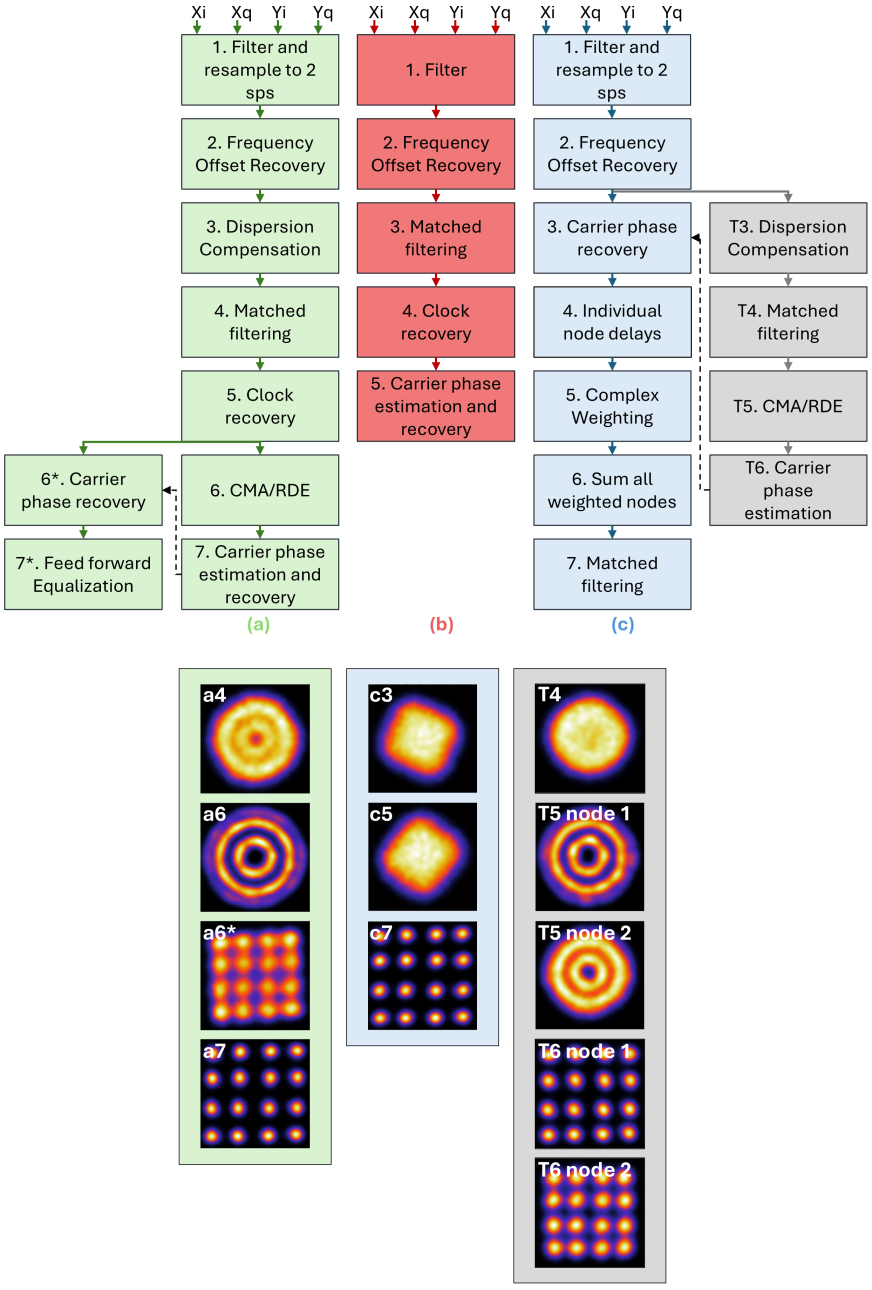


Figure 6.2: Processing pipelines of [a] benchmark system, [b] RC with parallel readout, [c] RC with serial readout. On the bottom, constellation diagrams after key steps are shown, where the first letter refers to the pipeline and the number refers to the step. Blocks around constellation diagrams are also color coded to match the pipeline they originated from.

6.3 Processing Benchmark Data

The DSP pipeline used for the benchmark system is shown in Figure 6.2a [153]. Steps 1 through 7 constitute the main pipeline, with an alternative path through steps 6* and 7*. A discussion of the details and relevance of this alternative path is deferred to Section 6.6.

In the main DSP pipeline, two equalizers are deployed. The first is an ideal chromatic dispersion compensating (CDC) filter used to compensate for the incurred dispersion. The second is an active equalizer that has a 2x2 multiple-input multiple-output (MIMO) architecture and is used to equalize residual linear impairments, including transceiver imperfections, as well as to perform polarization demultiplexing. Furthermore, in long-haul transmission, this equalizer is used for polarization mode dispersion compensation. For its implementation, four 17-tap finite-impulse response (FIR) filters are used, two for each polarization, as shown in Figure 6.3. Each of the four FIR filters is implemented through a T/2 fractionally-spaced architecture like that of Figure 4.5.

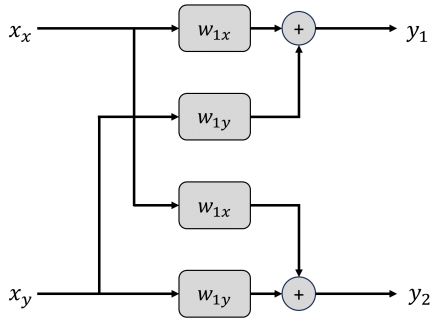


Figure 6.3: Block diagram of the 2x2 multiple-input multiple-output equalizer. Four 17-tap finite-impulse response filters are used, denoted by their weight matrices w_{1X}, w_{2X}, w_{1Y}, and w_{2Y}. The complex-valued input signals X_X and X_Y are sent to the equalizer to generate compensated and de-multiplexed outputs y_1 and y_2

The two MIMO outputs obtained are given by [22]

$$y_1[k] = w'_{1X}[k]X_X[k] + w'_{1Y}[k]X_Y[k]$$

$$y_2[k] = w'_{2X}[k]X_X[k] + w'_{2Y}[k]X_Y[k]$$

where $X_X[k]$ and $X_Y[k]$ are matrices sized $17 \times N$ that originate from the FIR filters, with N being the number of samples. The subscripts X and Y denote which polarization gave rise to the respective matrix. Four 17-element weight vectors multiply these matrices, namely w_{1X}, w_{2X}, w_{1Y}, and w_{2Y},. The subscripts 1 and

2 indicate which output this weight vector will yield, while the subscripts X and Y indicate which polarization matrix it will multiply. Outputs $y_1[k]$ and $y_2[k]$ are the two polarization demultiplexed and linearly compensated signals.

Since both $X_X[k]$ and $X_Y[k]$ carry data from a single transmitted polarization, with proper initialization, the polarization demultiplexing results in weights w_{2X} and w_{2Y} converging to yield an output $y_2[k]$ that is 0. For output $y_1[k]$, w_{1X} and w_{1Y} converge to yield a linearly compensated signal.

To find these weight vectors, a Constant Modulus Algorithm (CMA) is used for 4-QAM. For 16-QAM, CMA is first applied in a pre-convergence stage, followed by a Radius Directed Equalization (RDE) algorithm [154]. These are iterative and blind optimization algorithms that attempt to minimize the radial distance between the distorted signal and the most likely radius from a 4- or 16- QAM alphabet.

The rest of the pipeline involves DSP procedures needed to track and compensate time-varying issues. Namely, the frequency offset between the CW carrier and the LO is found through raising the signal spectrum to the fourth power [153]. The signal is then frequency-shifted based on the generated frequency tone. Furthermore, the phase and frequency offsets of the real-time oscilloscope's (RTO) sampling clock are compensated in a clock recovery algorithm, where a Gardner timing error detector [155] is used for estimation and an interpolator for compensation. Finally, the signals are compensated for the carrier phase shifts through a BPS algorithm [149].

6.4 Processing Reservoir Data

The photonic reservoir supports the migration of expensive and optically implementable equalization blocks to the photonic realm. As such, both CDC and active equalization procedures are replaced by the photonic RC. This migration offers considerable speed and power advantages, since both these operations happen in one shot, at the speed of transmission, and bypass the need for complex-valued digital operations at twice the baud rate. On the other hand, DSP procedures correcting for the time variable errors described in the previous section are still necessary. This results in an opto-electronic, co-integrated solution with low DSP requirements, as seen in the red pipeline in Figure 6.2b. These DSP procedures are implemented when the photonic reservoir when is operated in the parallel readout scheme of Figure 2.7. However, since the use of the parallel readout was not possible due to excessive losses in the current generation, the DSP pipeline had to be adapted temporarily to accommodate for the sequential nature of the measurements in the serial readout.

