



UNIVERSITY OF TRENTO

DEPARTMENT OF PHYSICS
Doctoral Thesis

PHOTONIC NEURAL NETWORKS FOR SIGNAL EQUALIZATION IN OPTICAL FIBER TRANSMISSION

Supervisor

Prof. Lorenzo Pavesi

Candidate

Emiliano Staffoli

February 2026

To my parents, Cristina and Alessandro

To my light, Anna

Abstract

Optical fibers represent the backbone of the telecommunication infrastructure, providing for long-distance transmission at large bandwidths. The massive use of the Internet in modern society, however, generates an unprecedented traffic demand, pushing research towards new solutions that adapt to different scenarios and costs. This involves an evolution in optical transceivers, which are modules responsible for the transmission and reception of optical signals in fiber. The implementation of higher modulation frequencies, complex modulation formats, and Dense Wavelength Division Multiplexing increases the transmitted bandwidth, but simultaneously worsens linear and nonlinear effects accumulated during fiber propagation. These create distortions in the propagated sequence, introducing the need for recovering its integrity after the transmission, operating an equalization process. Nowadays, this is mainly provided at the transceiver level via Digital Signal Processing, requiring huge computational effort, power consumption, and latency.

In this work, I explore the use of an integrated photonic neural network for efficient signal equalization in Intensity Modulation/Direct Detection links. The device under test implements an 8-tap time-delayed complex perceptron equipped with tunable amplitude and phase weights on each tap. These determine the optical response of the device and are optimized to provide equalization in various transmission scenarios. Contrary to electronics-based solutions, signal processing is operated entirely in the optical domain, thus minimizing latency and power consumption. The device is realized in a Silicon photonic platform, thus compatible with standard CMOS processes for reduced unit cost in mass production. Its small footprint makes it suitable for scalability and integration in the next-generation transceivers.

This design is experimentally tested against Chromatic Dispersion and Self-Phase modulation in 10 Gbaud Pulse amplitude-modulated signals. The training is performed offline by maximizing the eye diagram aperture, verifying the equalization in single- and multi-span links. Simulations allowed studying the scalability of the device to 100G applications, adapting the existing layouts and testing new architectures for future implementations. A comparison with the current state-of-the-art reveals the promising benefits introduced by the proposed all-optical approach, encouraging further steps towards photonic implementations for telecom.

Acknowledgements

Technical Acknowledgments

Among the past members of the Nanoscience Laboratory, I acknowledge Dr. Mattia Mancinelli and Dr. Davide Bazzanella for providing the basis of what has become my Ph.D. work. Among the current members, my gratitude is for Prof. Lorenzo Pavesi, Dr. Alessio Lugnan, and Dr. Stefano Biasi for the constant support offered throughout the years. I am deeply thankful to my colleagues Gianpietro Maddinelli, Stefano Gretter, and Alessandro Foradori for sharing life in the laboratory with me, and for the mutual supervision that has established between us. My gratitude is also to Prof. Pierpaolo Boffi and Prof. Alberto Gatto, whose experience in telecommunications guided me on different occasions. I feel deep gratitude to all the people of the LabTel at the Federal University of Espirito Santo in Brazil, with whom a precious collaboration has developed. A special thanks goes to Prof. Marcelo E.V. Segatto and Dr. Pablo R.N. Marciano for sharing their precious insights about OFDM signals and for the great hospitality demonstrated.

This work has been supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 788793, BACKUP and No 963463, ALPI), and Fondazione CARITRO (POC R2M 2024 project).

Personal Acknowledgments

Durante questi anni, e in particolare nel periodo di scrittura di questo elaborato, ho ricevuto supporto da parte di molte persone. Il loro contributo è stato fondamentale, ma rischia di passare in secondo piano tra i tanti tecnicismi qui riportati. Questi ringraziamenti sono quindi l'occasione per far emergere in questa ricerca la bellezza del lato umano, tanto importante quanto quello scientifico.

Innanzitutto, desidero ringraziare tutti i ricercatori e le ricercatrici del Nanolab con cui ho avuto la fortuna di condividere questo percorso. In particolare, ringrazio ancora

una volta il prof. Lorenzo Pavesi per le possibilità di crescita lavorativa e umana che mi ha offerto nel corso degli anni. Lo ringrazio, inoltre, per l'enorme pazienza e supporto dimostrati durante quest'ultimo periodo, in cui ho visto vacillare molte delle mie certezze.

Tra i membri del Nanolab, ringrazio il mio amico e compagno di laboratorio Alessandro Foradori, che in questi anni mi ha regalato innumerevoli momenti di leggerezza. Ringrazio anche Clara Zaccaria, che è stata per me come una sorella maggiore, donandomi il suo ascolto genuino e i suoi preziosi consigli.

Il mio grazie va anche agli amici mantovani della PGV, ai gonzaghesi del gruppo *Usciamo* e ai membri del club di beach volley (indoor e outdoor). La loro compagnia ha per me un sapore di casa e il tempo trascorso insieme serve a ricordarmi il valore delle relazioni. Ogni forma di interessamento che hanno rivolto verso di me e la mia ricerca, soprattutto nell'ultimo periodo, è stata come una carezza. Forse non ne sono consapevoli, ma sono stati un dono concreto in molte occasioni.

Ringrazio poi tutti gli atleti e le atlete del movimento della pallavolo trentina che ho incontrato in questi anni. In particolare, i miei allenamenti sono stati resi indimenticabili da Giacomo e Federico Chiech, la cui amicizia ha per me un valore inestimabile. Insieme a Giulia Olivieri, si sono presi cura di me senza chiedere nulla in cambio. A loro va la mia riconoscenza.

Tutte le parole del mondo non basterebbero per rendere onore a Chiara Cecchini, Giacomo Lamporesi e Chiara Gioia. Oltre ad essere eccellenti ricercatori e ricercatrici, essi sono ai miei occhi come fratelli e sorelle. Sono stati una presenza costante, fatta di pause caffè, cene saporite e viaggi di ritorno in macchina verso casa. Spero di poter ricambiare, anche solo in minima parte, tutto ciò che mi hanno donato.

I miei genitori, Cristina e Alessandro, sono tra i dedicatari di questa tesi perché tutto, dal mio primo giorno in università alla stesura di questo elaborato, è stato possibile grazie a loro. I sacrifici, la disponibilità ad ascoltare, il loro essere per me *famiglia* non sono di certo passati inosservati. Li considero un esempio da seguire e voglio loro un mondo di bene.

Infine, questa tesi è dedicata anche ad Anna, che in questi mesi mi è stata accanto con un amore incondizionato. La ringrazio di cuore per essersi presa cura di me e aver creduto nelle mie capacità anche quando io non avevo la lucidità per fare altrettanto. Lei è per me *luce*.

Contents

List of Acronyms	xiii
1 Introduction	1
2 Fiber-optic communication	7
2.1 Networks' architecture	7
2.1.1 Core networks	8
2.1.1.1 Coherent modulation	10
2.1.1.2 Towards Space Division Multiplexing	12
2.1.2 Metro networks	13
2.1.3 Access networks	14
2.1.4 Data Center networks	16
2.1.5 Optical Transceivers	17
2.2 Fiber propagation and impairment sources	20
2.2.1 Chromatic Dispersion	21
2.2.2 Self-Phase Modulation	25
2.2.3 Cross-Phase Modulation	26
2.3 Equalization techniques overview	27
2.3.1 Dispersion Compensating Fibers	27
2.3.2 Fiber Bragg Gratings	28
2.3.3 Digital Signal Processing	30
2.3.3.1 FIR filters for Chromatic Dispersion equalization . .	32
2.3.3.2 DSP-based nonlinearity mitigation	34
2.3.3.3 Forward Error Correction	36
3 Photonic Neural Networks	39
3.1 Artificial Neural Networks	40
3.2 Training procedures	42
3.2.1 Backpropagation	43
3.2.2 Adam algorithm	43
3.2.3 Particle Swarm Optimizer	46

3.3	Photonic Neural Networks	47
3.4	Applications to telecommunications	49
3.4.1	Overview of the implementations for signal equalization	50
3.4.2	All-optical time-delayed complex perceptron	53
4	Chromatic Dispersion equalization	57
4.1	Experimental Setup	57
4.1.1	Separation loss function	59
4.2	Premises on experimental measurements	62
4.3	Application to PAM formats	63
4.3.1	Application to 2-level 10 and 20 Gbaud signals	63
4.3.2	Application to 4-level 10 Gbaud signals	65
4.3.2.1	Investigation on the optimal training conditions	65
4.3.2.2	Equalization results	67
4.3.2.3	Physical interpretation of optimal weights	72
4.4	Application to OFDM	75
4.4.1	Optical OFDM basics	76
4.4.2	Experimental results	77
4.5	Comparison with other photonic approaches	78
5	Self-phase modulation equalization	81
5.1	Fiber loop implementation	82
5.1.1	Experimental setup	82
5.2	Experimental results	84
5.2.1	Linear propagation regime	84
5.2.2	Nonlinear propagation regime	85
5.2.2.1	Analysis of outcome weights	87
5.3	Simulated study for future 100G applications	88
5.3.1	Linear propagation regime	90
5.3.2	Nonlinear propagation regime	92
5.4	Comparison with other approaches	96
5.5	Supplementary activities towards the next-generation devices	97
5.5.1	Perceptron with electro-optic phase shifters	97
5.5.2	Exploring new implementations: the Mach-Zehnder cascade . .	100
5.5.3	Equalization with unsupervised learning	101
6	Conclusions	105
6.1	Comparison with the current state-of-the-art	105
6.2	Next-generation optical devices for applications at 100G and above . .	107
6.3	Conclusions	111
	List of publications	115

Appendices	117
Appendix A Experimental Setups and techniques	119
A.1 Optical transmission line up to 125 km	119
A.1.1 Experimental conditions	122
A.1.2 Eye diagram	122
A.2 Fiber loop implementation	123
A.2.1 Measurement procedures	124
A.3 Chip characterization	125
Appendix B Simulated Setup	129
B.1 Code block diagram	129
B.2 PNN devices computational modeling	131
B.2.1 Time-Delayed Complex Perceptron	131
B.2.2 Mach-Zehnder cascade	132
B.2.3 Photonic reservoir with all-optical readout	132
B.3 Simulation parameters	135
Appendix C Photonic Integrated Circuits for 100G applications	137
C.1 Time-delayed complex perceptron	137
C.2 Mach-Zehnder cascade	138
C.3 Photonic Reservoir with all-optical readout	139
Bibliography	141

List of Acronyms

ADC	Analog-to-Digital Converter
ANN	Artificial Neural Network
ASIC	Application-Specific Integrated Circuit
AWG	Arbitrary Waveform Generator
BER	Bit Error Rate
BTB	Back-To-Back
CD	Chromatic Dispersion
CWDM	Coarse Wavelength Division Multiplexing
DAC	Digital-to-Analog Converter
DBP	Digital Back-Propagation
DCF	Dispersion Compensating Fiber
DSP	Digital Signal Processing
DWDM	Dense Wavelength Division Multiplexing
EDFA	Erbium-Doped Fiber Amplifier
FBG	Fiber Bragg Grating
FEC	Forward Error Correction
FFE	Feed Forward Equalizer
FIR	Finite Impulse Response
GVD	Group Velocity Dispersion
IMDD	Intensity Modulation/Direct Detection
ISI	Inter-Symbol Interference
LAN	Local Area Network

MAN	Metropolitan Area Network
MCF	Multi-Core Fiber
MMF	Multi-Mode Fiber
MMI	Multi-Mode Interferometer
MRR	Micro-Ring Resonator
MZC	Mach-Zehnder Cascade
MZI	Mach-Zehnder Interferometer
NMZI	Nested Mach-Zehnder Interferometer
NRZ	Non-Return-to-Zero
OFDM	Orthogonal Frequency Division Multiplexing
OSNR	Optical Signal-to-Noise Ratio
PAM	Pulse Amplitude Modulated
PCB	Printed Circuit Board
PNN	Photonic Neural Network
PON	Passive Optical Network
PR	Photonic Reservoir
PRBS	Pseudo-Random Binary Sequence
PRX	Power at the Receiver
PS	Phase Shifter
PSO	Particle Swarm Optimizer
QAM	Quadrature Amplitude Modulation
RC	Reservoir Computing
RNN	Recurrent Neural Network
ROADM	Reconfigurable Optical Add/Drop Multiplexer
ROSA	Receiver Optical Sub-Assembly
RX	Receiver
SDM	Spatial Division Multiplexing
SMF	Single-Mode Fiber
SNR	Signal-to-Noise Ratio

SOA	Semiconductor Optical Amplifier
SPM	Self-Phase Modulation
SSFM	Split-Step Fourier Method
TDC	Tunable Dispersion Compensator
TDCP	Time-Delayed Complex Perceptron
TLS	Tunable Laser Source
TOF	Tunable Optical Filter
ToR	Top-of-the-Rack
TOSA	Transmitter Optical Sub-Assembly
TX	Transmitter
VCSEL	Vertical Cavity Surface Emitting Laser
VOA	Variable Optical Attenuator
WDM	Wavelength Division Multiplexing
XPM	Cross-Phase Modulation

Chapter 1

Introduction

Since the origin of telecommunications, research has continuously been driven by increasing transmission distance and data rate. A significant boost has come with the introduction of optical fiber-based communications, which represented a ground-breaking paradigm change for efficient data transfer. Since the first commercial implementations of lightwave systems in 1976-1979 [1], traditional transmission lines based on copper wires started being replaced by optical fibers. Using light instead of voltage pulses allowed overcoming the bandwidth limitations imposed by metal wires, mainly induced by interference between adjacent cables and high losses [2]. The Bandwidth-Distance product $B \times L$ was increased by more than six orders of magnitude already in the first 20 years [3]. The technology supporting optical fiber communications has undergone rapid development in the last few decades [4]. Significant milestones are represented by the introduction of low-loss optical fibers [5], and by the invention of Erbium-Doped Fiber Amplifiers (EDFAs) [6, 7], enabling long-distance propagation at wavelengths around 1550 nm (C-band), and Wavelength Division Multiplexing (WDM) for enhanced data rates [8]. As of today, the optical network is composed of more than 6 billions kilometers of fiber (in expansion) [9], and it represents the backbone of the Internet and the associated online services, such as e-commerce, streaming platforms, finance, online payments, cloud computing, storage, and, ultimately, AI tools [10–14]. Nowadays, optical fibers are deployed all around the world and cover a huge variety of transmission scenarios, from short-reach links in intra-data center applications to ultra-long-haul submarine transmission [15].

A central role in the telecom infrastructure is played by the optical transceivers [16]. As the name suggests, these devices integrate the functions of a transmitter and a receiver in a single platform: on one side, a modulated laser source sends optical signals in a fiber link for propagation, and on the other side, a detector acquires the incoming optical information. These appear as plug-in devices for Ethernet switches, routers, firewalls, and network interface cards. In this sense, they represent the main interface between the optical and the electrical domain. Transceivers operating in short-reach applications generally work with Intensity Modulation/Direct Detection (IMDD) protocols [17], namely, the information is encoded in the optical signal simply by modulating

its intensity. These devices feature lower costs and lower complexity in the hardware implementation compared to their coherent counterpart. By contrast, as the name suggests, this latest generation of transceivers exploits coherent modulation formats, where both the amplitude and the phase of the optical signal are used for encoding. This provides more versatility and bandwidth efficiency, which are requested for long-haul propagation, at the price of higher energetic power consumption and more sophisticated hardware implementations [18].

Other than the canonical transmission and reception functions, transceivers also feature Application-Specific Integrated Circuits (ASICs) implemented on board. These drive optical components, operate Digital Signal Processing (DSP) on the incoming and outgoing data streams, and regulate the data transfer with the hosting platform [19]. Central to DSP functionality, sophisticated algorithms are designed to address channel equalization, compensating for signal impairments arising from both linear effects, such as Polarization Mode Dispersion, Symbol Timing Offset, Optical filtering, and Chromatic Dispersion (CD), and nonlinear effects such as Self-Phase Modulation (SPM) and Cross-Phase Modulation (XPM) [20]. These induce distortions in the propagating optical pulses, whose associated information is wrongly decoded at the receiver and translates into an incorrect bit identification. Before the introduction of DSP, other techniques have been adopted to compensate for the distortions. Against CD, for example, Dispersion Compensating Fibers (DCFs) are still widely used [21], despite their application being affected by the high production costs, high losses, added latency, and non-tunability (being static devices after the implementation). Another solution is provided by Chirped Fiber Bragg Gratings (FBGs) [22], which drastically reduce the latency, but often require high power consumption for thermal or electrical tuning [23]. Among these and other alternatives, DSP represents the most solid approach, providing versatility, high accuracy, and customizable solutions to equalization problems [24, 25]. However, the onboard electronics represent the major contribution to the transceiver's cost and operational power consumption, becoming more and more significant with the bit rate supported by the device. The computational effort required by equalization algorithms significantly contributes to the overall energy needs [26], which for commercially available 800 G transceivers reaches up to 18 W, and it is forecast to increase up to 30 W for the next generation of 1.6 T devices [27]. The transceivers' market value is increasing year by year, as presented in Figure 1.1, thus addressing their drawbacks in terms of production costs, power consumption, optical-to-electrical, and analog-to-digital conversions becomes of primary importance to ensure a sustainable development of the telecommunication infrastructure and cope with the increasing bandwidth request.

In recent years, an alternative to the standard equalization algorithms has emerged. This exploits the paradigms and methodologies of Machine Learning, and, more specifically, of Artificial Neural Networks (ANNs). These consist of computational models bio-inspired by the structure and function of neurons in the human brain, mimicking the typical learning process based on experience [28]. Biological neurons find their analog

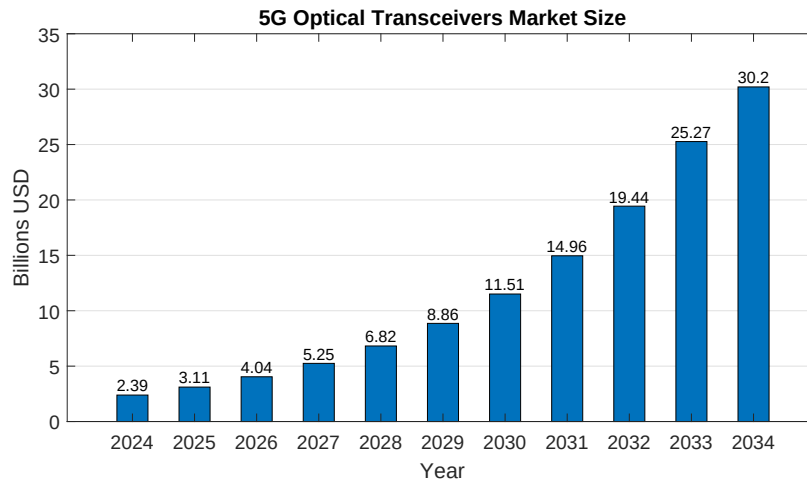


FIGURE 1.1: Global 5G optical transceivers market size. Source: <https://www.precedenceresearch.com/5g-optical-transceiver-market>

in the artificial neuron, which weights and combines multiple input signals and produces a single output via a nonlinear activation function [29]. Multiple artificial neurons connected form ANNs, which can feature different levels of complexity and recurrence, depending on the neurons' topology [30–32]. Given a specific architecture, the overall behavior is determined by the weights applied within each neuron, and the optimal setting for a specific task is found during the learning process. Many software-based implementations of ANNs have been proposed in the context of channel equalization [33–37], proving the validity of this alternative approach to traditional algorithms. However, they still feature the limitations of DSP in terms of power consumption, complexity, and latency introduction. In this sense, a further step towards efficient equalization is obtained by moving the computational platform to the optical domain and implementing Photonic Neural Networks (PNNs). Despite sharing the same paradigm as the standard ANNs, the linear and nonlinear stages of each neuron are implemented by optical structures and processes, thus allowing for signal processing directly performed on optical signals [38]. In this way, PNNs can exploit all the properties offered by light-based computation in terms of bandwidth, speed of propagation, latency reduction, and parallelization (e.g., WDM). These features gave rise to many applications of PNNs, which nowadays are used as accelerators for machine learning [39], speech-recognition [40], optical computing [41], real-time analysis in bio-medical applications [42], and image classification [43]. A straightforward implementation follows in telecommunications, where they share the same optical domain offered by optical fibers, and therefore, the processing is operated directly on propagated signals [44]. In this sense, PNNs provide a suitable platform for channel equalization, thus avoiding, or at least limiting, the power consumption and latency associated with DSP.

In this work, we propose a fully optical solution for linear and nonlinear distortion equalization in IMDD systems. Our solution, implemented via a PNN, aims to

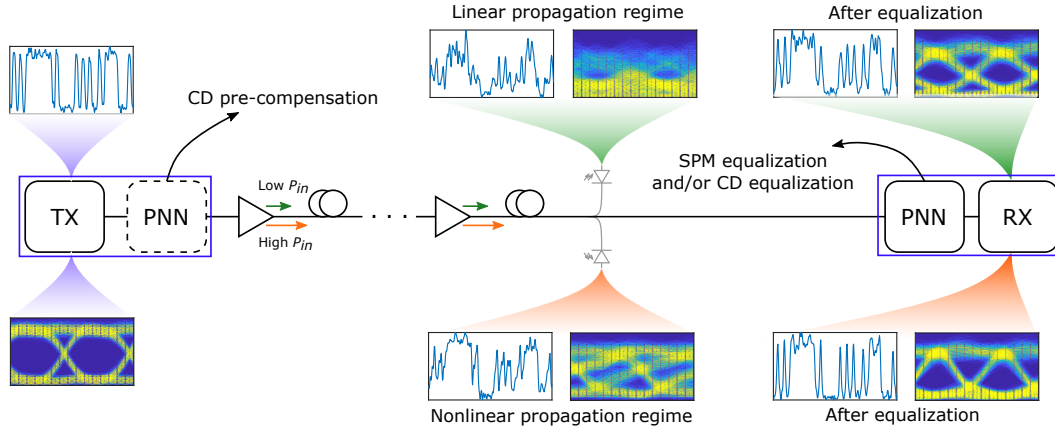


FIGURE 1.2: Schematics of a typical Transmitter/Receiver (TX/RX) optical link featuring a PNN as equalizing unit. The top-half plane describes a linear propagation regime (low input power P_{in} in fiber), where CD-related distortions accumulate. The lower-half plane describes a nonlinear propagation regime (high P_{in}), where SPM contributes to channel impairments. The optical traces and corresponding eye diagrams are presented in different observation points across the line. Possible implementations of the PNN are also shown: when used on the TX side, it allows for CD pre-compensation, while on the RX side, it acts against both linear and nonlinear impairments.

eliminate the need for DSP in these systems, significantly reducing costs and energy demands while preserving their simplicity and practicality for short-distance communication. This shift can redefine the economic and technical viability of the IMDD paradigm, unlocking new possibilities for efficient and sustainable telecommunications. Silicon photonics represents the platform of implementation for our integrated circuit, offering full compatibility with the standard CMOS processes for minimized unitary cost and scalability in large-scale production. Its miniaturized design makes it suitable for integration in the next generation of optical transceivers, with possible implementations in a typical Transmitter (TX)/Receiver (RX) line as illustrated in Figure 1.2. It can be used at the TX side for signal pre-distortion to tackle CD, or as an equalizer at the RX side, tackling CD, SPM, and XPM.

This work explores, from both an experimental and a simulative point of view, the equalization capabilities provided by the implemented PNN design, which aims for simplicity, transparency, and performance. The results are compared with the current state-of-the-art, finally tracing the path for the next-generation devices and future implementations in a real-case scenario.

The content of this manuscript is organized as follows.

Chapter 2 presents an introduction to fiber-optic communications. It provides an overview of the telecom infrastructure, divided into Core, Metro, Access, and Data Center networks, together with the main technological implementations and needs specific to each level. Starting from the pulse propagation equation in fiber, the main linear

and nonlinear impairment sources are then presented, followed by an insight into well-established equalization techniques, distinguished in hardware and software implementations.

Chapter 3 introduces ANNs drawing an analogy with the biological counterpart. Subsequently, some optimization algorithms used for training are introduced, with particular focus set on Particle Swarm Optimizer (PSO) and the Adam algorithm, whose performances are later compared in this work. The chapter then restricts to PNNs, highlighting the advantages of light as a message conveyer with respect to traditional electronics-based versions of ANNs. An overview of PNNs layouts adopted for telecom applications is presented, finally introducing the Time-Delayed Complex Perceptron (TDCP), which is the main object of study of this work.

Chapter 4 presents the results for CD equalization provided by the same TDCP. It illustrates the main features of the setup used for linear transmission and describes the custom-built loss function used for training. An in-depth analysis of the experimental outcomes for 10 Gbaud 2- and 4-level Pulse Amplitude Modulated (PAM), and Orthogonal Frequency Division Multiplexing (OFDM) formats is presented, including a physical interpretation of the optimal weights.

Chapter 5 extends the use of the TDCP to SPM equalization at 10 Gbaud. The optical fiber loop used to emulate a multi-span link is described, followed by the associated experimental results. Subsequently, the analysis focuses on future 100G applications, providing the simulated equalization results of an adapted PNN device to higher modulation frequencies. These outcomes and the previous ones obtained at 10 Gbaud are compared in terms of performance with the literature. In the end, some experimental spin-off activities are illustrated, aiming to provide further indications for the next-generation devices for a reduction in insertion losses, power consumption, and training complexity.

Chapter 6 analyzes the current state-of-the-art technologies and the potential innovations of our approach, also targeting the final application scenario. The new PNN designs for 100G applications, resulting from the previous chapter, are then presented. These include an adapted TDCP, a lattice filter-based implementation, and a reservoir-based approach with all-optical readout. Finally, the chapter summarizes the content of this manuscript.

The results presented in this work were obtained during my research as a Ph.D. student at the Nanoscience Laboratory of the University of Trento. During my stay, I was involved in both experimental and simulation activities. On the experimental side, I was in charge of the apparatus for optical transmission, which I maintained in operation for the whole duration of the research. Overall, the setup allowed me to test 3 generations of photonic integrated circuits implementing 4 different PNN layouts for channel equalization. On the simulation side, I developed the code for optical transmission in arbitrary conditions, starting from a very early draft by Dr. Mattia Mancinelli. I progressively enriched it with new features, exploiting it to access the equalization capabilities of different PNNs layouts. Ultimately, I used the simulated environment to optimize the PNN designs for 100G applications, which I then inserted in a new mask

for the next-generation integrated circuits. The mask has been implemented in a joint project together with other members of the Nanoscience laboratory.

Chapter 2

Fiber-optic communication

This chapter defines the working context for this research, which, in this case, is represented by optical transmission lines. These are described first from a hardware point of view, introducing the different optical network segments (Core, Metro, Access, and Data Center) and optical transceivers. The problem of distortions in optical signals is then treated starting from the propagation equation. Different impairment sources, as CD, SPM, and XPM, are described, emphasizing the role of the bitrate in each phenomenon. An overview of the currently implemented equalization techniques is then provided, distinguishing between hardware solutions such as DCFs or FBGs, and DSP-based solutions.

2.1 Networks' architecture

Nowadays, the massive Internet utilization in our society is reflected in the data exchange trends. Since 2020, global data traffic has more than doubled, and in 2024 the volume increased by 15% compared to the previous year [45]. To support such massive requirements, huge optical fiber networks are deployed all around the world, satisfying the needs on different levels, from ultra-long distance and ultra-large bandwidth transmission to widespread fiber-to-home services.

Networks can be organized into four broad segments based on geography and functions [46], as depicted in Figure 2.1. Core networks are the highest in hierarchy and provide the largest bandwidth and transmission reach, making them suitable for interconnections between far-apart Central Offices or data centers. Access networks are the lowest hierarchy level and provide Internet connections to the final users; for this, they are also identified as the "last mile" segment. Metro networks transport data within a single city or region, thus serving as a bridge between core and access. Besides these three segments, data center networks exist, handling traffic between servers and storage units within these facilities. Each segment faces different challenges and needs, therefore requiring different technological implementations, which are overviewed in the following subsections.

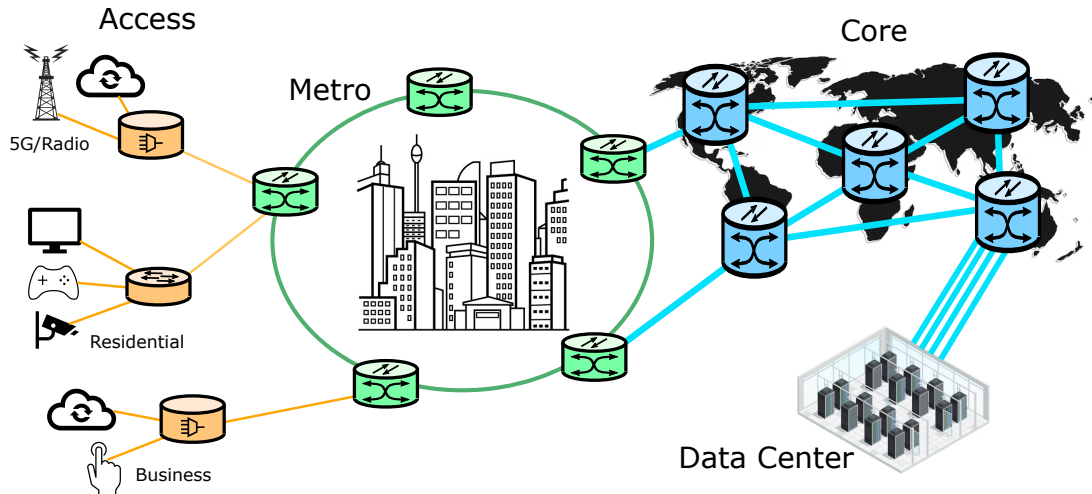


FIGURE 2.1: A typical optical network architecture with its segmentation in Core, Metro, Access, and Data Center networks. Colored cylinders generically represent network nodes, mainly constituted by Reconfigurable Add/Drop Multiplexers, Routers, and switches, used across the networks to handle optical traffic.

2.1.1 Core networks

The core (or long-haul) network consists of high-speed links and advanced routers and switches, which together form the backbone of global Internet traffic. Among the four network segments, this provides the largest transmission bandwidth (Multi Terabit per second) and the longest reach (> 1000 km), allowing for ultra-fast data transmission between distant geographic areas [46]. The transmission infrastructure is constituted by terrestrial and submarine optical fibers deployed all over the world, ensuring peak performance suitable for applications in Data Center Interconnects, and links between large-scale Local Area Networks (LANs) and Metropolitan Area Networks (MANs) [47]. Optical fiber links are implemented to satisfy the requirements of resilience, reliability, and redundancy, so that even in the case of failure or outage, the data transfer is still guaranteed. An interruption in the data flow would have a massive impact on business and users [48, 49], since the connections at lower levels rely on core segments. Given their strategic role in many economic sectors, Internet Service Providers constantly seek expansion and technological upgrades on these networks to satisfy the market needs.

In the core segment, transmission over wide ranges at large bandwidths is enabled by the combination of many technological progresses. First, long-distance data transfer is based on amplified multi-span links [46], which is typically implemented as schematized in Figure 2.2. The transmission line is constituted by repeated units (separated by dashed lines) made of a fiber span and an EDFA, which is used to compensate for accumulated losses and restore the optical power. In this context, the mentioned EDFAs represented true game changers, allowing for an all-optical amplification in the third telecom window (around 1550 nm), which exactly corresponds to the spectral region

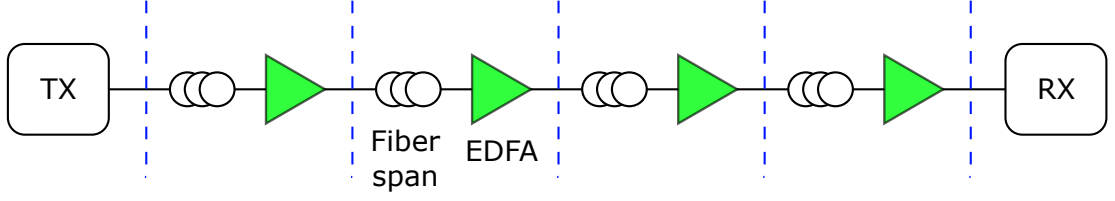


FIGURE 2.2: Optical multi-span transmission line implementation. The optical signal originates at the transmitter (TX) and propagates through multiple units made of a fiber span and an EDFA, finally reaching the receiver (RX).

of minimum losses for standard Silica Single-Mode Fibers (SMFs) [6, 7]. Other than providing for signal amplification, EDFAs are also a source of optical noise, which is introduced in the line via amplified spontaneous emission. After a propagation across n_l units (each composed of fiber span and EDFA), the optical signal's spectrum features a modified Optical Signal-to-Noise Ratio (OSNR) condition predicted by [19, 50]

$$OSNR = P_{out} - l_s - NF - 10\log_{10}(n_l) - 10\log_{10}(h\nu_0\Delta\nu_{bw}). \quad (2.1)$$

Here, P_{out} is the output power from each amplification stage, l_s is the loss accumulated in a fiber span, NF is the noise figure of the amplifiers, $\Delta\nu_{bw}$ is the bandwidth chosen for OSNR evaluation, ν_0 is the carrier frequency, and h is the Planck constant. The specific features of each line have, thus, to be carefully designed, since the optical signal must comply with precise OSNR specifications at the end-of-line receiver to ensure proper quality in the decoding.

Another enabling technology for core networks is represented by WDM, whose working principle appears in Figure 2.3. Many transmitters are used in parallel to encode information in separate wavelengths $\lambda_1, \dots, \lambda_N$ (or channels) with a fixed detuning between each other. The individual signals are then multiplexed in a single fiber, which they share during propagation. At the receiver side, signals are demultiplexed, sent to different detectors, and processed separately [51]. Depending on the spacing between the wavelengths, WDM can be distinguished in two main patterns, namely Dense Wavelength Division Multiplexing (DWDM) and Coarse Wavelength Division Multiplexing (CWDM). The former is generally operated with 40 or 80 channels spaced by a 100 or 50 GHz grid, providing, as the name suggests, the highest information density and thus the highest efficiency. The ITU-T fixes the standards for channels' features in terms of wavelengths and spacing, which for the long-haul transmission are set to C-band (1530-1565 nm), and possibly L-band (1565-1625 nm), with spectral spacing of 100/50 GHz (0.8-0.4 nm) [52–54]. On the other hand, CWDM features 16 channels spanning across the second and third transmission windows (1310-1550 nm) [55], thus corresponding to a much wider detuning, up to 3750 GHz. Since fewer wavelengths have to be handled simultaneously, this solution favors lower implementation costs and reduced hardware complexity [56], however, sacrificing bandwidth efficiency. Moreover, the amplification

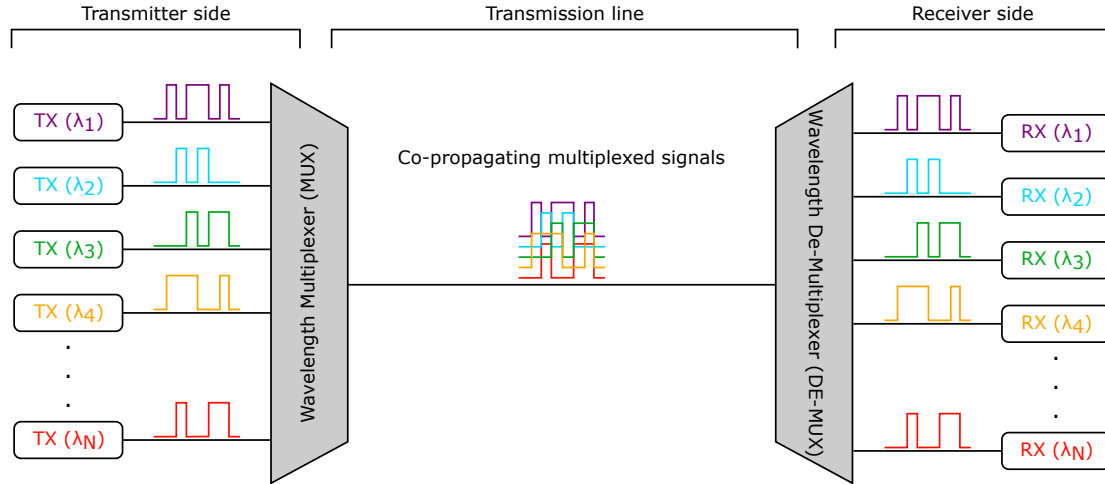


FIGURE 2.3: WDM transmission scheme. At the transmitter (TX) side, different sources work in parallel to encode optical sequences on equally spaced wavelengths ($\lambda_1, \dots, \lambda_N$). These are then combined (multiplexed) and propagated in a single fiber. At the receiver (RX) side, a demultiplexer separates the various wavelength components, which are then detected and processed separately.

of optical channels so separated in the spectrum is still not possible with a single ultra-broadband optical amplifier, in contrast to DWDM applications where a single EDFA provides for amplification across the whole spectral region of interest. In view of peak performances sought in core networks, these two aspects make the CWDM paradigm unsuitable compared to DWDM for long-haul and large bandwidth applications.

2.1.1.1 Coherent modulation

The choice of the modulation format represents another critical aspect of long-haul, highly-efficient transmission. In the first optical lines implementations, the standard was represented by IMDD protocols, namely, the information is encoded in the optical sequence by modulating its amplitude, resulting in formats such as Non-Return-to-Zero (NRZ) or PAMm, depending on the number of adopted levels. Retrieving the information at the receiver follows straightforwardly after applying the square modulus $|\cdot|^2$ at the end-of-line detector.

More efficient formats have later been adopted, so that the current standard for core network applications is represented by coherent schemes. Here, the information is encoded in the optical sequence by modulating both the amplitude and the phase, thus increasing the transmitted bits per symbol. The detection process is, in turn, more elaborated, and it is based on a precise phase relation (i.e., coherence) between the incoming optical signal and a local oscillator placed at the receiver [57], as depicted in Figure 2.4. At the end of the line, the electric field associated with the propagated optical signal can be written as

$$E_S = A_S \exp[-i(\omega_S t + \phi_S)], \quad (2.2)$$

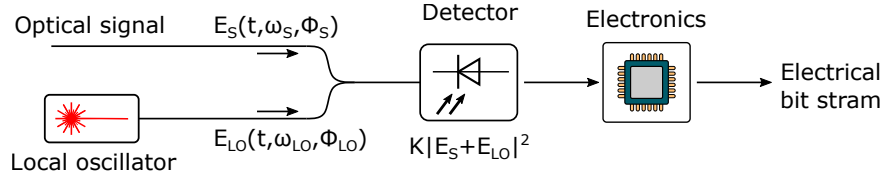


FIGURE 2.4: Coherent receiver scheme. The received optical signal is combined with a CW contribution from a Local oscillator. The signal resulting from the interference is detected and electronically processed to extract the bit stream.

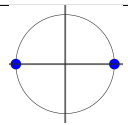
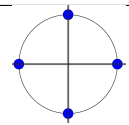
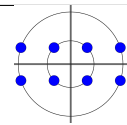
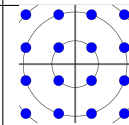
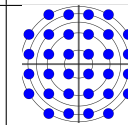
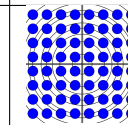
Modulation format	BPSK	4QPSK	8QAM	16QAM	32QAM	64QAM
Bits per symbol	1	2	3	4	5	6
Constellation						

TABLE 2.1: Examples of coherent modulation formats. BPSK: Binary Phase Shift Keying, QPSK: Quadrature Phase Shift Keying, QAM: Quadrature Amplitude Modulation.

where A_S is the amplitude, ϕ_S is the phase, and ω_S is the carrier frequency. This signal is combined with a CW laser generated by a local oscillator at a frequency ω_{LO} in the form

$$E_{LO} = A_{LO} \exp[-i(\omega_{LO}t + \phi_{LO})]. \quad (2.3)$$

The vector notation for E_S and E_{LO} has been dropped, assuming both have the same polarization. After the combination, the optical signal reaches the photodetector, which returns an electrical trace proportional to the optical power $P(t)$ resulting from the interference, namely

$$P(t) = K|E_S + E_{LO}|^2 = KA_S^2 + KA_{LO}^2 + 2KA_SA_{LO} \cos(\omega_{IF}t + \phi_S - \phi_{LO}) \quad (2.4)$$

where $\omega_{IF} = \omega_S - \omega_{LO}$ is known as intermediate frequency and K is a proportionality constant. The so-obtained profile $P(t)$ is then electronically processed to extract the final bit stream [57]. From Equation 2.4, it is evident that the electrical trace contains information both on the amplitude and the phase of the incoming optical signal E_S , contrary to IMDD where any dependence on ϕ_S disappears. At the TX side, various multi-level modulation formats (presented in Table 2.1) can be implemented, such as Binary Phase Shift Keying (BPSK), Quadrature Phase Shift Keying (QPSK), and Quadrature Amplitude Modulation (QAM), often combined with polarization multiplexing.

Compared to IMDD, coherent modulation features increased implementation and

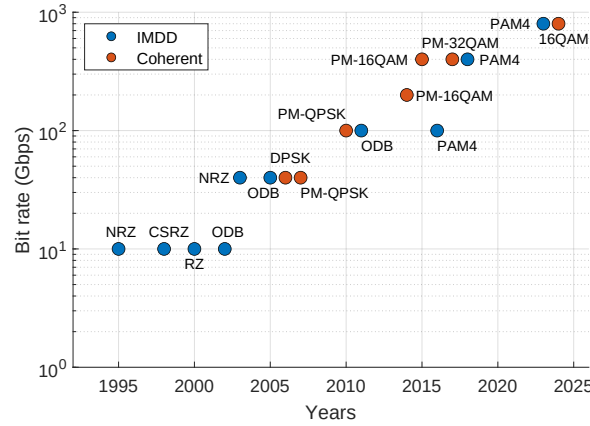


FIGURE 2.5: Historical trend of the bit rate achievable by optical transceivers equipped with IMDD (blue) and coherent (orange) transmission lines. Data is taken from [19, 64–66].

operative costs due to the complexity of the TX/RX scheme. Higher expenses are justified, however, by a receiver sensitivity improvement up to 20 dB compared with IMDD systems [58], and a higher bit rate obtained by encoding data in two dimensions. Furthermore, the use of coherent detection allows a more efficient use of fiber bandwidth by increasing the spectral efficiency of DWDM lines [59]. A strong push to coherent light-wave systems has been provided in the early 2000s by the advancements in DSP [60, 61]. With it, the development of coherent-compatible Forward Error Correction (FEC) [62] and polarization tracking [63] dictated the rapid ascent of coherent modulation in long-haul transmission. Figure 2.5 shows a historical trend of the bit rate achievable by optical transceivers with different modulation formats. In general, coherent systems reached high data rates earlier than IMDD technologies, but in recent years their performances in this sense have largely converged. Nevertheless, a comparison between the two references for 800G in the plot highlights that coherent solutions offer dramatically longer transmission reach, on the order of 500 km [64] against the 10 km for an equivalent bit rate IMDD implementation [65]. For this reason, coherent modulation is the preferred solution for long-haul, high-demanding scenarios.

2.1.1.2 Towards Space Division Multiplexing

Fiber optics capacity in core segments nowadays exceeds 50 Tbps, which is achieved with a combination of all the previously described key-enabling technologies [67]. However, current trends suggest that even such impressive values won't soon be able to sustain the enormous load applied to long-haul ultra-wide bandwidth links. It is therefore necessary to introduce new transmission paradigms that not only increase the capacity even more, but also reduce the cost per bit of transmitted data. In recent years, research has been exploring more and more the Spatial Division Multiplexing (SDM),

which aims to increase the number of parallel spatial channels used to transmit information. SDM can appear in different implementations [67], two of which are presented here.

SMFs nowadays represent the standard for long-distance transmission. Their narrow core supports the transmission of a single mode, which corresponds to a stationary solution of the propagation equation (see Section 2.2). This approach is preferred for minimizing signal loss over long distances and for preventing distortions induced by modal dispersion [67]. This occurs when multiple co-propagating modes are characterized by different propagation velocities, resulting in an enlargement of the signal in time. The counterpart of SMFs is represented by Multi-Mode Fibers (MMFs), constituting a first example of SDM. These fibers are typically engineered with a larger core or a higher refractive index to support the propagation of many optical modes. Each of them is used as an independent information carrier, drastically increasing the link capacity, with experimental demonstration already available for data rates above 270 Tbps in a 1000 km link [68]. The main drawbacks for Multi-mode propagation are represented by inter-modal cross-talk and the aforementioned modal dispersion, whose handling requires huge DSP effort that still prevents their commercial implementations [69].

Side to MMFs, recent years saw the development of Multi-Core Fibers (MCFs). These realize SDM by aggregating multiple independent cores within a single, shared outer cladding, instead of the single core found in traditional fibers. Each core, thus, acts as an independent channel, allowing for parallel propagation with minimal cross-talk between the light guides. This technology results largely compatible with current cabling technology, promising increased transported capacity thanks to SDM without extra footprint compared to the standard cables [70]. Transmission capacities above 45 Tbps in 12000 km links have already been demonstrated [71], with theoretical estimations reaching more than 500 Tbps in 8000 km [72]. Further advantages brought by MCFs are in the integration of key components, such as amplifiers and switches, to manage the multiple spatial channels efficiently. This integration is crucial for maintaining the cost and space advantages of SDM while avoiding the deployment of separate, bulky components for every single core [70].

2.1.2 Metro networks

The second network segment is represented by Metropolitan Area Networks (MANs, or Metro networks), whose primary role is to connect the access to the core networks. In this sense, they represent a bridge between the long-haul ultra-high bandwidth transmission proper of the backbone segment and the final Internet users. MANs aggregate data from multiple access networks, facilitating communications within local areas, and also supports data injection from the core segment [73]. Since MANs are called to interface with different telecom branches and to comply with all of the standards involved, they result in being one of the most heterogeneous segments in terms of architecture and traffic characteristics.

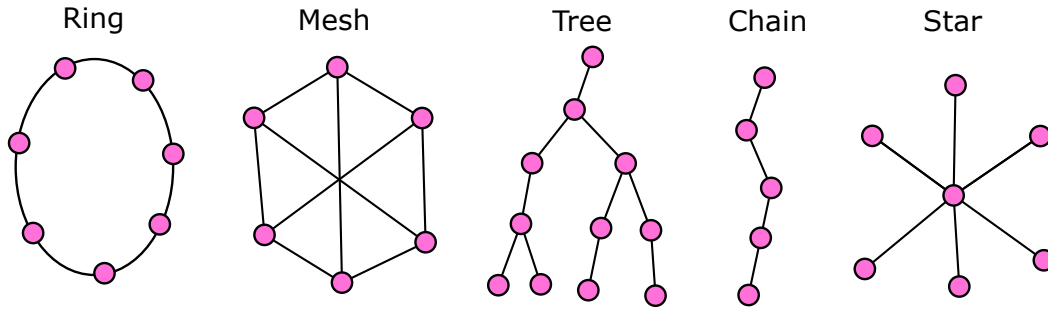


FIGURE 2.6: Examples of optical network topologies. Ring and mesh topologies are widely used in metro applications, while tree, chain, and star structures are more common in access segments.

Metro networks are organized in different topologies, namely, connections between network nodes are arranged differently according to the specific needs [74]. The most widely used topology is the Ring, reported in Figure 2.6, which serves the purpose of reliability (if a link breaks, traffic can still be routed through the other direction of the ring), redundancy (having multiple paths for data transmission increases resilience in case of failures), and scalability (ring networks can be easily expanded by adding more nodes to the loop). Another common network topology is the mesh, which shares similar features and purposes with the ring architecture. On the other hand, the tree, chain, or star topologies are more suitable for access networks, setting a link between the high-capacity metro rings and the final users [74]. Fundamental components in MANs are the Reconfigurable Optical Add/Drop Multiplexers (ROADMs), which handle signal adding or dropping to a specific fiber link in correspondence of the network's nodes [75]. In the context of WDM, they operate by separating different spectrum components via wavelength splitters and then routing the signal towards other fibers via wavelength-selective switches.

MANs are engineered for short- and medium-range connections, typically under 200 km and often confined to a single optical span without amplification, roughly 80 km. The traffic they carry spans a broad spectrum of bandwidths, from rates below 1 Gbps, still the most common among client ports, to 800 Gbps, towards which many providers are evolving. The heterogeneity typical of metro networks also interests the modulation formats, since coherent and IMDD schemes co-exist [76]. The former is employed at the higher end of performance, maintaining its role as the prevailing transmission strategy, frequently paired with DWDM in the C-band for greater efficiency. The latter offers a perfect integration with access networks, and it is used for low data rate applications, keeping costs and system complexity low. It often implements O-band transmission (1260-1360 nm), as it will be discussed for access applications in the next section.

2.1.3 Access networks

The final connection between users and their Internet service providers is constituted by access networks. Among the various segments, they are the most widespread in

urban areas, featuring high capillarity in bringing the last-mile connection to private houses, business services, companies, etc. A fundamental vehicle for data transfer is still represented by optical fibers, which drive the information from the metro segments to the users. However, within this context, the optical domain must interoperate with legacy technologies from existing installations, such as wired connections (DSL), as well as with wireless systems commonly deployed in urban environments, including Wi-Fi and LTE [77].

In an access network, connections can be included in two main categories, based on the number of links set between an access node and the terminal user. In a Point-to-Point (P2P) connection, these two are directly connected, which is typical of a star network topology (see Figure 2.6). Since no re-routing options are available and the starting and end points are always well defined, it is not necessary to introduce any concept of multiplexing or multiple accesses, as the various connections are reserved for the single user [78]. On the other hand, in a Point-to-Multi-Point (P2M) connection, the access node features a single connection to N terminal users, which is typical of a tree network topology. Multiplexing techniques are necessary in this case, since the optical information directed to different users shares some segments during propagation. In the end, the terminal user is reached via a cascade of optical power splitting stages [78].

Despite the backbone transport in access networks being represented by optical fibers, users receive, in many cases, the Internet signal via connections other than optical (e.g., copper cables, LTE). This last segment acts as a bottleneck, strongly affecting the download and upload bit rate. A good practice is to keep non-optical links as short as possible in an attempt to maintain a large bandwidth along the whole line, but the maximum performance is undoubtedly reached with a direct optical fiber connection to the users [79]. Three distinct technologies are available nowadays [80]: with Fiber to the Cabinet (FTTC), the optical link reaches a Serving Area Interface where it is converted to the electrical domain and proceeds to the final user via wired electrical connections, limiting the speed to 200 Mbps. Fiber to the Building (FTTB) reduces the electrical connection length, thus allowing for higher speeds from 100 Mbps to 1 Gbps. Finally, the maximum performance is achieved via Fiber to the Home (FTTH), with a speed greater than 1 Gbps.

In access networks, optical links are much shorter compared to those deployed in other segments, and thus, they require lower hardware complexity to carry the signals. This results in the implementation of Passive Optical Networks (PONs), namely, they use only unpowered devices for the transport, possibly having active components exclusively at both ends of the connection (respectively at the access node and the end user). Lower complexities make PONs more robust against disruptions and failures, simplify maintenance, and reduce operational power consumption [81].

PONs standards are fixed by ITU-T and IEEE [82], resulting in various levels of implementations and bit rates that adapt to the traffic requests of specific areas [77]:

- G-PON is currently the most applied standard, with bit rates up to 2.5 Gbps;
- XG-PON is the evolution of G-PON, with bit rates up to 10 Gbps;

- NG-PON introduces the use of 4 wavelengths WDM to enlarge the transmitted bandwidth for a total bit rate of 40 Gbps;
- HS-PON is under discussion, foreseeing the upgrade to higher bit rates up to 25/50 Gbps for each channel.

Further lowering of costs can be obtained by properly selecting the operational wavelengths for the optical links. For reaches longer than a few hundred meters, transmission is done in the O-band (1260-1360 nm), which corresponds to the zero-dispersion region for standard SMFs. In these conditions, CD, which in general acts as a source of distortions as we'll see in Section 2.2.1, is naturally absent, thus minimal use of DSP is required for the equalization. On the other hand, for distances up to a few hundred meters, as is the case also for the last connection to terminal users or intra-data center applications, the first telecom window (800-900 nm) propagating in multi-mode fibers represents the standard [83]. The implementation of such a wavelength range for longer distances is inconvenient due to the high propagation losses, but in short links where losses are contained, it provides a cost-effective solution. Optical components operating at these wavelengths (e.g., Vertical Cavity Surface Emitting Lasers (VCSELs)) are, indeed, economical and suitable for low unitary cost in mass production [84].

In addition, the use of the IMDD paradigm with two- and four-level PAM in PONs contributes even more to simplifying hardware requirements, especially compared to coherent detection. However, the recent advances and maturity in DSP techniques are raising discussions about a gradual transition to coherent PONs [85]. Higher power consumption and costs are justified by the high spectral efficiency provided by coherent modulation, which could help in satisfying the increasing data traffic demand [86].

2.1.4 Data Center networks

Nowadays, every web-based application, from cloud computing to storage, communications, and streaming services, is made possible thanks to the data centers. These consist of physical facilities that house an organization's IT infrastructure. They include servers, storage drives, and networking equipment, overall allowing to archive, process, and manage data [87]. An enormous amount of information is handled in such facilities, both in terms of fan-in/fan-out streams and in-site processing, thus requiring the appropriate technology for each scope. Inter-data center connections are set by the core network (Section 2.1.1), providing long-haul efficient transmission towards other facilities or network segments. Intra-data center connections, on the other hand, link devices, servers, and storage systems at short distances (up to 1 km) within a single facility, thus facing much different challenges.

Data centers aim for the maximally efficient distribution of tasks and storage among different computing units. This requires the maximum level of interconnection between them, which is typically achieved via local networks as presented in Figure 2.7. A standard implementation consists of a tree topology [88], which shows a layered structure similar to that already presented for terrestrial networks in Figure 2.1. Within a local data center network, the backbone is represented by its core level, which is designed to

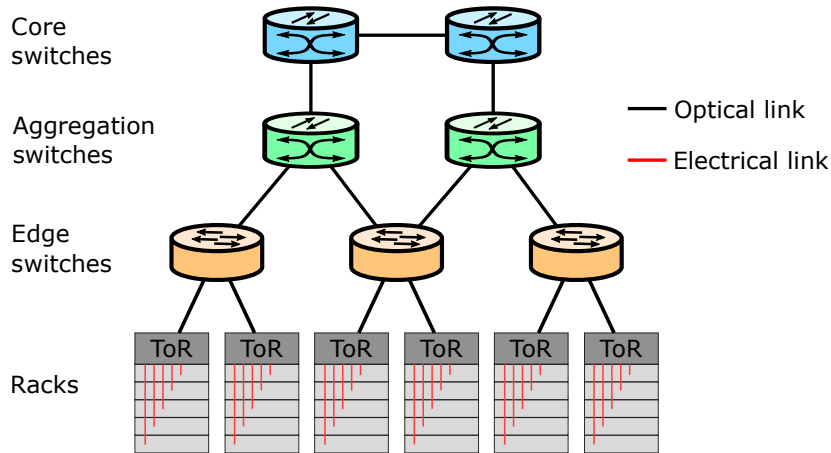


FIGURE 2.7: Typical data center tree architecture organized in core, aggregation, and edge layers. Top-of-Rack (ToR) switches interface the servers with the edge switches [88].

handle large volumes of traffic and forward it quickly to the destination across the facility. At the bottom of the topology is the edge layer, which is the closest to the servers and connects them to the network. The servers are housed in racks and connected to Top-of-the-Rack (ToR) switches, which are at the interface with the edge layer. Some network implementations also feature an intermediate aggregation level, which handles traffic from multiple edge switches before forwarding it to the core.

