INSTITUTO
UNIVERSITÁRIO
DE LISBOA

Point-to-point advanced self-coherent 200 Gb/s multicore fiber links supported by neural networks

Lucas Filipe Rodrigues Oliveira

Master's in Telecommunications and Computer Engineering

Supervisor:
PhD Tiago Manuel Ferreira Alves, Assistant Professor
Iscte - Instituto Universitário de Lisboa

Co-Supervisor:
PhD Adolfo da Visitação Tregeira Cartaxo, Full Professor
Iscte - Instituto Universitário de Lisboa

October, 2023

TECNOLOGIAS
E ARQUITETURA

Department of Information Science and Technology

Point-to-point advanced self-coherent 200 Gb/s multicore fiber links supported by neural networks

Lucas Filipe Rodrigues Oliveira

Master's in Telecommunications and Computer Engineering

Supervisor:
PhD Tiago Manuel Ferreira Alves, Assistant Professor
Iscte - Instituto Universitário de Lisboa

Co-Supervisor:
PhD Adolfo da Visitação Tregeira Cartaxo, Full Professor
Iscte - Instituto Universitário de Lisboa

October, 2023

Aos meus pais e irmã.

Alzira Rodrigues e Válter Rodrigues.

Ackowledgements

I would like to thank my supervisors Tiago Alves and Adolfo Cartaxo for their support and availability to answer all my questions. I would also like to thank the support provided by Instituto de Telecomunicações through the DigCORE project (BIL/Nº9/2023, DigCORE - UIDB/50008/2020), and for allowing me to use the facilities of their branch in Iscte-IUL.

Thanks to all my college friends for the stress, advice, support and fun we had throughout the years, it was all worth it. Thank you to all my friends from Pernes who were always there for me when I needed them. To my housemates in Lisbon, thank you for becoming part of my life. Thanks for all the support and fun you have provided me.

Thanks to Associação Duarte Tarré for the scholarship provided these past 4 years, it truly helped me a lot. I would also like to thank Huawei for investing in me with one of the fifty scholarships awarded in the first edition of the Huawei Scholarship program.

A big thank you to my parents and my sister for the continued love and support. Thank you for the effort you made so that I could study in Lisbon.

Resumo

Este trabalho propõe um sistema de fibra multi-núcleo (MCF) de curto alcance de 200 Gb/s considerando lasers reais com ruído de fase (LPN) e recetores Kramers Kronig. Os lasers usados neste sistema consideram larguras de linha típicas de lasers de cavidade externa (ECL) e lasers de feedback distribuído (DFB). Uma rede neural feed-forward (FFNN) é implementada para mitigar os efeitos do LPN e da diafonia entre núcleos (ICXT). O objetivo principal deste trabalho é avaliar o impacto do LPN na melhoria de desempenho proporcionada pela FFNN proposta. Em primeiro lugar, um sistema sem LPN é estudado como referência. Posteriormente, o LPN é introduzido no sistema, com o sinal ótico injetado no núcleo interferente a considerar um laser com largura de linha típica de ECLs e lasers DFB. O sinal ótico injetado no núcleo interferido considera um laser ideal sem LPN. Concluiu-se que, aplicando a FFNN, o BER médio obtido com o ECL e o laser DFB, em comparação com o caso de referência, aumentou mais de uma ordem de grandeza. O BER médio obtido com ambos os lasers manteve-se abaixo do limite de BER quando o intervalo de tempo entre a fase de treino e uso da FFNN (ΔT) não excedeu 20% do tempo de coerência (Tc). Considerando $\Delta T/Tc$=0.1, a FFNN proporcionou uma melhoria de 25% e 22% na probabilidade de indisponibilidade em comparação com a probabilidade de indisponibilidade antes da FFNN, ao considerar um ECL e um laser DFB, respetivamente. Essa melhoria diminui com o aumento de $\Delta T/Tc$.

Palavras-chave: Sistemas de comunicação por fibra ótica, data centers, fibras multi-núcleo, recetores auto-coerentes, redes neuronais, ruído de fase do laser.

Abstract

This work proposes a 200 Gb/s short-reach multi-core fiber (MCF) system considering real lasers with laser phase noise (LPN) and Kramers Kronig receivers. The lasers employed in this system consider linewidths typical of external cavity lasers (ECL) and distributed feedback lasers (DFB). A feed-forward neural network (FFNN) is implemented to mitigate the effects of the LPN and inter-core crosstalk (ICXT). The main objective of this work is to assess the impact of the LPN on the performance improvement provided by the proposed FFNN. Firstly, a system without LPN is studied as reference. Afterwards, the LPN is introduced in the system as the optical signal injected in the interfering core considered a laser with linewidth typical of ECLs and DFB lasers. The optical signal injected in the interfered core considered an ideal laser without LPN. It was concluded that, when employing the FFNN, the mean BER obtained with the ECL and DFB laser, compared to the reference case, increased more than one order of magnitude. Furthermore, the mean BER obtained with both lasers was kept below the BER threshold when the time interval between the training phase and the use of the FFNN (ΔT) did not exceed 20% of the coherence time (Tc). Considering $\Delta T/Tc$=0.1, the FFNN provided a 25% and a 22% improvement on the outage probability when compared with the outage probability before the FFNN, while considering an ECL and a DFB laser, respectively. This improvement will decrease with the increase of the $\Delta T/Tc$.

Keywords: Optical fiber communication systems, data centers, multi-core fibers, self-coherent receivers, neural networks, laser phase noise.

Contents

Ackowledgements i

Resumo iii

Abstract v

List of figures xi

List of Tables xiii

List of acronyms xv

List of symbols xix

CHAPTER 1: Introduction 1

1.1. Background and motivation 1

1.2. Research questions 2

1.3. Objectives 2

1.4. Methodology 3

1.5. Dissertation structure 4

1.6. Main contributions 4

CHAPTER 2: Literature review 5

2.1. Optical networks 5

2.2. Receiver detection techniques 6

2.2.1. Direct detection 7

2.2.2. Digital coherent 7

2.2.3. Self-coherent direct detection 8

2.3. Data centers 9

2.4. Space division multiplexing 11

2.4.1. Mode division multiplexing 11

2.4.2. Multi-core fiber 12

2.4.2.1. Single-mode multi-core fiber 13

2.4.2.2. Few-mode multi-core fiber 14

2.5. Performance degradation contributors 15
2.5.1. Inter-core crosstalk 15
2.5.2. Laser phase noise 16
2.6. Machine learning 17
2.6.1. Unsupervised learning and Reinforcement learning 18
2.6.2. Supervised learning 18
2.6.2.1. Feed-forward neural networks 18
2.6.2.2. Recurrent neural networks 19
2.6.2.3. Deep learning 20
2.6.2.4. Deep neural networks 21
2.6.2.5. Convolutional neural networks 21
2.6.3. Applications in optical communications 22
2.7. Summary 24
CHAPTER 3: Architecture of the NN-assisted self-coherent MCF system 25
3.1. System architecture 25
3.2. Optical transmitter 26
3.2.1. Root-raised cosine filter 27
3.2.2. Dual-Parallel Mach-Zender Modulator 27
3.2.3. Laser phase noise 28
3.2.4. Formatting the SSB signal 29
3.3. MCF model 31
3.3.1. ICXT statistic validation 33
3.4. KK receiver model 36
3.4.1. PIN photodetector 36
3.4.2. KK field reconstruction 36
3.4.3. KK receiver validation 38
3.5. Summary 40
CHAPTER 4: Performance of the NN-assisted self-coherent MCF system 41
4.1. System without laser phase noise 41
4.1.1. Study of the FFNN with 4 input features 41
4.1.2. Performance assessment 45
4.2. System with laser phase noise 47
4.2.1. Study of the FFNN with 4 input features – ECL with 100 kHz linewidth 47

4.2.2. Study of the FFNN with 4 input features – DFB laser with 1 MHz linewidth 50
4.2.3. Performance assessment 53

CHAPTER 5: Conclusion and future work 60

5.1. Final conclusion 60

5.2. Future work 61

Appendix 63

A. Feed-forward neural network with 2 input features 63

B. Evolution of the mean BER considering an ECL and a DFB laser 66

C. Electrical Noise 67

Bibliography 69

List of figures

Figure 1: DSRM model. 3
Figure 2: Optical Network's hierarchy. 6
Figure 3: Hyperscale DC's network architecture. 10
Figure 4: Weakly coupled FMF. 12
Figure 5: Strongly coupled FMF. 12
Figure 6: Weakly coupled SM-MCF. 13
Figure 7: Strongly coupled SM-MCF. 13
Figure 8: Architecture of a regression FFNN. 19
Figure 9: Architecture of a RNN. 20
Figure 10: Architecture of a classification DNN. 21
Figure 11: Architecture example of a classification CNN. 22
Figure 12: Architecture of the short-reach self-coherent 2-core fiber system employing NNs. 26
Figure 13: Evolution of the phase noise and phase noise variance with time, considering a linewidth of 1 MHz. 28
Figure 14: Spectrum of the optical field at the laser output. 29
Figure 15: PSDs of the information-bearing signal before and after being spectrally shifted. 30
Figure 16: PSDs of the SSB signals with and without the LPN, considering a linewidth of 1 MHz. 30
Figure 17: Evolution of the STAXT over 100 000 time fractions. 34
Figure 18: PDFs of the amplitude of the ICXT field components. 35
Figure 19: Block diagram of the KK receiver. 36
Figure 20: PSD of the information-bearing signal at the output of the KK receiver (without LPN). 37
Figure 21: PSD of the information-bearing signal without carrier before and after the shift in frequency. 38
Figure 22: Time trajectory of the signal and constellation for a CSPR=4 dB (< PAPR). 39
Figure 23: Time trajectory of the signal and constellation for a CSPR=PAPR. 39
Figure 24: Time trajectory of the signal and constellation for a CSPR=13 dB (> PAPR). 40
Figure 25: FFNN with 4 input features. 42
Figure 26: EVM vs #Training samples (10 neurons) – 4 inputs without LPN. 42
Figure 27: EVM vs #Neurons (256 samples) – 4 inputs without LPN. 43
Figure 28: EVM vs #Neurons (3000 and 20 000 samples) – 4 inputs without LPN. 44
Figure 29: Evolution of the BER as a function of the time fractions with and without the FFNN – 4 inputs without LPN. 45
Figure 30: Mean BER vs ICXT level without LPN – 4 input features. 46

Figure 31: Constellations before and after using the FFNN. 46
Figure 32: Phase noise variation as a function of time for a 100 kHz linewidth. 47
Figure 33: EVM vs #training samples (10 neurons) - 100 kHz linewidth. 48
Figure 34: EVM vs #neurons with 1000, 6000 and 20 000 training samples – 100 kHz linewidth. 49
Figure 35: Evolution of the BER as a function of the time fractions with and without the FFNN – 4 inputs with LPN (100 kHz). 50
Figure 36: Phase noise variation as a function of time for a 1 MHz linewidth. 51
Figure 37: EVM vs #training samples (10 neurons) - 1 MHz linewidth. 51
Figure 38: EVM vs #neurons considering 1000 and 2000 training samples - 1 MHz linewidth. 52
Figure 39: Evolution of the BER as a function of the time fractions with and without the FFNN – 4 inputs with LPN (1 MHz). 53
Figure 40: Diagram of the time basis of the simulations. 54
Figure 41: Evolution of the mean BER and EVM as a function of ΔT normalized by the coherence time T_c, for a linewidth of 0, 100 kHz and 1 MHz. 55
Figure 42: Estimated CCDFs of the BER for several ΔT/Tc considering a laser with 100 kHz of linewidth. 57
Figure 43: Estimated CCDFs of the BER for several $\Delta T/T_c$ considering a laser with 1 MHz of linewidth. 58
Figure 44: Constellations of the received signal in core m considering a DFB laser with a 1 MHz linewidth. 58
Figure 45: Constellations of the received signal in core n for ΔT/Tc=0.2, considering a DFB laser with a 1 MHz linewidth. 59
Figure 46: Constellations of the received signal in core n for ΔT/Tc=0.5, considering a DFB laser with a 1 MHz linewidth. 59
Figure 47: FFNN with 2 input features. 63
Figure 48: EVM vs #Training samples (10 neurons) – 2 inputs. 64
Figure 49: EVM vs #Neurons – 2 inputs. 64
Figure 50: BER evolution before and after the FFNN – 2 inputs. 65
Figure 51: Mean BER vs ICXT level – 2 inputs. 66
Figure 52: Evolution of the mean BER as a function of the time fractions while employing a) an ECL with a 100 kHz linewidth, and b) a DFB laser with a 1 MHz linewidth. 67

List of Tables

Table 1: Parameters for the validation of the ICXT statistic. 34
Table 2: Parameters for the validation of the KK receiver. 38
Table 3: MCF parameters. 41
Table 4: Signal parameters. 41
Table 5: FFNN configurations – 4 inputs without LPN. 44
Table 6: Coherence time and time window – 100 kHz of linewidth. 47
Table 7: FFNN configurations - 4 inputs with LPN (100 kHz). 49
Table 8: FFNN configurations - 4 inputs with LPN (1 MHz). 50
Table 9: FFNN configurations - 4 inputs with LPN (1 MHz). 52
Table 10: Parameters of the simulation. 54
Table 11: EVM gain for different ΔTs considering the ECL and DFB laser. 56
Table 12: FFNN configurations – 2 inputs without LPN. 65

List of acronyms

AR-RNN	Auto-regressive recurrent neural network
BER	Bit error rate
BP	Backpropagation
CCDF	Complementary cumulative distribution function
CD	Chromatic dispersion
CNN	Convolutional neural networks
CSPR	Carrier-to-signal power ratio
DAC	Digital to analog converter
DBP	Digital backpropagation
DC	Data center
DCF	Dispersion compensating fiber
DD	Direct detection
DDPG	Deep deterministic policy gradient
DFB	Distributed feedback laser
DL	Deep learning
DNN	Deep neural networks
DP-MZM	Dual-parallel Mach-Zehnder modulator
DQN	Deep Q learning
DSP	Digital signal processing
DSRM	Design science research methodology
DWDM	Dense wavelength division multiplexing
ECL	External cavity laser
EDFA	Erbium-doped fiber amplifier
ELM	Extreme learning machine
EVM	Error vector magnitude
FEC	Forward error correction

FFNN	Feedforward neural network
FMF	Few modes fiber
FM-MCF	Few-mode multi-core fiber
HPF	High-pass filter
ICP	Internet content providers
ICXT	Inter-core crosstalk
IM	Intensity modulation
IM/DD	Intensity modulation/direct detection
IM-DD	Intensity modulation and direct detection
ISI	Inter-symbol interference
KK	Kramers-Kronig
KNN	K nearest neighbour
LAN	Local area network
LO	Local oscillator
LPN	Laser phase noise
L-RNN	Layer-recurrent neural network
LSTM	Long short-term memory networks
MCF	Multi-core fiber
MIMO	Multiple-input multiple-output
ML	Machine learning
MMF	Multi-mode fiber
MSE	Mean squared error
MZM	Mach-Zehnder modulator
NEP	Noise equivalent power
NLC	Nonlinear compensation
NLSE	Nonlinear Schrödinger equation
NN	Neural network
OFDM	Orthogonal frequency division multiplexing
OOK	On-off keying

OOK-DD	On-off keying direct detection
OSNR	Optical signal-to-noise ratio
PAM	Pulse amplitude modulation
PAPR	Peak-to-average-power ratio
PCA	Principal component analysis
PDF	Probability density function
PDI	Propagation-direction interleaving
PIN	Positive-Intrinsic-Negative
PMP	Phase matching points
PON	Passive optical network
PS	Probabilistic shaping
PSD	Power spectral density
PSM	Parallel single-mode
QAM	Quadrature amplitude modulation
QPSK	Quadrature phase shift keying
RBF-NN	Radial basis function neural network
RL	Reinforcement learning
RNN	Recurrent neural networks
ROP	Received optical power
RPS	Random phase shifts
RRC	Root-raised-cosine
SCC	Spatial channel count
SCF	Single-core fiber
SDM	Space division multiplexing
SELU	Scaled exponential linear unit
SMF	Single-mode fiber
SM-MCF	Single-mode multi-core fiber
SSB	Single-side band
SSBI	Signal-to-signal beat interference

STAXT	Short-term average inter-core crosstalk
TOR	Top-of-rack
TPS	Truncated probabilistic shaping
ULAF	Ultra large effective area fiber
UWB	Ultra-wide band
VNLE	Volterra non-linear equalization
WDM	Wavelength division multiplexing

List of symbols

$\boldsymbol{\alpha}$	Fiber attenuation in Np/m
$\boldsymbol{\beta_0}$	Propagation constant
$\boldsymbol{\beta_1}$	Propagation time delay
$\boldsymbol{\beta_2}$	Group velocity dispersion
$\boldsymbol{\beta_3}$	Higher order dispersion
$\Delta\overline{\boldsymbol{\beta}}_{\boldsymbol{mn}}(\boldsymbol{\omega})$	Difference between the averages of the propagation constants in cores *m* and *n*
$\Delta\overline{\boldsymbol{\beta}}_{\boldsymbol{0,mn}}$	Difference between the averages of the propagation constants in cores *m* and *n*, for ω=0
$\Delta\boldsymbol{D_{mn}}$	Difference between the dispersion parameters of cores *m* and *n*
$\Delta\boldsymbol{\nu_l}$	Laser linewidth
$\boldsymbol{\lambda_0}$	Channel wavelength
$\boldsymbol{\mu(\tau)}$	Frequency noise
$\boldsymbol{\rho}$	Roll-off factor
$\boldsymbol{\varphi(t)}$	Phase of the signal reconstructed in the KK receiver
$\boldsymbol{\phi(t)}$	Laser phase noise
$\boldsymbol{\phi_{p,b}^{(k)}}$	RPS of the *k*-th PMP
$\boldsymbol{\psi, \Gamma, \theta}$	Random processes, constant along each time fraction
$\boldsymbol{\omega}$	Angular frequency
$\boldsymbol{A}$	Amplitude of the optical field of the carrier
$\boldsymbol{B_g}$	Guard band between the carrier and the information-bearing signal
$\boldsymbol{B_{signal}}$	Bandwidth of the information-bearing signal
$\boldsymbol{c}$	Speed of light in vacuum
$\boldsymbol{D_{\lambda_0}}$	Dispersion parameter
$\boldsymbol{d_{mn}}$	Walkoff parameter measured between cores *m* and *n*
$\boldsymbol{e_{1,2}(t)}$	Optical field at the output of each inner MZM
$\boldsymbol{e_{core,x}(t)}$	SSB signal at the transmitter output for core $\in \{m, n\}$ and polarization *x*
$\boldsymbol{e'_{core,x}(t)}$	SSB signal with polarization rotation at the output of core $\in \{m, n\}$ for polarization *x*
$\boldsymbol{e_{core,y}(t)}$	SSB signal at the transmitter output for core $\in \{m, n\}$ and polarization *y*
$\boldsymbol{e'_{core,y}(t)}$	SSB signal with polarization rotation at the output of core $\in \{m, n\}$ for polarization *y*
$\boldsymbol{e_{ICXT,n,p}(t)}$	The ICXT optical field induced by core *m* into core *n* for each polarization $p \in \{x, y\}$
$\boldsymbol{E_{in}}$	Optical field of the laser at the DP-MZM input
$\boldsymbol{E_{in,1,2}}$	Optical field of the laser at the input of each inner MZM
$\boldsymbol{e_{out}(t)}$	Optical field of the information-bearing signal at the output of the DP-MZM

$e_{out,n,p}(t)$	Optical field with ICXT at the output of core *n* for each polarization $p \in \{x, y\}$
$e_{SSB}(t)$	Optical field of the SSB signal at the output of the transmitter
f	Frequency
$F_{p,b}(\omega)$	Transfer function that models the ICXT induced by core *m* into core *n* with $p,b \in \{x, y\}$
F^{-1}	Inverse Fourier transform
H	Hilbert transform
$H_f(\omega)$	Transfer function of the SMF
$H_{HPF}(f)$	Transfer function of the HPF
$H_{RRC}(f)$	Transfer function of the RRC filter
$i(t)$	Electrical current at the output of the PIN
$i_n(t)$	Electrical noise
$\overline{K_{nm}}$	Discrete coupling coefficient
L	Length of the MCF in meters
m	Interfering core
n	Interfered core
n_{eff}	Effective refractive index of the fiber
N_{ic}	Number of interfering cores
N_{PMP}	Number of phase matching points
p	Polarization
$p_{ICXT}(t)$	Average ICXT power
P_m	Average power at the input of core *m*
P_n	Average power at the input of core *n*
$P_{out\ laser}$	Average power at the laser output
$P_{out,m}$	Average power at the input of core *m*
$P_{out,n}$	Average power at the input of core *n*
$P_{total,m}$	Total power injected in core *m*
R_λ	PIN responsivity
R_s	Symbol rate
S_{λ_0}	Slope of the dispersion parameter
$S(t)$	Information-bearing signal
S_{mn}	Skew between cores *m* and *n*
t	Time
T	Symbol period
T_c	Coherence time
$v_{1,2}(t)$	Voltage of the signal applied to either the upper or lower arm of the MZM
$v_{AC,1,2}(t)$	Voltage of the electrical signal that contains the information for the upper or lower arm

$V_{b,1,2}$	Bias voltage applied to each inner MZM
$V_{b,3}$	Bias voltage of the outer MZM
v_g	Group velocity
V_{sv}	Switching voltage
X_C	ICXT level
z_k	*K*-th random coordinate

CHAPTER 1

Introduction

1.1. Background and motivation

Current optical fiber networks are reaching the so-called capacity crunch of 100 Tb/s per single core fiber [1][2]. Space division multiplexing (SDM) has been indicated as a powerful solution to provide an ultimate capacity increase as it explores the only known physical dimension left to be exploited in optical networks [3][4]. There are several types of SDM such as few mode fibers (FMF) where multiple spatial modes propagate through the fiber, or multi-core fibers (MCF) where *N* independent cores provide a capacity increase of *N*-fold compared with standard single-mode fibers (SMF) used in current networks. These MCFs can be further distinguished based on the number of modes propagating in each core, resulting in single-mode MCFs (SM-MCF) and few-mode MCFs (FM-MCF) [1][5]. MCF-based SDM systems have been mainly proposed to respond to:

(i) The growing capacity demands in core networks, through new advanced transmission techniques and/or modulation formats.
(ii) The space limitations in short-reach networks as intra or inter data center (DC) communications.

Due to cost purposes, short-reach MCF-based systems should employ off-the-shelf lasers and DD receivers. However, these systems may present significant performance degradation due to the laser phase noise (LPN), photodetection nonlinearity, chromatic dispersion (CD) and random evolution of the inter-core crosstalk (ICXT), with the latter being caused by the optical power inserted in one core being coupled with neighboring cores during propagation.

In DC communications, the single-mode fiber's potential is not fully utilized due to strict limitations in transceiver size, cost and power dissipation [6]. In contrast, MCFs and FMFs are a good fit to increase the available bandwidth in these systems because they can effectively achieve a higher capacity than standard SMF. This type of communications prioritizes low cost, power consumption and latency, making direct detection (DD) systems the preferred solution. However, conventional intensity modulation and direct detection (IM-DD) systems have some limitations in terms of the bitrate. Although 4-PAM was proposed to increase the bitrate, it presented a 10 km limit in reach due to the intolerance of these systems to chromatic dispersion, so, focus was placed on systems that make use of the complex optical field [8]. Thus, advanced self-coherent receivers emerged as a DD solution that has a higher reach and allows higher-order modulation formats, resulting in higher bitrates and requiring a single photodiode [7]. The Kramers Kronig (KK) receiver is an example of this type of receivers, having been proposed to mitigate some effects such as the photodetector nonlinearity. The combined use of

these technologies will therefore result in higher bitrates and reach than current DC systems utilizing 4-PAM and parallel fibers.

As computing power increased over the years, machine learning (ML) began being used in multiple areas of investigation, including optics, in order to improve system performance. Multiple experiments have been conducted, with ML algorithms having been effectively used in fiber nonlinearity compensation [9], as equalizers to compensate the photodetector nonlinearity of DD receivers [10], and for the compensation of ICXT induced by MCFs [11].

In this work, ML algorithms based on neural networks (NN) are proposed to mitigate the combined effect of the ICXT and phase noise introduced by the laser source in 200 Gb/s MCF systems employing self-coherent KK receivers.

1.2. Research questions

Within the scope of this dissertation, the following questions are to be answered:

- What parameters should the chosen NN algorithm have, in order to improve the overall system's performance (number of neurons and training samples)?
- How does the LPN impact the performance improvement provided by the NNs?
- How long does the FFNN mitigate the combined effect of the LPN and ICXT?

1.3. Objectives

The general purpose of this work is to unlock the capacity provided by next generation optical networks. This is accomplished by proposing an SDM system for point-to-point transmission with a dedicated capacity of >200 Gb/s using KK receivers and NNs for system performance optimization. Particularly, the following objectives are pursued:

(i) To integrate a software platform for simulation of over 200 Gb/s SDM optical fiber systems employing self-coherent KK receivers, MCF transmission, LPN and NNs.

(ii) To identify the main advantages and challenges of different classes of NN algorithms.

(iii) To implement a NN technique in the simulator of step (i) for end-to-end performance optimization in point-to-point >200 Gb/s QAM-SDM optical fiber links employing DD and impaired by the combined effect of the ICXT and LPN.

(iv) To assess the impact of the random nature of the LPN and ICXT along time on the performance of the proposed ML technique.

1.4. Methodology

The Design science research methodology (DSRM) described in [12] will be followed in the elaboration of this dissertation, on account of the engineering nature of the subject. Its process model is composed of several activities which have different possible entry points, as seen in figure 1.

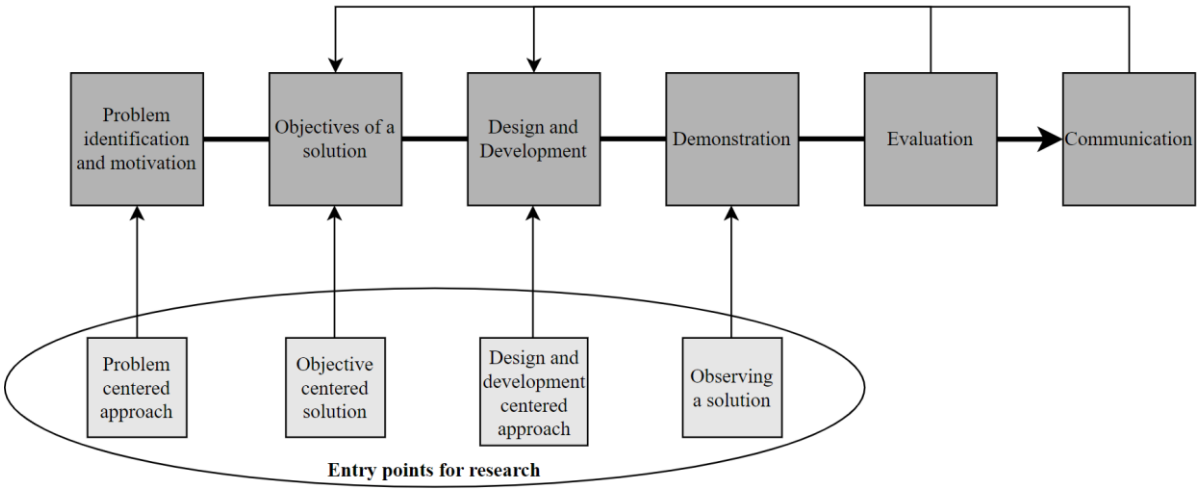


Figure 1: DSRM model.