This adjustment is shown in Figure 6.2c. In this case, all time-varying issues must be compensated for each node individually, since each node would have incurred different values for these issues. However, DSP procedures for these compensations are devised for input signals of certain criteria. For example, the Gardner timing error detection algorithm requires clear transitions between symbols for proper operation, and as such is typically implemented after CDC. Naturally, with readout signals that are formed from dispersed signals that interfere within the reservoir, clock recovery is very challenging. As such, this step is omitted from the pipeline, and its implication will be discussed later as part of the timing ambiguities discussion in Section 6.5.

Similarly, the carrier phase estimation by BPS requires an equalized signal before the algorithm is applied and as such is normally the last block in the chain. However, with a serial RC readout, each node would have different time-varying phase rotations and these must be compensated before the nodes are summed. Since these phase rotations cannot be extracted immediately from the uncompensated node data, a temporary chain of DSP blocks, greyed in Figure 6.2c, is implemented. In this pipeline, the node data are dispersion compensated and then equalized through the same MIMO equalizer discussed in the previous section. This equalizer can perform linear equalization and can therefore undo the reservoir mixing, thus retrieving – to some extent – a clean constellation. After the phase offset array is obtained using BPS, all these temporary DSP signals are discarded. This phase offset array is then used to compensate the unprocessed node data. We stress again that the entire pipeline of Figure 6.2c is only needed for the temporary fallback serial readout scheme, and not for the ultimate parallel scheme that should be facilitated by our next chip generation, which will utilize pipeline 6.2b.

To further clarify the role of the DSP blocks in the temporary reservoir pipeline, Figure 6.2 also shows several 16- QAM constellation insets after key steps for the different pipelines, where the first letter refers to the pipeline and the number refers to the step. Blocks around constellation diagrams are also color coded to match the pipeline they originated from.

Insets [a4] and [T4] in Figure 6.2 show signals after the matched filtering step for the standard pipeline 3a and the temporary pipeline. Constellation [a4] is to be expected at this stage of the DSP, where we can note distinct power levels, seen as data points distributed on two noisy radii (the third is obscured in the noise). The constellation still shows linear distortion and phase rotation, but those are compensated respectively through CMA/RDE as seen in [a6] and BPS as seen in [a7]. (inset [a6*] is related to an alternative DSP pipeline that will be discussed later.)

On the other hand, constellation [T4] in Figure 6.2 does not show a similar radial behaviour, which is a result of the signal mixing in the reservoir. Since the mixing is linear in nature, it can be partially undone by linear equalization using the CMA/RDE algorithms. This is seen in insets [T5], which show two different nodes after equalization. While the first node seems to have been entirely compensated, the latter still exhibits distortions indicating that the linear unmixing of the CMA/RDE algorithm is not fully successful. Nonetheless, both can then be sent to the BPS algorithm, which extracts the time-varying phase rotation arrays to yield the constellations in [T6]. These steps are then discarded and only the resulting phase rotation information is used to compensate phase rotations in 16 node measurements. An example of the node data after phase rotation compensation is shown in inset [c3] in Figure 6.2. The data is still dispersed and linearly mixed in the reservoir, hence the distorted form is seen. However, in the absence of phase rotations, the overall shape of the constellation is not circular. The weights for the readout states are then calculated, and an example of weighted node data is shown in inset [c5]. Finally, when all weighted nodes are summed, the result is passed to a matched filter that yields the constellations in [c7].

Besides these time-varying issues, the sequential measurement of reservoir nodes also resulted in several timing ambiguities. These ambiguities and their solutions are addressed in the following section.

6.5 Resolving Timing Ambiguities in Reservoir Data

The captured data streams from each separately measured node exhibit uncertainties in their symbol-based starting points. While cross-correlation is commonly used for symbol timing recovery, it fails here due to the dispersed and distorted nature of the node data. To address this, we synthetically attach a long header of zeros to the data, which serves as a flag marking the start of the symbol stream. By aligning the node data based on this header, the symbol timing ambiguity can be resolved.

However, this introduces a second ambiguity concerning the alignment of nodes relative to one another. Since all nodes are aligned to the common header, their relative alignment is lost, and this must be restored before summing the node data. To resolve this issue, we extract timing information from a simulated reservoir network model. By inputting an impulse signal into the network, we capture the timing delay based on when the impulse first appears at the output of each node. An example is shown in Figure 6.4, where the impulse response of three nodes is illustrated, with the impulse first detected at t = 1, 0.5, and 0, respectively. This timing information is extracted for all nodes and used to appropriately shift their outputs, restoring their relative alignment.

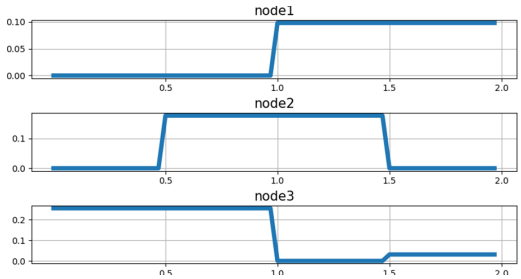


Figure 6.4: Simulated impulse response of reservoir showing time alignment of the first three nodes with respect to each other. Timing information extracted from this simulation is used to resolve the relative timing ambiguity arising from sequential node measurements.

Finally, we address the timing ambiguity caused by the absence of a clock recovery algorithm. Typically, DSP-based clock recovery corrects for mismatches in the sampling clock's frequency, phase, and jitter by adjusting early or late sampling instances on a symbol-by-symbol basis. However, as previously mentioned, this approach is difficult to apply to uncompensated node data. To mitigate this, we attempt to resolve the phase mismatch across the entire node stream by identifying an optimal sampling phase for each measured node. This method corrects for instances where sampling is one sample too early or too late. Although this solution may not be ideal for every symbol in the stream, it identifies a sampling point that minimizes errors across the entire dataset. Furthermore, since each node was sampled independently, the required phase adjustments may vary between nodes, turning this into a combinatorial challenge that is impractical to explore exhaustively. Consequently, optimization heuristics are employed to tackle this problem. To this end, we use the genetic algorithm CMA-ES to explore clock phase errors by applying sample-wise delays to all 16 nodes. The algorithm selects sample delays that minimize the training mean squared error (MSE) by evaluating different combinations of sample-wise shifts. Importantly, it also allows for the possibility that no adjustment is needed, meaning the correct sampling phase may have been used initially. Thus, for each node, the optimization algorithm selects a sample delay of +1, -1, or 0. In each iteration, a set of sample shifts is applied to the nodes, followed by linear regression to compute node weights. The resulting training error is recorded, and the algorithm proposes a new set of sample-wise delays. This process is repeated over 200 iterations to identify the optimal set of phase corrections.

Once the nodes are correctly aligned and their timing ambiguities are resolved, they are ready to be weighted and summed. These weights will allow correct compensation of the channel and transceiver impairments and their computation method is discussed in the following section.

6.6 Computing Weights

The experiment was done based on a training and testing scheme, where 95,000 symbols were transmitted, out of which 25,000 were partitioned for training and 70,000 were for testing. To compute the necessary node weights, the training data was used to generate an input signal matrix, X, with size 32 by 25,000. This corresponds to data collected from 16 nodes, with each node having an X and a Y polarization. Recall that the dual-polarization receiver setup will yield two signals that pertain to the same transmitted polarization. However, depending on the SOP at the receiver, either of these two signals would have a better SNR. As such, they were both used in the input signal matrix to yield the training matrix. Note that these two signals are treated exactly the same when performing the sampling frequency correction discussed in the previous section. As such, they practically only vary in their SNR, but we choose to keep both, since manual selection of the polarization with SNR for each node is cumbersome. Furthermore, utilizing both polarizations matches the behaviour of the MIMO equalizer, which similarly assigns different weights to each of the polarizations.

Besides the input matrix, the data labels for the training set, which are the symbol alphabets, are needed to form the target vector $\hat{y}$. The weight vector can then be calculated using a closed-form solution of Equation 2.7 .

Note that the reservoir's limited nonlinearity resulted in a relatively simple network structure. Furthermore, the intrinsic noise introduced by the EDFAs in the setup provided an implicit regularization effect. These aspects led to the reduced need of hyperparameter tuning, which we observed in the minimal effect that changing the regularization parameter had on the results. As such, no validation set was provisioned, since we did not need to compare results of different regularizations or other hyperparameters.

Computing the reservoir weights through linear regression utilizes a data-aided training algorithm, where labelled target data is used to compute the weights. On the other hand, blind optimization algorithms like CMA/RDE do not utilize such labels, but rather find ways to iteratively adjust weights through probability of symbol occurrence. These algorithms are very helpful when random phase rotations obscure the use of a target signal or when data is not labelled. However, to level the comparison ground between the DSP and reservoir pipelines, a comparable, but not standardly used, DSP alternative is devised that utilizes a data-aided training algorithm. This is shown in Fig 6.2 in steps 6* and 7*. Here, the same phase noise array computed in step 7 is used to compensate the signal post clock recovery. This generates the constellation seen in inset [a6*] of Figure 6.2, which shows a clear 16-QAM constellation but requires further equalization. To achieve that, two FFEs

are used, one for each polarization, and each with 16 taps to match the reservoir's 16 nodes. The FFE output is again a 32 by 25,000 matrix for the training set and is trained using the same linear regression method to compute the appropriate equalizer weights.