In this context, electronics and optics work in symbiosis to provide peak performance in data handling. ASICs implement the switches at the ToR level, and standard copper cables provide for short intra-rack connections up to a few meters. On the other hand, longer links are covered in the optical domain, which allows overcoming the known limitations of electronic transport [87]. As already introduced in the previous section, optical transceivers for data centers typically operate in the first telecom window (800-900 nm), using low-cost modulation VCSELs as light sources and Multi-mode fibers as the propagation medium. This combination provides the best overall link cost for up to 25G per lane and reaches up to a few hundred meters. Low unitary cost per link is a key component for data center applications, for which a crucial aspect is represented by scalability. Nowadays, a single data center can achieve bandwidths above 10 Pbps [89]; however, Internet service providers are constantly seeking its expansion. In this sense, major challenges consist of maintaining a high performance while addressing costs (hardware, power consumption,...) and energy efficiency.

2.1.5 Optical Transceivers

Transport networks at every level need an interface between the optical and electrical domains. This is provided by optical transceivers, which enable the transmission and reception of optical signals at the edge of fiber links [19]. All the transceivers feature the same internal structure as illustrated in the schematics of Figure 2.8, with the main

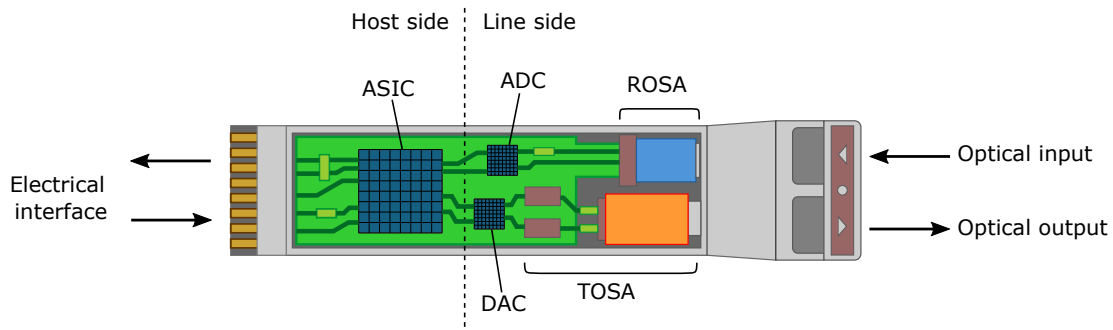


FIGURE 2.8: Schematic of a typical optical transceiver module with the main electrical and optical components in SFP28 form factor.

components being the Receiver Optical Sub-Assembly (ROSA), the Transmitter Optical Sub-Assembly (TOSA), and ASICs. This arrangement forms the optical transmit and receive interfaces on the line side, and a bidirectional electrical interface on the host side.

Transceivers appear in multiple packaging specs, which define their internal architecture and main features [19]. Let us briefly describe a transceiver, taking the SFP28 paradigm as an example. Two fiber cables directly connected to the metal housing are responsible for data transfer in the optical domain towards and away from the transceiver. The input optical signal enters the ROSA, whose main elements consist of a photodiode, optical interface, and electrical interface with more internal components. The ROSA implements the receiver stage according to the used modulation format, whether IMDD or coherent-based. The signal is then digitized via an Analog-to-Digital Converter (ADC), which is characterized by a sampling frequency, an analog bandwidth, and a physical bit resolution. While at the receiver side an optical sequence is converted into a digital sequence, the opposite happens at the transmitter side, where a digital data stream has to be prepared for optical transmission. A Digital-to-Analog Converter (DAC) shares the same feature as the aforementioned ADC, and the resulting analog sequence drives the modulation of a laser source in the TOSA. Again, the specific implementations depend on the adopted formats, with the coherent paradigm requiring much more hardware complexity. The interface between the TX/RX and the external server/rack is represented by the on-board ASIC, the main actor of electronic computations within the transceiver. It manages the incoming and outgoing data stream between the optical cables and the external rack, fulfilling DSP and data reception/transmission management. On the receiver side, it operates demodulation, deserialization, clock recovery, and possibly FEC. On the transmitter side, it drives the modulation and serialization of the output stream, as well as pre-distortion [19].

Depending on their application field, optical transceivers severely differ in terms of data rate, packaging form factor, transmission reach, operational wavelength, modulation format, and implemented hardware. A quick overview of mature technologies is provided in Table 2.2: discussing all the aspects is out of the scope of this work, but

Optical transceivers specifications for Core applications						
Data rate	Form Type	Transmission Distance	Wavelength	Modulation Format	Cost (€)	Power Cons. (W)
200 Gbps	CFP2	80 km	DWDM (1550 nm)	DP-QAM8	8k	23
400 Gbps	QSFP-DD	80-120 km	DWDM (1550 nm)	DP-QAM16	7.5k	18.5
Optical transceivers specifications for Metro applications						
Data rate	Form Type	Transmission Distance	Wavelength	Modulation Format	Cost (€)	Power Cons. (W)
25 Gbps	SFP28	20-40 km	1310 nm	NRZ	350	2.2
100 Gbps	QSFP28 CFP4	40-80 km	DWDM (1310 nm)	NRZ/PAM4	2k-3k	4-6
Optical transceivers specifications for Access and Data Center applications						
Data rate	Form Type	Transmission Distance	Wavelength	Modulation Format	Cost (€)	Power Cons. (W)
10 Gbps	SFP+	300 m	850 nm	NRZ	25	1
10 Gbps	SFP+	2-10 km	1310 nm	NRZ	35	1
25 Gbps	SFP28	70-100 m	850 nm	NRZ	50	1
25 Gbps	SFP28	300 m-10 km	1310 nm	NRZ	70-100	1.2

TABLE 2.2: Comparison of optical transceiver features for different transmission scenarios. Data are taken from [18, 90].

it's worth mentioning some crucial differences that best highlight the close relation between constraints and technical solutions. Transceivers for long-reach applications in core networks implement DWDM operated in the C-band, combining optimized spectral efficiency and minimal fiber losses, as previously discussed. Coherent modulation, often combined with a Dual-Polarization scheme, is exploited for maximum data rate. Their form factor increases to host more complex coherent detectors and to dissipate the heat generated by DSP, which otherwise could lead to device failure. These solutions feature the highest complexity level among the other implementations, and their cost is indeed relevant. Devices for mid-range applications, generally used in metro or access networks, use the carrier frequency at 1310 nm (O-band), namely opting for slightly higher fiber losses but reducing the need for CD equalization. These less demanding scenarios, combined with simpler IMDD modulation, pose the transceivers under less stress, allowing for a smaller form factor. Finally, in short-reach access and intra-data center applications, the major constraint is represented by the unitary cost, which has to be reduced to facilitate mass production. This leads to the adoption of the first telecom window around 850 nm, which is well-suited for multimode fiber and offers a cost-effective implementation of the optical components.

Overall, optical transceivers play a primary role in telecommunications since they handle data traffic at every transport network level, from data centers to long-haul transmission. The development of the telecom infrastructure to meet the constantly increasing traffic demand is, then, strongly dependent on the technological developments of the transceivers. A lot of efforts and economic resources are invested in improving their specifications, with a particular concern for power efficiency [91]. Indeed, the electrical I/O and the on-board operated DSP represent the major contribution to the energy consumption of a transceiver [26], with many resources dedicated to channel equalization. The next sections introduce the main distortion sources in terms of linear (CD, polarization rotation,...) and nonlinear (SPM, XPM,...) effects occurring during propagation, which the transceivers are called to compensate.

2.2 Fiber propagation and impairment sources

The description of the pulse propagation equation in a confined medium follows the approach proposed in [20, 92]. The full derivation starting from the Maxwell equations is outside the scope of this work and can be found in the mentioned references; however, we present some of the most fundamental steps and approximations that highlight the roles played by different linear and nonlinear effects.

Consider an optical signal traveling in a medium and polarized along the x axis; it obeys the Helmholtz equation

$$\nabla^2 \tilde{E}(\mathbf{r}, \omega - \omega_0) + \varepsilon(\omega) k_0^2 \tilde{E}(\mathbf{r}, \omega - \omega_0) = 0. \quad (2.5)$$

Here, $\tilde{E}(\mathbf{r}, \omega - \omega_0)$, that is the Fourier transform of the electric field amplitude, has a dependence on the position $\mathbf{r}$ within the optical medium, the frequency ω , and the carrier frequency ω_0 . In the equation, $k_0 = \omega/c$ is the free-space wave number, and $\varepsilon(\omega)$ is the dielectric constant of the medium. Given all the specifications related to the considered optical fiber, Equation 2.5 can be solved for propagation, and the corresponding optical modes are returned. An optical mode is defined as a specific solution of the wave equation that satisfies the appropriate boundary conditions and retains its spatial distribution unchanged with propagation (in ideal conditions). Each mode is identified with its propagation constant β , which depends on the fiber parameters. SMFs are designed to allow the propagation of only one mode, namely, β can assume a single value at the operating wavelength. In general, to each β it is possible to associate an effective refractive index $\bar{n} = \beta/k_0$ with $n_2 < \bar{n} < n_1$, with n_1 and n_2 being the refractive index of the core and cladding, respectively. The parameter $\bar{n}$ tells us that the mode is propagating in the fiber as if that refractive index characterized the material. Moreover, the group velocity $v_g = d\omega/d\beta$ describes the velocity of propagation associated with the envelope of an optical pulse.

Equation 2.5 provides the solution for the propagation of each single frequency component ω in the spectrum, and it can be solved using the method of separation of

variables, namely, assuming that the solution has the form of

$$\tilde{E}(\mathbf{r}, \omega - \omega_0) = F(x, y) \tilde{A}(z, \omega - \omega_0) \exp(i\beta_0 z). \quad (2.6)$$

Here, $F(x, y)$ is the modal distribution in the fiber section, $\tilde{A}(z, \omega - \omega_0)$ is the amplitude of the specific spectral component, and β_0 is the first term in the expansion of $\beta(\omega)$. In particular, the term $\tilde{A}(z, \omega - \omega_0)$ is directly linked to the amplitude of the propagating electric field in the time domain by the inverse Fourier transform, as

$$A(z, t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \tilde{A}(z, \omega - \omega_0) \exp[-i(\omega - \omega_0)t] d\omega. \quad (2.7)$$

For $A(z, t)$, the slowly-varying approximation is assumed, namely, the envelope of a forward-traveling wave pulse varies slowly in time and space compared to its period or wavelength. This appears in the time-domain expression for the propagating electric field as

$$\mathbf{E}(\mathbf{r}, t) = \frac{1}{2} \hat{x} \{ F(x, y) A(z, t) \exp[i(\beta_0 z - \omega_0 t)] + \text{c.c.} \}. \quad (2.8)$$

The final expression for $\tilde{A}(z, \omega - \omega_0)$, from which the other presented expressions depend, is obtained by substituting Equation 2.6 into 2.5. Before converting the expression to the time domain, it is necessary to introduce explicit functional forms for $\beta(\omega)$ and $\varepsilon(\omega)$, which are responsible for the main sources of distortions happening during optical fiber propagation.

2.2.1 Chromatic Dispersion

The frequency dependence of the propagation constant β follows from the refractive index. This dependence is named Chromatic Dispersion (CD) and is described by the Sellmeier equation

$$n^2(\omega) = 1 + \sum_{j=1}^M \frac{B_j \omega_j^2}{\omega_j^2 - \omega^2}. \quad (2.9)$$

Parameters B_j are characteristic for each material and are linked to the resonance frequencies ω_j at which the material absorbs electromagnetic radiation. CD induces the so-called Group Velocity Dispersion (GVD), namely, the frequency dependence of the group velocity in a medium.

For optical pulses for which $\omega_0 \gg |\omega - \omega_0|$, it is possible to expand the expression in a Taylor series around the carrier frequency as

$$\beta(\omega) = \bar{n} \frac{\omega}{c} \approx \beta_0 + \beta_1(\omega - \omega_0) + \frac{\beta_2}{2}(\omega - \omega_0)^2 + \frac{\beta_3}{6}(\omega - \omega_0)^3 + o[(\omega - \omega_0)^3], \quad (2.10)$$

where $\beta_m = (d^m \beta / d\omega^m)|_{\omega=\omega_0}$. In the expansion, the first-order term is linked to the group velocity as $\beta_1 = 1/v_g$, while the second-order term, also called GVD parameter, expresses the variation of the group velocity for different frequency components.

Limiting our analysis to a propagation in a SMF with length L , the temporal broadening ΔT in an optical pulse can be estimated as

$$\Delta T = \frac{dT}{d\omega} \Delta\omega = \frac{d}{d\omega} \left(\frac{L}{v_g} \right) \Delta\omega = L \frac{d^2\beta}{d\omega^2} \Delta\omega = L\beta_2 \Delta\omega \quad (2.11)$$

where $\Delta\omega$ is the spectral width of the pulse. The last relation can also be expressed as a function of the wavelength, namely

$$\Delta T = \frac{d}{d\lambda} \left(\frac{L}{v_g} \right) \Delta\lambda = DL\Delta\lambda \quad (2.12)$$

where the relation $\Delta\omega = -(2\pi c/\lambda^2)\Delta\lambda$ has been used. The so-called Dispersion parameter D has been introduced as

$$D = \frac{d}{d\lambda} \left(\frac{1}{v_g} \right) = -\frac{2\pi c}{\lambda^2} \frac{d}{d\omega} \left(\frac{1}{v_g} \right) = -\frac{2\pi c}{\lambda^2} \beta_2 \quad (2.13)$$

and allows for linking the time pulse broadening with the spectral width expressed in terms of wavelength. Finally, introducing the group index as

$$n_g = \bar{n} + \omega \frac{d\bar{n}}{d\omega}, \quad (2.14)$$

it is possible to explicitly state the relation between D and CD, namely

$$D = -\frac{2\pi c}{\lambda^2} \frac{d}{d\omega} \left(\frac{n_g}{c} \right) = -\frac{2\pi}{\lambda^2} \left(2 \frac{d\bar{n}}{d\omega} + \omega \frac{d^2\bar{n}}{d\omega^2} \right). \quad (2.15)$$

The temporal broadening induced by CD is visualized in Figure 2.9(a), where the same 100 ps pulse is visualized at different stages of propagation, corresponding to different link lengths. Despite an evident widening, the approximately Gaussian shape is preserved along the propagation, allowing for distinguishing the presence and absence of light, and thus recognizing the arrival of the information at the receiver. However, when a train of pulses is propagating, as in Figure 2.9(b), each of them is widening in time and thus affecting the neighboring ones. Optical contributions from adjacent symbols combine, creating the so-called Inter-Symbol Interference (ISI) and consequently the distortions on the optical signal. A measure of the impact of CD is provided by $\Delta T/T_0 = BL\beta_2\Delta\omega$ [20], which expresses the ratio between the temporal broadening ΔT and the initial pulse width $T_0 = 1/B$, with B being the baud rate. As B and $\Delta\omega$ increase, shorter propagation distances are sufficient to raise the ratio, which estimates the amount of induced ISI. This highlights the role of CD as a major impairment source, which becomes more and more relevant for large bandwidth signals, even in short-reach applications. It is worth mentioning that, when applying Equation 2.11 to n -level PAM signals, the spectral width term can be estimated in a conservative way as $\Delta\omega = 2\pi B$.

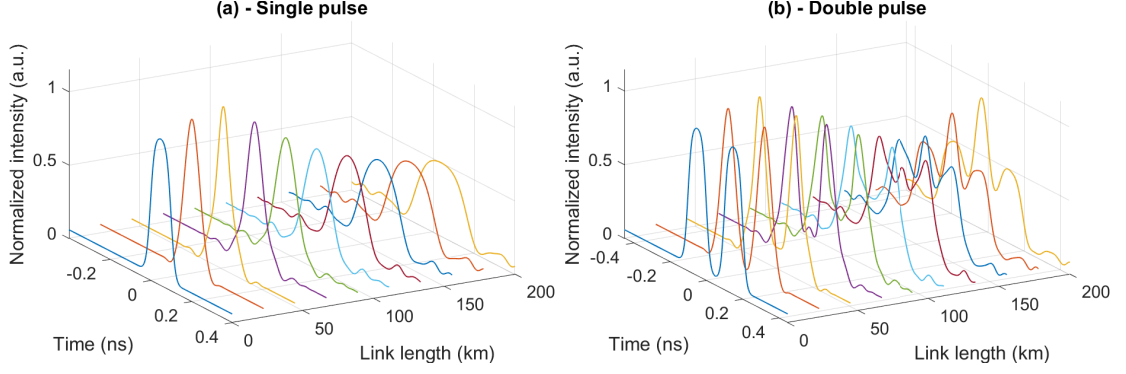


FIGURE 2.9: Intensity profile of a single (a) and double (b) 100 ps Gaussian pulse after propagation in lossless standard SMF links of different lengths.

However, it is standard to adopt the rule of thumb given by $\Delta\omega = 2\pi(B \times 0.7)$, which reflects that the majority of the signal power is concentrated within the main lobe of the spectrum, making the effective bandwidth smaller than the total baud rate [19].

The effects deriving from CD also appear in the frequency response of an optical fiber. Substituting Equation 2.6 into 2.5 and introducing the hypothesis of a lossless fiber and linear propagation regime, the transfer function describing the propagation inside a link of length L results [20, 93]

$$H_{of}(\omega) = e^{j\beta(\omega)L}. \quad (2.16)$$

With the expansion of $\beta(\omega)$ around ω_0 up to the second order, Equation 2.16 can be approximated as

$$H_{of}(\omega) = e^{j(\beta_0 + \beta_1\Omega + \frac{1}{2}\beta_2\Omega^2)L}. \quad (2.17)$$

where $\Omega = \omega - \omega_0$. Here, β_0 and β_1 can be removed from the expression as they affect the signal with a constant phase shift and a propagation delay, respectively; therefore, in a reference system moving with the propagating pulse, we have [20]

$$H_{of}(\Omega) \approx e^{j\frac{1}{2}\beta_2L\Omega^2}. \quad (2.18)$$

Equation 2.18 shows that CD affects the propagation by applying a phase mask to the different frequency components of the signal. The associated impulse response is proportional to

$$h_{of}(t) \propto e^{-j\frac{1}{2\beta_2L}t^2}, \quad (2.19)$$

thus, corresponding to a time-dependent phase mask.

Clearly, the effects of CD occurring in the optical domain have implications in the transmission line's electrical response measured at the receiver. Indeed, in IMDD links, CD is known to cause the fading of specific frequencies of the transmitted signal [94]. Let $s_{in}(t)$ be the amplitude of an input optical signal to a fiber of length L . According to

Equations 2.7 and 2.18, the output signal amplitude $s_{out}(t)$ can be written as

$$s_{out}(t) = \int d\Omega S_{in}(\Omega) e^{j\frac{1}{2}\beta_2 L \Omega^2} e^{-j\Omega t}, \quad (2.20)$$

where $S_{in}(\Omega)$ is the Fourier transform of $s_{in}(t)$. Assuming that $s_{in}(t) = a + b \cos(\bar{\omega}t)$, with a and b real numbers, the optical output transforms in

$$s_{out}(t) = a + b e^{j\frac{1}{2}\beta_2 L \bar{\omega}^2} \cos(\bar{\omega}t). \quad (2.21)$$

The signal detected at the input is proportional to

$$|s_{in}(t)|^2 = a^2 + 2ab \cos(\bar{\omega}t) + b^2 \cos^2(\bar{\omega}t), \quad (2.22)$$

while the signal detected at the output is proportional to

$$|s_{out}(t)|^2 = a^2 + 2ab \cos(\beta_2 L \bar{\omega}^2 / 2) \cos(\bar{\omega}t) + b^2 \cos^2(\bar{\omega}t). \quad (2.23)$$

The CD-induced penalty of the transmission line at $\bar{\omega}$ is defined as the ratio between the frequency component at $\bar{\omega}$ of $|s_{in}(t)|^2$ and that of $|s_{out}(t)|^2$ [93],

$$P(\bar{\omega}) = 20 \log_{10} \left| \frac{\mathcal{F}\{|s_{in}(t)|^2\}(\bar{\omega})}{\mathcal{F}\{|s_{out}(t)|^2\}(\bar{\omega})} \right| = -20 \log_{10} |\cos(\beta_2 L \bar{\omega}^2 / 2)|, \quad (2.24)$$

where $\mathcal{F}$ denotes the Fourier transform. The same result is found in [94], where a sinusoidal intensity profile for the input signal is considered. The formula predicts the presence of CD-induced notches in the electrical response, with the first located at

$$\omega_m = \sqrt{\pi / |\beta_2| L}. \quad (2.25)$$

As long as this value is larger than the bandwidth of the signal, the effects of CD are minimal, but when this condition is not satisfied, CD severely impairs the transmission.

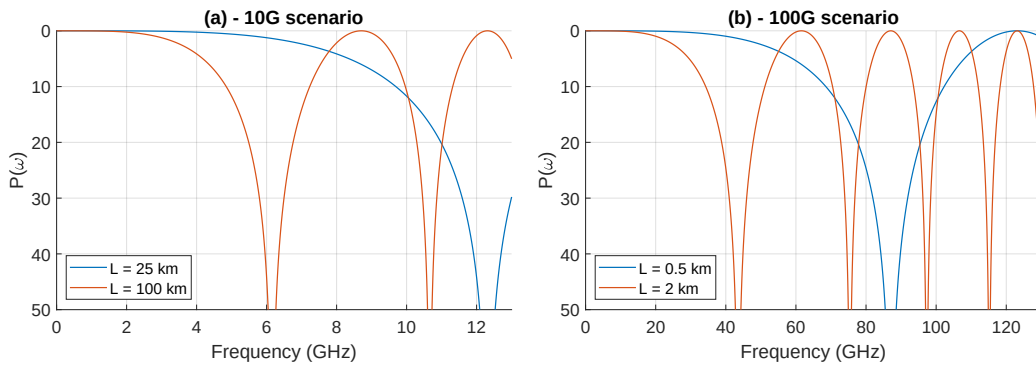


FIGURE 2.10: CD-induced power penalty behavior predicted by Equation 2.24 for $\beta_2 = -21 \text{ ps}^2/\text{km}$. Results for different propagation distances are presented in the two panels to underline the impact of CD on 10G (a) and 100G (b) signals.

This concept is illustrated in Figure 2.10, which presents the $P(\bar{\omega})$ predicted in Equation 2.24 obtained for $\beta_2 = -21 \text{ ps}^2/\text{km}$ typical for a standard SMF-28. Both panels show that the first notch of the power penalty moves towards lower frequencies as the fiber length increases. In particular, for $L = 25 \text{ km}$ (a), the first notch still occurs at sufficiently large frequencies not to alter the spectrum of a 10G signal. On the other hand, for $L = 100 \text{ km}$, the notch moves into the signal bandwidth, thus causing distortions. In a 100G scenario, much shorter propagation distances ($< 2 \text{ km}$) are sufficient to trigger this condition, highlighting that at high modulation frequencies CD drastically impacts the transmission quality already at short reach applications.

2.2.2 Self-Phase Modulation

In the derivation of the equations for CD, we omitted the dependence of the refractive index on the optical power P traveling in the fiber. This dependence originates from the optical Kerr effect [95] in which electrons in the propagating medium have an anharmonic response to intense optical fields, which translates into a nonlinear susceptibility. Therefore, a more complete treatment must consider that

$$n' = n + n_2 \left(\frac{P}{A_{eff}} \right) \quad (2.26)$$

where n_2 is the nonlinear index coefficient and A_{eff} is the effective mode area defined as

$$A_{eff} = \frac{\left(\int_{-\infty}^{\infty} |F(x,y)|^2 dx dy \right)^2}{\int_{-\infty}^{\infty} |F(x,y)|^4 dx dy}, \quad (2.27)$$

which describes the level of light confinement within the core. Via a perturbative treatment of Equation 2.5 at the first order, it is possible to include the dependence introduced by n_2 , which affects the propagation constant as

$$\beta' = \beta + k_0 n_2 \frac{P}{A_{eff}} = \beta + \gamma P \quad (2.28)$$

Here, we define γ as the nonlinear coefficient. Considering a constant input power P_{in} in a fiber of length L with loss coefficient α , Equation 2.6 tells us that the optical signal accumulates a nonlinear phase shift described as

$$\phi_{NL} = \int_0^L (\beta' - \beta) dz = \int_0^L \gamma P_{in} \exp(-\alpha z) dz = \gamma P_{in} \frac{1 - \exp(-\alpha L)}{\alpha} = \gamma P_{in} L_{eff}. \quad (2.29)$$

where L_{eff} is defined as the effective fiber length, and $P(z) = P_{in} \exp(-\alpha z)$ accounts for the fiber losses. The nonlinear phase shift accumulated by a propagating pulse depends on its associated launch power. Therefore, in a typical propagation of a train of pulses, namely where P_{in} assumes a time dependence, each pulse will accumulate a self-induced phase shift, which gives the phenomenon the name of Self-Phase Modulation (SPM). The induced frequency chirp, i.e., an instantaneous change in the signal

frequency $\delta\omega = -d\phi/dt$, causes a spectral broadening, which contributes to additional distortions due to its interplay with CD. These effects become more pronounced as the modulation frequency of the transmitted signal increases. Indeed, higher bit rates generate faster transients in the transmitted optical pulses, thus increasing the induced chirp and the related spectral broadening.

For a given fiber link, the SPM can be limited by keeping the accumulated nonlinear phase shift below a specified value ϕ_{NL}^{max} . According to Equation 2.29, this can be achieved by limiting the input power in fiber below a certain threshold P_{SPM} , which is estimated as

$$P_{SPM} = \frac{\phi_{NL,max}}{\gamma L_{eff} N_A} \quad (2.30)$$

with N_A accounting for the number of in-line optical amplifiers restoring the optical power to its original value. For typical values in a SMF28 fiber of $\gamma = 2 \text{ W}^{-1}/\text{km}$, $\alpha = 0.2 \text{ dB/km}$, $L = 50 \text{ km}$, $N_A = 10$, and $\phi_{NL,max} = 0.1$, it results $P_{SPM} = 0.26 \text{ mW}$.

From the combination of Equation 2.6 into 2.5 and after performing the inverse Fourier transform, the final form of the propagation equation in the time domain reduces to

$$i \frac{\partial A}{\partial z} + \frac{i\alpha}{2} A - \frac{\beta_2}{2} \frac{\partial^2 A}{\partial T^2} + \gamma |A|^2 A = 0. \quad (2.31)$$

In particular, this is written in the propagating pulse-based reference system, giving the possibility to drop the β_1 term. CD enters the equation via the β_2 parameter, SPM via the γ parameter, and the fiber losses via the α parameter.

2.2.3 Cross-Phase Modulation

When multiple wavelengths are co-propagating in the same fiber, as in a typical WDM scenario, the nonlinear phase shift induced on each channel is influenced by the power of the others. This takes the name of Cross-Phase Modulation (XPM). In a more complete treatment, Equation 2.29 becomes

$$\phi_{NL,m} = \gamma L_{eff} \left(P_m + 2 \sum_{n \neq m} P_n \right), \quad (2.32)$$

which now expresses the nonlinear phase shift accumulated by the m th channel. The factor of 2 arises from the form of the nonlinear susceptibility implicitly present in Equation 2.5 via the dielectric constant ϵ , and implies that for the same intensity, the contribution of the co-propagating wavelengths is twice as effective with respect to the self-induced amount [20]. In the case of modulated optical sequences, the phase accumulated in a specific propagation distance by a single pulse of the m th channel strictly depends on the modulation patterns of the other channels. Moreover, the crossed contributions are imprinted only when the co-propagating pulses overlap, which is a short-lasting condition since different wavelengths propagate at different speeds. In

this sense, the closer the channel spacing, the longer the overlap lasts, generating an effective nonlinear phase imprinting.

To include the XPM effect on the pulse propagation, additional terms have to be included in Equation 2.31 for the m th channel. Considering two propagating envelopes, it results in

$$\frac{\partial A_m}{\partial z} + \beta_{1,j} \frac{\partial A_j}{\partial t} + \frac{i\beta_{2,m}}{2} \frac{\partial^2 A_m}{\partial t^2} + \frac{\alpha_m}{2} A_m = \frac{in_2\omega_m}{c} (f_{mm}|A_m|^2 + 2f_{mn}|A_n|^2) A_m \quad (2.33)$$

where f_{mn} is the overlap integral defined as

$$f_{mn} = \frac{\int \int_{-\infty}^{\infty} |F_m(x,y)|^2 |F_n(x,y)|^2 dx dy}{\left(\int \int_{-\infty}^{\infty} |F_m(x,y)|^2 dx dy \right) \left(\int \int_{-\infty}^{\infty} |F_n(x,y)|^2 dx dy \right)}. \quad (2.34)$$

and is linked to the effective mode area by $A_{eff} = 1/f_{mm}$. Notice that in Equation 2.33, the term $\beta_{1,j}$ must be considered since it accounts for the temporal walk-off between the signals that arises due to the difference in group velocities. This limits the duration of the temporal overlap between the pulses and consequently diminishes the magnitude of cross-phase modulation (XPM) during propagation.

Equation 2.33 describes the temporal evolution of the envelope of two co-propagating optical signals, including the effect of CD, SPM, and XPM. Other effects, such as Polarization-Mode Dispersion, Modal Dispersion, Stimulated Brillouin Scattering, and Stimulated Raman Scattering, are well recognized as further sources of distortions in fiber optic transmission, but they are not accounted for in the equation. Their treatment is out of the scope of this work, which will be limited to the analysis of the aforementioned phenomena. The interested reader is referred to [20].

2.3 Equalization techniques overview

Actual data transfer is made possible by the compensation of the distortions induced during optical propagation. For this, various equalization techniques, both software- and hardware-based, are deployed within the optical networks to compensate for linear and nonlinear effects that accumulate in fibers. Ideally, equalization devices should operate with minimal latency, minimal power consumption, reduced implementation and operational costs, featuring, in addition, adaptive properties to adhere to the needs of each specific optical link. The following subsections provide an overview of the main solutions currently implemented, which are far from the ideality condition and generally result in a trade-off between the mentioned aspects.

2.3.1 Dispersion Compensating Fibers

One of the most used techniques for CD equalization is represented by DCFs [21]. Discovered in the 1990s, they are still widely deployed in telecom lines, especially in

long-haul and submarine links [96]. In a typical configuration, they are placed in series with a standard optical fiber, effectively being part of the connection path between two points. Considering the approximated fiber transfer function previously introduced in Equation 2.18, DCFs are characterized by a GVD parameter $\beta_{2,DCF}$ which allows the total transfer function to satisfy the condition

$$H_{of}(\Omega)H_{DCF}(\Omega) = \exp\left(j\frac{1}{2}\beta_2 L \Omega^2\right) \exp\left(j\frac{1}{2}\beta_{2,DCF} L_{DCF} \Omega^2\right) = 1. \quad (2.35)$$

The GVD accumulated in the propagation over a fiber span of length L gets fully equalized by a DCF of length L_{DCF} such that $\beta_2 L + \beta_{2,DCF} L_{DCF} = 0$, namely the total GVD across the whole fiber link averages to zero. As an example, for a typical $L = 80$ km across a standard SMF with $\beta_2 = -21$ ps²/km, another 20 km of DCF with $\beta_{2,DCF} = 80$ ps²/km are necessary to restore the original pulse shape [97].

DCFs are produced with a precise control of the core-cladding structure and the use of dopants that generate the refractive index difference between the two [96]. The special manufacturing techniques allow typical negative dispersion coefficients of $D \sim -100$ ps/nm/km, but have higher production costs compared to standard fibers. DCFs are also affected by higher propagation losses, which can be compensated by increasing the gain in EDFAs at the price of a higher amplified spontaneous emission contribution. Again, the production process creates small effective mode areas around ~ 20 μ m, drastically enhancing the presence of nonlinear effects compared to standard fibers. From an implementation perspective, their installation introduces further latency in the data transmission, and their non-compact nature makes them difficult to substitute or update. Despite their effective broadband equalization properties, most importantly, they represent a non-tunable solution poorly applicable to metro scenarios involving continuous signal re-routing [97].

2.3.2 Fiber Bragg Gratings

An alternative to DCFs in CD equalization is based on optical filters, among which we find Fiber Bragg Gratings (FBGs) [97, 98]. These are obtained by controlling the refractive index profile $n(z)$ along an optical fiber to create wavelength-dependent reflection patterns. In uniform-period gratings, the refractive index varies periodically as in $n(z) = \bar{n} + \Delta n \cos(2\pi z/\Lambda)$, where $\bar{n}$ is the average over the fiber, Δn is the modulation depth, and Λ is the grating period. Overall, such a device reflects specific wavelengths, creating a stop-band where the central wavelength (or Bragg wavelength) λ_B is directly linked to the grating period by $\lambda_B = 2\bar{n}\Lambda$. In case of a uniform period grating, an analytical form of the frequency response is found by solving the coupled-mode equations

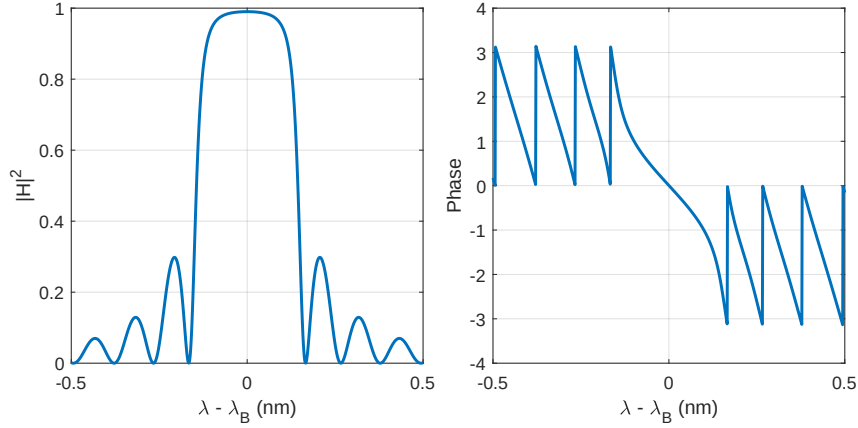


FIGURE 2.11: Magnitude (a) and phase (b) of the frequency response for reflected waves in a uniform-period FBG. The simulation is performed using $\lambda_B = 1550$ nm, $L_g = 10$ cm, and $\kappa L_g = 3$.

describing the propagation along the x axis in an FBG of length L_g [97]:

$$\begin{aligned}\frac{dA_f}{dx} &= i\delta A_f + i\kappa A_b \\ \frac{dA_b}{dx} &= -i\delta A_b - i\kappa A_f \\ \delta &= \frac{2\pi}{\lambda} - \frac{2\pi}{\lambda_B} \\ \kappa &= \frac{\pi\Delta n\Gamma}{\lambda_B}\end{aligned}$$

Here, A_f and A_b are the amplitudes of the forward and backward propagating waves, δ is the detuning from the Bragg wavelength, κ is the coupling coefficient between the two waves, and Γ is the confinement factor, which measures the fraction of optical power contained in the core. The FBG transfer function for the reflected waves results in [97]

$$H(\omega) = \frac{A_b(x=0)}{A_f(x=0)} = \frac{i\kappa \sin(qL_g)}{q \cos(qL_g) - i\delta \sin(qL_g)}, \quad \text{with } q^2 = \delta^2 - \kappa^2 \quad (2.36)$$

and appears graphically in Figure 2.11(a). The phase imposed on the counter-propagating waves is linear with the frequency within the stop band, as appears in Figure 2.11(b): this makes the design poorly functional for CD equalization, which, indeed, according to Equation 2.18, requires a quadratic phase profile with the frequency. For a uniform-period FBG, this behavior appears at the edges of the stop band, with the drawback of increased insertion losses.

A more effective solution is represented by Chirped FBGs [22, 99], which feature a grating period that varies along the fiber and sets a precise reflection point for each

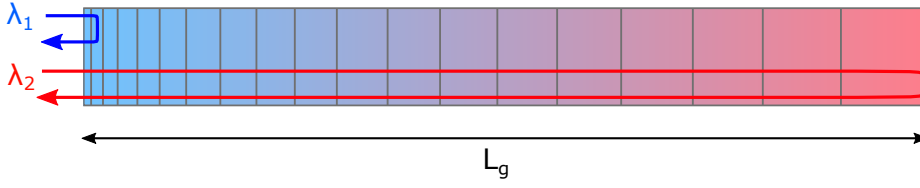


FIGURE 2.12: Inner structure of a Chirped FBG producing anomalous dispersion. The wavelength λ_1 meets the Bragg condition immediately after entering the device from the left port. As the grating period increases, longer wavelengths are reflected deeper in the FBG, with λ_2 meeting the Bragg condition at the extreme opposite to the input port, thus accumulating the maximum delay with λ_1 .

wavelength within the medium. Each segment of a chirped FBG locally approximates a uniform-period device; thus, the overall stop band results from the combination of narrow stop bands provided by each segment. The optical path varies for each wavelength depending on where the Bragg condition is satisfied within the device; each wavelength accumulates a different delay before the recombination, which is used to null the contribution imposed by CD. In a chirped FBG as that presented in Figure 2.12, the component λ_1 entering the device from the left meets the Bragg condition immediately, thus spending virtually no time $\Delta T_1 = 0$ within the grating. Since the grating period increases in the left-right direction, longer wavelengths will be reflected further in the device, with λ_2 accumulating the largest delay $\Delta T_2 = 2L_g\bar{n}/c$ corresponding to a double propagation along the whole grating length L_g . The total amount of dispersion provided by the device is estimated as the delay accumulated per unit length per unit wavelength, namely

$$D_g = \frac{\Delta T_2 - \Delta T_1}{(\lambda_2 - \lambda_1)L_g} = \frac{2\bar{n}}{\Delta\lambda \cdot c} \quad (2.37)$$

with $\Delta\lambda = \lambda_2 - \lambda_1$. The sign of $\Delta\lambda$ matches the sign of dispersion applied by the device, which, in the case of an increasing refractive index profile, as that presented in Figure 2.12, results in an anomalous dispersion. The same device can provide normal propagation by simply flipping the refractive index profile.

Compared to DCFs, chirped FBGs features many advantages provided by compactness, low costs, and ease in integration into existing systems [96]. On the other side, limitations are given by a relatively narrow bandwidth, a limited temperature operating range, and the presence of ripples in the amplitude and phase response [96, 100].

2.3.3 Digital Signal Processing

The counterpart to full-optical techniques for channel equalization is represented by DSP. It consists of an electronic-based approach that makes use of ASICs to treat the digitized version of the received optical signal for error compensation and, in general,

to monitor the performance of the transmission line. This approach has been largely enabled by the development of ADCs based on a CMOS technology [101]: this provides low cost per unit in mass production and large bandwidth, compatible with the modulation frequencies in use in the optical domain. The early 2000s represented the advent of DSP for telecom applications with the first 10 Gbps implementations on IMDD systems [102], until reaching the current research state-of-the-art [103], enabling data processing up to 1 Tbps and beyond. A major concern in DSP is associated with the high power consumption and implementation costs [104], which, under the present conditions, prevent a fruitful implementation for short-reach transport networks. DSP has also represented a key-enabling technology for the development of coherent systems, where nowadays it finds the majority of its applications. In a typical coherent transceiver implementation, the optical and electronic domains work in synergy to provide the maximum bandwidth efficiency, which is especially needed in long-haul transmission [105].

DSP provides many functionalities which are, definitely, not limited to channel equalization [108]. An overview of the operations enabled by DSP is presented in Figure 2.13 and Table 2.3, which describe, respectively, a typical DSP work-chain and the specific functionalities implemented within an optical transceiver [24, 109]. The algorithms are broadly distinguished into two classes according to the description provided in [108]. A first class of algorithms focuses on synchronization problems, dealing with timing errors, frequency, and phase impairments that arise between the transmitter and the receiver. A second class of algorithms operates channel equalization at different levels. The associated DSP counteracts a wide variety of phenomena, such as imperfections in the detection process, ISI induced by CD, Polarization Mode Dispersion, Polarization rotation, and nonlinear effects, thus, in general, counteracting distortions induced in the optical signals. A detailed dive into all of the algorithms is out of the scope of this work, but it is worth providing an overview of those of interest for this manuscript. These regard the implementations for CD equalization, nonlinear equalization, and Forward Error Correction, whose operational principles are treated in the next paragraphs.

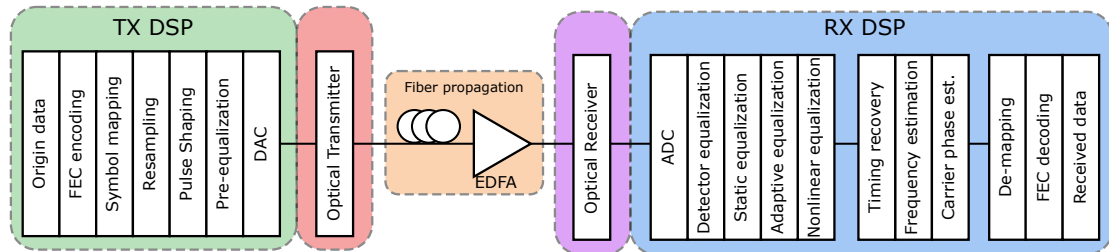


FIGURE 2.13: Block diagram of a typical DSP implementation in optical transceivers.

Synchronization Algorithms		
Group	Functionality	Main Algorithms
Timing Recovery	Ensure matching of the TX and RX sample clocks	Gardner Algorithm, Interpolation
Frequency estimation algorithms	Monitor and adjust possible detunings between the frequencies of the carrier (where information is encoded) and the Local Oscillator	Freq. offset estimation [106], phase increment estimation [107]
Carrier phase estimation	Track the mean phase evolution of TX laser and local oscillator to reduce the impact of phase noise	Viterbi and Viterbi
Equalization Algorithms		
Group	Functionality	Main Algorithms
Orthonormalization and de-skew	Correct for imperfections in coherent detection (non-idealities in 90° hybrids, timing and amplitude differences between paths,...)	Gram–Schmidt algorithm, Finite Impulse Response filter
Static channel equalization	Pulse shaping, CD compensation	Finite Impulse Response filters
Adaptive channel equalization	Track and counteract dynamic time-varying phenomena (polarization rotation, Polarization Mode Dispersion,...)	Constant Modulus Algorithm
Nonlinear channel equalization	Correct for fiber nonlinearities affecting signals during propagation	Digital Backpropagation, Volterra, ANNs

TABLE 2.3: Overview of DSP algorithms for optical telecommunications. The categorization follows the one proposed in [108].

In current technology, up to 50% of the total transceiver power consumption is absorbed by DSP [110], with the major resources allocated for channel equalization indeed. In a 100G coherent transmission scenario up to 2400 km, CD and FEC are estimated to account for 18% and 10% of the total consumption, respectively [111]. Finding alternative, less power-hungry ways to alleviate these functions is, thus, of primary importance to reduce the environmental and economic impact of future telecommunications.

2.3.3.1 FIR filters for Chromatic Dispersion equalization

The modeling of CD in pulse propagation treated in Section 2.2.1 results in the transfer function $H_{of}(\omega)$ of Equation 2.18 and the associated impulse response h_{of} of

Equation 2.19. An algorithm that operates CD equalization has, therefore, to approximate $H_{of}^{-1}(\omega)$, which corresponds to an impulse response in the form

$$h_{cd} \propto \exp\left(j \frac{t^2}{2\beta_2 z}\right) \quad (2.38)$$

where z is the propagation distance, and β_2 is the GVD parameter. Ideally, the equalized output sequence $y(t)$ will result from the convolution between the distorted sequence $x(t)$ and $h_{cd}(t)$, which, however, cannot be implemented in its continuous form. Indeed, after the detection, the distorted sequence undergoes a finite sampling occurring every T ; the spectrum of the digitized sequence results, thus, limited between $\pm\omega_N$, with this being the Nyquist frequency obtained as $\omega_N = \pi/T$. It is, therefore, necessary to implement a discrete convolution in the form [112]

$$y[n] = (h_{eq}[n] * x[n]) = \sum_{m=0}^{\bar{m}} h_{eq}[m]x[n-m] \quad (2.39)$$

where the time boundary $\bar{m}$ is set to prevent aliasing in the filter design. A straightforward implementation of Equation 2.39 is provided by a Finite Impulse Response (FIR) filter [108]. As the name suggests, it implements a filtering operation whose impulse response $h_{FIR}(t)$ is limited in time. For a generic FIR filter acting on a discrete time sequence, this condition can be described by a weighted sum of Dirac's Deltas as

$$h_{FIR}(t) = \sum_{n=0}^{N_T-1} h_{FIR}[n]\delta(t-nT) \quad (2.40)$$

where N_T is the number of taps, which determines how long the response lasts in time, thus the memory of the filter. In the context of CD equalization, the weights appearing in the FIR impulse response become [108]

$$h_{eq}[n] = \frac{1}{\sqrt{\rho}} \exp\left[\frac{j\pi}{\rho} \left(n - \frac{N_T-1}{2}\right)^2\right] \quad (2.41)$$

with

$$\begin{aligned} \rho &= \frac{2\pi\beta_2 z}{T^2}, \\ N_T &= \lfloor |\rho| \rfloor, \\ n &= 0, 1, \dots, N_T - 1, \end{aligned}$$

and $\lfloor \cdot \rfloor$ representing the integer part of the argument.

FIR filters represent a versatile approach to CD equalization, with an adjustable number of taps and associated weights depending on the transmission scenarios. The implementation described by Equations 2.39 and 2.41 represents the starting point in

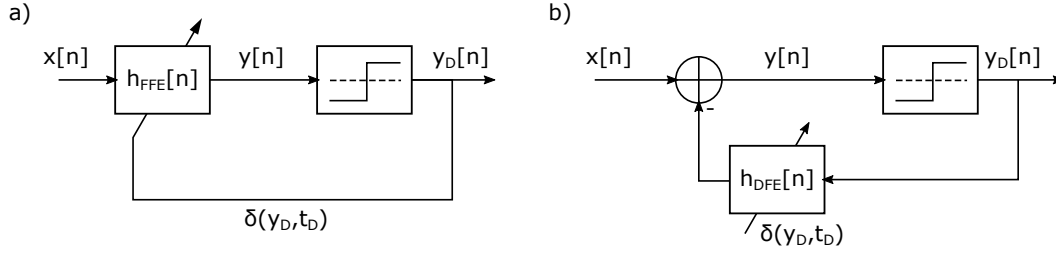


FIGURE 2.14: Structures of a Feed-Forward Equalizer (a) and a Decision Feedback Equalizer (b).

this context, but more sophisticated solutions prove to be more effective in counteracting ISI generated by GVD. For example, FIR filters are implemented in Feed-Forward Equalizers (FFE) and Decision Feedback Equalizers (DFEs), whose schematics are depicted in Figure 2.14 [25]. A FFE represents a linear equalizer that basically implements a FIR filter with a tunable impulse response $h_{FFE}[n]$. In a general treatment, this operates directly on the distorted sequence $x[n]$ to produce $y[n]$, which then passes through a decision circuit for bit identification to obtain $y_D[n]$. Finally, this is compared to a target sequence $t_D[n]$ to determine the level of discrepancy δ . The tap weights are adjusted via specific algorithms, which act both on the pre-cursor and post-cursor taps, aiming to minimize δ . On the other hand, a DFE represents a nonlinear equalizer, in which the output of a decision circuit $y_D[n]$ represents the input to a feedback FIR filter characterized by an impulse response $h_{DFE}[n]$. Its output, in turn, is subtracted from the received signal $x[n]$, and the result $y[n]$ is provided as an input to the decision function. Compared to FFE, this solution is very effective against post-cursor ISI and requires fewer taps for the equalization. However, it is sensitive to error propagation, meaning that if a past decision is wrong, the error propagates to the subsequent iteration and corrupts future symbols. This strongly limits its applicability in the context of optical communications [25].

When adopted for CD equalization, FFEs are used in a static configuration, and can be placed both at the transmitter and at the receiver as illustrated in Figure 2.13. Despite the implementation might appear simple, computational complexity is driven by the number of taps. In turn, this is determined by the amount of ISI to be equalized and by the operative frequency (symbol-spaced or fractionally-spaced) of the filters.

2.3.3.2 DSP-based nonlinearity mitigation

Contrary to what has been presented for linear transmission, when nonlinear effects are considered in the propagation equation, it is not possible to determine analytically the impulse response to be applied to compensate for the distortions. The equalization problem becomes more complex, with further contribution given by Kerr effect distortions which can be both intra-channel (SPM) and inter-channel (XPM). Especially in this last case, their equalization requires operating on multiple channels simultaneously,

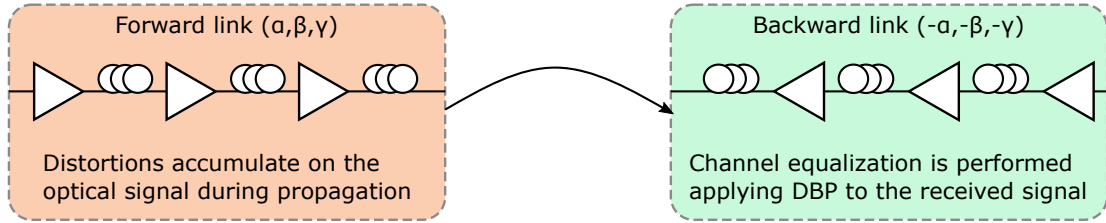


FIGURE 2.15: Digital Back-Propagation scheme for channel equalization.

which translates into more sophisticated algorithms and a greater computational effort.

A first solution is provided by Digital Back-Propagation (DBP), which implements a numerical model to replicate the optical signal propagation in the backward direction [113, 114]. This algorithm is strictly linked to the Split-Step Fourier Method (SSFM) [92], which provides an approximated solution to Equation 2.33 based on the alternation of propagation steps involving linear or nonlinear effects only. From Equation 2.33, we have that the forward propagation of an optical signal in a fiber span is characterized by a loss coefficient α , a group velocity dispersion parameter β_2 , and a nonlinear coefficient γ . Therefore, to approximate the backwards propagation the DBP uses the SSFM featuring $-\alpha$, $-\beta_2$, and $-\gamma$, as illustrated in Figure 2.15. DBP is not limited to the equalization of a specific effect, but multiple origins of distortions (CD, SPM, XPM,...) can be equalized at once, at the price of a high computational effort. Indeed, the SSFM is known for providing a good level of approximation with the drawback of reduced computational efficiency. This is further reduced in the context of WDM equalization, where multiple coupled differential equations must be solved simultaneously (adopting multi-channel DBP), or when polarization-mode dispersion is also considered. The entire DBP is based on the knowledge at the receiver of the transmission conditions encountered by the propagated signal, which might not be best suited for reconfigurable optical paths in metro networks.

An alternative approach to nonlinear equalization is based on the Volterra Series Transfer Functions [113]. While the SSFM tackles Equation 2.33 via a numerical approach, this method provides an analytical approximation to the solution. It is widely adopted in the context of telecommunications to simulate the propagation of optical pulses in the nonlinear regime efficiently. Time domain implementations of this method are possible [115], but in this section, we are focusing on its formulation in the frequency domain. Given a signal at the receiver $x(t)$, and $X(\omega)$ its Fourier transform, the Volterra Series Transfer Functions provide a functional expansion that models its response $Y(\omega)$ as a series of linear and nonlinear terms, as graphically illustrated in Figure 2.16. The response is expressed in terms of Volterra kernels $H_i(\omega)$, which characterize the system's behavior at different nonlinearity orders. In the context of optical propagation in Silica SMFs, the first-order kernel accounts for linear distortions induced by CD, while the higher-order terms describe nonlinearities and their interplay with CD, as well as the memory effects induced during the propagation. Overall, the

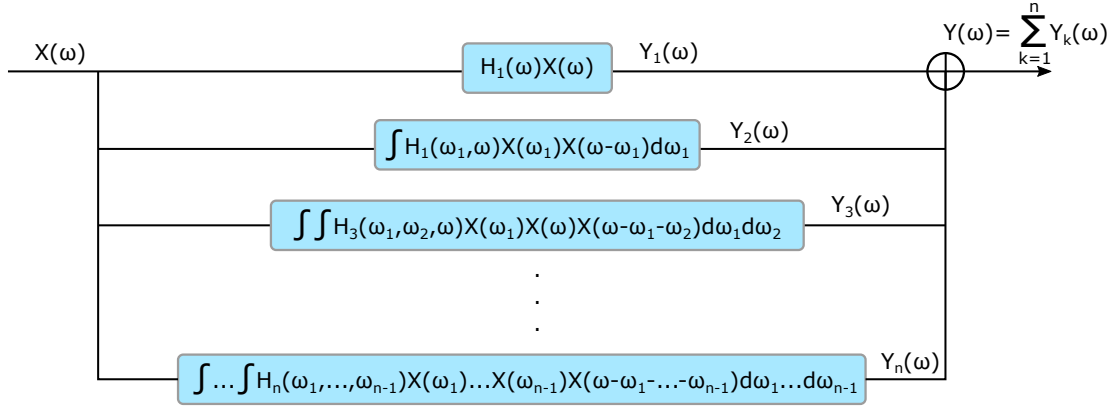


FIGURE 2.16: Block diagram of the Volterra Series Transfer Function for channel equalization.

system's output after a single span propagation can be described as

$$Y(\omega) = H_1(\omega)X(\omega) + \sum_{n=2}^{\infty} \int \cdots \int H_n(\omega_1, \omega_2, \dots, \omega_{n-1}, \omega) \times \quad (2.42)$$

$$\times X(\omega_1)X(\omega_2) \cdots X(\omega_{n-1})X(\omega - \omega_1 - \cdots - \omega_{n-1})d\omega_1 \cdots d\omega_{n-1}$$

Generally, the third-order expansion is used to get sufficiently accurate approximations in the standard propagation conditions, with the acknowledgment that, due to symmetry properties in the Silica fibers, the even-labeled terms can be discarded.

The nonlinearity equalization in the context of the Volterra Series is solved by finding the inverse kernels to those appearing in Equation 2.42. The inverse transformations are applied to the received optical signal to remove the linear and nonlinear contributions induced by the propagation and restore the original information. While DBP, which is implemented via SSFM, necessitates continuous switching between the time and frequency domains, the Volterra method can be fully implemented in the frequency domain. Overall, when dealing with short-memory systems, these factors imply an overall reduction in computational effort spent for equalization compared to DBP.

Here we presented the most intuitive DSP-based methods for nonlinear effects equalization. Many other techniques exist [113, 114], but their detailed treatment is out of the scope of this work.