In the context of this work, a problem centered approach is taken since the problem, which is the impact of the LPN and ICXT on the performance of MCF systems, has already been observed and the main goal of this work is to create a software platform that employs NNs and KK receivers in order to assess the impact of the LPN on the performance improvement provided by the neural networks on the ICXT. Considering this, the first development stage of this dissertation is the problem identification and motivation. In this report, the problem and motivation are firstly introduced in section 1.1 and further explained in the literature review. Afterwards the solution's objectives were defined and are represented in section 1.3. The next activity involves the design and development of the intended simulator that was concisely described in the first objective of section 1.3. Here, a proper architecture of the simulator has to be determined before starting the development, resorting, for example, to a schematic diagram. After the simulator has been successfully developed, in the demonstration activity, it is going to be run with all the required parameters.

Subsequently, in the evaluation activity, with all the results obtained from the simulation, the impact of the LPN on the performance improvement provided by the chosen neural network technique is going to be assessed. Finally, in the communication phase, when all the desired results have been analyzed, the work done is going to be published in the form of a master's dissertation.

1.5. Dissertation structure

The dissertation is organized as follows:

- Chapter 2 includes the literature review of several key concepts in this work, namely self-coherent detection (KK receivers), MCFs and its main drawback - the ICXT, the phase noise present in real lasers and the use of ML in optical communications.
- Chapter 3 presents the system architecture, providing a thorough description and validation of all the models to be used in the simulation of each element in this system.
- Chapter 4 studies the optimum parameters for the intended NNs and the impact of the LPN on their performance improvement, considering external cavity and distributed feedback lasers (ECL and DFBs).
- Chapter 5 provides the main conclusions of these studies, while also suggesting some future work.

1.6. Main contributions

The main contributions of this work are:

- Implementation of a FFNN that mitigates the combined effect of the LPN and ICXT induced by the neighboring cores in systems with skew×symbol rate << 1.
- Demonstration of how long the FFNN is able to mitigate the combined effect of the LPN and ICXT without having to be trained again.
- Demonstration of the impact the LPN has on the performance improvement provided by the FFNN.

CHAPTER 2

Literature review

2.1. Optical networks

Optical networks are divided into different segments in a hierarchical structure with three layers: core, metropolitan, and access (figure 2) [14][17]. Access networks are on the edge of the network and are the ones that actually serve the end users. These networks are usually of smaller distances up to 20 km, although recent research on passive optical networks (PON) have shown a maximum coverage of 60 km [13].

Metropolitan networks essentially collect business, residential and mobile traffic of end users at central offices (COs) spread across cities and towns, i.e., from the access networks, and span between dozens and hundreds of kilometers [14]. Most of the traffic collected will go to the core, to reach, e.g., content distribution network services [17].

Finally, core networks span large distances from hundreds to thousands of kilometers, serving metropolitan networks offering them connectivity and transporting large volumes of aggregated traffic [14]. This is the network segment that transports the most amount of traffic, so it is very important to use the available spectrum efficiently. Initially, in the dense wavelength division multiplexing (DWDM) era, core networks used fixed-grid channel spacings of 50 GHz in the C-band reaching bitrates of 10 Gb/s per channel, while utilizing the intensity modulation and direct detection scheme (IM-DD). With the continuous growth of traffic, technology evolved to reach data rates of 100 Gb/s on a single wavelength carrier, employing coherent technology. Using two polarizations with quadrature phase shift keying (QPSK) as the modulation scheme, a data rate of 100 Gb/s was attained, achieving a spectral efficiency of 2 bit/s/Hz [18][19]. Research was later focused on higher order quadrature amplitude modulation (QAM) to reach higher bitrates and spectral efficiencies, with results revealing that there is a higher optical signal to noise ratio (OSNR) limit, therefore reaching lower distances [19]. In fixed grid systems, spectrum was a lot of times wasted, so to avoid that, flexible grid systems were introduced and essentially eliminated that problem, since the channel spacings are multiples of 12.5 GHz, therefore having higher granularity and lower waste of spectrum [19]. All this combined with sophisticated forward error correction (FEC) schemes and the use of digital signal processing (DSP) enabled systems to operate close to their theoretical Shannon limits, reaching aggregate per-fiber capacities of up to 70 Tbps, using the C (1530-1565 nm) and L (around 1565-1625 nm) bands. Furthermore, experiments based on the C+L band single mode fiber (SMF) transmission over 1000 km have barely surpassed 100 Tbps, thus marking this number as the capacity limit/crunch of SMF [2].

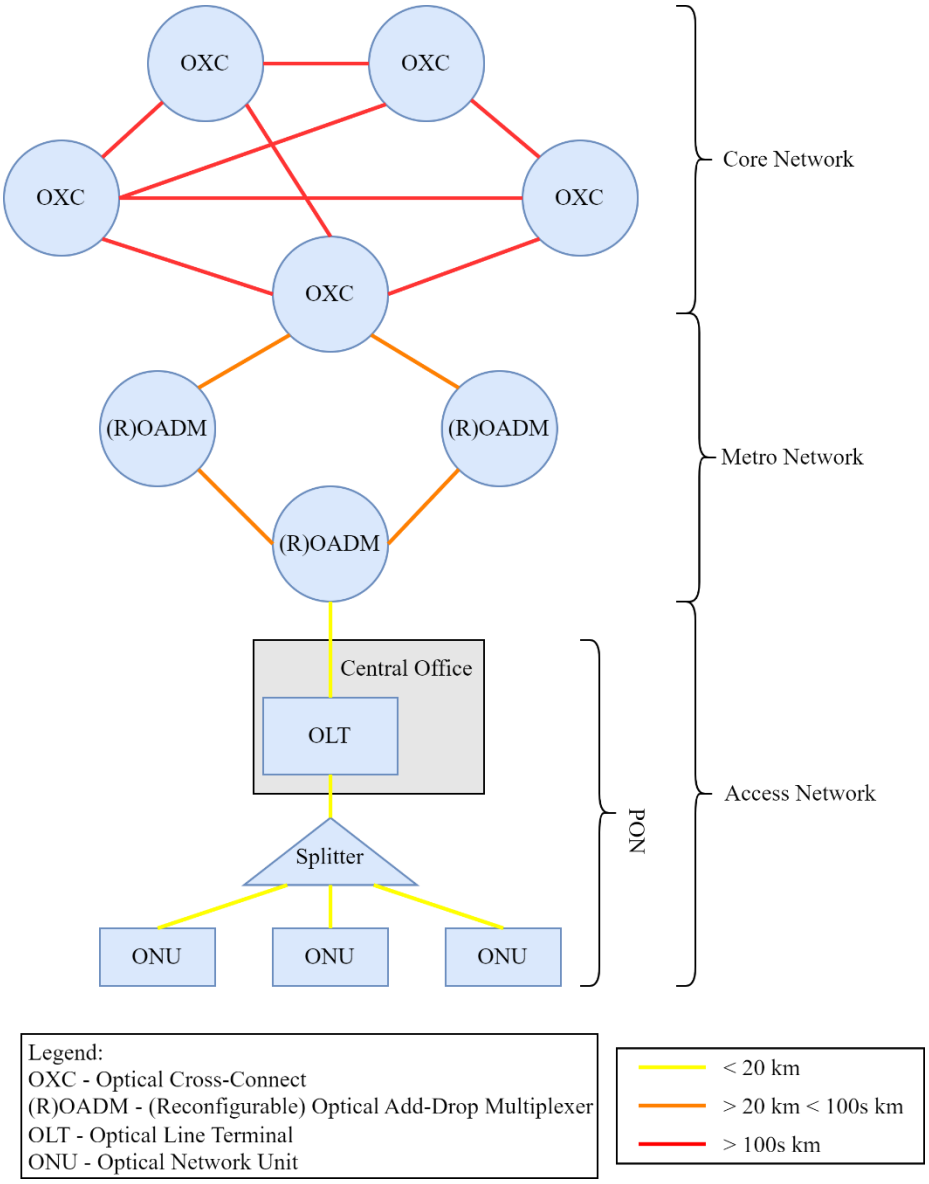


Figure 2: Optical Network's hierarchy.

Even though multi-core fibers (MCF) could be used on these networks segments, the system proposed in this dissertation is intended for data centers (DC), which will be introduced in section 2.3.

2.2. Receiver detection techniques

This section presents the main detection techniques used in today's receivers, such as direct detection (DD), coherent detection and self-coherent detection.

2.2.1. Direct detection

DD receivers dominated optical communication networks since the first generation [15]. The information is sent only through optical field intensity, using the on-off keying (OOK) modulation scheme, thus having a limitation in the maximum bit rate achievable. In the beginning of the 2010s there was a big focus on developing transceivers based on the coherent technology, which eventually replaced IM-DD based systems in some segments of the optical networks, namely the metro and core systems [3] [15]. DD systems are still used in access networks mainly to bring down cost, since the receiver design is much simpler, requiring a single photodiode [20]. However, with this need for higher bit rates, these networks must follow this evolution, or take the chance of becoming obsolete. With that in mind, several new proposals for DD systems appeared.

One proposal reported in [16] was to use pulse amplitude modulation (PAM) in order to send more bits per symbol, encoding the information in M intensity levels. The performance of this scheme is determined by the noise variance in each level, since if one level has a high noise variance, its symbols can be interpreted in the receiver as a level above or below. With a bit rate of 112 Gb/s using 4-PAM (M=4) modulation, results showed that the dispersion has an effect on the receiver sensitivity, with the penalty imposing a limit in the reach of the link, if the dispersion is left uncompensated. This modulation is proposed for short reach links such as intra data center links. Higher bit rates of 200 Gb/s were also accomplished with 4-PAM, although the reach was very limited, being up to 10 km in the O-band (1260-1360 nm) [21]. However, these systems could still be used in DCs and access networks.

2.2.2. Digital coherent

This technique was a breakthrough in optical fiber systems, allowing information to be coded in amplitude, phase, and polarization. Digital coherent receivers are the most advanced type of receivers and are characterized by having a local oscillator at the receiver for extracting the phase information of the signal [22]. They are also able to compensate channel impairments such as CD via DSP, which is one of the key enablers of this technique [23]. Although these are many advantages of using coherent technology, some disadvantages are associated with it, such as cost and complexity, requiring complex designs [22].

It has been experimentally demonstrated a 400 Gb/s per channel wavelength division multiplexing (WDM) transmission on the 50 GHz grid, over 990 km standard SMF with erbium-doped fiber amplification (EDFA) [24]. The WDM signal consisted of 8 channels using probabilistic shaping (PS) 64-QAM together with polarization multiplexing, resulting in a 414 Gb/s net bit rate per channel. This approach brings OSNR gain comparing to regular 64-QAM, while also providing a transmission distance improvement of about 83% over regular 64-QAM. This type of technology shows potential for longer-reach links such as the ones found in core networks.

In [25], it was demonstrated a 5 channel WDM transmission on the 125 GHz grid over 6400 km of ultra large effective area fiber (ULAF) which has a lower attenuation. Signal amplification was made via Raman amplifiers which, together with ULAF, increases the transmission distance more than three times than that of the standard SMF with EDFA. Dual polarization PS-16-QAM was employed, having a net bit rate of 640 Gb/s per carrier. As seen in the preceding experiment, PS-16-QAM also has a sensitivity and transmission gain when compared to regular 16-QAM. The reach improvement can be of about 48% when using Volterra non-linear equalization (VNLE) at the receiver, thus reaching the 6400 km mark. This type of system can be used effectively in ultra-long-haul WDM transmissions, such as transoceanic.

Long distance transmission of a WDM signal consisting of 5 channels with a net bit rate of 800 Gb/s each has been accomplished in [26]. Truncated probabilistic shaping (TPS) 64-QAM was used along with polarization multiplexing. This signal was transmitted along 2000 km of ULAF, being amplified with Raman amplifiers to maximize reach and using a Volterra equalizer. Similarly to the previous stated experiments, a sensitivity gain was obtained from using 100 Gbaud TPS-64QAM instead of 100 Gbaud regular 32-QAM. On that account, this type of system is appropriate for long-haul WDM transmissions.

2.2.3. Self-coherent direct detection

Between the traditional IM-DD and digital coherent solutions there is a technique that is a combination of both, and it is called self-coherent. The transmission of an optical carrier along with the data signal makes it possible for, in the receiver, using filters with a very narrow bandwidth, the recovery of the optical carrier, therefore employing it as a local oscillator [7]. Digital coherent systems require a LO at the receiver, while these ones do not, hence why they are called self-coherent. Essentially, it is a compromise between the two main solutions, having a trade-off between performance and complexity. These systems make higher order modulation possible in a direct detection manner.

A single sideband (SSB) 16-QAM system with signal-signal beat interference (SSBI) cancellation technique was proposed in [20] for metro and regional networks. The experiments were carried out with a channel bitrate of 25 Gb/s occupying only 8.75 GHz of bandwidth, over 560 km and 800 km of fiber. In DD systems, the phase information is lost because of the square-law photodiode so, to avoid that, the QAM signal is sent alongside an optical carrier at a nearby frequency, enabling the receiver to obtain the amplitude and phase information from the beating of the signal with the optical carrier. The SSBI cancellation technique is required to mitigate the signal-signal products, and these experiments showed that it was able to effectively reduce the nonlinear distortion caused by the square-law detection.

In [8], it was possible to transmit and correctly receive an 8 channel WDM signal, with a channel bit rate of 256 Gb/s, over 200 km of Corning TXF fiber. This was accomplished by sending, on each channel, a SSB 16-QAM signal together with a digital tone (virtual carrier), which is generated together

with the information and simplifies the implementation. Unlike the preceding proposal, this one does not have a gap between the information and the carrier. In the receiver, both two stage SSBI cancellation technique and Kramers-Kronig (KK) field reconstruction algorithm were experimented, with both working very well based on the system's performance margin, although the KK algorithm slightly outperformed the SSBI cancellation technique. This type of system shows potential for DC interconnect applications.

2.3. Data centers

The infrastructure for many online services such as on-demand video delivery, storage and file sharing are provided by DCs. These provide flexible access to scalable computing and storage resources, which today is a big necessity for cloud computing [27].

Large internet content providers (ICPs) have started to host and process a lot of data in hyperscale DCs in recent years. Virtualization and cloud computing caused the evolution of traffic patterns, resulting in a shift from north-south traffic, i.e., traffic from outside to the servers in the DC, to east-west traffic, i.e., traffic from servers to other servers in the same DC or other DCs nearby [16], which according to [28] would represent 85% of DC traffic by 2021. To better accommodate this type of traffic, these hyperscale DCs have shifted from a traditional three tier architecture to a flatter two tier architecture illustrated in figure 3. This requires high-capacity optical links inside the datacenters and between them, while keeping cost, power consumption and latency down [16][29]. DCs still use IM-DD technology, and for a long time the solution to scale the link's capacity has been to employ multiple wavelengths or parallel fibers while using OOK [16]. However, this is considered a limitative technology due to its inability to reach high bit rates like the coherent technology. Also, this combination of technologies is not scalable resulting in high cost, complexity, and power consumption [16], while the multiple parallel fibers solution takes too much space. So, to scale the fiber capacity while still using DD as the baseline technology, to minimize power consumption and cost, research was focused on more spectrally efficient modulation formats such as 4-PAM and orthogonal frequency division multiplexing (OFDM) [16].

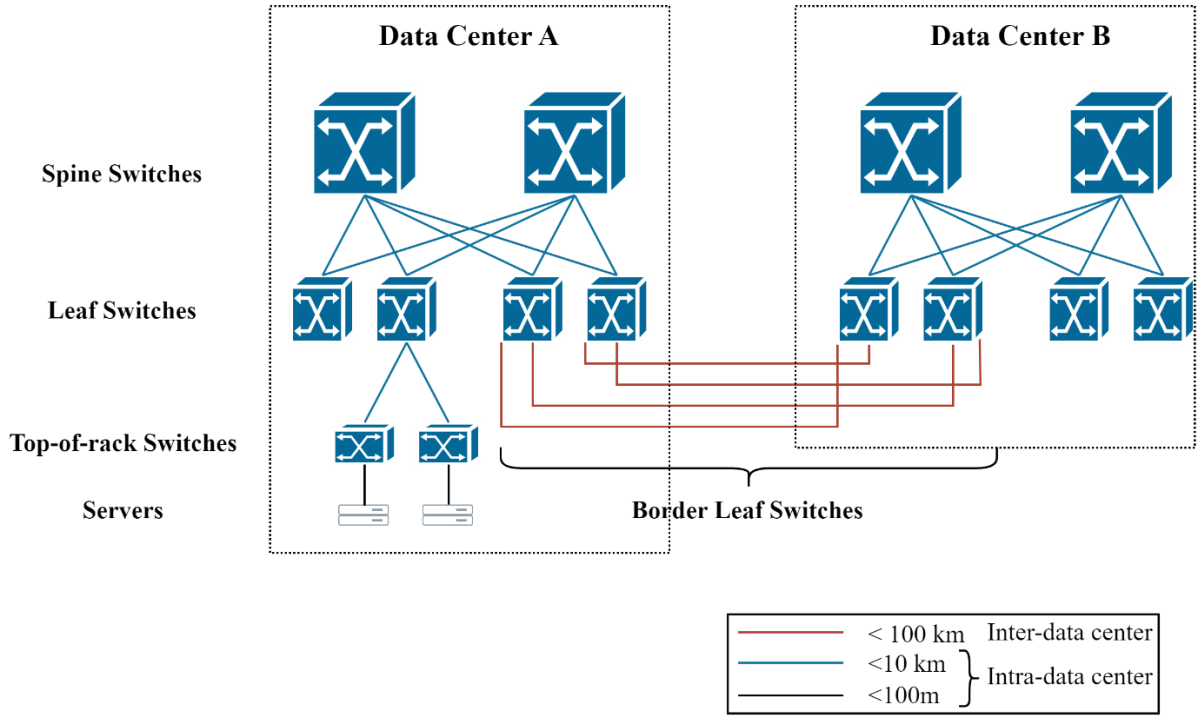


Figure 3: Hyperscale DC's network architecture.

There are two types of DC links: the intra-data center links which are of short reach and are used to connect servers to top-of-rack (TOR) switches (<100 m), TOR switches to leaf switches (>100 m and <10 km), and leaf switches to spine switches (>100 m and <10 km), and inter-data center links which are of longer reach and are used to connect the border leaf switches (<100 km). Intra-DC links have no need for amplification due to the short link lengths (up to 10 km), which keeps the cost down. To minimize the CD in links above 100 m and below 10 km, these systems operate in the O Band at around 1310 nm. In addition, several channels can be multiplexed with local area network (LAN) WDM or coarse WDM to increase system bitrate [16]. Intra-DC links are further differentiated depending on its distance [30]: short reach for 100 m distances based on multi-mode vertical-cavity-surface-emitting lasers (VCSELs) and parallel multi-mode fibers (MMFs), which will limit the data rate and reach due to modal dispersion caused by the multiple transverse modes not travelling at the same speed, thus causing inter-symbol interference (ISI) at the receiver; DC reach which generally reach up to 500 m distances using single-mode parallel fibers; far reach and long reach for up to 2 km and 10 km, respectively, using WDM and SMFs. As an example, by 2017, according to [31], Google's intra-data center optical interconnects of up to 100 m, operated with 50 Gb/s 4-PAM modulation using up to 8 lanes via space division multiplexing (SDM) using parallel MMFs, therefore being able to achieve a bit rate of 400 Gb/s. For longer reach intra-data center connections (> 100 m), SMF was used, while the 8 lanes were achieved via coarse WDM, thus eliminating the need for multiple fibers.

Inter-DC links have a longer reach of up to 100 km, so amplification is required. These systems operate on the C-Band at around 1550 nm for the ability of being amplified by EDFAs [16]. Operating

in this band has the disadvantage of having significant CD that needs to be compensated for using, for instance, dispersion compensating fibers (DCF). In this particular case, DD technology is still the better choice, due to rack space and power limitations, with 4-PAM modulation being the standard. This will allow a DWDM transmission with 40 channels at 100 Gb/s, thus reaching a total bit rate of 4 Tbps per fiber pair [32].

2.4. Space division multiplexing

There are several ways to increase fiber capacity such as the use of multiple transmission bands. This is currently done by transmitting in both C and L bands, while recent research is being focused on also including other bands such as the O, E, S and U-bands, in order to realize the so-called ultra-wide band (UWB) transmission. However, this is primarily indicated for long-haul systems where it is paramount to scale the fiber capacity as much as possible, unlike short-reach applications, where cable deployment costs are much lower, therefore it is preferable to multiply the capacity in another way, such as via SDM [2].

SDM is another solution to increase system bandwidth. Currently, it is used in the form of multiple parallel fibers, which takes too much space [16]. Parallel spatial paths incorporated into a single fiber are an evolution of SDM and are still compatible with WDM and signal formatting, allowing for high flexibility, increasing the bandwidth available and increasing the spatial efficiency [4]. There are a few types of SDM solutions for different applications such as MCFs and few-mode fibers (FMF) [1][4].

2.4.1. Mode division multiplexing

Mode division multiplexing is accomplished via FMFs or MMFs. FMFs are a simpler alternative to MMFs, utilizing only a small number of modes propagating in the fiber, theoretically being able to scale up to a few dozen modes within a standard cladding diameter of 125 μm, thus achieving a high spatial density, meaning a high number of spatial channels per unit area. They can be either weakly coupled (figure 4) or strongly coupled (figure 5), differing on whether it is designed to suppress mode coupling or to reduce differential mode group delay, respectively [5]. Its core diameter is larger than SMFs while strongly coupled FMFs have an even larger core diameter than weakly coupled ones. Figures 4 and 5 display the core diameters of commercially available 6-mode weakly coupled (step-index) and strongly coupled (graded-index) YOFC FMFs [55]. It is important to note that these values vary depending on the number of modes and manufacturer. Weakly coupled FMFs are designed to suppress mode coupling between different mode groups so, in order to recover the signals, only low complexity multiple input multiple out (MIMO) DSP is necessary at the receiver. The fiber may use a step index profile which will limit the channel count. Strongly coupled FMFs use graded-index profiles, therefore reducing the differential mode group delay and increasing the channel count. However, this will require complex

MIMO signal processing [5]. Along with MIMO signal processing, there needs to be a mode multiplexer/demultiplexer to separate the several transmitted modes into individual SMFs [29]. Core diameters for both types may vary but, in general, strongly coupled FMFs have larger core diameters compared do weakly coupled ones.

It has been demonstrated in [33], a transmission of 402.7 Tbps over a 48 km 10-mode weakly coupled FMF with standard 125 μm cladding. Each mode carried a WDM signal consisting of 747 channels in the C and L bands, employing polarization multiplexing and PS-16-QAM modulation, at a symbol rate of 12 Gbaud. A low complexity MIMO DSP was used at the receiver. This is four times more than the so-called capacity crunch.

More recently, a record transmission of 1.53 Pbps over a 25.9 km 55-mode MMF with a standard 125 μm cladding has been realized. In this case, only the C-band was used, which makes it a much more spectrally efficient approach than previous ones. Each mode carried a WDM signal consisting of 184 channels employing polarization multiplexing and 16-QAM modulation. A complex 110 × 110 MIMO DSP was required at the receiver [34].

The above-described experiments show that FMFs and MMFs are an interesting approach to increase capacity. Although the reach is limited due to the modal dispersion, high-capacity transmission is attainable, reaching bitrates 15.3 times more than the capacity crunch. However, these systems require more complex receivers, on account of the need of MIMO signal processing and mode multiplexers/demultiplexers.

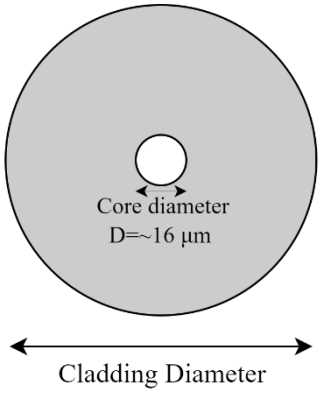


Figure 4: Weakly coupled FMF.

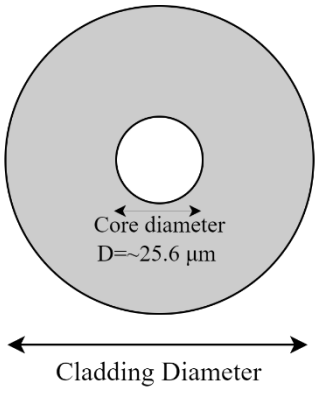


Figure 5: Strongly coupled FMF.

2.4.2. Multi-core fiber

MCFs gained more attention in the late 2000s due to the capacity of the SMF getting close to reaching its fundamental limit at around 100 Tbps [1][2]. MCF is applicable for short distance interconnects and long-haul links. Multiple experiments have been conducted for long haul links, however, due to some technology constraints, viable commercial options are still to be available for this type of systems [2]. Nevertheless, MCF-based systems are well-suited for short reach links, given that they do not require

optical amplification (the losses of MCFs can be kept low). The main impairments are inter-core crosstalk (ICXT) and relative signal propagation delay between cores (skew) [4].

Similarly to FMFs, there are two types of multi-core fibers: weakly-coupled and strongly-coupled. Figures 6 and 7 illustrate these fibers, referencing the core diameters of ideal homogeneous MCFs [5]. In weakly-coupled MCFs, multiple cores are incorporated in a single fiber with enough distance between each other to guarantee low ICXT. The typical core to core distance is around 40 μm [1], thus sacrificing core density, and the cladding diameter tended to be bigger than the standard 125 μm [5]. However, some experiments have already been carried out employing SM-MCFs in a 125 μm cladding [34][36][37][38], though the number of cores is kept low (4 or less). This type of MCFs can have low complexity MIMO DSPs at the receiver to recover the signal but work well without it [1]. Strongly-coupled fibers have crosstalk purposefully introduced between cores by decreasing the core to core distance, which is less than 30 μm [1], meaning it improves core density while also being able to maintain the 125 μm cladding [5]. This type of MCF needs MIMO signal processing at the receiver to recover the signal [1].

Spatial efficiency in MCFs should be higher than conventional SMF. The number of cores to be placed in a fiber is determined by the distances between each other, which in turn is determined by the maximum allowable crosstalk. Another limitation to the number of cores is the cladding diameter, since a smaller one is preferable due to its reliability for bending [1].

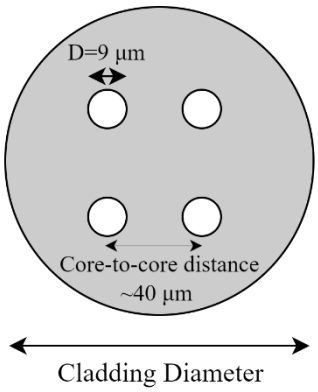


Figure 6: Weakly coupled SM-MCF.

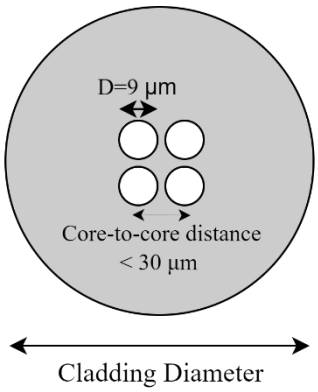


Figure 7: Strongly coupled SM-MCF.