After weight computations are completed, the testing set is used to evaluate the performance of the equalization methods. These results are presented and discussed in the following section.

6.7 Results and Discussion

The role of the photonic reservoir is to equalize signals from fiber and transceiver impairments. To visualize the impact of this processing, Figure 6.5 shows a 4-QAM signal's time-domain waveforms, with the real and imaginary components plotted separately. The ideal, distorted, and compensated signals are represented by black, red, and green waveforms, respectively. To isolate the impact of the reservoir, the distorted signal was pre-compensated for frequency and phase offsets. Note that the distorted and compensated waveforms are plotted with 2 samples per symbol, and as such rigid transitions are seen. The effectiveness of the compensation is further visualized through constellation diagrams, comparing the distorted (top) and compensated (bottom) signals.

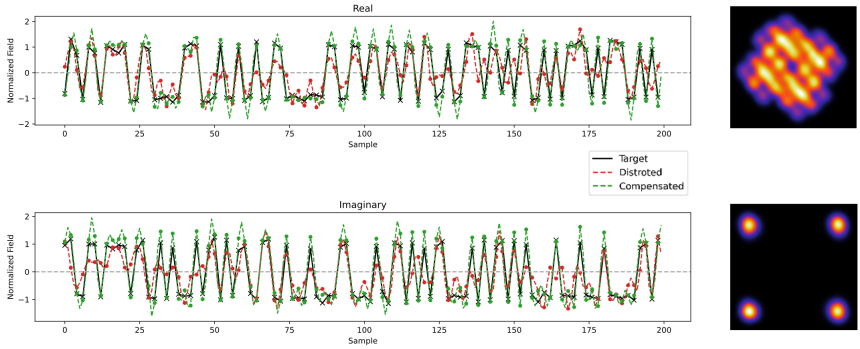


Figure 6.5: Time-domain waveforms (left) and constellation diagrams (right) of a 4-QAM signal at 2 dBm input power before and after compensation by the 4-port reservoir. On the left, the top panel shows the real component while the bottom panel shows the imaginary component of the signals. Black, red, and green waveforms represent ideal, distorted, and compensated signals, respectively. The constellation diagrams on the right show the distorted signal (top) and the reservoir compensated (bottom)

To benchmark the performance of the reservoir, Figure 6.6 shows the post-compensation results from the CMA/RDE DSP pipeline, the FFE pipeline, and the RC system. The BER is reported for the 16-QAM signals while the MSE is used as a metric for the 4-QAM signals, since their BER is negligible. In the 16-QAM plot, the results show very similar performance for all three systems, with a slight edge for the FFE pipeline in the linear regime and for the RC system in the more nonlinear regime. Similarly, the improved performance in the nonlinear regime is visible in the MSE plot for the 4-QAM signals.

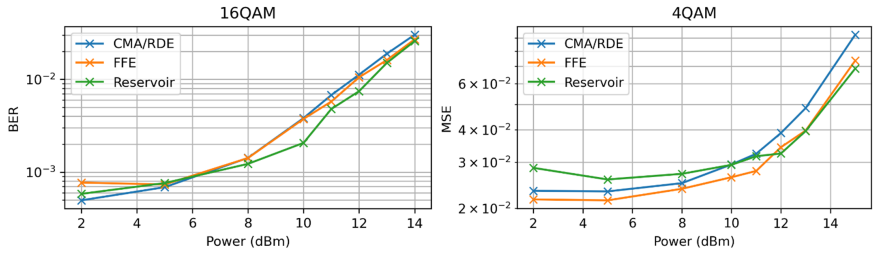


Figure 6.6: Equalization results over a range of transmission powers from RC system (green) benchmarked against full DSP pipelines equalized using CMA/RDE (blue) and FFE (orange). BER results are reported for 16-QAM (left) and MSE results are reported for 4-QAM (right)

This can also be seen in Figure 6.7, which shows three constellation diagrams for a 4 QAM signal at a 15 dBm fiber input power. Since the CMA/RDE algorithm performs equalization by minimizing the distance between the radius of a distorted symbol and an ideal radius, this algorithm shows the least curbing of nonlinearities. On the other hand, both the FFE and RC minimize the distances to the symbol itself, therefore achieving better nonlinear equalization.

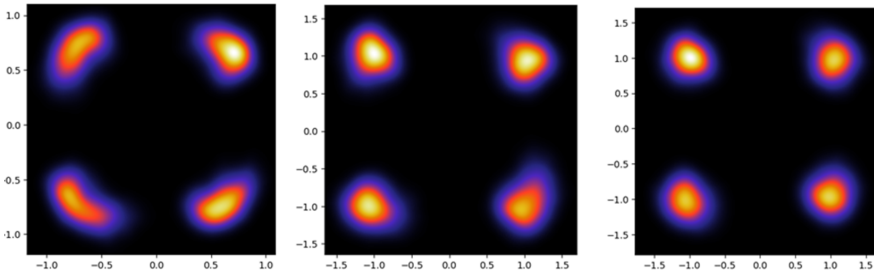


Figure 6.7: 4-QAM constellation diagrams for 15 dBm fiber input power after equalization in the CMA/RDE pipeline (left), FFE pipeline (center), and RC system (right)

The reservoir however seems to perform worse than the benchmark pipelines in the linear regime of the 4-QAM signals. This can be attributed to the higher losses of most reservoir nodes, where many were as high as 50 dB. Recall that the benchmark setup was set to match the loss of the highest output power node and as such would have a better overall SNR performance. This penalty is only visible in low-error scenarios, like those for the low power 4-QAM signals, where SNR becomes the main contributor. Furthermore, since MSE is the metric used, even slight noise penalties are readily observable. To verify this, Figure 6.8 shows the reservoir errors for the 4-QAM signal compared to the benchmark pipeline when losses of 35 dB, 45 dB, and 50 dB are introduced. For conciseness, only the FFE benchmark pipeline is shown, which is the better performing of the two DSP benchmarks. As seen, the reservoir closely matches the benchmark at 45 dB loss and outperforms it at 50 dB loss.

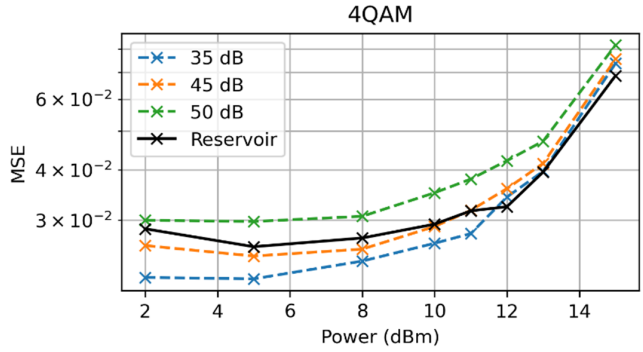


Figure 6.8: Errors of the 4-QAM signal from the reservoir pipeline (black) compared to the benchmark pipeline when losses of 35 dB (blue), 45 dB (orange), and 50 dB (green) are introduced.

Finally, we note that the choice of fiber length is common for the targeted short- and mid- reach applications, but experimental expansion to longer lengths was restricted by the equipment available at the time of the experiment. However, independent simulations showed that the reservoir can compensate for up to around 100 km of dispersion at the 28 Gbaud transmission rates. This limitation arises from the amount of memory in the reservoir, which can be expanded by adding more nodes and reducing the loss. Other hyper-parameters, including the length of the inter-node delay lines and the addition of delay lines before the summation tree, can also be optimized to increase the memory of the network.

6.8 Conclusion

In this chapter, we experimentally demonstrated the use of an integrated photonic reservoir for the equalization of 4- and 16-QAM signals at 28 Gbaud. This photonic system replaces the dispersion compensation and linear equalization blocks in the DSP pipeline, with results reported showing that it can achieve BER values on par with its DSP counterparts. Furthermore, the results show that the photonic reservoir has inherent parallelism on both quadratures and one generic chip has the programmability to address errors of different modulation formats at various linear and nonlinear transmission powers. This approach surpasses conventional computationally expensive equalization procedures, since it operates directly at the transmission speed, eliminating the need for computationally demanding high-baud-rate operations in the DSP at two samples per symbol.

7

Conclusion and Outlook

7.1 Conclusion

The increasing demand for higher data rates has driven substantial interest in new technologies that can enhance the capacity and efficiency of short- and mid-reach optical communication systems. Meeting these demands depends on the system constraints, however, maximizing the channel capacity per wavelength, rather than expanding the wavelength dimension, is generally a desired approach. Such efforts not only address immediate needs but also position the networks to accommodate future, albeit uncertain, generations of communication technology.

However, coherent transmission comes with high receiver costs and significant processing requirements, which hinder its application in high-density networks. To bridge this coherent gap, common research directions investigate reducing receiver complexity through alternative architectures, including self-coherent receivers, alongside more efficient methods for necessary signal processing.