2.3.3.3 Forward Error Correction

Another fundamental block in DSP techniques to reduce errors in data transmission is represented by Forward Error Correction (FEC) codes. These protect the data to be transmitted (or stored) from the channel impairments by adding redundancy [116]. Not only do they act against residual distortions after channel equalization, but also against optical noise and any source of impairments that can occur during transmission. As

FEC name	Decoding	OH (%)	NCG (dB)	pre-FEC thresh.
Algebraic LDPC	SD	5.05	3.35	4.1×10^{-3}
CFEC	SD	14.8	10.8	1.2×10^{-2}
Convolutional LDPC	SD	20	11.51	2.76×10^{-2}
Nokia SDFEC-G2	SD	25	11.7	3.3×10^{-2}
SC-LDPC	SD	25	12	3.73×10^{-2}
RS+Hamm. prod.	HD	6.7	8.5	1.9×10^{-3}
Concat. BCH	HD	6.7	9	3.1×10^{-3}
Swizzle	HD	6.7	9.4	4.84×10^{-3}
CI-BCH	HD	12	10	8.76×10^{-3}
Orth. concat. BCH	HD	24.3	10.06	1.3×10^{-2}

TABLE 2.4: Overview of different implementations of FEC codes and their features in terms of Decoding (Hard [HD] or Soft [SD] Decision), Overhead (OH), Net Coding Gain (NCG), and pre-FEC threshold. Data are taken from [116, 117].

specified in the block diagram of Figure 2.13, the FEC encoding process occurs at the transmitter side and introduces redundancy in the data stream. The redundancy is then removed at the last stage of the receiver DSP chain with the FEC decoding.

FEC codes can be broadly grouped into two main categories based on the type of input data provided to the algorithm and the decoding method used to correct for errors [116]. In Hard Decision-FEC (HD-FEC), the input data to the FEC decoder is represented by a digital sequence (generally 0s and 1s), which has been generated after a threshold-based (thus hard) decision at the receiver. The decoder estimates the original message from this hard-detected sequence, exploiting schemes usually implemented in blocks. This was the first implemented FEC method, featuring lower complexity and associated power consumption compared to other approaches [118]. In contrast, in Soft Decision-FEC (SD-FEC), decoders are fed with analog (discretized) sequences, which allows the decision process to deal with more pieces of information from the received trace. Convolutional algorithms process the received sequence and determine the final decision on the bit value based on probabilistic details, including the confidence in the possible transitions. Overall, SD-FEC provides better performance compared to HD-FEC, and is generally best suited for high-demand applications, such as long-haul transmission, where reliability is a crucial aspect [119]. The improved performance in error correction is, however, partially balanced by an increased algorithm complexity and higher computational effort [120].

Many different implementations exist of both SD and HD decoders, some of which are listed in Table 2.4 to present the state-of-the-art in the field and to compare them based on three different standard figures of merit. In particular, the Overhead (OH) denotes the relative portion of redundant bits that are added to the information bits during the encoding at the transmitter. The pre-FEC Bit Error Rate (BER), on the other hand, indicates the maximum error rate (i.e., the maximum amount of impairments in the signal) that the decoder can deal with to guarantee a final BER value below a given threshold, conventionally set to 10^{-15} . Finally, the Net Coding Gain (NCG) measures

the beneficial effects on the BER reduction provided by the FEC. It is evaluated as the ratio (in dB) between the OSNR values corresponding to a fixed BER, in general corresponding to 10^{-15} , measured before and after FEC introduction.

Chapter 3

Photonic Neural Networks

An Artificial Neural Network (ANN) or neural network is defined as a massively parallel distributed processor consisting of processing units that have a natural propensity to store experiential knowledge and make it available to use [121]. The initial concept of an ANN was introduced in 1943 by Walter Pitts [122], who developed a model grounded in mathematics and algorithms to mimic the behavior of biological brain cells. This groundbreaking idea later proved to be revolutionary, sparking significant advancements and a remarkable surge of interest in the field over the past few decades. In the very recent past, with the availability of large massively parallel computers, Artificial Intelligence (AI) has made its unbridled entrance into our society, posing a radical change in our way of approaching problems on many levels. This new paradigm is widely appreciated in many fields since it can provide rapid and personalized solutions that satisfy end-users' needs in terms of speed and precision. A clear example in this sense is provided by AI tools based on large language models [123] such as ChatGPT [124], Gemini [125], and others, which allow us to access information or produce content that other methods would have required much more time and knowledge from the user.

A profound synergy is present between ANNs and photonics, since both domains can benefit from the other. Access to the resources provided by AI models is possible thanks to the backbone optical transport network and photonic technologies implemented in data centers [13]. In this sense, optical fibers and photonics represent a key enabling technology, allowing for large bandwidth data transfer, processing speed, and parallelization. Conversely, Machine Learning found applications in data treatment at many levels, including optimization of optical routing [126], data center interconnections [127], and channel equalization [33, 34]. In this context, PNNs appears as a true game changer, bringing together the advantages of optical processing (discussed later in Section 3.3) with the versatility of the machine learning paradigm.

This chapter introduces ANNs, starting from the fundamental processing unit and exploring the possible network topologies. Also, some training algorithms are presented. Finally, Neuromorphic computing and the photonic implementations are reviewed, focusing especially on telecom applications.

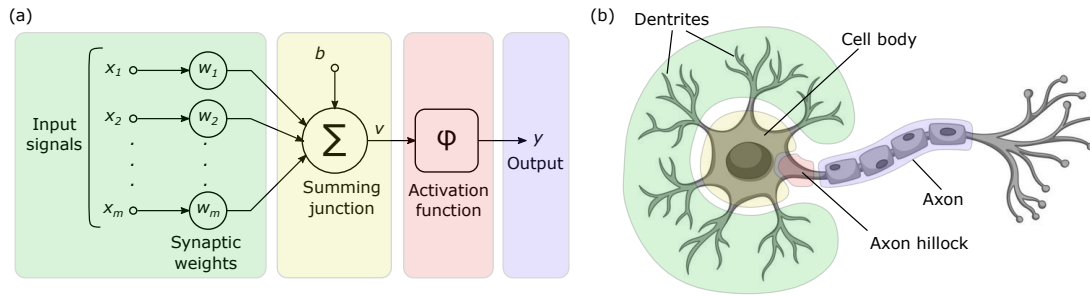


FIGURE 3.1: Comparison of an Artificial Neuron (a) and a biological neuron (b). Areas with the same color highlight artificial and biological structures providing similar functions.

3.1 Artificial Neural Networks

ANNs are inspired by the functioning of the brain. In broad terms, they are capable of adjusting their internal state to adapt to specific tasks. Just like in a biological brain, the core unit of ANNs is the neuron, which serves as the fundamental information-processing element in both systems. The structure of a neuron within an ANN is illustrated schematically in Figure 3.1(a). A neuron receives multiple inputs $x_1, x_2, \dots, x_m$, which are then multiplied by the corresponding weights w_i , and then summed together, often with the addition of a bias term b . This produces an intermediate value v , which is passed through an activation function ϕ to get the neuron's response. Mathematically, this can be expressed as:

$$y = \phi \left(\sum_{i=1}^m x_i w_i + b \right) = \phi(v) \quad (3.1)$$

Figure 3.1(b) also shows the schematic of a biological neuron, highlighting its analogies with its computational counterpart. Input channels in the artificial unit function like dendrites, collecting signals from other neurons. The inputs are applied with weights, which find their biological counterpart in the synaptic strength, corresponding to different connections' strengths between cells. The weighted signals are then brought to a summing stage, equivalent to the cell body. The result of the sum undergoes a nonlinear transformation, which on a biological level is operated by the axon hillock, and finally, the resulting signal is transmitted through an output channel, similar to an axon, which can then connect to multiple other neurons.

In ANNs, each neuron functions as a network node and can be connected to numerous others through multiple links. By combining several nodes, a wide range of network architectures can be formed, ranging from simple models like the single-layer perceptron [128] to highly complex configurations. In building the different topologies, it is possible to distinguish between different types of neurons, whose denomination depends on their position within the network. Input neurons receive data from the world

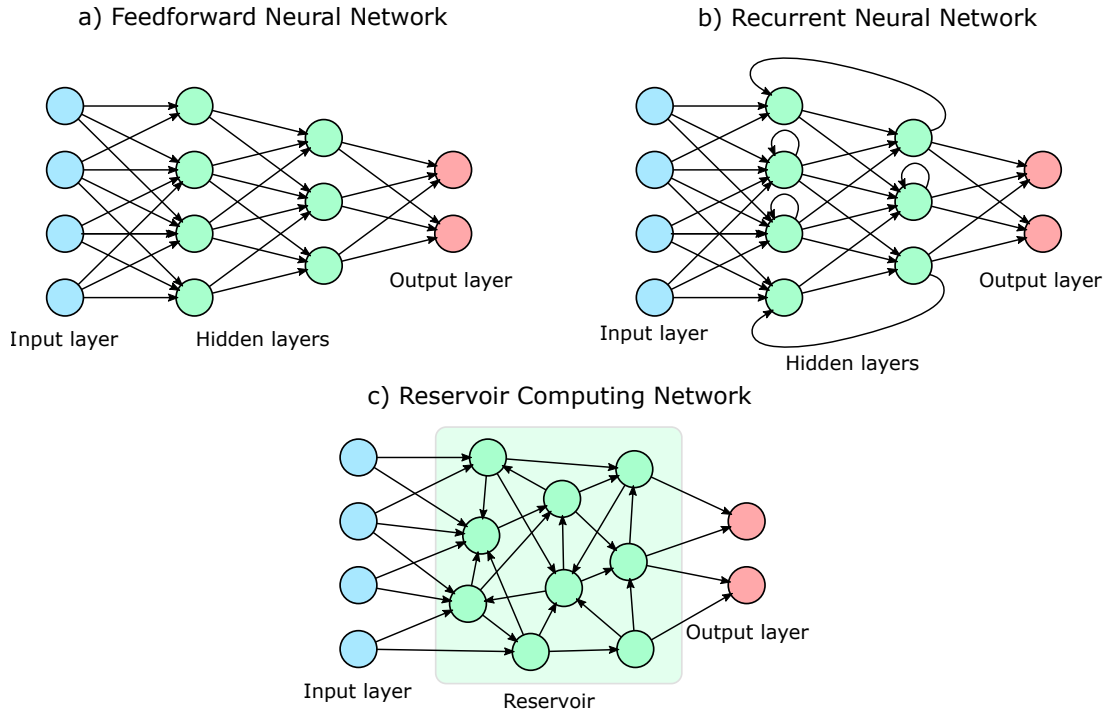


FIGURE 3.2: Schematics of different ANNs architectures: (a) Feedforward, (b) Recurrent, and (c) Reservoir Computing Networks.

external to the network and feed it into subsequent units for processing. Output neurons produce the outgoing data departing the network and represent the final stage of the processing. Finally, the hidden neurons receive and provide data from and to other neurons, thus appearing buried within the network with no external contact. A layer is constituted by a group of neurons that simultaneously receive the same type of information. In principle, neurons can be organized into virtually any topology, with the optimal structure depending on the nature of the task and the type and quantity of data being processed. Despite the vast diversity of possible designs, neural network architectures can generally be classified into two main categories [28].

In Feedforward Neural Networks, data travels in one direction only, from the input to the output layer, possibly connected by hidden layers. The connections between nodes never include feedback signals, namely, a communication is set between neurons belonging to subsequent layers, without links between units of the same layer or previous ones, as Figure 3.2(a) illustrates. The data flow in feedforward networks makes them particularly suitable for tasks that involve pattern recognition, classification, and regression. The simplest implementation of a Feedforward Neural Network is represented by the single-layer perceptron, which is equivalent to a single artificial neuron [129] and operates exactly according to Equation 3.1. After its first theorization by Warren McCulloch and Walter Pitts in 1943 [122] and the first implementation in 1957 by Frank

Rosenblatt [130], many new realizations have been proposed, including multilayered versions to deal with problems of a higher degree of complexity.

Recurrent Neural Networks represent the counterpart of feedforward implementations. In this scheme, at least a feedback loop is present, and data flow is not forced in the forward direction, but it can recirculate multiple times within the neurons, creating complex dynamics and allowing for the installation of memory in the network. In addition to forward connections, links between neurons of the same layer or even with previous ones can be created, as illustrated in Figure 3.2(b). This type of architecture is particularly suitable for time-series predictions due to the implemented Long Short-term memory features. Particularly relevant in the context of recurrent topologies is the implementation of Reservoir Computing (RC) Networks [131]. Figure 3.2(c) visualizes the working principle in which data, at a certain point during the propagation within the network, enters a fixed nonlinear system, named the Reservoir, which is made of nodes with random sparse connections. The Reservoir represents a non-trainable part of the network that maps input signals into higher-dimensional computational spaces. The output is then fed to a simple algorithm, generally a trainable linear classifier, that performs the readout and provides the final output. The reservoir is usually seen as a black box that transforms the data according to a complex dynamics. This serves the purpose of enriching the data features, making the later classification task easier for the algorithms.

3.2 Training procedures

The solution of a specific task is not determined only by choosing the proper network architecture, but also by its union with a suitable loss function and a training algorithm. In a wide sense, the loss function measures the performance level of the network in accomplishing a specific operation. On the other hand, the training is done via an algorithm that adjusts the network's weights, aiming to minimize the loss function. In supervised learning, the loss function evaluation is based on a comparison between the network's output and the associated target. In unsupervised learning, the loss function is agnostic of the target, and the training aims at unveiling hidden structures or patterns already present in the input data, without having a clue about whether the proposed answer matches the correct one [132]. Finally, a different paradigm is proposed with reinforcement learning, in which an agent learns to make decisions by performing actions and receiving feedback from an environment in the form of rewards or penalties [133]. Multiple factors drive the training algorithm choice, such as the tradeoff between convergence speed and number of configurations explored, the linear or nonlinear nature of the problem, and the discrete or continuous nature of the loss function. Of the huge variety of training algorithms that exist, only a few of them will be presented below, which are particularly relevant in the context of this work.

3.2.1 Backpropagation

Backpropagation offers an algorithmic method grounded in gradient evaluation for identifying the maxima and minima of a differentiable function [134]. This technique operates iteratively: at each step, the function's input parameters are adjusted based on the gradient computed at the current point. This gradient guides the parameters toward minimizing or maximizing the function. The foundation of gradient-based optimization was first laid by Cauchy [135], and today, it is a cornerstone of training ANNs due to its efficiency and adaptability. Over time, both the theory and practical implementations of this method have significantly evolved to address increasingly complex problems involving large-scale data. Despite these advancements, all variants retain the same core principle: at every iteration, the gradient is used to update the parameters in the direction that minimizes the loss function.

In a standard backpropagation, the inputs to the algorithm are the function $f(\vec{w})$ to be minimized, the starting point $\vec{w}_0$ for the search of the minimum, a learning rate parameter γ , and a stopping criterion. At each i -th step the weights vector $\vec{w}$ is updated according to

$$\vec{w}_{i+1} = \vec{w}_i - \gamma \nabla f(\vec{w}_i) \quad (3.2)$$

where the learning rate γ is set by the user. A too-small value for γ could lead to a slow descent, while a too-large value may induce a non-stable solution with the research point jumping back and forth around the minimum. The iterations proceed until the stopping criterion is satisfied: many possible choices for the named criterion can be adopted, the optimal one depending on the nature of the problem and on the amount of information that is required by the algorithm.

In the context of ANNs training, the function to be minimized becomes the loss function associated with the considered problem. Notice that a limiting factor for the applicability of backpropagation is that the functions under minimization must be differentiable in $\vec{w}$. Moreover, the effectiveness of gradient descent as an optimization method is highly sensitive to the choice of the initial starting point. Once this point, along with other parameters, is specified and the algorithm runs without any added randomness, its deterministic nature becomes evident, namely, the convergence outcome is essentially predetermined. However, this final point is not necessarily the global minimum of the function; it may only represent a local minimum. This tendency to stop prematurely is one of the key limitations of gradient descent. The algorithm is constrained to follow the trajectory defined by the gradient and cannot venture beyond it. Exploring a wider space of possible configurations could potentially yield better results and mitigate the risk of early convergence to suboptimal solutions.

3.2.2 Adam algorithm

Introduced in 2015 by Diederik P. Kingma and Jimmy Ba [136], the Adam algorithm was developed as an enhancement of the Stochastic Gradient Descent method.

It integrates key aspects of other optimization techniques, such as RMSprop [137] and AdaGrad [138], making it particularly well-suited for optimizing complex functions with multiple parameters, especially when gradients are sparse or highly noisy. Adam is easy to implement, computationally efficient, and exhibits invariance to diagonal rescaling of the gradients [136]. Its robustness to noise makes it especially relevant in experimental contexts, where measurement noise can significantly undermine the accuracy of traditional gradient-based methods.

As mentioned above, the Adam optimization approach can be seen as a particular expression of the Stochastic Gradient Descent (SGD) algorithm [139]. This last method can be adopted as a substitute for the standard gradient descent, being computationally efficient when the data set employed for the training is large [140]. The method relies on the same iterative procedure described in Equation 3.2, but here the gradient is substituted with an estimation evaluated on a randomly chosen fraction of the data set. In the context of ANNs training, the loss function $L(\vec{w})$ is often evaluated as

$$L(\vec{w}) = \frac{1}{N} \sum_{i=1}^N |y_i - y_{i,targ}|^2 \quad (3.3)$$

where y_i is the prediction of the network, $y_{i,targ}$ is the target value, both corresponding to the same input x_i , and N is the number of training points. This is a quite general form of the loss function since it measures the discrepancies between the predictions and the reference points for every element of the data set from $i = 1$ to N . SGD approach estimates the gradient by selecting a fraction of the N points for its evaluation: when the data set is very populated, this approach allows for maintaining a good precision while reducing the computational time.

Knowing that the Adam algorithm operates in this context of stochastic optimization, we return now to its specific features. Its name derives from *Adaptive Moment Estimation*, alluding to the fact that it uses the estimates of the first and second moments of the gradients for the generation of an adaptive step for each of the tunable parameters. The algorithm performs exponential moving averages m and v , respectively, of the gradient and the square gradient. These averages are updated at each iteration, and they enter into the generation of the adaptive step, as we will see in a moment. A distinctive characteristic of Adam is that the descent proceeds with a memory of the previous iterations, since the present step is proposed by considering the history of the descent itself. The amount of memory the algorithm has to keep track of is determined through the parameters β_1 and β_2 (both in $[0, 1)$), representing the exponential decay rate for the moving averages.

Notice that, in the context of this work, noise in gradient evaluation is not a consequence of the restriction on the number of points on which it is estimated, as in SGD. On the contrary, the gradient is always evaluated on the whole set of points at disposal at each iteration, while it is the function itself that is noisy in an experimental context.

Going more into the details, it is possible to consider a function $L(\vec{w})$ (e.g., the loss function for an ANN) that must be differentiable with respect to its input parameters. Since the function has a stochastic nature associated with a source of noise, the task is then reduced to finding the best combination of parameters that minimizes the expected value of the function $\mathbb{E}[L]$. At each i -th iteration of the descent, the gradient ∇L is evaluated and the exponential averages m and v are updated according to

$$\begin{aligned} m_{i+1} &= m_i \beta_1 + (1 - \beta_1) \nabla L(\vec{w}_i) \\ v_{i+1} &= v_i \beta_2 + (1 - \beta_2) [\nabla L(\vec{w}_i)]^2 \end{aligned} \quad (3.4)$$

with $m_0 = v_0 = 0$ fixed as initial conditions and $\vec{w}_i$ are the weights at the i -th iteration. Setting these initial conditions creates discrepancies of $\mathbb{E}[m]$ with $\mathbb{E}[g]$ and $\mathbb{E}[v]$ with $\mathbb{E}[g^2]$ (being g the gradient), which have then to be corrected. We will focus now on the case of the second moment estimation, by knowing that the procedure is analogous for the first moment.

As shown in [136], the moving average for $[\nabla f]^2 = g^2$ evaluated at the i -th iteration can be written as

$$v_i = (1 - \beta_2) \sum_{j=1}^i \beta_2^{(i-j)} g_j^2 \quad (3.5)$$

The expectation value for v_i results in

$$\begin{aligned} \mathbb{E}[v_i] &= \mathbb{E} \left[(1 - \beta_2) \sum_{j=1}^i \beta_2^{(i-j)} g_j^2 \right] \\ &= \mathbb{E} [g_i^2] (1 - \beta_2) \sum_{j=1}^i \beta_2^{(i-j)} + \xi \\ &= \mathbb{E} [g_i^2] (1 - \beta_2^i) + \xi \end{aligned} \quad (3.6)$$

where the term ξ contains the discrepancies introduced by equating the first and second lines. This term results to be null if $\mathbb{E}[g_i^2] = \mathbb{E}[g_j^2] \forall j < i$, otherwise we can assume ξ to be small since the parameter β_2 should be chosen to guarantee that gradients evaluated far away in the past are taken less into consideration while evaluating the average. Thus possible discrepancies between $\mathbb{E}[g_i^2]$ and $\mathbb{E}[g_j^2]$ are less important for small j . Neglecting then ξ , we are left with

$$\mathbb{E}[v_i] = \mathbb{E}[g_i^2] (1 - \beta_2^i) \quad (3.7)$$

which clarifies the factor that has to be inserted in order to correct the bias in the evaluation of the second moment. With an analogous procedure, it is possible to demonstrate that

$$\mathbb{E}[m_i] = \mathbb{E}[g_i] (1 - \beta_1^i) \quad (3.8)$$

holds for the evaluation of the first moment. Returning to the averages evaluated within the descents, the corrected estimations $\hat{m}_i$ and $\hat{v}_i$ assume the following form

$$\hat{m}_i = \frac{m_i}{1 - \beta_1^i} \quad (3.9)$$

$$\hat{v}_i = \frac{v_i}{1 - \beta_2^i} \quad (3.10)$$

The corrected estimations of the first and second moments of the gradient enter the proposal for the new step as follows

$$\vec{w}_{i+1} = \vec{w}_i - \alpha \frac{\hat{m}_i}{(\sqrt{\hat{v}_i} + \epsilon)} \quad (3.11)$$

where α is the step size provided by the user and ϵ is a small number that prevents divisions by zero. From this equation, it is evident how memory enters the descent since both the direction and the magnitude of the step proposed for a given weight depend on both the actual gradient and on the history of the descent itself.

Being still a gradient-based algorithm, it shows all the limitations associated with the simple gradient descent mentioned above. However, the introduction of memory allows the descent to proceed towards the minimum even when noise does not allow for a precise estimation of the gradient. When considering the simple gradient algorithm, the generation of the step is entirely based on the gradient value at the current iteration. At a certain point during the descent, the presence of noise in the loss function will generate fluctuations in the evaluation of the incremental ratio, with the consequent premature arrest of the research. On the contrary, introducing memory in the descent allows proceeding along the path outlined from the previous iterations, even when the gradient evaluation becomes imprecise. Indeed, the algorithm assigns much importance to the history of the descent rather than to the current evaluation of the gradient for generating the step.

3.2.3 Particle Swarm Optimizer

The Particle Swarm Optimizer (PSO), introduced by Kennedy and Eberhart in 1995 [141], is a population-based technique for solving nonlinear optimization problems. Since its inception, PSO has been extensively studied and improved, gaining widespread use across various domains, including multi-objective optimization, electronic circuit design, ANN training, and biometric security systems [142].

PSO draws inspiration from the social behavior observed in groups of animals, such as flocks of birds or schools of fish [141]. In these systems, individuals share information and benefit from one another's discoveries. While each group member explores its environment independently, the emergence of proper behaviors or advantageous positions by some agents influences the entire group. This cooperative learning process accelerates the group's progression toward more optimal states.

In computational terms, PSO seeks to minimize a given function, such as a loss function in ANN training, using a collection of candidate solutions known as particles. These form a swarm, where each of them represents a point in the function's configuration space. Each particle is characterized by a position and a velocity vector, both of which span the same dimensionality as the function's input space.

The algorithm begins by initializing the swarm with random positions and zero velocities. During each iteration, the objective function is evaluated for every particle. Based on individual and shared experiences, particles adjust their velocities and positions. These updates are influenced by three main components: the particle's own best-known position, the best-known position of the entire swarm, and a stochastic term that encourages exploration. When a particle identifies a promising region of the configuration space, it effectively communicates this to others, guiding the swarm's collective movement toward better solutions.

The optimization process continues until a stopping condition is met, such as reaching a predefined number of iterations or achieving sufficient convergence. This socially driven, feedback-based mechanism makes PSO a powerful and flexible tool for tackling a wide range of complex optimization problems. Compared to gradient-based alternatives, this minimization method applies to any function, without restrictions imposed by differentiability. It also allows for a wider exploration of the parameters' space, limiting the possibility of ending the research in a local minimum. This, however, comes at the expense of a longer minimization time.

3.3 Photonic Neural Networks

Nowadays, the most widely developed platform for supporting ANNs is the electronic domain, principally implemented as a classical von Neumann architecture [143]. However, this shows some intrinsic constraints related, for example, to the high power consumption, the performance limitations induced by the interference between electric signals, and the effort needed to handle a large amount of floating point operations, especially in parallel computing [144, 145]. These limitations are pushing the research in this field to the extreme in the attempt to find new solutions for future and even more demanding scenarios (telecom, cloud computing, AI data centers,...) [146, 147]. Despite the efforts, the industry is close to the limit in the transistor dimension [148], and electronic processing is nowadays the bottleneck in ultra-fast data processing applications [149].

A possible solution to overcome these constraints is to move the computing platform into the optical domain, thus implementing Photonic Neural Networks (PNNs). These are based on the same elements and architectures described above for a generic ANN, namely, each neuron still receives multiple inputs processed with some linear operations, whose result is then sent to an activation function. However, operations that are difficult to realize in the electrical domain become straightforward to implement optically. Light becomes the information carrier, enabling many advantages that

counteract the limitations imposed in electronics [150], as it will be discussed in a moment. Many different implementations and usages of PNNs can be found, but pertinent to this work are the applications of Silicon photonics platforms in data processing for telecommunications.

In this context, the main advantages consist of:

- **Large bandwidth:** the optical domain allows for an enlarged bandwidth, which is achieved both via high modulation frequencies and complex encoding formats. This enables the transport of more information per unit time compared to the 2-level formats typical of electronic computing;
- **Parallelization:** light shows an intrinsic predisposition to parallelism, which appears both at the transport level, for example with WDM propagation, and at the processing levels, where many operations can be performed simultaneously on a single dedicated photonic platform;
- **Scalability:** Silicon photonics technology is compatible with the standardized CMOS fabrication process used for electronic processors. This allows the integration of many components on a single device, with no need for a new manufacturing process to implement. The integration with other materials enables the co-implementation of active components on the same platform, such as photodetectors, modulators, and light sources. High propagation losses in Silicon represent the main limitation to scalability;
- **Complex domain operations:** light is described as a complex field, and thus it allows for performing efficient computations in this domain, exploiting natural phenomena such as interference. Further versatility is provided by the use of complex weights, which are naturally implemented in PNNs. Overall, handling complex numbers is computationally expensive in the electronic digital domain, while much less effort has to be provided in light-based platforms;
- **Nonlinear dynamics:** the photonic domain offers many sources of nonlinearities, which can be exploited in the context of PNNs as activation functions to enrich the system behavior.

A typical design of an integrated PNNs is based on building blocks such as Phase Shifters (PSs), Mach-Zehnder Interferometers (MZIs), Micro-Ring Resonators (MRRs), and Multi-Mode Interferometers (MMIs) [151]. These implement the primary optical operations of phase tunability, interference, filtering, splitting, and often represent the tunable degrees of freedom of the PNN. The connections between these building blocks are established through waveguides in what is called a Waveguide Optical Interconnection. Enlarging the field of view, different PNN architectures are determined by the topology of the connections between single neurons, analogously to the standard ANN case.

Light injected into an optical circuit can undergo both linear and nonlinear processing. On the linear side, as mentioned above, weighting and summing operations are easily implemented in the optical domain, giving the PNNs a great advantage in terms of speed and energy consumption compared to the electronic counterpart [151]. The total number of linear operations to be performed in a full ANN is very large, since a linear stage characterizes each neuron. Thus, the advantages provided by a transposition in the optical domain become more and more pronounced as the network scales in dimension. In a very general treatment, one can think about the linear component of a neuron as a series of parallel waveguides. Once the optical signal is injected into these lines, it can be modified using tunable waveguide elements (e.g., PSs) and then recombined at the end of the linear stage. Particularly relevant in this context is the capability of optical systems to perform matrix multiplication operations, whose electronic implementation is particularly demanding [39].

On the nonlinear stage, many possible choices can be made, as anticipated above, but some of them are found to be more straightforward to implement than others. A suitable activation function, especially in the context of telecommunications and thus of interest for this work, is the square modulus $|\cdot|^2$, since its implementation comes for free by detecting the output signal with a photodetector. Other sources of nonlinearity are represented by Semiconductor Optical Amplifiers (SOAs) operated in saturated regime [152], phase-change materials [153], and free-carrier dynamics in MRRs [154].

3.4 Applications to telecommunications

Machine Learning techniques are gaining popularity in the context of optical communications. These find applications in the optical transport network, in the physical layer with amplifier control, modulation format recognition, and Quality of Transmission estimation, as well as in the network layer with traffic prediction, traffic classification, and fault detection [155]. Their usage also extends to nonlinearity mitigation, where they are used as more efficient and versatile alternatives to the traditional equalization methods presented in Section 2.3.3.2. Machine learning algorithms such as Support Vector Machines [156], k-nearest neighbors [157], clustering methods (k-means, Gaussian Mixture Models,...) [158, 159], and ANNs [160] find applications in this context, receiving as input the transmitted symbols or constellations and returning as outputs their equalized version, or quality indicators such as OSNR, BER, or Q-factor. However, all of these techniques still rely on DSP, thus sharing the same drawbacks as the standard electronic equalization implementations.

In this context, PNNs are being increasingly considered candidates for substituting, or at least supporting, DSP algorithms. The photonic platform perfectly fits in the context of telecom networks, allowing signal processing in the optical domain where the propagation occurs. Optical signals can be directly fed to the PNNs without any prior need for conversion to the digital domain, enabling fast and efficient processing and reducing the computational effort provided by electronics. Photonics platforms appear

as a promising support in terms of performance, costs, and sustainability to supply the constantly increasing market demand. Many transceiver providers are pushing towards electronic-photonic integration, with the first examples already available in the market [161, 162].

3.4.1 Overview of the implementations for signal equalization

Since the early stages of the research in PNNs for telecom applications, many different solutions have been experimentally proven to be effective in linear and nonlinear effects mitigation [38, 44, 163, 164]. The various implementations span different architectures, with an increasing level of complexity for more demanding transmission scenarios. Common features among the different implementations are the tunable weights, which can operate directly on the optical signal or at the readout level, and the presence of memory sources within the optical circuit. These can be represented by delay lines, nonlinear dynamics, and recursivity, which provide the possibility of storing optical information from different time instants, even if for a very short amount of time. This property is fundamental in counteracting CD- and SPM-related distortions, which indeed arise from the interference between temporally adjacent symbols and thus require accessing optical samples close in time to reconstruct the original symbol shape. Some architectures are presented here, but a more detailed comparison of their performances is proposed in Section 4.5.

Among the devices that implement memory via optical delay lines, it is possible to find integrated lattice filters, which provide a suitable platform to tackle the problem of CD equalization. A typical design, such as that taken from [165, 166] and schematized

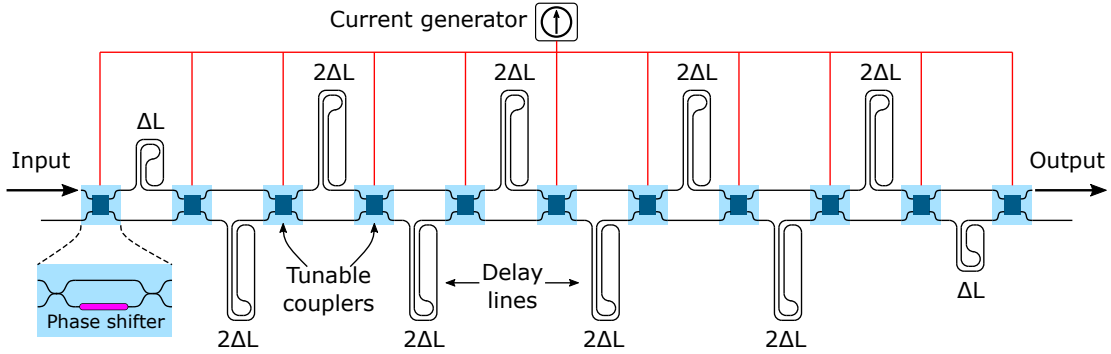


FIGURE 3.3: Schematics of a tunable lattice filter implemented in SiN and used in [165, 166] for CD equalization. The input signal propagates through subunits constituted by a tunable coupler and an unbalanced MZI. The imbalance is provided by delay lines implemented via spirals of lengths ΔL (first and last units) or $2\Delta L$. The inset shows the internal structure of a tunable coupler, which is represented by a balanced MZI with a tunable PS. The couplers' response is driven by a common external current generator, which controls all of them simultaneously.

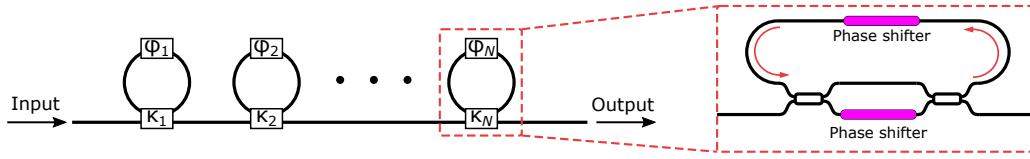


FIGURE 3.4: Schematics of a cascade of N MRRs used in [167, 168] for CD equalization. As shown in the inset, within each subunit, the coupling between the bus waveguide and the resonant structure is provided by an MZI, which can be thermally tuned to regulate the corresponding coupling coefficient κ_i with $i = 1, \dots, N$. Each MRR is then equipped with another PS that drives the phase weight ϕ_i .

in Figure 3.3, is based on a cascade of repeated units constituted by a tunable coupler and an unbalanced MZI. The couplers, in turn, are constituted by balanced MZIs with one of the arms equipped with a PS driven by an external current generator. At each tunable coupler, spectral components are sent respectively to the upper and lower arms of the subsequent unbalanced MZI, thus imposing a longer propagation path to specific wavelengths. These accumulate a delay with respect to those traveling in the straight waveguide, balancing the effect generated by CD. This platform is suitable for addressing both the normal and anomalous dispersion problems, which can be selected by retarding the shorter or longer wavelengths in the signal spectrum. It is also compatible with WDM operations, provided that all the delay lines are multiple in length of a given ΔL . This value and the number of stages determine the whole structure's behavior in terms of Free Spectral Range, maximum achievable dispersion, and bandwidth.

Other filter implementations are based on MRRs architectures, which can be arranged in different topologies to mimic the inverse transfer function of CD. References [167] and [168] propose sequences of, respectively, 7 and up to 9 MRRs coupled to the same bus waveguide, following the scheme proposed in Figure 3.4. The overall frequency response of the structure is determined by the waveguide-MRR coupling coefficients κ_i and the phase shifts ϕ_i imposed in each structure, setting the maximum achievable dispersion and the bandwidth. When used in a linear regime, as is the case for the cited works, memory is provided by the resonant nature of the MRRs, which allows storing the optical information in the signal coupled to the rings.

It is worth noting that the approaches presented so far as programmable optical filters can be interpreted as the linear stages of PNNs whose activation function is the square modulus applied at the end-of-line photodetector. Indeed, the parameters related to each structure require a proper regulation to provide the desired amount of dispersion. When these structures are inserted in a transmission line, a training procedure provides for the tuning, driven, for example, by maximizing the signal quality at the receiver (BER, OSNR, eye diagram aperture,...).

Other approaches to channel equalization implement PNNs based on the Reservoir

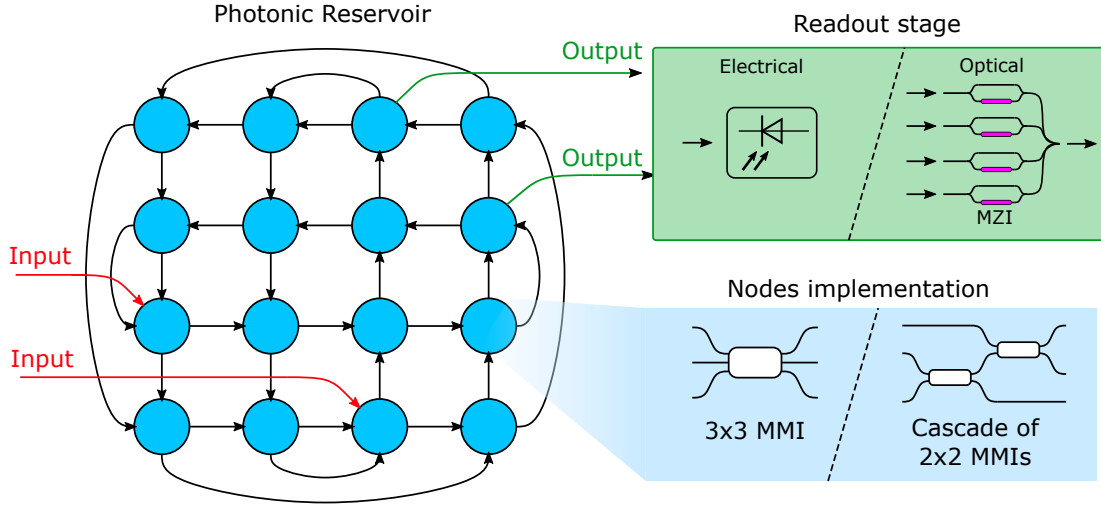


FIGURE 3.5: Schematics of a 16-node photonic reservoir implementation featuring the same building blocks used in [169–172] for nonlinear equalization. Possible node implementations as 3×3 structures are presented (blue inset). Two nodes in the photonic reservoir receive the input optical signal (red arrows), and two other nodes provide the readout stage with the elaborated optical output (green arrows). The readout (green inset) can be implemented in the electrical domain through the detection process [169–171] or in the optical domain as a weighted summation tree [172].

Computing scheme described in Section 3.1. Numerous implementations of the Reservoir exist, each exploiting the randomness and recursivity features offered by different photonic platforms. In [173], the reservoir consists of a semiconductor laser operating in the C-band and an optical fiber delay loop from which the laser receives delayed feedback. The optical processing provided by the Reservoir projects the input signal to a higher-dimensional space, where the action of a linear classifier completes the equalization. A very similar workflow is adopted in a series of works where the reservoir is implemented as a passive integrated photonic circuit [169–172]. In a general treatment, these structures receive the distorted optical signal and operate an optical processing analogous to that schematized in Figure 3.5. Each circle represents a node, which is implemented as a 3×3 MMI (or similar structures with 3 input and 3 output ports), and the arrows indicate the inter-node connections, which are realized via spiralized waveguides. The input signal can be fed to the reservoir at different nodes, as well as the output can be extracted from many ports at the readout stage. Notice that in these cases, no intrinsic nonlinearity is incorporated in the reservoir at each node. All of it is enclosed in the square nonlinear response of the photodetector that collects one or multiple outputs from as many nodes across the photonic chips. In a first implementation proposed in [169], the output of a 32-node reservoir is post-processed via software to maximize the similarity with a digital target sequence. This allows reaching nonlinear equalization of a 32 Gbaud PAM2 signal after 25 km of fiber. The same hybrid approach featuring optical signal processing followed by an electronic readout stage has also proven effective in a WDM scenario [170], as well as for nonlinear equalization

in coherent systems [171]. Ultimately, an all-optical equalization has been proposed, combining a reservoir structure with a weighted summation tree [172].

3.4.2 All-optical time-delayed complex perceptron

Another PNN topology that finds many applications in the telecom domain is the so-called Time-Delayed Complex Perceptron (TDCP) [174]. This implements a Feed-forward PNN that represents an all-optical transposition of the artificial neuron, with the addition of delay lines on the input channels. Specifically, in the photonic implementation under study in this work, the TDCP design is composed of an N -line FIR filter in the optical domain, followed by a nonlinear activation function. The FIR filter operates the linear processing as illustrated in Figure 3.6, acting on the input optical signal $x(t)$ (complex field) to generate an output $y(t)$ according to

$$y(t) = \sum_{i=1}^N x[t - (i-1)\Delta t] a_i k_i e^{j\phi_i}. \quad (3.12)$$

The input signal $x(t)$ is first split into N separate waveguides (channels, or taps), where each duplicate undergoes a delay that is an integer multiple of Δt . The delay is imposed by propagating the copies through spirals of different lengths. The channels are numbered from 1 to N , the first hosting no spirals and the N -th corresponding to the maximum delay, namely $(N-1) \times \Delta t$. After accumulating the respective time-shift, each copy is modified by applying a trainable amplitude weight a_i ($0 \leq a_i \leq 1$) and a trainable phase shift $e^{j\phi_i}$. Each channel is then affected by another fixed contribution to amplitude weights k_i , which in a real implementation can be induced, for example, by the power splitters' ratio and by propagation losses in the optical medium. Once all

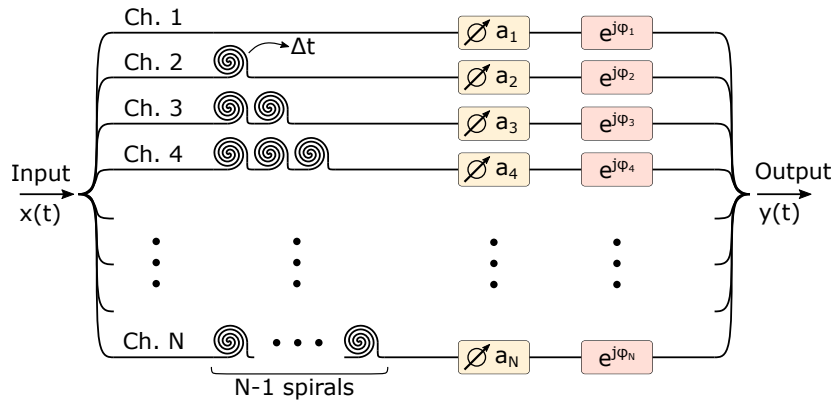


FIGURE 3.6: Schematics of an N -channel Time-Delayed Complex Perceptron. The optical input time sequence $x(t)$ is split into N copies, which then propagate through spiralized waveguides to accumulate a relative delay multiple of Δt . Copies traveling in each channel are then applied with independent amplitude a_i and phase weights $e^{j\phi_i}$ ($i = 1, \dots, N$). Finally, they reach the recombination stage where they are complex-summed to produce the optical output sequence $y(t)$.

delayed and weighted copies have been processed, they are recombined and summed together to produce the optical output $y(t)$ (complex field), subsequently sent to the nonlinear stage. In comparison with the artificial neuron described in Figure 3.1(a), the inputs to the linear stage there indicated as x_i are here represented by delayed copies $x[t - (i - 1)\Delta t]$ of the same input signal. Moreover, the weights defined as w_i for the standard neuron (Equation 3.1), now contain three contributions (a_i , k_i , and $e^{j\phi_i}$), assuming, thus, a complex nature. In the photonic implementation, crucial design parameters are the unitary delay Δt , which regulates the density of the optical samples reaching the recombination stage simultaneously, and the number of taps N , determining the total observation window (i.e., the memory) of the device as $(N - 1) \times \Delta t$.

For this work, the device under test is represented by a Photonic Integrated Circuit (PIC) that implements the linear stage of a TDCP. The PIC is used in combination with a photodetector placed at the end of an optical line, which provides for the square modulus to act as the nonlinear activation function. Figure 3.7(a) presents the layout of the PNN device (i.e., the PIC), which was realized by Applied Nanotools foundry [175] and designed by D. Bazzanella and M. Mancinelli. It is fabricated in a Silicon-on-Insulator platform, and made by Si waveguides with a $220 \times 500 \text{ nm}^2$ cross-section surrounded by Silica cladding. The input and output ports are located respectively at the leftmost and rightmost edges of the chip. The input signal $x(t)$ enters the PNN device via edge coupling and propagates through a cascade of 50/50 y-branches and spiralized optical paths. Overall, this creates $N = 8$ time-delayed copies of the input signal, ending in as many parallel waveguides (channels). Spirals are available in 3 different lengths, respectively $L_A = 7.10 \text{ mm}$, $L_B = 3.55 \text{ mm}$, and $L_C = 1.77 \text{ mm}$, each corresponding to an added propagation time of $\Delta t_A = 100 \text{ ps}$, $\Delta t_B = 50 \text{ ps}$, and $\Delta t_C = 25 \text{ ps}$. The channels are labeled from the first ($i = 1$) at the bottom edge to the eighth ($i = 8$) at the top edge. The i -th channel hosts a delayed copy $x_i(t) \propto x[t - (i - 1)\Delta t]$ of the input signal, where the delay is measured with respect to the first channel. The quantity $\Delta t = \Delta t_C$ is the delay unit imposed by the shortest spirals between copies traveling in adjacent channels. The eighth channel is associated with the maximum delay of $\Delta t_A + \Delta t_B + \Delta t_C = 7\Delta t$. Each channel hosts an MZI with a footprint of $400 \times 150 \mu\text{m}^2$, and two y-branches provide the initial splitting and final recombination. Each MZI is controlled via the thermo-optic effect produced by a current-driven TiN micro-heater placed on top of one of the arms. The final aperture level (i.e., the transmittance) of the MZI determines the amplitude weight a_i applied to the optical signal traveling in that channel. A phase weight $e^{j\phi_i}$ is then generated for each channel by another current-controlled micro-heater acting as a phase shifter (PS). The weighted copies are then complex-summed via an 8×1 combiner realized via a cascade of 50/50 y-branches. The results of a characterization of the main components of the chip are reported in Appendix A.3.

Figure 3.7(b) illustrates the working principle of the PNN device upon the arrival of a 100 ps optical pulse, which is pictured in the top-right corner. Colored dots equally spaced by $\Delta t = 25 \text{ ps}$ individuate some optical samples in the input pulse, with the purple one being the first entering the PNN device at $t = 0$ and the red one being the last at $t = t_0 = 175 \text{ ps}$. Once the pulse has entered the PNN device, it is split along 8 lanes,

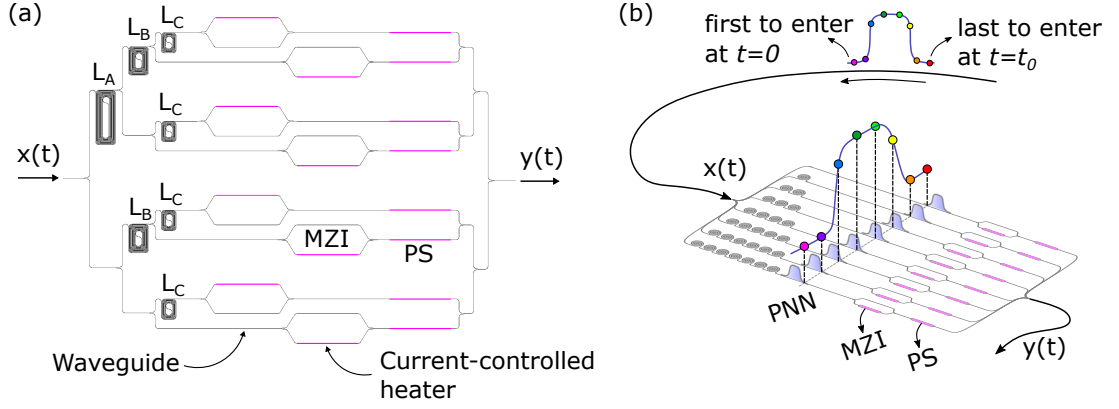


FIGURE 3.7: (a) Layout of the 8-channel PNN device. Y-branches implement 1×2 and 2×1 structures, used in each splitting and combination stage, respectively. Spirals are available in 3 lengths: $L_A = 7.10$ mm, $L_B = 3.55$ mm, and $L_C = 1.77$ mm. These correspond to an added propagation time, respectively, of $\Delta t_A = 100$ ps, $\Delta t_B = 50$ ps, and $\Delta t_C = 25$ ps. Amplitude and phase weights are implemented via an MZI and a PS hosted in each channel. These are controlled via current-driven micro-heaters placed on top of the waveguides (pink lines). (b) Illustration of the PNN device working principle upon the arrival of a 100-ps optical pulse. The optical samples (colored dots) from the impinging pulse enter the chip in sequence, but then they are parallelized in time by the combined action of the splitting stage and the delay lines. As a result, they proceed simultaneously through both the weighting stage (MZI and PS) and the recombination.

where the 8 copies travel in parallel, accumulating a relative delay of Δt . This results in the colored optical samples associated with the different copies of the input signal being temporally aligned at time $t = t_0$ after the delay stage. The optical samples serially inserted in the PNN device are now parallel propagating in the corresponding channels, finally reaching the summation stage simultaneously. In other words, the delay lines parallelize the input information, allowing for the simultaneous processing of pieces of optical information initially separated in time.

In the context of this work, the PNN device is used to compensate for distortions induced by CD and SPM on IMDD optical signals, with the currents driving the amplitude and phase weights representing the trainable parameters. Since the optical response of the device is defined but for one of the phase weights $e^{j\phi_i}$, the PS associated with the null-delay channel (Channel 1) is left unconnected and chosen as reference. Moreover, unfortunately, during a measurement session, the electrical connection controlling the MZI of the 5th channel accidentally broke, leaving us with a total of 14 tunable weights (7 MZIs and 7 PSs). A prior calibration showed that the rest state of the 5th MZI (i.e., with no applied current) corresponds to a fully open configuration, resulting in a fixed amplitude weight $a_5 = 1$. Within the context of signal equalization, this constraint does not represent a limitation to the device performance. Indeed, as illustrated in Figure 3.7(b), when the optical samples associated with a symbol and its two nearest neighbors

reach the recombination stage, the main contribution to the information of the symbol itself is primarily carried by the device's central channels. These delay contributions are therefore expected to be essential for accurate symbol reconstruction, and consequently, the corresponding channels are in any case foreseen to operate in an open state, namely, near-unity transmission.

The design specifics in terms of the number of taps N and unitary delay Δt for the device have been determined starting from the following empirical formula:

$$N = \text{int} \left(\frac{1/B + |L\beta_2\Delta\omega|}{\Delta t} \right). \quad (3.13)$$

Here, the numerator represents an estimate of the increased time width of a Gaussian pulse after propagation in a fiber of length L , obtained as the sum of the initial baud time slot $1/B$ (with B being the baud rate) and the pulse broadening $\Delta T = |L\beta_2\Delta\omega|$ induced by CD (see Equation 2.11). We consider the GVD parameter $\beta_2 = -0.021 \text{ ps}^2/\text{m}$ of a standard SM G.652D fiber, and $\Delta\omega$ is the signal bandwidth. In a target application for 10G signals in $L = 100 \text{ km}$, the selected $N = 8$ and $\Delta t = 25 \text{ ps}$ are a trade-off between a sufficiently dense optical sampling (4 samples per baud), insertion loss minimization, and a reduced number of trainable parameters. Finally, the design of the photonic structures implemented on the PNN device is optimized for the Transverse Electric mode propagation, making the PNN device polarization sensitive.

Chapter 4

Chromatic Dispersion equalization

This chapter treats the use of the 8-channel PNN device (Figure 3.7) for CD equalization in optical links up to 125 km. The linear nature of the device well matches the linear equalization problem posed by GVD, making its implementation straightforward in this context. The TDCP used for CD compensation candidates as a substitute for digital filter-based equalizers, as well as for other all-optical techniques, such as DCFs and chirped FBGs.

Here, a description of the setup and the loss function definition is provided, followed by the experimental results obtained for Pulse Amplitude Modulated (PAM) and Orthogonal Frequency Division Multiplexed (OFDM) signals at different propagation distances and bit rates. Further, we treat an interpretation of the optimal weights based on the theoretical model, and a study of the system's response in the frequency domain, both with and without the active equalization. Finally, a comparison with the existing literature is performed. The discussion presented in this chapter reflects the content of [176, 177].

4.1 Experimental Setup

The PNN device has been tested with the experimental setup presented in a simplified version in Figure 4.1. An extensive description of the apparatus and the experimental procedures can be found in Appendix A.1. In the transmission stage, a laser source operating at 1550 nm is modulated as a 10 Gbaud 2 or 4-level PAM signal based on periodic Pseudo-Random Binary Sequences (PRBSs), producing an overall period of 2^{10} symbols. A Fiber Optic Coupler sends part of the optical signal to a 20 GHz-bandwidth fast photodiode (RX1), where the input signal is acquired for reference. The other fraction proceeds to the PNN device for optical processing, as described in Equation 3.12, with the amplitude and phase weights being regulated by an external DC generator. The output signal from the PNN is then sent through an optical fiber span with variable length, from 0 km (Back-To-Back (BTB) configuration) to 125 km with steps of 25 km. The position of the PNN device within the transmission line makes it, in fact, a pre-compensator for CD effects. Placing the PNN device before the propagation stage

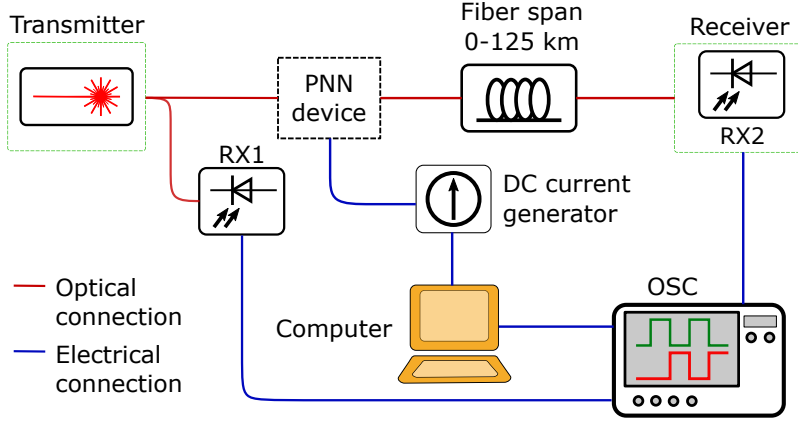


FIGURE 4.1: Experimental setup (simplified) used to test the PNN device for CD equalization up to 125 km.

allows for getting rid of polarization diversity effects since the device is polarization-sensitive. The optical power in the setup remains low enough to avoid nonlinear effects and damage in any component. In particular, it stays below 3 dBm at the input of the PNN and below 2 dBm at the input of each optical fiber span. After fiber propagation, the optical signal is detected by another 20 GHz bandwidth fast photodiode (RX2). Both RX1 and RX2 are connected to a 16 GHz bandwidth oscilloscope with sampling frequency $SF_{osc} = 80 \text{ GSa/s}$ (otherwise differently specified).

For each measure (i.e., transmission test and quality assessment), the DC generator sends an array of pre-set currents to the PNN device to drive the amplitude and phase weights. A triggering signal is then sent to the oscilloscope to acquire the periodic trace at the end-of-line receiver resulting from the PNN processing and fiber propagation. The output trace is aligned via the cross-correlation method with the digital target sequence obtained as a periodic version of the digital input sequence. Each temporally aligned output trace y serves as an input to a properly defined loss function that assesses the quality of the transmission. The same loss function is also provided to the minimization algorithms for the PNN training. In this context, both the PSO (Section 3.2.3) and the Adam algorithm (Section 3.2.2) are tested for a comparison in terms of accuracy and convergence speed.