2.4.2.1. Single-mode multi-core fiber

A possible application for single-mode MCFs (SM-MCFs) is using transceivers based on parallel single-mode fiber (PSM) transmission in data centers like the PSM4. In this type of transmission there are 4 links associated with each direction, so it is achievable with an MCF with 8 cores or two MCFs with 4 cores instead of 8 individual fibers, thus saving space [4].

In [34], a 4-core weakly coupled SM-MCF in a 125 μm cladding with 45 km was employed for a post-FEC error free single channel 200 Gb/s transmission using coherent technology. The modulation

format used was 16-QAM with polarization multiplexing at 32 Gbaud. A fan-in was used in the transmitter side, while a fan-out was used in the receiver side. The ICXT was significantly low, and the performance of this fiber was compared with a traditional SMF, showing similar performance. This type of system could therefore be used in the context of inter-data center links.

In [36], a 4-core weakly coupled SM-MCF was successfully employed to achieve an error free transmission of an aggregate 400 Gb/s bit rate (4 × 106 GbE) with FEC using 50 Gbaud 4-PAM modulation, reaching up to 12 km and making it applicable for intra-data center links.

Some high-capacity and long-distance experiments were also realized utilizing digital coherent technology. In [37], a transmission of 3 WDM signals with 359 channels, employing polarization multiplexing and 16-QAM modulation at a symbol rate of 24.5 Gbaud was realized over 2040 km of a strongly coupled 3-core SM-MCF, with a standard cladding diameter of 125 μm. The C-band was combined with the L-band to allow the increase of the number of multiplexed channels, thus reaching a total bit rate of 172 Tbps.

In [38], a transmission of 4 WDM signals with 152 channels employing polarization multiplexing and 4-QAM modulation at a symbol rate of 24 Gbaud was realized over 9150 km of a strongly coupled 4-core SM-MCF, with standard cladding of 125 μm. In this case, only the C-band was used, achieving a 50.47 Tbps total bit rate. Even though it is unable to achieve the same capacity as the above referenced experiment, it has more than 4 times the reach. This MCF could be used in submarine cable systems scenarios, where there are space limitations, and high capacity and reach are necessary.

In [1], experiment results for weakly-coupled SM-MCFs have shown that there is a tradeoff between high spatial efficiency and low crosstalk. While one experiment led to a spatial channel count (SCC) of 7 and a worst-case crosstalk of -64 dB in a 100 km link at 1550 nm, another one managed to reach an SCC of 30 with a worst-case scenario of -42 dB crosstalk also in a 100 km link at 1550 nm, thus confirming the tradeoff scenario. The experiments reported in this subsection demonstrate that, unlike FMFs, high-distance and high-capacity transmission is achievable on MCFs, thus making this type of fiber more versatile. Furthermore, MCF transmission does not require complex receivers.

2.4.2.2. Few-mode multi-core fiber

Few-mode multi-core fibers (FM-MCF) are another type of SDM. They are similar to figures 6 and 7, but the cores are larger in order to accommodate more modes. It is a result of the combination of core multiplexing with mode multiplexing. With it come more limitations, namely in long-distance links where complex MIMO DSP is necessary to separate coupled modes during propagation [1]. Basically, the complexity of the DSP will determine the number of modes allowed, low complexity DSP means lower number of modes and vice-versa. The distance between cores needs to be larger than in SM-MCFs to achieve the same amount of crosstalk, since there are several modes of propagation in each core which is a cause for higher crosstalk.

In 2019, a 1.2 Pbps single-span transmission was achieved over 3.37 km FM-MCF [39]. This fiber had 4-cores with 3-modes each and a 160 μm cladding, therefore being larger than the standard ones. In the L-band at around 1608.8 nm, the ICXT between modes of adjacent cores ranged between -54 dB/km and -31 dB/km. The C+L bands were used with channels having a 25 GHz spacing and containing a 24.5 Gbaud polarization multiplexed 256-QAM signal, which were combined to form a WDM signal with 368 channels that would be replicated over 12 spatial channels (4 cores times 3 modes), thus reaching the 1.2 Pbps bit rate.

A 10.66 Pbps transmission was reported in [40]. In this transmission, a weakly-coupled 38-core 3-mode FM-MCF with a length of 13 km was used. This fiber was designed to have low differential-mode delay within each core, in order to achieve higher throughputs. Furthermore, being weakly coupled with 38 cores, resulted in having a much larger cladding diameter of 312 μm. Both C and L bands were used in each spatial channel. i.e., each mode of a core, to transmit 369 WDM channels in a 25 GHz grid, each with a 24.5 Gbaud 64-QAM or 256-QAM signal. Employing FEC, an aggregated 10.66 Pbps data-rate was attained, while each spatial channel carried, on average, a 93.5 Tbps signal. This means that a bit rate 106.6 times greater than the capacity-crunch was achieved. It is important also to note that the ICXT was kept below -35 dB.

In a 100 km link at 1550 nm, experiment results for weakly-coupled FM-MCFs showed a SCC of 12 (4 cores with 3 modes each) with a crosstalk of -39 dB and an SCC of 114 (19 cores with 6 modes each) with a crosstalk of -19 dB. There were also experiments achieving lower crosstalk of around -55 dB to -58 dB with an SCC of 36 [1]. These results, despite demonstrating FM-MCFs are able to transport a higher number of channels, also show a much higher crosstalk than in SM-MCFs that needs to be compensated for, in another way. Nevertheless, this type of fiber allows for the highest capacity transmissions, although the reach is still very limited and complex receivers are required.

2.5. Performance degradation contributors

This work studies the use of neural networks to compensate the ICXT on multi-core optical fiber systems also impaired by LPN, making it crucial to understand both these concepts.

2.5.1. Inter-core crosstalk

ICXT is an important problem in MCFs and is caused by the optical power launched in one core being coupled with neighboring cores during propagation [1]. It varies randomly along the fiber over time and frequency [41].

Studies have shown that the impact of the ICXT depends on modulation format, data-rate, temporal skew between cores (time delay) and type of optical receiver employed. DD systems that use carrier

supported signals are impaired by ICXT power, which may fluctuate significantly over time [41]. In [49], the impact of the ICXT on the performance evolution of OOK-DD systems along time was evaluated for low and high skew×symbol rate. It was shown that the impact of the ICXT on the received symbols depends on the skew×symbol rate. When the skew×symbol rate << 1, the ICXT induced in the interfered core only depends on the symbol transmitted in the interfering core at the same time instant, and the received amplitudes showed two levels for each bit, with the levels' spacing becoming larger with the increase in short-term average crosstalk (STAXT), which is the average ICXT power measurement over a certain short period of time. When skew×symbol rate >> 1, the ICXT induced in the interfered core results from the contributions of several symbols transmitted in the interfering core, and the received amplitudes presented a noise-like behavior. Considering the same outage probability, i.e., the probability that the BER exceeds a given BER threshold, it was shown that when skew×symbol rate >> 1, there is an additional ICXT tolerance of around 3 dB, so, the rule of thumb when using a OOK-DD MCF-based system impaired by ICXT, is to use cores with a skew longer than the symbol period to transmit information.

There are different techniques to suppress ICXT. One way of decreasing it is reducing the coupling coefficient. In weakly coupled MCFs this coupling coefficient is much lower than in strongly coupled MCFs, because it is directly related to the distance between cores. To reduce it, one proposal used trench assisted MCFs. In trench-assisted MCFs, a low-index trench layer of a determined thickness is put around the core, therefore suppressing the overlap of electromagnetic fields between cores. Results have shown a 30 dB crosstalk reduction. Another approach is to have the multiple cores with intrinsic index difference between them, also known as heterogenous MCFs, and a third way is by using the propagation-direction interleaving (PDI), where adjacent cores have a different direction. The combination of the aforementioned techniques results in low crosstalk and high core count [1].

2.5.2. Laser phase noise

Light emission is done by the photon generation, which has two fundamental processes: stimulated emission, where all photons have similar energy, phase and propagation direction, and spontaneous emission, where photons are emitted in random directions without any relation between its phases. Ideally, in the case of optical systems, a laser source emits light through stimulated emission, resulting in a monochromatic (coherent) oscillator with a certain frequency. However, in reality, all light sources, including lasers, generate photons though stimulated and spontaneous emission, resulting in a noisy oscillator. This noise results in a non-monochromatic (incoherent) oscillator, whose power spectral density (PSD) of the optical field at the laser's output is not represented by a single line, instead having a certain linewidth which is associated with the phase fluctuation of the laser [43][44].

There are several types of lasers, such as single-mode and multi-mode lasers. Fabry-Perot lasers are multi-mode lasers with a linewidth between tens and hundreds of GHz, having a limited bit-rate×reach

product [43]. Commonly used single-mode lasers include distributed feedback lasers (DFB) and external cavity lasers (ECL), which usually have a linewidth of a few MHz and tens of kHz, respectively.

In [50], the impact of the LPN in the instantaneous ICXT power, STAXT and performance of OOK-DD weakly coupled MCFs was studied for a skew×bit rate >> 1. Simulation results showed that the level of the fluctuations of the STAXT power were higher for smaller linewidths. They also showed that the amplitude of the fluctuations of the instantaneous ICXT power increased with the increase of the laser linewidth. The experimental results corroborated the simulation results, showing a higher outage probability for lasers with higher linewidth, owing to the fact that the errors on the OOK bit stream are dependent on the instantaneous ICXT power, which can be higher for higher LPN. For that reason, the level of the fluctuations of the STAXT was shown to be a bad metric to characterize the system performance. It was also concluded that this type of system operating with a DFB laser may need an extra ICXT margin of up to 8 dB comparatively to ECL lasers. In [51] similar results were obtained, revealing that when the skew×symbol rate << 1 and skew×linewidth << 1, the effect of the LPN on the instantaneous ICXT and STAXT is negligible. In contrast, when skew×linewidth and skew×symbol rate >> 1, considering a laser with a 10 MHz of linewidth, the instantaneous ICXT power reveals much larger fluctuations when compared with an ideal laser, proving that the LPN can have a big impact on the ICXT. When analyzing the BER for the case when skew×symbol rate and skew×linewidth << 1 - lower variation of the instantaneous ICXT and insignificant impact of the LPN on the ICXT - it was also concluded that the dispersion only effects the BER when a laser of high linewidth is used in the interfered core. Moreover, higher laser linewidths result in higher BERs, with results implying that when these are used in the interfered core instead of the interfering one, a slightly higher BER is observed.

2.6. Machine learning

Machine Learning (ML) is based on the idea that patterns and trends in a set of data can be learned automatically through algorithms [45]. This learning process makes it possible that in the future, predictions and decisions be made with other data but with the same objective. Over the years, as a result of the increase in computing power, ML began being used in multiple areas of investigation, including optical communications. It has applications in different layers of optical communications systems, namely in the physical and network layer [46].

ML algorithms mainly perform two types of pattern recognition: trying to find some functional description in the given data, with the aim of predicting values for new inputs in the future – regression problem; trying to find decision boundaries in the data so that different data is indeed classified differently – classification problem. In the context of optical communications, parameter estimation and symbol detection are regression and classification problems, respectively, therefore being possible to resolve via ML. As high-capacity links are increasingly getting limited by transmission impairments like fiber nonlinearity, explicit statistical characterizations of inputs/outputs become difficult.

Consequently, there has been an increasing focus on ML techniques to compensate system impairments, for example: fiber nonlinearity, CD, laser chirp and transceiver components imperfections [45].

2.6.1. Unsupervised learning and Reinforcement learning

In unsupervised learning, the algorithm does not know what the correct output should be, instead needing to come up with one. It requires unlabeled data, i.e. no context is provided to the ML algorithm, with that being the main difference with supervised learning, which requires labeled data. The algorithm will learn the similarities between the various inputs to either group them together or determine a better representation of the original data. This type of algorithm is becoming increasingly more important since in real circumstances it is hard to obtain labeled data for the training process. They are mainly used for clustering and features extraction but could also be used in combination with supervised algorithms, with the raw data going through it beforehand to extract some useful features. Algorithm examples include k-means clustering, which is able to cluster information, and principal component analysis (PCA), which is a technique for features extraction [45].

In reinforcement learning (RL), the algorithm interacts with the environment in order to learn something, with every action affecting the environment and returning a reward that will quantify its success [46]. The input of the algorithm, called observation, is associated with the reward or reinforcement signal. The output, also called the action, will determine the next observation. Every action has a direct relationship with the reward, with the model being rewarded with a good output and punished with a bad one. It will try different actions until it learns a set of parameters that lead to better rewards, therefore learning to choose the actions that will give better results [45]. Examples include deep Q learning (DQN) that uses deep neural networks (DNN) to estimate the optimal action-value function, and deep deterministic policy gradient (DDPG) which is inspired on DQN [46].

2.6.2. Supervised learning

In supervised leaning, labeled data is required so that the algorithm is able to use it as a target when learning [46]. The ML algorithm will generalize them to provide the best possible output to new input data. It can be considered a closed-loop feedback system since the error between the algorithm's output and the target output will be used in the learning process [45].

2.6.2.1. Feed-forward neural networks

Neural Network's (NN) structure is inspired by an animal's brain, being formed by several layers of neurons that are connected via edges with a determined weight. The most commonly used type of NN in ML is the feed-forward NN (FFNN). This type has an input layer, whose inputs consist of features previously obtained from the data by humans (human crafted features), followed by one hidden layer

and then an output layer (figure 8). The hidden layer neurons have a function associated with them called the activation function, which is applied to all its inputs, with the result being forwarded to the next layer. The output layer can have one or multiple neurons [46].

The choice of structure for a NN, like the choice of the activation function or the number of hidden layer neurons, is often done in trial and error, turning NNs into black box like systems due to the difficulty of interpretation by humans [46]. The most commonly used activation functions have for a long time been the sigmoid and hyperbolic tangent also known as the tangent sigmoid [45].

The training process determines the optimal set of weights that resolve a given problem and are able to resolve other similar problems in the future that have different data from the training one. There are different possibilities for the training process, the most common being backpropagation (BP) and gradient descent. A loss function also must be chosen to measure how far the prediction is from the target output, with the most common being the mean squared error (MSE) [46]. Several datasets are used in this process. The training dataset typically has the size of ten times the number of weights. The validation dataset, usually a third of the size of the training dataset, is employed to check on how well the NN is working, looking for possible cases of overfitting, in which the NN does not generalize, instead fitting perfectly to the training data, or underfitting, which is the opposite of overfitting. Finally, the testing dataset will evaluate the performance of the NN [45].

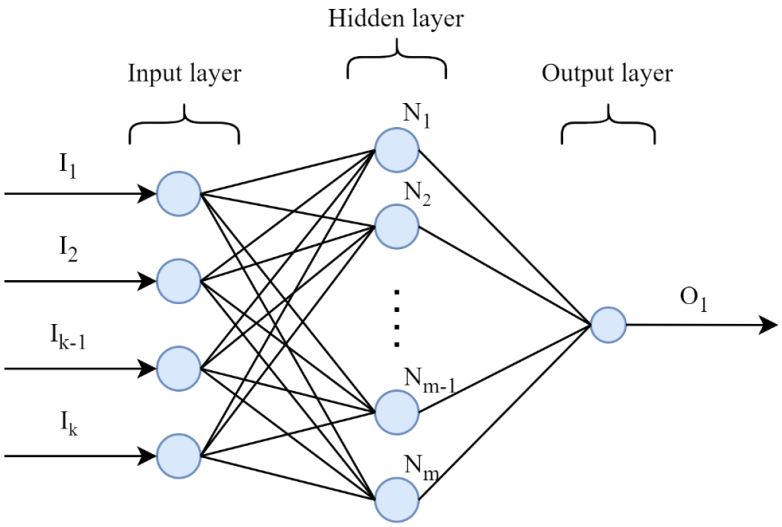


Figure 8: Architecture of a regression FFNN.

2.6.2.2. Recurrent neural networks

Recurrent Neural Networks (RNN) are characterized by performing pattern recognition on sequential data. This data has important spatial/temporal dependencies to be learned, so to enable the memorization of information, these networks have feedback connections. This will allow the RNNs to analyze sequential data by taking advantage of their inherent memory. A RNN (figure 9) has an input i(t), a hidden state h(t) in the middle that represents the memory of the network, and an output o(t). The edges

that connect these nodes all have different weights, but these weights will be replicated in the unfolding process across all time periods, since the same task is being performed at each time period, but with different inputs. These are limited by the vanishing gradient problem but also by the so-called exploding gradient problem during training. The solution is to use a special type of RNN architecture, the long short-term memory networks (LSTM), which can model and learn temporal sequences and their dependencies more accurately. It will decide to store or delete the information based on the importance it attributed to it [45].

These are suitable for channel characterization and data detection in nonlinear channels with memory such as long-haul transmission links with fiber nonlinearity, direct detection systems with CD, chirp or other component nonlinearities. It can also be useful in network traffic prediction [45].

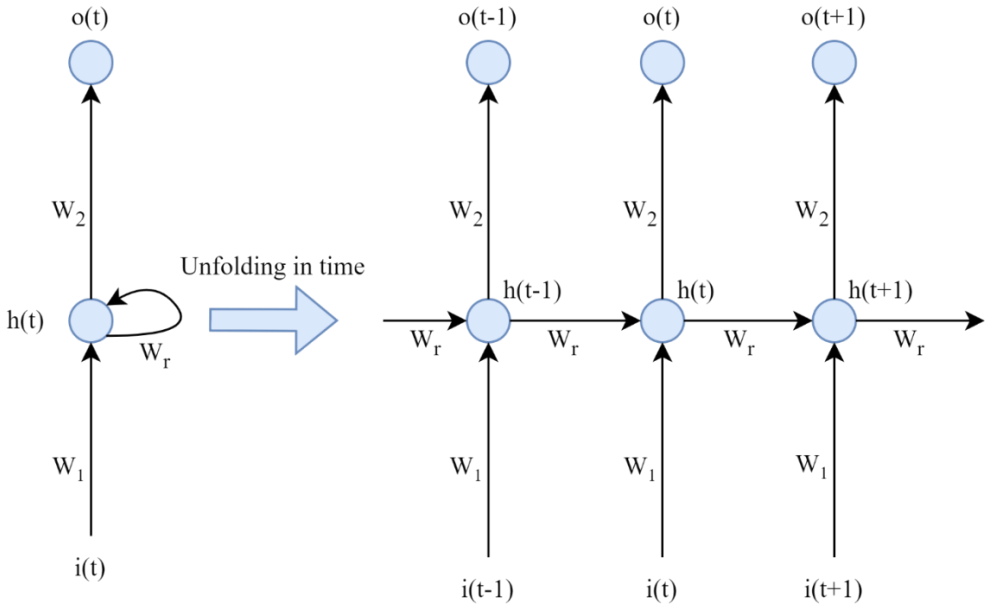


Figure 9: Architecture of a RNN.

2.6.2.3. Deep learning

The increase in computational power, recent breakthroughs in theory and algorithms together with the abundance of information to work with contributed to an advancement in ML, originating deep learning, a new type of ML that shows superior performance to previously mentioned supervised learning concepts. These algorithms have a deeper architecture, meaning more than one hidden layer and possibly containing more complex structures such as feedback and memory [45]. Examples of deep learning are DNNs and Convolutional Neural Networks (CNN).

Deep learning (DL) algorithms automatically learn and extract high level features of data from lower-level ones as the input propagates through the algorithm's various layers. Different learning stages are distributed through its various layers at an increasing level of complexity, essentially splitting a certain problem in different parts, thus allowing the system to learn complex relationships between inputs and outputs directly from the original data itself, instead of human crafted features. For example,

considering a conventional NN, such as the FFNN, it is possible to obtain the OSNR from a given eye diagram. The FFNN's inputs, however, are the variances of the "1" and "0" levels, and the eye's height and width (human crafted features). With a DNN, which is essentially a DL based multilayer NN, the input can just be an image of the eye diagram since the algorithm will learn by itself OSNR features without human intervention [45].

2.6.2.4. Deep neural networks

One of the main differences between DNNs and conventional NNs is that the former contains multiple hidden layers (figure 10). Regularly used BP and gradient-based learning methods used for training are not effective in multiple layer networks due to the vanishing gradient problem. This results in different layers learning at different speeds which will lead, in a worst-case scenario, to the network stopping its learning process. Some solutions have been proposed, including: using the rectified linear unit (ReLu) activation function; pretraining of the network layers one at a time, later fine-tuning the entire network with the BP algorithm; and avoiding gradient methods [45].

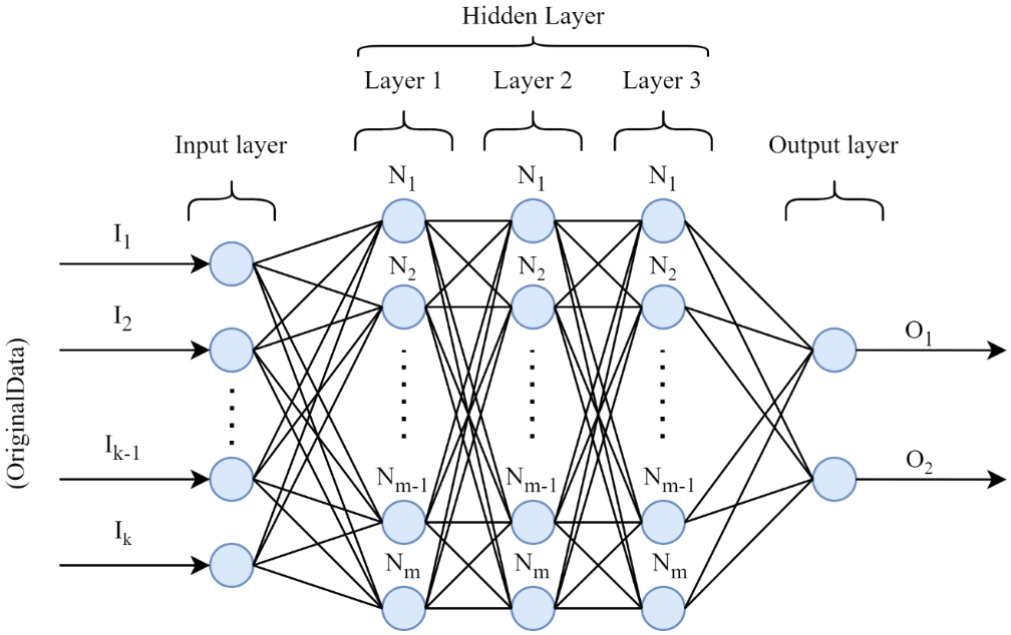


Figure 10: Architecture of a classification DNN.

2.6.2.5. Convolutional neural networks

CNNs are mainly used for pattern recognition within images, but also have their use in other areas such as speech recognition, natural language processing and video analysis. The name derives from the fact that in some of the hidden layers, convolution is used instead of general matrix multiplication. Its typical structure (figure 11) comprises of a few alternating convolutional and pooling layers, with a standard

FFNN-like structure at the end that will provide the output. The convolutional layer usually has a multidimensional array of data as the input, for example an image, and will output a feature map after the convolution operation. This operation is done between the input (the image) and a certain filter (or kernel), where the filter is obtained after the training phase and will manage to extract features. It is important to acknowledge that the more filters a CNN has, the more feature maps it will return. After each convolutional layer, there is a pooling layer that will replace the outputs with a summary statistic of the nearby outputs. A typically used operation in this layer is the max pooling, which will operate over the feature maps picking the largest value out of a certain number of neurons, essentially down-sampling the feature maps reducing its size, therefore resulting in a reduction of computational load. The training process uses a modified BP algorithm which updates convolutional filters' weights [45][47].

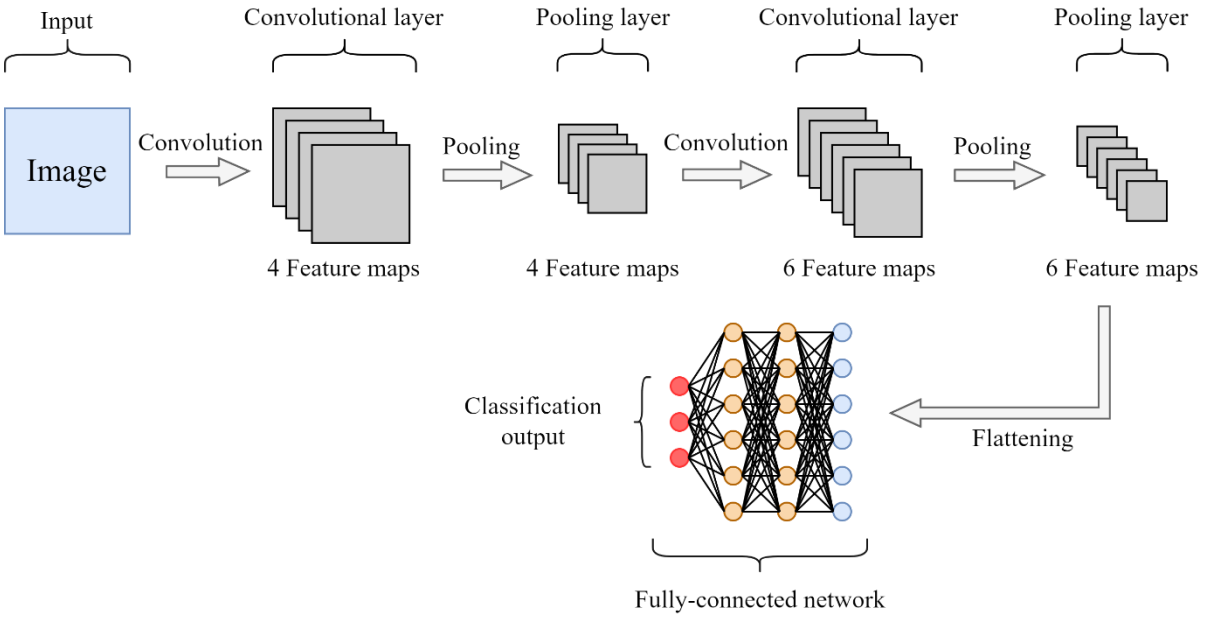


Figure 11: Architecture example of a classification CNN.

2.6.3. Applications in optical communications

In optical communications, supervised algorithms are more predominantly used but RL algorithms also have some use. For physical layer applications like the compensation of nonlinear effects, supervised regression algorithms are usually employed, whereas network layer applications tend to use supervised classifiers or RL algorithms [46].

Optical fiber communications suffer from nonlinearity due to the Kerr effect, rendering classical linear systems equalization suboptimal. The majority of nonlinear compensation (NLC) methods work by solving or approximating the nonlinear Schrödinger equation's (NLSE) solutions, which models the signal propagation in the presence of Kerr nonlinearity in the fiber, with a tradeoff existing between

complexity and the degree of compensation [9]. Building a black-box model of the transmission link is a simpler task and is possible recurring to ML algorithms, therefore ignoring the complicated interactions between nonlinearity, CD and noise in the NLSE. With this in mind, a NN solution with one hidden layer that applies a technique called extreme learning machine (ELM) was proposed in [48]. This ML model has as inputs the *I* and *Q* components of the signal after CD compensation in the coherent receiver. Simulations have shown that this ML model achieves similar results to other NLC techniques such as digital back propagation (DBP), a capable procedure but of complex computation, due to the interactions between signal, nonlinearity, CD and noise. In [9], a DNN with two hidden layers was studied, where the scaled exponential linear unit (SELU) activation function was used in order to avoid the vanishing and exploding gradient problems referenced before. This DNN has as inputs the received symbols that were affected by the nonlinearities in the fiber (real and imaginary parts are fed separately into the DNN) and other relevant information related to nonlinearities, outputting the real and imaginary parts of the estimated nonlinear perturbation, which will make the compensation possible. It is preferable than the ones which existed before due to it being system agnostic, making it versatile and adaptable.