To this end, this thesis explored the use of photonic processing, realized through an integrated reservoir computing architecture, to facilitate broader deployment of coherent modulation in short- and mid-reach communication systems. This was achieved by simplifying processing requirements for both coherent and self-coherent receivers, as well as enabling self-coherent receiver architectures to operate under less stringent conditions. The approach was validated through both simulations and experimental demonstrations across various modulation formats and data rates.

The thesis first introduced important preliminaries for these results by discussing the concept of reservoir computing in the context of physically implementable networks. With its reduced programmability requirements, the photonic reservoir facilitates the exploitation of optical properties, such as high bandwidth and inherent support for complex operations-valued, in performing signal processing.

In the following chapter, coherent and self-coherent receiver architectures were also introduced, highlighting the potential advantages of simplified designs and the challenges associated with their operation. Receiver simulation models were initially assessed in back-to-back scenarios to build a comprehensive understanding of their behavior.

The thesis then demonstrated the use of photonic-based reservoir equalization across two self-coherent receivers operating under various conditions. Specifically, the KK receiver was evaluated at low CSPR levels with low oversampling, while the SiP SSBI-canceling receiver was evaluated with shifted frequency responses. Given the stringent operational conditions required for these receivers to maintain linearity, the thesis showed that optical compensation of the link impairments could enhance overall system performance, enabling operation under relaxed conditions while achieving BERs below FEC thresholds. Furthermore, equalization of the SiP SSBI-based receiver errors were shown. These results were simulated with 64 QAM signals and in fiber lengths of up to 90 km.

For equalizing the KK receiver errors, the following chapter presented an opto-electronic network, comprising an integrated photonic reservoir and an electronic feed-forward equalizer, aimed at linearizing a KK receiver operating at 3 samples/symbol and low CSPRs. This network was used for relaxing the KK receiver's operational conditions, such that the DSP procedures following detection could be implemented with higher effectiveness. Simulation results indicated up to a fourfold reduction in BER with 64 QAM signals in links of 250 km and shorter, without the need for explicit training of the reservoir on these channels and with seamless DSP integration. The network's fully linear components facilitate practical implementation, leveraging the receiver's inherent nonlinearity to achieve desired performance enhancements.

Finally, the thesis experimentally demonstrated the application of the photonic reservoir for the equalization of 4- and 16-QAM signals at 28 Gbaud. This photonic system effectively replaced the dispersion compensation and linear equalization blocks in the DSP pipeline, achieving BER performance comparable to conventional DSP methods. The results underscore the inherent parallelism of the photonic reservoir across both quadratures and its programmability, enabling it to address errors in various modulation formats and under different linear and nonlinear transmission conditions. This approach surpasses traditional, computationally intensive equalization procedures by operating directly at transmission speed, eliminating the need for demanding high-baud-rate DSP operations at two samples per symbol.

7.2 Outlook

The findings presented in this thesis provide a foundation for advancing the use of photonic reservoir computing to simplify systems based on both coherent and self-coherent receivers. This approach shows promise in reducing the complexity of traditional DSP tasks by leveraging the benefits of photonic systems such as parallel processing capabilities and large bandwidths. However, further experimental validation using an integrated optical readout remains an open direction for future research.

Moreover, experimental validation with self-coherent receivers is needed to confirm and expand upon the promising simulation results obtained in this work. Additionally, it is crucial to investigate the suitability of self-coherent architectures for various transmission scenarios. While these architectures offer potential solutions to the complexities of conventional coherent receivers, the need to propagate a reference carrier, especially in polarization- or wavelength-multiplexed systems, introduces nonlinear interactions that can complicate processing.

Furthermore, the continued migration of the remaining DSP procedures to the optical domain is an important objective. This would significantly reduce the reliance on traditional DSP, thereby minimizing the processing overhead and latency associated with electronic components. Achieving this would be a major step towards fully optical processing units, which, when combined with simplified receiver architectures, can deliver superior performance and improved energy efficiency. Such advancements would enable the creation of scalable and cost-effective solutions for a wide range of optical communication applications, addressing the growing demand for high-speed, high-capacity networks.

Despite these promising directions, several challenges remain, particularly regarding power consumption and operational requirements. For the real-time deployment of optical processing solutions in practical, real-world optical networks, it is essential to ensure low-loss PIC components and effective heat dissipation. Achieving these objectives may necessitate the optimization of the reservoir's architecture to enhance efficiency and minimize energy demands. To facilitate this, future research should focus on bridging the gap between theoretical models and practical implementations, enabling the assessment of different designs based on relevant criteria before fabrication. This approach would help avoid unnecessary costs by determining the viability of a design in advance.

Another key consideration is the robustness and adaptability of these systems under diverse network conditions, which is crucial for widespread adoption. Network operators are often cautious about replacing established technologies due to the significant overhead costs and potential disruptions involved. Therefore, any new solution must offer strong performance assurances and support early error detection, including features like closed-loop feedback and predictive warnings to ensure stable system behavior. Moreover, existing systems have been carefully optimized for essential tasks such as monitoring, quality control, and maintenance. As a result, new optical processing solutions must demonstrate long-term reliability that not only matches or exceeds the performance of traditional electronic DSP systems but also provides comparable levels of observability and monitoring. This need is especially critical when adopting analog or less deterministic approaches, which may introduce new dynamics into system behavior and require more nuanced control mechanisms.

Finally, effective solutions for processing polarization-multiplexed signals, a fundamental transmission technology, are essential. This is because integrated silicon-based components are typically designed with polarization-optimized performance, making it necessary to develop approaches that can efficiently handle dual-polarization signals. Addressing this need is crucial for ensuring compatibility and performance in modern optical communication systems.

Continued research and development efforts are essential to address these challenges. By focusing on optimizing performance, efficiency, and reliability, the integration of optical processing technologies can become a viable alternative to conventional electronic DSP. This progress would pave the way for a new era in optical communications, characterized by higher scalability, reduced costs, and improved energy efficiency.

References

[1] P. J. Winzer, D. T. Neilson, and A. R. Chraplyvy, "Fiber-optic transmission and networking: the previous 20 and the next 20 years [Invited]," *Optics Express*, vol. 26, no. 18, p. 24190, 2018.

[2] B. Chomycz, *Planning Fiber Optic Networks*. McGraw-Hill, 2009.

[3] R. Ramaswami, K. N. Sivarajan, and G. H. Sasaki, *Optical Networks: A Practical Perspective*. ELSEVIER, third ed., 2010.

[4] M. Nance Hall, K. T. Foerster, S. Schmid, and R. Durairajan, "A Survey of Reconfigurable Optical Networks," *Optical Switching and Networking*, vol. 41, no. February, p. 100621, 2021.

[5] G. Simon, "Signal Processing Opportunities in Passive Optical Networks (Workshop)," in *European Conference on Optical Communication (ECOC)*, (Glasgow), 2023.

[6] Orange, "Net zero carbon by 2040." https://www.orange.com/en/net-zero-carbon-2040#:~:text=Asaleadingoperator%2COrange,zerocarbonemissionsby2040. Accessed: 10. 07. 2024.

[7] S. Abreu, I. Boikov, M. Goldmann, T. Jonuzi, A. Lupo, S. Masaad, L. Nguyen, E. Picco, G. Pourcel, A. Skalli, L. Talandier, B. Vettelschoss, E. Vlieg, A. Argyris, P. Bienstman, D. Brunner, J. Dambre, L. Daudet, J. Domenech, I. Fischer, F. Horst, S. Massar, C. Mirasso, B. Offrein, A. Rossi, M. Soriano, S. Sygletos, and S. Turitsyn, "A photonics perspective on computing with physical substrates," *Reviews in Physics*, vol. 12, no. June, p. 100093, 2024.

[8] J. Laydevant, L. G. Wright, T. Wang, and P. L. McMahon, "The hardware is the software," *Neuron*, vol. 112, no. 2, pp. 180–183, 2024.

[9] K. Zhong, X. Zhou, J. Huo, C. Yu, C. Lu, and A. P. T. Lau, "Digital Signal Processing for Short-Reach Optical Communications: A Review of Current

Technologies and Future Trends," *Journal of Lightwave Technology*, vol. 36, no. 2, pp. 377–400, 2018.

[10] A. S. Karar, A. R. E. Falou, J. M. H. Barakat, Z. N. Gürkan, and K. Zhong, "Recent Advances in Coherent Optical Communications for Short-Reach: Phase Retrieval Methods," *Photonics*, vol. 10, no. 3, p. 308, 2023.

[11] X. Zhou, R. Urata, and H. Liu, "Beyond 1 Tb/s Intra-Data Center Interconnect Technology: IM-DD or Coherent?," *Journal of Lightwave Technology*, vol. 38, no. 2, pp. 475–484, 2020.

[12] S. Kumar, G. Papen, K. Schmidtke, and C. Xie, "Intra-data center interconnects, networking, and architectures," in *Optical fiber telecommunications VII* (A. E. Willner, ed.), ch. 14, Elsevier, 2019.