First, the loss function evaluates the BER, here defined as the sum of the discrepancies between the binary demapped version of the digitized target and output sequences. As discussed in Sections 3.2.2 and 3.2.3, the discrete-valued nature of the BER function makes it not well-suited for a gradient-based approach to minimization. In addition, a BER value of 0 can be obtained statistically even in sub-optimal equalization conditions, possibly leading the PSO algorithm to end the research prematurely. Therefore, it is necessary to define a new analog loss function that allows reaching minimum BER but with higher versatility and applicability.

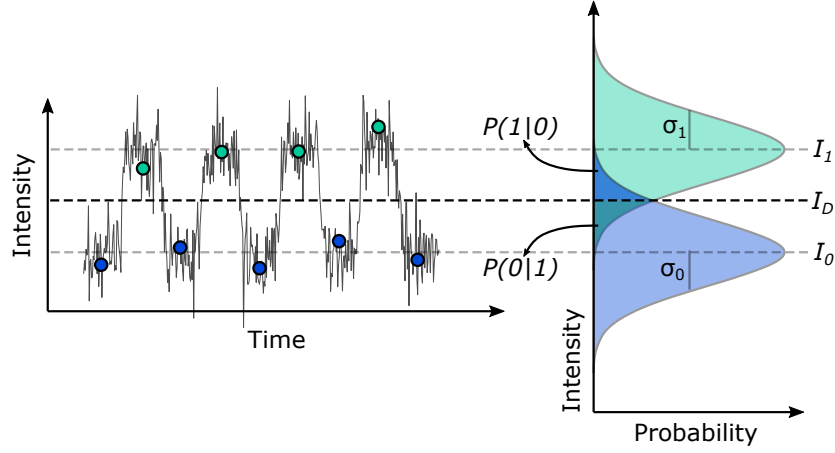


FIGURE 4.2: Graphical representation of the quantities entering in the BER evaluation based on incorrect bit identification probability. In the intensity profile (left plot) of a received optical signal, samples associated with 0s (blue) or 1s (green) oscillate around I_0 and I_1 , which represent the mean values of the associated distributions (right plot). A threshold set by I_D applied to the samples discriminates between the two classes. The probability of incorrect bit identification is given by the overlap integral between the distributions, namely the sum of the probability $P(0|1)$ of identifying as 0 a bit originally sent as 1 and $P(1|0)$ vice versa.

4.1.1 Separation loss function

Figure 4.2 illustrates the procedure leading to the theoretical evaluation of the BER starting from a trace acquired at the receiver. Considering a PAM2 modulation, the samples associated with 0s or 1s (blue and green dots, respectively) oscillate around the values I_0 and I_1 . In the presence of a random Gaussian noise characterized by standard deviations σ_0 and σ_1 affecting 0s and 1s in the bit sequence, the acquired samples generate the two corresponding distributions on the right side of Figure 4.2. I_D represents a decision threshold, namely, received samples lying above or below it are labeled as 0s and 1s during the digitization. In this process, errors occur when samples in the green distribution leap below the threshold I_D , leading to identifying as 0 a sample that originally was sent as a 1. Analogously, original 0s are identified as 1s when the blue distribution gets above the threshold. The probability of these two events is expressed by two terms: $P(0|1)$ accounts for possible identifications of a 0 in place of an original 1, and $P(1|0)$ vice versa. Their sum provides the overall probability of incorrect bit identification, which corresponds to the overlap integral between the two distributions. The BER can thus be computed as [178]

$$\text{BER} = \frac{1}{2} [P(0|1) + P(1|0)] = \frac{1}{4} \left[\text{erfc} \left(\frac{I_1 - I_D}{\sigma_1 \sqrt{2}} \right) + \text{erfc} \left(\frac{I_D - I_0}{\sigma_0 \sqrt{2}} \right) \right]. \quad (4.1)$$

with erfc being the complementary error function.

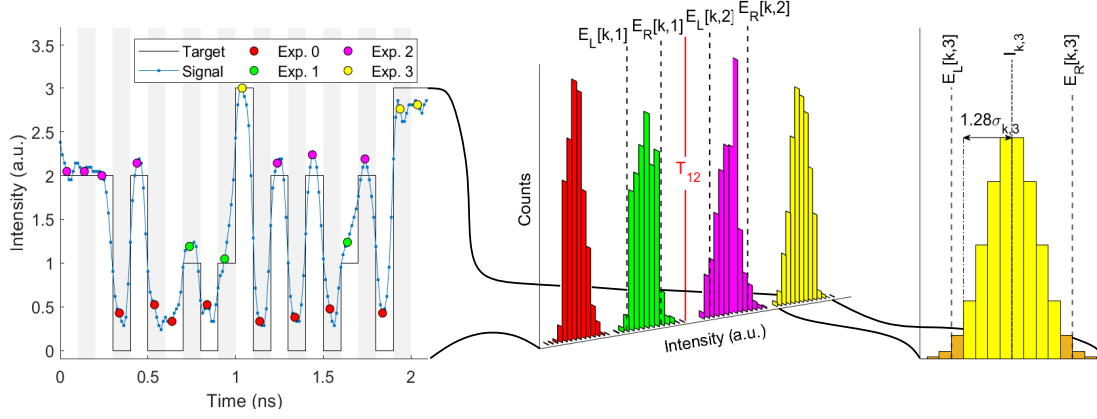


FIGURE 4.3: Illustration of the one-sample separation loss function $\mathcal{L}_1$ evaluation. The left panel presents the trace acquired at the receiver (blue line) overlapped with the digital target (black line). The sub-sequence on which the loss function is evaluated is represented by the large colored dots, which lie close to the center of the symbol and correspond to the maximum contrast between the levels. The points produce the distributions presented in the central panel. For each of these, the positions E_L and E_R of the left and right tails are estimated, respectively, as the average over the 10% of the leftmost and rightmost points in the distribution, as illustrated in the right panel. Finally, the relative position between the tails enters into the definition of the separation loss function.

In view of these considerations, an analog loss function $\mathcal{L}$, from now on called the separation loss function, is created to measure the distance between the adjacent tails of the distributions of the expected optical levels. Its maximization quickly produces a BER reduction, since creating a distance between the tails (i.e., pushing the distributions far apart) minimizes the overlap region and produces a clearance for a safe threshold application and bit identification. In the following paragraphs, the definition of $\mathcal{L}$ is provided for a 4-level signal for an improved generality compared to the simpler 2-level case.

Figure 4.3 provides a visual representation of the procedures leading to the loss function evaluation. Consider a modulated optical sequence with a baud rate B acquired in a time window t_w by the oscilloscope sampling at SF_{osc} . Each measurement (i.e., oscilloscope acquisition) produces an $N_s = t_w \times SF_{osc}$ samples-long output trace y (blue line), with $N_{sps} = SF_{osc}/B$ samples per symbol. Typical values of $t_w = 1 \mu s$, $SF_{osc} = 80$ GSa/s, and $B = 10$ Gbaud in use during our experiments lead to $N_s \approx 8 \times 10^4$ and $N_{sps} = 8$. The alignment between y and the target identifies the samples in each symbol, labeling them from $k = 1, \dots, N_{sps}$. A sequence y_k (colored large dots) is then obtained by sub-sampling y at the k -th sample in each symbol and normalizing its maximum to $N_L - 1$, with N_L being the number of levels in the chosen PAM format. The points in y_k are grouped into $\{y\}_{k,n}$ based on their expected level $n = 0, \dots, N_L - 1$ determined by comparison with the target sequence. In Figure 4.3, the points belonging to each $\{y\}_{k,n}$ are marked with a different color and generate the distributions shown in the central

panel. Each distribution is populated by N_n samples $\{y^i\}_{k,n}$, with $i = 1, \dots, N_n$. The rightmost panel focuses on a single histogram, showing the typical Gaussian envelope identified by its average $I_{k,n}$ and variance $\sigma_{k,n}$. The position of the left and right tails, respectively $E_L[k, n]$ and $E_R[k, n]$, can be estimated as

$$E_L[k, n] = \frac{1}{n_L} \sum_{i=1}^{n_L} \{y^i\}_{k,n}, \quad (4.2)$$

$$E_R[k, n] = \frac{1}{n_R} \sum_{i=1}^{n_R} \{y^i\}_{k,n}. \quad (4.3)$$

In Equation 4.2, the index i is such that $\{y^i\}_{k,n} < I_{k,n} - 1.28\sigma_{k,n}$, namely the sum is performed over the n_L points corresponding to the 10% of the population in the leftmost area of the distribution. Analogously, in Equation 4.3 i is such that $\{y^i\}_{k,n} > I_{k,n} + 1.28\sigma_{k,n}$, namely the sum is performed over the n_R points corresponding to the 10% of the population in the rightmost area of the distribution. These areas are identified by the orange bars in the third panel of Figure 4.3. The separation between the distributions is measured as

$$\mathcal{S}(k) = \max_{n=0, \dots, N_L-2} \{E_R[k, n] - E_L[k, n+1]\}, \quad (4.4)$$

which returns negative values when all the distributions are correctly separated. At this point, we define the one-sample separation loss function as

$$\mathcal{L}_1 = \min_{k=1, \dots, N_{sps}} \{\mathcal{S}(k)\}, \quad (4.5)$$

namely, we select the minimum value of $\mathcal{S}(k)$ among those obtained for each possible sub-sampling choice. The minimum $\mathcal{S}(k)$ is registered for $k = k_{opt}$, which corresponds to the highest contrast in the sequence, a condition that is usually satisfied for samples close to the center of the symbol, as it is shown in Figure 4.3(a). A minimization of $\mathcal{L}_1$ produces the maximum separation between the optical levels, that is, the maximum aperture of the eye diagram (described in Appendix A.1.2). Since the aperture level is measured for a single sample (the k_{opt} -th) close to the center of the symbol, the minimization may lead to a local separation of the levels, losing the symmetry in the eye diagram aperture around its vertical axis. Therefore, it is convenient to define a two-sample separation loss function as

$$\mathcal{L}_2 = \mathcal{L}_1 + \min_{k'=k_{opt} \pm 1} \{\mathcal{S}(k')\}. \quad (4.6)$$

This measures the separation between the optical levels at two different samples (the optimal one k_{opt} and $k_{opt} \pm 1$) within each symbol, respectively symmetrical to its center. In particular, with this definition of $\mathcal{L}_2$, we implicitly assume that the actual center of the symbol is comprised between the two samples that produce the minimum values for $\mathcal{S}$. A minimization of $\mathcal{L}_2$ reduces the risk of asymmetric eye diagrams both on the horizontal and vertical axes.

Finally, the BER evaluation starts with the digitization of the output trace. Returning to a generic sub-sampled sequence y_k , the separation threshold T_n between the sample distributions associated with the n -th and the $n + 1$ -th levels are defined as

$$T_{n,n+1} = \frac{1}{2} (E_R[k, n] + E_L[k, n + 1]). \quad (4.7)$$

with $n = 0, \dots, N_L - 2$. These are then applied to y_k producing the corresponding 4-level digitized sequence y_{dig} , which is then de-mapped into the binary domain y_{bin} . The BER is evaluated as an error counting between y_{bin} and the corresponding de-mapped binary target sequence $\bar{y}_{bin}$. Again, a scan over different samples $k = 1, \dots, N_{sps}$ is performed, finally selecting the lowest BER value among those provided by the different sub-samplings. Notice that this includes the possibility of having two different optimal samples which minimize, respectively, $\mathcal{L}_1$ and the BER.

4.2 Premises on experimental measurements

During data acquisition, the experimental process was divided into two distinct phases. Considering a given link length L , the *training* phase was carried out first, during which the PNN device's weights were tuned to minimize the chosen loss function. The optimization was always performed with the PSO, otherwise differently specified. Subsequently, the *testing* phase took place, during which the quality of the transmission was effectively assessed with the trained PNN in the line. The evaluation consisted of a scan over the Power at the Receiver (PRX), or equivalently over the Signal-to-Noise Ratio (SNR), performed by varying an optical attenuator placed in front of the end-of-line photodiode (see Figure A.1 in Appendix A.1). For each PRX, the BER was acquired as the average over 50 acquisitions, with the error bars on each point corresponding to a 68% confidence interval estimated by the Bootstrap method with 10^3 iterations [179, 180]. From now on in this manuscript, every null BER value appearing in experimental plots or maps is substituted with 10^{-6} , which is lower than the minimum measurable BER due to the finite dimension of the datasets. The testing phase was also performed by removing the PNN from the setup, to compare the equalized and unequalized transmissions.

In this work, when dealing with CD equalization, no distinction is made between the PRBS sequence used for the training and the testing phases. Consider, indeed, Equation 2.11, which measures how much information about a symbol has spread to the adjacent ones. The chosen worst-case scenario for a 10 Gbaud signal in the linear propagation regime is represented by $L = 125$ km, leading to $\Delta T = 115$ ps and a total symbol time width of $100 \text{ ps} + \Delta T = 215$ ps. Similarly, in a 20 Gbaud scenario, the limit is set to $L = 75$ km, corresponding to $\Delta T = 138$ ps and a total symbol time width of $50 \text{ ps} + \Delta T = 188$ ps. It follows that for a 10 Gbaud signal, the worst generated ISI involves 3 adjacent symbols (the central one and its left and right prime closest), while for a 20 Gbaud propagation, the span is increased to 5 bauds. In both cases, the distortions result

from the interaction of a limited number of symbols, making the chosen PRBS (2 or 4 levels) input sequence sufficient to let the PNN explore all the possible cases.

4.3 Application to PAM formats

The first tests have been performed on the simple 2-level PAM modulation to prove the applicability of the PNN device to the equalization task, even at different modulation frequencies. Subsequently, more efforts have been devoted to the PAM4 case, being this of more interest for the commercial applications. In this case, other than presenting the bare equalization results in terms of BER and power penalty reduction, a more detailed analysis is performed. This includes an investigation of the optimal loss function, a comparison between PSO and Adam as optimizers, an insight into the role of the amplitude weights, a comparison with a commercial device, and with an ideal PNN design.

4.3.1 Application to 2-level 10 and 20 Gbaud signals

First, we present the results for the unconstrained optimization performed with 14 free parameters (7 amplitude and 7 phase weights, see Section 3.4.2) using the 1-sample separation loss function $\mathcal{L}_1$. Figure 4.4 shows the BER versus PRX profiles obtained for bare fiber (colored stars) and equalized transmissions (colored circles). The reference TX/RX performance is measured by the BTB curve (black stars) corresponding to a 0 km optical link. For low PRX values, all the curves overlap because the SNR at the

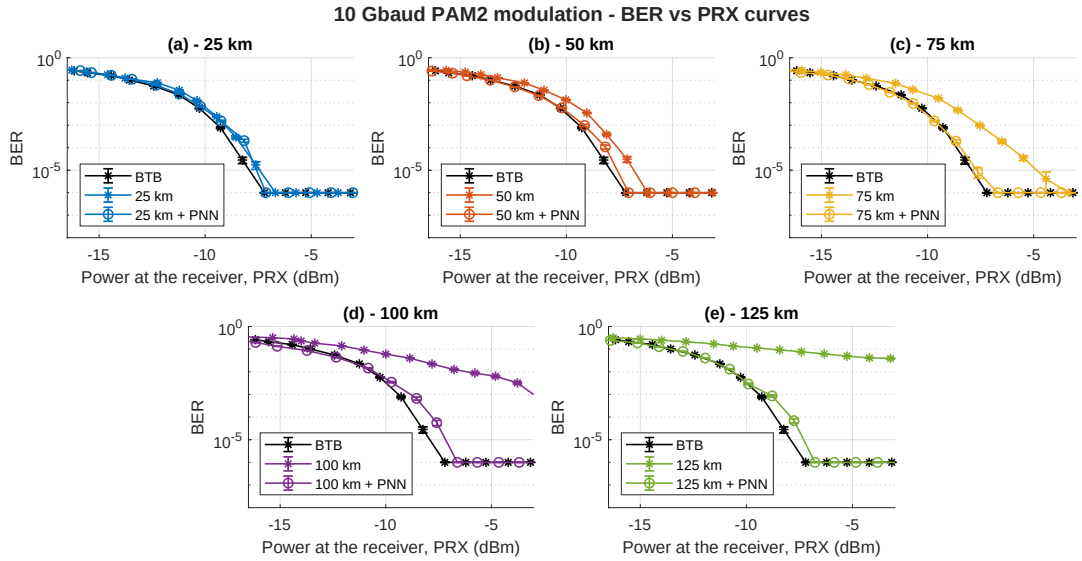


FIGURE 4.4: BER versus PRX profiles for 10 Gbaud PAM2 signal in links of different lengths ($SF_{osc} = 40$ GSa/s). Black lines set the reference BTB transmission, while colored lines with stars and circles refer to the bare and equalized transmission.

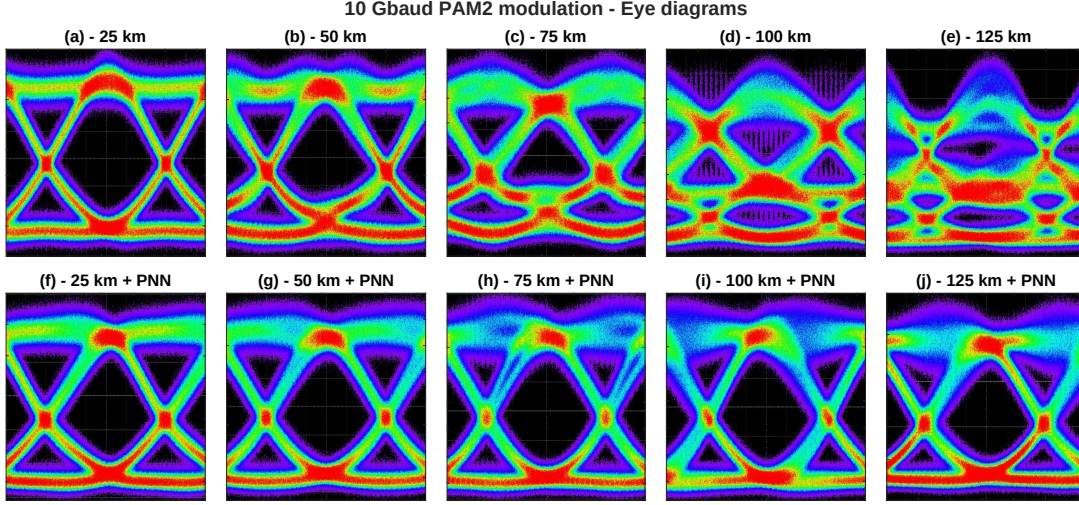


FIGURE 4.5: Eye diagrams at the receiver for 10 Gbaud PAM2 ($SF_{osc} = 40$ GSa/s). Bare (a-e) and equalized (f-l) transmission in links of different lengths. Jiggles in (d) are due to timing errors in the oscilloscope during the eye diagram construction.

receiver decreases and the BER evaluation is dominated by noise, which is present regardless of the length of the fiber. On the contrary, increasing the power at the receiver, and thus the SNR, the most dominant contribution to the BER is associated with the accumulated ISI due to signal propagation in the fiber. CD increases with the fiber length, causing a worsening of the BER even at high PRX. The effect of accumulated ISI is evident in the closing of the eye diagram with the length of the fiber, as illustrated in Figures 4.5(a-e). The trained PNN recovers the signal and, thus, brings the BER curves close to the reference BTB curve in Figure 4.4 and opens the eye diagrams in Figures 4.5(f-j). The beneficial effects are more evident for longer fiber links, namely where the amount of ISI is higher, with a 5 order of magnitude BER reduction at $PRX = -7$ dBm between the equalized and unequalized case at 125 km.

The same considerations reported so far apply also to the equalization of a 20 Gbaud signal, whose results in terms of BER versus PRX curves and eye diagrams are reported in Figure 4.6. The unequalized BER profiles depart faster from the BTB profiles compared to those obtained at 10 Gbaud for the same fiber length. Indeed, a larger bandwidth generates a more severe ISI condition, thus requiring a larger number of taps N and a smaller unitary delay Δt for an optimal equalization. Despite the PNN design being optimized for 10 Gbaud signals with the unitary delay corresponding to a quarter of a symbol width, good performance in CD equalization is obtained even at 20 Gbaud, namely where Δt corresponds to half the symbol width. At 75 km, the PNN ensures a BER reduction of 4 orders of magnitude at $PRX = -2$ dBm.

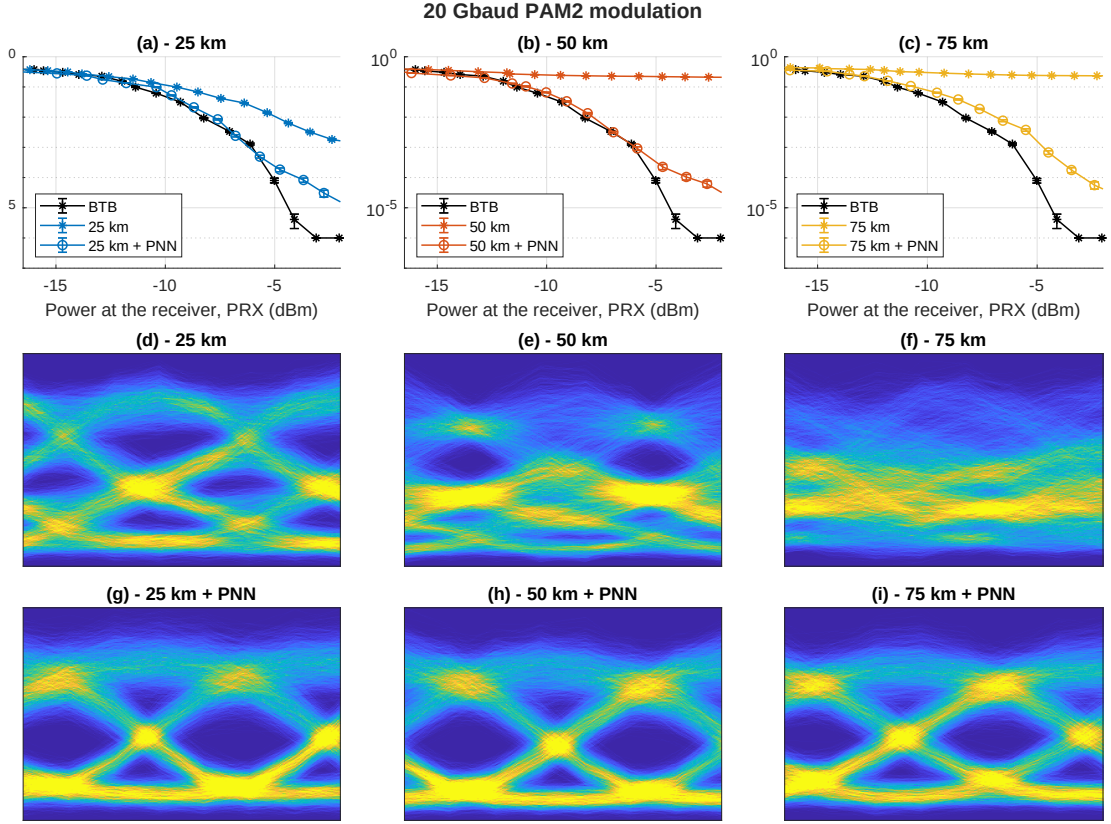


FIGURE 4.6: Results for CD equalization in 20 Gbaud PAM2 propagation. (a-c) BER versus PRX profiles for different link lengths. Black lines set the reference BTB transmission, while colored lines with stars and circles refer to the bare and equalized transmission. (d-i) Eye diagrams at the receiver for bare (d-f) and equalized (g-i) transmission in links of different lengths.

4.3.2 Application to 4-level 10 Gbaud signals

Greater efforts have been dedicated to the application of a more complex modulation format, switching from a 2-level to a 4-level PAM. The studio has been limited to the 10 Gbaud case, which is more adherent to the design specifications of the PNN and avoids problems in the experimental signal generation. Indeed, we verified that the RF amplifiers (see Appendix A.1) connected to the modulator are a considerable source of noise, which drastically affects the optical signal already at the transmitter stage. On a 20 Gbaud PAM4, the noise contribution becomes the main source of distortions, enlarging the optical levels, and preventing reaching $\text{BER} < 10^{-3}$ even in BTB transmission. For these reasons, we avoided symbol rates larger than 10 Gbaud.

4.3.2.1 Investigation on the optimal training conditions

A first exploratory phase has been devoted to finding optimal training conditions in terms of the number of parameters and the loss function definition. Starting from

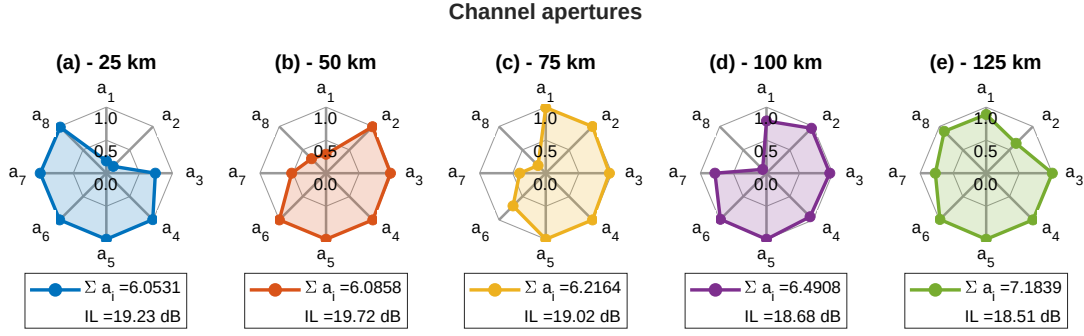


FIGURE 4.7: Optimal amplitude weights a_i with $i = 1, \dots, 8$ obtained after the training of the PNN in *Full* configuration (7 amplitude and 7 phase weights) for channel equalization at different link lengths. The boxes below each diagram report the sum of the aperture levels $\sum_{i=1}^8 a_i$ for that transmission scenario and the related insertion losses (IL).

the most general case and adopting $\mathcal{L}_1$ as the loss function, all 14 available currents (7 PSs, 7 MZIs) have been used as training parameters (*Full* configuration). Figure 4.7 presents the optimal amplitude weights a_i , the sum of the apertures $\sum_{i=1}^8 a_i$, and the measured insertion loss for each transmission length between 25 km and 125 km. The graphs highlight that, for $L \geq 50$ km, the number of open channels ($a_i \neq 0$) increases with the fiber length, and the insertion losses progressively decrease. Indeed, the CD-induced temporal broadening of the transmitted symbols increases with the propagation distance; thus, the PNN needs more channels to recombine pieces of optical information that are spread over a larger time window. Training results for $L = 25$ km break this trend. Considering $\sum_{i=1}^8 a_i$ as a measure of the overall channel opening, we note that the sum results in a value higher than the corresponding one for $L = 50$ km. Since for short propagation distances the impact of cumulated CD is limited, a small number of open channels is expected for the compensation. However, limiting the number of open channels increases the insertion loss, with a consequent degradation of the SNR at the receiver. Therefore, in the presence of reduced CD, the PNN favors an a_i configuration that maintains a high SNR rather than using the minimum amount of open channels needed for the compensation.

In a second set of measurements, the tunable parameters have been restricted to the phasors only (*PO* configuration), while setting the amplitude weights to the open state fixed for every channel ($a_i = 1$ for $i = 1, \dots, N$). Figure 4.8 summarizes the comparison between the different training configurations with the two tested loss functions $\mathcal{L}_1$ and $\mathcal{L}_2$. In particular, Figure 4.8(a) shows the BER reached after a PSO run for different fiber lengths. Each colored curve is associated with a particular combination of the loss function and the number of parameters. The points within each curve are obtained as the average of 10 measurements with the optimal current configuration provided by the training. Since CD acts by applying a phase mask to the signal (as seen with Equation 2.19), we found an overall better performance achieved with *PO* configuration. An

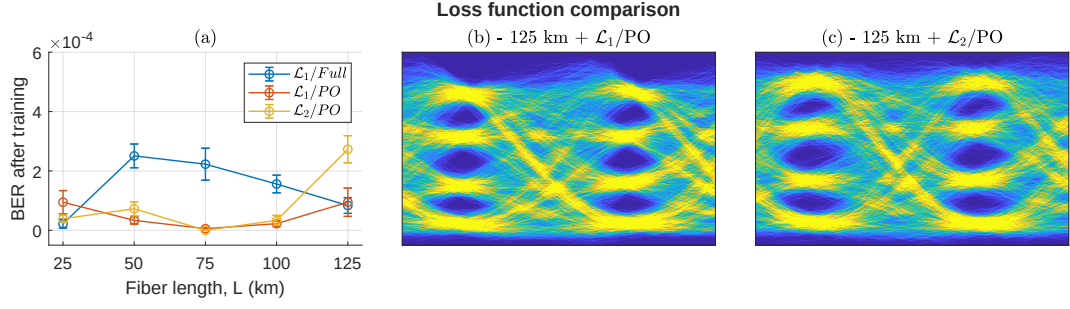


FIGURE 4.8: Comparison between different training strategies. (a) BER values measured after the training performed for equalization of different link lengths. Different colored curves refer to different training strategies, obtained as different combinations of the loss function ($\mathcal{L}_1$ or $\mathcal{L}_2$) and the number of trained weights (*Full* or *PO*). Each BER value results from the average over 10 measurements, with the error bars evaluated as the standard error (Bootstrap method could not be applied due to lack of data). (b-c) Eye diagrams at the receiver after the training performed at 125 km in PO configuration with (b) $\mathcal{L}_1$ and (c) $\mathcal{L}_2$ respectively.

exception is made with $\mathcal{L}_2/PO$ in the $L = 125$ km case, which, however, is reasonably attributed to a suboptimal descent. The minimization of $\mathcal{L}_1/PO$ and $\mathcal{L}_2/PO$ for $L = 125$ km produces the equalized eye diagrams reported in Figure 4.8(b-c). It is possible to appreciate how the use of $\mathcal{L}_2/PO$ allows, in general, to retrieve the symmetry of the eye, while some ripples are present in the $\mathcal{L}_1/PO$ case. Therefore, the 2-sample version has been used for all the subsequent training processes across this work. In the case of PAM2 modulation treated in the previous section, the introduction of $\mathcal{L}_2/PO$ would not have brought significant improvements to the BER reduction, as, in fact, an eye diagram asymmetry negatively impacts transmission quality more in a multi-level format than in a 2-level one.

4.3.2.2 Equalization results

The BER versus PRX profiles obtained for $L = 25$ km, $L = 75$ km, and $L = 125$ km are presented in Figure 4.9(a-c), together with the corresponding unequalized (d-f) and equalized (g-i) eye diagrams. Again, the ISI generated by CD raises the unequalized BER curves with respect to the BTB condition and causes a progressive closure in the corresponding eye diagrams (Figure 4.9(d-f)) as L increases. The trained PNN lowers the BER curves close to the BTB performance, reaching the HD pre-FEC threshold of 2×10^{-3} (see Table 2.4) in correspondence of $PRX \approx -3$ dBm, as well as the opening of the eye diagrams (Figure 4.9(g-i)).

PSO as a minimization algorithm has been compared with other optimization methods. Strategy ST_1 consists of a gradient descent performed by the Adam algorithm with the starting point chosen as the configuration that provides the lowest loss among 20 randomly generated ones. Similarly, the strategy ST_2 again exploits the Adam algorithm,

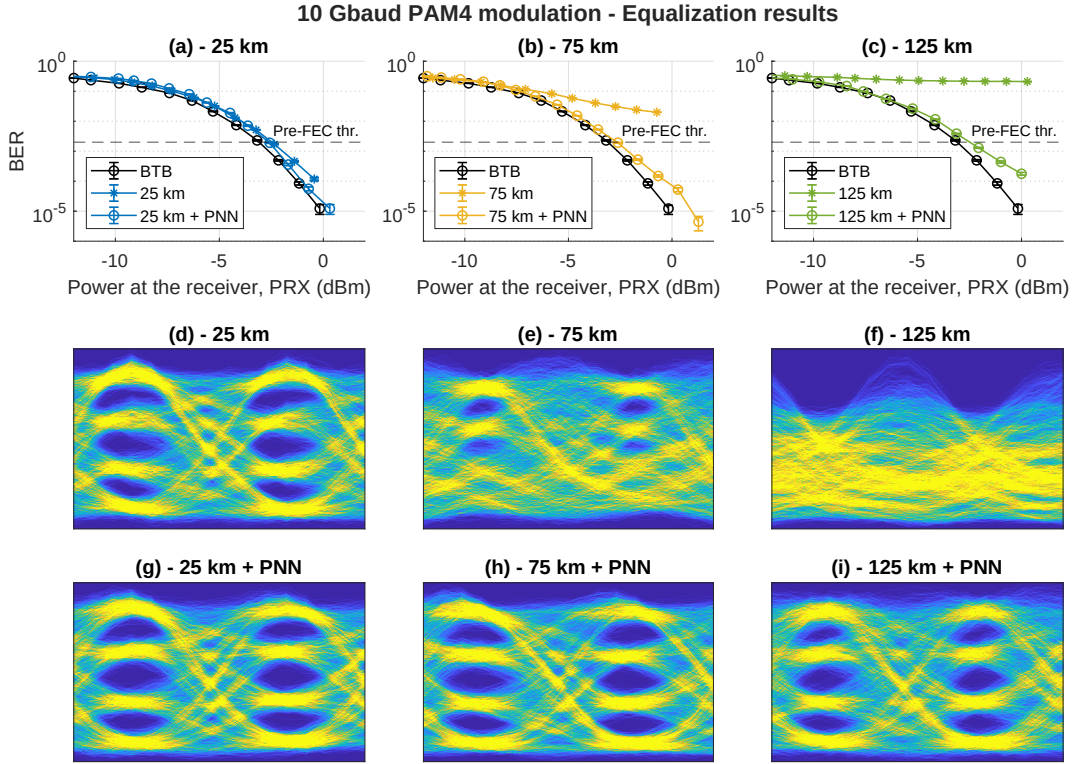


FIGURE 4.9: (a-c) BER versus PRX profiles and corresponding (d-f) unequalized and (g-h) equalized eye diagrams for propagation in different optical links. In (a-c), the unequalized (colored stars) and equalized (colored circles) profiles are compared with the Back-to-Back performance (BTB, black circles).

but the initial condition is fixed to the optimal configuration (found with PSO) that optimizes the BTB transmission. Finally, strategy ST_3 represents the standard training configuration that produced the results of Figure 4.9, namely the parameters are optimized via PSO starting from a random initial condition. The optimization ends when the loss fails to improve by more than 0.02 for 10 consecutive iterations in ST_1 and ST_2 , and for 15 consecutive iterations in ST_3 . The final loss value after each training is evaluated as the average of 10 measures acquired with the optimal weights. Each strategy has been executed 15 times to optimize the equalization of 125 km of fiber with the intent to assess its performance in terms of speed, precision, and repeatability. Figure 4.10(a-c) reports the best loss at every epoch as a function of time for the three strategies, while Figure 4.10(d) compares the final loss values against the number of loss function evaluations spent in the minimization process. Strategy ST_1 fails to reach an optimal solution in most cases, while strategy ST_2 is more consistent in reaching low loss values. The PSO-based strategy ST_3 achieves the lowest loss values on average, but the number of iterations needed for the minimization is increased by one order of magnitude compared to ST_1 and ST_2 . The Adam algorithm is thus faster than the PSO, but

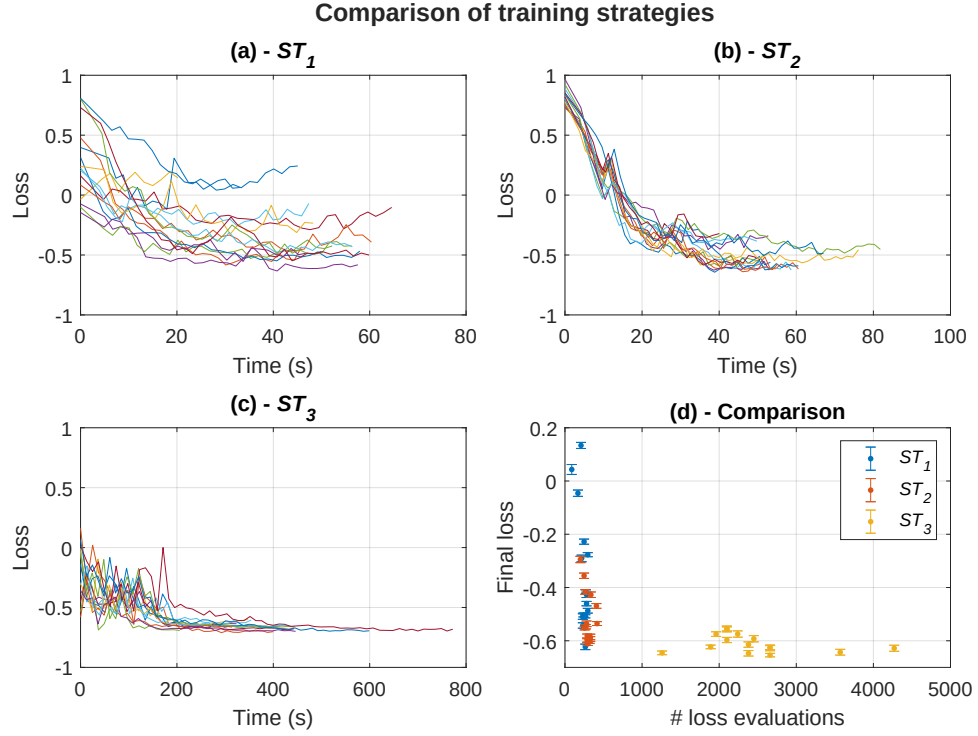


FIGURE 4.10: Comparison of different training strategies to minimize $\mathcal{L}_2$ for signal equalization after propagation in 125 km fiber. (a-c) Loss value as a function of epoch time during the training performed with strategy (a) ST_1 (Adam with random starting point), (b) strategy ST_2 (Adam with fixed starting point), and (c) strategy ST_3 (PSO). (d) Final loss function value against the number of evaluations. Each value and its error bars are obtained, respectively, as the average and standard error over 10 measurements in the optimal configuration found at the end of each training.

the reliability strongly depends on a proper choice of the starting point. These experimental results prove that a favorable starting configuration for an equalization problem is the one that minimizes the loss function at BTB, namely the one that makes the PNN as transparent as possible to the signal. PSO guarantees convergence to a near-optimal configuration at the cost of increased complexity.

The performance of the PNN has been compared against a commercial Tunable Dispersion Compensator (TDC) based on fiber-Bragg gratings¹. The device features an insertion loss of 5 dB and a dispersion range of ± 900 ps/nm, corresponding to a complete equalization of CD accumulated in a 50 km propagation. In the experimental setup, the PNN was replaced by the TDC, producing the results of Figure 4.11, where the BER curves obtained at 50 km and 125 km using the two equalization methods are compared. The two devices are equally effective in retrieving the signal after 50 km of fiber, while the PNN demonstrates a larger dispersion range with good performance up

¹TeraXion ClearSpectrum™ Model: TDCMB-C000-J.

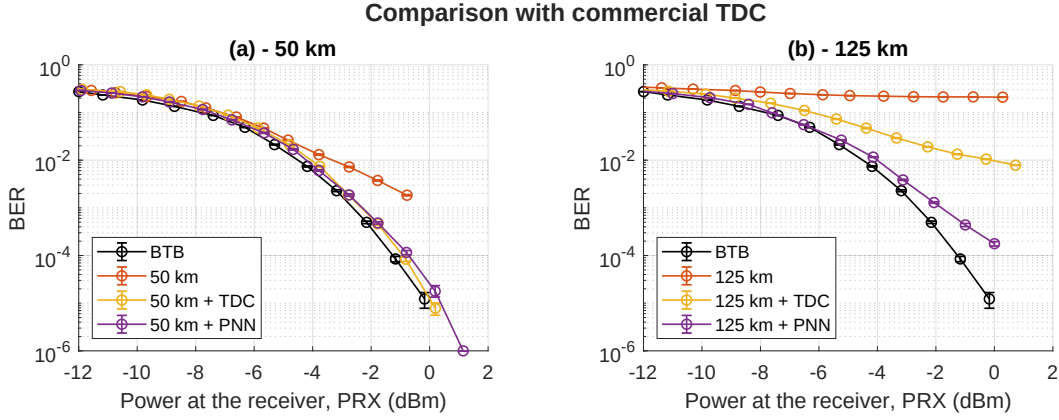


FIGURE 4.11: BER versus PRX profiles for equalized and unequalized transmission in (a) 50 km and (b) 125 km. The profiles refer to transmission in BTB (black), bare fiber (orange), and compensated with TDC (yellow) and PNN (purple).

to 125 km.

A special experimental session has been devoted to measuring the CD-induced penalty of our transmission line, as defined in Equation 2.24. The operation is performed by modulating the intensity of the laser with a cosine of frequency ω , which in our case ranged from 50 MHz to 15 GHz, taking steps of 50 MHz. Figure 4.12 shows the corresponding results of the equalized (yellow) and unequalized (orange) 25 km, 100 km, and 125 km long optical link. The reported parameter $\omega_m = \sqrt{\pi/|\beta_2|L}$ (dashed line) corresponds to the position of the first notch in the unequalized profile predicted by Equation 2.25. The CD-induced penalty at BTB (blue) is reported for comparison, as well as the same quantity measured after removing the fiber span from the link and keeping just the PNN (violet). Note that the yellow curves describe the frequency response of the whole system formed by the PNN and the fiber span, and they do not result from the summation of the red and purple ones. In the 25 km case, bare fiber propagation does not significantly alter the power penalty within the bandwidth of the modulation. In this case, the PNN is useless, and actually, its action moves the first notch of the profile around 10 GHz, worsening the transmission. On the other hand, at 100 km and 125 km, the differences between BTB and bare fiber profiles are significant within the bandwidth of the transmitted signal. In these cases, the PNN reduces the penalty by shifting the first notch back to higher frequency values. One can observe that when the fiber span is removed from the link, keeping just the trained PNN, the first minimum of the CD-induced penalty is approximately where it would be with the fiber in place and without the PNN. This observation suggests that the frequency response $H(\omega)$ of the trained PNN approximates the inverse of the optical fiber frequency response (see Equation 2.18) within the bandwidth of the signal so that

$$H(\omega)H_{of}(\omega) = 1. \quad (4.8)$$

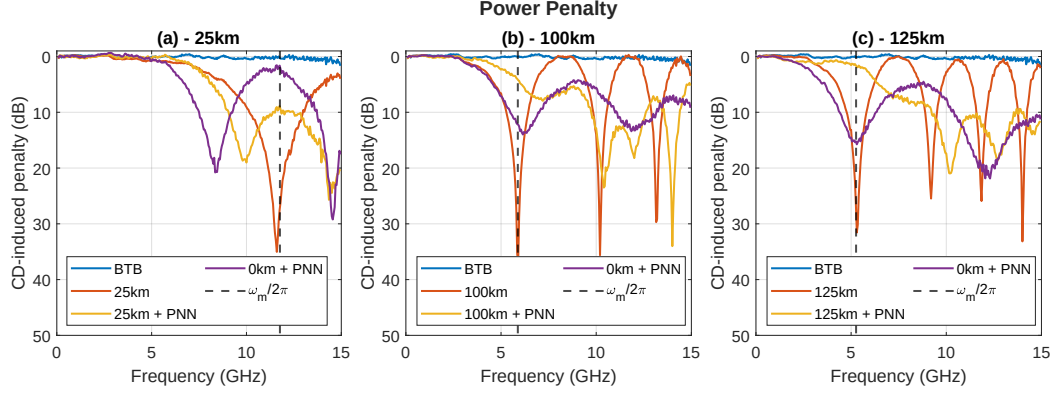


FIGURE 4.12: CD-induced penalty of the optical link for (a) 25 km, (b) 100 km, and (c) 125 km propagation. The BTB profile (blue) is reported for a comparison with the unequalized (orange) and equalized (yellow) profiles. The frequency response of the trained PNN is reported (violet) along with the ω_m parameter (dashed vertical line) for each configuration.

Finally, to test the applicability of our PNN to a DWDM link, we tested its equalization capabilities by varying the laser frequency at the transmitter. The experiment was performed with 125 km of optical fiber using the trained configuration of the PNN at $f_0 = 193$ THz. The weights of the PNN were left unaltered throughout the frequency sweep, which covers the interval $f_0 \pm 90$ GHz with steps of 1 GHz. The blue line in Figure 4.13(a) shows how the loss function $\mathcal{L}_2$ changes with the transmitted frequency. Each point and its error bar are obtained, respectively, as the average and the standard error of 10 experiments. It can be seen that as the signal frequency changes by a few GHz from the training frequency f_0 , the loss increases, namely, the PNN becomes ineffective at compensating CD. This is explained by the fact that the phases accumulated by the signal along the spirals change with the optical frequency. Since the phase shifts induced by the micro-heaters remain the same, the optimal interference condition for the equalization is lost. The same observation is made by training the PNN at 193.04 THz (orange line) and 192.96 THz (yellow line). The Free Spectral Range (FSR) $\Delta\lambda$ of the PNN is given by

$$\Delta\lambda = \frac{c}{n_g L_C} \approx 40 \text{ GHz},$$

where $n_g = 4.2$ is the group index of the PNN Silicon waveguides, and L_C is the length of the unitary delay spiral. Therefore, we expected that the optimal weights found by training the PNN at 193 THz would also work at 192.96 THz and 193.04 THz. However, this is not the case. A plausible reason lies in fabrication imperfections: the periodicity in frequency is lost if the delays are not multiples of the same unit as a result of variations in the spiral lengths or, more plausibly, in the group velocity. The electron beam lithography technique used to trim the waveguide induces local changes in the waveguide width (Line Width Roughness), which are perceived as variations in the

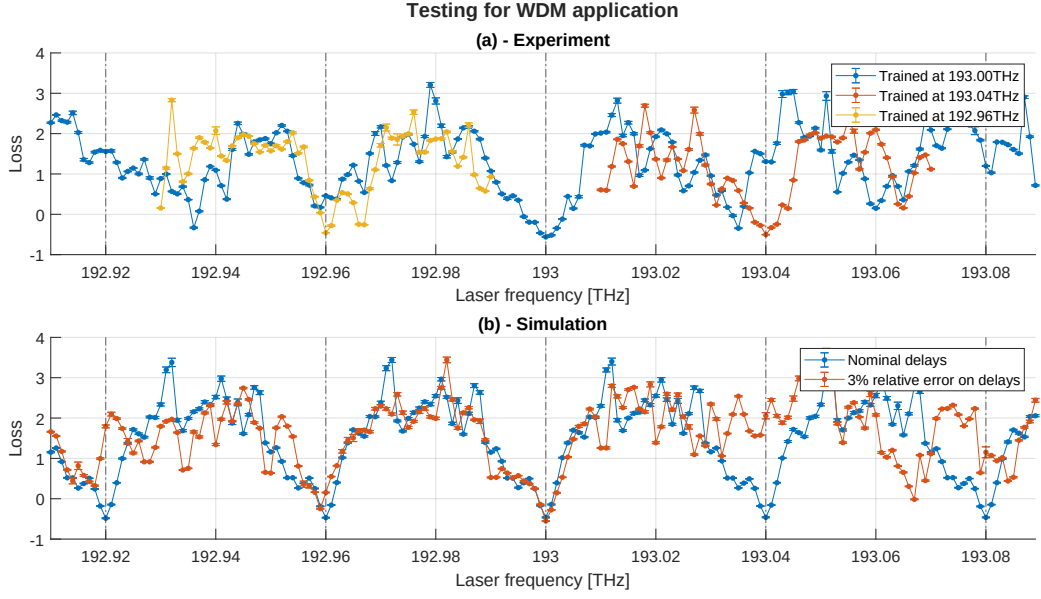


FIGURE 4.13: Two-sample separation loss function $\mathcal{L}_2$ measured for different optical carrier frequencies. (a) Experiment with the PNN trained at 193 THz (blue), 193.04 THz (orange), and 192.96 THz (yellow). Each point and its error bar are obtained, respectively, as the average and standard error of 10 repeated measures. (b) Simulation with the PNN trained at 193 THz, assuming the nominal spiral lengths (blue) or a 3% relative error on those values.

average group index at different frequencies. The impact of delay uncertainty is further confirmed through numerical simulations performed in the environment described in Appendix B. In the results of Figure 4.13(b), the periodicity is well visible when spirals with nominal values are simulated (blue line), while it disappears when 3% random errors in delays are considered.

4.3.2.3 Physical interpretation of optimal weights

As anticipated in Equation 2.38, the impulse response of an ideal device that counteracts CD is

$$h_{cd}(t) = e^{j\frac{1}{2\beta_2 L}t^2}. \quad (4.9)$$

and the output signal $y(t)$ from such a device is obtained as the convolution of the input signal $x(t)$ with $h_{cd}(t)$:

$$y(t) = \int_{-\infty}^{\infty} d\tau x(t - \tau) e^{j\frac{1}{2\beta_2 L}\tau^2}. \quad (4.10)$$

Electronic FIR filters implement this equation in a discrete form in the digital domain, as presented in Section 2.3.3.1, with all the drawbacks of DSP. On the other hand, our PNN device moves the computation in the optical domain, still performing a discrete convolution of its input over a finite window of time. Comparing the TDCP response of Equation 3.12 with the continuous convolution of Equation 4.10, it emerges that perfect

CD compensation can be achieved by a delayed complex perceptron featuring infinite taps ($N \rightarrow \infty$), an infinitesimal delay unit ($\Delta t \rightarrow 0$), and lossless channels ($k_i = 1$, see Equation 3.12). Furthermore, all MZIs are to be kept open ($a_i = 1$), and the PSs must induce a relative phase equal to

$$\phi_i = \frac{1}{2\beta_2 L} (i - N/2)^2 \Delta t^2 \quad (4.11)$$

referred to the central channel ($i = N/2$). Real device implementations have a limited tap number and a finite delay unit, making them non-ideal. Therefore, it is reasonable to expect the optimal phases provided by a training procedure to depart from those given by Equation 4.11. A comparison between the ideal and experimental values is carried out as follows.

Each training procedure returns a set of optimal current values I_i , which can be translated into a set of phase shifts according to

$$\phi_i = \alpha_i I_i^{2\gamma_i} + \beta_i, \quad (4.12)$$

with $i = 1, \dots, 8$ referring to the various taps. These phase values appear in Equation 3.12 and determine the optical response of the PNN. The coefficients α_i , β_i , and γ_i are experimentally characterized for every micro-heater and reported in Appendix A.3. Notice that the exponent γ_i is expected to be close to 1, since in the thermo-optic effect the induced phase shift is proportional to the temperature variation in the waveguides, and thus to the heat generated by the resistors. We deliberately left γ_i as a fit parameter to include possible deviations from the predicted behavior. The phases ϕ_i for each channel are measured relative to the channel with no delay, chosen as reference.

Figure 4.14(a) shows the so-obtained phases of the PNN trained to compensate for different lengths of optical fiber. Figure 4.14(b) focuses on the 125 km case, showing the outcomes of 15 different minimization procedures. The phases measured in a particular trial are adjusted by $\pm 2\pi$, and the average value is subtracted to superimpose the curves obtained in different repetitions. This is feasible because the response of the PNN is invariant under a constant phase shift over all channels. It is shown that the trend of the phases resulting from a repeated minimization is not fully compatible with Equation 4.11, especially in the first and last channels. Many factors can contribute to these discrepancies, first of all, the variability in the training outcomes. This is a consequence of the statistical fluctuations in the loss evaluation, as well as the intrinsic mechanisms driving the PSO algorithm, which involve randomness in the initialization and evolution of the swarm. In addition to this, the possibility of ending up in local minima, and thus in phase configurations that do not match with Equation 4.11, cannot be completely ruled out. Let us also stress that the configuration described by Equation 4.11 was derived under the assumption of a PNN with a large number of channels and small Δt , while the PNN under test does not satisfy these assumptions. Equation 4.11

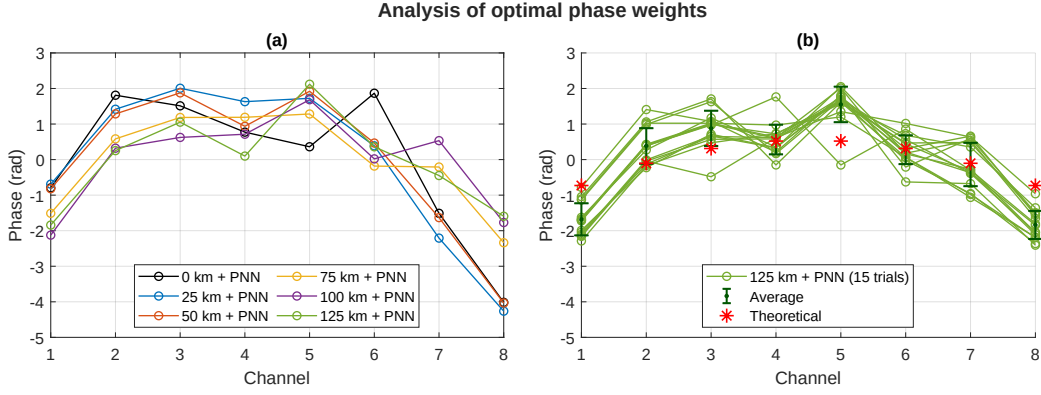


FIGURE 4.14: (a) Optimal phases resulting from PSO for different transmission distances. (b) Optimal phases resulting from 15 different training for $L = 125$ km (using PSO) compared with the optimal phases for a perceptron with infinite channels and an infinitesimal delay unit (see Equation 4.11). Error bars on average values represent the standard deviation of the underlying points.

also neglects the insertion loss of the PNN, which, on the other hand, strongly influences the training outcome. Indeed, in a real implementation, a larger insertion loss of the device implies a smaller SNR at the receiver, which in turn translates into a narrower separation of the optical levels and a larger value of $\mathcal{L}_2$. Therefore, when considering a non-ideal device, the actual configuration that minimizes the loss function consists of a trade-off between the compensation of CD and the reduction of the insertion loss. These factors contribute to determining the training outcome, possibly inducing a departure from the ideal behavior.