Since all linear channel effects such as CD are converted into nonlinear ones by the square-law detection in DD, it is challenging to properly compensate the signal distortion. In [10], different neural-network-based nonlinear equalizers were investigated in an experiment with 20 km of standard SMF and a 50 Gb/s 4-PAM signal using a DD system, therefore being a possible solution for data center interconnections. The four types of neural networks used were the FFNN, radial basis function neural networks (RBF-NN), auto-regressive RNNs (AR-RNN) and layer RNNs (L-RNN). Only one hidden layer was considered for all NNs since it is enough to achieve efficient equalization. In terms of the output, a single neuron was employed to reduce the computational complexity. The inputs contained the 4-PAM symbol received and some other past and post symbols. The activation function employed in the input and hidden layer was the hyperbolic tangent function (tanh) while the output layer used a pure-linear function. Results showed that there is a tradeoff between computational complexity and BER performance. Considering all NNs have the same number of inputs and hidden neurons, it was found that AR-RNN achieved the best performance, while the FFNN had the lowest computational complexity. With 11 input and 7 hidden layer neurons, all NNs showed good BER performance while an acceptable performance was also attained with 5 and 4 respectively. Overall, the AR-RNN performed better, with the FFNN being the worst. Furthermore, it was proven that all of them are superior in terms of performance to conventional DD without resorting to ML algorithms.

As mentioned before, ICXT is one of the main impairments of MCFs, varying randomly along the fiber length, time and frequency therefore affecting the system's performance. In order to minimize this impact, some machine learning techniques were proposed in [11], namely one unsupervised learning technique - k-means clustering - and also two supervised ones – FFNN and k nearest neighbor (KNN). The NN employed had four inputs consisting of the *I* and *Q* components of both the interfered and interfering cores at the end of the KK receivers, as well as 10 hidden layer neurons with the output layer

predicting the *I* and *Q* components without the transmission impairments. The experiment involved the transmission of a 16-QAM 64 Gbaud signal on a 35 km 2-core MCF with an ICXT level of -13 dB. Results indicated that k-means clustering and KNN algorithms do not provide any notable performance improvement. However, the FFNN is able to mitigate the ICXT-induced BER degradation, allowing a BER below the defined threshold. It was also shown that for received optical powers (ROP) below -14 dBm the FFNN did not provide any performance improvement, while a ROP above -10 dBm allowed the system to achieve a mean BER below the defined threshold. Regarding the skew×symbol rate, in [54], results showed that these NNs perform better when the skew×symbol rate << 1. This work considered ideal lasers which do not exist in real systems. As explained on subsection 2.5.2, real lasers are affected by phase noise, which can significantly impact the performance of a system. On account of that, the impact of real noisy lasers on the performance improvement provided by these NNs needs to be assessed, with that being the main objective of the work to be developed in this dissertation.

2.7. Summary

In this chapter, the main concepts of this work were presented as well as some studies regarding them. Self-coherent detection is shown to have potential for DC interconnects since it allows for the transmission of higher order modulation formats, therefore higher bitrates, as well as reaching distances of hundreds of kms, while keeping complexity and cost down. MCFs are applicable for either short or long-distance transmissions, making it a versatile solution, and suitable for DCs. The ICXT is introduced and some existing techniques to suppress it are mentioned. The LPN is also introduced, and the main types of lasers are mentioned, including their typical linewidth. Furthermore, when the skew×linewidth << 1 (the case considered in this work), the LPN has an insignificant impact on the ICXT. FFNNs, which will be used in this work, are shown to have been successfully employed in optical communications, namely compensating transmission impairments, optimizing network performance, and mitigating the ICXT in systems employing ideal lasers when skew×symbol rate << 1.

CHAPTER 3

Architecture of the NN-assisted self-coherent MCF system

This chapter describes the architecture of the 200 Gb/s short-reach self-coherent MCF system employing NNs. Section 3.1 gives a brief overview of the system to be studied by numerical simulation. Section 3.2 details the model of the transmitter, namely the filter and the modulator, while also presenting and validating the simulation model for the laser phase noise of real lasers. Section 3.3 presents and validates the transmission model for weakly coupled MCFs. Section 3.4 introduces the KK algorithm of the KK receivers employed in this work. Finally, section 3.5 presents a summary of the chapter.

3.1. System architecture

Figure 12 illustrates the architecture of the 200 Gb/s short-reach self-coherent MCF system employing NNs, considered in this work. This system is composed of two optical transmitters that produce different SSB optical signals and inject them into each of the two cores of the weakly coupled MCF, the interfering and interfered cores *m* and *n*. Once the optical signals have travelled through the MCF, they will reach the receivers, one for each core, where the received optical signal is photodetected and digitally processed.

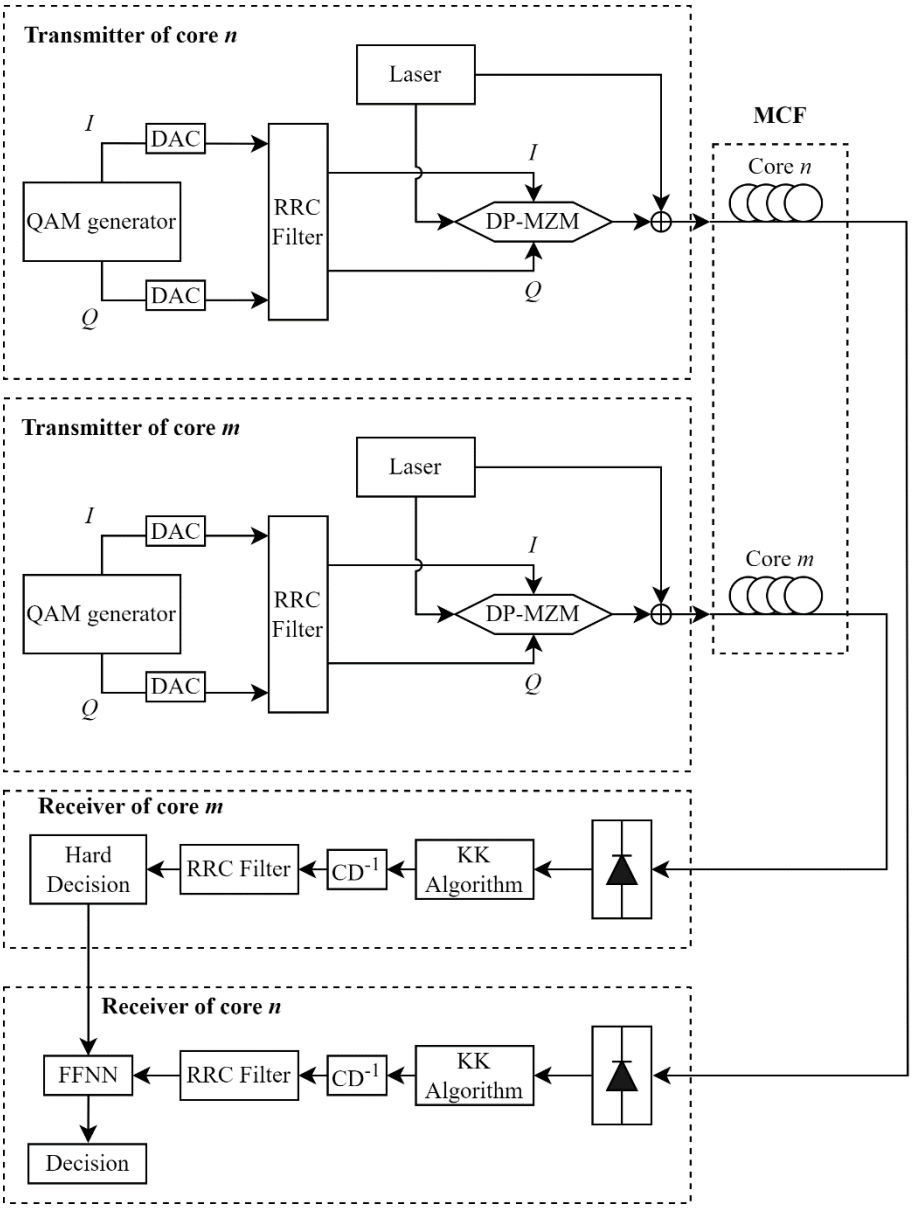


Figure 12: Architecture of the short-reach self-coherent 2-core fiber system employing NNs.

3.2. Optical transmitter

This section describes the blocks of the optical transmitter. In particular, the equations implemented in the simulator to model each component of the transmitter are presented. The transmitter block, which is shown in figure 12, includes a QAM signal generator that generates 16-QAM symbols, which are separated into the in-phase and quadrature components, subsequently going through a digital-to-analog converter (DAC). Once this is done, Nyquist shaping is applied to the signal waveform by using a root-raised cosine (RRC) filter. The *I* and *Q* components of the signal waveform at the filter output are then fed separately into the dual-parallel Mach-Zehnder modulator (DP-MZM). The optical signal injected in the DP-MZM is obtained through a non-ideal laser characterized by the laser linewidth.

3.2.1. Root-raised cosine filter

The RRC filter is present in both transmitter and receiver, in order to generate pulses with Nyquist shaping at the output of the receiver, thus minimizing the ISI at the optimal sampling time. The filter is implemented according to the following:

$$H_{RRC}(f) = \begin{cases} 1, \ \left(0 \le |f| < \frac{1-\rho}{2T}\right) \\ \sqrt{\frac{1}{2}\left\{1 + \cos\left[\frac{\pi T}{\rho}\left(|f| - \frac{1-\rho}{2T}\right)\right]\right\}}, \ \left(\frac{1-\rho}{2T} \le |f| \le \frac{1+\rho}{2T}\right) \\ 0, \ \left(|f| > \frac{1+\rho}{2T}\right) \end{cases} \quad (1)$$

where ρ is the roll-off factor, T is the symbol period and f is the frequency. It is also important that the receiver has the filter with the same parameters as the ones used at the transmitter side, in order to attain minimal ISI at the optimal sampling time.

3.2.2. Dual-Parallel Mach-Zender Modulator

The DP-MZM has the ability to perform the modulation of the *I* and *Q* components in order to have QAM signals at its output. It is composed of two arms, each of which having a Mach-Zehnder modulator (MZM), one for each component, biased at the null-point in order to suppress the carrier. The DP-MZM still includes an outer MZM, biased at the quadrature-point, to guarantee an optical phase difference of 90 degrees between the lightwaves at the output of each inner MZM. The optical field at the output of the DP-MZM is defined as:

$$e_{out}(t) = \frac{E_{in}}{2}\left[e^{j\frac{\pi}{2V_{sv}}V_{b,3}}\ \frac{e_1(t)}{E_{in,1}} + e^{-j\frac{\pi}{2V_{sv}}V_{b,3}}\ \frac{e_2(t)}{E_{in,2}}\right] \quad (2)$$

where E_{in} is the optical field of the laser at the DP-MZM input, V_{sv} is the switching voltage which varies for different MZMs, $V_{b,3}$ is the bias voltage of the outer MZM, $e_1(t)$ and $e_2(t)$ are the optical fields at the output of each inner MZM, and $E_{in,1}$ and $E_{in,2}$ are the optical fields at the input of each inner MZM.

$$e_{1,2}(t) = \frac{E_{in,1,2}}{2}\left(e^{j\frac{\pi}{2V_{sv}}v_{1,2}(t)} + e^{-j\frac{\pi}{2V_{sv}}v_{1,2}(t)}\right) \quad (3)$$

where $v_{1,2}(t)$ is the signal applied to either the upper or lower arm, being given by the following equation:

$$v_{1,2}(t) = V_{b,1,2} + v_{AC,1,2}(t) \quad (4)$$

where $V_{b,1,2}$ is the bias voltage applied to the inner MZMs, which should correspond to the null-bias point, and $v_{AC,1,2}(t)$ is the electrical signal that contains the information for either the upper or lower arm.

3.2.3. Laser phase noise

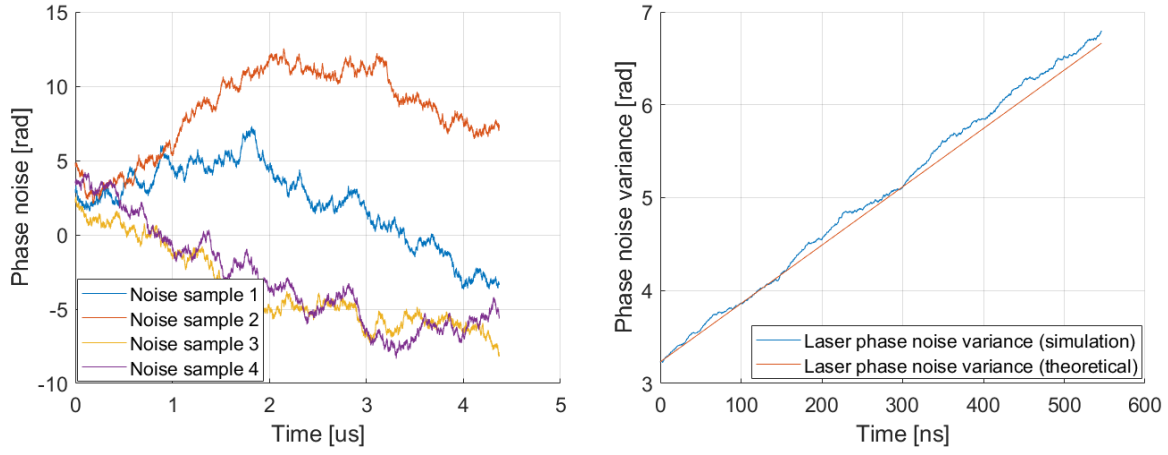
The purpose of this work is to evaluate the impact of the LPN on the performance improvement provided by NNs. To do this, a general LPN model is implemented, as described in [53]. No intensity noise is considered. This model to describe the LPN is characterized by a Wiener (Brownian motion) process with zero mean and a variance that increases with time, given by $2\pi\Delta\nu_l t$, where t is time and $\Delta\nu_l$ is the linewidth of the laser spectrum at half power (-3 dB). The phase noise is the integral of the frequency noise, as illustrated in the following equation:

$$\phi(t) = 2\pi \int_0^t \mu(\tau)\, d\tau, \quad t > 0 \tag{5}$$

where $\mu(\tau)$ is the frequency noise. This frequency noise is gaussian, having zero mean and a PSD given by $\frac{\Delta\nu_l}{2\pi}$. For t=0, $\phi(t)$ is uniformly distributed between 0 and 2π.

The LPN is characterized by the coherence time of the laser field, which is a time interval where the phases of the laser field at different time instants have some correlation. If the time interval between those time instants exceeds the coherence time, the phases become weakly correlated or uncorrelated [50]. The coherence time is given by [50]:

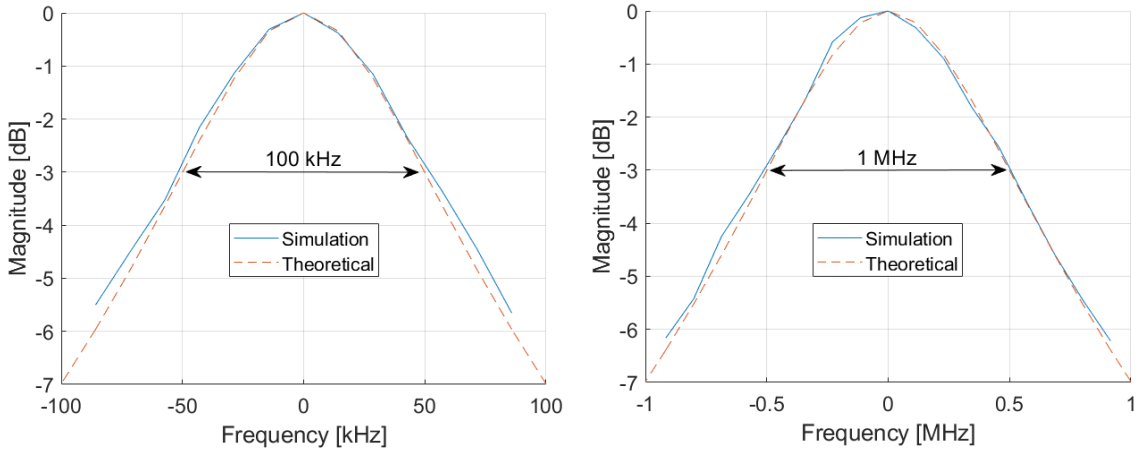
$$T_c = \frac{1}{\pi \cdot \Delta\nu_l} \tag{6}$$



a) Evolution of the phase noise with time. *b) Evolution of the phase noise variance with time.*

Figure 13: Evolution of the phase noise and phase noise variance with time, considering a linewidth of 1 MHz.

Figure 13a) presents the evolution of the LPN with time. Considering a DFB laser with a linewidth of 1 MHz, figure 13a) shows that the LPN, for different noise samples, has independent walks due to the random nature of the model. For the first time sample, each noise sample has different phase noises because this first time sample is uniformly distributed between 0 and 2π, with the following time samples depending on the ones before. Figure 13b) presents the evolution of the phase noise variance with time. To obtain the variance, 1500 noise samples were considered, and it is possible to observe that, in fact, the variance increases linearly with time, according to the theoretical expression mentioned in the beginning of this section. Theoretically, the variance of a uniformly distributed number between 0 and 2π is given by $\frac{\pi^2}{3}$, which is approximately the value of the variance for the first time sample.



a) 100 kHz linewidth. *b) 1 MHz linewidth.*

Figure 14: Spectrum of the optical field at the laser output.

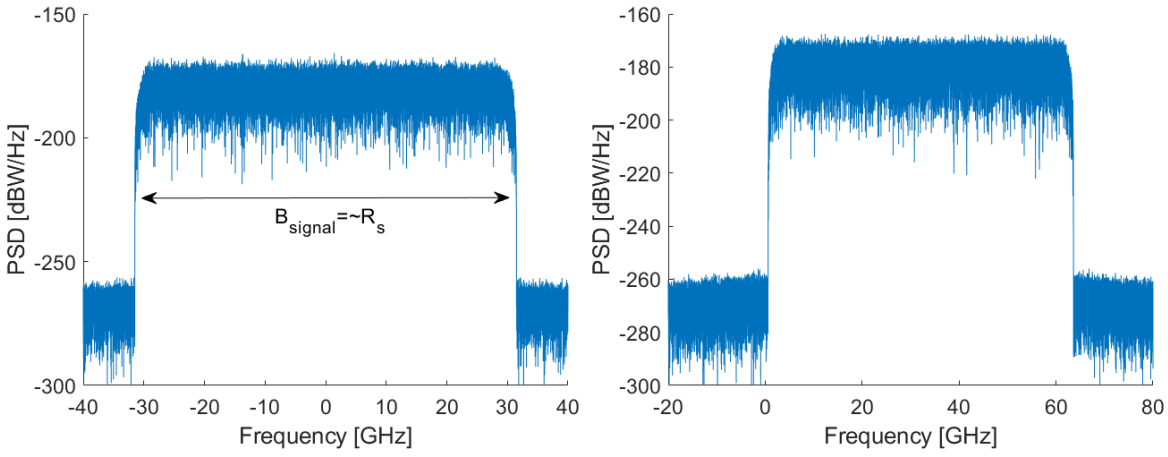
Figure 14 shows the spectrum of the laser fields considering a linewidth of a) 100 kHz and b) 1 MHz, which was averaged over 1500 noise samples. The spectrum follows the theoretical Lorentzian curve [56], and the linewidth, measured at -3 dB, is 100 kHz and 1 MHz as expected.

3.2.4. Formatting the SSB signal

In self-coherent systems such as the one considered in this work, a strong optical tone must be sent alongside the signal, while the information-bearing signal must also be spectrally shifted, thus forming the SSB signal [8]. In this work, the tone is sent on the left edge of the signal. The following equation, represents the optical field at the output of the transmitter:

$$e_{SSB}(t) = \left[A + e_{out}(t) \cdot e^{j2\pi\left(\frac{B_{signal}}{2}+B_g\right)t}\right] e^{j\phi(t)} \quad (7)$$

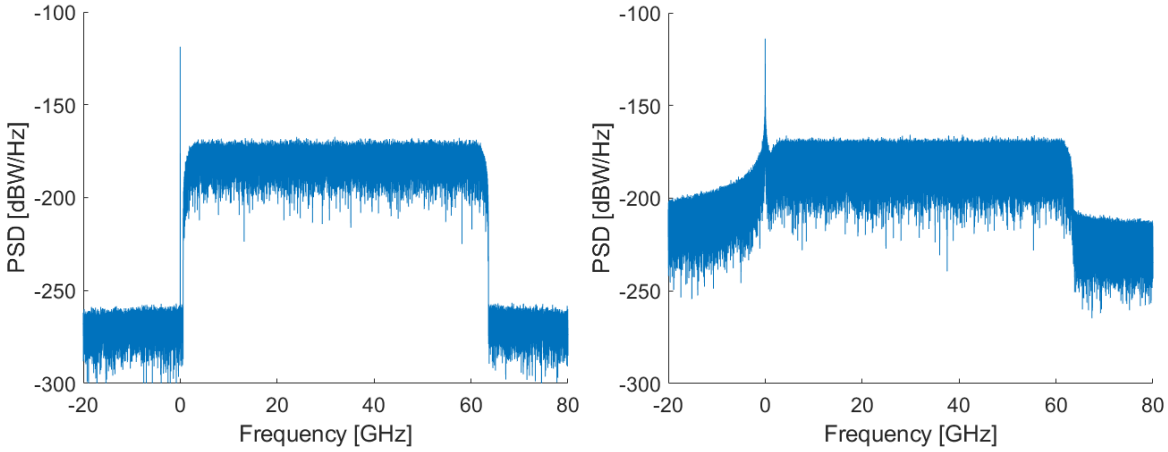
where A is the amplitude of the optical field of the carrier, $e_{out}(t)$ is the information-bearing signal obtained in (2), B_{signal} corresponds to the bandwidth of the information-bearing signal, which should be equal to the symbol rate R_s, B_g is a guard band between the carrier and $e_{out}(t)$, and $\phi(t)$ is the LPN. Considering a R_s of 60 GBaud and a guard band corresponding to 7% of B_{signal}, figures 15 and 16 present PSDs of the signal at various stages of the SSB formatting.



a) Information-bearing signal. *b) Information-bearing signal spectrally shifted.*

Figure 15: PSDs of the information-bearing signal before and after being spectrally shifted.

Figure 15a) presents the PSD of the information-bearing signal $e_{out}(t)$ at the DP-MZM output. The bandwidth of the signal is approximately the same as R_s, as expected. Figure 15b) presents the PSD of the information-bearing signal after being spectrally shifted, i.e., $e_{out}(t) \cdot e^{j2\pi(B_{signal}/2+B_g)t}$ from (7).



a) SSB signal without LPN. *b) SSB signal with LPN.*

Figure 16: PSDs of the SSB signals with and without the LPN, considering a linewidth of 1 MHz.

Figure 16a) presents the SSB signal without LPN, i.e., looking at (7), it represents $A + e_{out}(t) \cdot e^{j2\pi(B_{signal}/2+B_g)t}$. Figure 16b) presents the SSB signal with LPN, considering a linewidth

of 1 MHz, which corresponds to $e_{SSB}(t)$ in (7). The carrier is shown on the left side of the signal. Due to the phase noise, in figure 16b) this carrier is not represented as a single line, instead spreading across the band, in contrast to figure 16a), which represents the ideal case where the laser is monochromatic (single line).

3.3. MCF model

This section explains how the MCF model presented in [42] is implemented in the simulator. Two cores are considered, the interfering and interfered cores, from here on out also referred as core *m* and *n*, respectively. The interfering core's purpose is to transport a signal with a different bitstream from the one in the interfered core, which is going to interfere in the form of ICXT. The interfered core is the one that transports the signal impaired by the ICXT. Furthermore, two polarizations are considered in each core, with the power of the modulated signal being split by the two. Each core has linear propagation, thus being modeled by the following linear SMF transfer function:

$$H_f(\omega) = e^{-j\beta_0 L - j\beta_1 \omega L - j\frac{\beta_2}{2}\omega^2 L - j\frac{\beta_3}{6}\omega^3 L} \cdot e^{-\frac{\alpha}{2}L} \quad (8)$$

where ω is the angular frequency, L is the length of the fiber in meters, α is the fiber attenuation in Np/m and β_0, β_1, β_2 and β_3 are given by the following equations:

$$\beta_0 = \frac{n_{eff} \cdot 2\pi}{\lambda_0} \quad (9)$$

$$\beta_1 = \frac{1}{v_g} \quad (10)$$

$$\beta_2 = -\frac{{\lambda_0}^2 D_{\lambda_0}}{2\pi c} \quad (11)$$

$$\beta_3 = \left(\frac{{\lambda_0}^2}{2\pi c}\right)^2 S_{\lambda_0} + \frac{{\lambda_0}^3 D_{\lambda_0}}{2\pi^2 c^2} \quad (12)$$

where n_{eff} is the effective refractive index of the fiber, λ_0 is the channel wavelength, v_g is the group velocity, D_{λ_0} is the dispersion parameter of the fiber and S_{λ_0} is the slope of the dispersion parameter, both for a certain wavelength λ_0 and, finally, c is the speed of light in vacuum which is 299792458 m/s.

Considering $core \in \{m, n\}$, the optical field at the output of a certain core with polarization rotation is given by [42]:

$$e'_{core,x}(t) = \left[e^{j\theta}\cos(\Gamma)\cdot e_{core,x}(t) - e^{-j\psi}\sin(\Gamma)\cdot e_{core,y}(t)\right] * F^{-1}\left[H_f(\omega)\right] \quad (13)$$

$$e'_{core,y}(t) = \left[e^{j\psi}\sin(\Gamma)\cdot e_{core,x}(t) + e^{-j\theta}\cos(\Gamma)\cdot e_{core,y}(t)\right] * F^{-1}\left[H_f(\omega)\right] \quad (14)$$

where F^{-1} is the inverse Fourier transform, $e_{core,x}(t)$ and $e_{core,y}(t)$ are the SSB signals at the transmitter output obtained in (7), for each core and polarization, and the phases ψ, Γ, θ are random processes assumed constant along each time fraction. This time fraction is a short time window with a much shorter duration than the ICXT decorrelation time, which is in the order of a few minutes [42].