[13] J. K. Perin, A. Shastri, and J. M. Kahn, "Coherent Data Center Links," *Journal of Lightwave Technology*, vol. 39, no. 3, pp. 730–741, 2021.

[14] J. Cheng, C. Xie, Y. Chen, X. Chen, M. Tang, and S. Fu, "Comparison of coherent and IMDD transceivers for intra datacenter optical interconnects," *Optics InfoBase Conference Papers*, vol. Part F160-, pp. 1–3, 2019.

[15] I. Alimi, R. Patel, N. Silva, C. Sun, H. Ji, W. Shieh, A. Pinto, and N. Muga, "A review of self-coherent optical transceivers: Fundamental issues, recent advances, and research directions," *Applied Sciences*, vol. 11, no. 16, pp. 1–85, 2021.

[16] C. Xie and J. Cheng, "Coherent Optics for Data Center Networks," *2020 IEEE Photonics Society Summer Topical Meeting Series, SUM 2020 - Proceedings*, pp. 10–11, 2020.

[17] L. Paraschis and K. Raj, "Innovations in DCI transport networks," in *Optical fiber telecommunications VII* (A. E. Willner, ed.), ch. 15, ELSEVIER, 2019.

[18] R. Nagarajan, I. Lyubomirsky, and O. Agazzi, "Low Power DSP-Based Transceivers for Data Center Optical Fiber Communications (Invited Tutorial)," *Journal of Lightwave Technology*, vol. 39, no. 16, pp. 5221–5231, 2021.

[19] S. T. Le, K. Schuh, R. Dischler, F. Buchali, L. Schmalen, and H. Buelow, "Beyond 400 Gb/s Direct Detection over 80 km for Data Center Interconnect Applications," *Journal of Lightwave Technology*, vol. 38, no. 2, pp. 538–545, 2020.

[20] C. F. Lam and S. Yin, "Evolution of fiber access networks," in *Optical Fiber Telecommunications VII* (A. E. Willner, ed.), ch. 19, pp. 827–865, Elsevier Inc., 2019.

[21] L. A. Neto, J. Maes, P. Larsson-Edefors, J. Nakagawa, K. Onohara, and S. J. Trowbridge, "Considerations on the Use of Digital Signal Processing in Future Optical Access Networks," *Journal of Lightwave Technology*, vol. 38, no. 3, pp. 598–607, 2020.

[22] D. A. de Arruda Mello and F. A. Barbosa, *Digital Coherent Optical Systems*. Springer, first ed., 2021.

[23] H. Sun, K.-t. Wu, and K. Roberts, "Real-time measurements of a 40 Gb / s coherent system," *Optics Express*, vol. 16, no. 2, pp. 873–879, 2008.

[24] A. Mecozzi, C. Antonelli, and M. Shtaif, "Kramers–Kronig coherent receiver," *Optica*, vol. 3, pp. 1220–1227, 11 2016.

[25] H. Chen, N. K. Fontaine, J. M. Gene, R. Ryf, D. T. Neilson, and G. Raybon, "Dual Polarization Full-Field Signal Waveform Reconstruction Using Intensity only Measurements for Coherent Communications," *Journal of Lightwave Technology*, vol. 38, no. 9, pp. 2587–2597, 2020.

[26] M. Lyu, W. Shi, and L. A. Rusch, "SiP Alternative to Enhanced KK for OFDM," *European Conference on Optical Communication, ECOC*, vol. 2018-Septe, no. 1, pp. 1–3, 2018.

[27] T. Gui, X. Wang, M. Tang, Y. Yu, Y. Lu, and L. Li, "Real-Time Demonstration of Homodyne Coherent Bidirectional Transmission for Next-Generation Data Center Interconnects," *Journal of Lightwave Technology*, vol. 39, no. 4, pp. 1231–1238, 2021.

[28] J. K. Perin, A. Shastri, and J. M. Kahn, "Design of Low-Power DSP-Free Coherent Receivers for Data Center Links," *Optics InfoBase Conference Papers*, vol. Part F84-O, no. 21, pp. 4650–4662, 2018.

[29] J. Cheng, C. Xie, M. Tang, and S. Fu, "A Comparative Study of Intradyne and Self-homodyne Systems for Next Generation Intra-datacenter Optical Interconnects," *OECC/PSC 2019 - 24th OptoElectronics and Communications Conference/International Conference Photonics in Switching and Computing 2019*, vol. 1, no. c, pp. 1–3, 2019.

[30] T. Hirokawa, S. Pinna, N. Hosseinzadeh, A. Maharry, H. Andrade, J. Liu, T. Meissner, S. Misak, G. Movaghar, L. A. Valenzuela, Y. Xia, S. Bhat, F. Gambini, J. Klamkin, A. A. Saleh, L. Coldren, J. F. Buckwalter, and C. L. Schow, "Analog Coherent Detection for Energy Efficient Intra-Data Center Links at 200 Gbps per Wavelength," *Journal of Lightwave Technology*, vol. 39, no. 2, pp. 520–531, 2021.

[31] K. Vandoorne, P. Mechet, T. Van Vaerenbergh, M. Fiers, G. Morthier, D. Verstraeten, B. Schrauwen, J. Dambre, and P. Bienstman, "Experimental demonstration of reservoir computing on a silicon photonics chip," *Nature Communications*, vol. 5, 2014.

[32] S. M. Ranzini, F. Da Ros, H. Bülow, and D. Zibar, "Tunable optoelectronic chromatic dispersion compensation based on machine learning for short-reach transmission," *Applied Sciences (Switzerland)*, vol. 9, no. 20, 2019.

[33] C. Mesaritakis, K. Sozos, D. Dermanis, and A. Bogris, "Spatial Photonic Reservoir Computing based on Non-Linear Phase-to-Amplitude Conversion in Micro-Ring Resonators," in *Optical Fiber Communication Conference*, 2021.

[34] C. Huang, S. Fujisawa, T. F. de Lima, A. N. Tait, E. Blow, Y. Tian, S. Bilodeau, A. Jha, F. Yaman, H. G. Batshon, H. T. Peng, B. J. Shastri, Y. Inada, T. Wang, and P. R. Prucnal, "Demonstration of photonic neural network for fiber nonlinearity compensation in long-haul transmission systems," *Optical Fiber Communications Conference*, 2020.

[35] S. Li and S. Pachnicke, "Optical equalization using photonic reservoir computing with optical analog signal injection," in *Asia Communications and Photonics Conference*, 2019.

[36] I. Estébanez, S. Li, J. Schwind, I. Fischer, S. Pachnicke, and A. Argyris, "56 GBaud PAM-4 100 km Transmission System with Photonic Processing Schemes," *Journal of Lightwave Technology*, vol. 40, no. 1, pp. 55–62, 2022.

[37] M. Sorokina, S. Sergeyev, and S. Turitsyn, "Fiber echo state network analogue for high-bandwidth dual-quadrature signal processing," *Optics Express*, vol. 27, no. 3, p. 2387, 2019.

[38] X. Zhang and Z. Wu, "Linear optical signal processing with optical filters: a tutorial," *Frontiers of Optoelectronics*, vol. 9, no. 3, pp. 377–389, 2016.

[39] G. E. Moore, "Cramming more components onto integrated circuits, Reprinted from Electronics, volume 38, number 8, April 19, 1965, pp.114 ff.," *IEEE Solid-State Circuits Society Newsletter*, vol. 11, no. 3, pp. 33–35, 2009.

[40] M. M. Waldrop, "More Than Moore," *Nature*, vol. 530, no. 7589, pp. 144–147, 2016.

[41] E. Agrell, M. Karlsson, F. Poletti, S. Namiki, X. Chen, L. A. Rusch, B. Puttnam, P. Bayvel, L. Schmalen, Z. Tao, F. R. Kschischang, A. Alvarado, B. Mukherjee, R. Casellas, X. Zhou, D. van Veen, G. Mohs, E. Wong, A. Mecozzi, M. S. Alouini, E. Diamanti, and M. Uysal, "Roadmap on optical communications," 2024.

[42] H. Jaeger, "Towards a generalized theory comprising digital, neuromorphic and unconventional computing," *Neuromorphic Computing and Engineering*, vol. 1, no. 1, p. 012002, 2021.

[43] Postdigital ITN, "PostDigital ITN." https://postdigital.astonphotonics.uk/. Accessed: 17. 07. 2024.

[44] A. Lugnan, A. Katumba, F. Laporte, M. Freiberger, S. Sackesyn, C. Ma, E. Gooskens, J. Dambre, and P. Bienstman, "Photonic neuromorphic information processing and reservoir computing," *APL Photonics*, vol. 5, no. 2, 2020.

[45] A. Kumar and T. Davenport, "How to Make Generative AI Greener (Harvard Business Review)." https://hbr.org/2023/07/how-to-make-generative-ai-greener, 2023. Accessed: 18. 07. 2024.