Further comparisons between an ideal and a non-ideal device are performed on a simulative level. The benchmark performance is set with a simulated PNN that matches the real implementation, with $N = 8$ taps and a delay unit $\Delta t = 25$ ps. This is trained to equalize CD after 125 km propagation, saving the eye diagram at the receiver and BER versus PRX profile. The results have been compared with those produced respectively by the configurations $N = 8$, $\Delta t = 25$ ps, and $N = 512$, $\Delta t = 0.7812$ ps (corresponding to $1/128$ of the original symbol width) but applied with the ideal phases derived from Equation 4.11. In particular, the last configuration adheres more to the ideal specifications under which Equation 4.11 was derived. The comparison of the BER versus PRX profiles for the three cases and the corresponding eye diagrams is reported in Figure 4.15. As expected, the lowest BER profile is obtained via the training procedure, namely when the phase weights are optimally adapted to the selected transmission scenario, resulting in the opened eye diagram of panel (b). The same PNN layout applied with the ideal phases does not perform similarly, bringing the BER profile close to the unequalized case and resulting in the degraded eye diagram of panel (c). Applying the ideal phases brings an advantage as soon as the PNN layout adheres to the ideal case ($N \rightarrow \infty$, $\Delta t \rightarrow 0$), for which an improvement in the eye diagram aperture is observed

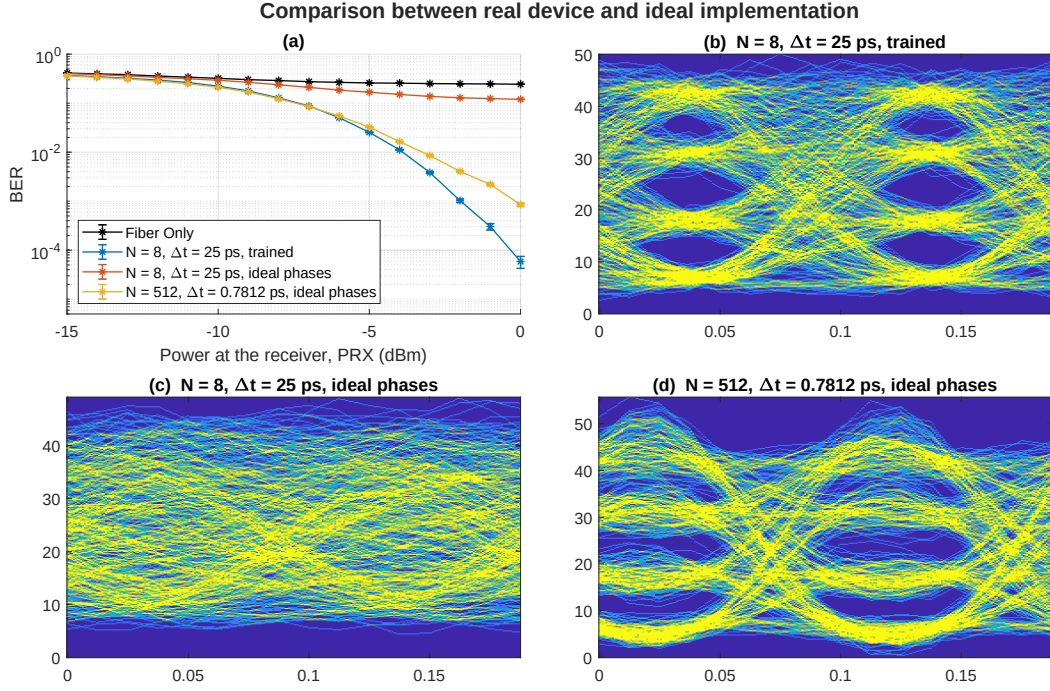


FIGURE 4.15: Simulation results for $L = 125$ km and different PNN configurations in terms of number of taps N , unitary delay Δt , and applied phase weights. (a) BER versus PRX profiles and (b-d) eye diagrams at the receiver for different PNN layouts and applied phases specified (a) in the legend and (b-d) in the titles, respectively. In (a), the error bars correspond to the standard error of the average.

in panel (d). Overall, this strengthens our hypothesis, according to which an ideal implementation requires the ideal phase predicted by Equation 4.11, while in a real-case scenario, the best option is always provided by a training procedure.

4.4 Application to OFDM

After the application to the standard PAM formats, the PNN has been tested for CD equalization in optical signals encoded with Orthogonal Frequency Division Multiplexing (OFDM). This is part of a wide class of multi-carrier modulation techniques and is very popular in wireless communications, where it finds applications in digital television and audio broadcasting, DSL internet access, power line networks, and 4G/5G mobile communications [181]. In OFDM, a bit stream is encoded in closely spaced carriers orthogonal to each other, which then undergo an inverse Fourier transform to produce a time sequence that propagates to the receiver. Here, the acquired signal is demodulated via the Fourier transform, recovering the information originally sent. So far, the OFDM modulation has not emerged as a suitable format in fiber communications because, among many reasons, of its complexity in the modulation/demodulation process [182]. Indeed, DSP implemented at the scope requires severe computational

power, which is even enhanced by the equalization needed for impairments induced by CD [183]. In this section, we present the application of the PNN device for all-optical CD equalization of a OFDM signal, proving the versatility of the proposed approach to more complex formats. Furthermore, the aim is to enable OFDM as a promising candidate for the next generation of optical networks, facilitating the wireless-to-optical compatibility and vice versa. This section partially reflects the content of [184], which results from a fruitful collaboration with the LabTel members from the Federal University of Espirito Santo in Brazil.

4.4.1 Optical OFDM basics

Figure 4.16 presents the optical OFDM transmission line used in the context of these experiments. In the OFDM transmitter (panel (a)), an input binary sequence is parallelized into N_c channels and then mapped into QAM symbols. The multiplexing is performed via the Inverse Fast Fourier Transform (IFFT), which assigns each point in the constellation to a specific sub-carrier oscillating at a multiple of a fundamental frequency f_c up to $N_c \times f_c$, as illustrated in panel (b). The Hermitian Symmetry [185], applied to the constellation points before the multiplexing, ensures obtaining a real-valued OFDM symbol, allowing us to operate still in the IMDD workspace. The Hermitian Symmetry, in fact, doubles the number of sub-carriers, introducing frequency components up to $-N_c \times f_c$ in the resulting signal, so that its total bandwidth results in

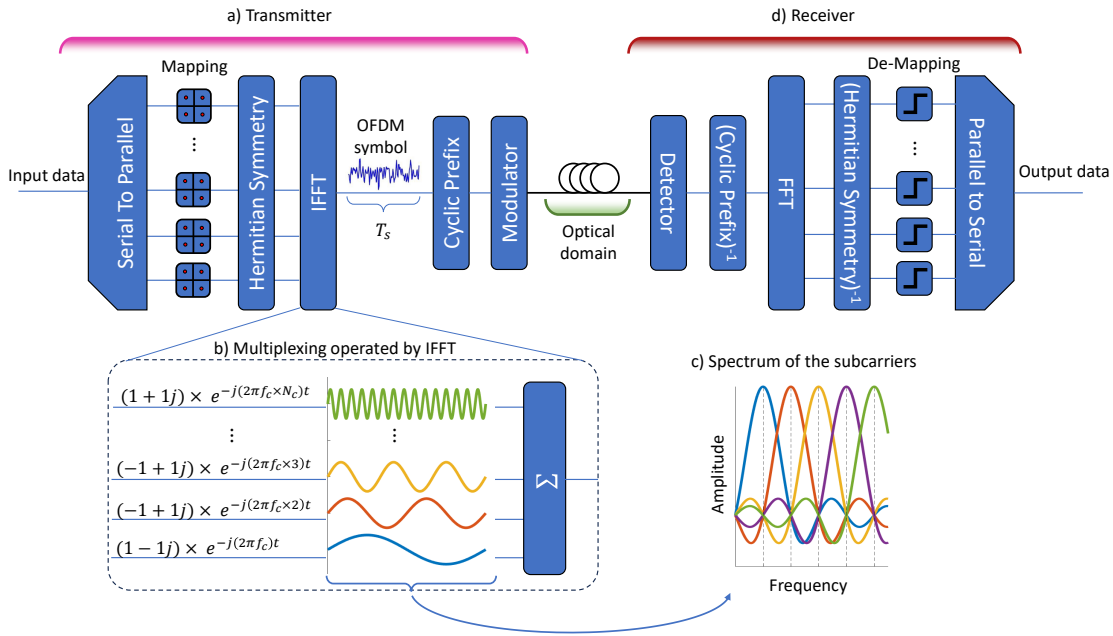


FIGURE 4.16: Block diagram of the optical transmission line implementing OFDM for IMDD signal modulation. (a) OFDM modulation scheme with (b) a detail on the multiplexing operated by the IFFT and (c) the spectrum of the optical carriers. (d) OFDM receiver, implementing the inverse operations of the transmitter.

$2 \times N_c \times f_c$. The value of f_c is linked to the time duration T_s of OFDM symbol by $f_c = 1/T_s$, while the upper limit on N_c is set by the available bandwidth in the experimental setup. Having the subcarriers to be integer multiples of the inverse of the OFDM symbol duration ensures their orthogonality, namely, in the spectrum of each sub-carrier the maximum is in correspondence with the null-points of the other spectra, as shown in panel (c). The symbol is then added with the Cyclic Prefix (CP), that is, a fraction of the symbol is copied and appended to the sequence itself. This acts as a guard interval used for distinguishing consecutive symbols and to mitigate ISI, with the drawback of reducing the transmitted information rate. The so-obtained sequence is then provided to an electro-optical modulator that starts the propagation in the fiber.

After the propagation, the optical sequence enters the OFDM receiver (panel (d)), which inverts the operations performed on the signal at the transmitter. After the detection and the CP removal, the FFT performs the de-multiplexing, moving back from the time to the frequency domain. The auxiliary components generated by the Hermitian Symmetry are then removed, leaving us with the QAM symbols populating the constellation at the receiver (possibly distorted by CD). Finally, the demapping and the subsequent serialization produce the outcome data stream.

At the foundation of the advantages provided by OFDM is the possibility of tuning the bandwidth properties of the signal to adapt to the specific needs of a transmission line. This includes performing an adaptive mapping or switching off specific channels so that less information (or no information at all) is sent in sensitive spectral regions. Moreover, the orthogonality ensures no interference between the subcarriers, maximizing the spectral efficiency.

4.4.2 Experimental results

The optical transmission of OFDM signals is performed via the same setup used for PAM formats, up to 125 km. The only difference is in the signal encoding, which does not affect the optical components, since the data processing for the multiplexing and de-multiplexing operations described in Figure 4.16 is performed at the software level. The transmission quality assessment is performed with the same modalities described for PAM modulations, carrying out a training and testing phase for each transmission scenario. Here, the separation loss function is substituted with the Error Vector Magnitude (EVM) [186], which measures the scattering of the constellation points with respect to the associated target. Different transmission scenarios in terms of data rate (10 or 20 Gbps), propagation distance (up to 125 km), and the constellation order (QAM4 or QAM16) have been explored, producing the results presented in Figure 4.17. In the unequalized diagrams, the points' position is modified with respect to their target by the action of CD, which induces in some of them a rotation with respect to the origin or a reduction in the distance from it. This second aspect reflects the frequency fading induced by CD, which has been treated in Section 2.2.1 and leads to the definition of

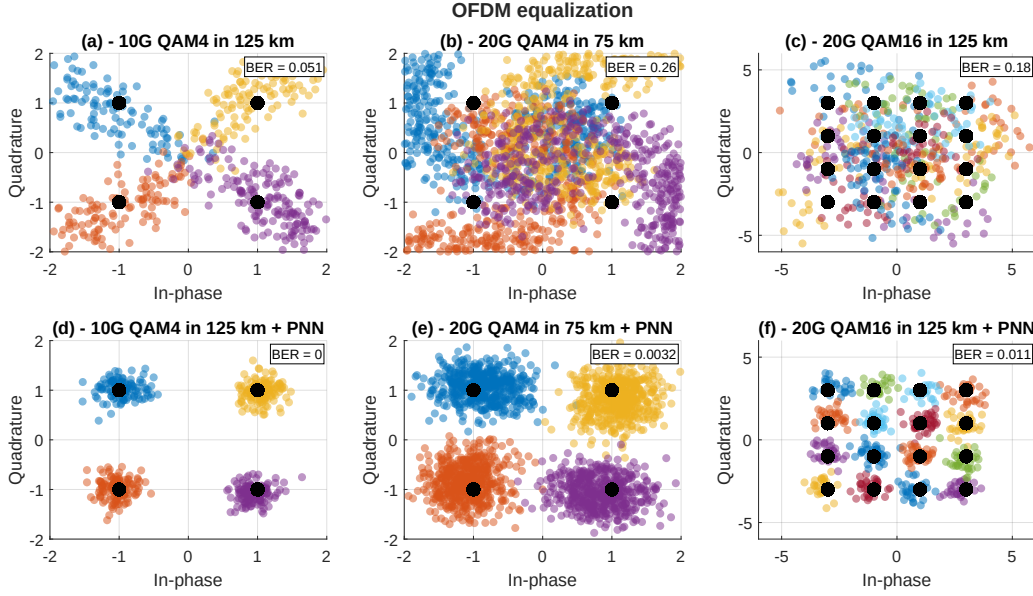


FIGURE 4.17: Constellation diagrams and corresponding measured BER (white insets) for unequalized (a-c) and equalized (d-f) transmission of OFDM optical signals. The different tested scenarios in terms of bit rate and fiber length are indicated in the titles of each panel.

Power Penalty in Equation 2.24. The PNN restores the original constellation shape, with a consequent reduction of the BER.

4.5 Comparison with other photonic approaches

Over the years, several photonic implementations have been proposed for CD equalization. Table 4.1 collects and compares many of them, giving priority to experimental realizations over simulated ones, which are abundant in the literature. The table includes some examples that have already been introduced in Section 3.4.1. Among these, we remind the implementation of optical filters based on a cascade of integrated microring resonators coupled to a single bus waveguide. In [168], a 9-stage architecture was proved to equalize a 100 Gbps Polarization Multiplexed Non-Return-To-Zero (NRZ) signal over 30 km of optical fiber, corresponding to a maximum provided dispersion of -540 ps/nm. In [167], the number of stages is set to 7, allowing the equalization of a 25 Gbps NRZ signal in 40 km propagation, corresponding to -720 ps/nm. The periodic nature of these filters' frequency response allows operating in a WDM scenario, which for the considered cases corresponds to a spacing of 100 GHz between the carriers. In general, the maximum dispersion that cascade filters can compensate grows with the number of stages, which, however, leads to a larger insertion loss.

A different approach to CD equalization consists of photonic Reservoir Computing implementations. In [173], the recovery of the transmitted sequence from the distorted

Ref.	Layout	Platform	Bitrate (GHz)	Encoding	Fiber (km)	Maximum dispersion (ps/nm)	WDM (GHz)
[168]	MR-based 9 stages filter	Si	100	PolMux NRZ	30	−540	100 (sim.)
[167]	MR-based 7 stages filter	Si	25	NRZ	40	−720	100 (sim.)
[173]	O-RC	LASER + OF	25	NRZ	45	−810+NL	-
[169]	O-RC	Si	32	NRZ	25	−450+NL	-
[172]	O-RC + optical readout	Si	28	NRZ	50	−900+NL	-
[165, 166]	MZI-based 10 stages O-FIR	SiN	53	PAM4	40	±500	100 (exp.)
[187]	4-tap O-FIR	Si	10	NRZ	125	±2200	80 (theo.)
[188]	4-tap O-FIR + DSP	SiN/ LNOI	28	QPSK-4	30	−570	28 (theo.)
[189]	MZM-based transm. + DSP	Si	256	PAM4	5	±16	3660 (exp.)
this work, [177, 184]	8-tap O-FIR	Si	20	PAM4, OFDM	125	±2200	40 (theo.)

TABLE 4.1: Comparison between different techniques for CD compensation. O-FIR: Optical Finite Impulse Response filter, MR: Microring, RC: Reservoir Computing, NL: Nonlinearities, SiN: Silicon Nitride, Si: Silicon, PLC: Planar Lightwave Circuit, LNOI: Lithium Niobate On Insulator, OF: Optical Fiber, WDM: Wavelength Division Multiplexing, MZM: Mach-Zehnder Modulator. WDM entrance refers to the spacing between channels that the device in a static configuration manages to compensate for. All the reported measurements were performed at 1550 nm, except for [189], which worked at 1270 nm.

signal is framed as a classification task. It is demonstrated that a reservoir comprising a semiconductor laser with a delayed feedback loop can successfully tackle it. Masking on the input signal is used to increase the dimension of the reservoir, implying the necessity of sampling more points per symbol, which is impractical at high baud rates. Moving to more advantageous and compact alternatives, in [169] (see Section 3.4.1) equalization is achieved with an integrated passive photonic reservoir featuring a swirl topology and 16 nodes. The non-linearity is provided by the detectors probing the nodes, meaning that the final classification (i.e., the linear regression operation) must still be implemented as part of a DSP, even though the reservoir allows for a simpler and more power-efficient realization. Further reduction of the computational effort is

brought by the all-optical readout proposed in [172]: in an 8-node reservoir implementation, the different contributions are extracted from the nodes and sent to a weighted re-combination while still in the optical domain. An external photodiode performs the acquisition, with no need for electronic processing afterward. This implementation allows dealing with both SPM and CD, the latter equalized in a 28 Gbaud NRZ signal up to 50 km.

Other photonic implementations are represented by optical FIR filters, for which it is possible to draw a classification into two groups [190]. In the so-called lattice-form filters, the architecture is represented by a series of identical or similar sections. This is the case, for example, of the already discussed implementation of [166], where an integrated filter consisting of 10 cascaded unbalanced MZIs was proved to equalize a 53.125 Gbps PAM4 signal over 40 km of optical fiber. In general, lattice-form filters allow for lower transmission losses, but setting the desired configuration of weights is not straightforward [191]. On the other hand, in transversal-form filters, the structure is composed of a series of stages (different from each other), each providing for a specific functionality (e.g., splitting, delay, weighting, combining). Our device falls into this category, since the signal goes through a $1 \times N$ splitter and delay stage, then amplitude and phase weights are applied to each branch, and finally a $N \times 1$ combiner produces the output. Transversal-form filters can differ among them in the way the splitter/combiner is implemented, with some architectures using, for example, a tree of y-branches/MMIs [177, 184, 187], while others implement a tree of MZIs [188].

Finally, a remarkable example in peak performance is provided in [189] with a 4×256 Gbps silicon transmitter with on-chip adaptive dispersion compensation. Equalization is achieved up to 5 km while operating at 1271 nm, thus corresponding to a provided dispersion of 16 ps/nm. Notice that large DSP efforts implemented by a Feed Forward Equalizer (FFE) are required at the receiver to cope with residual ISI. Although the proposed approach is based on an integrated photonic device, at the current state, an all-optical solution is still not sufficient to deal with such large bandwidths. The path to a truly efficient and sustainable equalization at high bitrates is still long.

In this context, our device proposes a competitive approach compared to other solutions to the problem of CD equalization. Its compact design offers a wide compensation range with great adaptability to different modulation formats. It allows maintaining the signal processing entirely at the optical level (except for the training phase), with minimized latency and power consumption. Manufacturing imperfections affect its applicability to DWDM and increase the insertion losses. Furthermore, the layout is easily scalable compared to other technologies, as proved by the simulation study for applications to 100G modulation proposed in Section 5.3.

Chapter 5

Self-phase modulation equalization

In this chapter, the focus moves from the problem of CD compensation to the equalization of nonlinear effects, specifically Self- and Cross-Phase Modulation (SPM, XPM). The observation of SPM in a 10 Gbaud signal is made possible thanks to a new experimental setup that allows for multi-span propagation and higher launch power. The same photonic device based on an 8-channel TDCP is tested for equalization, enlarging its applicability to more challenging tasks. The application to nonlinear distortions compensation also implies a change in the interpretation of the PNN device's role. In the previous chapter, it is described as the linear stage of a PNN, whose activation function is represented by the square modulus applied at the end-of-line photodetector. However, in a simplified view, it could also be interpreted as a trainable FIR filter that yields CD equalization due to its linear nature. In that case, the role of the nonlinear activation function remained marginal, since the linear action of the device alone was sufficient to compensate for the ISI. However, the PNN interpretation must be recovered in the context of nonlinear equalization, where the activation function becomes of primary importance. In this situation, the PNN device returns to be the linear stage of a TDCP equipped with the detection process as the activation function. The PNN device alone cannot compensate for SPM, but it is the combined action of the linear and nonlinear stages that counteracts the signal distortions. While applying the FIR filter design to CD equalization seemed straightforward, the application of an all-optical TDCP to nonlinear equalization appears less obvious.

In what follows, a description of the new setup implementation and the explored transmission scenarios is provided. The experimental results are then presented, starting with an extension of the CD equalization reach to 200 km in the linear propagation regime. Subsequently, it is the turn of the novel applications to SPM equalization up to 450 km in the nonlinear propagation regime. Simulated results on a 100G scenario are presented afterward, rescaling the PNN device design, and attempting the equalization of CD, SPM, and XPM. Finally, a comparison of performance with the literature is provided.

5.1 Fiber loop implementation

5.1.1 Experimental setup

Figure 5.1 presents a schematized version of the experimental setup featuring the fiber loop implementation. A more detailed description is provided in Appendix A.2. In the transmission stage, a laser source operating at 1550 nm is modulated as a 10 Gbaud PAM2 signal based on a PRBS. The optical sequence then reaches the propagation stage, where the distortions induced by linear and nonlinear effects are accumulated. The propagation stage emulates a multi-span transmission by recirculating the signal through a fiber loop, whose elements are enclosed in the blue rectangle in Figure 5.1. At every loop cycle, light propagates in a 50 km standard SMF-28 fiber span preceded by an EDFA that amplifies the input power in the fiber up to 9 dBm. A TDC can be used to impose a dispersion $d_{TDC} \in [-900, 0]$ ps nm⁻¹ to remove the accumulated CD. Each propagation is then made of an arbitrary number of loop cycles n_L , namely multiple passages across the elements enclosed in the blue rectangle. This corresponds to a total propagation distance of $L = n_L \times 50$ km, during which the distortions accumulate in the optical signals. Also, due to the multiple amplifications provided by the EDFA, an OSNR degradation occurs, as already discussed in Section 2.1.1. After the loop, the signal proceeds to the PNN device. After the optical processing, a final amplification stage compensates for its losses and sends the signal to the receiver stage (RX). The oscilloscope (OSC) limits the overall bandwidth of the apparatus to 16 GHz and a sampling frequency of 80 GSa s⁻¹.

The PNN equalization capabilities are tested in different transmission scenarios determined by the number of loops $n_L = 0, \dots, 9$, input power in fiber $P_{in} = \{0, 3, 6, 9\}$

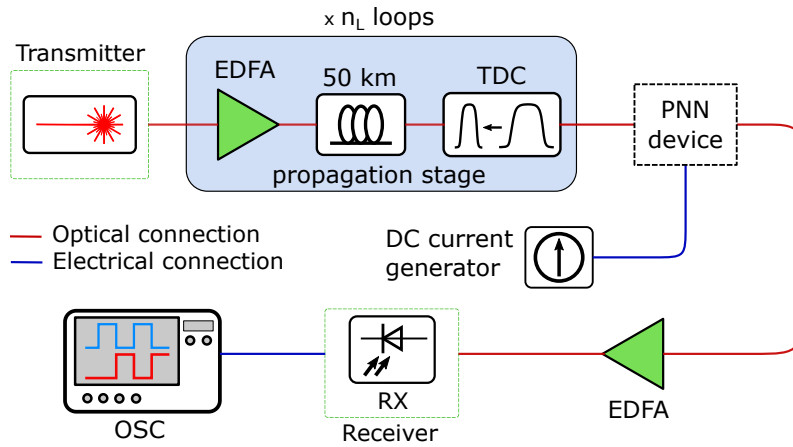


FIGURE 5.1: Experimental setup (simplified) used to test the PNN device for SPM equalization up to 450 km. The propagation occurs as multiple passages across an optical fiber loop (blue rectangle) enclosing an Erbium-Doped Fiber Amplifier (EDFA), a 50 km fiber span, and a Tunable Dispersion Compensator (TDC).

dBm at the first loop, and the dispersion $d_{TDC} = \{-900, 0\}$ ps nm⁻¹ imposed by the TDC. In this work, the reported P_{in} values serve as labels for different propagation regimes, grouping together measurements performed under comparable input-power conditions. Despite small variations in the actual experimental applied power, the measurements within each group are well described by the corresponding P_{in} value.

In each scenario identified by (n_L, P_{in}, d_{TDC}) , the gain inside the loop has been fine-tuned to reach RX with the same reference power set by the value measured after a 0 km propagation. These regulations also take into account the fast transients occurring at the final amplification stage before the receiver [192] (see Appendix A.2). Overall, the gain regulation reflects in small changes of the input power in the fiber at every loop, which is thus not maintained at the value set by P_{in} . A subsequent calibration revealed that, for example, in the worst-case scenario of $P_{in} = 9$ dBm and $n_L = 9$, the input power in the fiber decreases by a total maximum of 4 dB across the 9 loops.

We can distinguish between two different regimes explored during the experiments. In the linear propagation regime, the input power is set to $P_{in} = 0$ dBm, and the PNN is called to compensate for CD ($d_{TDC} = 0$ ps nm⁻¹) or its residuals ($d_{TDC} = -900$ ps nm⁻¹). On the other hand, in the nonlinear propagation regime, $P_{in} > 0$ dBm and $d_{TDC} = -900$ ps nm⁻¹, and the PNN is called to equalize SPM-related distortions. Unfortunately, XPM observation was prevented by setup limitations, including the limited available bandwidth and a minimum achievable detuning of 200 GHz between two available laser sources, corresponding to a too-short walk-off distance [193].

The transmission quality is accessed via a loss function, whose input parameters are represented by the number of loops n_L and the array of currents driving the PNN weights. A loss function call propagates the optical pulses for the assigned number of loops, sets the current to the PNN, and acquires the corresponding trace at the receiver. The outputs are the BER and the result of the 2-sample Separation loss function. The PSO is still chosen as the training algorithm. The unequalized and equalized transmissions are compared via the corresponding BER versus PRX profiles, obtained according to the same modalities described in Section 4.1.1. Again, in the experimental context, all the measurements maintain the same periodic input sequence represented by a 2¹⁰ bit-long PRBS of order 10. The proposed approach has no distinction between the training and testing data sets since the chosen input sequence allows the PNN to experience all the observable distortion conditions generated by adjacent symbols. Following the same reasoning proposed in Section 4.2, the chosen worst-case scenario in the linear propagation regime is represented by $L = 200$ km and $d_{TDC} = 0$ ps nm⁻¹. According to Equation 2.11, this leads to $\Delta T = 184$ ps and a total symbol time width of $1/B + \Delta T = 284$ ps, with B being the baud rate. The generated ISI involves 3 adjacent symbols (the central one and its left and right prime closest). This represents a limit case for the PNN since the total bit width is more significant than the maximum time window of $(N - 1) \times \Delta t = 175$ ps that the PNN can parallelize and process simultaneously. On the other hand, in the nonlinear propagation regime where $d_{TDC} = -900$ ps nm⁻¹, CD acts within every 50 km segment before being equalized. At each loop, the broadening reaches $\Delta T = 46$ ps and is subsequently counteracted by the TDC. Therefore, in both

propagation regimes, the distortions result from the interaction of at most 3 adjacent bauds, making the chosen PRBS input sequence sufficient to let the PNN explore all the possible distortions.

5.2 Experimental results

5.2.1 Linear propagation regime

Figure 5.2 reports the results obtained in the linear propagation regime, featuring $P_{in} = 0$ dBm at every loop. The BER versus PRX profiles (panel (a)) are obtained setting $d_{TDC} = 0$ ps nm⁻¹, so the PNN is called to compensate for CD. The trainings are performed in Phase-Only (PO) configuration (see Section 4.3.2.1), namely by setting the amplitude weights to $a_i = 1$ for $i = 1, \dots, 8$ and optimizing the phases only. The so-obtained results extend those shown in the previous chapter since the propagation distance has increased over 125 km, up to 200 km. The unequalized BER profiles (colored stars) depart from the BTB curve (black stars), with the ISI generated by CD becoming increasingly severe with the propagation distance. The trained PNN lowers the BER profiles, restoring the eye diagram aperture from panel (c) to panel (d). On the other hand, the results reported in panel (b) are obtained by setting $d_{TDC} = -900$ ps nm⁻¹, removing the most contributions of CD at every loop. The departure of the BER versus PRX profile (dark red stars) from the BTB curve (black stars) becomes evident only after long propagation due to CD residual accumulation, which the trained PNN

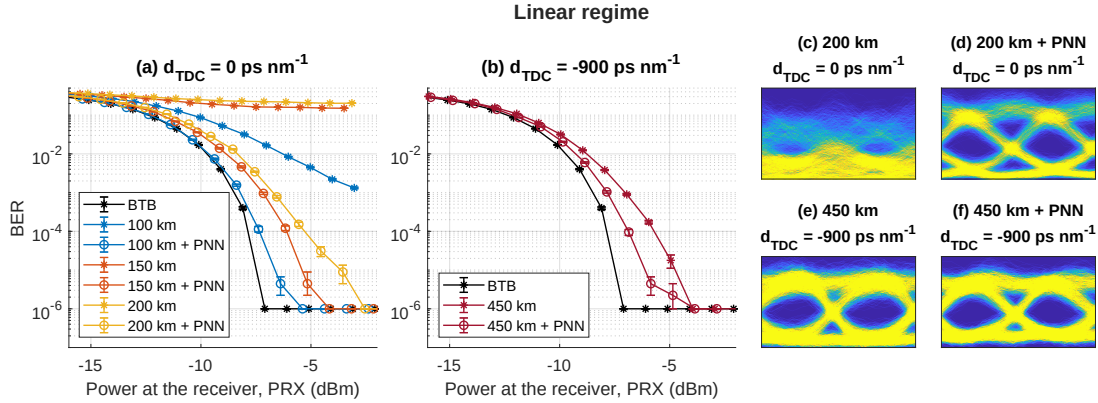


FIGURE 5.2: Experimental results for equalization at 10 Gbaud in linear propagation regime ($P_{in} = 0$ dBm). (a-b) Unequalized (stars) and equalized (circles) BER versus PRX profiles obtained in the context of (a) CD and (b) CD residuals compensation. Different transmission scenarios are specified in the legends. The curve generated for BTB transmission (black stars) sets the reference performance. (c,d) Eye diagrams for the unequalized (c) and equalized (d) transmission in $L = 200$ km and $d_{TDC} = 0$ ps nm⁻¹. (e,f) Eye diagrams for the unequalized (e) and equalized (f) transmission in $L = 450$ km and $d_{TDC} = -900$ ps nm⁻¹.

mitigates (dark red circles). This improvement also appears in the corresponding eye diagrams in panels (e) and (f).

5.2.2 Nonlinear propagation regime

In correspondence with a higher input power in fiber ($P_{in} > 0$ dBm), the PNN is trained to equalize SPM-induced distortions, which have been isolated from CD contributions by setting $d_{TDC} = -900$ ps nm⁻¹. The impact of nonlinear effects on the transmission quality is presented in Figure 5.3, obtained in the worst-case scenario of $P_{in} = 9$ dBm. The accumulation of nonlinear distortions with the propagation distance appears in panel (a) as an increasing gap between the unequalized BER versus PRX profiles (colored stars) and the BTB performance (black stars). This is reflected in panels (b-d) as a progressive tilting of the eye diagram. Again, the PNN beneficial action restores the transmission's quality, lowering the BER profiles (colored circles) closer to the BTB curve and opening the eye diagrams (panels (e-g)).

To quantify the figure of merit brought by the PNN, we introduce the BER reduction factor R , whose derivation is illustrated in Figure 5.4(a). Considering the unequalized BER versus PRX profile obtained in a given transmission scenario, BER_{neq} represents the linearly interpolated value at a reference power at the receiver PRX_{ref} . Analogously, BER_{eq} represents its homologous obtained from the equalized profile generated after the introduction of the PNN. The R -factor results from the ratio (expressed in dB) of the two as $R = BER_{eq}/BER_{neq}$, with lower values representing better equalization performance of the PNN. Notice that the choice of PRX_{ref} value aims to produce informative BER maps with meaningful variations between equalized and unequalized BERs. Taking

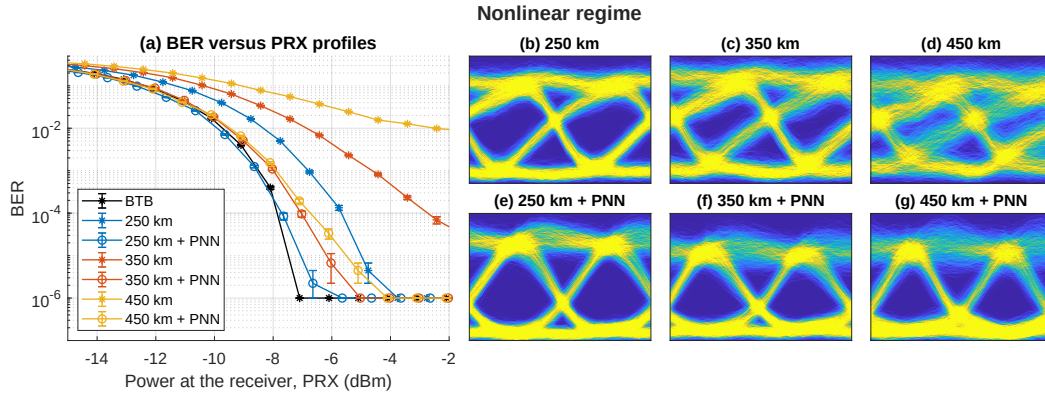


FIGURE 5.3: Experimental results for equalization at 10 Gbaud in nonlinear propagation regime ($P_{in} = 9$ dBm, worst case scenario). (a) BER versus PRX profiles obtained with $d_{TDC} = -900$ ps nm⁻¹ for unequalized (stars) and equalized (circles) transmission. Different colors in the lines indicate different propagation lengths L . The curve generated for BTB transmission (black stars) sets the reference performance. (b-g) Eye diagrams for the unequalized (b-d) and equalized (e-g) transmission in (b,e) $L = 250$ km, (c,f) 350 km, and (d,g) 450 km.

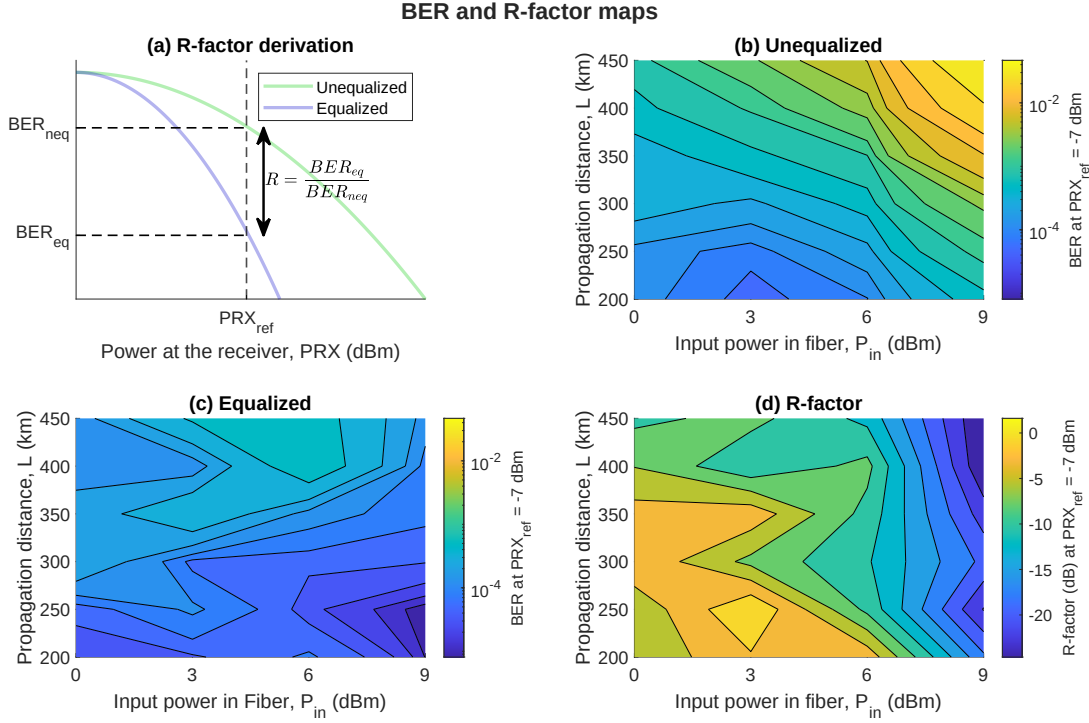


FIGURE 5.4: (a) Illustration of the BER reduction factor (R-factor) derivation at a given reference power PRX_{ref} from two unequalized (green) and equalized (blue) BER versus PRX curves. (b) Unequalized and (c) equalized BER measured at $PRX_{ref} = -7$ dBm mapped as a function of the input power in fiber P_{in} and the total propagation distance L . Null BER points have been replaced with the minimum measurable value. (d) R-factor map at $PRX_{ref} = -7$ dBm as a function of the input power in fiber P_{in} and the propagation distance L .

Figure 5.3(a) as a reference, a too-low PRX_{ref} explores the left-most region of a typical BER versus PRX curve, where noise is the dominant contribution. On the other hand, a too-high PRX_{ref} may result in low BER estimates corresponding to the statistical limit. Thus, the tradeoff to obtain informative maps from a family of curves is found with PRX_{ref} that intercepts the last non-null (i.e., higher than the statistical limit) BER value in the lowest curve.

Figure 5.4(b) reports the BER_{neq} map for $PRX_{ref} = -7$ dBm as a function of P_{in} and the propagation distance L . The BER_{neq} values increase along both axes, and the worst-case scenario is identified by the yellow region in the top-right angle, corresponding to the combination of the highest input power and longest propagation distance. The introduction of the PNN makes the BER_{eq} map of Figure 5.4(c) more uniform, eliminating the presence of the peak in the top-right angle. Overall, the PNN leads to an improved transmission quality, also testified by the R-factor map of Figure 5.4(d). Configurations in which $R > 0$ are negligible since the corresponding BER_{neq} is far below the pre-FEC threshold of 2×10^{-3} .

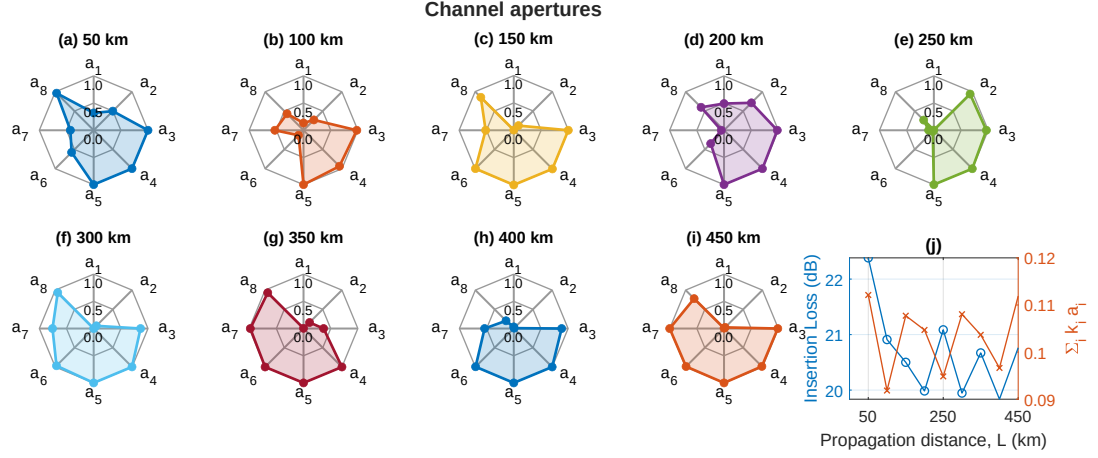


FIGURE 5.5: (a-i) Optimal amplitude weights a_i after PNN training obtained for channel equalization in nonlinear regime ($P_{in} = 9$ dBm) at different propagation distances. (j) Measured insertion loss (left axis) and overall channel apertures obtained as $\sum_{i=1}^8 k_i a_i$ (right axis) for the PNN device while operating (i.e., with active optimal currents) signal equalization in different propagation distances.

5.2.2.1 Analysis of outcome weights

Figure 5.5 presents the optimal amplitude weights (a-i), the corresponding overall channel apertures ($\sum_{i=1}^8 k_i a_i$), and insertion loss (IL) (j) obtained for $P_{in} = 9$ dBm at different propagation distances. We expected to see a trend where, as L increases, the PNN device features more adjacent channels in the open state (i.e., $a_i \simeq 1$), resulting in a partial reduction of the insertion loss. Indeed, longer L allows more SPM-related distortions to accumulate and possibly interact with residual CD. This increases the task's complexity, making it necessary to combine more and more pieces of optical information from different time instants. However, many factors influence the training procedures (noise, polarization rotation during propagation, PNN device misalignment, etc.), making the expected trend less evident from the measured data. Support for our expectations comes from simulations, which reproduce similar experimental conditions as those associated with Figure 5.5. Multiple training sessions for SPM equalization are performed, imposing different constraints on the number n_{op} of adjacent channels set in the open state while the others are closed. It is found that a condition with $n_{op} = 3$ provides equalization for $L = 100$ km, in accordance with Figure 5.5(b). For $L = 450$ km, an increased value of n_{op} is needed to equalize the transmission with the same performance.

Subsequently, we analyze the role of amplitude weights, following the discussion started in Section 4.3.2.1. In the context of CD equalization, setting $a_{1,\dots,8} = 1$ and training the phase weights only (*PO* configuration) is sufficient to counteract the linear distortions. On the contrary, training the amplitude weights too (*Full* configuration) becomes fundamental against nonlinear effects, although this makes the training time longer. Figure 5.6(a-b) shows the eye diagrams and the corresponding BER values after

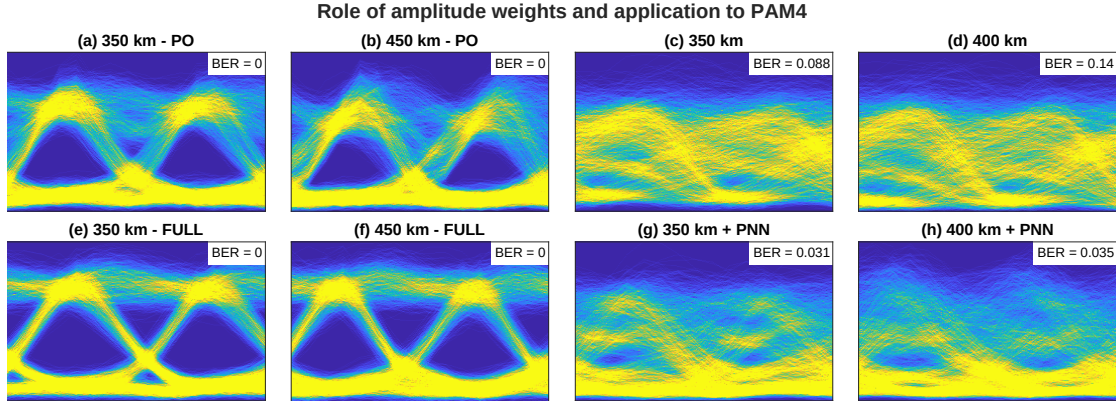


FIGURE 5.6: Eye diagrams at the receiver and corresponding BER (white frames) after signal propagation at $P_{in} = 9$ dBm. (a-b) Equalized eye diagrams after propagation in (a) 350 km and (b) 450 km, with the PNN trained in Phase-Only (PO) configuration. (e-f) Equalized eye diagrams after propagation in (e) 350 km and (f) 450 km, with the PNN trained in Full configuration. (c,d,g,h) Unequalized (c,d) and equalized (g,h) eye diagrams for 10 Gbaud PAM4 modulation in (c,g) 350 km and (d,h) 400 km.

350 km and 450 km propagation equalized with the PNN in *PO* configuration. Despite the BER being null, distortions and ripples appear, causing a broadening of the high level and the transient lines. On the contrary, the equalized eye diagrams obtained in *Full* configuration appear wider and symmetric to the vertical axis, as shown in Figure 5.6(e-f).

Finally, the eye diagrams for unequalized and equalized PAM4 modulation for $P_{in} = 9$ dBm are reported in Figure 5.6(c-d) and (g-h), respectively. Transmission is limited by the high noise level introduced by RF amplifiers already at the transmitter stage, which generates non-null BER even in BTB configuration. However, despite the low SNR, the PNN provides up to an order of magnitude of BER reduction, proving its potential in SPM equalization even with multi-level modulation formats.

5.3 Simulated study for future 100G applications

The implemented PNN device has been designed in terms of the number of taps N and unitary delay Δt for treating 10 Gbaud signals. This constraint is imposed by bandwidth limitations in the experimental setup and makes the current design not applicable in modern optical transmission lines, which are operated at much higher baud rates, reaching up to 100G per lane. Scaling the device layout is thus necessary to make it suitable for commercial applications in a real case scenario. The limitations of the experimental setup have been circumvented by implementing a simulated version of the optical line. This allowed overcoming technical challenges and performing a study on the scalability of the proposed approach to different transmission scenarios, including 100 Gbaud modulation and XPM-induced distortions. A detailed description of the

code and the parameters used in the simulation is presented in Appendix B.

The simulated setup is implemented via MATLAB software, and represents a simplified version of the experimental optical link described in Section 5.1.1 and Appendix A.2. It features two laser sources, Probe and Pump, separated by a detuning $\Delta\nu = 50$ GHz and independently modulated. The so-obtained optical signals are then sent to the loop for multi-span propagation. Inside the loop, the optical propagation through an arbitrary long fiber span is simulated via the SSFM [92], subsequently removing the CD contribution to mimic the action of the TDC when active. The OSNR degradation introduced by the multiple passages through the EDFA is accounted for by applying Equation 2.1, following the approach proposed in [50]. The simulated setup includes a parametric version of the PNN device, with tunable unitary delay Δt and number of taps N . The processed optical signal is then sent to the end-of-line photodetector, whose noise generation is modeled as the sum of thermal and shot noise contributions [178].

The experimental results obtained with 10 Gbaud signals in terms of BER versus PRX profiles and eye diagram apertures served as a benchmark for code calibration. After that, the bandwidth of each instrument (modulator, RX, OSC, ...) is properly rescaled to accommodate 100 Gbaud signals. Each transmission scenario is identified by the number of loops $n_L = 0, \dots, 9$, the probe input power in fiber $P_{pr} = [0, 10]$ dBm, and the pump input power in fiber $P_{pu} = [-30, 10]$ dBm, with $P_{pu} = -30$ dBm standing for the pump being off. Contrary to the experimental context, here the input power in the fiber is restored exactly to P_{pr} and P_{pu} at every loop. Each simulated transmission returns the BER and the 2-sample separation loss function, which is provided to the PSO for PNN training. The loss function receives as input an optical sequence (complex field) to be processed, the associated target sequence, and the PNN weights $a_i = [0, 1]$ and $\phi_i = [0, 2\pi)$. In linear regime exploration, the PNN is trained to compensate for CD. In this scenario, P_{pu} is set to 0 dBm, the Pump is switched off, and the propagation is performed in a single loop ($n_L = 1$) of length up to 6 km, emulating short-reach applications. On the other hand, in the nonlinear regime, the span length is set to 50 km, using the TDC to compensate for CD at every loop. In both the explored regimes, different PNN layouts have been tested, including 8 and 16 taps designs with unitary delay scanning from $T_0/8$ to T_0 , this last corresponding to a full baud width of 10 ps. Other cases are not an object of this study since $N = 4$ would have represented a too-simplistic layout, while $N > 16$ would drastically increase the insertion losses.

The experimental approach proposed with a 10 Gbaud modulation, which always uses the same 2^{10} bit-long PRBS sequence for every measure, no longer applies to higher frequencies. Indeed, for a 100 Gbaud signal propagating in a 50 km link, Equation 2.11 estimates a pulse time-broadening of around 460 ps, corresponding to 46 times the original baud time width. It would require the propagation of an incredibly long sequence to allow the PNN to explore all the possible distortion conditions generated by the interference of 46 symbols in a single measure. The unfeasibility of this solution necessitates reintroducing the distinction between training and testing data sets, whose

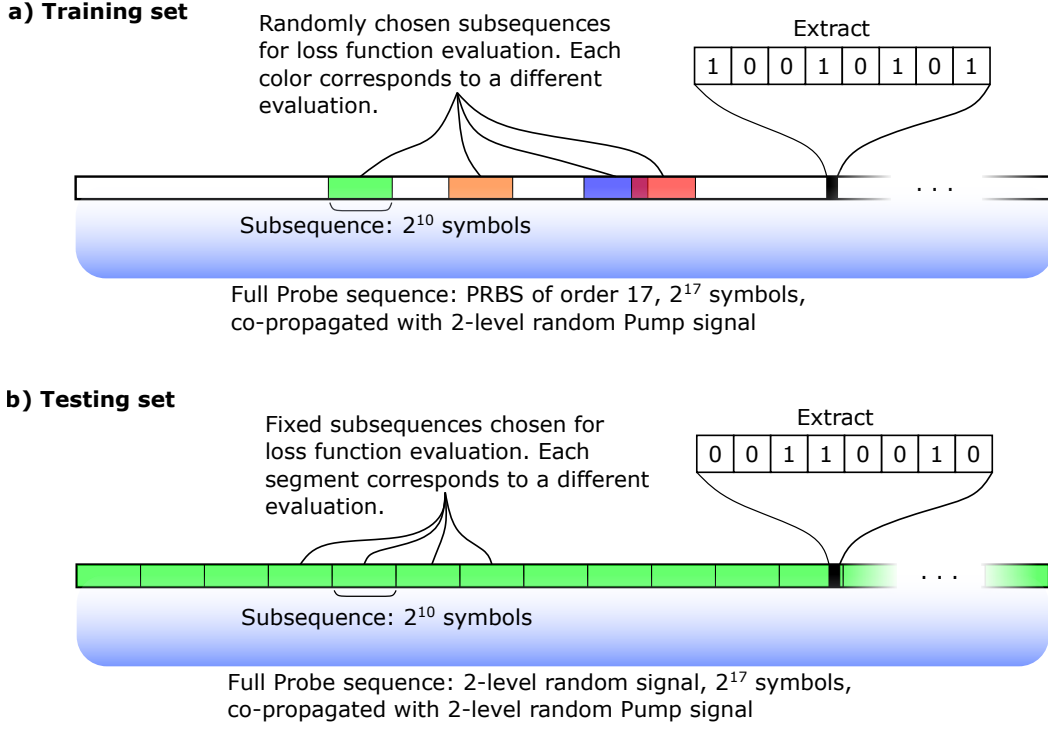


FIGURE 5.7: Illustration of training (a) and testing (b) datasets creation used in simulations at 100 Gbaud. A detailed description is provided in the text.

generation is depicted in Figure 5.7. The training set in panel (a) is based on a 2^{17} symbol-long PRBS sequence of order 17 (enclosed in blue shadow) co-propagated with a 2-level random pump sequence. The same two sequences are propagated in every transmission configuration regarding P_{pr} , P_{pu} , and n_L , obtaining the data set specific to each scenario. Subsequently, each loss function evaluation during the training is operated on a sub-sequence of 2^{10} symbols (colored segments) derived from the main one by randomly choosing a starting point and selecting the subsequent 2^{10} symbols. Similarly, the generation of the testing data set in panel (b) originates from the propagation of the probe and pump signals, this time both encoding a 2-level random sequence. The resulting probe trace is divided into ~ 100 consecutive subsequences of 2^{10} symbols (green segments), each separately used for loss function evaluation. The testing procedure outputs a BER versus PRX profile where each point is obtained as the average loss function evaluation over the mentioned 100 subsequences. During the following results presentation, null BER points have been replaced with 5×10^{-6} , which is lower than the minimum measurable value.

5.3.1 Linear propagation regime

For each combination of number of taps N and unitary delay Δt , the code reproduces the training in PO configurations of the corresponding PNN device, and subsequently

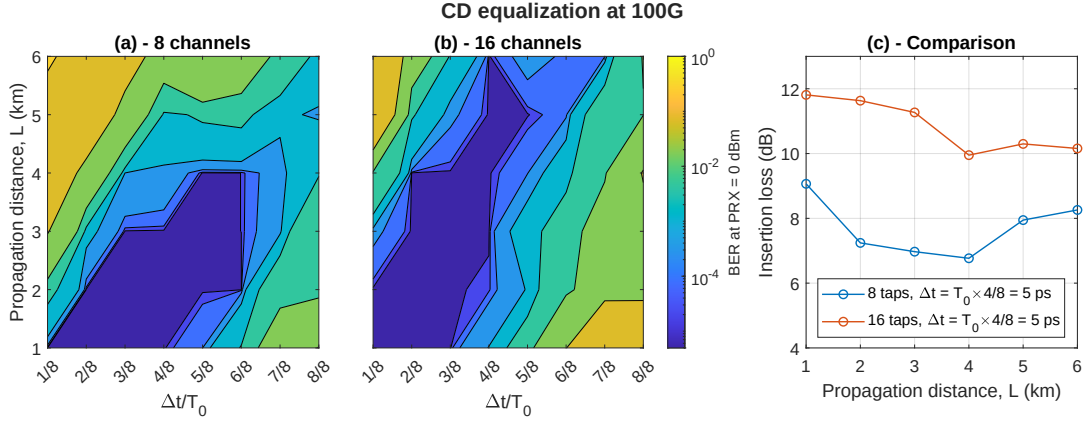


FIGURE 5.8: Simulated results for PNN layout optimization to operate CD equalization on 100 Gbaud signals. The tested PNN layouts feature (a) $N = 8$ and (b) $N = 16$, respectively. Trainings are performed in *PO* configuration. (a-b) Equalized BER mapped as a function of the propagation distance L and unitary delay Δt . This is expressed as the corresponding fraction of the initial baud width $T_0 = 10$ ps. (c) Insertion Loss as a function of the propagation distance measured for the optimized PNN devices in operation. The curves refer to the 8- and 16-channel configurations equipped with $\Delta t = 5$ ps.

the testing phase, after which a final BER estimate measured at a PRX of 0 dBm is returned. The corresponding results populate the maps reported in Figure 5.8(a-b), where the unitary delay Δt is expressed as the corresponding fraction of the total baud width T_0 . In the maps, blue regions correspond to lower BER, suggesting the optimal configurations for CD equalization. For both the 8 and 16 taps layouts, a small unitary delay limits the memory of the PNN device, thus reducing the maximum achievable equalization distance. On the other hand, larger Δt increases the overall memory, but reduces the density of optical samples per baud brought to recombination and, therefore, the quality of symbol reconstruction. The optimal configuration stays between these two extremes. The 16-channel layout equipped with $\Delta t = 5$ ps appears to offer better performance for L between 2 km and 6 km, allowing the extension of the equalization reach compared to the 8-channel counterpart. Indeed, for the same Δt , an increased number of channels leads to larger memory for the device, allowing for dealing with larger pulse broadening induced by CD. Moreover, a PNN with more degrees of freedom can better approximate the inverse transfer function of the optical fiber. Figure 5.8(c) compares the insertion loss for different propagation distances measured for the two layouts with $\Delta t = 5$ ps. It appears evident that a larger number of channels implies a larger insertion loss, possibly limiting the applicability of this configuration. For an estimation of the electrical power consumption of the device, one has to consider that, in a real case scenario, the amount of phase shift to be applied with each resistor depends on the specific interference conditions present on the device in its rest state (i.e., no currents are applied). These conditions are influenced, for example, by chip fabrication and temperature variations, and thus differ across the devices. Assuming the need to apply an average π

phase shift at each heater, the total electrical power can be estimated as $N \times 2 \times P_\pi$, where the factor of 2 accounts for the two weights (amplitude and phase) in each channel and P_π is the dissipation required to generate a π shift. Foundry specifications for efficient thermo-optic phase shifters provide an estimate of $P_\pi = 2$ mW [194], resulting in a total of 32 mW and 64 mW for the 8- and 16-channel designs, respectively.