The ICXT is induced by core *m* into core *n* and is uncorrelated between time fractions, which means that, in different time fractions, the ICXT has a different impact on the signal [42]. Considering $p \in \{x, y\}$, the ICXT optical field induced by core *m* into core *n* for each polarization is given by [42]:

$$e_{ICXT,n,p}(t) = F^{-1}\left[e_{m,x}(\omega)\, F_{x,p}(\omega)\right] + F^{-1}\left[e_{m,y}(\omega)\, F_{y,p}(\omega)\right] \quad (15)$$

where $e_{m,x}(\omega)$ and $e_{m,y}(\omega)$ are the SSB signals at the transmitter output obtained in (7), for core *m* and each polarization, ω is the angular frequency, and $F_{p,b}(\omega)$ with $b \in \{x, y\}$ is a transfer function given by:

$$F_{p,b}(\omega) = -j\frac{\overline{K_{nm}}}{\sqrt{2}} H_f(\omega) \sum_{k=1}^{N_{PMP}} e^{-j\,\Delta\bar{\beta}_{mn}(\omega)\, z_k}\, e^{-j\phi_{p,b}^{(k)}} \quad (16)$$

where N_{PMP} is the number of phase matching points (PMP), $\overline{K_{nm}}$ is the discrete coupling coefficient given by:

$$\overline{K_{nm}} = \sqrt{\frac{X_C \cdot P_n}{N_{PMP} \cdot P_m}} \quad (17)$$

where X_C is the ICXT level, which is defined as the ratio between the mean ICXT power and signal power, at the output of the interfered core [49], and P_m and P_n are the average powers at the input of the interfering and interfered cores, respectively.

The parameter $\Delta\bar{\beta}_{mn}(\omega)$ is given by:

$$\Delta\bar{\beta}_{mn}(\omega) = \Delta\bar{\beta}_{0,mn} + d_{mn}\omega - \frac{\Delta D_{mn}\lambda^2\omega^2}{4\pi c} \quad (18)$$

where $\Delta\bar{\beta}_{0,mn}$ is the difference between the averages of the propagation constants, ΔD_{mn} is the difference between the dispersion parameters of the cores, d_{mn} is the walkoff parameter measured between cores *m* and *n* and is obtained from the skew, S_{mn}, as:

$$d_{mn} = \frac{S_{mn}}{L} \quad (19)$$

where *L* is the fiber length in meters.

Furthermore, $\phi_{p,b}^{(k)}$ are random phase shifts (RPS) that model random fluctuations of the ICXT induced by perturbations of the physical structure of the MCF. These RPSs are random variables with uniform distribution between 0 and 2π, with the *k*-th contribution being introduced at the *k*-th random coordinate z_k, which is uniformly distributed between $(k\text{-}1)L/N_{PMP}$ and kL/N_{PMP}, with N_{PMP} corresponding to the number of PMPs. Furthermore, these RPSs are uncorrelated between time fractions, meaning that for each one, a new value has to be calculated.

Finally, considering $p \in \{x, y\}$, the optical field at the output of core *n* for each polarization is given by:

$$e_{out,n,p}(t) = e'_{n,p}(t) + e_{ICXT,n,p}(t) \quad (20)$$

where $e'_{n,p}(t)$ is obtained in (13) and (14) for polarization *x* and *y*, respectively, and $e_{ICXT,n,p}(t)$ is obtained in (15).

The instantaneous power of the ICXT at the output of core *n* is given by:

$$p_{ICXT}(t) = \left|e_{ICXT,n,x}(t)\right|^2 + \left|e_{ICXT,n,y}(t)\right|^2 \quad (21)$$

where $e_{ICXT,n,x}(t)$ and $e_{ICXT,n,y}(t)$ are obtained from (15).

Another important concept is the STAXT, which was briefly introduced in section 2.5.1. This value is the average ICXT power (p_{ICXT}) measured over a short period of time usually around 100 to 200 ms [42]. In this work, it is simply calculated as the mean value of $p_{ICXT}(t)$ over the time window of the simulation, which depends on the number of symbols sent.

3.3.1. ICXT statistic validation

To validate the ICXT statistic of the MCF model referred in [42] and implemented in the simulator, a continuous wave signal was injected in the interfering core, while no signal was launched into the interfered core. This way, it is possible to measure the ICXT power at the output of the interfered core.

Table 1 indicates the system parameters considered in the simulation, some of which, namely, $P_{total,m}$, S_{mn} and N_{PMP}, were retrieved from [41]. An L of 35 km, typical of DC interconnects, is considered.

Table 1: Parameters for the validation of the ICXT statistic.

Parameter	Value
$P_{total,m}$ [dBm]	0
L [km]	35
S_{mn} [ns]	2.4
N_{PMP}	1000
ICXT level [dB]	-15
# Time fractions	100 000

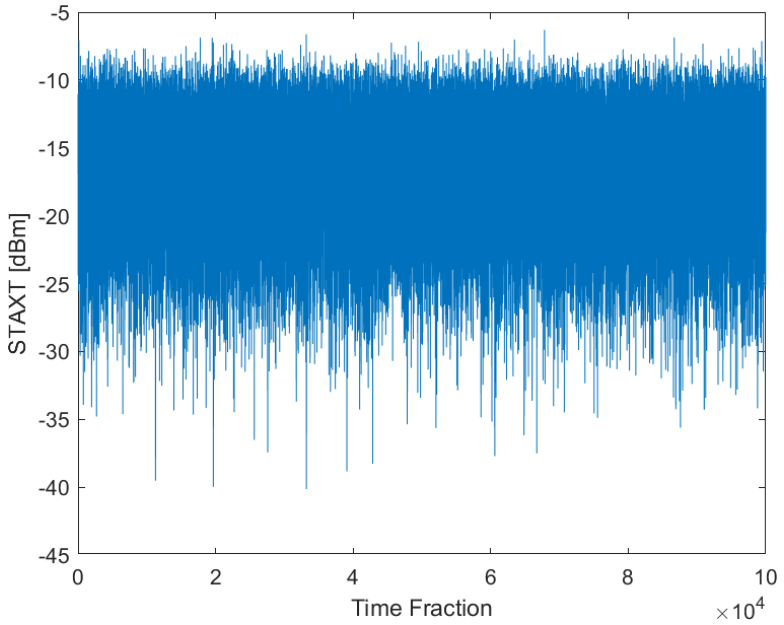


Figure 17: Evolution of the STAXT over 100 000 time fractions.

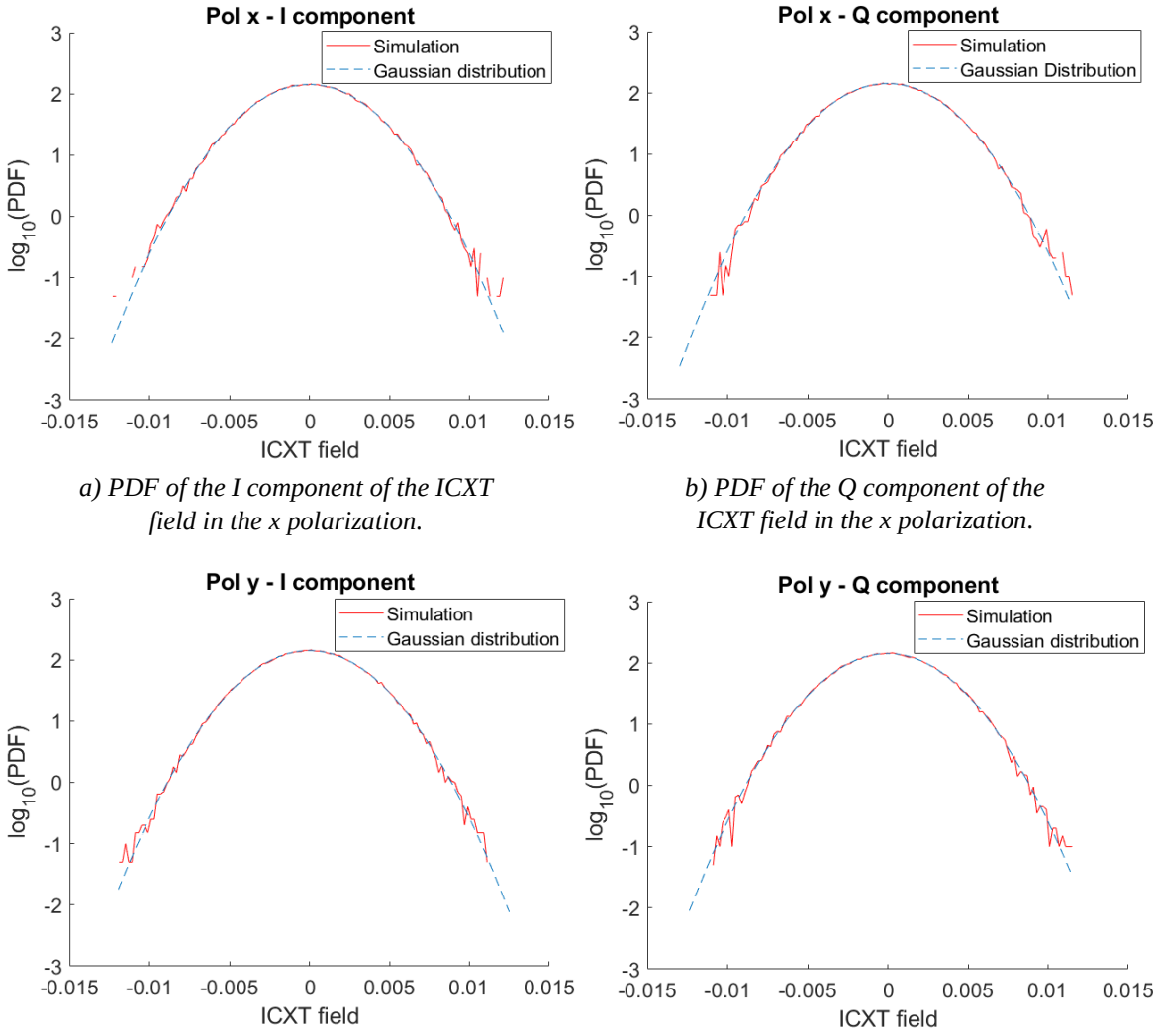
Figure 17 illustrates the evolution of the STAXT over 100 000 time fractions for an ICXT level of -15 dB. It is shown that, in almost every time fraction, the STAXT varies more than 15 dB, sometimes even exceeding 20 dB, as was also reported in [41]. According to [41], the theoretical mean STAXT is obtainable by the following equation:

$$E[STAXT(t)] = N_{PMPs} \sum_{m=1}^{N_{ic}} p_{total,m} |\overline{K_{nm}}|^2 \quad (22)$$

where N_{PMPs} is the number of phase matching points, N_{ic} is the number of interfering cores, $p_{total,m}$ is the total power injected in core m and $|\overline{K_{nm}}|$ is the absolute value of the average discrete coupling

coefficient. To obtain $|\overline{K_{nm}}|$ from (17), the optical power at the input of cores *m* and *n* is considered the same at 0 dBm. The simulation and theoretical mean values of the STAXT were both computed, returning very similar results of about -15 dBm.

In [52] it was concluded that the probability density functions (PDF) of the amplitude of the ICXT field components follow a Gaussian distribution. In order to validate the ICXT model employed in this work, the amplitudes of the ICXT field components were obtained by simulation and overlapped with the Gaussian prediction presented in [52].



a) PDF of the I component of the ICXT field in the x polarization.

b) PDF of the Q component of the ICXT field in the x polarization.

c) PDF of the I component of the ICXT field in the y polarization.

d) PDF of the Q component of the ICXT field in the y polarization.

Figure 18: PDFs of the amplitude of the ICXT field components.

Figure 18 shows the PDFs of the amplitude of the ICXT field components for both polarizations overlapped with a Gaussian distribution, which was obtained from the mean and variance of the simulation results for each field component and polarization. All of the results above follow the Gaussian distribution, as presented in [52], therefore we can assume that the ICXT model is well implemented.

3.4. KK receiver model

In this section, the blocks that comprise the self-coherent KK receiver are detailed. Figure 19 displays a block diagram of the KK receiver employed in this work.

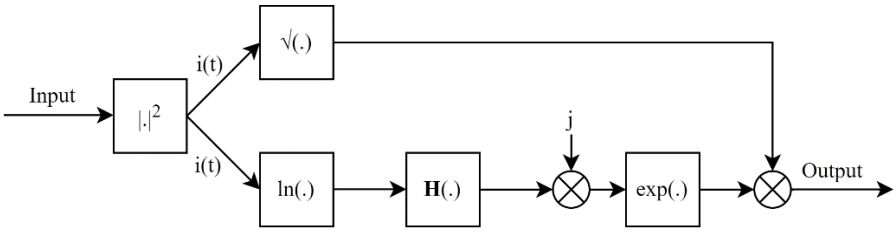


Figure 19: Block diagram of the KK receiver.

3.4.1. PIN photodetector

After propagating in the MCF, the QAM signal arrives at the receiver impaired by the ICXT and the LPN. Considering that the KK receiver employs DD, a single photodiode is required. The positive-intrinsic-negative (PIN) photodetector is responsible for converting the optical power at its input into electrical current in a process called photodetection. The electrical current at the output of the PIN is given by:

$$i(t) = R_\lambda |e(t)|^2 + i_n(t) \quad (23)$$

where R_λ is the PIN responsivity, $e(t)$ is the optical field at the output of core *n* of the MCF, obtained in (20) (with $|e(t)|^2$ representing the instantaneous power of the optical signal at the PIN input) and $i_n(t)$ is the electrical noise, presented in appendix C.

3.4.2. KK field reconstruction

This scheme allows for the reconstruction of the phase of the SSB signal, lost in the photodetection process. However, the minimum phase condition must be met and, for that to happen, the carrier-to-signal power ratio (CSPR) must be larger than the peak-to-average power ratio (PAPR) [8]. A way to ensure this condition is verified is to check if the time trajectory of the SSB signal in the constellation does not include the origin [8].

The CSPR and PAPR are given by the following equations:

$$CSPR = \frac{|A|^2}{\langle |S(t)|^2 \rangle} \quad (24)$$

$$PAPR = \frac{\max\left(|S(t)|^2\right)}{\langle |S(t)|^2 \rangle} \quad (25)$$

where *A* represents the amplitude of the carrier and *S(t)* is the information-bearing signal obtained in (2). Section 3.4.3 demonstrates the importance of the minimum phase condition.

The detected SSB signal after the photodetection, considering $R_\lambda = 1$ A/W, no electrical noise and no LPN, can be written as:

$$i(t) = |A|^2 + A^* \cdot S(t)\, e^{j2\pi(B_{signal}/2+B_g)t} + A \cdot S^*(t)\, e^{-j2\pi(B_{signal}/2+B_g)t} + |S(t)|^2 \quad (26)$$

where $|S(t)|^2$ is the SSBI. The information-bearing signal is $A^* \cdot S(t)\, e^{j2\pi(B_{signal}/2+B_g)t}$, which is in band with the SSBI, thus interfering with each other. If the minimum phase condition is met, this component does not degrade the signal, and the phase and absolute values of the SSB signal are related by the Hilbert transform [8]. Thus, the phase of the signal can be obtained from the intensity of the SSB signal as follows:

$$\varphi(t) = 0.5 \cdot \boldsymbol{H}(\ln(i(t))) \quad (27)$$

where $\boldsymbol{H}$ is the Hilbert transform.

After the reconstruction of the complex-valued signal in (28), the information-bearing signal at the output of the KK receiver (figure 19) still has a carrier and is not in baseband. Figure 20 presents the PSD of the information-bearing signal at the output of the KK receiver (without LPN).

$$S(t)_{modulated\ with\ carrier} = \sqrt{i(t)} \cdot e^{j\varphi(t)} \quad (28)$$

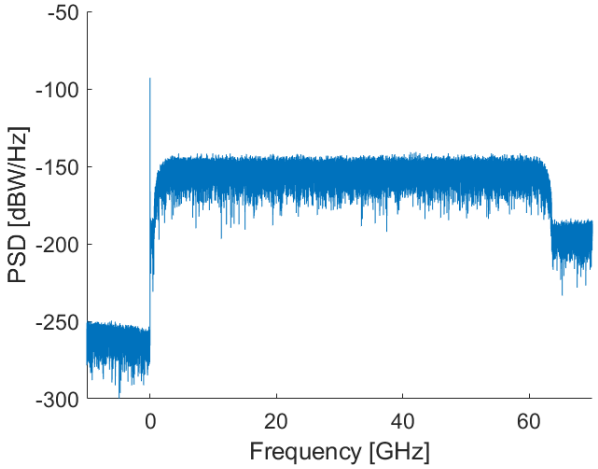
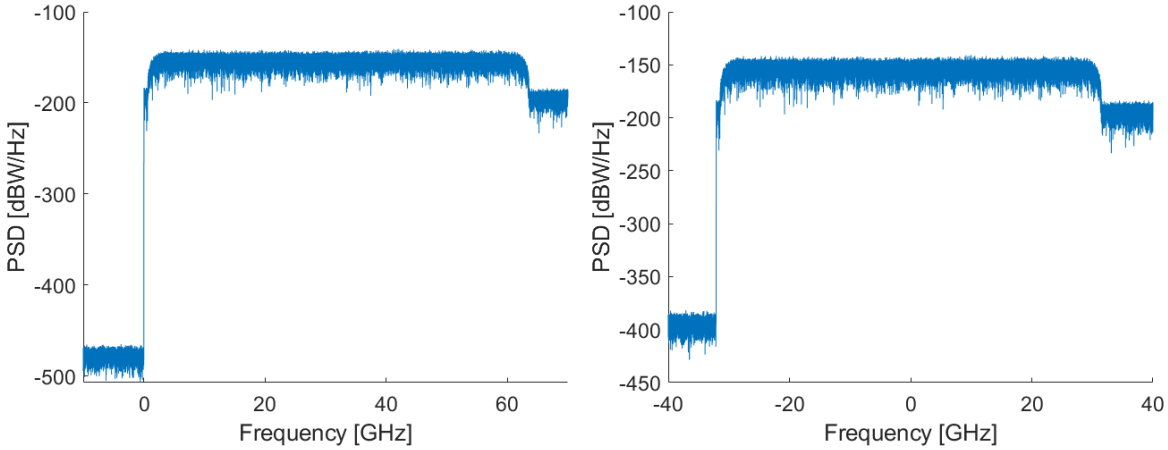


Figure 20: PSD of the information-bearing signal at the output of the KK receiver (without LPN).

Filtering the signal from (28) and figure 20 with a high-pass filter (HPF), removes the carrier. Then, the information-bearing signal is shifted to baseband (figure 21):

$$S(t)_{modulated\ without\ carrier} = F^{-1}[S(f)_{modulated\ with\ carrier} \cdot H_{HPF}(f)] \quad (29)$$

$$S(t) = S(t)_{modulated\ without\ carrier} \cdot e^{-j2\pi(\frac{B_{signal}}{2}+B_g)t} \quad (30)$$



a) Information-bearing signal modulated without carrier.

b) Information-bearing signal in baseband without carrier.

Figure 21: PSD of the information-bearing signal without carrier before and after the shift in frequency.

Once *S(t)* is reconstructed in (30), CD compensation can be performed. Afterwards, the signal passes through an RRC filter, equal to the transmitter, after which the symbols will finally be recovered.

3.4.3. KK receiver validation

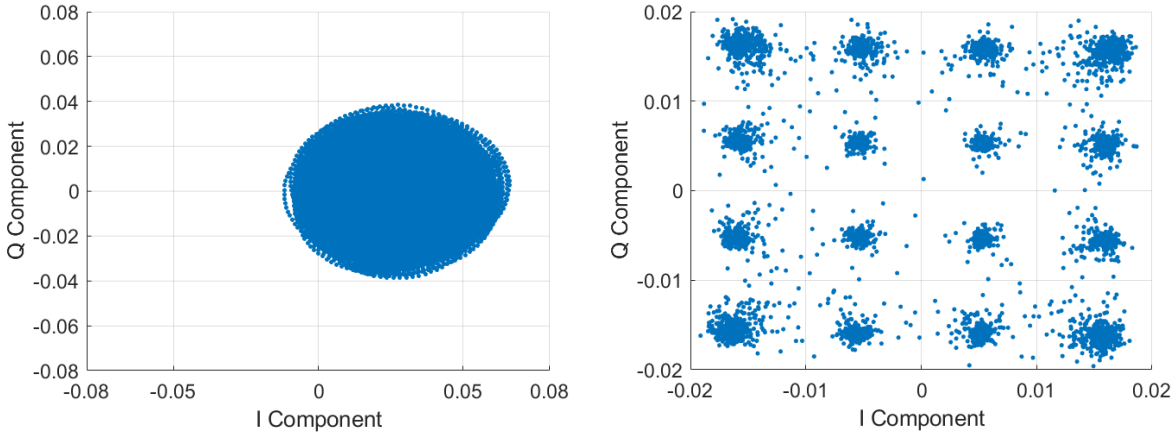
For the validation of the KK receiver, the tests were executed in a B2B operation, i.e., without fiber, thus without ICXT, no LPN is considered, and the electrical noise is considered in the form of gaussian noise, for a given noise equivalent power (NEP). Table 2 summarizes the values considered in the simulation process.

Table 2: Parameters for the validation of the KK receiver.

Parameter	Value
Bit rate [Gb/s]	240
Modulation	SSB 16-QAM
$P_{out\ laser}$ [dBm]	0
PAPR [dB]	∈ [6, 8]
CSPR [dB]	Variable
B_g [% of R_s]	7
NEP [pW/$\sqrt{\text{Hz}}$]	10

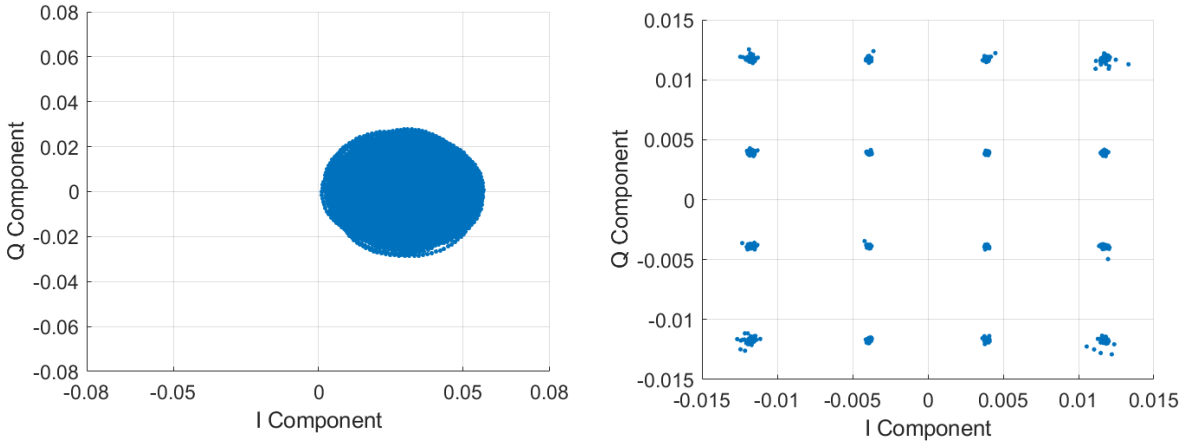
A B_g of 7% of R_s is considered, since it is shown to be within the range of values that achieve optimum performance [54]. Multiple tests were carried out for various CSPRs, in order to show the importance of the fulfillment of the minimum phase condition. Figures 22, 23 and 24 show the time trajectories of the signal and constellations at the KK receiver so as to verify a) if the minimum phase condition is met and b) the distortion of the symbols in the constellation.

In figure 22, a CSPR of 4 dB is considered for the purpose of verifying that, in fact, when the CSPR is lower than the PAPR, the minimum phase condition is not met, and the SSBI degrades the SSB signal. As it is shown in figure 22a), the time trajectory of the signal includes the origin, which means that the condition does not hold. The constellation at the receiver shows that the symbols have a considerable amount of distortion, meaning that they are not well received due to the SSBI, making it undesirable since the LPN and the fiber's associated effects (ICXT, attenuation) are yet to be introduced.



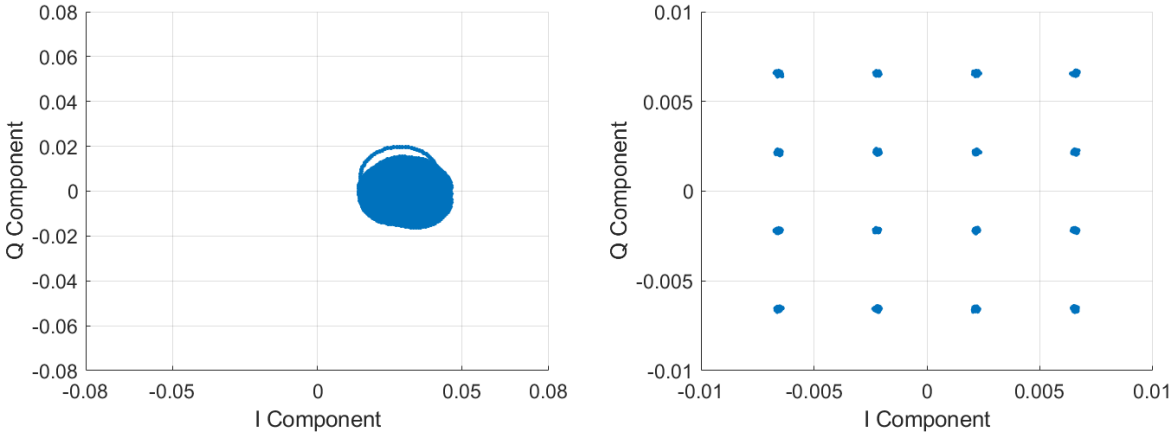
a) Time trajectory of the signal. *b) Constellation at the receiver.*

Figure 22: Time trajectory of the signal and constellation for a CSPR=4 dB (< PAPR).



a) Time trajectory of the signal. *b) Constellation at the receiver.*

Figure 23: Time trajectory of the signal and constellation for a CSPR=PAPR.



a) Time trajectory of the signal. *b) Constellation at the receiver.*

Figure 24: Time trajectory of the signal and constellation for a CSPR=13 dB (> PAPR).

Figures 23 and 24 show the cases where the CSPR is the same as the PAPR, and higher than the PAPR, respectively. Both meet the minimum phase condition, as illustrated in the time trajectories of the signals in figures 23a) and 24a). At the receiver, figure 23b) shows that the symbols are well received but have some distortion due to the SSBI not being completely cancelled. Figure 24b) shows that when the CSPR is, indeed, high enough, the SSBI is completely cancelled, and the symbols are perfectly received.

3.5. Summary

Chapter 3 described the architecture of the 200 Gb/s short-reach self-coherent MCF system employing NNs, namely the simulation models considered for the transmitter, the weakly coupled MCF and KK receiver. In the LPN model, it was shown that different noise samples have independent walks due to the random nature of the model. Furthermore, the spectrums of the laser fields were shown to follow the theoretical Lorentzian curve. For the KK receiver model it was shown that, at the transmitter, the CSPR must be larger than the PAPR of the signal at the output of the DP-MZM, and only when the CSPR was set at 13 dB, did the KK receiver perfectly detect the SSB signal, with the constellations displaying the received symbols without any distortion. When validating the ICXT statistic, the STAXT was shown to sometimes vary more than 20 dB, which is in accordance with the theory. Moreover, it was demonstrated that the PDFs of the amplitude of the ICXT field components follow a Gaussian distribution, as already reported in theory.