[46] AWS, "Amazon EC2 Update – Inf1 Instances with AWS Inferentia Chips for High Performance Cost-Effective Inferencing." https://aws.amazon.com/blogs/aws/amazon-ec2-update-inf1-instances-with-aws-inferentia-chips-for-high-performance-cost-effective-inferencing/. Accessed: 18. 07. 2024.

[47] S. Hooker, "The hardware lottery," *Communications of the ACM*, vol. 64, no. 12, pp. 58–65, 2021.

[48] D. V. Christensen, R. Dittmann, B. Linares-Barranco, A. Sebastian, M. L. Gallo, A. Redaelli, S. Slesazeck, T. Mikolajick, S. Spiga, S. Menzel, I. Valov, G. Milano, C. Ricciardi, S.-J. Liang, F. Miao, M. Lanza, T. J. Quill, S. T. Keene, A. Salleo, J. Grollier, D. Marković, A. Mizrahi, P. Yao, J. J. Yang, G. Indiveri, J. P. Strachan, S. Datta, E. Vianello, A. Valentian, J. Feldmann, X. Li, W. H. P. Pernice, H. Bhaskaran, E. Neftci, S. Ramaswamy, J. Tapson, F. Scherr, W. Maass, P. Panda, Y. Kim, G. Tanaka, S. Thorpe, C. Bartolozzi, T. A. Cleland, C. Posch, S.-C. Liu, A. N. Mazumder, M. Hosseini, T. Mohsenin, E. Donati, S. Tolu, R. Galeazzi, M. E. Christensen, S. Holm, D. Ielmini, and N. Pryds, "Roadmap on Neuromorphic Computing and Engineering," *Neuromorphic Computing and Engineering*, vol. 2, 2021.

[49] F. Zhang, C. Li, Z. Li, L. Dong, and J. Zhao, "Recent progress in three-terminal artificial synapses based on 2D materials: from mechanisms

to applications," *Microsystems and Nanoengineering*, vol. 9, no. 1, pp. 1–17, 2023.

[50] V. A. Demin, A. V. Emelyanov, D. A. Lapkin, V. V. Erokhin, P. K. Kashkarov, and M. V. Kovalchuk, "Neuromorphic elements and systems as the basis for the physical implementation of artificial intelligence technologies," *Crystallography Reports*, vol. 61, no. 6, pp. 992–1001, 2016.

[51] L. Wang, S. R. Lu, and J. Wen, "Recent Advances on Neuromorphic Systems Using Phase-Change Materials," *Nanoscale Research Letters*, vol. 12, no. 1, 2017.

[52] C. D. Schuman, T. E. Potok, R. M. Patton, J. D. Birdwell, M. E. Dean, G. S. Rose, and J. S. Plank, "A Survey of Neuromorphic Computing and Neural Networks in Hardware," *Arxiv*, pp. 1–88, 2017.

[53] B. Schrauwen, D. Verstraeten, and J. Van Campenhout, "An overview of reservoir computing: Theory, applications and implementations," *ESANN 2007 Proceedings - 15th European Symposium on Artificial Neural Networks*, pp. 471–482, 2007.

[54] J. Touch, A. H. Badawy, and V. J. Sorger, "Optical computing," *Nanophotonics*, vol. 6, no. 3, pp. 503–505, 2017.

[55] C. Li, X. Zhang, J. Li, T. Fang, and X. Dong, "The challenges of modern computing and new opportunities for optics," *PhotoniX*, vol. 2, no. 1, 2021.

[56] P. L. McMahon, "The physics of optical computing," *Nature Reviews Physics*, vol. 5, no. 12, pp. 717–734, 2023.

[57] J. Touch, Y. Cao, M. Ziyadi, A. Almaiman, A. Mohajerin-Ariaei, and A. E. Willner, "Digital optical processing of optical communications: Towards an Optical Turing Machine," *Nanophotonics*, vol. 6, no. 3, pp. 507–530, 2017.

[58] M. Kissner, L. Del Bino, F. Päsler, P. Caruana, and G. Ghalanos, "An All-Optical General-Purpose CPU and Optical Computer Architecture," *Arxiv*, pp. 1–14, 2024.

[59] R. Athale and D. Psaltis, "Optical Computing: Past and Future," *Optics and Photonics News*, vol. 27, no. 6, p. 32, 2016.

[60] J. Tanida, "Revival of Optical Computing," in *Photonic Neural Networks with Spatiotemporal Dynamics* (H. Suzuki, J. Tanida, and M. Hashimoto, eds.), ch. 1, pp. 3–23, Springer Nature Singapore, 2024.

[61] S. Greengard, "Photonic processors light the way," *Communications of the ACM*, vol. 64, no. 9, pp. 16–18, 2021.

[62] R. S. Tucker, "The role of optics in computing," *Nature Photonics*, vol. 4, no. 7, p. 405, 2010.

[63] D. A. Miller, "Are optical transistors the logical next step?," *Nature Photonics*, vol. 4, no. 1, pp. 3–5, 2010.

[64] W. Bogaerts, D. Pérez, J. Capmany, D. A. Miller, J. Poon, D. Englund, F. Morichetti, and A. Melloni, "Programmable photonic circuits," *Nature*, vol. 586, no. 7828, pp. 207–216, 2020.

[65] Z. Gao, X. Chen, Z. Zhang, U. Chakraborty, W. Bogaerts, and D. S. Boning, "Automatic synthesis of light-processing functions for programmable photonics: theory and realization," *Photonics Research*, vol. 11, no. 4, p. 643, 2023.

[66] B. J. Shastri, A. N. Tait, T. Ferreira de Lima, W. H. Pernice, H. Bhaskaran, C. D. Wright, and P. R. Prucnal, "Photonics for artificial intelligence and neuromorphic computing," *Nature Photonics*, vol. 15, no. 2, pp. 102–114, 2021.

[67] L. De Marinis, M. Cococcioni, P. Castoldi, and N. Andriolli, "Photonic Neural Networks: A Survey," *IEEE Access*, vol. 7, pp. 175827–175841, 2019.

[68] G. Wetzstein, A. Ozcan, S. Gigan, S. Fan, D. Englund, M. Soljačić, C. Denz, D. A. Miller, and D. Psaltis, "Inference in artificial intelligence with deep optics and photonics," *Nature*, vol. 588, no. 7836, pp. 39–47, 2020.

[69] G. Van Der Sande, D. Brunner, and M. C. Soriano, "Advances in photonic reservoir computing," *Nanophotonics*, vol. 6, no. 3, pp. 561–576, 2017.

[70] H. Wang, J. Hu, Y. Baek, K. Tsuchiyama, M. Joly, Q. Liu, and S. Gigan, "Optical next generation reservoir computing," *Arxiv (Optics)*, 2024.

[71] C. Huang, A. Jha, T. F. De Lima, A. N. Tait, A. N. Tait, B. J. Shastri, B. J. Shastri, and P. R. Prucnal, "On-Chip Programmable Nonlinear Optical Signal Processor and Its Applications," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 27, no. 2, 2021.

[72] K. Sozos, A. Bogris, P. Bienstman, and C. Mesaritakis, "Photonic Reservoir Computing based on Optical Filters in a Loop as a High Performance and Low-Power Consumption Equalizer for 100 Gbaud Direct Detection Systems," in *2021 European Conference on Optical Communication, ECOC 2021*, pp. 1–4, IEEE, 2021.

[73] J. Hennessy, "The End of Moore ' s Law, CPUs (as we know them), and the Rise of Domain Specific Architectures." https://opennetworking.org/wp-content/uploads/2020/12/9_2.05pm_John_Hennessey.pdf, 2019. Lecture Notes - Stanford.

[74] S. Y. Siew, B. Li, F. Gao, H. Y. Zheng, W. Zhang, P. Guo, S. W. Xie, A. Song, B. Dong, L. W. Luo, C. Li, X. Luo, and G. Q. Lo, "Review of Silicon Photonics Technology and Platform Development," *Journal of Lightwave Technology*, vol. 39, no. 13, pp. 4374–4389, 2021.

[75] L. Thylén and L. Wosinski, "Integrated photonics in the 21st century," *Photonic Research*, vol. 2, no. 2, pp. 75–81, 2014.

[76] R. Baets, A. Z. Subramanian, S. Clemmen, B. Kuyken, P. Bienstman, N. Le Thomas, G. Roelkens, D. Van Thourhout, P. Helin, and S. Severi, "Silicon photonics: Silicon nitride versus silicon-on-insulator," *2016 Optical Fiber Communications Conference and Exhibition, OFC 2016*, pp. 3–5, 2016.

[77] A. Rahim, J. Goyvaerts, B. Szelag, J. M. Fedeli, P. Absil, T. Aalto, M. Harjanne, C. Littlejohns, G. Reed, G. Winzer, S. Lischke, L. Zimmermann, D. Knoll, D. Geuzebroek, A. Leinse, M. Geiselmann, M. Zervas, H. Jans, A. Stassen, C. Dominguez, P. Munoz, D. Domenech, A. L. Giesecke, M. C. Lemme, and R. Baets, "Open-access silicon photonics platforms in Europe," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 25, no. 5, pp. 1–18, 2019.