Similar to other integrated photonic devices, the scalability comes at the cost of a larger footprint, higher power consumption by the microheaters, enhanced thermal cross-talk, and a longer training duration. Consistent efforts are being put into determining the tradeoff between complexity and performance. In this direction, starting from the results of the simulations presented above, the next generation of PNN device (presented in Appendix C) will implement both the 8 and 16-channel designs for an ultimate comparison between the two layouts.

5.3.2 Nonlinear propagation regime

A similar analysis to that proposed for CD equalization is devoted to adapting the PNN layout to nonlinear effects compensation at 100G. Again, the 8- and 16-tap designs equipped with different unitary delays Δt are compared, evaluating the performance in SPM equalization in the worst-case scenario, which is obtained by fixing the probe input power in fiber to $P_{pr} = 10$ dBm. The corresponding results after training in *Full* configuration are reported in Figure 5.9(a-b). Each curve, associated with a different Δt , reports the equalized BER obtained after the PNN training for different propagation distances. The reference performance is set by the unequalized curve (black stars). Considering the 8-tap layout, in panel (a), it appears that a PNN device featuring a large unitary delay $\Delta t/T_0 \geq 6/8$ generates a higher BER compared to the unequalized transmission. The same condition is present in a 16-tap design for $\Delta t/T_0 \geq 4/8$ up to $L \leq 150$ km, extending to $L = 200$ km for $\Delta t/T_0 \geq 6/8$. This behavior is justified by considering that, for long propagation distances, the interplay between SPM and CD (that is, the origin of the distortions) in each 50 km span accumulates, and a mutual influence between adjacent symbols is enforced. The equalization process requires, therefore, more memory, namely, the simultaneous access to an increasing number of consecutive symbols. On the contrary, for short propagation distances, mutual interference is minimal, and equalization requires a limited amount of memory. In our case, PNN devices featuring small Δt result particularly suitable for low-memory tasks: indeed, they allow maintaining a small time window for recombination (i.e., memory) while having multiple channels in the open state, and thus limiting the insertion losses. In other words, they can ensure high transmittance, and thus a high SNR at the receiver, and still avoid bringing to recombination unnecessary pieces of information which would degrade the equalization performance. On the other hand, this optimal condition cannot be achieved for PNN devices featuring larger Δt : having more channels in the open state reduces the insertion losses but inevitably introduces unwanted optical samples in the recombination process that cause distortions rather than eliminating them.

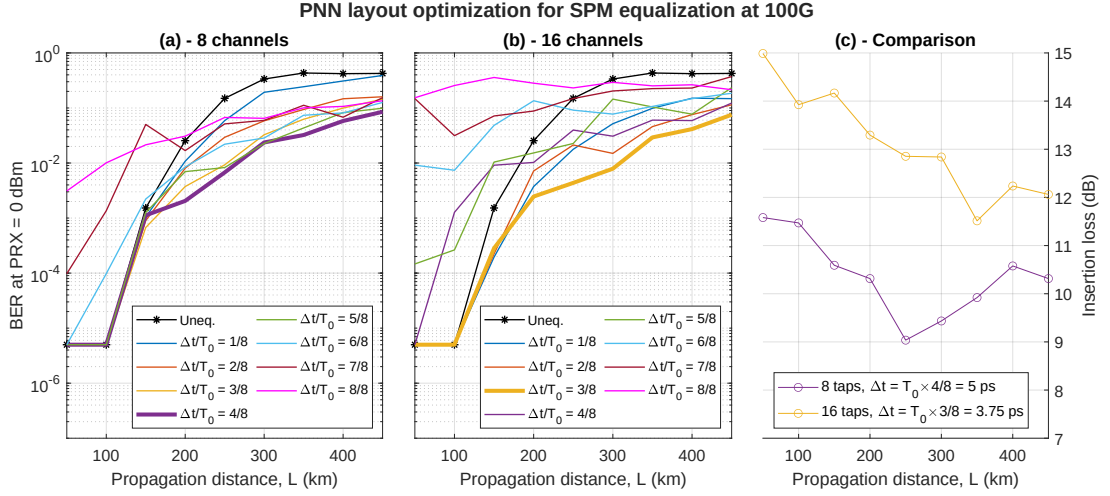


FIGURE 5.9: Simulated results for PNN layout optimization to operate SPM equalization on 100 Gbaud signals. The tested PNN layouts feature (a) $N = 8$ and (b) $N = 16$, respectively. Trainings are performed in *Full* configuration. (a-b) Equalized BER as a function of the propagation distance L . The unitary delay Δt is expressed as the corresponding fraction of the initial baud width $T_0 = 10$ ps. Error bars are not plotted for the sake of clarity in data visualization. Bold lines refer to the optimal PNN configurations. (c) Insertion Loss as a function of the propagation distance measured for the optimized PNN device in operation. The two curves refer to a PNN featuring $N = 8$, $\Delta t = 5$ ps, and $N = 16$, $\Delta t = 3.75$ ps, namely the PNN configurations offering the best performances in panels (a) and (b).

Moving back to the Δt optimization process, in both panels (a) and (b) of Figure 5.9, the best performance is identified by the lowest curve (bold line), revealing the optimal unitary delays to be $\Delta t = 5$ ps for $N = 8$ and $\Delta t = 3.75$ ps for $N = 16$. Despite the lower BER values provided by the optimized 16-taps layout, its better performance is obscured by the higher insertion loss (Figure 5.9(c)) and the more significant number of parameters to be trained that result in longer training times and higher chances of training failure. These critical aspects do not justify the slightly improved BER reduction compared to the 8-tap design, which remains the preferred option. In light of these considerations and those obtained from the previous section, the next-generation PNN layout will feature an 8-tap structure with $\Delta t = 5$ ps, which results in optimal performance for both CD and SPM equalization. The same unitary delay will also be used for the 16-tap design, thus focusing on its implementation for linear equalization.

The optimal 8-tap layout is then tested in various configurations, training the PNN for different combinations of probe input power P_{pr} and propagation distance L . The results are reported in Figure 5.10 in terms of unequalized and equalized BER at $\text{PRX}_{ref} = 0$ dBm in panels (a) and (b), and the corresponding BER reduction factor in panel (c). The BER values in both the maps show a general trend where the isolines describe a reversed U-shape: for fixed L , the BER decreases up to $P_{pr} = 3$ dBm, and then the trend

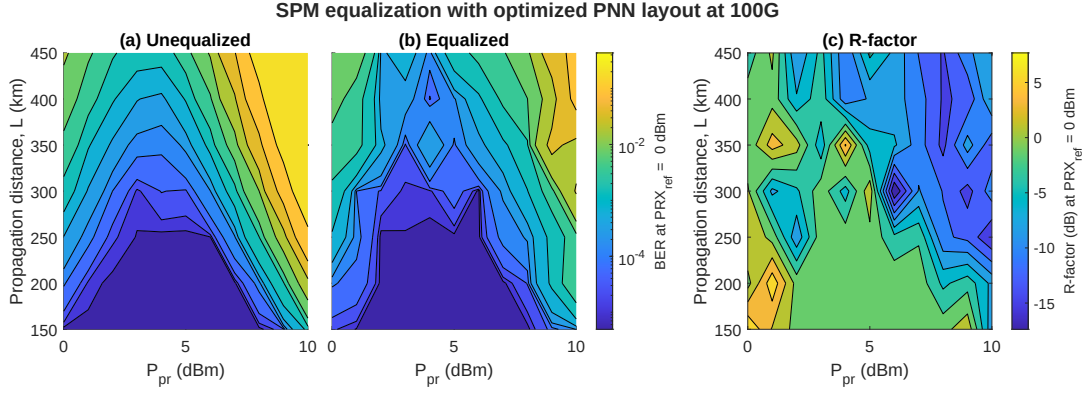


FIGURE 5.10: Simulated results for SPM equalization in 100 Gbaud modulation with an optimized PNN design featuring $N = 8$ and $\Delta t = 5$ ps. (a) Unequalized and (b) equalized BER values and (c) corresponding R -factor measured at $\text{PRX}_{ref} = 0$ dBm as a function of the probe input power in fiber P_{pr} and the propagation distance L . Null BER points have been replaced with the minimum measurable value.

is inverted for higher P_{pr} . The shape results from the interplay between the OSNR and the SPM effect: as the input power in the fiber increases, the OSNR increases too, and in turn, the BER diminishes. However, for $P_{pr} > 3$ dBm, the distortions induced by SPM become predominant over the augmented OSNR, causing a BER rise. The highest values are reached in the top-right corner of panel (a), corresponding to the unequalized transmission at high input power and long propagation distance. The trained PNN partially equalizes the SPM-induced distortions, imposing a BER reduction of up to 15 dB. In some cases, the PNN presence can even become detrimental rather than beneficial, as testified by positive R -factor values in panel (c) (yellow spots). Sub-optimal training processes may induce these, or, for low P_{pr} , SPM-related distortions are quasi-absent, thus not requiring equalization. In this regime, the PNN fails to transmit the signal without modifications, and the optical processing increases the BER.

Finally, the same optimized PNN layout has been applied to XPM equalization. The training process relies solely on the probe beam's initial bit sequence, without accounting for the random power fluctuations in the pump beam, which introduce distortions on the probe beam. As a result, the model is challenged to correct for these unpredictable variations. The first exploratory phase determines the threshold value for the pump input power in fiber P_{pu} that triggers XPM. Two random sequences of 2^{10} symbols have been co-propagated up to 450 km in different conditions of P_{pr} and P_{pu} , and the corresponding BER at $\text{PRX} = 0$ dBm are reported in Figure 5.11(a). For fixed P_{pr} , the mapped values assume a dependence on the pump power for $P_{pu} > 8$ dBm when the transmission quality is affected by XPM, with a consequent BER increase. This exploratory mapping has been performed by propagating relatively short sequences since using the long-traces approach described in Figure 5.7 would have been extremely time-consuming. Switching back to the long-traces approach for the following simulations,

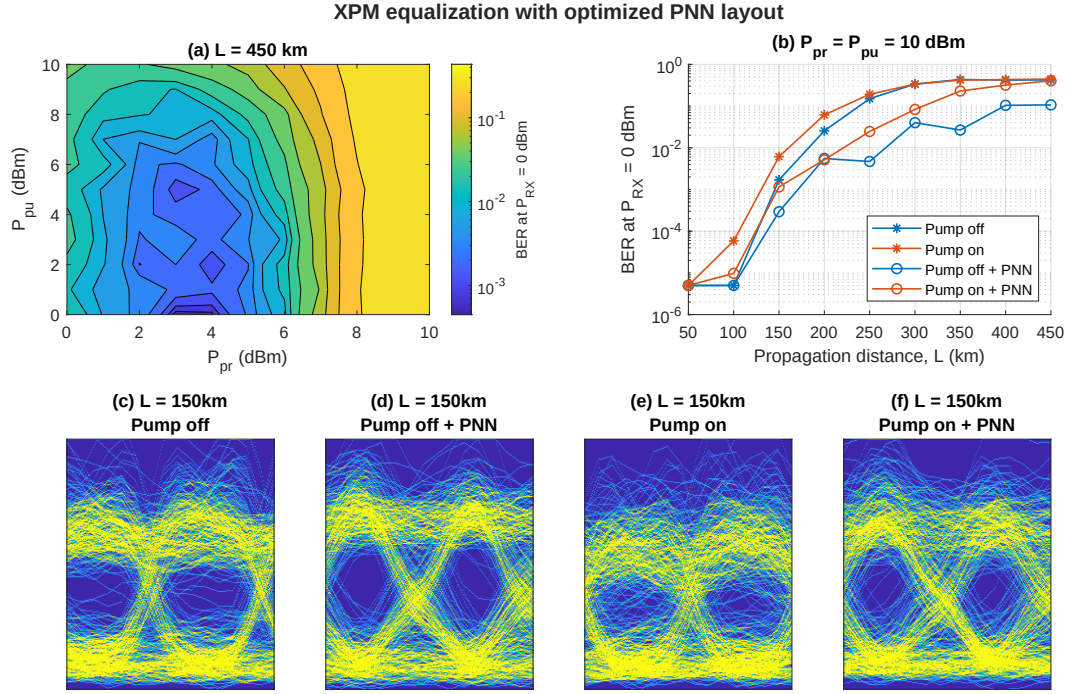


FIGURE 5.11: Simulated results for XPM equalization in a 100 Gbaud signal. (a) Simulated BER measured for $L = 450$ km as a function of the probe P_{pr} and pump P_{pu} input power in fiber. (b) Unequalized (stars) and equalized (circles) BER as a function of the propagation distance for $P_{pr} = 10$ dBm. Blue curves are obtained with the pump signal turned off, while orange curves are obtained by co-propagating the probe with a pump signal at $P_{pu} = 10$ dBm. Error bars are not plotted for the sake of clarity in data visualization. (c-f) Eye diagrams associated with the four profiles presented in (b) for $L = 150$ km.

the PNN is trained to equalize the transmission for $P_{pr} = P_{pu} = 10$ dBm; an interplay between SPM and XPM generates the accumulated distortions. Figure 5.11(b) presents the unequalized (stars) and equalized (circles) BER values at $PRX = 0$ dBm. Even for short propagation distances, the presence of the pump signal (orange stars, pump on) worsens the transmission compared to the probe-only propagation (blue stars, pump off). The trained PNN manages to improve the transmission quality even in the presence of the XPM effect despite being unable, in general, to reach the performance obtained against the SPM only. The eye diagrams of Figure 5.11(c-f), obtained for $L = 150$ km, show that XPM acts as an additional noise source that manifests as random fluctuations in the transmitted bit, thus causing a closure of the diagram. The PNN still manages, at least partially, to counteract this phenomenon while being completely agnostic about the encoding of the pump signal.

5.4 Comparison with other approaches

Applying the presented feed-forward PNN to nonlinear equalization represents a true novelty, allowing a switch from a linear FIR interpretation to an all-optical time-delayed complex perceptron. This work is inserted in a broad research spectrum where different approaches, summarized in Figure 5.12, are used to tackle the problem of nonlinear signal equalization. As of today, DSP-based technologies offer the highest reliability, robustness, and adaptability to different transmission scenarios [195–199], with the drawback of high power consumption, latency, and cost. A partial mitigation to this energy request and complexity is provided by hybrid approaches [169, 200–202], where the transmitted signal undergoes an optical processing, usually implemented as a Photonic Reservoir Computing, before being detected. The resulting sequences are then fed to a linear classifier or other algorithms for correct bit identification. A further step to DSP-free technologies is represented by all-optical approaches to signal equalization [166, 172, 203–205], where this work is inserted.

In Figure 5.12, the features of all the implementations presented in this section are summarized. Each of them is identified with the coordinates corresponding to the bit rate and the transmission reach it can handle. Our implemented PNN device lies in the bottom-right region of the plane (blue stars) since it can provide equalization after large propagation distances, but the symbol rate remains limited. Simulations performed at

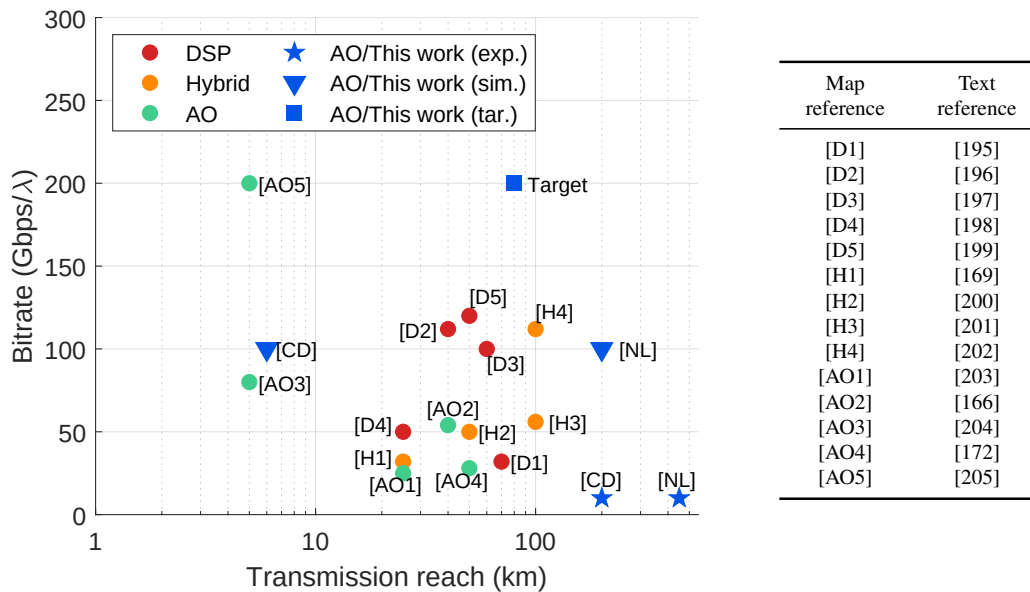


FIGURE 5.12: Map of the best equalization performance regarding bitrate per single wavelength λ and transmission reach obtained in the references (in table) via the different approaches based on DSP (red dots), all-optical processing (green dots), and hybrid techniques (orange dots). The results of the present work in terms of experimental (blue stars) and simulated (blue triangles) outcomes are presented, as well as the specifics of the foreseen target application (blue square).

100G with an adapted layout allow overcoming the limitation on the bit rate (blue triangles), taking our approach closer to a real-case scenario. The final target application, however, aims at the implementation into the next-generation optical transceivers, therefore dealing with 200 Gbps/ λ data rates in links up to 80 km (blue square).

5.5 Supplementary activities towards the next-generation devices

The experimental and simulated outcomes presented so far regarding the 8-channel TDCP are at the heart of this work. Other than these, further activities have been carried out as preparation for the future implementations of photonic chips. Specifically, these involve the characterization and testing of two new Si-photonics structures, as well as the exploration of a new loss function to perform PNN training in an unsupervised regime.

5.5.1 Perceptron with electro-optic phase shifters

In the PNN device tested so far, amplitude and phase weights are controlled via thermal phase shifters, which exploit the thermo-optic effect to modify the refractive index and induce a phase change in the underneath waveguide. Not only do the resistors used as phase shifters represent the only source of power consumption of the PNN device, but they also cause thermal cross-talk between the weights, due to the heat spreading across the whole chip. To circumnavigate these limitations, we explored the implementations of electro-optic phase shifters based on PN-junctions. These were inserted in the 4-channel TDCP presented in Figure 5.13, realized by Advanced Micro Foundry [206]. After the splitting and delay stage, which imposes a relative time shift of $\Delta t = 40$ ps, the two top channels are equipped with MZIs driven by thermal phase shifters and acting as amplitude weights. However, these are set in the open state for the subsequent training procedures, leaving the electro-optic phase shifters as the only free parameters. Their implementation consists of two PN-junctions sharing a p-doped region, as shown in the figure inset. Two parallel waveguides containing delayed copies of the input signal are enclosed within the p- and n-doped regions of each junction. The relative phase shift between the waveguides is generated via the Plasma Dispersion Effect, for which the concentration of free charges in Silicon changes the real and imaginary parts of the refractive index [207]. The free-carriers concentration is controlled via V_{drive} and V_{bias} , with the constraints that $V_{drive} > V_{bias}$ and $V_{bias} < 0$, to maintain the reverse polarization. In principle, this implementation features lower power consumption compared to thermal heaters, and also avoids the problem of thermal cross-talk between the weights.

The test bench for the 4-channel device is represented by CD equalization on 10 Gbaud PAM2 signals, using both V_{bias} and V_{drive} for a total of 4 tunable parameters. The corresponding results, obtained in collaboration with Elisabetta Ferri (master student),

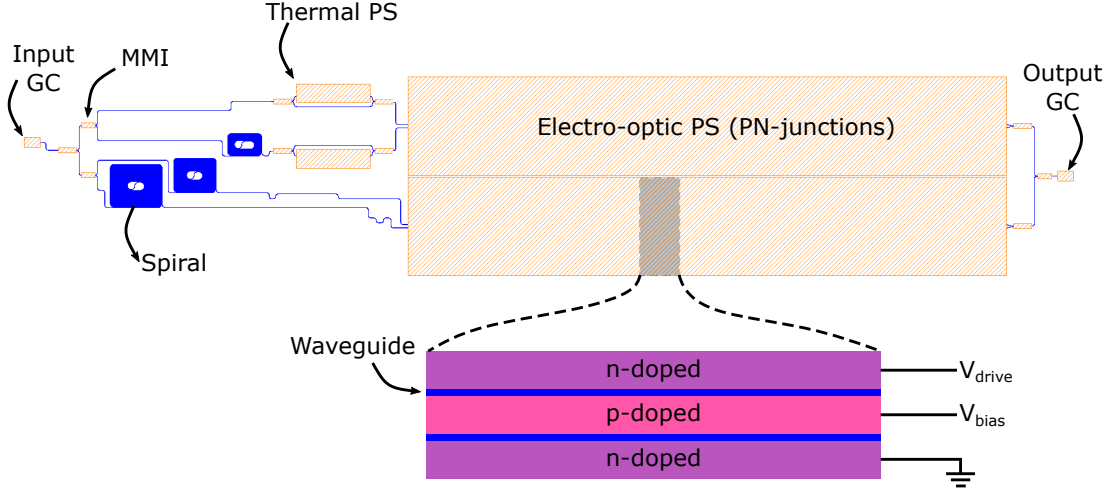


FIGURE 5.13: Layout of the 4-channel TDCP circuit equipped with electro-optic phase shifters (PS). Grating couplers (GCs) represent the input and output ports from the device, while the spirals provide delays from 0 to $3 \times \Delta t$, with $\Delta t = 40$ ps. Two thermal phase shifters are inserted to control amplitude weights on the two upper channels, but were kept open during the experiments. The inset shows the internal structure of the electro-optic phase shifters, constituted by two PN-junctions sharing a p-doped region. Orange boxes are part of the proprietary process design kit of Advanced Micro Foundry [206].

are reported in Figure 5.14 in terms of BER versus PRX profiles obtained for various propagation distances. The trained PNN manages to bring the BER under the pre-FEC threshold of 2×10^{-3} up to 150 km, showing, however, a performance reduction compared to the 8-channel version tested so far. Considering the 150 km case, for example, the old device reduced the BER by more than 5 orders of magnitude at $\text{PRX} = -3$ dBm (see Figure 4.4), while in the new version this is limited to 3 orders. The limited performances are determined by multiple factors, the first of which is the layout features in terms of a restricted number of taps and a larger unitary delay, corresponding to a less dense optical sampling. In addition, during the training, the driving voltages were conservatively limited to $V_{\text{drive}} > 0$, thus restricting the explorable configurations of the device. This was done to prevent any accidental occurrence of a direct polarization of the upper junction in each couple, which could possibly damage the device. However, one also has to consider that the electro-optic phase shifters are designed to work within a Si-based modulator and thus to be operated by high-frequency signals. Large swings in the driving signals allow reaching up to 2π phase shifts between the arms, while, unfortunately, this is prevented in DC operations when the voltage stays within the failure range provided by the foundry. Therefore, even in the absence of the imposed range limitation on V_{drive} during the training, not all the phase configurations could have been explored, limiting the device performances in any case.

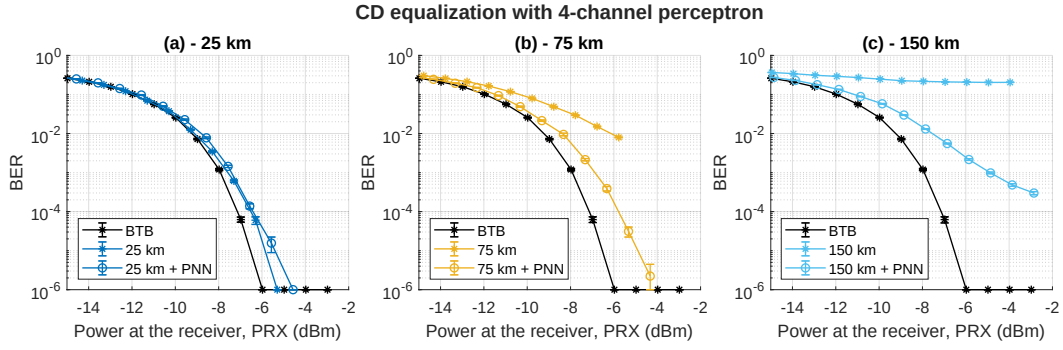


FIGURE 5.14: BER versus PRX profiles for CD equalization on 10 Gbaud PAM2 signals with a 4-channel PNN device implementing electro-optic phase shifters. The unequalized (colored stars) and equalized (colored circles) profiles are compared with the Back-to-Back performance (BTB, black stars). For these measurements, the fiber length within the loop was set to 25 km.

	8 taps (thermo-optic)	4 taps (electro-optic)
Foundry	Applied Nanotools [175]	Advanced Micro Foundry [206]
Footprint	3.77 mm ²	3.68 mm ²
Channels	8	4
Coupling	Edge	Grating
Max. reach for CD equal.	200 km	150 km
Training parameters	14	4
Power consumption	~ 350 mW	< 60 mW
Insertion loss	~18 dB	~25 dB

TABLE 5.1: Comparison between two PNN devices featuring 8 taps with thermal phase shifters and 4 taps with electro-optic phase shifters, respectively. The power consumption for the 4-channel implementation refers to the contributions coming from PN-junctions only.

Table 5.1 summarizes the main features of the two PNN layouts, highlighting useful elements for a final comparison between the phase shifters implementations. PN-junctions offer lower power consumption, but are affected by large dimensions, which would constitute the almost entire footprint in an 8-tap device. For this reason, combined with the limited equalization performance, the PN-junction-based implementation has been discarded for the next-generation devices, which will still be equipped with thermal phase shifters.

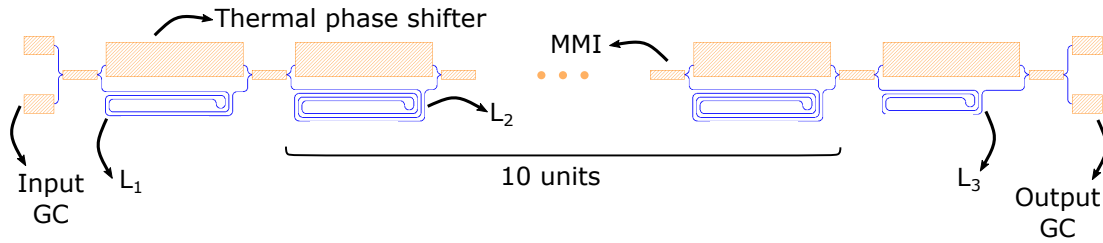


FIGURE 5.15: Schematics of the Mach-Zehnder Cascade (MZC) design. Input and output ports are represented by Grating Couplers (GC), while the splitting stages are 2×2 Multi-Mode Interferometers (MMIs). In the design, the spirals appear in 3 lengths: $L_1 = 1236 \mu\text{m}$ at the beginning, $L_3 = 930 \mu\text{m}$ at the end, and $L_2 = 1866 \mu\text{m}$ repeated for 10 units in between. Orange boxes are part of the proprietary process design kit of Advanced Micro Foundry [206].

5.5.2 Exploring new implementations: the Mach-Zehnder cascade

One of the main limitations imposed by the design of the 8-channel TDCP introduced in Section 3.4.2 is represented by the high insertion losses, which in the experiments have always been measured around 20 dB. In an attempt to reduce this contribution, a different PNN device layout, from now on called Mach-Zehnder Cascade (MZC), has been implemented, taking inspiration from the lattice filter presented in Figure 3.3 [166]. It's worth specifying that, as in the case of the TDCP implementation, the MZC still represents just the linear stage of the PNN, while the activation function is provided by the end-of-line photodetector. Figure 5.15 presents a schematization of the device layout, whose implementation was realized by Advanced Micro Foundry [206]. The optical signal is coupled in and out of the device via Grating couplers, while the main structure is constituted by a series of sub-units (from here the name *cascade*). Each of these contains an unbalanced MZI, with one arm hosting a tunable phase shifter (thermal), and the other a spiral, which can be found in three different lengths: $L_1 = 1236 \mu\text{m}$ at the beginning, $L_3 = 930 \mu\text{m}$ at the end, and $L_2 = 1866 \mu\text{m}$ repeated for 10 units in between. As already said for the lattice filter presented in Section 3.4.1, at each tunable coupler, spectral components are sent respectively to the upper and lower arms of the subsequent unbalanced MZI, thus imposing a longer propagation path to specific wavelengths. These accumulate a delay with respect to those traveling in the straight waveguide, which can be used to balance the effect generated by CD. The first and last units act as a demultiplexer/multiplexer, while the central units provide the proper amount of time delay to the different frequency components [165]. This design promises lower insertion losses compared to the TDCP implementation, since the optical power is preserved at every stage but the last.

Figure 5.16 compares the CD equalization results and overall performances obtained with the two PNN implementations (TDCP and MZC). Panels (a-c) present the BER versus PRX profiles for unequaled and equalized (blue for TDCP and orange for MZC) transmission at different distances. The two layouts perform similarly up to

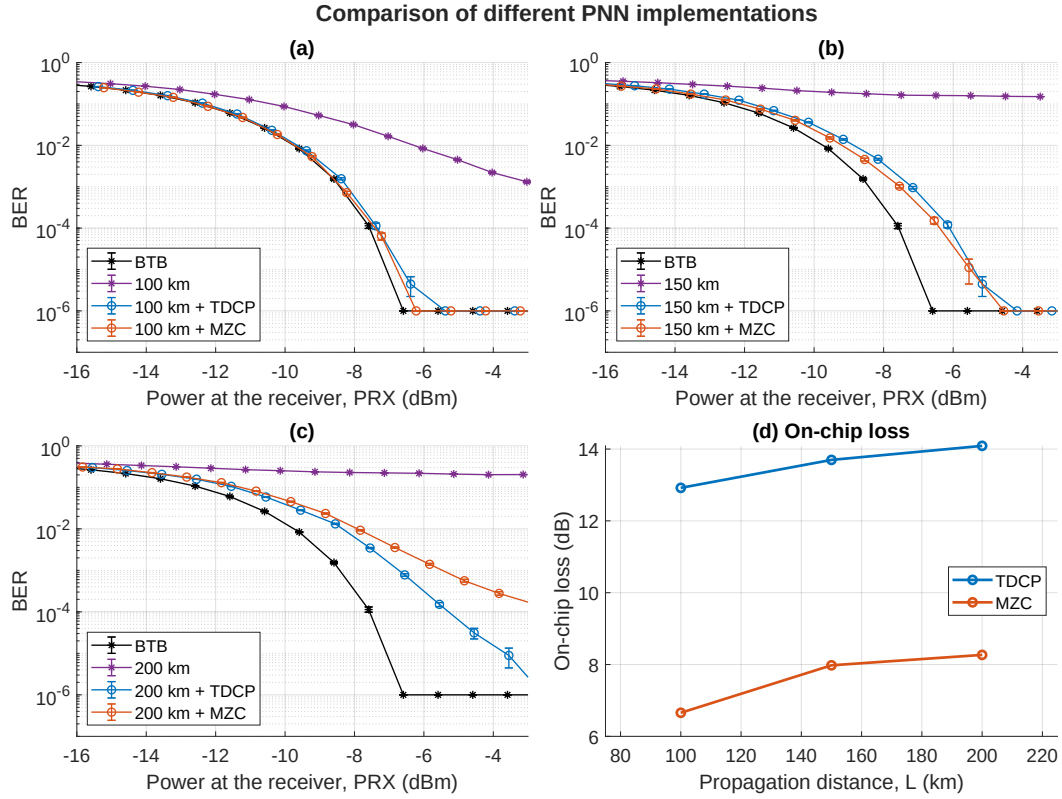


FIGURE 5.16: Comparison of the equalization performance of the 8-channel TDCP and the MZC layouts in CD equalization. (a-c) BER versus PRX profiles for unequalized (violet) and equalized transmission at 100 (a), 150 (b), and 200 km (c) TDCP (blue) and MZC (orange). The reference is set by a 0 km propagation (BTB, black curve). (d) Measured on-chip loss of the trained devices when in operation.

$L = 150$ km, and then the MZC provides worse equalization when the distortions become more severe. A possible cause lies in the shorter memory provided in this design compared to the TDCP, but more detailed explanations will follow from currently ongoing tests. The lack of performance in equalization for the MZC is, however, compensated by the reduced excess loss, as it appears in panel (d). As expected, this represents a major advantage provided by the MZC, which keeps it among the possible candidates for future applications in commercial devices. For this reason, a 14-stage version of the layout has been adapted to a 100G scenario and inserted in the next fabrication mask for further comparisons with the TDCP (see Appendix C).

5.5.3 Equalization with unsupervised learning

Up to this point, the presented learning procedures of the PNN devices have been carried out in a supervised regime. In the context of Machine Learning, this condition is verified when the evaluation of the loss function to be minimized during the training requires a comparison with a target [132]. It is the case, for example, of the one-sample

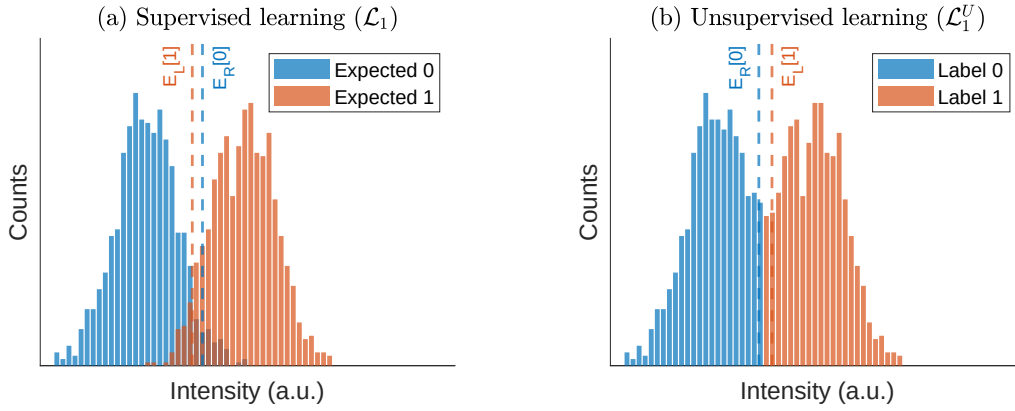


FIGURE 5.17: Distributions of samples in the received signals based on different labeling criteria. In both panels, $E_R[0]$ and $E_L[1]$ identify the neighboring tails of the two distributions. (a) In a supervised learning regime, the clustering is given by the expected value, 0 or 1, for each sample determined after comparing the received trace with the target sequence. This clustering is operated within $\mathcal{L}_1$. (b) In an unsupervised learning regime, the clustering is given by the belonging of each sample of the received trace to the lower or upper 50% of the overall population, thus avoiding any comparison with the target sequence. This clustering is operated within $\mathcal{L}_1^U$.

separation loss function $\mathcal{L}_1$, defined in Section 4.1.1: the received trace (down-sampled and normalized) is, indeed, overlapped with the digital target sequence to label the expected value for each symbol, which can be 0 or 1 in case of a 2-level modulation. Referring to that same notation used in Section 4.1.1, we assume now to deal with the optimal sub-sampling condition. Thus, from now on, any explicit reference to the optimal sub-sampling performed in correspondence of k_{opt} will be dropped. The labeling provided by the target generates two distributions, blue and orange, in Figure 5.17(a). $E_R[0]$ estimates the position of the right tail in the Expected 0s distribution and corresponds to the average of the highest 10% of that population. Analogously, $E_L[1]$ estimates the position of the left tail in the Expected 1s distribution and corresponds to the average of the lowest 10% of that population. Returning to the definition of Equation 4.5, we have that the loss function measures the relative distance between the tails as

$$\mathcal{L}_1 = E_R[0] - E_L[1], \quad (5.1)$$

and the training procedure aims to maximize the separation between the distributions, which corresponds to minimizing $\mathcal{L}_1$.

On the other hand, in an unsupervised regime, the loss function evaluation does not require any comparison with a target; therefore, pattern identification or classification tasks are solely based on the data distributions. Translating this into the current work context, it is possible to define a new loss function that measures the transmission quality without any a priori knowledge of the target sequence. Specifically, consider a trace at the receiver and its sub-sampled version s (equivalent to $y_{k_{opt}}$ defined in Section 4.1.1, but without any normalization). The points in s generate the overall distribution

described in Figure 5.17(b), in which the points corresponding to the lowest 50% of the population are labeled as 0s, while those in the upper 50% are labeled as 1s. This creates the distinction between the blue and orange sub-distributions, which are characterized by their variances σ_0^2 and σ_1^2 and the position of the two neighboring tails $E_R[0]$ and $E_L[1]$. These are estimated as described above, starting from the rightmost and leftmost 10% of the two populations. The new loss function is defined as

$$\mathcal{L}_1^U = E_R[0] - E_L[1] + \beta(\sigma_0^2 + \sigma_1^2) - \alpha\bar{s} \quad (5.2)$$

with $\bar{s}$ being the average of s and $\alpha = 0.1$ and $\beta = 0.1 \text{ mV}^{-1}$ being two fixed corrective factors applied to avoid the presence of dominating terms in the loss function definition. Notice that $\mathcal{L}_1^U$ is implicitly defined on a single sample, equivalently to Equation 4.5, and the generalization to the two-sample case $\mathcal{L}_2^U$ follows straightforwardly from Equation 4.6. In this new framework, the training is performed with an $\mathcal{L}_2^U$ minimization, during which the reduction of each of the terms appearing in Equation 5.2 accounts for a specific contribution to the eye diagram aperture. Overall, minimizing $\mathcal{L}_2^U$ simultaneously allows separating the distributions ($E_R[0] - E_L[1]$), reducing their width ($\sigma_0^2 + \sigma_1^2$), and maintaining a considerable signal average ($\bar{s}$). These additional terms appearing in Equation 5.2 are intended to ensure that the training leads to well-separated and narrow distributions, a condition that is fulfilled by an equalized signal for which the labeling proposed in $\mathcal{L}_2^U$ matches the requirements of the target. The only bias present in the definition of $\mathcal{L}_2^U$ is introduced with the labeling procedure, which is operated by selecting the points contributing to the lower or upper 50% of the overall population. This implicitly assumes that the original sequence (and thus the target) contains an equal number of 0s and 1s: in this case, the probability of a match between the so-obtained labeling and that proposed by the target increases, so that a separation of the lower and upper half of the distribution leads to an actual BER reduction. On the other hand, the higher the imbalance between the number of 0s and 1s in the input sequence, the higher the probability of incorrect labeling. Fortunately, the mentioned condition of an approximately equal number of low and high bits is totally reasonable (and also desirable) in a real case scenario of propagating optical sequences [208].

In $\mathcal{L}_2$, the loss function is based on a comparison between the received information and the target sequence. In contrast, in $\mathcal{L}_2^U$ the evaluation is agnostic of the target information. In a real-case implementation with the PNN placed at the receiver, this solution allows continuous monitoring of the PNN performance and weights adaptation without transmitting pilot symbols, being equivalent to optical blind equalization [209].

A comparison between the supervised and unsupervised approaches is proposed here. Notice that the BER is evaluated according to the same modalities described in Section 4.1.1, with the threshold estimation provided by the standard separation loss function used in the supervised regime (Equation 4.7). This allows for a fair comparison in terms of BER values with the supervised case. Figure 5.18 shows the equalized eye diagrams obtained in different transmission scenarios. Panels (a,e) and (b,f) focus

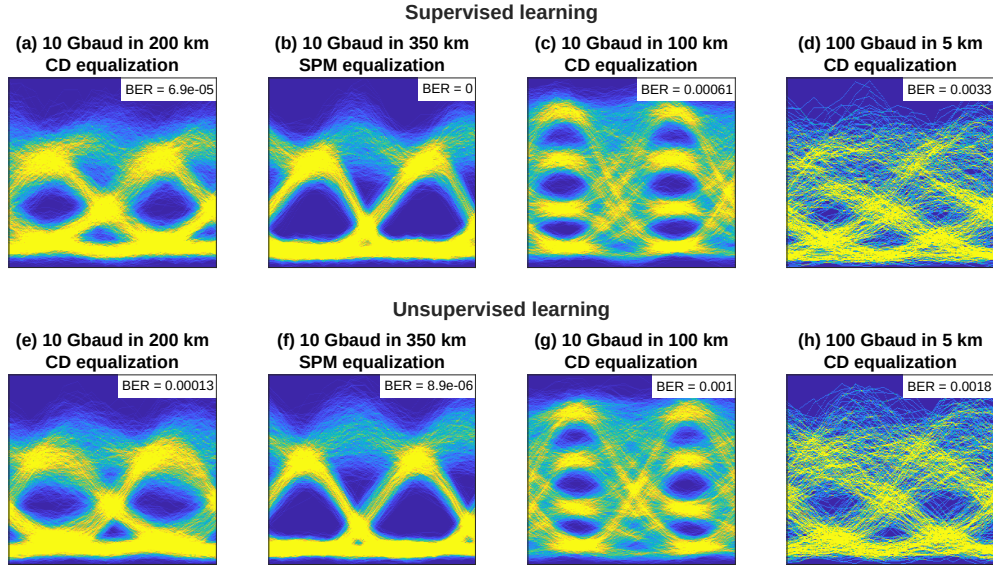


FIGURE 5.18: Eye diagrams after equalized transmission with the PNN trained in supervised (a-d) and unsupervised (e-h) regimes. Data obtained with a 10G modulation are experimentally produced, using the MZC as a PNN layout. The results presented in panels (b,f) are obtained with $P_{in} > 10$ dBm. On the other hand, results for 100G modulations are simulated, implementing the 8-tap TDCP with $\Delta t = 5$ ps.

on a 2-level modulation at 10 Gbaud propagated in linear and nonlinear regimes, respectively. Panels (c,g) demonstrate the generalization of the new loss function $\mathcal{L}_2^U$ to multi-level formats. All of these are experimental results obtained while implementing the MZC as the PNN layout. Panels (d,h), on the other hand, present simulated results obtained in a 100G application using the 8-tap TDCP with $\Delta t = 5$ ps found during the optimization process of Section 5.3.1. The propagation distances associated with the various eye diagrams correspond to the maximum reach for which the unsupervised learning manages to produce compatible performance with the supervised. For longer links, the distortions become too severe for the minimization of $\mathcal{L}_2^U$ to produce effective equalization, leading to training failure in many cases. However, the proposed approach still represents a suitable candidate for future applications, demonstrating the applicability to different modulation formats, propagation regimes, modulation frequencies, and PNN implementations.

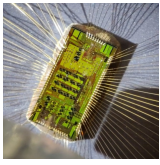
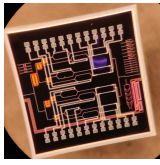
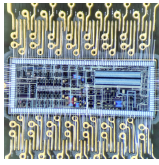
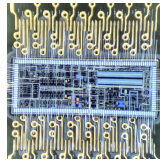
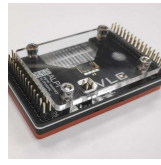
Chapter 6

Conclusions

This chapter concludes the manuscript, providing an overall review of the presented results and defining the next steps of this project. First, the current state-of-the-art technologies are put in relation to the presented solutions for optical signal equalization, highlighting their innovative contributions and the potential criticalities. Subsequently, future developments of this research are outlined, reviewing the results that led to the design of the next generation of photonic circuits for 100G applications. Finally, an overall summary of this work is presented.

6.1 Comparison with the current state-of-the-art

Although coherent modulation is becoming more and more popular in many applications, most data centers and access links still rely on IMDD schemes. These are favored over coherent formats due to their advantages in terms of hardware simplicity, power consumption, and implementation and maintenance costs [213]. The standard in these scenarios is nowadays represented by 25G per lane DWDM transmission, cumulated to 100G in PAM4 modulation, which is used to cover different spans, from 100 m to 80 km [214]. The increasing bandwidth request interests these applications too, but scaling IMDD links beyond current rates poses formidable challenges. The hypothetical 100G PAM2 scenario treated in the simulations of Section 5.3 would consent a maximum 400 m transmission to stay below a 5×10^{-3} pre-FEC threshold, and this under the conditions of having $\text{PRX} = -1$ dBm and DSP-based 9-tap feedforward and feedback equalizers (see Section 2.3.3.1) at the receiver. The requests become even more demanding when switching to a 200G PAM4 modulation, for which the needs scale to a 7 dBm PRX, a 71-tap feedforward and 15-tap feedback equalizers [215]. DSP engines must, therefore, evolve as well to sustain such a computational effort. ASICs implemented for this purpose, even those based on the latest 3 nm CMOS technology, are pushed to the limit by equalizing 100G signals for reaches above 400 m. The heavy reliance on DSP introduces unsustainable energy consumption and unacceptable latency overheads ($\sim$ microseconds). Present technology is thus incompatible with the requirements imposed for the GPU-based and AI-driven next-generation data centers, in which peak performance must coexist with sustainability. Overall, two different growth rates exist:

Overview of the tested PNN implementations					
Picture					
Generation	1.0	2.0	2.1	2.1	2.2
Platform	Silicon	Silicon	Silicon	Silicon	InP
Foundry	IMEC	ANT	AMF	AMF	HHI
Structures	TDCP	TDCP	TDCP	MZC	TDCP + SOA-based act. fun.
# taps	4	8	4	12*	8
Unitary delay Δt	50 ps	25 ps	40 ps	25 ps*	25 ps
Coupling	Grating	Edge	Grating	Grating	Edge
Weights control	Thermo- optic	Thermo- optic	Electro-optic	Thermo- optic	Electro-optic
Pow. cons.	70 mW	~ 350 mW	< 60 mW	< 30 mW	-
Insertion loss	~16 dB	18-22 dB	~25 dB	~20 dB	-
Max. tested bitrate	10G	20G	10G	10G	-
Tested modulation formats	PAM2	PAM2, PAM4, OFDM	PAM2	PAM2	-
Article ref.	[187]	[176, 177, 184, 210]	-	-	-

(*) For the MZC, data for # taps and unitary delay correspond to # sub-units and delay induced by inner spirals.

TABLE 6.1: Overview of the tested PNN implementations and their main features. The main experimental results presented in this work refer to Version 2.0. TDCP: time-delayed complex perceptron; MZC: Mach-Zehnder Cascade. Foundries: IMEC [211]; ANT: Applied Nanotools [175]; AMF: Advanced Micro Foundry [206]; HHI: Fraunhofer HHI: Fraunhofer Heinrich-Hertz-Institut [212].

on one side, the expansion of multi-datacenter infrastructures focused on larger bandwidth, and on the other side, the bottleneck represented by current electronic and optical

technologies that cannot meet the requirements. Apparently, overcoming the hard ceiling set by 100G transmission over 2-3 km requires a paradigm change. For many data center management facilities, this consists of switching to coherent technologies over IMDD, but they still face prohibitive costs for widespread deployment. A revolutionary third path represented by photonic technologies has been proposed as a viable solution to reconcile performance and sustainability. Integrated photonic platforms are used to implement various AI paradigms (Feed-Forward, Reservoir Computing, ...), pioneering optical signal processing for ultra-fast applications. As introduced in Sections 4.5 and 5.4, both full-optical and electro-optical hybrid approaches are proposed with the intent of breaking the technological deadlock.

The approach described in this work is conceived within this context. It implements a feed-forward neuromorphic photonic equalizer that processes optical signals without the energy-intensive optical-electrical-optical conversions that bottleneck today's systems. Signal processing is moved from the digital electronic to the analog photonic domain, unlocking previously impossible performance levels. The use of electronic computation is limited to the training process, which is performed offline. After training, the presented PNN performs real-time, all-optical compensation of both linear and nonlinear signal distortions with unprecedented efficiency. Operations are performed with an ultra-low latency of less than 100 ps compared to the 800,000 ps typical of conventional DSP systems [205]. In addition to latency advantages, the proposed approach dramatically reduces energy demand, bringing down the consumption to ~ 300 mW from the $\sim W$ required by DSP engines. On a forecast 100G application, the PNN device equipped with more efficient thermal phase shifters offers sub-pJ/bit performance for processing and massive scalability across WDM channels. The realization on a Silicon-On-Insulator platform makes it compatible with standard CMOS process, thus further reducing the costs per unit in mass production. The small dimensions make it prone to integration in an optical transceiver without the need for altering the form factor.

The PNN design presented in this work is already at the second iteration of devices, as illustrated in Table 6.1. This work validates the use of the selected layout for both linear and nonlinear equalization on a 10 Gbaud signal, acting as a proof of concept for future implementations targeted at higher bit rates. Simulative studies on the next-generation devices have been performed, oriented both to the optimization of the current layout to 100G and to the exploration of new architectures.

6.2 Next-generation optical devices for applications at 100G and above

The first experimental tests for equalization at commercial baud rates will be performed with the next fabricated chip, expected for April 2026, and hosting the structures reported in Appendix C. As anticipated, this will implement a novel version of a TDCP, with adapted parameters following the simulations of Figure 5.8. The selected layout

allows choosing between an 8 or 16-tap design with a unitary delay of 5 ps, corresponding to half the baud width. This design allows for an ultimate comparison between the two architectures: an 8-tap design limits the insertion losses and reduces the number of parameters, thus easing the training process; on the other hand, a 16-tap design offers higher performances, with CD equalization in reaches above 6 km. A second structure is represented by a 14-stage MZC, for which low insertion losses are expected, and thus a comparison with the TDCP will follow.

Finally, a third structure inserted in the new mask allows testing the Recurrent Neural Network (RNN) and RC paradigms, which in previous applications [172, 205] resulted in effective compensation for linear and nonlinear distortions. The new structure is conceived taking inspiration from the design proposed in [172] and adapted to the 100G scenario (see Appendix B.2.3). The layout consists of two parts. First, a Photonic Reservoir (PR) is implemented as a 4-node architecture depicted in Figure 6.1, with each node constituted by a 3×3 structure resulting from the combination of two 2×2 MMIs, giving a directionality to each node. A generic internode connection is constituted by a waveguide, providing a delay between 7.5 ps and 15 ps, and a tunable phase shifter, which allows for testing different configurations of the PR. Each node leads to the second part of the network, which consists of an all-optical readout stage that ensures ultra-fast data processing and minimal DSP usage. The readout is implemented in

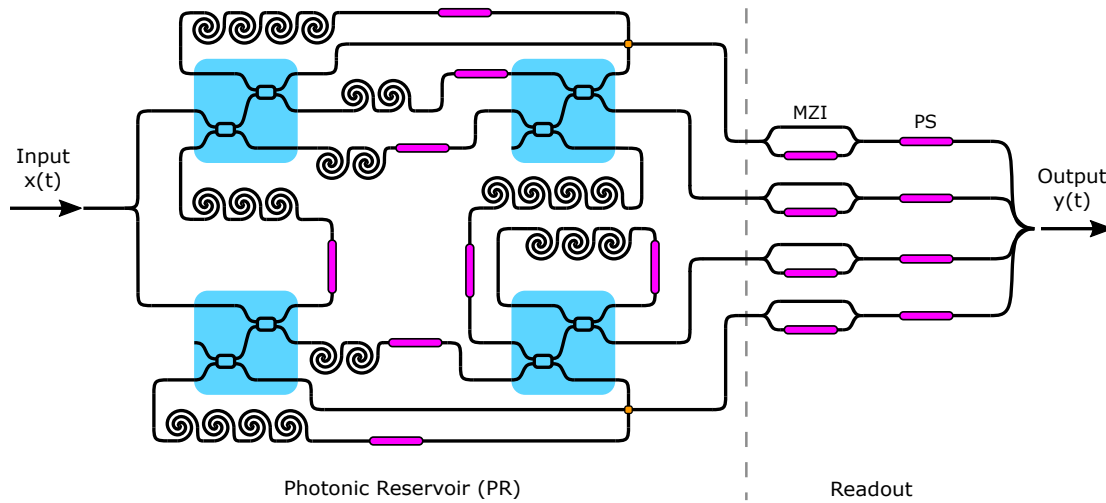


FIGURE 6.1: Schematics of the PNN design featuring a Photonic Reservoir (PR) with 4-nodes (blue rectangles) followed by an all-optical readout for 100G applications. The input signal $x(t)$ is split into two copies and enters the PR in two nodes. These are constituted by a cascade of 2×2 MMIs, and the inter-nodes connections are provided by delay lines and thermal phasors. Each spiral adds 3.75 ps delay in the propagation. In the readout stage, 4 channels connected to each node collect the corresponding signal, which then propagates through a MZI and a phase shifter PS accounting for an amplitude and a phase weight, respectively. The weighted signals are then summed together to produce the final output $y(t)$.

four channels equipped with amplitude and phase weights ending in a summation stage, similarly to a TDCP.

In our design, the presence of recurrence generates a fading memory (see Appendix B.2.3), making this layout suitable for addressing memory-requiring problems such as CD and SPM equalization. The tunable phases in the internode connections represent a true novelty: the reduced number of nodes limits the design complexity, but the versatility of the PNN is maintained thanks to the tunable phase weights. Overall, the structure allows for testing two different paradigms. If all the weights (internode and readout) are tuned during the training, the device operates as a generic RNN. On the other hand, if the internode weights are set to a random configuration and only the readout parameters are tuned, it switches to an RC approach. It is worth specifying that, as it is for [172], the PR does not feature any intrinsic nonlinearity in the nodes. All of it will be enclosed in the response of an external device (e.g., photodetector, nonlinear amplifier,...) that collects the optical output of the integrated circuit.

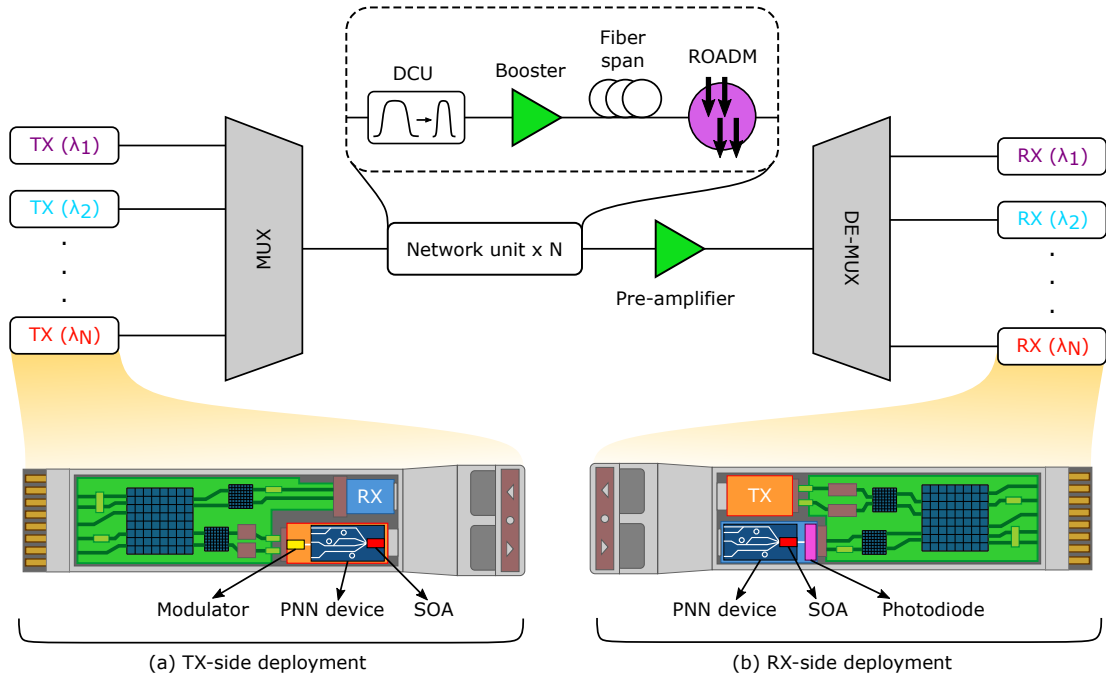


FIGURE 6.2: Foreseen application scenarios in a typical WDM metro network for the PNN device implemented into the transceivers. At the transmitter side (TX), different carriers are multiplexed together (MUX) and subsequently propagate in multiple network units. These are constituted by a Dispersion Compensating Unit (DCU), a booster, and a fiber span, followed by a ROADM for signal routing. After a last amplification stage, the carriers are de-multiplexed (DE-MUX) and sent to the receivers (RX). The PNN can be deployed at the TX (a) or at the RX (b) side, counteracting distortions originating from different phenomena. An SOA integrated within the PNN device acts as the nonlinear activation function.