CHAPTER 4

Performance of the NN-assisted self-coherent MCF system

In this chapter, the impact of the LPN on the performance improvement provided by the NNs is assessed. In section 4.1, the system without LPN employing a 4 feature FFNN is studied, as reference. Section 4.2 introduces the LPN in the system. Firstly, the optimum number of training samples and neurons of the FFNN is studied for an ECL with a linewidth of 100 kHz, and a DFB laser with a linewidth of 1 MHz. Afterwards, the impact of the LPN on the performance improvement provided by the 4 feature FFNN is evaluated through the mean error vector magnitude (EVM), mean BER and the estimated complementary cumulative distribution function (CCDF) of the BER. The parameters from Tables 3 and 4 were considered in all studies.

Table 3: MCF parameters.

Parameter	Value
$n_{eff,n}$	1.4453
$n_{eff,m}$	1.4455
L [km]	35
ICXT level [dB]	-10
$S_{mn} \times R_s$	0.001

Table 4: Signal parameters.

Parameter	Value
$P_{out,m}$ [dBm]	0
$P_{out,n}$ [dBm]	0
λ [nm]	1550
Modulation	SSB 16-QAM
R_s [GBaud]	60

4.1. System without laser phase noise

In this section, the performance of the 200 Gb/s short-reach self-coherent MCF system employing NNs without LPN is assessed. In this work, a FFNN is studied, which requires some input features in order to learn and produce the desired outcome, which is the mitigation of the ICXT caused by neighboring cores.

4.1.1. Study of the FFNN with 4 input features

In order to reduce the system complexity, firstly a FFNN with only 2 input features – the *I* and *Q* components of the received signal from the interfered core - is considered. This study is detailed in appendix A and shows that 2 input features are definitely not enough for the FFNN to compensate for the ICXT. To do that, the FFNN must also have some information regarding the interfering core, since it is this core that induces the ICXT on the interfered core. It is known from [11] and [54] that a FFNN with 4 input features – the *I* and *Q* components of the received signal from both interfering and interfered

cores - is able to efficiently compensate for this degradation caused by weakly coupled MCFs. However, these works consider ideal lasers in both transmitters when, in fact, real lasers are affected by LPN. As was said before, the objective of this work is to evaluate the impact that these real lasers have on the performance of the NNs but, to do that, the system without LPN must be studied as reference. Figure 25 illustrates the structure of the FFNN to be studied.

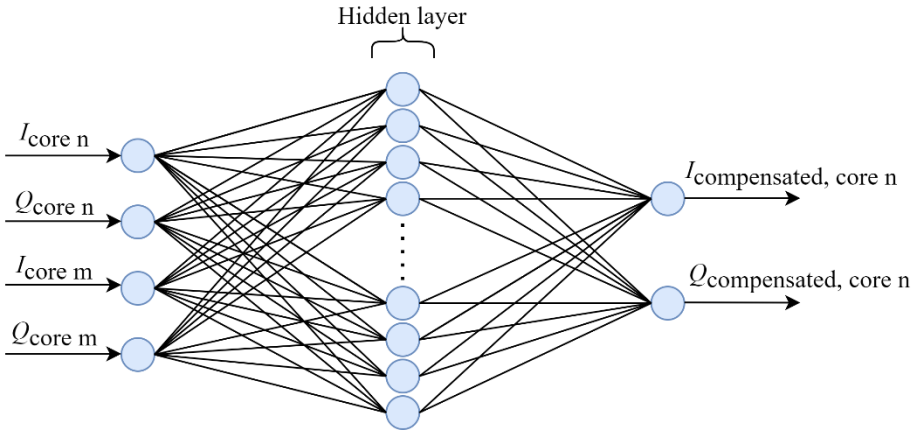


Figure 25: FFNN with 4 input features.

The algorithms used in this FFNN are: the Levenberg-Marquardt for the training phase, the hyperbolic tangent and pure-linear for the activation functions of the hidden layer and output layer, respectively, and the MSE for the loss function. Furthermore, to train the network, the optimum number of neurons and training samples must be obtained. To do this, firstly, a few studies considering the EVM as a function of the number of training samples and number of neurons are conducted considering 4 time fractions, all with an EVM above -14 dB, thus ensuring that the ICXT significantly affects the signal.

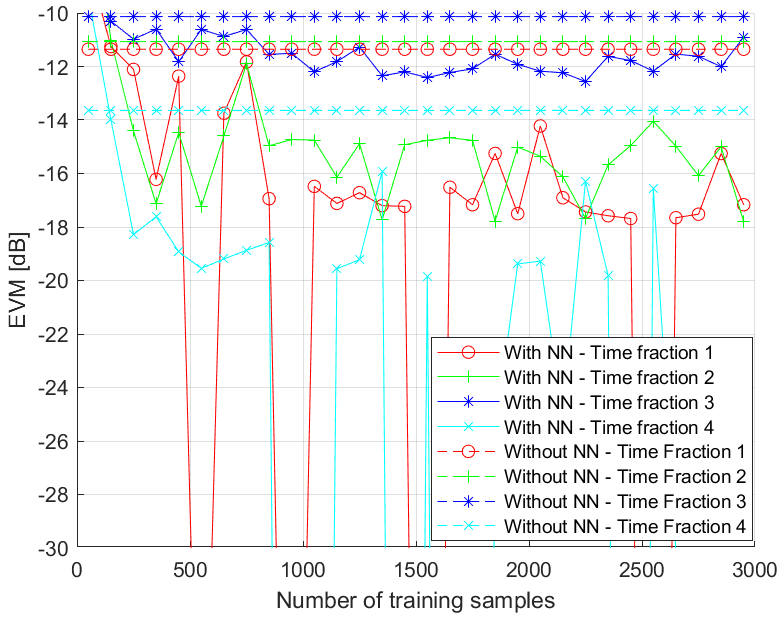


Figure 26: EVM vs #Training samples (10 neurons) – 4 inputs without LPN.

Figure 26 presents the EVM results with and without the FFNN with 4 input features, 10 neurons and a variable number of training samples for 4 time fractions, in order to quantify the performance improvement. Time fractions 1 and 4 show frequent and large fluctuations, with EVMs dropping below -30 dB sometimes, representing at least a 16 dB improvement. Time fraction 2 shows some fluctuations as well, with the EVM dropping to, as low as, -18 dB, representing a 7 dB improvement. The FFNN displays the worse performance for time fraction 3, with a maximum EVM improvement of, approximately, 2 dB. Time fractions 1, 2, 3 and 4, for a number of training samples above 1000, show an improvement of, at least, 2.5 dB, 3 dB, 1 dB and 2 dB, respectively. It is not possible to decide on the optimum number of training samples yet, so, 3 possibilities are going to be further evaluated, namely, 256 (according to the theory, for a skew×symbol rate << 1, there are 16^2=256 combinations of symbols transmitted into the two cores, which should be sufficient for training [54]), 3000 and 20 000 samples, to better understand if increasing the number of samples will result in a better FFNN performance.

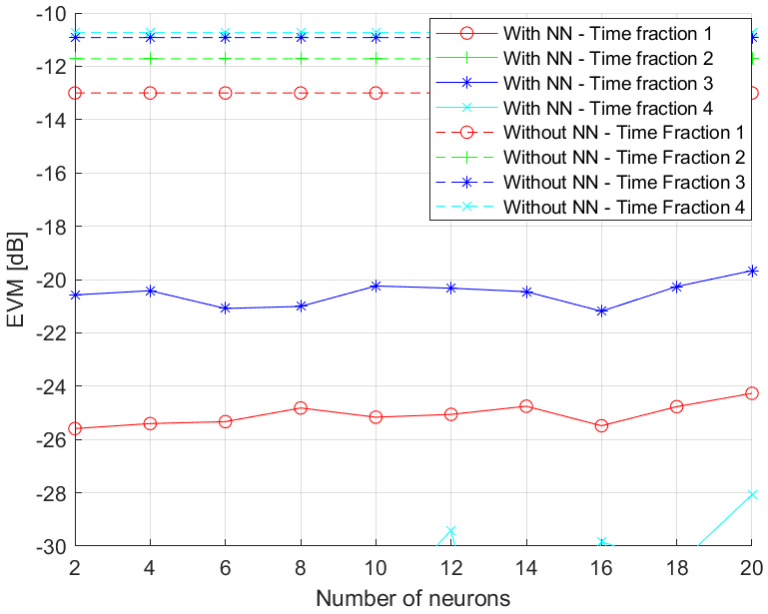
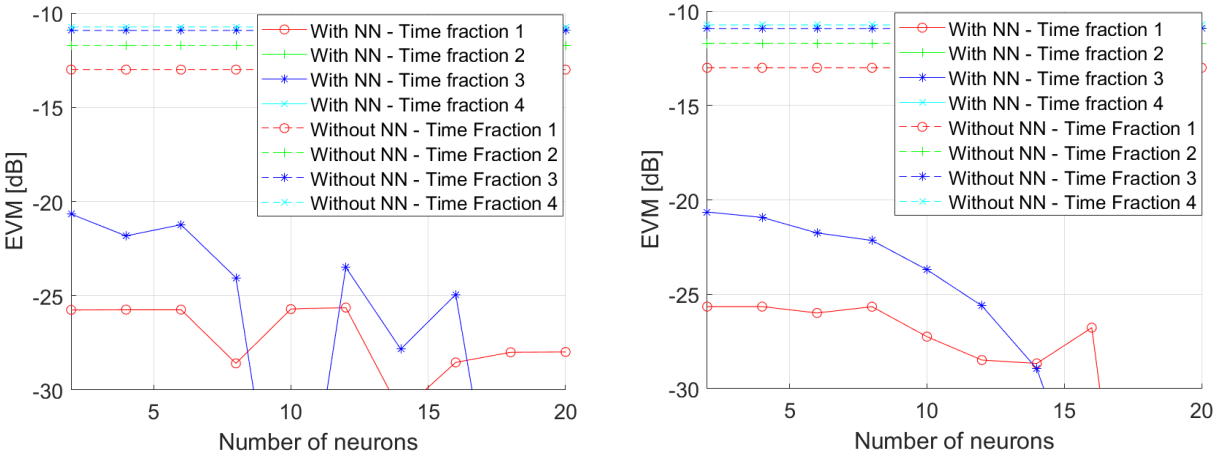


Figure 27: EVM vs #Neurons (256 samples) – 4 inputs without LPN.

Figure 27 presents the EVM as a function of the number of neurons to be employed in the hidden layer, considering only 256 training samples, and 4 time fractions, different than those used in the simulation of figure 26. After employing the FFNN, time fractions 2 and 4 have EVMs below -28 dB, which means that the ICXT is effectively compensated for. Looking at time fractions 1 and 3, it is possible to see that increasing the number of neurons will not actually improve the EVM. Considering only 2 neurons, time fraction 1 shows a performance improvement of at least 12 dB, while time fraction 3 presents an improvement of around 9 dB. The other two have improvements of at least 18 dB, which represents promising results. Thus, a possible configuration is a FFNN with 2 neurons and 256 training

samples. These results suggest that increasing the number of neurons when considering a low number of training samples, does not lead to an effective performance improvement.



a) EVM vs #Neurons – 3000 samples.

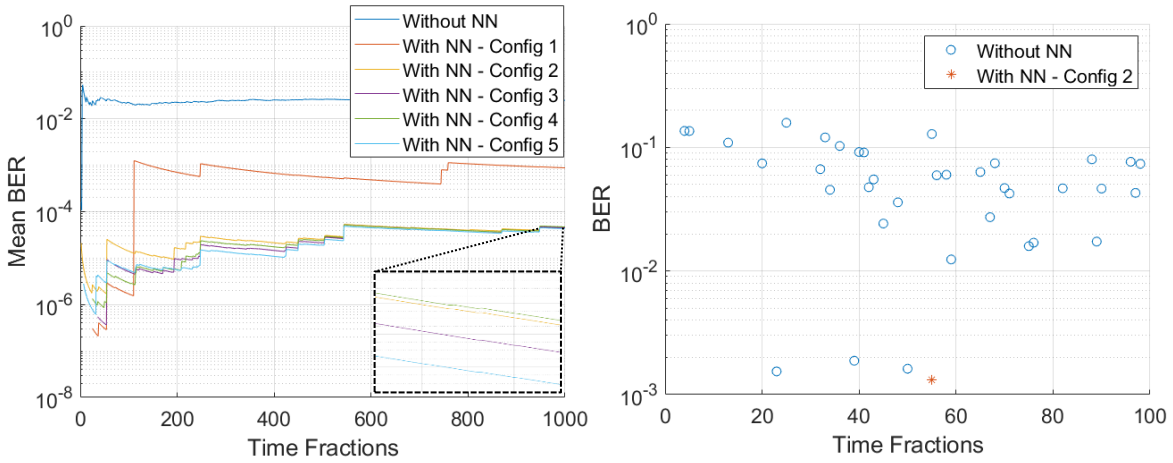
b) EVM vs #Neurons – 20 000 samples.

Figure 28: EVM vs #Neurons (3000 and 20 000 samples) – 4 inputs without LPN.

Figure 28 presents the EVM as a function of the number of neurons, considering a) 3000 samples and b) 20 000 samples, for the same 4 time fractions from figure 27. It can be seen that, considering a higher number of training samples, increasing the number of neurons results in an actual decrease of the EVM, in contrast to figure 27. Looking at figure 28a), in time fraction 1, a number of neurons higher than 14 will result in an EVM below, approximately, -28 dB and, in time fraction 3, disregarding some fluctuations, using a NN with 10 or more neurons can result in an EVM lower than -27 dB, which is at least a 6 dB improvement comparing to the FFNN with 2 neurons and 256 training samples. Furthermore, increasing the number of samples from 3000 to 20 000 (figure 28b)), does not result in a significant improvement of the EVM, with similar values being obtained for 10 and 14 neurons. These studies only considered 4 time fractions. However the ICXT is uncorrelated between time fractions and, for that reason, the performance improvement provided by the FFNN must be analyzed more deeply. Therefore, based on these studies, a detailed analysis of the dependence of the system performance on the time fractions will be accomplished considering 5 configurations, which should make us able to confidently decide on the optimum configuration. Table 5 aggregates the 5 configurations.

Table 5: FFNN configurations – 4 inputs without LPN.

Configuration	Number of training samples	Number of neurons
1	256	2
2	3000	10
3	20 000	10
4	3000	14
5	20 000	14



a) *Mean BER vs 1000 time fractions – 4 inputs without LPN.*

b) *BER vs 100 time fractions – 4 inputs without LPN, configuration 2.*

Figure 29: Evolution of the BER as a function of the time fractions with and without the FFNN – 4 inputs without LPN.

Figure 29a) presents the mean BER of 5 FFNN configurations as a function of the number of time fractions used for the average. It shows that, when employing a NN, the BER stabilizes at around 1000 time fractions. Configuration 1 is clearly the worst one, while the other four return very similar results. From configurations 2, 3, 4 and 5 it is possible to observe that significantly increasing the number of training samples from 3000 to 20 000, or the number of neurons from 10 to 14 will not lead to a significantly better performance. Therefore, since all four configurations have very similar mean BERs, the configuration chosen for the FFNN will be configuration 2 which requires only 3000 training samples and 10 neurons. Figure 29b) presents the BER for 100 time fractions while considering the chosen FFNN configuration (2), in order to show the performance of the FFNN for each time fraction. From these 100 time fractions, several resulted in a BER below 10^{-3}, reason why they are not shown in the figure. Only on one occasion, the BER while employing the chosen FFNN was above 10^{-3} which shows that this configuration with 10 neurons on the hidden layer and 3000 training samples is able to effectively compensate for the ICXT.

4.1.2. Performance assessment

In this subsection, the performance of the 200 Gb/s short-reach self-coherent MCF system employing NNs is going to further evaluated for several ICXT levels. A Monte Carlo simulation was used to calculate the BER and, for each ICXT level, 1000 time fractions were considered to calculate the mean BER. A sequence of 2^{19} bits was considered in each time fraction to present BERs in the order of 10^{-4} with a reasonable confidence interval. Considering 20% FEC, the line bitrate is 240 Gb/s and net bitrate is 200 Gb/s, while the threshold for the system outage is set at BER=$10^{-1.8}$.

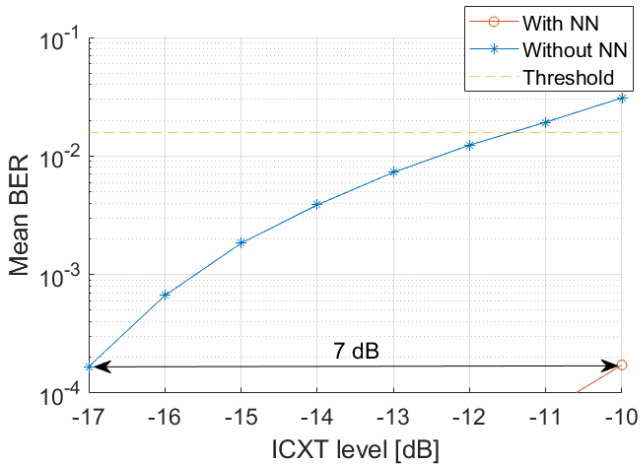
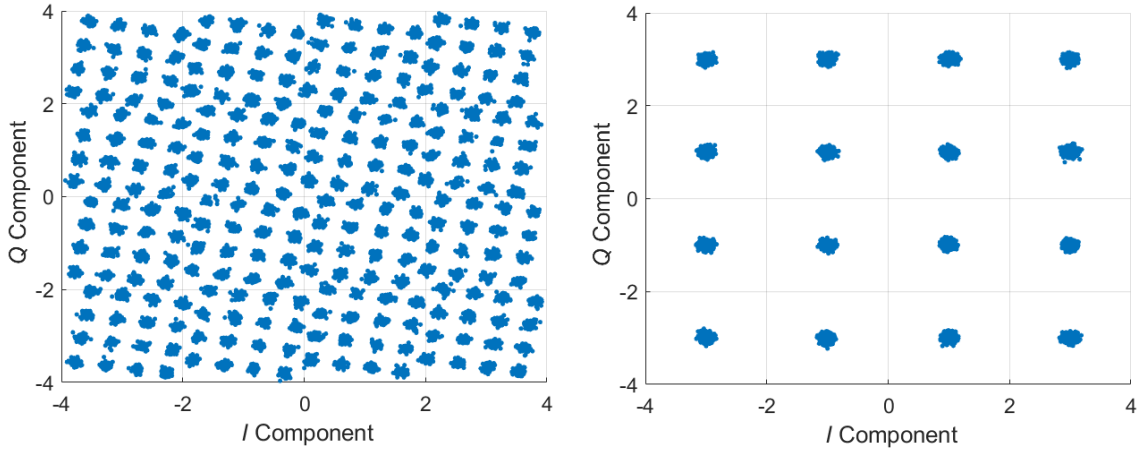


Figure 30: Mean BER vs ICXT level without LPN – 4 input features.

Figure 30 presents the mean BER as function of the ICXT level. The BER was averaged over 1000 time fractions in order to get a stabilized BER estimate. These results indicate that the FFNNs provide a 7 dB tolerance, i.e., the system employing the FFNN can achieve the same mean BER with an ICXT level 7 dB higher than the system without the FFNN. Considering an ICXT level of -10 dB, the mean BER of the system employing the FFNN is, approximately, 2×10^{-4}, which is well below the defined threshold ($10^{-1.8}$) and represents a significant performance improvement compared to the system without the ML algorithm.



a) Constellation before using the FFNN. *b) Constellation after using the FFNN.*

Figure 31: Constellations before and after using the FFNN.

Figure 31a) and b) show the constellations of the received signal before and after using the FFNN, respectively. As seen in figure 31a), each 16-QAM symbol is affected by 16 different symbols due to the ICXT, resulting in high BERs and EVMs. Figure 31b) perfectly illustrates the improvement provided by the FFNN on this system, revealing a constellation with little distortion. All the results in this section are in accordance with what was reported in [11] and [54].

4.2. System with laser phase noise

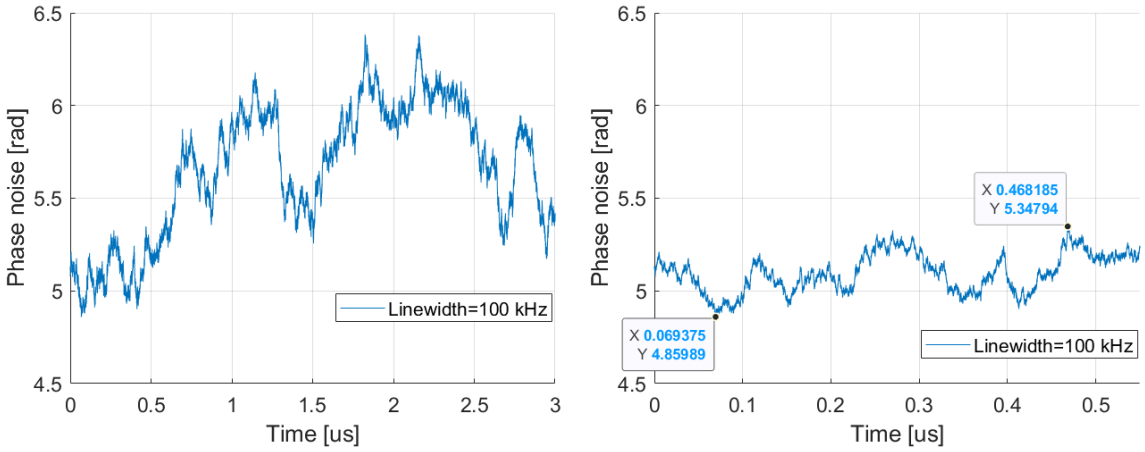
In this section, the laser of the interfering core is affected by LPN characterized by two linewidths: 100 kHz and 1 MHz, typical of ECLs and DFB lasers, respectively. An ideal laser, with no phase noise, is considered to generate the optical signal to be injected in the interfered core.

4.2.1. Study of the FFNN with 4 input features – ECL with 100 kHz linewidth

The LPN is characterized by the coherence time of the laser field, which is a time interval where the phases of the laser field have some correlation [50]. The coherence time of this ECL is 3.18 μs as is shown in Table 6. Figure 32a), which presents the phase noise variation as a function of time, shows us that the phase noise for different time instants during the coherence time can have significant variations. However, to guarantee an adequate training of the FFNN, it must be employed in a time window where the phase noise variation is limited, since the FFNN requires symbols from the interfering core, which are affected by the LPN. That way, the FFNN is trained and subsequently employed with symbols whose phase noise is highly correlated, thus being able to improve the performance of the system.

Table 6: Coherence time and time window – 100 kHz of linewidth.

T_c - Coherence time [μs]	~3.18
Number of symbols	2^{15}
Time window [μs]	$2^{15}\times 1/R_s$=~0.55



a) Phase noise variation as a function of time along the coherence time.

b) Phase noise variation as a function of time along the time window of the simulation.

Figure 32: Phase noise variation as a function of time for a 100 kHz linewidth.

Figure 32 presents the phase noise variation along a) the coherence time and b) the time window of the simulation considered for a linewidth of 100 kHz. As it can be seen from figure 32b), when considering a time window 5 times smaller than the coherence time, the phase noise of this sample varies around the same values, with a maximum variation of 27.5 degrees, which is much smaller than the maximum variation if we considered a time window the size of the coherence time (figure 32a)). For that reason, the FFNN is studied for a time window considerably smaller than the coherence time.

With the LPN introduced in the system, a new study of the optimum number of training samples and neurons is conducted considering the number of symbols and time window presented in Table 6.

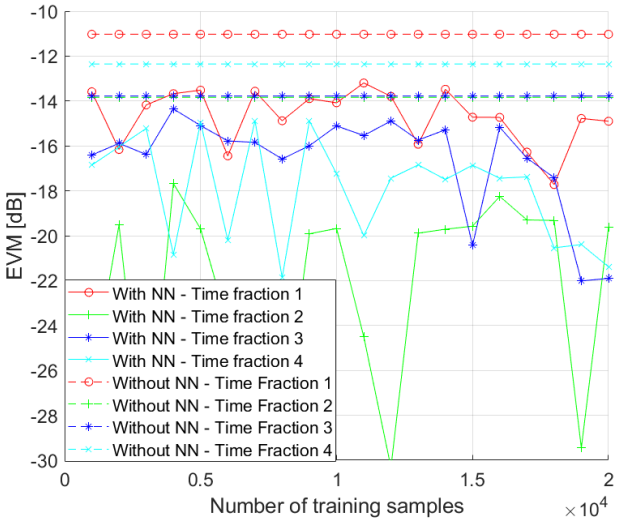
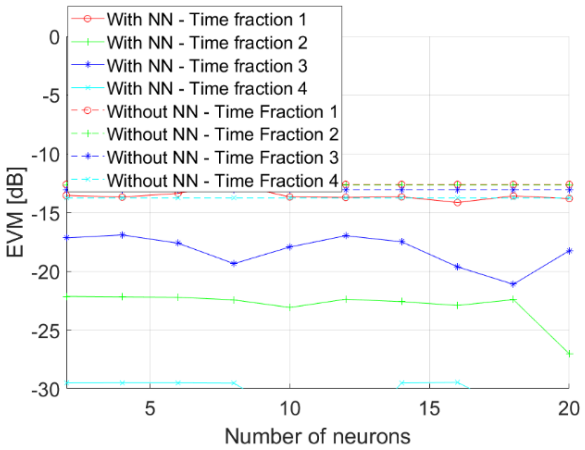
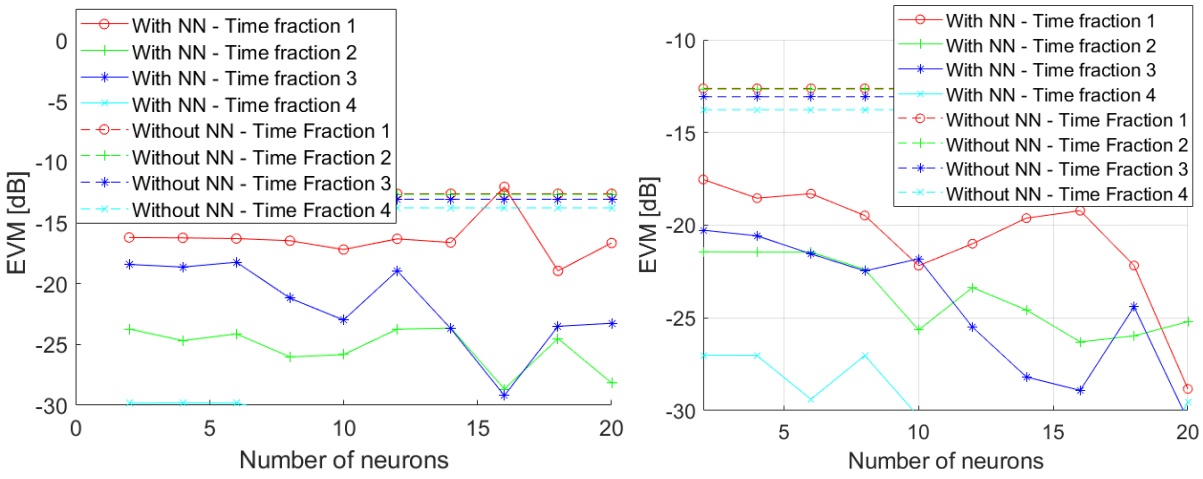


Figure 33: EVM vs #training samples (10 neurons) - 100 kHz linewidth.