[78] R. Baets and A. Rahim, "Heterogeneous integration in silicon photonics: opportunities and challenges: opinion," *Optical Materials Express*, vol. 13, no. 12, p. 3439, 2023.

[79] N. M. Fahrenkopf, C. McDonough, G. L. Leake, Z. Su, E. Timurdogan, and D. D. Coolbaugh, "The AIM Photonics MPW: A Highly Accessible Cutting Edge Technology for Rapid Prototyping of Photonic Integrated Circuits," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 25, no. 5, pp. 1–6, 2019.

[80] W. Bogaerts and A. Rahim, "Programmable Photonics: An Opportunity for an Accessible Large-Volume PIC Ecosystem," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 26, no. 5, pp. 1–17, 2020.

[81] F. Laporte, J. Dambre, and P. Bienstman, "Highly parallel simulation and optimization of photonic circuits in time and frequency domain based on the deep-learning framework PyTorch," *Scientific Reports*, vol. 9, no. 1, pp. 1–9, 2019.

[82] S. Sackesyn, C. Ma, A. Katumba, J. Dambre, and P. Bienstman, "A Power-Efficient Architecture for On-Chip Reservoir Computing," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11731 LNCS, pp. 161–164, 2019.

[83] C. Bishop, *Pattern recognition and machine earning*. Springer, 2006.

[84] A. Géron, *Hands-on Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly, 1 ed., 2017.

[85] L. Jaurigue and K. Lüdge, "Connecting reservoir computing with statistical forecasting and deep neural networks," *Nature Communications*, vol. 13, no. 1, pp. 5–7, 2022.

[86] T. F. De Lima, H. T. Peng, A. N. Tait, M. A. Nahmias, H. B. Miller, B. J. Shastri, and P. R. Prucnal, "Machine Learning with Neuromorphic Photonics," *Journal of Lightwave Technology*, vol. 37, no. 5, pp. 1515–1534, 2019.

[87] R. Pascanu, T. Mikolov, and Y. Bengio, "On the difficulty of training recurrent neural networks," in *30th International Conference on Machine Learning, ICML 2013*, pp. 2347–2355, 2013.

[88] M. Lukoševičius and H. Jaeger, "Reservoir computing approaches to recurrent neural network training," *Computer Science Review*, vol. 3, no. 3, pp. 127–149, 2009.

[89] E. Bollt, "On explaining the surprising success of reservoir computing forecaster of chaos? The universal machine learning dynamical system with contrast to VAR and DMD," *Chaos*, vol. 31, no. 1, pp. 1–34, 2021.

[90] H. Zhang and D. V. Vargas, "A Survey on Reservoir Computing and its Interdisciplinary Applications beyond Traditional Machine Learning," *IEEE Access*, vol. 11, pp. 81033–81070, 2023.

[91] G. Tanaka, T. Yamane, J. B. Héroux, R. Nakane, N. Kanazawa, S. Takeda, H. Numata, D. Nakano, and A. Hirose, "Recent advances in physical reservoir computing: A review," *Neural Networks*, vol. 115, pp. 100–123, 2019.

[92] D. Brunner, B. Penkovsky, B. A. Marquez, M. Jacquot, I. Fischer, and L. Larger, "Tutorial: Photonic neural networks in delay systems," *Journal of Applied Physics*, vol. 124, no. 15, pp. 1–16, 2018.

[93] F. Duport, B. Schneider, A. Smerieri, M. Haelterman, and S. Massar, "All-optical reservoir computing," *Optics Express*, vol. 20, no. 20, 2012.

[94] D. Brunner, M. C. Soriano, C. R. Mirasso, and I. Fischer, "Parallel photonic information processing at gigabyte per second data rates using transient states," *Nature Communications*, vol. 4, 2013.

[95] R. Modeste Nguimdo, G. Verschaffelt, J. Danckaert, and G. Van Der Sande, "Simultaneous computation of two independent tasks using reservoir computing based on a single photonic nonlinear node with optical feedback," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, no. 12, pp. 3301–3307, 2015.

[96] Q. Vinckier, F. Duport, A. Smerieri, M. Haelterman, and S. Massar, "Information processing using a photonic reservoir computer based on a coherently driven passive cavity with an analog readout layer," *Optics InfoBase Conference Papers*, vol. 2, no. 5, 2015.

[97] D. Brunner and I. Fischer, "Reconfigurable semiconductor laser networks based on diffractive coupling," *Optics Letters*, vol. 40, no. 16, p. 3854, 2015.

[98] P. Antonik, N. Marsal, D. Brunner, and D. Rontani, "Human action recognition with a large-scale brain-inspired photonic computer," *Nature Machine Intelligence*, vol. 1, no. 11, pp. 530–537, 2019.

[99] M. Rafayelyan, J. Dong, Y. Tan, F. Krzakala, and S. Gigan, "Large-Scale Optical Reservoir Computing for Spatiotemporal Chaotic Systems Prediction," *Physical Review X*, vol. 10, no. 4, p. 41037, 2020.

[100] K. Vandoorne, W. Dierckx, B. Schrauwen, D. Verstraeten, R. Baets, P. Bienstman, and J. V. Campenhout, "Toward Optical signal processing using photonic reservoir computing," *Journal of Modern Optics*, vol. 61, no. 17, pp. 1442–1451, 2008.

[101] K. Vandoorne, J. Dambre, D. Verstraeten, B. Schrauwen, and P. Bienstman, "Parallel reservoir computing using optical amplifiers," *IEEE Transactions on Neural Networks*, vol. 22, no. 9, pp. 1469–1481, 2011.

[102] D. J. Gauthier, E. Bollt, A. Griffith, and W. A. Barbosa, "Next generation reservoir computing," *Nature Communications*, vol. 12, no. 1, pp. 1–8, 2021.

[103] A. Orvieto, S. L. Smith, A. Gu, A. Fernando, C. Gulcehre, R. Pascanu, and S. De, "Resurrecting Recurrent Neural Networks for Long Sequences," in *International Conference on Machine Learning*, pp. 1–30, 2023.

[104] H. Ma, D. Prosperino, and C. Räth, "A novel approach to minimal reservoir computing," *Scientific Reports*, vol. 13, no. 1, pp. 1–9, 2023.

[105] J. M. Senior, *Optical Fiber Communications Principles and Practice*. Pearson, 3rd ed., 2009.

[106] Ligentec, "Ligentec Silicon Nitride." https://www.ligentec.com/technology/. Accessed: 15. 08. 2024.

[107] S. Sackesyn, "Nebula 3rd Review Meeting," tech. rep., Ghent University, 2022.

[108] E. Gooskens, *The Wavelength Dimension in Photonic Reservoir Computing Emmanuel*. PhD thesis, Ghent University, 2024.

[109] S. Shekhar, W. Bogaerts, L. Chrostowski, J. E. Bowers, M. Hochberg, R. Soref, and B. J. Shastri, "Roadmapping the next generation of silicon photonics," *Nature Communications*, vol. 15, no. 1, 2024.

[110] Floris Laporte, "Photontorch." https://docs.photontorch.com/. Accessed 23.07.2024.

[111] T. W. Hughes, M. Minkov, Y. Shi, and S. Fan, "Training of photonic neural networks through in situ backpropagation and gradient measurement," *Optica*, vol. 5, no. 7, p. 864, 2018.

[112] S. Pai, Z. Sun, T. W. Hughes, T. Park, B. Bartlett, I. A. Williamson, M. Minkov, M. Milanizadeh, N. Abebe, F. Morichetti, A. Melloni, S. Fan, O. Solgaard, and D. A. Miller, "Experimentally realized in situ backpropagation for deep learning in photonic neural networks," *Science (New York, N.Y.)*, vol. 380, no. 6643, pp. 398–404, 2023.

[113] M. J. Filipovich, Z. Guo, M. Al-Qadasi, B. A. Marquez, H. D. Morison, V. J. Sorger, P. R. Prucnal, S. Shekhar, and B. J. Shastri, "Silicon photonic architecture for training deep neural networks with direct feedback alignment," *Optica*, vol. 9, no. 12, p. 1323, 2022.

[114] G. Hinton, "The Forward-Forward Algorithm: Some Preliminary Investigations," *Arxiv*, 2022.

[115] C. Ma, J. Van Kerrebrouck, H. Deng, S. Sackesyn, E. Gooskens, B. Bai, J. Dambre, and P. Bienstman, "Integrated Photonic Reservoir Computing with All-Optical Readout," *Optics Express*, 2023.

[116] S. M. Buckley, A. N. Tait, A. N. McCaughan, and B. J. Shastri, "Photonic online learning: a perspective," *Nanophotonics*, vol. 12, no. 5, pp. 833–845, 2023.

[117] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning Representations by Back-Propagating Errors," *Nature*, no. 323, p. 533–536, 1986.

[118] S. Ruder, "An overview of gradient descent optimization algorithms," *Arxiv*, pp. 1–14, 2016.