The tests carried out on this new mask will provide fundamental information on the application of these photonic devices to high-speed scenarios. Successive demonstrators will be more and more oriented to real-case applications, focusing first on the reach extension to 80 km for 200G PAM4 signals, and then exploring the application to multi-span PAM4 links in meshed networks using fixed dispersion compensating units. This transmission scenario, illustrated in Figure 6.2, involves a multi-span amplified architecture with add/drop capabilities. ROADM nodes and dispersion compensation units (DCUs) define the unit network block, with SPM representing a severe impairment source. In this scenario, the PNN modules can be strategically deployed either before propagation (TX-side, panel (a)), after propagation (RX-side, panel (b)), or both, depending on the nature of the impairments. The TX-side deployment addresses modulator-induced distortion and CD, while the RX-side focuses on residual CD and SPM. This dual placement provides flexibility and separation of functions, performing everything with dramatic energy efficiency. Huge reduction compared to transceiver-implemented DSP is forecast, as illustrated in Table 6.2. Unsupervised learning represents a decisive ingredient for the mentioned applications, removing the need for pilot symbol transmission to train the PNN parameters. In turn, this results in adaptive equalization to channel variations without sacrificing bandwidth efficiency. Another

Competitive analysis of PAM4 transceivers							
Company	FS	Dell	Juniper	Arista	CISCO	Brocade	UniTn
Reach (km)	10	10	25	40	80	100	80
Wavelength (nm)	1310	1310	1310	1310	1310	1310	Any
Single lane bit rate (Gbps)	50	100	100	100	25	25	200
Aggregate rate (Gbps)	200	400	800	400	100	100	1600
Cost (\$)	1800	780	3000	1800	1800	5000	200
Power consumption (W)	10	10	18	12	5.5	6	3
In-built equalization	No	DSP	DSP	DSP	DSP	DSP	PNN
Form factor	QSFP56	QSFP-DD	OSFP	QSFP-DD	QSFP28	QSFP28	QSFP28/ OSFP

TABLE 6.2: Comparison between PAM4 optical transceivers with traditional DSP implementation and the PNN approach proposed in this work (last column). Data are taken from the web pages of each vendor. The power consumption estimation for the PNN approach is obtained by summing the contributions from SOA (300 mW), TX Laser Diode (1.5 W), RX Photodiode and Amplifier (1 W), and PNN (200 mW). The cost predictions are based on [216].

key enabler lies in the heterogeneous integration of InP-based SOAs with passive silicon photonic circuits via micro-transfer printing [217]. This approach combines the high-performance optical gain and nonlinear characteristics of InP with the scalability, compactness, and low-cost fabrication of silicon photonics. InP-based SOAs provide optical amplification to reduce the device's insertion losses, and simultaneously acts as a nonlinearity source, especially when used in a saturated regime. A first generation of InP-based devices (Table 6.1, last column) has been tested with poor results due to fabrication imperfections, but promising results come from the performed simulations, which include the nonlinear effects provided by an integrated SOA (credits to S. Gretter, unpublished). The final goal is to go beyond equalization to enable truly cognitive optical networks capable of self-optimization and autonomous adaptation, revolutionizing telecommunications infrastructure while creating foundational technology for neuro-morphic computing and optical AI acceleration.

6.3 Conclusions

In this work, a time-delayed complex perceptron has been used as PNN for channel equalization in telecom applications. The device is implemented on a Silicon photonics platform and features 8 channels time-spaced with a relative delay of $\Delta t = 25$ ps, corresponding to a quarter of a bit width in the tested 10 Gbaud transmission scenario.

The device is first implemented in a linear transmission regime to equalize CD in an IMDD link. In this context, it works as a tunable optical FIR filter, which is trained to approximate the inverse of the impulse response of a fiber. The device shows excellent equalization capabilities in different scenarios, obtained by varying the transmission rate, the modulation format, and the transmission reach up to 125 km (maximum offered by the setup). With standard PAM formats, the 2-sample separation loss function is used as a minimization function, providing a valid analog alternative to the digital BER for PNN training. This allows for achieving equalization of a 20 Gbaud 2-level signal up to 75 km and a 10 Gbaud 4-level signal up to 125 km. Similar performance is reached with 10 Gbaud OFDM signals up to 125 km, this time using the Error Vector Magnitude (EVM) function to measure the transmission quality. In the tested PAM4 scenarios, the PNN device manages to keep the BER far below the 2×10^{-3} pre-FEC threshold, operating with an average insertion loss of ~ 18 dB and a power consumption of ~ 350 mW.

In the context of PAM4 equalization, PSO and Adam algorithms are tested for optimization. The former appears more precise and accurate, with less probability of ending the research in a local minimum, but featuring the drawback of an increased computational cost and convergence time. In a comparison with a commercial TDC, the proposed PNN proved to be as effective in compensating CD while requiring less power consumption during the operation (~ 350 mW against a few W). The PNN device's effectiveness is confirmed by the flattening induced in the CD power penalty within the bandwidth of the transmitted signal. Unfortunately, due to manufacturing

imperfections, the 40 GHz periodicity in the response of the PNN predicted by the ideal model was not observed, limiting its applicability to WDM schemes.

The implementation of a new experimental setup featuring a 50 km fiber loop allowed for testing the PNN equalization capabilities in new and more demanding cases. In a linear regime, the PNN results effective against CD, extending the reach up to 200 km, and CD residuals left by the TDC up to 450 km. In a nonlinear propagation regime, equalization against SPM is achieved even in the most demanding transmission scenarios, represented by a 450 km propagation and an input power in fiber $P_{in} = 9$ dBm at the first loop. The BER is again maintained below the 2×10^{-3} pre-FEC threshold with average insertion loss of ~ 20 dB and power consumption of ~ 290 mW.

In parallel with the experimental activity, a simulated optical link is used to overcome the setup bandwidth limitations and study the scalability of the PNN approach to higher baud rates, up to 100G. First, two TDCP layouts featuring 8 and 16 taps are tested against linear and nonlinear distortions, optimizing the unitary delay Δt value to each scenario. CD equalization is provided up to 6 km (single span), while SPM equalization (isolated from CD) up to 200 km (multi-span), with $P_{in} = 10$ dBm restored every 50 km. In general, it is observed that the 16-tap design provides better equalization performances and flexibility, with the drawback of an increased training time and higher insertion losses (+3 dB on average) compared to the 8-tap version. An experimental comparison between the two layouts will be performed with the next-generation devices, which implement both of them. The optimized layout obtained from CD and SPM equalization (8 taps and $\Delta t = 5$ ps) is tested against XPM too, partially managing to counteract this phenomenon while being completely agnostic about the encoding of the pump signal.

Further experimental studies prone to the next-generation devices have been carried out, starting with exploring the use of electro-optic phase shifters as a less power-hungry alternative to thermo-optic ones. These have been implemented on a 4-channel TDCP, which has been tested in CD compensation up to 150 km. In front of a remarkably low energy demand (< 60 mW), the device provided limited equalization performance, partly due to the limited exploitable phase range. This reason, combined with the large footprint of the PN-junctions, made us exclude this approach for future devices. Proceeding, the PNN implementation based on a MZC produced slightly worse equalization results against CD for $L > 150$ km compared to the 8-channel TDCP, satisfying, however, the expectations of lower on-chip losses (~ 7 dB against ~ 14 dB). To proceed with the comparison, an analogous version scaled to a 100G scenario has been included in the next mask, together with an RC-based PNN design followed by an all-optical readout stage. Compared to a similar implementation proposed in [172], our design features a reduced number of nodes, but the versatility is maintained thanks to phase shifters included in the PR stage, making it reconfigurable. Finally, the aforementioned MZC served as the platform for benchmark tests on unsupervised equalization. This has been operated with a newly defined loss function, which measures the separation between optical levels without needing a comparison with the target sequence. The

proposed approach provided equalization results comparable with those obtained in the supervised regime, adapting to different modulation formats, baud rates, and distortions. This poses the basis for further developments, aiming at continuous online monitoring of the PNN performance and weights adaptation without transmitting pilot symbols.

From this study, photonics platforms emerge as integrated, scalable, and tunable solutions that perform signal processing with minimized latency and power consumption. They appear as key-enabling technologies for future applications in telecommunications, removing (or at least alleviating) the computational effort required by traditional DSP. Studies on the subsequent devices development are already in place, targeting applications in 200G per lane scenarios in single and multi-span WDM links. Particular effort is devoted to studying the role of InP-based SOAs integrated on Silicon platforms via transfer printing. These could enable active on-chip amplification, reducing the problem of high insertion losses, and provide nonlinearity when used in the saturation regime. This solution provides further flexibility to the platform, preparing it for real-case scenario implementation in 6G networks.

List of publications

Research and review papers

E. Staffoli, M. Mancinelli, P. Bettotti, and L. Pavesi, "Equalization of a 10 Gbps IMDD signal by a small silicon photonics time delayed neural network," *Photonics Research*, vol. 11, issue 5, 2023.

E. Staffoli, D. Bazzanella, S. Biasi, G. Donati, M. Mancinelli, P. Bettotti, and L. Pavesi, "Nonlinear response of Silicon Photonics microresonators for reservoir computing neural network", in CNIT Technical Report, Vol. 12, ISBN: 9788894982794, 2023.

S. Biasi, G. Donati, A. Lugnan, M. Mancinelli, E. Staffoli, and L. Pavesi, "Photonic Neural Networks Based on Integrated Silicon Microresonators", *Intelligent Computing*, vol. 3, id. 0067, 2024.

E. Staffoli, G. Maddinelli, and L. Pavesi, "A Silicon Photonic Neural Network for Chromatic Dispersion Compensation in 20 Gbps PAM4 Signal at 125 km and its Scalability up to 100 Gbps," in *Journal of Lightwave Technology*, vol. 43, no. 2, 2025.

P.R.N. Marciano, E. Staffoli, G. Maddinelli, M.S. Coelho, L.C. Silva, J.A. Silva, M.J. Pontes, M.E.V. Segatto, and L. Pavesi, "Chromatic Distortion Precompensation in OFDM-Based Optical Systems Through an Integrated Silicon Photonic Neural Network", *Journal of Lightwave Technology*, vol. 43, no. 7, 2025.

E. Staffoli, E. Ferri, S. Gretter, and L. Pavesi, "Nonlinear Distortion Equalization in Multi-Span Optical Links Via a Feed-Forward Photonic Neural Network", *Laser Photonics Reviews*, vol. 19, no. 23, 2025. (Correction article: *Laser Photonics Reviews*, e70909, 2026)

Conference proceedings

E. Staffoli, M. Mancinelli, 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), 2023.

E. Staffoli, M. Mancinelli, P. Bettotti, and L. Pavesi, "IM/DD Transmission Equalization with an All-Optical Photonic Neural Network", 2023 International Conference on Photonics in Switching and Computing (PSC), 2023.

E. Staffoli, M. Mancinelli, P. Bettotti, and L. Pavesi, "An All-optical delayed complex perceptron for signal equalization in IMDD fiber transmission", 2023 IEEE Photonics Conference (IPC), 2023.

E. Staffoli, G. Maddinelli, M. Mancinelli, P. Bettotti, and L. Pavesi, "Chromatic dispersion compensation via an all-optical perceptron", Next-Generation Optical Communication: Components, Sub-Systems, and Systems XIII, SPIE Digital Library, 2024.

Appendices

Appendix A

Experimental Setups and techniques

A.1 Optical transmission line up to 125 km

Figure A.1 depicts a schematic of the complete experimental setup used to produce the results appearing in Chapter 4 in the context of Chromatic Dispersion (CD) equalization. Manufacturing companies and models of each instrument are provided in Table A.1. Notice that the setup has undergone some modifications throughout the various experiments performed over the years. However, such modifications do not alter the overall functioning with respect to that described here.

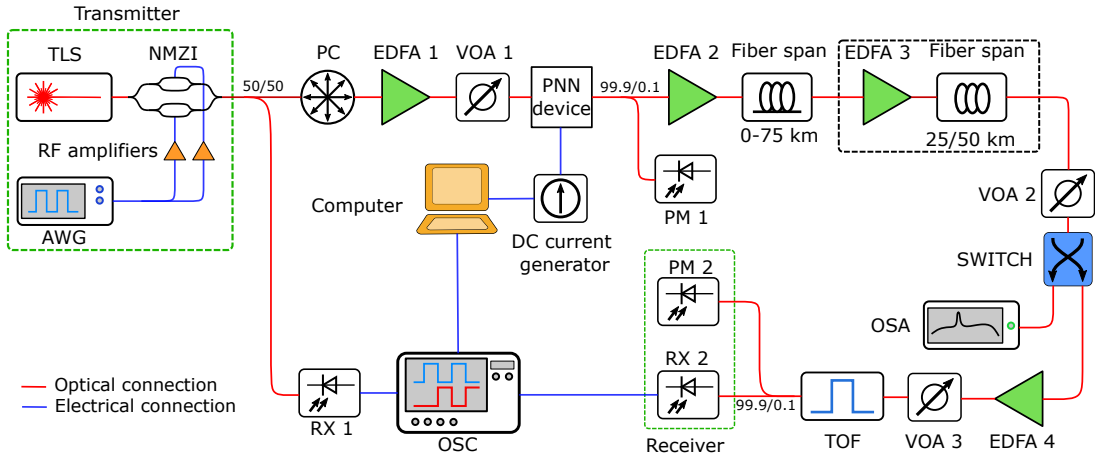


FIGURE A.1: Experimental setup used to produce the results of Chapter 4 in the context of CD equalization up to 125 km. TLS: Tunable Laser Source; NMZI: Nested Mach-Zehnder Interferometer; AWG: Arbitrary Waveform Generator; RF: Radio Frequency; PC: Polarization Controller; EDFA: Erbium-Doped Fiber Amplifier; VOA: Variable Optical Attenuator; PNN: Photonic Neural Network; PM: Power Monitor; OSA: Optical Spectrum Analyzer; TOF: Tunable Optical Filter; OSC: Oscilloscope; RX: Fast Photodiode.

A CW Tunable Laser Source (TLS) operating at 1550 nm is externally modulated via a Nested Mach-Zehnder Interferometer (NMZI). The digital signal is software-generated and based on a periodic Pseudo-Random Binary Sequence (PRBS) of order 10 and period 2^{10} bits. The pattern is produced by a 30 GHz-bandwidth Arbitrary Waveform Generator (AWG) connected to two 20 GHz-bandwidth RF amplifiers, which in turn drive the modulator. A 50/50 Fiber Optic Coupler sends half of the optical power to a 20 GHz-bandwidth fast photodiode (RX 1), where the input signal is acquired for reference. The other half of the optical power proceeds to a polarization controller, which is needed since the Photonic Neural Network (PNN) device placed later in the line is polarization-sensitive. An Erbium-Doped Fiber Amplifier (EDFA) and a Variable Optical Attenuator (VOA) (VOA 1) regulate the input power to the PNN device so that it never exceeds 3 dBm in order not to damage it and not to trigger nonlinear effects in Silicon.

The PNN is placed before the fiber to eliminate polarization-dependent losses, and its position within the transmission line makes it, in fact, a pre-compensator for CD effects. Light enters and exits the chip in an edge coupling configuration via two tapered fibers held by a 3-axis motion piezo-controller, providing 4.9 dB of coupling losses. The 0.1% of the collected optical power is sent to a Power Meter (PM 1) for alignment monitoring. The PNN processes the optical signal according to Equation 3.12, with the weights driven by a DC current generator with 16 independent channels. Since the PNN optical response is defined but for one of the phase weights $e^{j\phi_i}$, the Phase Shifter (PS) associated with the null-delay channel (Channel 1) is left unconnected and chosen as reference. Therefore, only 15 of the 16 channels of the DC current generators are used to tune a complete PNN device, while the remaining channel is used as a triggering signal for the oscilloscope. The currents in the PNN are held on for 220 μ s to allow the local thermal relaxation of the micro-heaters and then immediately switched off to avoid the thermalization of the whole chip with consequent cross-talk between the different weights. The PNN device is soldered to a Printed Circuit Board (PCB) and electrically connected via wire bonding. A Copper plate in direct thermal contact with the PNN device is placed underneath the PCB and connected to a Temperature Controller for thermal stabilization (see Figure A.4).

The remaining 99.9 % of the PNN optical power transmission is sent to an EDFA (EDFA 2) to compensate for the PNN insertion losses. Their amount depends on the selected weight configuration, with values ranging between 17 dB and 21 dB after the training procedures. Then, the optical signal is coupled to a first fiber span with a variable length between 0 and 75 km with steps of 25 km. Larger propagation distances are reached by implementing another 25 km or 50 km fiber span preceded by another EDFA (EDFA 3) to prevent Optical Signal-to-Noise Ratio (OSNR) degradation. Standard SM G.652D fibers have been used (0.2 dB/km losses), with the input optical power to each span never exceeding 2 dBm to prevent the triggering of nonlinear effects. The fiber link output optical power is attenuated by a VOA (VOA 2) and, via an optical switch, is

Instrument	Ref. in Figure A.1	Ref. in Figure A.3
PurePhotonics PPCL200	TLS	TLS
Keysight M9502A	AWG	AWG
IXBlue DR-AN-20-HO	RF amplifiers	RF amplifiers
IXBlue MXIQ-LN-30	NMZI	NMZI
Newport F-POL-APC	PC	PC
IPG Photonics EAD-1K-C	EDFA 1	EDFA 1
Viavi mVOA-C1	VOA 1	VOA 1
mOPM-C1	PM 1,2	PM 1,2
Optilab EDFA-PA-MSA	EDFA 2, 3, 4	EDFA 2, 3, 4
Thorlabs V1550A	VOA 2, 3	VOA 2, 3
OZ-Optics TF-100-11	TOF	TOF
Thorlabs DXM20AF	RX 1,2	RX 1,2
National Instrument NI 9266	DC generator	DC generator
Teledyne LeCroy SDA816Zi-A	OSC	OSC
mOSW-C1	SWITCH	-
Yokogawa AQ6360	OSA	-
Teraxion ClearSpectrum TDCMB	-	TDC
Agiltron NanoSpeed Ultra-Fast 1×2	-	UFS 1,2
GMSVOA MEMS VOA	-	VOA 3

TABLE A.1: Manufacturing company and model of the instruments used in the experimental setups described in Figure A.1 (linked to the results of Chapter 4) and Figure A.3 (linked to the results of Chapter 5).

addressed alternatively to an Optical Spectrum Analyzer (OSA) or to a final amplification stage (EDFA 4), which compensates for fiber losses. Subsequently, another VOA (VOA 3) regulates the amount of optical power sent to a 30 GHz Tunable Optical Filter (TOF) and then to the receiver. A fiber splitter routes part of the signal towards a second Power Meter (PM 2, 0.1%) and part to another 20 GHz-bandwidth fast photodiode (RX 2, 99.9%). Both RX 1 and RX 2 are connected to an oscilloscope (OSC) with a 16 GHz bandwidth and a maximum sampling frequency of 80 GSa/s. The acquired trace is then processed via the terminal, evaluating the separation loss function and the Bit Error Rate (BER) according to the same modalities described in Section 4.1.1.

A.1.1 Experimental conditions

In case of a 10G 2-level modulation, at the end of the line, the OSNR for a Power at the Receiver (PRX) of -3 dBm at RX2 is maintained at (39 ± 1) dB (0.1-nm resolution bandwidth) for the whole measurement process. Since the optical gain of the end-of-line EDFA is left unaltered for the whole experiment duration, the mentioned condition at the receiver reflects into a (29 ± 1) dB OSNR (0.1-nm resolution bandwidth) and (-18 ± 1) dBm optical power at 1550 nm measured at the OSA in its setup placement of Figure A.1. This condition is chosen as the reference working point for these experiments, and it is also maintained for the different tested fiber lengths. The same power and OSNR conditions must be set even when the PNN device is inserted in the line. Indeed, this ensures a fair comparison between equalized and unequalized results, making them an indicator of the actual PNN performance. In view of this, when the chip is installed in the line, the aforementioned conditions measured by the OSA are reproduced while keeping all currents at 0 mA in the device. Some deviations from this working point continuously occur during the training, since the insertion loss of the device significantly changes with the current configuration set in each specific measure. Unfortunately, these changes cannot be monitored directly by the OSA due to the too-low duty cycle of its measurement. Regardless of this, a noise characterization of the experimental setup revealed that these variations in PRX and OSNR minimally impact the Signal-to-Noise Ratio (SNR) profile at the receiver; thus, the consistency between the measurements performed with and without the PNN is preserved. The same concepts apply when switching to different signal baud rates or modulation formats, with case-specific reference points selected for each scenario.

A.1.2 Eye diagram

In the context of telecommunications, a typical indicator of the quality of an optical Intensity Modulation/Direct Detection (IMDD) signal and system performance is represented by the eye diagram. It consists of a visual representation of a digital signal in the time domain obtained by superimposing in a specific time window $T_w = 2/B$, with B being the baud rate, all the segments of that duration that compose the signal. Figure A.2 illustrates this concept, starting from a trace acquired at the receiver (panel (a)). Trace portions of duration T_w marked with different colors are overlapped during the eye diagram construction (panel (b)), whose final shape resulting from the folding of the whole trace appears as in panel (c). In a wide-open diagram, all the possible transitions between the optical levels appear, with regions with a brighter color indicating a higher density of lines. Overall, they describe an eye-like shape, and from here, the name.

Eye diagrams help in assessing signal integrity by providing information on issues like jitter, noise, and Inter-Symbol Interference (ISI) [218]. In particular, the eye width is related to timing margin or jitter tolerance, where a wider eye means the system can tolerate more timing variation without errors. The height of an eye diagram reflects

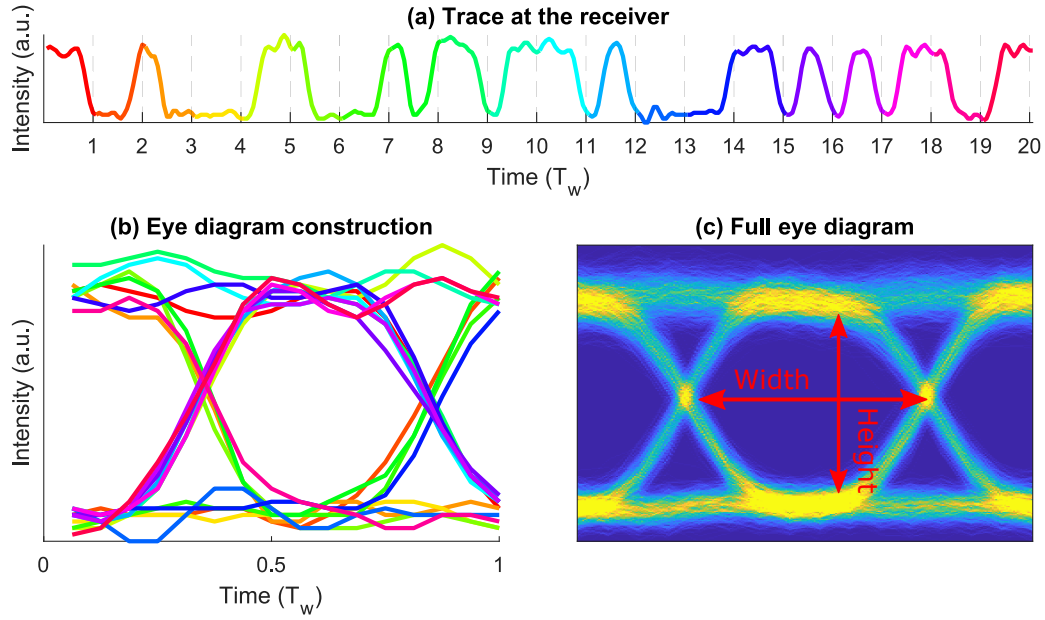


FIGURE A.2: Illustration of the eye diagram construction and main features. Considering a trace at the receiver (panel (a)), different segments marked in colors are overlapped over a time window $T_w = 2/B$, where B is the baud rate (panel (b)). The final result is presented as a heatmap (panel (c)), with brighter regions marking a higher density of lines.

the signal's voltage margin, noise tolerance, and the presence of ISI, with a taller eye indicating a stronger, less noisy signal.

A.2 Fiber loop implementation

Figure A.3 illustrates a schematic of the complete experimental setup used to produce the results of Chapter 5 in the context of CD and Self-Phase Modulation (SPM) equalization. Manufacturing companies and models of each instrument appear in Table A.1. Components and modalities used for signal generation, from the TLS to the first VOA (VOA 1), and acquisition, from the last amplifier (EDFA 4) to the receiver stage, are shared with the previous setup. Differences appear right after VOA 1, where, in this version, light is directly injected in the propagation stage realized via a fiber loop. This is implemented via two Ultrafast optical Switches (UFSs, rise/fall time of 8 ns), which allow to fill and close the loop for propagation and, later, to empty it and direct light to the PNN. At every loop cycle, light propagates in a 50 km standard SMF-28 fiber span (corresponding to a propagation time of $\tau_F \sim 250 \mu s$), preceded by an EDFA (EDFA 2) that amplifies the input power in the fiber P_{in} up to 9 dBm. A Tunable Dispersion Compensator (TDC) (max dispersion: -900 ps nm^{-1}) can be used to remove the accumulated CD, and a VOA (VOA 2) is used to fine-tune the optical gain in a single loop (discussed below). The total propagation distance (i.e., the number of loop cycles) is

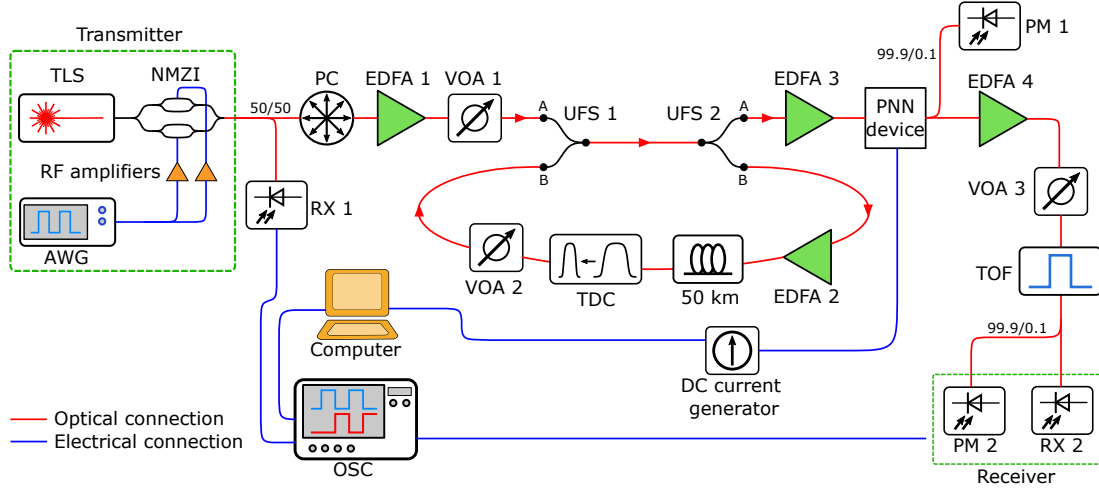


FIGURE A.3: Experimental setup implementing the optical fiber loop and used to produce the results of Chapter 5 in the context of CD and SPM equalization. TLS: Tunable Laser Source; NMZI: Nested Mach-Zehnder Interferometer; AWG: Arbitrary Waveform Generator; RF: Radio Frequency; PC: Polarization Controller; EDFA: Erbium-Doped Fiber Amplifier; VOA: Variable Optical Attenuator; UFS: Ultra-Fast Switches; TDC: Tunable Dispersion Compensator; PNN: Photonic Neural Network; PM: Power Monitor; TOF: Tunable Optical Filter; OSC: Oscilloscope; RX: Fast Photodiode.

set by properly delaying the change of state in the switches.

After the propagation, the distorted optical pulses reach an intermediate amplification stage (EDFA 3) and proceed to the PNN for optical processing. The coupling to the chip, weights control, and oscilloscope triggering are performed according to the same modalities described in Section A.1. In the fiber loop configuration, during the measurements for unequalized transmission, the PNN is removed from the setup and replaced with a VOA (not depicted in Figure A.3) set to 20 dB attenuation to mimic the device losses and maintain the overall OSNR budget. After the PNN, a fused coupler with a 99.9/0.1 splitting ratio sends a minor fraction of the optical power to a Power Monitor (PM 1) to constantly check the device's alignment. The remaining fraction of optical power proceeds through EDFA 4, which compensates for the insertion losses, and then to the receiver stage. The acquired trace is then processed via the terminal, evaluating the separation loss function and the BER according to the same modalities described in Section 4.1.1.

A.2.1 Measurement procedures

A typical measurement procedure that leads to the separation loss function and BER evaluation occurs in different steps. Following the nomenclature appearing in Figure A.3, at the beginning of each measurement, the switches are set in the $A \rightarrow A$ configuration, which is used for Back-To-Back (BTB) propagation. In this status, the loop is

completely isolated from the BTB propagation line, and the EDFA 2 is left in an amplified spontaneous emission (ASE) condition. Light from the transmitter is then diverted into the loop when the switches are set in configuration $A \rightarrow B$. EDFA 2 goes from an ASE condition to an active amplification of the incoming signal, which is associated with a transient in the device response [219]. Therefore, the $A \rightarrow B$ configuration is maintained for approximately $5\tau_F$, with $\tau_F \sim 250 \mu\text{s}$. During this time, light propagates through the dead-end loop and the EDFA's dynamics stabilize. Subsequently, the recirculation is enabled by switching to $B \rightarrow B$ configuration, which is held for an entire multiple of τ_F . After the propagation, the switches are set in $B \rightarrow A$ configuration, emptying the loop and directing light towards the PNN for further processing. The measurement procedure relies on precise timing relations between triggering the switches and obtaining a controlled propagation, setting the PNN weights, and triggering the oscilloscope.

During the propagation in the loop, EDFAs 3 and 4 are left in an ASE condition, thus present transients in the gain upon the arrival of the signal exiting the loop [192]. In view of this, in each transmission scenario, the gain inside the loop has been fine-tuned to reach RX with the same reference power set by the value measured after a 0 km propagation, thus to compensate for the mentioned transients. Overall, this reflects in small changes of the input power in the fiber at every loop, which is thus not maintained at the value set by P_{in} . A subsequent setup calibration revealed that, for example, in the worst-case scenario of $P_{in} = 9 \text{ dBm}$ and $n_L = 9$, the input power in the fiber decreases by a total maximum of 4 dB across the 9 loops, thus still preserving the nonlinear propagation regime.

A.3 Chip characterization

The chip presented in Figures 3.7(a) and A.4 has been fabricated by the Applied Nanotools Foundry [175] on a Silicon-On-Insulator platform, for a total footprint of $2.8 \times 2.3 \text{ mm}^2$. Its waveguides feature a $220 \times 500 \text{ nm}^2$ cross-section and are surrounded by a Silica cladding. Light is inserted into the device via edge coupling, featuring losses of $4.9 \pm 0.1 \text{ dB}$ (measured by D. Bazzanella, with 68% confidence). Spirals are present in three lengths, namely $L_A = 7.10 \text{ mm}$, $L_B = 3.55 \text{ mm}$, and $L_C = 1.77 \text{ mm}$, with an estimated propagation loss of $-1.7 \pm 0.2 \text{ dB/cm}$ (measured by D. Bazzanella, with 68% confidence). Y-branches provide for the splitting and combination of optical signals, characterized by a $(-0.07 \pm 0.08) \text{ dB}$ excess loss and a $(0.1 \pm 0.2) \text{ dB}$ imbalance (measured by D. Bazzanella, with 68% confidence). The phasors are implemented as thermal heaters, consisting of $300 \mu\text{m}$ -long TiN wires placed on top of the Si waveguides. The heaters are connected via integrated $20 \mu\text{m}$ -thick metal wires to $100 \mu\text{m}$ -wide electrical pads located at the top and bottom edges of the chip (Figure A.4(a)). Golden filaments implemented via a wire bonding technique link these pads to the electrical connections of the PCB on which the chip is hosted (Figure A.4(b)). The wire bonding has been performed by the Bruno Kessler Foundation in Trento, while the

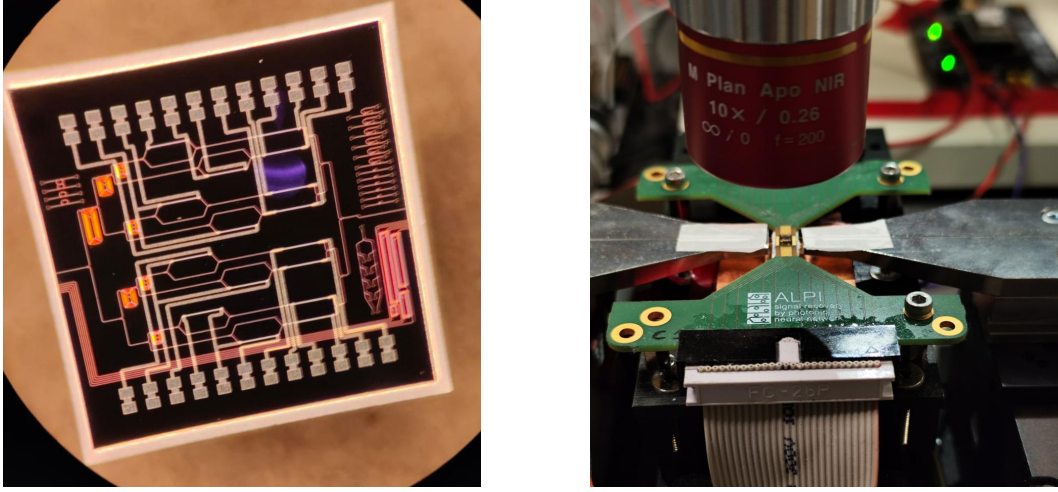


FIGURE A.4: Implementation of the 8-channel time-delayed complex perceptron. (left) Details of a bare chip. Pink paths are the waveguides, yellowish paths the metal wires used for electrical connections between heaters (buried) and pads. Test structures surround the perceptron. (right) Full implementation on the hosting Printed Circuit Board.

PCB has been made by the electronic design service of the University of Trento.

Considering the Mach-Zehnder Interferometers (MZIs) providing for the amplitude weights a_i , the current values $I_{0,i}$ and $I_{1,i}$ corresponding to minimum and maximum aperture are experimentally determined. The intermediate values for a generic current $I_{MZI,i}$ are extrapolated from the typical response of a MZI as

$$a_i = \left| \cos \left(\frac{I_{MZI,i}^2 - I_{1,i}^2}{I_{0,i}^2 - I_{1,i}^2} \frac{\pi}{2} \right) \right|.$$

Overall, the electrical power that generates a π phase shift is measured to be, in terms of average and standard deviation, (24 ± 2) mW. For the sake of completeness, we recall from Section 3.4.2 that, during a measurement session, the electrical connection controlling the MZI of the 5th channel accidentally broke, thus preventing its further tuning. A prior calibration showed that $I_{1,5} \approx 0$, thus applying no current to the 5th MZI corresponds to a fully open configuration, namely, to a fixed amplitude weight $a_5 = 1$.

Proceeding, a finer characterization has been performed for the PSs, determining experimentally the coefficients α_i , β_i , and γ_i appearing in Equation 4.12 and reported in Table A.2. Data are obtained by fitting the interference curves produced during a scan over the current of each phasor.

Moreover, the fixed amplitude weights k_i for each channel appearing in Equation 3.12 have been extrapolated. The estimation begins with a condition of entirely constructive interference among all the channels, determined by setting $a_i = 1$ for every i

Channel #	α_i	β_i	γ_i	$k_i \times 10^3$	Channel Loss (dB)
1	-	-	-	12 ± 7	-19.1 (+1.9, -3.5)
2	0.032	2.09	1.01	20 ± 7	-17.0 (+1.2, -1.7)
3	0.025	2.70	1.06	20 ± 2	-16.9 (+0.5, -0.5)
4	0.010	0.00	1.24	21 ± 2	-16.9 (+0.4, -0.4)
5	0.030	1.92	1.00	31 ± 2	-15.1 (+0.2, -0.2)
6	0.034	1.11	1.01	17 ± 5	-17.8 (+1.1, -1.5)
7	0.022	2.64	1.10	15 ± 3	-18.2 (+0.9, -1.1)
8	0.018	4.50	1.11	11 ± 6	-19.6 (+1.9, -3.3)

TABLE A.2: Results of the PNN characterization procedure in terms of phase shifters' response to current as described in Equation 4.12 and channel losses k_i (linear and dB values) entering in Equation 3.12. Some parameters for Channel 1 are not reported since the channel is chosen as a reference for phase shift evaluation, thus the corresponding phase shifter is left unconnected during the measurements. Values for α_i , β_i , and γ_i are truncated to the third non-null digit and provided with no error due to the lack of statistics on these fit parameters. A description of the measurement procedures leading to k_i and their uncertainties is provided in the text.

and training the phase weights to maximize the output signal. Then, while maintaining these fixed phase weights, the channels were progressively closed, namely, we set $a_1 = 0$, then $a_1 = a_2 = 0$, and so on, measuring the output optical power after each closing. Notice that the 5th channel always remained open, due to the broken electrical bond. The experimental data were then provided to a PNN device model to fit the contributions of the individual channels. In the model, the actual circuit of the PNN (shown in Figure 3.7) was used. The same procedure has been performed by closing the channels in progressive order (from the 1st to the 8th) and in reverse order (from the 8th to the 1st), respectively returning the estimates k_i^P and k_i^R . Table A.2 reports the final values for k_i , obtained as the average of the estimates obtained from the two measurements. The large uncertainties on some values of k_i reflect the differences between the corresponding k_i^P and k_i^R . These discrepancies, in turn, can be caused by a non-perfect constructive interference condition, which can indeed trigger a dependence of the k_i estimates on the specific closure order of the channels.

Appendix B

Simulated Setup

B.1 Code block diagram

Figure B.1 presents the block diagram of the setup reproduced via the simulation software. The proper parameters of each block and their used value or range of values are reported in Table B.1. The simulations happen in two phases: first occurs the propagation phase (orange blocks), where the optical signals are generated and propagated through the fiber loop. Second, the processing phase (blue blocks) occurs, where the propagated sequences are accessed, treated by the PNN, and detected, leading to the evaluation of the loss function.

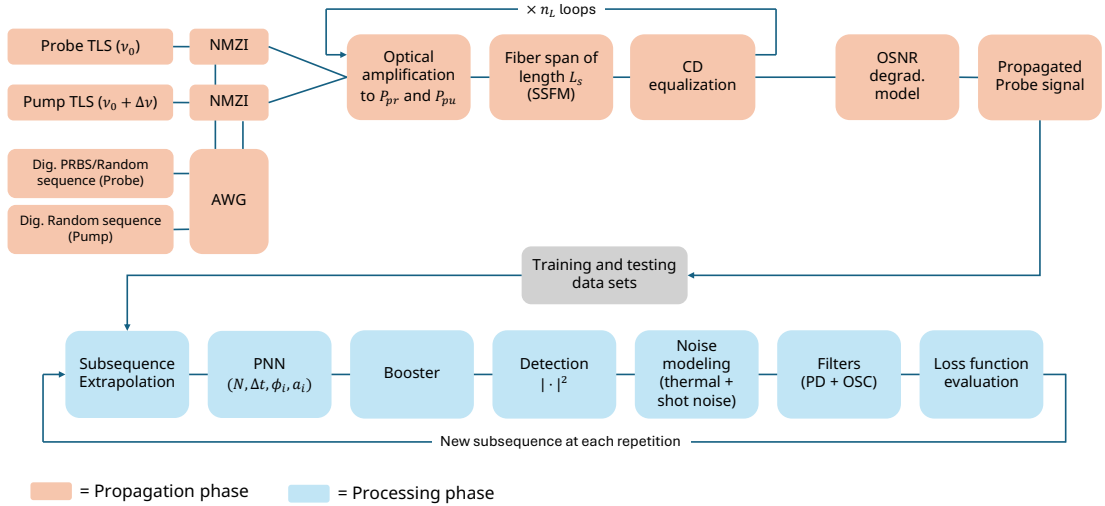


FIGURE B.1: Block scheme for the simulated setup. A complete walk-through of the different blocks is provided in the text. TLS: Tunable Laser Source; AWG: Arbitrary Waveform Generator; NMZI: Nested Mach-Zehnder Modulator; SSFM: Split-Step Fourier Method; CD: Chromatic Dispersion; OSNR: Optical Signal-To-Noise Ratio; PNN: Photonic Neural Network; PD: Photodiode; OSC: Oscilloscope.

In the propagation phase, two TLSs operating at the frequencies ν_0 and $\nu_0 + \Delta\nu$ generate the Probe and Pump CW signals, respectively. These are then modulated via two NMZIs driven by an AWG with bandwidth B_{AWG} , resulting in a signal baud rate B . In our specific 100G application, the probe is encoded with a 2^{17} bit-long sequence, which can be a PRBS of order 17 or a random sequence, respectively used to produce the training and testing data sets. On the other hand, the pump always encodes a random sequence of the same length and bitrate. A very high sampling frequency f_{prop} is imposed on the two signals in this first phase to ensure accuracy during propagation. This is realized via a simulated fiber loop represented by a 3-block sequence, which is repeated for n_L times to mimic a multi-span propagation. An optical amplification stage implemented as a rescaling factor within the loop restores the optical power to P_{pr} and P_{pu} at every recirculation. The propagation through a fiber span of length L_s is reproduced via the Split-Step Fourier Method (SSFM) [92]. The nomenclature for the SSFM parameters proposed in the text and Table B.1 follows the one presented in Section 2.2 and [220]. The subsequent CD equalization stage is described by the operator $\exp\{-jL_s\beta_2\omega^2/2 - jL_s\beta_3\omega^3/6\}$ in the frequency domain, representing the inverse of the linear propagation term in a lossless fiber. The pump signal is discarded at this point since it only co-propagates with the probe and triggers the Cross-Phase Modulation (XPM) effect. On the other hand, the probe signal is added with optical noise according to the model proposed in [19, 50] to account for the OSNR degradation imposed by the optical amplification at each loop. In particular, after a n_L loop propagation, the probe signal's spectrum is added with Gaussian noise to match the OSNR condition predicted by

$$OSNR = P_{out} - l_s - NF - 10\log_{10}(n_l) - 10\log_{10}(h\nu_0\Delta\nu_{bw}). \quad (B.1)$$

Here, P_{out} is the output power from the amplification stage (i.e., P_{pr}), l_s is the loss accumulated in a loop as the sum of the contributions from fiber, TDC, VOA 2, and switches, NF is the noise figure of the amplifiers, $\Delta\nu_{bw}$ is the bandwidth chosen for OSNR evaluation, and h is the Planck constant. The so-obtained propagated probe signal is stored for future utilization, and this step ends the propagation phase. The process described so far is repeated for different transmission scenarios in terms of P_{pr} , P_{pu} , and propagation distance $L = L_s \times n_L$, thus creating the training and testing data sets.

Subsequently, the processing phase starts, consisting of a whole run through the steps described in the blue rectangles of Figure B.1. Each run leads to a loss function evaluation, and thus, it is iteratively repeated in the context of PNN training or testing. Each run starts by accessing a stored propagated sequence and extrapolating a sub-sequence according to the procedures described in Figure 5.7. The next block emulates the optical processing of a PNN device, adapting the computational modeling to the different designs adopted in this work. An insight into the modeling of each device is treated in Section B.2. The resulting optical sequence is then applied with a rescaling factor to boost the optical power to the desired PRX (during the testing phase) or to emulate a constant-gain amplifier (during the training phase). After the detection, implemented with the square-modulus operation and a conversion gain R_{PD} (Watts-to-Volts),

the resulting signal $|s|^2$ (expressed in mV) is added with Gaussian noise according to the model suggested in [178]. Its variance is given by a thermal σ_T and a shot noise σ_s contribution, namely

$$\sigma^2 = \sigma_s^2 |s|^2 + \sigma_T^2. \quad (\text{B.2})$$

The code calibration with 10 Gbaud signals provides the model parameters' value. The noise-added signal undergoes a double filtering process implemented with a 5th order Bessel Filter accounting for the photodiode bandwidth B_{PD} and the oscilloscope bandwidth B_{OSC} . A down-sampling at a frequency f_{osc} is then operated, together with an n_{vert} -bit vertical quantization of the full scale V_{max} . The 2-sample Separation Loss function and BER are finally evaluated on the so-obtained output sequence.

B.2 PNN devices computational modeling

Each tested PNN architecture is associated with a specific function that models its optical processing. All the functions receive as input an optical sequence $x(t)$, namely, a complex time series, generally resulting from fiber propagation described by Equation 2.33. Other inputs to the function are the complex weights that set the working conditions for the device, and the output represents the processed optical sequence. Notice that Equation 2.33 is never used to simulate signal propagation inside the chips, since effects deriving from Group Velocity Dispersion (GVD) can be neglected for such short optical paths.

B.2.1 Time-Delayed Complex Perceptron

A device characterized by N taps and a unitary delay Δt is modeled starting from Equation 3.12. First, N copies of the input signal $x(t)$ are created as columns of a matrix X , and they are mutually shifted to reproduce the propagation through the delay lines. A matrix multiplication is then performed between the rows of X and a column of N complex weights w_i , thus joining in a single operation the weighting of each channel and the summation stage. The weights array is obtained as

$$w_i = a_i e^{j\phi_i} k_i \quad \text{with} \quad i = 1, \dots, N.$$

Here, $a_i = [0, 1]$ and $\phi_i = [0, 2\pi)$ account for the amplitude and phase weights introduced by the MZIs and heaters in each channel. Then, k_i accounts for the factors $1/\sqrt{2}$ assigned at each splitting/recombination between two waveguides, as well as the amplitude weights deriving from the propagation through the delay lines. These are evaluated from the propagation loss coefficient $\alpha_s = -2$ dB/cm, typical for Si-integrated waveguides.

B.2.2 Mach-Zehnder cascade

The device is modeled as a series of N units representing unbalanced MZIs. In each of them, a 2×2 Multi-Mode Interferometer (MMI) is described by the matrix

$$M_{2 \times 2} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & j \\ j & 1 \end{bmatrix},$$

sending the optical signals in the upper and lower arms. Considering the i th unit, with $i = 1, \dots, N$, the signal traveling in the upper arm is delayed by Δt_i to account for the propagation in the spiral, and applied with an amplitude weight derived from the propagation losses ($\alpha_s = -2$ dB/cm). Simultaneously, the signal traveling in the lower arm is applied with a phase weight $e^{i\phi_i}$. The two signals reach another 2×2 MMI where they are recombined and sent to the subsequent unit.

B.2.3 Photonic reservoir with all-optical readout

PNN implementations similar to that of Figure 6.1 are constituted by two stages: the first represents the Photonic Reservoir (PR), constituted by N_n nodes with $N_p = 3$ input and output ports, and an optical readout stage, constituted by N_c channels, then combined in a summation stage.

The input frame $x(t)$, represented by a complex time series of N_f samples, is sent to the input nodes of the PR, from which the propagation across the whole structure starts. After the signal injection, the PR is described at each time instant t by the optical signals outgoing from the ports of each node. These sequences may then become inputs for other nodes or reach the readout, depending on the chosen architecture. The information on the PR is stored in a $N_n \times N_p \times N_f$ matrix, where each entry $M_{n,p}(t)$ represents the optical sample found in correspondence of node $n \in [1, N_n]$ at the output port $p \in [1, N_p]$ at time $t \in [1, N_f]$. All the matrix entries are initially set to 0, and then progressively modified as the input frame is fed to the PR. Notice that $x(t)$ is composed of three segments sharing the same modulation (PAMn): a header sequence x_h , a footer sequence x_f , and a pilot sequence x_p , which is encapsulated between the other two, and represents the only segment used for transmission quality evaluation. The random nature of the header and the footer serves the purpose of filling up and emptying the PR, ensuring that the pilot sequence experiences unbiased optical processing through the structure. The length of x_h and x_f can be arbitrarily varied depending on the estimated memory for the PR.

At each time instant, the new state $M_{n,p}(t')$ of each port is evaluated by feeding the corresponding node n with the proper delayed contributions, as described in Figure B.2.

Each node is modeled as an $N_p \times N_p$ matrix,

$$M_{\text{node}} = \begin{bmatrix} M_{2 \times 2} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & M_{2 \times 2} \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{j}{\sqrt{2}} & 0 \\ \frac{j}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & \frac{j}{\sqrt{2}} \\ 0 & \frac{j}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{j}{2} & -\frac{1}{2} \\ \frac{j}{\sqrt{2}} & \frac{1}{2} & \frac{j}{2} \\ 0 & \frac{j}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

(where $j = \sqrt{-1}$) which is obtained as a combination of the $M_{2 \times 2}$ matrices describing the action of MMIs constituting the node itself. The new output of the n -th node corresponds to

$$\begin{bmatrix} M_{n,1}(t') \\ M_{n,2}(t') \\ M_{n,3}(t') \end{bmatrix} = M_{\text{node}} \begin{bmatrix} M_{a,b}(t' - \Delta t_{ab}) e^{j\phi_{ab}} l_{ab} \\ M_{c,d}(t' - \Delta t_{cd}) e^{j\phi_{bc}} l_{cd} \\ M_{d,e}(t' - \Delta t_{ef}) e^{j\phi_{de}} l_{ef} \end{bmatrix} \quad (\text{B.3})$$

where the indices (a-f) are determined depending on the chosen topology. Considering, for example, the first row, this formulation expresses the fact that a connection from a sending port b of node a to the receiving port 1 of node n contains a spiral providing for the delay Δt_{ab} and a thermal heater driving an arbitrary phase weight ϕ_{ab} . Spirals induce losses too, which are accounted for via the weights l_{ab} , overall producing the connection weight $w_{ab} = e^{j\phi_{ab}} l_{ab}$. The architecture, therefore, determines the connection map between the ports and the features specific to each link. At each time step, the update described in Equation B.3 must be performed for every $n = 1, \dots, N_n$. After the propagation in the PR, the output of N_c selected ports is sent to the readout stage, where the processing is operated according to the same modalities described for the weighting and summation stage of a Time-Delayed Complex Perceptron (TDCP) (Section B.2.1).

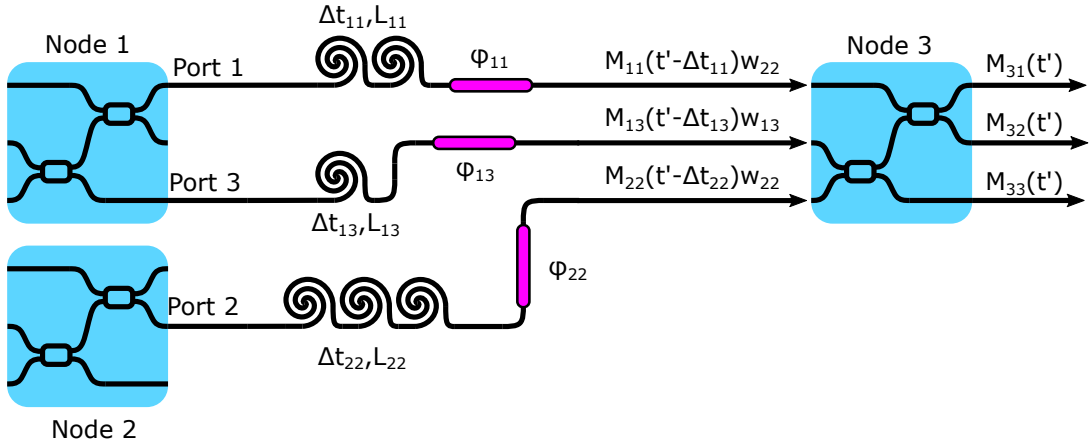


FIGURE B.2: Illustration of the operations performed during the propagation of optical signals through the PR in a simulated environment. Output signals from different ports of nodes 1 and 2 are delayed and weighted during the propagation through the links to node 3. This is reached simultaneously at time t' by the optical samples $M_{xy}(t' - \Delta t_{xy})w_{xy}$ (with xy specified in the picture), which are elaborated by the matrix M_{node} to produce the output $M_{3z}(t')$.

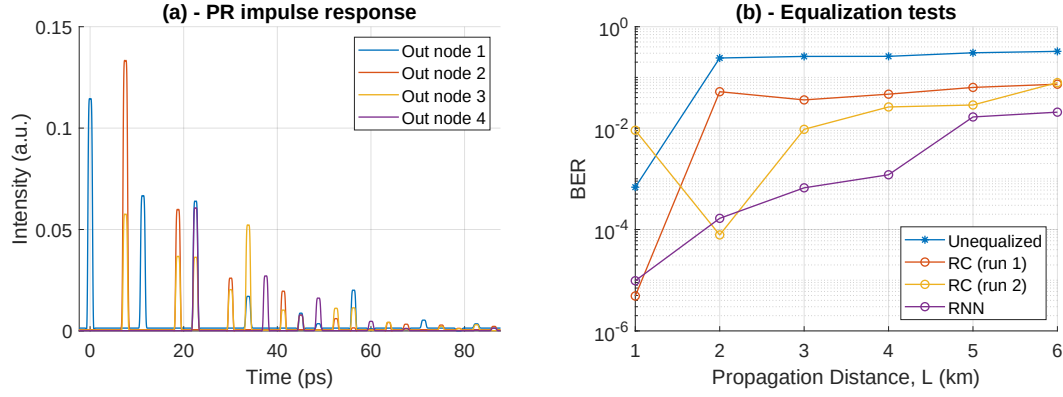


FIGURE B.3: (a) Time response at each node of the PR design (Figure 6.1) to a 1 ps input pulse entering the structure at $t = 0$. The traces are normalized to the peak of the input pulse (not shown in the plot). (b) BER versus PRX profiles as a function of the propagation distance for CD equalization. The reference is set by the unequalized profile (blue curve), while the equalized ones are obtained using the PNN in a Reservoir Computing approach (RC, orange and yellow lines) or in a Recurrent Neural Network approach (RNN, purple line). The profiles for RC correspond to two different sets of random internode phase weights (run 1 and 2).