Figure 33 presents the EVM as a function of the number of training samples to be employed in the FFNN, considering 10 neurons on the hidden layer. The EVM of time fractions 1 and 3 seem to improve and to be roughly stable from 1000 to 13 000 training samples at around -14 dB (~3 dB improvement) and -15 dB (~1 dB improvement), respectively. Time fractions 2 and 4 show large fluctuations having a minimum improvement of 4 dB and approximately 2.5 dB, respectively, and a maximum improvement of at least 16 dB and 9 dB, respectively. This study from figure 33 did not provide any definite possibilities so, the dependence of the system performance on the number of neurons is optimized considering 1000, 6000 and 20 000 samples.



a) EVM vs #neurons (1000 samples) – 100 kHz linewidth.



b) EVM vs #neurons (6000 samples) – 100 kHz linewidth.

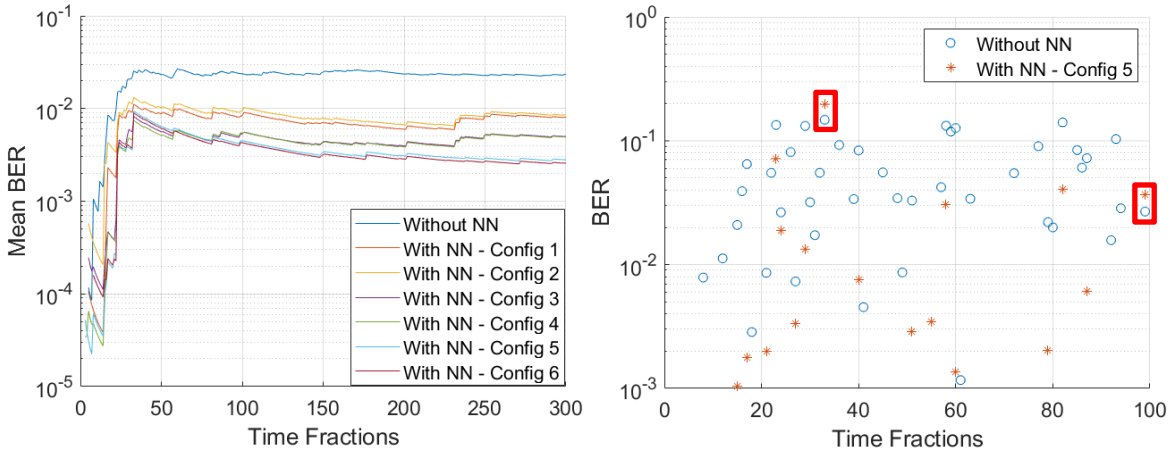
c) EVM vs #neurons (20 000 samples) – 100 kHz linewidth.

Figure 34: EVM vs #neurons with 1000, 6000 and 20 000 training samples – 100 kHz linewidth.

Figure 34 presents the EVM as a function of the number of neurons considering a) 1000 b) 6000 and c) 20 000 training samples, for 4 time fractions, different than those used in the simulation of figure 33. Figure 34a) exhibits the same sort of behavior as observed before in figure 27, in which the EVM stays roughly stable with the increase of number of neurons for all 4 time fractions. In this case the number of training samples is not as low as before (in figure 27 only 256 training samples were employed), and the EVM of time fraction 3 varies around 2-3 dB, but these results also suggest that when the number of training samples is low, increasing the number of neurons does not result in a significantly better performance. Figures 34b) and c) show that time fractions 2 and 3 present the best performance improvement when considering 16 neurons. Time fractions 1, 2 and 4 also have notable improvements when employing only 10 neurons. Hence, a detailed analysis of the dependence of the system performance on the time fractions will be accomplished considering 10 and 16 neurons. Table 7 displays all 6 configurations to be considered.

Table 7: FFNN configurations - 4 inputs with LPN (100 kHz).

Configuration	Number of training samples	Number of neurons
1	1000	10
2	1000	16
3	6000	10
4	6000	16
5	20 000	10
6	20 000	16



a) *Mean BER vs 300 time fractions – 4 inputs with LPN (100 kHz).*

b) *BER vs 100 time fractions – 4 inputs with LPN (100 kHz).*

Figure 35: Evolution of the BER as a function of the time fractions with and without the FFNN – 4 inputs with LPN (100 kHz).

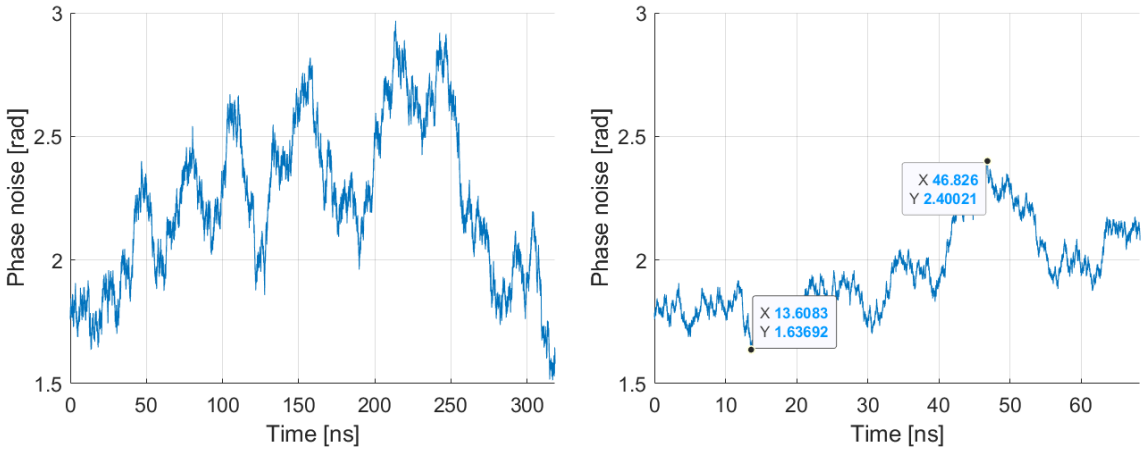
Figure 35a) presents the evolution of the mean BER for the 6 configurations over 300 time fractions. The best configuration is number 6, however, configuration number 5 with 20 000 samples and 10 neurons is the chosen one for the FFNN, since it requires less neurons and achieves approximately the same performance as configuration 6. Figure 35b) presents the BER over 100 time fractions, for configuration 5, and it can be observed that the FFNN is always able to provide some sort of improvement, except for time fractions 33 and 99 (highlighted in red), in which it worsens the performance. These results are a lot different than those observed in figure 29b), illustrating the impact that the LPN has on the performance of the FFNN. It is important to note that, from these 100 time fractions, several resulted in a BER below 10^{-3}, hence why they are not shown.

4.2.2. Study of the FFNN with 4 input features – DFB laser with 1 MHz linewidth

For the same reasons mentioned in section 4.2.1, a time window 5 times smaller than the coherence time of the laser was, once again, considered. Table 8 presents the coherence time of the DFB laser, number of symbols and time window of the simulation.

Table 8: FFNN configurations - 4 inputs with LPN (1 MHz).

T_c - Coherence time [ns]	~318.31
Number of symbols	2^{12}
Time window [ns]	$2^{12}\times 1/R_s$=~68.27



a) Phase noise variation in function of time along the coherence time.

b) Phase noise variation in function of time along the time window of the simulation.

Figure 36: Phase noise variation as a function of time for a 1 MHz linewidth.

Figure 36 presents the phase noise variation along a) the coherence time and b) the time window of the simulation considered for a linewidth of 1 MHz. Figure 36b) shows that the phase noise of this sample varies around the same values until 40 ns. Considering the full time window, the phase noise varies a maximum of, approximately, 44 degrees. Although this is a significant variation, it is still smaller than the maximum variation considering a time window the size of the coherence time, as was already concluded in section 4.2.1. Since we are considering a different laser with a different linewidth and a different time window of simulation, a new study for the optimum number of training samples and neurons is conducted considering the number of symbols and time window presented in Table 8.

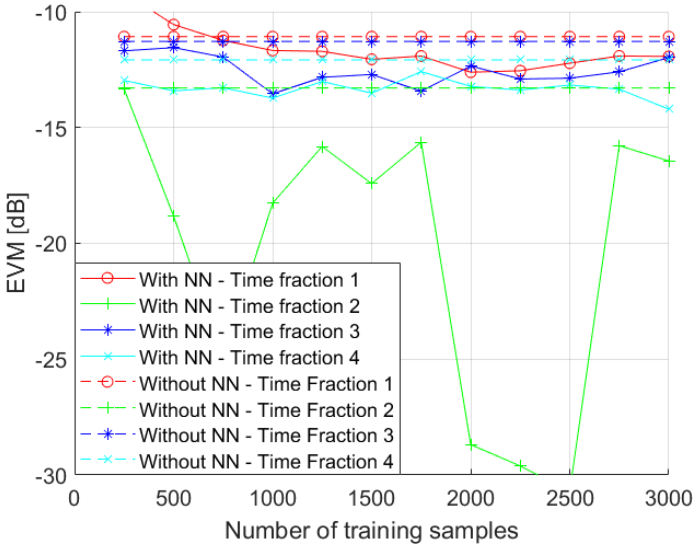
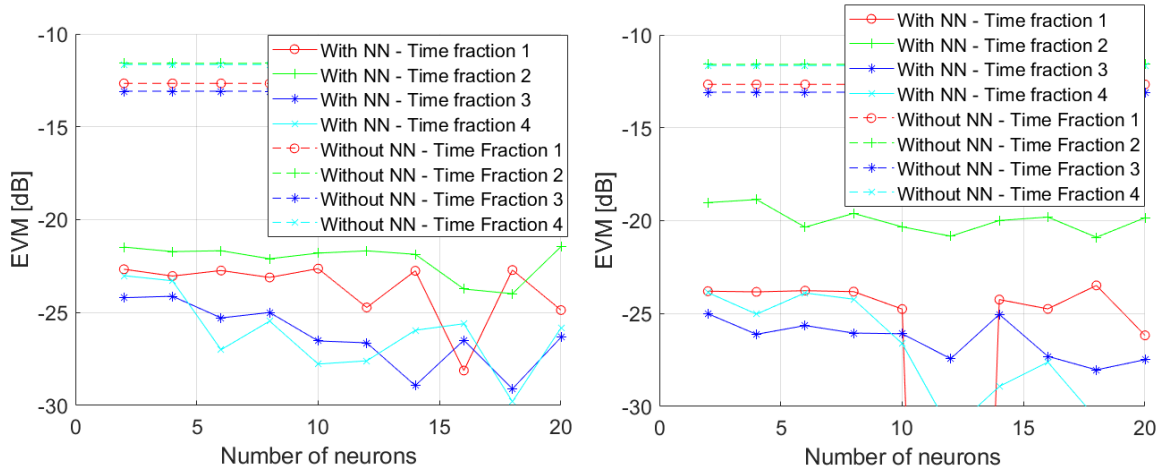


Figure 37: EVM vs #training samples (10 neurons) - 1 MHz linewidth.

Figure 37 presents the EVM as a function of the number of training samples for a linewidth of 1 MHz, considering 10 neurons. For time fraction 1, the EVM only seems to stabilize at around 1500 training samples, showing a maximum improvement of 1.6 dB. Time fraction 2 displays some large fluctuations and a maximum improvement of more than 6.7 dB. The EVM seems to be stable at around 1000 training samples for time fractions 3 and 4, showing an improvement of 2.2 dB and 1.7 dB, respectively. For these reasons, the dependence of the system performance on the number of neurons is optimized considering 1000 and 2000 training samples.



a) EVM vs #neurons (1000 samples) – 1 MHz linewidth.

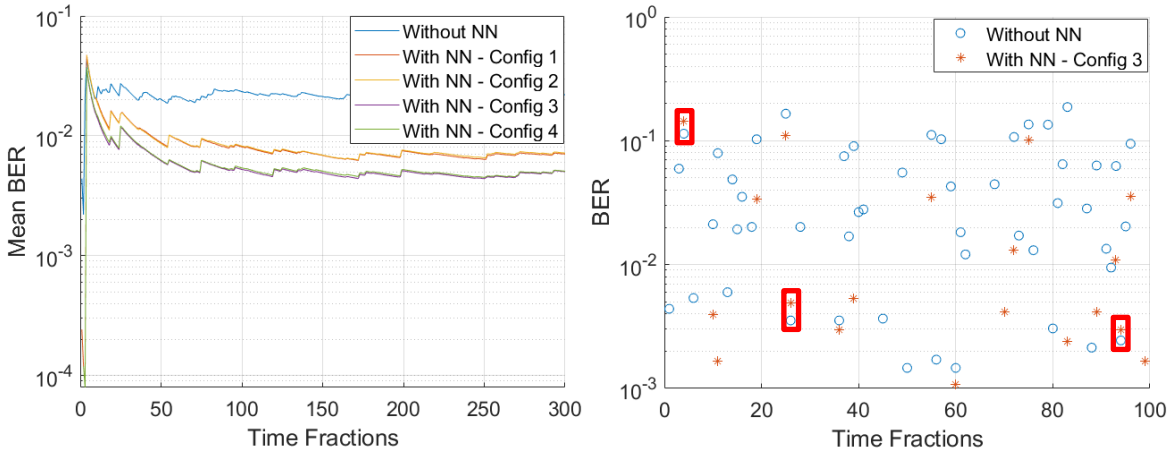
b) EVM vs #neurons (2000 samples) – 1 MHz linewidth.

Figure 38: EVM vs #neurons considering 1000 and 2000 training samples - 1 MHz linewidth.

Figure 38 presents the EVM as a function of the number of neurons considering a) 1000 and b) 2000 training samples, for 4 time fractions, different than those used in the simulation of figure 37. Looking at time fraction 1 in figure 38a), 16 neurons seem to achieve the best performance. Time fractions 2, 3 and 4 indicate that 18 neurons provide the best performance. Figure 38b) seems to suggest the same, although, for time fraction 1, when employing the FFNN, the EVM drops below -30 dB for 12 and 14 neurons. As was seen before in figures 27 and 34a), increasing the number of neurons with a low number of training samples does not necessarily lead to notable performance improvement. On that account, a detailed analysis of the dependence of the system performance on the time fractions will be carried out considering a FFNN with 10 and 16 neurons. Table 9 indicates all 4 configurations to be studied.

Table 9: FFNN configurations - 4 inputs with LPN (1 MHz).

Configuration	Number of training samples	Number of neurons
1	1000	10
2	1000	16
3	2000	10
4	2000	16



a) *Mean BER vs 300 time fractions – 4 inputs with LPN (1 MHz).*

b) *BER vs 100 time fractions – 4 inputs with LPN (1 MHz).*

Figure 39: Evolution of the BER as a function of the time fractions with and without the FFNN – 4 inputs with LPN (1 MHz).

Figure 39a) shows the evolution of the mean BER over 300 time fractions. The mean BER reveals that either configuration 3 or 4 is the best, since the curves are overlapped. Considering that configuration 3 requires less neurons, it will be the chosen one for the FFNN, and it is confirmed that increasing the number of neurons with a low number of training samples, does not improve the performance. Figure 39b) presents the BER over 100 time fractions for configuration 3, and it can be seen that the FFNN is always able to somewhat improve the performance, except for time fractions 4, 26 and 94 (highlighted in red), which reveal that the FFNN can sometimes also worsen the BER. It is important to note that, from these 100 time fractions, several resulted in a BER below 10^{-3}, hence why they are not shown.

4.2.3. Performance assessment

In this subsection, several studies considering the mean BER and mean EVM as a function of time, and the estimated CCDF of the BER as a function of BER are conducted. The main goals of these studies are to i) quantify the impact of the LPN on the performance improvement provided by the FFNN through the mean BER, mean EVM and estimated CCDF of the BER, and ii) understand how long the FFNN is able to mitigate the combined effect of the ICXT and LPN without having to be trained again. Firstly, the FFNN is trained with $N_{training\ samples}$ during a time interval $T_{training}$. Then, the EVM and BER are evaluated for a fixed number of symbols $N_{BER/EVM}$ and time interval $T_{BER/EVM}$, while varying the time interval ΔT, which represents the time interval between the training phase and the symbol sequence in which the FFNN is employed. Figure 40 represents an illustrative diagram of the time basis of the simulations.

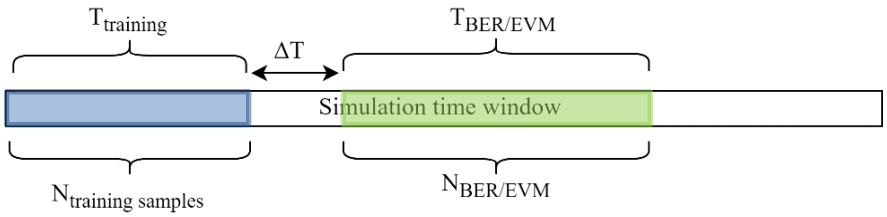


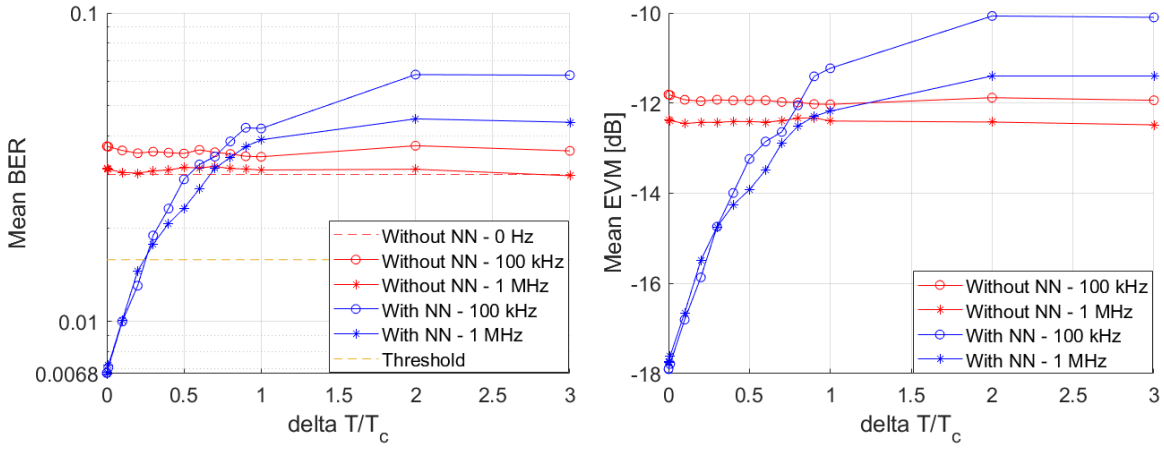
Figure 40: Diagram of the time basis of the simulations.

In the simulation results, ΔT is normalized by the coherence time T_c. When $\Delta T/T_c << 1$, the phase noise considered in the FFNN training phase (highlighted in blue in figure 40) is correlated with the phase noise associated to the symbol sequence in which the FFNN is employed (highlighted in green in figure 40). When $\Delta T/T_c >> 1$ the phase noise considered in these two stages becomes highly uncorrelated. Consequently, we will be able to assess the impact that this uncorrelation has on the performance of the FFNN.

The number of training samples was already obtained from sections 4.2.1 and 4.2.2. The number of symbols considered for the BER and EVM is obtained from the time window considered, which is 5 times smaller than the coherence time of the respective laser. The time interval ΔT is proportional to the coherence time. Table 10 presents the values from figure 40 used in the simulation.

Table 10: Parameters of the simulation.

Laser - Linewidth	**ECL - 100 kHz**	**DFB - 1 MHz**
$N_{training\ samples}$	20 000	2000
$T_{training}$	~333.33 ns	~33.33 ns
T_c	~3.18 μs	~318.31 ns
$\Delta T/T_c$	[0, 3]	[0, 3]
$T_{BER/EVM}$	~0.55 μs	~68.27 ns
$N_{BER/EVM}$	2^{15}	2^{12}
ICXT level	-10 dB	-10 dB



a) Mean BER vs $\Delta T/T_c$. *a) Mean EVM vs $\Delta T/T_c$.*

Figure 41: Evolution of the mean BER and EVM as a function of ΔT normalized by the coherence time T_c, for a linewidth of 0, 100 kHz and 1 MHz.

Figure 41 presents the evolution of the mean BER and mean EVM over the time interval ΔT normalized by the coherence time for a linewidth of 0, 100 kHz and 1 MHz. It is known from [50] and [51] that, when the linewidth×skew << 1, the LPN does not impact the ICXT power. In this work, a symbol rate of 60GBaud and a skew×symbol rate << 1 are considered so, the linewidth×skew << 1. Figure 41a) shows that the mean BER for all the lasers without the FFNN is approximately the same between 3×10^{-2} and 4×10^{-2}. It was expected that it would be closer for the ECL and DFB laser, however, the mean BER of the systems employing the ECL and DFB laser was averaged over only 200 time fractions due to time constraints and is not completely stable (appendix B), while the mean BER when employing the ideal laser was averaged over 1000 time fractions (figure 30).

Looking at figure 41a), for a ΔT of 0, it is possible to say that, while employing the FFNN, the mean BER rises more than one order of magnitude from 2×10^{-4} when employing an ideal laser (as seen in figure 30) to, approximately, 6.8×10^{-3} when employing an ECL or DFB laser. The performance of the FFNN is approximately the same for the ECL and DFB laser employed, up to $\Delta T/T_c=1$. After, the FFNN starts to perform notably worse when an ECL with a 100 kHz linewidth is employed. Nevertheless, we are able to conclude that, for the ECL and DFB laser, the mean BER after the FFNN is kept below the BER threshold when the time interval between the training phase and the use of the FFNN does not exceed 20% of the coherence time (ΔT_{max}). Therefore, the FFNN will have to be trained every $T_{training} + \Delta T_{max,100\ kHz} + T_{BER/EVM} = 0.333 + 0.2\times T_{c,100\ kHz} + 0.55 = 1.52$ µs, when employing an ECL with a 100 kHz linewidth, and $T_{training} + \Delta T_{max,1\ MHz} + T_{BER/EVM} = 33.33 + 0.2\times T_{c,1\ MHz} + 68.27 = 165.26$ ns, when employing a DFB with a 1 MHz linewidth. Furthermore, from figure 41b), it is possible to see that, with the EVM, the FFNN can provide some performance improvement with ΔT up to 80% of the coherence time, for both lasers. Although the performance of the FFNN is similar for both lasers, the ECL provides

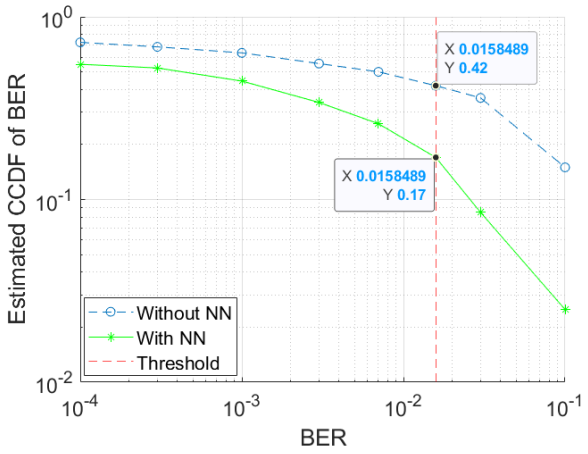
the advantage of the FFNN having to be trained less frequently, due to the longer coherence time when compared to the DFB laser.

In order to quantify the improvement in the EVM, the EVM gain is calculated between the cases when the FFNN is not employed, and when the FFNN is employed. Table 11 presents those results for both the ECL and DFB laser.

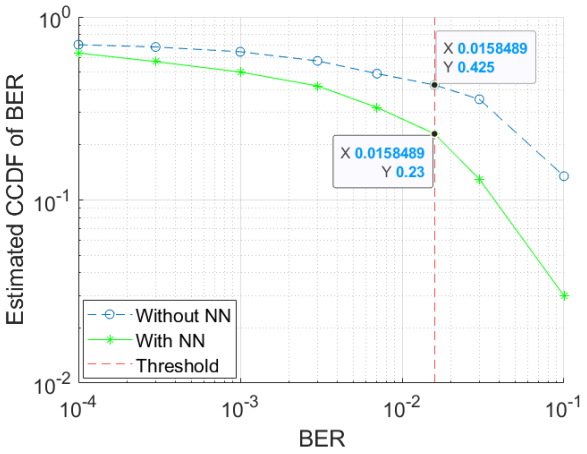
Table 11: EVM gain for different ΔTs considering the ECL and DFB laser.

ECL		**DFB laser**	
$\Delta T/T_c$	**EVM gain [dB]**	$\Delta T/T_c$	**EVM gain [dB]**
0.001	6.1	0.001	5.3
0.01	6	0.01	5.2
0.1	4.9	0.1	4.2
0.5	1.3	0.5	1.5
1	-0.8	1	-0.2

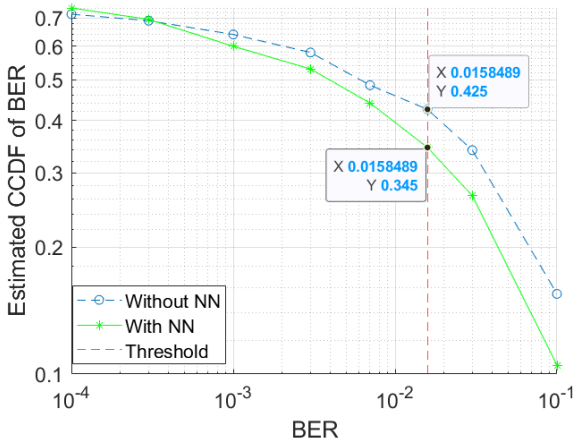
Looking at Table 11 we are able to conclude that, for both lasers, when $\Delta T/T_c << 1$, the FFNN performs the best since the phase noise of the training phase and the phase noise associated to the symbol sequence in which the FFNN is employed is highly correlated, thus the FFNN can properly generalize and produce better results. Once the time interval between the training phase and the use of the FFNN is of the same order as the coherence time - $\Delta T/T_c=1$, the FFNN will no longer provide any performance improvement, due to the uncorrelation between the phase noises considered in both stages of the simulation.



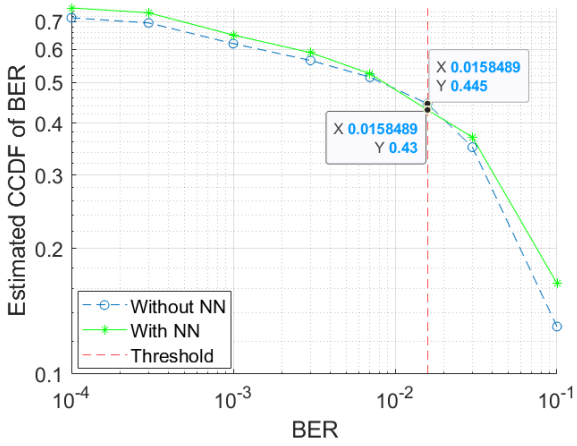
a) Estimated CCDF of the BER for $\Delta T/T_c=0.1$ – 100 kHz.



b) Estimated CCDF of the BER for $\Delta T/T_c=0.2$ – 100 kHz.



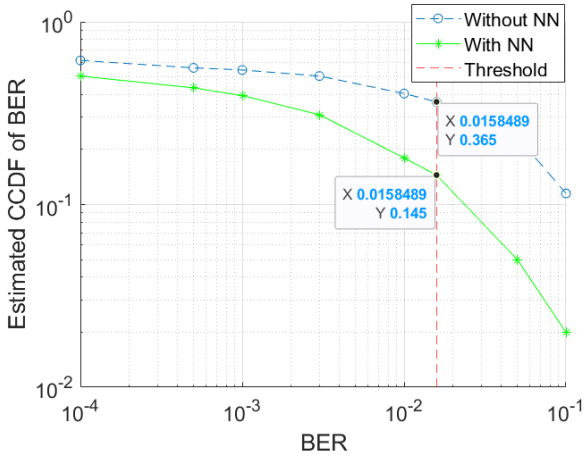
c) Estimated CCDF of the BER for $\Delta T/T_c$=0.5-100 kHz.