[119] A. Dayne, "Gradient Descent Visualization." https://medium.com/intuition/gradient-descent-visualization-285d3dd0fe00. Accessed 27.07.2024.

[120] N. Hansen, "The CMA evolution strategy: A comparing review," *Studies in Fuzziness and Soft Computing*, vol. 192, no. 2006, pp. 75–102, 2006.

[121] K. Gentile, "The care and Feeding of Digital, Pulse-Shaping Filters," *RF Design*, 2002.

[122] A. Mecozzi, C. Antonelli, and M. Shtaif, "Kramers–Kronig receivers," *Advances in Optics and Photonics*, vol. 11, no. 3, p. 480, 2019.

[123] A. Mecozzi, C. Antonelli, and M. Shtaif, "Kramers-Kronig receivers: supplementary material," *Optical Society of America*, vol. 11, no. 3, pp. 480–517, 2019.

[124] C. Fullner, M. M. H. Adib, S. Wolf, J. N. Kemal, W. Freude, C. Koos, and S. Randel, "Complexity Analysis of the Kramers–Kronig Receiver," *Journal of Lightwave Technology*, vol. 37, no. 17, pp. 4295–4307, 2019.

[125] M. Lyu, W. Shi, and L. Rusch, "SIP-based SSBI cancellation for OFDM," *IEEE Photonics Journal*, vol. 11, no. 5, 2019.

[126] W. Bogaerts, P. de Heyn, T. van Vaerenbergh, K. de Vos, S. Kumar Selvaraja, T. Claes, P. Dumon, P. Bienstman, D. van Thourhout, and R. Baets, "Silicon microring resonators," *Laser and Photonics Reviews*, vol. 6, no. 1, pp. 47–73, 2012.

[127] S. Masaad, E. Gooskens, S. Sackesyn, J. Dambre, and P. Bienstman, "Photonic Reservoir Computing for Nonlinear Equalization of 64-QAM Signals with a Kramers-Kronig Receiver," in *2022 European Conference on Optical Communication, ECOC 2022*, pp. 1–4, 2022.

[128] S. Masaad, E. Gooskens, S. Sackesyn, J. Dambre, and P. Bienstman, "Photonic reservoir computing for nonlinear equalization of 64-QAM signals with a Kramers-Kronig receiver," *Nanophotonics*, vol. 12, no. 5, pp. 925–935, 2023.

[129] A. Zelaci, "Thesis: Photonic reservoir computing for equalizing coherently modulated signals." https://libstore.ugent.be/fulltxt/RUG01/003/149/818/RUG01-003149818_2023_0001_AC.pdf, 2023. Master Thesis.

[130] Y. Osadchuk, O. Jovanovic, D. Zibar, and F. Da Ros, "Reservoir Computing-based Multi-Symbol Equalization for PAM 4 Short-reach Transmission," in *2023 Conference on Lasers and Electro-Optics, CLEO 2023*, pp. 2–3, 2023.

[131] Z. Xu, C. Sun, J. H. Manton, and W. Shieh, "Computational complexity analysis of neural network-based nonlinear equalization for short reach direct detection systems," *Optics InfoBase Conference Papers*, vol. Part F138-, no. 2, pp. 2–4, 2019.

[132] P. Gou and J. Yu, "A nonlinear ANN equalizer with mini-batch gradient descent in 40Gbaud PAM-8 IM/DD system," *Optical Fiber Technology*, vol. 46, no. October, pp. 113–117, 2018.

[133] F. Da Ros, S. M. Ranzini, H. Bulow, and D. Zibar, "Reservoir-computing based equalization with optical pre-processing for short-reach optical transmission," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 26, no. 5, 2020.

[134] E. Staffoli, P. Bettotti, and L. Pavesi, "Recovery of 40 Gbps PAM4 in a 50 km optical link through a delayed complex perceptron," *2023 IEEE Silicon Photonics Conference (SiPhotonics)*, pp. 1–2, 2023.

[135] S. Li and S. Pachnicke, "Photonic Reservoir Computing in Optical Transmission Systems," *2020 IEEE Photonics Society Summer Topical Meeting Series, SUM 2020 - Proceedings*, pp. 0–1, 2020.

[136] S. Sackesyn, C. Ma, J. Dambre, and P. Bienstman, "Experimental realization of integrated photonic reservoir computing for nonlinear fiber distortion compensation," *Optics Express*, vol. 29, no. 20, p. 30991, 2021.

[137] X. Zuo, L. Pei, B. Bai, J. Wang, J. Zheng, T. Ning, F. Dong, and Z. Zhao, "Integrated Silicon Photonic Reservoir Computing with PSO Training Algorithm for Fiber Communication Channel Equalization," *Journal of Lightwave Technology*, vol. 41, no. 18, pp. 5841–5850, 2023.

[138] V. S. Suite, "VPI Transmission Maker." https://www.vpiphotonics.com/Tools/OpticalSystems/. Accessed 12.08.2024.

[139] Pytorch, "Pytorch: Python Library." pytorch.org. Accessed 26.08.2024.

[140] S. J. Savory, "Digital filters for coherent optical receivers," *Optics Express*, vol. 16, no. 2, p. 804, 2008.

[141] T. Xu, G. Jacobsen, S. Popov, J. Li, A. T. Friberg, and Y. Zhang, "Digital chromatic dispersion compensation in coherent transmission system using a time-domain filter," *2010 Asia Communications and Photonics Conference and Exhibition, ACP 2010*, vol. 18, no. 15, pp. 132–133, 2010.

[142] B. A. Wichmann and I. D. Hill, "Algorithm AS 183: An Efficient and Portable Pseudo-Random Number Generator," *Journal of the Royal Statistical Society (Applied Statistics)*, vol. 31, pp. 188–190, 3 1982.

[143] Pytorch, "Softplus Function." https://pytorch.org/docs/stable/generated/torch.nn.Softplus.html. Accessed 22.08.2024.

[144] S. Masaad and P. Bienstman, "Opto-electronic machine learning network for Kramers-Kronig receiver linearization," *Optics Express*, vol. 32, no. 13, p. 23561, 2024.

[145] D. Orsuti, C. Antonelli, A. Chiuso, M. Santagiustina, A. Mecozzi, A. Galtarossa, and L. Palmieri, "Deep Learning-Based Phase Retrieval Scheme for Minimum-Phase Signal Recovery," *Journal of Lightwave Technology*, vol. 41, no. 2, pp. 578–592, 2023.

[146] D. Orsuti, M. Santagiustina, A. Galtarossa, and L. Palmieri, "Edge-Carrier-Assisted Phase Retrieval At Low CSPR and Low Dispersion Diversity With Deep Learning," *Journal of Lightwave Technology*, vol. PP, no. 8, pp. 1–11, 2023.

[147] X. Li, S. An, H. Ji, J. Li, W. Shieh, and Y. Su, "Deep-learning-enabled high-performance full-field direct detection with dispersion diversity," *Optics Express*, vol. 30, no. 7, p. 11767, 2022.

[148] S. Li, C. Häger, N. Garcia, and H. Wymeersch, "Achievable Information Rates for Nonlinear Fiber Communication via End-to-end Autoencoder Learning," *European Conference on Optical Communication, ECOC*, 2018.

[149] T. Pfau, S. Hoffmann, and R. Noé, "Hardware-efficient coherent digital receiver concept with feedforward carrier recovery for M-QAM constellations," *Journal of Lightwave Technology*, vol. 27, no. 8, pp. 989–999, 2009.

[150] I. Sutskever, J. Martens, G. Dahl, and G. Hinton, "On the importance of initialization and momentum in deep learning," *30th International Conference on Machine Learning, ICML 2013*, pp. 2176–2184, 2013.

[151] S. Masaad, S. Sackesyn, S. Sygletos, and P. Bienstman, "Experimental Demonstration of 4-Port Photonic Reservoir Computing for Equalization of 4 and16 QAM signals," in *European Conference on Optical Communication*, (Frankfurt), pp. 2–5, IEEE, 2024.

[152] S. Masaad, S. Sackesyn, S. Sygletos, and P. Bienstman, "Experimental Demonstration of 4-Port Photonic Reservoir Computing for Equalization of 4 and16 QAM signals," *Journal of Lightwave Technology*, vol. PP, no. 1, pp. 2–5, 2024.

[153] M. S. Faruk and S. J. Savory, "Digital Signal Processing for Coherent Transceivers Employing Multilevel Formats," *Journal of Lightwave Technology*, vol. 35, no. 5, pp. 1125–1141, 2017.

[154] P. J. Winzer, A. H. Gnauck, C. R. Doerr, M. Magarini, and L. L. Buhl, "Spectrally efficient long-haul optical networking using 112-Gb/s polarization-multiplexed 16-QAM," *Journal of Lightwave Technology*, vol. 28, no. 4, pp. 547–556, 2010.

[155] F. M. Gardner, "A BPSK/QPSK Timing-Error Detector for Sampled Receivers," *IEEE Transactions on Communications*, vol. 34, no. 5, pp. 423–429, 1986.