This code was used for the optimization of the layout presented in Figure 6.1, targeting 100G applications against CD. In reference to that design, Figure B.3(a) presents the PR impulse response obtained by injecting a 1 ps pulse into it and collecting the corresponding signal at the output ports of each node. It is possible to appreciate the presence of fading memory, which is induced by the internode spirals, providing for different delay contributions, and overall network topology, providing for recurrence. Finally, Figure 6.1(b) presents the results of CD equalization in terms of BER values at PRX = 0 dBm as a function of the propagation distance L , in the same conditions associated with the results of Figure 5.8. The reference is set by the unequalized profile (blue curve). The orange and yellow curves are obtained using the PNN in a Reservoir Computing (RC) approach, namely, training the readout weights only, while the internode phases are set to random values, with different sets for the two runs. As it appears in panel (b), when exploiting the RC paradigm, the PNN performances are strongly influenced by these weights, which make the device suitable for applications at specific L , but compromise its performance in other scenarios. In this sense, the readout weights cannot compensate for the bias set by the internode phases. Better overall equalization is achieved by training all the available parameters, thus exploiting the Recurrent Neural Network approach (RNN, purple line).

B.3 Simulation parameters

Parameter Description	Symbol	Value
Baud rate	B	10 to 100 Gbps
Probe Laser Frequency	ν_0	193.55 THz
Detuning between probe and pump laser sources	$\Delta\nu$	50 GHz
Sampling frequency for propagation phase	f_{prop}	$B \times 80$
Arbitrary Waveform Generator bandwidth	B_{AWG}	$3 \times B$
Losses in a single loop round-trip	l_s	$16 \text{ dB} + L_s \times \alpha_{dB}$
Noise figure in OSNR degradation model	NF	5 dB
Bandwidth for OSNR evaluation	$\Delta\nu_{bw}$	$1.2 \times B$
Photodiode bandwidth	B_{PD}	$2 \times B$
Photodiode Conversion Gain	R_{PD}	23 V W^{-1}
Oscilloscope sampling frequency	f_{osc}	$B \times 8$
Oscilloscope vertical full scale	V_{max}	160 mV
Oscilloscope vertical resolution	n_{vert}	8 bits
Oscilloscope bandwidth	B_{osc}	$1.6 \times B$
Number of loops	n_L	0 to 9
Fiber span length	L_s	1 to 50 km
Fiber dispersion	D	$18 \text{ ps nm}^{-1} \text{ km}^{-1}$
Group velocity dispersion coefficient	β_2	$-0.022 \text{ ps}^2 \text{ m}^{-1}$
Third-order dispersion coefficient	β_3	$0.1 \text{ ps}^3 \text{ km}^{-1}$
Effective mode area	A_{eff}	$80 \times 10^{-12} \text{ m}^2$
Kerr coefficient	n_2	$2.7 \times 10^{-20} \text{ m}^2 \text{ W}^{-1}$
Fiber Attenuation constant	α_{dB}	0.2 dB km^{-1}
Split-Step Fourier Method propagation step	h_s	100 m (100G) - 1000 m (10G)
Thermal noise contribution	σ_T^2	0.65 mV^2
Shot noise contribution	σ_S^2	0.05 mV
Delay lines losses per unit length	α_s	2 dB cm^{-1}

TABLE B.1: Parameters used in the simulated setup. Section B provides a more detailed description of each table entry. The values presented here are associated with the most recent version of the code.

Appendix C

Photonic Integrated Circuits for 100G applications

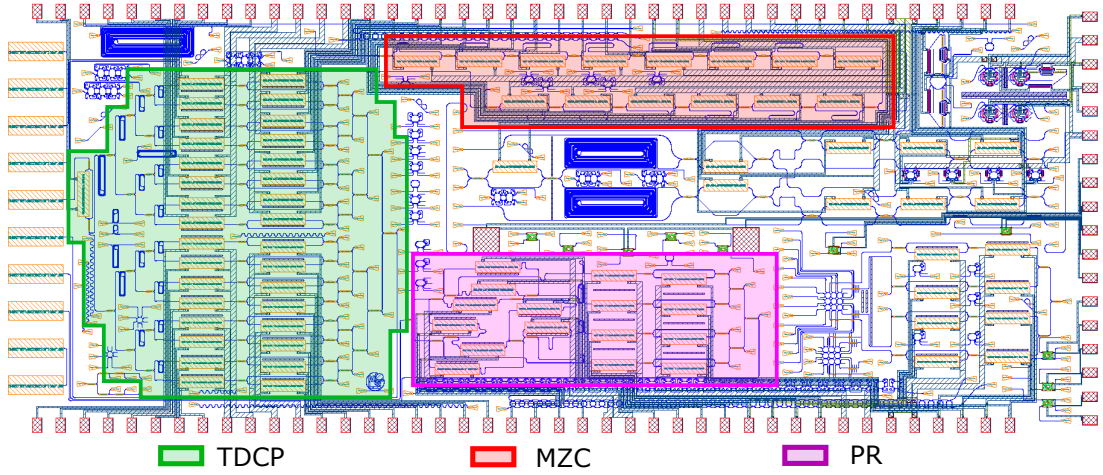


FIGURE C.1: Full circuit design for the new mask layout oriented to 100G applications. The colored frames highlight the Time-Delayed Complex Perceptron (TDC, green), Mach-Zehnder Cascade (MZC, red), and Photonic Reservoir (PR, purple) implementations.

Figure C.1 shows the full circuit design for the new mask layout. The colored frames highlight the three structures mentioned in this work. I designed the TDCP and the PR, while S. Gretter designed the Mach-Zehnder Cascade (MZC). Other structures and the electrical wiring were provided by A. Foradori, S. Biasi, D. Olivieri, A. Baldazzi, and X. Chen.

C.1 Time-delayed complex perceptron

The implementation for the TDCP is shown in Figure C.2. Light enters the structure from the leftmost port and propagates through a tunable MZI. This is realized with a

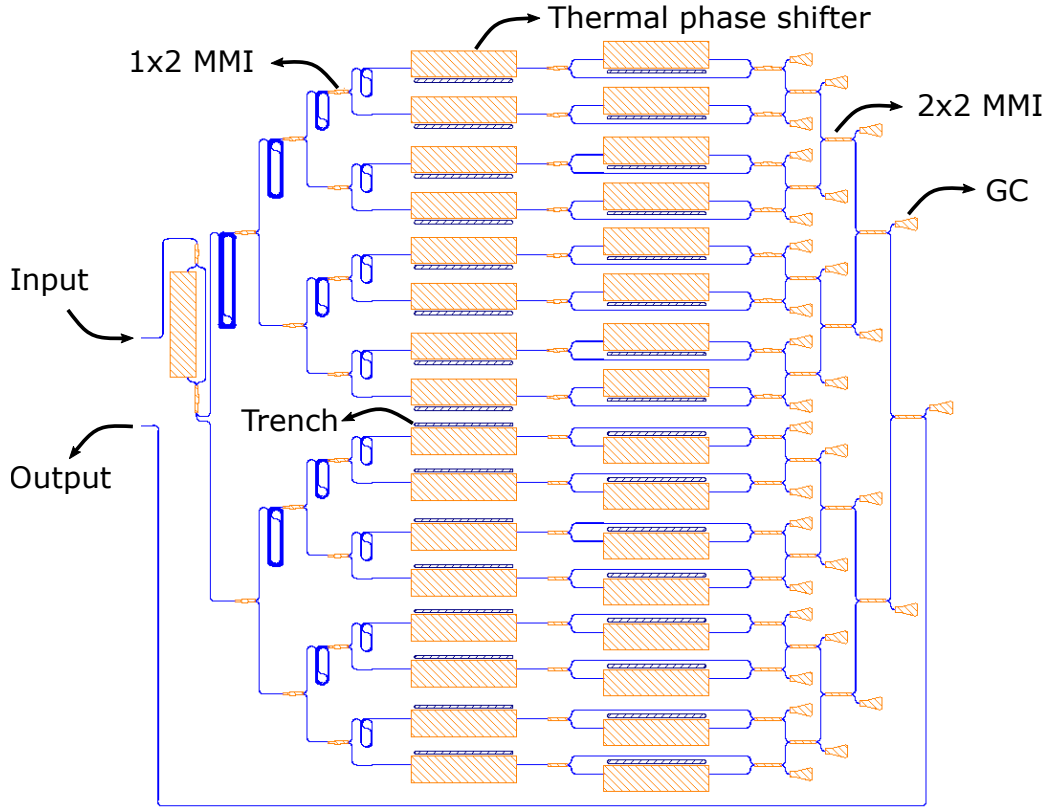


FIGURE C.2: Real mask layout of the reconfigurable 8- or 16-tap TDCP for 100G applications. Spirals provide a relative delay of 5 ps between each tap. Orange blocks are part of the proprietary Process Design Toolkit of the foundry.

1×2 MMI at the input, and a 2×2 MMI at the output, and one of the two arms is equipped with a tunable phase shifter. The structure allows for the selection between two different layouts for the PNN: in a 0:100 configuration, light exits entirely from one of the two output ports, thus allowing the use of the PNN as an 8-tap device. On the contrary, a 50:50 splitting ratio allows the propagation to all the delay lines at disposal, corresponding thus to a 16-tap device. Spirals provide a relative delay of 5 ps between the channels, which are then equipped with amplitude and phase weights controlled via thermal heaters. In the summation stage, channels are progressively recombined via 2×2 MZIs. At each summation, one of the output ports is connected to a grating coupler used as a monitor port, while the other port proceeds to the subsequent summation layer.

C.2 Mach-Zehnder cascade

The implementation for the MZC is shown in Figure C.3. The core of the structure consists of a sequence of sub-units, each of them including an unbalanced MZI: one arm contains a tunable (thermal) phase shifter, while the other incorporates a spiral waveguide. The spiral section appears in three distinct lengths, therefore providing for

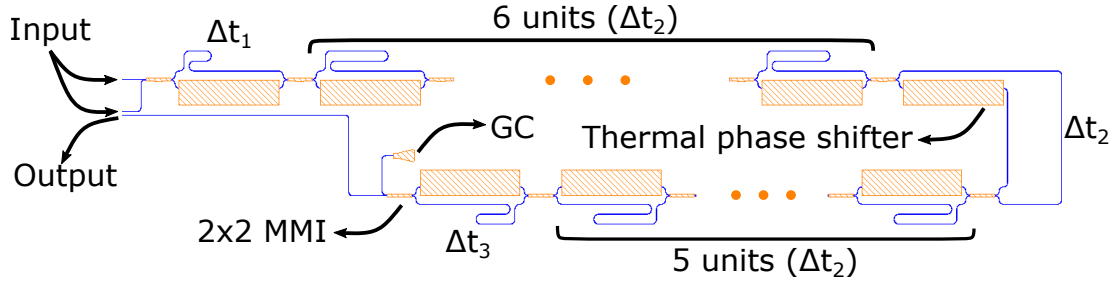


FIGURE C.3: Real mask layout of the 14-unit MZC for 100G applications. Spirals appear in 3 lengths, providing, respectively, a delay of $\Delta t_1 = 2.17$ ps, $\Delta t_2 = 6.25$ ps, and $\Delta t_3 = 4.33$ ps. Orange blocks are part of the proprietary Process Design Toolkit of the foundry.

delays $\Delta t_1 = \Delta t_2/3$ at the input, $\Delta t_3 = 2\Delta t_2/3$ at the output, and $\Delta t_2 = 6.5$ ps across the twelve intermediate units. The splitting and coupling within each unit are processed by 2×2 MMIs. Credits for optimization of the parameters and the design are to S. Gretter.

C.3 Photonic Reservoir with all-optical readout

The implementation for the Reservoir Computing (RC) circuit is shown in Figure C.4. It is composed of a first stage represented by a PR with 4 nodes, and a second stage constituted by 4 parallel channels with amplitude and phase weights, ending up in a summation stage. Light is injected into the structure from the leftmost port and then enters the photonic reservoir in two different nodes. Each node is composed of two MMIs connected to provide a 3×3 structure with directionality, namely, in each node all the ports on the left side are used as input, while those on the right side provide the outputs. Connections between the nodes imply a propagation time of 7.5, 11.25, or 15 ps, and host tunable independent phase shifters. In each node, an output port sends the processed signal to one of the 4 parallel channels of the readout stage, with spirals of

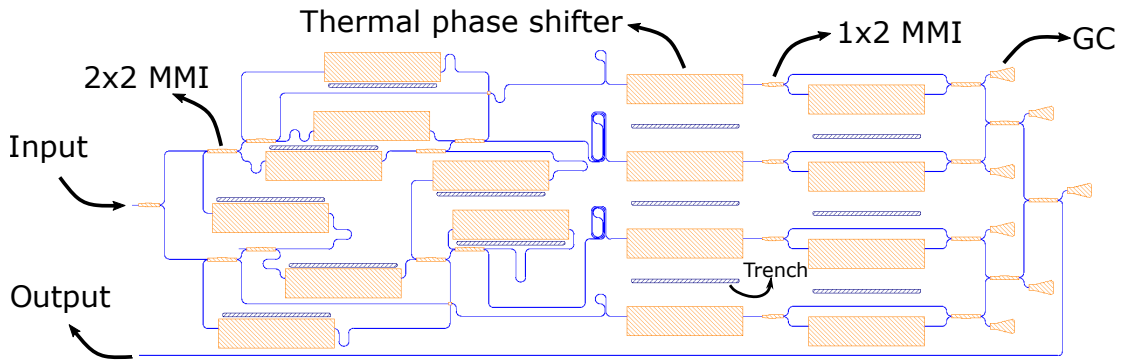


FIGURE C.4: Real mask layout of the PR implementation with 4-channel all-optical readout stage. Orange blocks are part of the proprietary Process Design Toolkit of the foundry.

different lengths inserted to compensate for the length differences in the optical paths to the readout. Amplitude and phase weights are placed in each channel and controlled via thermal phase shifters, finally ending in a summation stage. Analogously to the TDCP, channels are progressively recombined via 2×2 MZIs and, at each summation, one of the output ports is connected to a grating coupler used as a monitor port, while the other port proceeds to the subsequent layer.

Bibliography

- [1] S. Miller. *Optical fiber telecommunications*. Elsevier, 2012.
- [2] J. M. Senior and M. Y. Jamro. *Optical fiber communications: principles and practice*. Pearson Education, 2009.
- [3] H. Kogelnik. “High-capacity optical communications: personal recollections”. *IEEE Journal of Selected Topics in Quantum Electronics* 6.6 (2002), pp. 1279–1286.
- [4] A. E. Willner et al. “Optics and photonics: Key enabling technologies”. *Proceedings of the IEEE* 100.Special Centennial Issue (2012), pp. 1604–1643.
- [5] T. Miya et al. “Ultimate low-loss single-mode fibre at 1.55 μm ”. *Electronics Letters* 15 (4 1979), pp. 106–108.
- [6] R. J. Mears et al. “Low-noise erbium-doped fibre amplifier operating at 1.54 μm ”. *Electronics Letters* 23.19 (1987), pp. 1026–1028.
- [7] E. Desurvire et al. “High-gain erbium-doped traveling-wave fiber amplifier”. *Optics letters* 12.11 (1987), pp. 888–890.
- [8] T. Li. “The impact of optical amplifiers on long-distance lightwave telecommunications”. *Proceedings of the IEEE* 81.11 (2002), pp. 1568–1579.
- [9] Corning. *A sustainable future with optical fiber*. White Paper. 2023.
- [10] C. Xie and B. Zhang. “Scaling optical interconnects for hyperscale data center networks”. *Proceedings of the IEEE* 110.11 (2022), pp. 1699–1713.
- [11] B. Wu et al. “Secure communication in fiber-optic networks”. *Emerging trends in ICT security*. Elsevier, 2014, pp. 173–183.
- [12] M. P. Fok et al. “Optical layer security in fiber-optic networks”. *IEEE Transactions on Information Forensics and Security* 6.3 (2011), pp. 725–736.
- [13] S. B. Yoo. “New trends in photonic switching and optical networking architectures for data centers and computing systems”. *Journal of Optical Communications and Networking* 15.8 (2023), pp. C288–C298.
- [14] P. Maniotis and D. M. Kuchta. “Exploring the benefits of using co-packaged optics in data center and AI supercomputer networks: a simulation-based analysis”. *Journal of Optical Communications and Networking* 16.2 (2024), A143–A156.

- [15] T. Sasaki et al. “Ultrahigh fiber count and high-density cables, deployments, and systems”. *Proceedings of the IEEE* 110.11 (2022), pp. 1760–1771.
- [16] V. Houtsma et al. “Transceiver technologies for passive optical networks: past, present, and future [Invited Tutorial]”. *Journal of Optical Communications and Networking* 13.1 (2021), A44–A55.
- [17] M. Chagnon. “Optical communications for short reach”. *Journal of Lightwave Technology* 37.8 (2019), pp. 1779–1797.
- [18] I. Alimi et al. “A review of self-coherent optical transceivers: Fundamental issues, recent advances, and research directions”. *Applied Sciences* 11.16 (2021), p. 7554.
- [19] G. Bosco and J. Elbers. “Optical Transponders”. *Springer Handbook of Optical Networks*. Ed. by B. Mukherjee et al. Cham: Springer, 2020.
- [20] G. Agrawal. “Optical fibers”. *Fiber-Optic Communication Systems*. John Wiley & Sons, Ltd, 2003. Chap. 2.
- [21] L. Grüner-Nielsen et al. “Dispersion-compensating fibers”. *Journal of Lightwave Technology* 23.11 (2005), p. 3566.
- [22] K. Hill et al. “Chirped in-fiber Bragg gratings for compensation of optical-fiber dispersion”. *Optics letters* 19.17 (1994), pp. 1314–1316.
- [23] J. Kwon et al. “Tunable dispersion slope compensator using a chirped fiber Bragg grating tuned by a fan-shaped thin metallic heat channel”. *IEEE photonics technology letters* 18.1 (2006), pp. 118–120.
- [24] K. Zhong et al. “Digital signal processing for short-reach optical communications: A review of current technologies and future trends”. *Journal of Lightwave Technology* 36.2 (2018), pp. 377–400.
- [25] H. Zhou et al. “Recent advances in equalization technologies for short-reach optical links based on PAM4 modulation: A review”. *Applied Sciences* 9.11 (2019), p. 2342.
- [26] C. Minkenberg et al. “Co-packaged datacenter optics: Opportunities and challenges”. *IET optoelectronics* 15.2 (2021), pp. 77–91.
- [27] FiberMall. *Technology Route from 400G to 800G to 1.6T Optical Transceivers*. <https://www.fibermall.com/blog/400g-to-800g-to-1600g-optical-transceivers.htm>. Accessed: 2025-10-08. July 2023.
- [28] O. A. Montesinos López et al. “Fundamentals of artificial neural networks and deep learning”. *Multivariate statistical machine learning methods for genomic prediction*. Springer, 2022, pp. 379–425.
- [29] W. Ertel. “Neural Networks”. *Introduction to Artificial Intelligence*. Wiesbaden: Springer Fachmedien Wiesbaden, 2025, pp. 253–309.
- [30] N. Ketkar and J. Moolayil. “Feed-Forward Neural Networks”. *Deep Learning with Python: Learn Best Practices of Deep Learning Models with PyTorch*. Berkeley, CA: Apress, 2021, pp. 93–131.

- [31] I. D. Mienye et al. "Recurrent neural networks: A comprehensive review of architectures, variants, and applications". *Information* 15.9 (2024), p. 517.
- [32] M. M. Taye. "Theoretical understanding of convolutional neural network: Concepts, architectures, applications, future directions". *Computation* 11.3 (2023), p. 52.
- [33] P. J. Freire et al. "Reducing computational complexity of neural networks in optical channel equalization: From concepts to implementation". *Journal of Lightwave Technology* 41.14 (2023), pp. 4557–4581.
- [34] D. A. Ron et al. "Experimental implementation of a neural network optical channel equalizer in restricted hardware using pruning and quantization". *Scientific Reports* 12.1 (2022), p. 8713.
- [35] S. Hassan et al. "Performance Evaluation of Machine Learning-Based Channel Equalization Techniques: New Trends and Challenges". *Journal of Sensors* 2022.1 (2022), p. 2053086.
- [36] X. Liu et al. "Bi-directional gated recurrent unit neural network based nonlinear equalizer for coherent optical communication system". *Optics Express* 29.4 (2021), pp. 5923–5933.
- [37] Z. Xu et al. "Low-complexity multi-task learning aided neural networks for equalization in short-reach optical interconnects". *Journal of Lightwave Technology* 40.1 (2022), pp. 45–54.
- [38] C. Huang et al. "Prospects and applications of photonic neural networks". *Advances in Physics: X* 7.1 (2022), p. 1981155.
- [39] H. Zhou et al. "Photonic matrix multiplication lights up photonic accelerator and beyond". *Light: Science & Applications* 11.1 (2022), p. 30.
- [40] L. Larger et al. "High-Speed Photonic Reservoir Computing Using a Time-Delay-Based Architecture: Million Words per Second Classification". *Phys. Rev. X* 7 (1 Feb. 2017), p. 011015.
- [41] D. Owen-Newns et al. "Photonic spiking neural networks with highly efficient training protocols for ultrafast neuromorphic computing systems". *Intelligent Computing* 2 (2023), p. 0031.
- [42] Z. Liang et al. "Real-time respiratory motion prediction using photonic reservoir computing". *Scientific Reports* 13.1 (2023), p. 5718.
- [43] F. Ashtiani et al. "An on-chip photonic deep neural network for image classification". *Nature* 606.7914 (2022), pp. 501–506.
- [44] A. Argyris. "Photonic neuromorphic technologies in optical communications". *Nanophotonics* 11.5 (2022), pp. 897–916.
- [45] DE-CIX. *Global Data Traffic Volume Hits New Record-Breaking High at Internet Exchanges*. Online; PDF. Press release, Frankfurt am Main, Germany. Jan. 2025.

- [46] J. M. Simmons. “Introduction to Optical Networks”. *Optical Network Design and Planning*. Cham: Springer International Publishing, 2014, pp. 1–23.
- [47] F. Rambach et al. “A multilayer cost model for metro/core networks”. *Journal of Optical Communications and Networking* 5.3 (2013), pp. 210–225.
- [48] A. L. Chiu et al. “Architectures and protocols for capacity efficient, highly dynamic and highly resilient core networks”. *Journal of Optical Communications and Networking* 4.1 (2011), pp. 1–14.
- [49] R. Potharaju and N. Jain. “When the network crumbles: An empirical study of cloud network failures and their impact on services”. *Proceedings of the 4th annual Symposium on Cloud Computing*. 2013, pp. 1–17.
- [50] G. Agrawal. “Optical Amplifiers”. *Fiber-Optic Communication Systems*. John Wiley & Sons, Ltd, 2003. Chap. 6.
- [51] G. E. Keiser. “A Review of WDM Technology and Applications”. *Optical Fiber Technology* 5.1 (1999), pp. 3–39.
- [52] G. Bennett et al. “A review of high-speed coherent transmission technologies for long-haul DWDM transmission at 100G and beyond”. *IEEE Communications Magazine* 52.10 (2014), pp. 102–110.
- [53] X. Zhou and L. E. Nelson. “400G WDM transmission on the 50 GHz grid for future optical networks”. *Journal of Lightwave Technology* 30.24 (2012), pp. 3779–3792.
- [54] International Telecommunication Union, Telecommunication Standardization Sector (ITU-T). *Spectral Grids for WDM Applications: DWDM Frequency Grid*. ITU-T Recommendation G.694.1. Edition 3.0, approved 2020-10-29. Oct. 2020.
- [55] International Telecommunication Union, Telecommunication Standardization Sector (ITU-T). *Spectral Grids for WDM Applications: DWDM Frequency April 2000 Grid (G.694.2)*. ITU-T Recommendation G.694.2. Approved December 2003. Dec. 2003.
- [56] E. B. Grann. “Low cost CWDM optical transceivers”. *2001 Proceedings. 51st Electronic Components and Technology Conference (Cat. No. 01CH37220)*. IEEE. 2001, pp. 26–29.
- [57] K. Kikuchi. “Fundamentals of coherent optical fiber communications”. *Journal of lightwave technology* 34.1 (2015), pp. 157–179.
- [58] J. R. Barry and E. A. Lee. “Performance of coherent optical receivers”. *Proceedings of the IEEE* 78.8 (2002), pp. 1369–1394.
- [59] J. M. Kahn and K.-P. Ho. “Spectral efficiency limits and modulation/detection techniques for DWDM systems”. *IEEE Journal of selected topics in quantum electronics* 10.2 (2004), pp. 259–272.
- [60] M. G. Taylor. “Coherent detection method using DSP for demodulation of signal and subsequent equalization of propagation impairments”. *IEEE Photonics Technology Letters* 16.2 (2004), pp. 674–676.

- [61] M. Kuschnerov et al. "DSP for coherent single-carrier receivers". *Journal of lightwave technology* 27.16 (2009), pp. 3614–3622.
- [62] I. B. Djordjevic et al. "Next generation FEC for high-capacity communication in optical transport networks". *Journal of Lightwave Technology* 27.16 (2009), pp. 3518–3530.
- [63] R. Noe et al. "Endless polarization control systems for coherent optics". *Journal of lightwave technology* 6.7 (2002), pp. 1199–1208.
- [64] I. Marvell Technology. *Marvell, Lumentum and Coherent Demonstrate Industry's First 800 G ZR/ZR+ Pluggable Modules for 500 km Data Center Interconnects*. <https://www.marvell.com/company/newsroom/marvell-800g-zr-zrplus-pluggable-modules-for-500km-data-center-interconnects.html>. Accessed: 2025-11-20. 2024.
- [65] P. Dykes. *ECOC 2023: Source Photonics Demonstrates Industry's First 800G 4x200G LR4 OSFP Transceivers*. <https://opticalconnectionsnews.com/2023/10/ecoc-2023-source-photonics-demos-800g-4x200g-lr4-osfp-transceivers/>. Accessed: 2025-11-19. 2023.
- [66] I. Lumentum Operations. *Lumentum Showcases 400G Transceiver Portfolio for Next-Generation Data Centers*. <https://www.lumentum.com/en/media-room/news-releases/lumentum-showcases-400g-transceiver-portfolio-next-generation-data-centers>. Accessed: 2025-11-20. 2018.
- [67] R. Ryf and C. Antonelli. "Space-Division Multiplexing". *Springer Handbook of Optical Networks*. Ed. by B. Mukherjee et al. Cham: Springer International Publishing, 2020, pp. 353–393.
- [68] M. v. d. Hout et al. "Transmission of 273.6 Tb/s over 1001 km of 15-mode multi-mode fiber using C-band only 16-QAM signals". *Journal of lightwave technology* 42.3 (2024), pp. 1136–1142.
- [69] N.-P. Diamantopoulos et al. "On the power consumption of MIMO processing and its impact on the performance of SDM networks". *2017 Optical Fiber Communications Conference and Exhibition (OFC)*. IEEE. 2017, pp. 1–3.
- [70] W. Chen et al. "Applications and development of multi-core optical fibers". *Photonics*. Vol. 11. 3. MDPI. 2024, p. 270.
- [71] G. Di Sciullo et al. "45.7 Tb/s over 12053 km Transmission in an All-Multi-Core-Amplified Recirculating Loop". *Journal of Lightwave Technology* (2025).
- [72] A. Turukhin et al. "Demonstration of 0.52 Pb/s potential transmission capacity over 8,830 km using multicore fiber". *ECOC 2016; 42nd European Conference on Optical Communication*. VDE. 2016, pp. 1–3.
- [73] A. Gumaste. "Metropolitan Networks". *Springer Handbook of Optical Networks*. Cham: Springer International Publishing, 2020, pp. 609–630.

- [74] L. S. de Sousa and A. C. Drummond. “Metropolitan optical networks: A survey on single-layer architectures”. *Optical Switching and Networking* 47 (2023), p. 100719.
- [75] M. D. Feuer et al. “ROADMs and their system applications”. *Optical Fiber Telecommunications VB*. Elsevier, 2008, pp. 293–343.
- [76] S. Al Zoubi and R. Gaudino. “Modeling and analysis of coherent metro+ PON converged networks for ultra-high speed applications”. *Journal of Optical Communications and Networking* 17.4 (2025), pp. 294–308.
- [77] B. Skubic and L. Wosinska. “Introduction to Optical Access Networks”. *Springer Handbook of Optical Networks*. Cham: Springer International Publishing, 2020, pp. 831–848.
- [78] S. Jay et al. “Comparing FTTH access networks based on P2P and PMP fibre topologies”. *Telecommunications Policy* 38.5-6 (2014), pp. 415–425.
- [79] J. Hecht. “The bandwidth bottleneck that is throttling the Internet”. *Nature* 536.7615 (2016), pp. 139–142.
- [80] K. Borzycki. “FTTx Access Networks: Technical Developments and Standardization”. *Broadband Communications Networks-Recent Advances and Lessons from Practice*. IntechOpen, 2017.
- [81] J.-i. Kani and D. van Veen. “Current TDM-PON Technologies”. *Springer Handbook of Optical Networks*. Cham: Springer International Publishing, 2020, pp. 849–870.
- [82] ITU-T. *Recommendation G.9804.3 (2021) Amendment 1 — 50-Gigabit-capable Passive Optical Network (50G-PON) System*. International Telecommunication Union, Telecommunication Standardization Sector. Approved 22 February 2023; in force. 2023.
- [83] H. Liu et al. “Evolving Requirements and Trends of Datacenters Networks”. *Springer Handbook of Optical Networks*. Cham: Springer International Publishing, 2020, pp. 707–724.
- [84] L. Jiang et al. “Integrated Components and Solutions for High-Speed Short-Reach Data Transmission”. *Photonics* 8.3 (2021).
- [85] A. Shahpari et al. “Coherent access: A review”. *Journal of Lightwave Technology* 35.4 (2017), pp. 1050–1058.
- [86] A. Yan et al. “Next-Generation Access Network Based on Coherent Optics With Hybrid Transceivers, Multi Formats, and Flexible Rates”. *IEEE Journal on Selected Areas in Communications* (2025).
- [87] H. Liu et al. “Evolving Requirements and Trends of Datacenters Networks”. *Springer Handbook of Optical Networks*. Ed. by B. Mukherjee et al. Cham: Springer International Publishing, 2020, pp. 707–724.

- [88] R. Proietti et al. "Intra-Datacenter Network Architectures". *Springer Handbook of Optical Networks*. Ed. by B. Mukherjee et al. Cham: Springer International Publishing, 2020, pp. 757–778.
- [89] A. Vahdat. *Speed, Scale and Reliability: 25 Years of Google Data-Center Networking Evolution*. <https://cloud.google.com/blog/products/networking/speed-scale-reliability-25-years-of-data-center-networking>. Accessed: 2025-11-20. Google Cloud Blog, 2024.
- [90] FS.com. *Optical Transceivers*. <https://www.fs.com/it/c/optical-transceivers-9>. Accessed: 2025-10-04. 2025.
- [91] L. Nadal et al. "Advanced optical transceiver and switching solutions for next-generation optical networks". *Journal of Optical Communications and Networking* 16.8 (2024), pp. D64–D75.
- [92] G. Agrawal. "Chapter 2 - Pulse Propagation in Fibers". *Nonlinear Fiber Optics (Fifth Edition)*. Ed. by G. Agrawal. Fifth Edition. Optics and Photonics. Boston: Academic Press, 2013, pp. 27–56.
- [93] U. Gliese et al. "Chromatic dispersion in fiber-optic microwave and millimeter-wave links". *IEEE Transactions on Microwave Theory and Techniques* 44.10 (1996), pp. 1716–1724.
- [94] F. Devaux et al. "Simple measurement of fiber dispersion and of chirp parameter of intensity modulated light emitter". *Journal of Lightwave Technology* 11.12 (2002), pp. 1937–1940.
- [95] R. Hui. "Chapter 2 - Optical fibers". *Introduction to Fiber-Optic Communications*. Ed. by R. Hui. Academic Press, 2020, pp. 19–76.
- [96] B. Gul and F. Ahmad. "Review of FBG and DCF as dispersion management unit for long haul optical links and WDM systems". *Optical and quantum electronics* 55.6 (2023), p. 557.
- [97] G. Agrawal. "Dispersion Management". *Fiber-Optic Communication Systems*. John Wiley & Sons, Ltd, 2003. Chap. 7.
- [98] A. B. Dar and R. K. Jha. "Chromatic dispersion compensation techniques and characterization of fiber Bragg grating for dispersion compensation". *Optical and quantum electronics* 49.3 (2017), p. 108.
- [99] M. Yamada and K. Sakuda. "Analysis of almost-periodic distributed feedback slab waveguides via a fundamental matrix approach". *Applied optics* 26.16 (1987), pp. 3474–3478.
- [100] A. Willner et al. "DISPERSION MANAGEMENT". *Encyclopedia of Modern Optics*. Ed. by R. D. Guenther. Oxford: Elsevier, 2005, pp. 353–365.
- [101] C. Laperle and M. O'Sullivan. "Advances in high-speed DACs, ADCs, and DSP for optical coherent transceivers". *Journal of lightwave technology* 32.4 (2014), pp. 629–643.

- [102] J. McNicol et al. “Electrical domain compensation of optical dispersion [optical fibre communication applications]”. *OFC/NFOEC Technical Digest. Optical Fiber Communication Conference, 2005*. Vol. 4. IEEE. 2005, 3–pp.
- [103] H. Isono. “Latest standardization trend and future prospects for 800G/1.6 T optical transceivers”. *Next-Generation Optical Communication: Components, Sub-Systems, and Systems XII*. Vol. 12429. SPIE. 2023, pp. 45–51.
- [104] F. Frey et al. “Estimation of trends for coherent DSP ASIC power dissipation for different bitrates and transmission reaches”. *Photonic Networks; 18. ITG-Symposium*. VDE. 2017, pp. 1–8.
- [105] M. Torbatian et al. “Performance oriented DSP for flexible long haul coherent transmission”. *Journal of Lightwave Technology* 40.5 (2022), pp. 1256–1272.
- [106] I. Fatadin and S. J. Savory. “Compensation of frequency offset for 16-QAM optical coherent systems using QPSK partitioning”. *IEEE Photonics Technology Letters* 23.17 (2011), pp. 1246–1248.
- [107] A. Leven et al. “Frequency estimation in intradyne reception”. *IEEE Photonics Technology Letters* 19.6 (2007), pp. 366–368.
- [108] S. J. Savory and D. S. Millar. “DSP for Optical Transponders”. *Springer Handbook of Optical Networks*. Cham: Springer International Publishing, 2020, pp. 155–176.
- [109] Y. Xie et al. “A decision feedback equalization algorithm based on simplified Volterra structure for PAM4 IM-DD optical communication systems”. *Applied Sciences* 13.14 (2023), p. 8125.
- [110] R. Nagarajan et al. “Low power DSP-based transceivers for data center optical fiber communications (invited tutorial)”. *Journal of Lightwave Technology* 39.16 (2021), pp. 5221–5231.
- [111] B. S. G. Pillai et al. “End-to-end energy modeling and analysis of long-haul coherent transmission systems”. *Journal of Lightwave Technology* 32.18 (2014), pp. 3093–3111.
- [112] Z. Wu et al. “Chromatic Dispersion Equalization FIR Digital Filter for Coherent Receiver”. *Photonics*. Vol. 8. 11. MDPI. 2021, p. 478.
- [113] S. K. O. Soman. “A tutorial on fiber Kerr nonlinearity effect and its compensation in optical communication systems”. *journal of Optics* 23.12 (2021), p. 123502.
- [114] A. Bononi et al. “Fiber Nonlinearity and Optical System Performance”. *Springer Handbook of Optical Networks*. Cham: Springer International Publishing, 2020, pp. 287–351.
- [115] F. P. Guiomar et al. “Time-domain Volterra-based digital backpropagation for coherent optical systems”. *Journal of Lightwave Technology* 33.15 (2015), pp. 3170–3181.

- [116] A. Graell i Amat and L. Schmalen. “Forward Error Correction for Optical Transponders”. *Springer Handbook of Optical Networks*. Cham: Springer International Publishing, 2020, pp. 177–257.
- [117] R. Eisenach. *What the FEC?* <https://www.nokia.com/blog/what-the-fec/>. Accessed: October 2025. Oct. 2022.
- [118] Z. Jia and L. A. Campos. “Coherent optics ready for prime time in short-haul networks”. *IEEE Network* 35.2 (2021), pp. 8–14.
- [119] R. Harada et al. “1200-km Transmission of 4096-ary Eigenvalue-modulated Signal Using a Neural Network-based Demodulator and SD-FEC”. *Optical Fiber Communication Conference*. Optica Publishing Group. 2024, M4K–7.
- [120] K. Mishina et al. “Long-Haul Optical-Eigenvalue Transmission Using a Neural Network Demodulator and SD-FEC”. *IEEE Journal on Selected Areas in Communications* (2025).
- [121] S. Haykin. *Neural Networks and Learning Machines*. Pearson International Edition. Pearson, 2009.
- [122] W. S. McCulloch and W. Pitts. “A logical calculus of the ideas immanent in nervous activity”. *Bulletin of Mathematical Biophysics* 5 (1943), pp. 115–133.
- [123] H. Naveed et al. “A comprehensive overview of large language models”. *ACM Transactions on Intelligent Systems and Technology* 16.5 (2025), pp. 1–72.
- [124] OpenAI. *ChatGPT*. <https://chat.openai.com>. Accessed: 2025-11-21. 2025.
- [125] Google. *Gemini*. <https://gemini.google.com>. Accessed: 2025-11-21. 2025.
- [126] R. Amin et al. “A survey on machine learning techniques for routing optimization in SDN”. *IEEE Access* 9 (2021), pp. 104582–104611.
- [127] H. Dong et al. “Next-generation data center network enabled by machine learning: Review, challenges, and opportunities”. *IEEE Access* 9 (2021), pp. 136459–136475.
- [128] C. M. Bishop. *Pattern Recognition and Machine Learning*. 1st ed. Information Science and Statistics. New York, NY: Springer, 2006.
- [129] D. Brunner et al. “Tutorial: Photonic neural networks in delay systems”. *Journal of Applied Physics* 124.15 (Oct. 2018), p. 152004.
- [130] F. Rosenblatt. “The perceptron: a probabilistic model for information storage and organization in the brain.” *Psychological review* 65.6 (1958), p. 386.
- [131] K. Nakajima and I. Fischer. *Reservoir computing*. Springer, 2021.
- [132] M. Alloghani et al. “A Systematic Review on Supervised and Unsupervised Machine Learning Algorithms for Data Science”. *Supervised and Unsupervised Learning for Data Science*. Cham: Springer International Publishing, 2020, pp. 3–21.

- [133] M. A. Wiering and M. Van Otterlo. “Reinforcement learning”. *Adaptation, learning, and optimization* 12.3 (2012), p. 729.
- [134] R. Rojas. “The Backpropagation Algorithm”. *Neural Networks: A Systematic Introduction*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1996, pp. 149–182.
- [135] A. Cauchy et al. “Méthode générale pour la résolution des systemes d’équations simultanées”. *Comp. Rend. Sci. Paris* 25.1847 (1847), pp. 536–538.
- [136] D. P. Kingma and J. Ba. *Adam: A Method for Stochastic Optimization*. 2017. arXiv: 1412.6980 [cs.LG].
- [137] T. Tieleman, G. Hinton, et al. “Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude”. *COURSERA: Neural networks for machine learning* 4.2 (2012), pp. 26–31.
- [138] S. Ruder. *An overview of gradient descent optimization algorithms*. 2017. arXiv: 1609.04747 [cs.LG].
- [139] H. Robbins and S. Monro. “A Stochastic Approximation Method”. *The Annals of Mathematical Statistics* 22.3 (1951), pp. 400–407.
- [140] D. Tran et al. *Stochastic gradient descent methods for estimation with large data sets*. 2015. arXiv: 1509.06459 [stat.CO].
- [141] J. Kennedy and R. Eberhart. “Particle swarm optimization”. *Proceedings of ICNN’95 - International Conference on Neural Networks*. Vol. 4. 1995, 1942–1948 vol.4.
- [142] A. G. Gad. “Particle swarm optimization algorithm and its applications: A systematic review.” *Archives of computational methods in engineering* 29.5 (2022).
- [143] I. Arikpo et al. “Von neumann architecture and modern computers”. *Global Journal of Mathematical Sciences* 6.2 (2007), pp. 97–103.
- [144] J. L. Smoyer and P. M. Norris. “Brief Historical Perspective in Thermal Management and the Shift Toward Management at the Nanoscale”. *Heat Transfer Engineering* 40.3-4 (2019), pp. 269–282.
- [145] A. Al-Shafei et al. “High-performance and parallel computing techniques review: Applications, challenges and potentials to support net-zero transition of future grids”. *Energies* 15.22 (2022), p. 8668.
- [146] X. Zou et al. “Breaking the von Neumann bottleneck: architecture-level processing-in-memory technology”. *Science China Information Sciences* 64.6 (2021), p. 160404.
- [147] M. R. Ashakin et al. “Enhancing Digital Signal Processing: Obstacles and Advancements in FPGA and ASIC Hardware Implementations”. *Pathfinder of Research* 2.2 (2024), pp. 25–36.
- [148] W. Cao et al. “The future transistors”. *Nature* 620.7974 (2023), pp. 501–515.

- [149] J. Meng et al. “Electronic bottleneck suppression in next-generation networks with integrated photonic digital-to-analog converters”. *Advanced photonics research* 2.2 (2021), p. 2000033.
- [150] C. Xiang et al. “Perspective on the future of silicon photonics and electronics”. *Applied Physics Letters* 118.22 (2021).
- [151] A. Tsakyridis et al. “Photonic neural networks and optics-informed deep learning fundamentals”. *APL photonics* 9.1 (2024).
- [152] O. Destras et al. “Survey on activation functions for optical neural networks”. *ACM Computing Surveys* 56.2 (2023), pp. 1–30.
- [153] Z. Fu et al. “Programmable low-power consumption all-optical nonlinear activation functions using a micro-ring resonator with phase-change materials”. *Optics Express* 30.25 (2022), pp. 44943–44953.
- [154] G. Donati et al. “Microring resonators with external optical feedback for time delay reservoir computing”. *Optics Express* 30.1 (2021), pp. 522–537.
- [155] F. Musumeci et al. “An overview on application of machine learning techniques in optical networks”. *IEEE Communications Surveys & Tutorials* 21.2 (2018), pp. 1383–1408.
- [156] T. Nguyen et al. “Fiber nonlinearity equalizer based on support vector classification for coherent optical OFDM”. *IEEE Photonics Journal* 8.2 (2016), pp. 1–9.
- [157] J. Zhang et al. “Non-data-aided k-nearest neighbors technique for optical fiber nonlinearity mitigation”. *Journal of Lightwave Technology* 36.17 (2018), pp. 3564–3572.
- [158] J. Zhang et al. “K-means-clustering-based fiber nonlinearity equalization techniques for 64-QAM coherent optical communication system”. *Optics express* 25.22 (2017), pp. 27570–27580.
- [159] F. Tian et al. “Gaussian mixture model-hidden Markov model based nonlinear equalizer for optical fiber transmission”. *Optics express* 28.7 (2020), pp. 9728–9737.
- [160] K. Burse et al. “Channel equalization using neural networks: A review”. *IEEE transactions on systems, man, and cybernetics, Part C (Applications and Reviews)* 40.3 (2010), pp. 352–357.
- [161] H. Wang et al. “Silicon photonic transceivers for application in data centers”. *Journal of Semiconductors* 41.10 (2020), p. 101301.
- [162] *400G OSFP DR4 Silicon Photonics Optical Transceiver*. <https://www.naddod.com/products/102511.html>. Accessed: 2025-10-07. NADDOD, 2025.
- [163] L. C. B. Silva et al. “Integrated Photonic Neural Networks for Equalizing Optical Communication Signals: A Review”. *Photonics* 12.1 (2025).

- [164] D. Brunner et al. “Roadmap on neuromorphic photonics”. *arXiv preprint arXiv:2501.07917* (2025).
- [165] R. Moreira et al. “Programmable eye-opener lattice filter for multi-channel dispersion compensation using an integrated compact low-loss silicon nitride platform”. *Opt. Express* 24.15 (July 2016), pp. 16732–16742.
- [166] G. M. Brodnik et al. “Extended Reach 40km Transmission of C-Band Real-Time 53.125 Gbps PAM-4 Enabled with a Photonic Integrated Tunable Lattice Filter Dispersion Compensator”. *2018 Optical Fiber Communications Conference and Exposition (OFC)*. 2018, pp. 1–3.
- [167] Y. Liu et al. “Silicon Integrated Continuously Tunable Dispersion Compensator Based on Cascaded Micro-Ring Resonators”. *2022 Asia Communications and Photonics Conference (ACP)*. 2022, pp. 1352–1355.
- [168] V. Sorianello et al. “100Gb/s PolMux-NRZ transmission at 1550nm over 30km single mode fiber enabled by a silicon photonics optical dispersion compensator”. *Optical Fiber Communication Conference*. Optica Publishing Group. 2018, W2A–31.
- [169] S. Sackesyn et al. “Experimental realization of integrated photonic reservoir computing for nonlinear fiber distortion compensation”. *Opt. Express* 29.20 (Sept. 2021), pp. 30991–30997.
- [170] E. Gooskens et al. “Experimental results on nonlinear distortion compensation using photonic reservoir computing with a single set of weights for different wavelengths”. *Scientific Reports* 13.1 (2023), p. 21399.
- [171] S. Masaad et al. “Experimental Demonstration of 4-Port Photonic Reservoir Computing for Equalization of 4 and 16 QAM Signals”. *Journal of Lightwave Technology* (2024).
- [172] R. V. Assche et al. *Real-time all-optical signal equalisation with silicon photonic recurrent neural networks*. 2025. arXiv: 2503.19911 [physics.optics].
- [173] A. Argyris et al. “Photonic machine learning implementation for signal recovery in optical communications”. *Scientific reports* 8.1 (2018), p. 8487.
- [174] M. Mancinelli et al. “A photonic complex perceptron for ultrafast data processing”. *Scientific Reports* 12.1 (2022), p. 4216.
- [175] I. Applied Nanotools. *Integrated Photonics Fabrication; X-Ray & EUV Optics; Metalens & Foundry Services*. <https://www.appliednt.com/>. Accessed: 2025-11-24. 2025.
- [176] E. Staffoli et al. “Chromatic dispersion compensation via an all-optical perceptron”. *Next-Generation Optical Communication: Components, Sub-Systems, and Systems XIII*. Vol. 12894. SPIE. 2024, pp. 84–93.
- [177] E. Staffoli et al. “A Silicon Photonic Neural Network for Chromatic Dispersion Compensation in 20 Gbps PAM4 Signal at 125 km and its Scalability up to 100 Gbps”. *Journal of Lightwave Technology* 43.2 (2025), pp. 557–571.

- [178] G. Agrawal. “Optical Receivers”. *Fiber-Optic Communication Systems*. John Wiley & Sons, Ltd, 2003. Chap. 4.
- [179] B. Efron. “Bootstrap methods: another look at the jackknife”. *Breakthroughs in statistics: Methodology and distribution*. Springer, 1992, pp. 569–593.
- [180] H. Adèr et al. *Advising on Research Methods: A Consultant’s Companion*. Johannes Van Kessel Publishing, 2008.
- [181] T. Hwang et al. “OFDM and its wireless applications: A survey”. *IEEE transactions on Vehicular Technology* 58.4 (2008), pp. 1673–1694.
- [182] J. Armstrong. “OFDM for optical communications”. *Journal of lightwave technology* 27.3 (2009), pp. 189–204.
- [183] X. Luo. “The application of OFDM in optical fiber communication systems”. *IOP Conference Series: Earth and Environmental Science* 332.4 (Oct. 2019), p. 042010.
- [184] P. R. N. Marciano et al. “Chromatic Distortion Precompensation in OFDM-Based Optical Systems Through an Integrated Silicon Photonic Neural Network”. *Journal of Lightwave Technology* 43.7 (2025), pp. 3034–3040.
- [185] H. Lin and P. Siohan. “OFDM/OQAM with hermitian symmetry: Design and performance for baseband communication”. *2008 IEEE International Conference on Communications*. IEEE. 2008, pp. 652–656.
- [186] R. A. Shafik et al. “On the error vector magnitude as a performance metric and comparative analysis”. *2006 International Conference on Emerging Technologies*. IEEE. 2006, pp. 27–31.
- [187] E. Staffoli et al. “Equalization of a 10 Gbps IMDD signal by a small silicon photonics time delayed neural network”. *Photon. Res.* 11.5 (May 2023), pp. 878–886.
- [188] X. Liu et al. “Adaptive dispersion compensation using a photonic integrated circuit finite impulse response filter”. *Opt. Express* 31.22 (Oct. 2023), pp. 35971–35981.
- [189] S. Ran et al. “A 4×256 Gbps silicon transmitter with on-chip adaptive dispersion compensation”. *Nature Communications* 16.1 (2025), p. 6268.
- [190] K. Takiguchi et al. “Optical-signal-processing device based on waveguide-type variable delay lines and optical gates”. *Journal of Lightwave Technology* 24.7 (2006), pp. 2593–2601.
- [191] K. Jinguji and M. Kawachi. “Synthesis of coherent two-port lattice-form optical delay-line circuit”. *Journal of Lightwave Technology* 13.1 (1995), pp. 73–82.
- [192] M. Hayee and A. Willner. “Transmission penalties due to EDFA gain transients in add-drop multiplexed WDM networks”. *IEEE Photonics Technology Letters* 11.7 (1999), pp. 889–891.

- [193] R. Hui et al. "Cross-phase modulation in multispan WDM optical fiber systems". *Journal of lightwave Technology* 17.6 (1999), pp. 1018–1026.
- [194] Y. Fang et al. "Comparison of thermo-optic phase shifters in imec's silicon photonics platform". *European Conference on Integrated Optics*. Springer. 2024, pp. 337–342.
- [195] Y. Dong et al. "Equalization of Short-Reach IMDD Transmission with Enhanced Reservoir Computing". *Journal of Lightwave Technology* (2024).
- [196] G. S. Yadav et al. "Computation efficient sparse DNN nonlinear equalization for IM/DD 112 Gbps PAM4 inter-data center optical interconnects". *Optics letters* 46.9 (2021), pp. 1999–2002.
- [197] J. Zhang et al. "Comparison of low-complexity sparse and weight-sharing nonlinear equalizers for C-band 100-Gbit/s DSB PAM-4 transmission over 60-km SSF". *Optics express* 30.20 (2022), pp. 36343–36357.
- [198] Z. Xu et al. "Joint equalization of linear and nonlinear impairments for PAM4 short-reach direct detection systems". *IEEE Photonics Technology Letters* 33.9 (2021), pp. 425–428.
- [199] W. Sheng et al. "Complex-valued recurrent neural network equalizer with low complexity for a 120-Gbps 50-km optical PAM-4 IM/DD system". *Optics Express* 32.16 (2024), pp. 27624–27634.
- [200] S. Wang et al. "Signal recovery based on optoelectronic reservoir computing for high speed optical fiber communication system". *Optics Communications* 495 (2021), p. 127082.
- [201] S. Li and S. Pachnicke. "Optical equalization using photonic reservoir computing with optical analog signal injection". *Asia Communications and Photonics Conference*. Optica Publishing Group. 2019, T4G–5.
- [202] I. Estébanez et al. "56 GBaud PAM-4 100 km transmission system with photonic processing schemes". *Journal of Lightwave Technology* 40.1 (2022), pp. 55–62.
- [203] X. Zuo et al. "Integrated silicon photonic reservoir computing with PSO training algorithm for fiber communication channel equalization". *Journal of Lightwave Technology* 41.18 (2023), pp. 5841–5850.
- [204] K. Takiguchi. "Integrated-optic chromatic dispersion compensator with completely passive operation and wide operational bandwidth". *Optics Continuum* 2.12 (2023), pp. 2529–2539.
- [205] B. Wang et al. "Beyond Terabit/s Integrated Neuromorphic Photonic Processor for DSP-Free Optical Interconnects". *arXiv preprint arXiv:2504.15044* (2025).
- [206] A. M. F. P. Ltd. *Dedicated Si Photonics Foundry Services and Solutions*. <https://www.advmf.com/>. Accessed: 2025-11-24. 2025.
- [207] G. T. Reed et al. "Silicon optical modulators". *Nature photonics* 4.8 (2010), pp. 518–526.

- [208] A. I. Abbas and G. E. Cowan. “Fast-Locking Burst-Mode Clock and Data Recovery for Parallel VCSEL-Based Optical Link Receivers”. *IEEE Access* 10 (2022), pp. 34306–34320.
- [209] I. Fatadin et al. “Blind equalization and carrier phase recovery in a 16-QAM optical coherent system”. *Journal of lightwave technology* 27.15 (2009), pp. 3042–3049.
- [210] E. Staffoli et al. “Nonlinear Distortion Equalization in Multi-Span Optical Links Via a Feed-Forward Photonic Neural Network”. *Laser & Photonics Reviews* (2025), e00855.
- [211] imec. *Custom Chip Manufacturing via IC-Link*. imec – What We Offer / Development. Online; accessed 2025-11-24. 2025.
- [212] Fraunhofer Heinrich-Hertz Institute (HHI). *Fraunhofer HHI — Institute for Telecommunications, Heinrich-Hertz Institute*. Fraunhofer HHI Website. Online; accessed 2025-11-24. 2025.
- [213] A. A. Juarez et al. “Design Considerations for 1.6 Tbit/s Data Center Interconnects: Evaluating IM/DD and Coherent Transmission over O-Band Transmission Window”. *Photonics* 11.12 (2024).
- [214] M. Sr. *Edge Data Center Interconnection*. <https://fiberroad.com/resources/articles/edge-data-center-interconnection/>. Accessed: 2025-10-20. 2021.
- [215] X. Pang et al. “200 Gbps/lane IM/DD technologies for short reach optical interconnects”. *Journal of Lightwave Technology* 38.2 (2020), pp. 492–503.
- [216] P. Groumas et al. “Enabling low-cost high-volume production compatible terabit transceivers with up to 1.6 Tbps capacity and 100Gbps per lane PAM-4 modulation for intra-data center optical interconnects up to 2km: The TERIPHIC project approach”. *Metro and Data Center Optical Networks and Short-Reach Links III*. Ed. by A. K. Srivastava et al. Vol. 11308. International Society for Optics and Photonics. SPIE, 2020, p. 113080C.
- [217] S. Qin et al. “Micro-transfer printing of InP SOAs on advanced silicon photonics platform for C-band pre-amplified receivers”. *Silicon Photonics XX*. Vol. 13371. SPIE. 2025, pp. 111–115.
- [218] M. D. Lorenzo. *What Is an Eye Diagram in Electronics, What Is It Used For?* Maker.io (Digi-Key Electronics). Online; accessed 2025-11-23. 2024.
- [219] D. Meena et al. “Mitigation of EDFA transient effects in variable duty cycle pulsed signals”. *Defence Technology* 15.3 (2019), pp. 276–281.
- [220] G. Agrawal. “Chapter 1 - Introduction”. *Nonlinear Fiber Optics (Fifth Edition)*. Fifth Edition. Optics and Photonics. Boston: Academic Press, 2013, pp. 1–25.