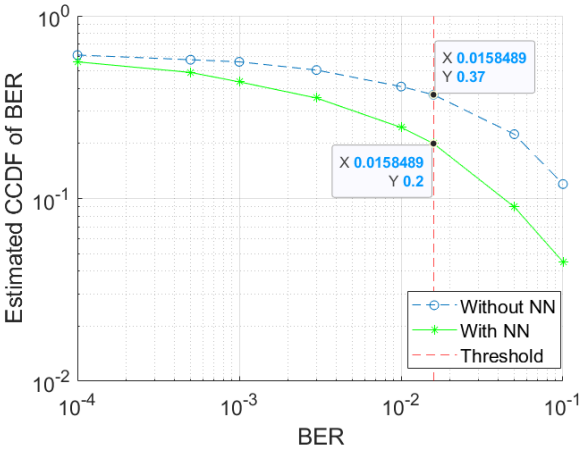
d) Estimated CCDF of the BER for $\Delta T/T_c$=1 – 100 kHz

Figure 42: Estimated CCDFs of the BER for several ΔT/Tc considering a laser with 100 kHz of linewidth.

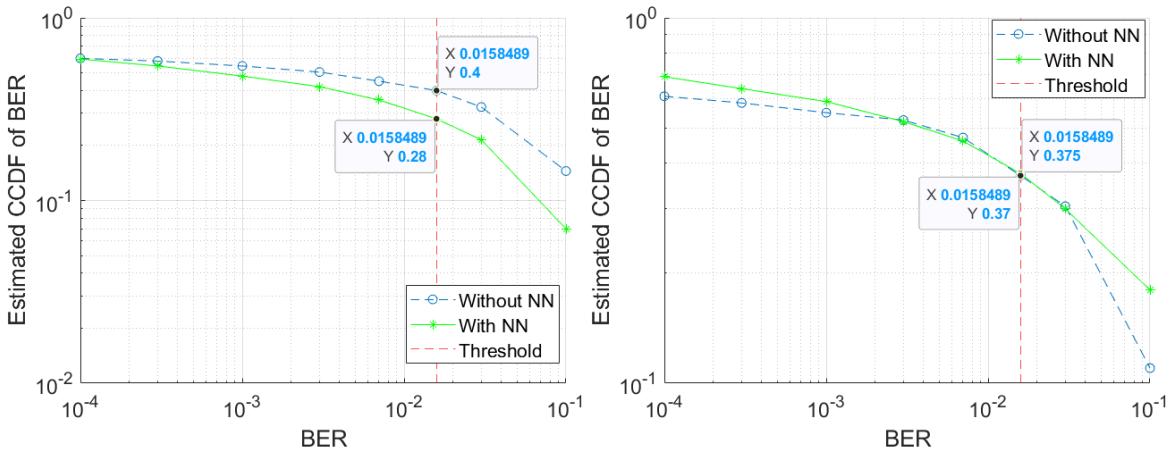
Figure 42 presents the estimated CCDF of the BER for several $\Delta T/T_c$, and 200 different time fractions, while employing an ECL with a 100 kHz linewidth. Figure 42a) shows, for $\Delta T/T_c$=0.1, an outage probability of 42% before the FFNN and 17% after the FFNN. This means that the FFNN provides a 25% improvement on the outage probability, compared with the outage probability before the FFNN. Figures 42b) and c) show that, when ΔT increases, the outage probability after the FFNN increases to 23% and 34.5%, representing an improvement of 19.5% and 8%, respectively. Figure 42d) tells us that, when ΔT is of the same order as coherence time, there is no improvement, with the outage probability being 1.5% worse after the FFNN. Therefore, the FFNN can only improve the performance when ΔT is kept as short as possible when, in reality, it was desired that it was as long as possible so that the FFNN did not have to be trained so often.



a) Estimated CCDF of the BER for $\Delta T/T_c$=0.1 – 1 MHz.



b) Estimated CCDF of the BER for $\Delta T/T_c$=0.2 – 1 MHz.

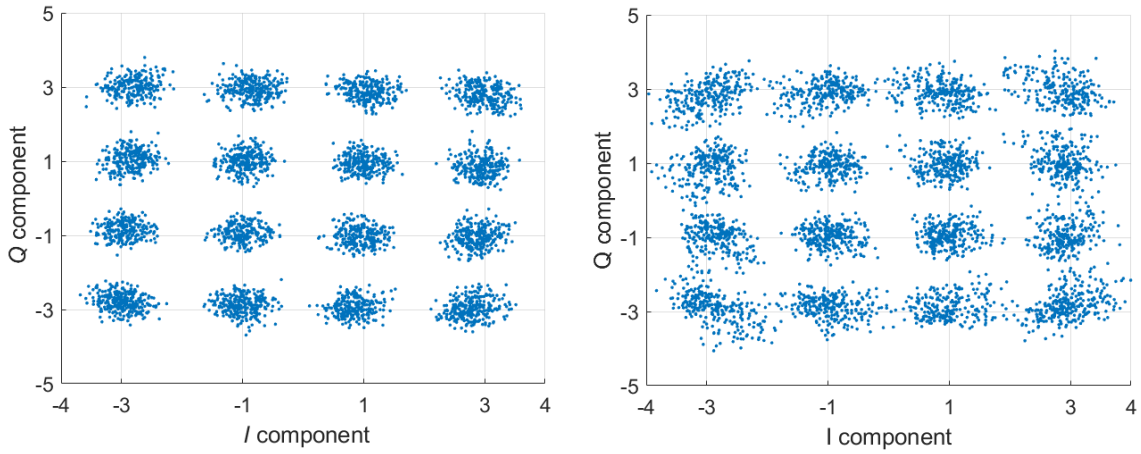


c) Estimated CCDF of the BER for $\Delta T/T_c$=0.5 – 1 MHz.

d) Estimated CCDF of the BER for $\Delta T/T_c$=1 – 1 MHz.

Figure 43: Estimated CCDFs of the BER for several $\Delta T/T_c$ considering a laser with 1 MHz of linewidth.

Figure 43 presents the estimated CCDF of the BER for several $\Delta T/T_c$, and 200 different time fractions, while employing a DFB laser with a 1 MHz linewidth. Figure 43a) shows, for $\Delta T/T_c$=0.1, an outage probability of 36.5% before the FFNN and 14.5% after the FFNN. This means that the FFNN provides a 22% improvement on the outage probability, compared with the outage probability before the FFNN. Figures 43b) and c) show that, when ΔT increases, the outage probability after the FFNN increases to 20% and 28%, representing an improvement of 17% and 12%, respectively. Figure 43d) tells us that, when ΔT is of the same order as coherence time, there is no improvement, with the outage probability being approximately the same before and after the FFNN. Figures 42 and 43 show that the performance of the system in general, when employing an ECL laser with a 100 kHz linewidth is worse than when a DFB laser with a 1 MHz linewidth is employed. This can be due to the fact that only 200 time fractions were considered, with the mean BER not being completely stable as shown in appendix B, resulting in a low confidence interval.

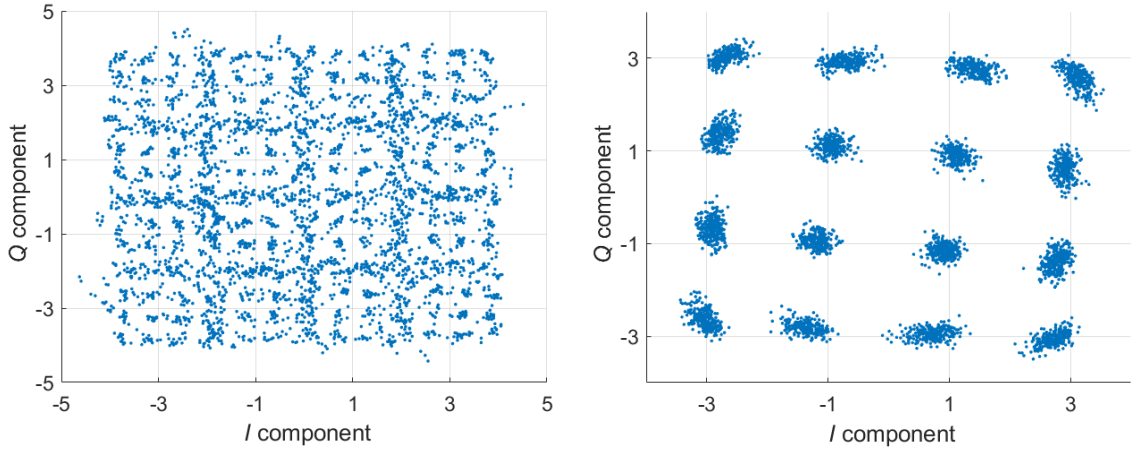


a) Constellation of the received signal in core m for $\Delta T/T_c$=0.2.

b) Constellation of the received signal in core m for $\Delta T/T_c$=1.

Figure 44: Constellations of the received signal in core m considering a DFB laser with a 1 MHz linewidth.

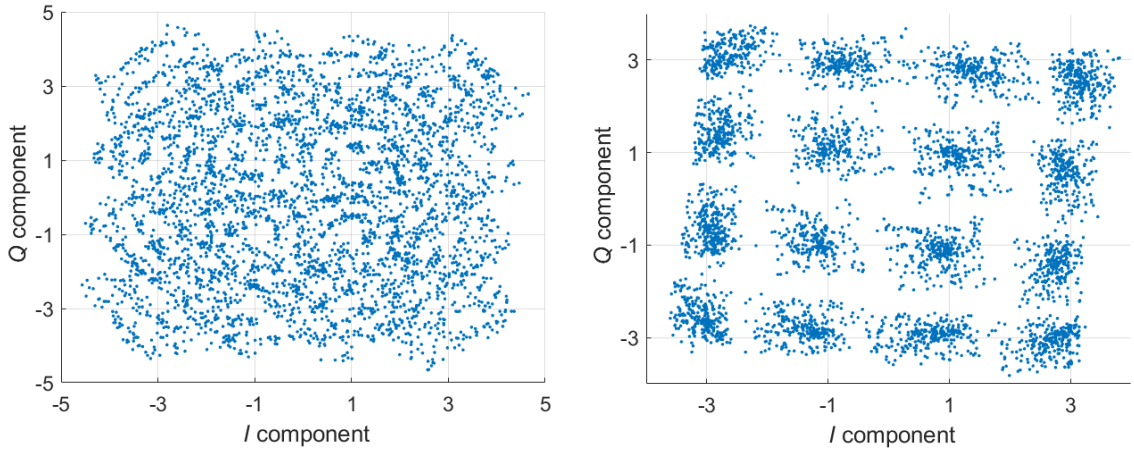
Figures 44a) and b) present the constellations of the received signal in core *m* for $\Delta T/T_c$=0.2 and $\Delta T/T_c$=1, respectively. These figures demonstrate the evolution of the LPN in the interfering core as ΔT increases. We can see the different effect the LPN has on the signal comparing figure 44a) and 44b).



a) Constellation of the received signal before the FFNN for $\Delta T/T_c$=0.2.

b) Constellation of the received signal after the FFNN for $\Delta T/T_c$=0.2.

Figure 45: Constellations of the received signal in core n for ΔT/Tc=0.2, considering a DFB laser with a 1 MHz linewidth.



a) Constellation of the received signal before the FFNN for $\Delta T/T_c$=0.5.

b) Constellation of the received signal after the FFNN for $\Delta T/T_c$=0.5.

Figure 46: Constellations of the received signal in core n for ΔT/Tc=0.5, considering a DFB laser with a 1 MHz linewidth.

Figures 45 and 46 present the constellations of the signal received in core *n* before and after the FFNN for $\Delta T/T_c$=0.2 and $\Delta T/T_c$=0.5, respectively. Figures 45a) to 46a) show that the constellations before the FFNN for both cases exhibit differences, due to the evolution of the LPN over time in the interfering core, as evidenced in figure 44. From the analysis of figures 45b) and 46b), we can see that, when the LPN is present, the FFNN introduces a slight rotation in the symbols. Furthermore, we can also see the performance deterioration of the FFNN, expressed by the larger variance of the received symbols in turn of each one of the 16 affixes of the optimum 16-QAM constellation in figure 46b).

CHAPTER 5

Conclusion and future work

5.1. Final conclusion

In this work, a 200 Gb/s short-reach self-coherent MCF system employing NNs was implemented in order to assess the impact of the LPN on the performance improvement provided by the NNs. The study has been performed for lasers with linewidths typical of ECLs and DFB lasers and considered the ICXT of the MCFs, KK receivers and a FFNN to mitigate the combined effects of the LPN and the ICXT.

In chapter 2, a literature review was conducted on the key elements of the system employed in this work, namely on self-coherent detection, space division multiplexing, LPN, ICXT, and ML applied to optical communications. Self-coherent detection and MCFs were shown to have potential for DC interconnects, which are of short distance and require increasingly higher bitrates at a low cost and complexity. It was also reported that, when skew×linewidth << 1 (the case considered in this work), the LPN has an insignificant impact on the ICXT, which are the two main impairments of the system considered in this work. Finally, NNs were reported to have been successfully employed in optical communications, namely at compensating transmission impairments such as the ICXT, and optimizing network performance.

Chapter 3 described the architecture of the 200 Gb/s short-reach self-coherent MCF system employing NNs, namely the simulation models considered for the transmitter, the weakly coupled MCF and KK receiver. In the LPN model, it was shown that different noise samples have independent walks due to the random nature of the model. Furthermore, the spectrums of the laser fields were shown to follow the theoretical Lorentzian curve. For the KK receiver model it was shown that, at the transmitter, the CSPR must be larger than the PAPR of the signal at the output of the DP-MZM, in order to fulfill the minimum phase condition. When the CSPR was set at 13 dB, the KK receiver perfectly detected the SSB signal, with the constellations displaying the received symbols without any distortion. The ICXT model was presented and validated through the evolution of the STAXT over time and the PDFs of the amplitude of the ICXT field components. It was shown that the STAXT may vary more than 20 dB, which is in accordance with the results previously reported. Moreover, it was demonstrated that the PDFs of the amplitude of the ICXT field components follow a Gaussian distribution, as already reported in theory.

In chapter 4, the impact of the LPN on the performance improvement provided by the NNs was evaluated. Firstly, the 200 Gb/s short-reach self-coherent MCF system employing NNs was studied considering ideal lasers, as reference. Afterwards, the LPN was introduced in the system, as the optical signal injected in the interfering core considered a laser with linewidth typical of ECLs and DFB lasers.

The optical signal injected in the interfered core considered an ideal laser without LPN. Due to the random evolution of the LPN over time, the study of the optimum number of training samples and neurons of the FFNN was accomplished considering a time window given by a fraction of the coherence time of the laser. With this, we want to guarantee that the variation of the phase noise field over time is limited, enabling adequate training of the proposed FFNN. The mean BER before the FFNN, with or without the LPN, was approximately the same, in accordance with the theory that the LPN does not impact the ICXT power. When studying the performance of the FFNN, it was concluded that the mean BER obtained with the ECL and DFB laser, compared to the reference case, increased more than one order of magnitude. The FFNN provided approximately the same performance for both lasers when the time interval between the training phase and the use of the FFNN (ΔT) did not exceed the coherence time, point where the system employing the ECL with a 100 kHz linewidth started to perform notably worse. Nevertheless, the mean BER obtained with both lasers was kept below the BER threshold when ΔT did not exceed 20% of the coherence time, although the ECL provides the advantage of the FFNN having to be trained less frequently, due to the longer coherence time when compared to the DFB laser. Finally, the outage probability of the system was evaluated and, considering a $\Delta T/Tc=0.1$, the FFNN provides a 25% and a 22% improvement on the outage probability when compared with the outage probability before the FFNN, while considering an ECL and a DFB laser, respectively. This improvement will decrease with the increase of the $\Delta T/Tc$.

5.2. Future work

This section presents suggestions for future work, based on the work developed in this dissertation:

- To assess the outage probability of the proposed system, considering more time fractions to obtain more precise outage probability levels.
- To study the possibility of adding another input feature to the FFNN related to the characteristics of the LPN.
- To evaluate the impact of the LPN on the performance improvement provided by NNs when employing lasers with LPN for both interfering and interfered cores with a skew×symbol rate << 1.
- To evaluate the impact of the LPN on the performance improvement provided by another type of ML algorithm when skew×symbol rate << 1.

Appendix

A. Feed-forward neural network with 2 input features

In [54], after the digital processing of the received signals from the interfered and interfering cores, the *I* and *Q* components from both signals were fed into the FFNN, resulting in 4 input features. To reduce the system complexity and understand if the information regarding the interfering core is expendable, it is important to study the possibility of only consider the *I* and *Q* components from the signal received in the interfered core as input features. Thus, a FFNN with 2 input features is studied in this appendix.

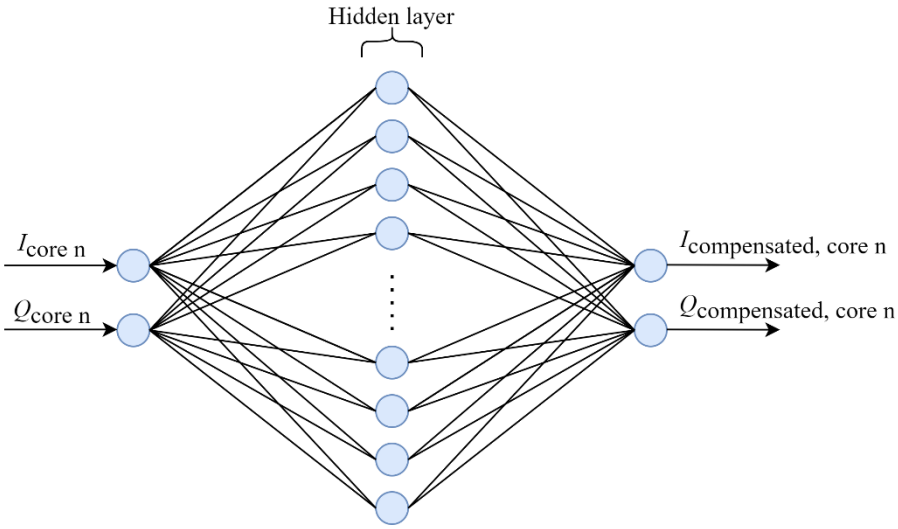


Figure 47: FFNN with 2 input features.

Figure 47 illustrates the structure of the FFNN with 2 inputs. First of all, the required number of neurons and training samples is evaluated. To do that, studies considering the EVM as a function of the number of training samples and number of neurons are conducted considering 4 time fractions, all with an EVM above -14 dB, thus ensuring that the ICXT significantly affects the signal. To finally decide on the optimum configuration of training samples and neurons, a study is conducted to choose the configuration that results in the lowest mean BER.

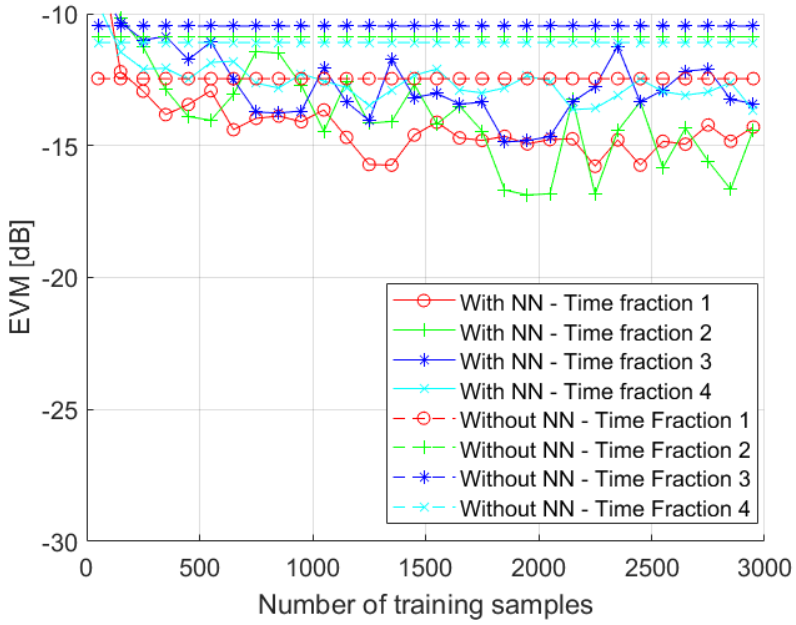
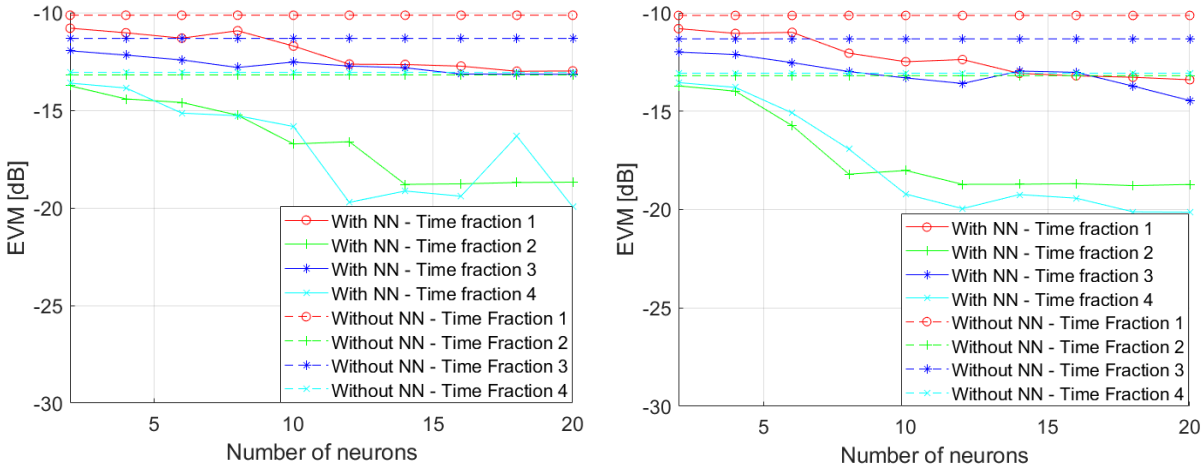


Figure 48: EVM vs #Training samples (10 neurons) – 2 inputs.

Figure 48 presents the EVM as a function of the number of training samples, initially considering 10 hidden layers neurons. All time fractions had an EVM of around -10 and -12 dB when ML was not considered. With the FFNN, results showed that the EVM for time fractions 1, 3 and 4 improves and starts to stabilize at around 1000 training samples. For time fraction 2, the EVM stabilizes at around 2000 samples. It is worth pointing out that, in time fractions 2, 3 and 4, considering 3000 training samples, the FFNN was able to reduce the EVM, compared with the case in which the FFNN is not used, at around 3 dB. In the following study, 3000 and 20 000 training samples are considered, so as to understand if drastically increasing the number of training samples shows better results.



a) EVM vs #Neurons (3000 samples).

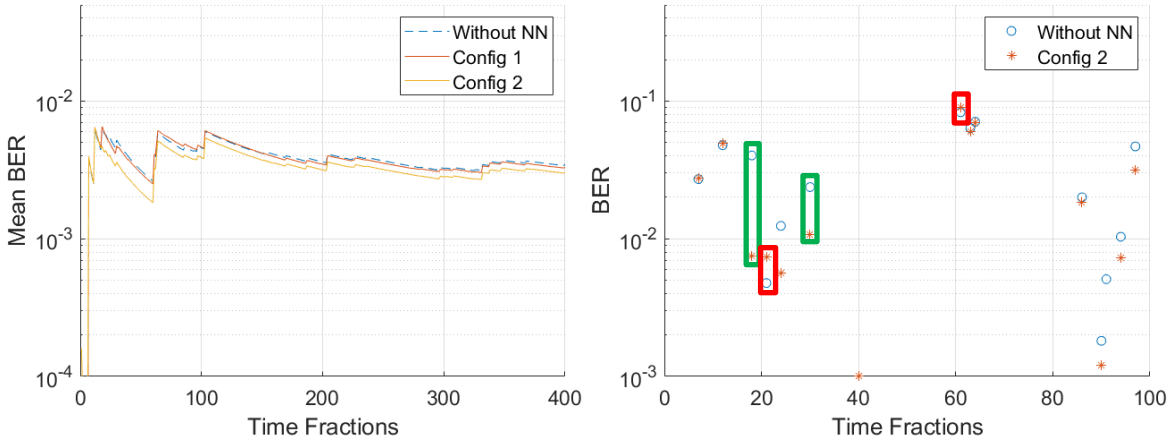
b) EVM vs #Neurons (20 000 samples).

Figure 49: EVM vs #Neurons – 2 inputs.

Figure 49 presents the EVM as a function of the number of neurons considering a) 3000 and b) 20 000 training samples. Figure 49a) shows that, a higher number of hidden layer neurons will result in a lower EVM. In this case, a number of neurons close to 12 or 14 seem to be the optimum number, with EVMs being as low as around -19 dB for time fractions 2 and 4. Compared to the previous study, considering a FFNN with 3000 samples and 10 neurons, time fractions 2 and 4 have the same EVM improvement of around 3 dB. However, time fractions 1 and 3 present an improvement of around 1 dB. In figure 49b), the same study is conducted, this time considering 20 000 training samples. It is possible to conclude that, considering 12 or 14 neurons, significantly increasing the number of training samples does not necessarily result in a significantly lower EVM, which means that the optimum number of training samples is 3000. To decide between 12 and 14 neurons, a study considering the mean BER along a few hundred time fractions with 2 possible FFNN configurations is going to be conducted. However, from these preliminary studies, a FFNN with only these two features does not appear to effectively compensate for the ICXT. Table 12 presents the 2 FFNN configurations in study.

Table 12: FFNN configurations – 2 inputs without LPN.

Configuration	Number of training samples	Number of neurons
1	3000	12
2	3000	14



a) Mean BER vs 400 time fractions. *b) BER vs 100 time fractions.*

Figure 50: BER evolution before and after the FFNN – 2 inputs.

Figure 50 shows the BER evolution before and after the use of the FFNN. Figure 50a) confirms that the optimum configuration will be number 2 with 3000 training samples and 14 neurons, while the mean BER seems to stabilize at around 400 time fractions. Figure 50b) shows the evolution of the BER along 100 time fractions, considering an ICXT level of -13 dB, a sequence of 2^{17} bits and the FFNN with the optimum configuration. These results show that this FFNN is sometimes able to provide some sort of

improvement, with the performance varying from time fraction to time fraction, as it can be seen in time fraction 18 and 30 in figure 50b) (highlighted in green). There are also some cases where the use of the NN results in performance deterioration such as time fraction 21 and 61 (highlighted in red). From these 100 time fractions, several resulted in a BER without a reasonable confidence interval or with less than 1 errored bit in a sequence of 2^{17} bits, reason why they are not shown in the figure 50b). Both these figures suggest that this configuration will, most likely, not lead to an effective ICXT compensation. Nevertheless, the following study considering the evolution of the mean BER (with or without the NN) according to the ICXT level will provide conclusive results.

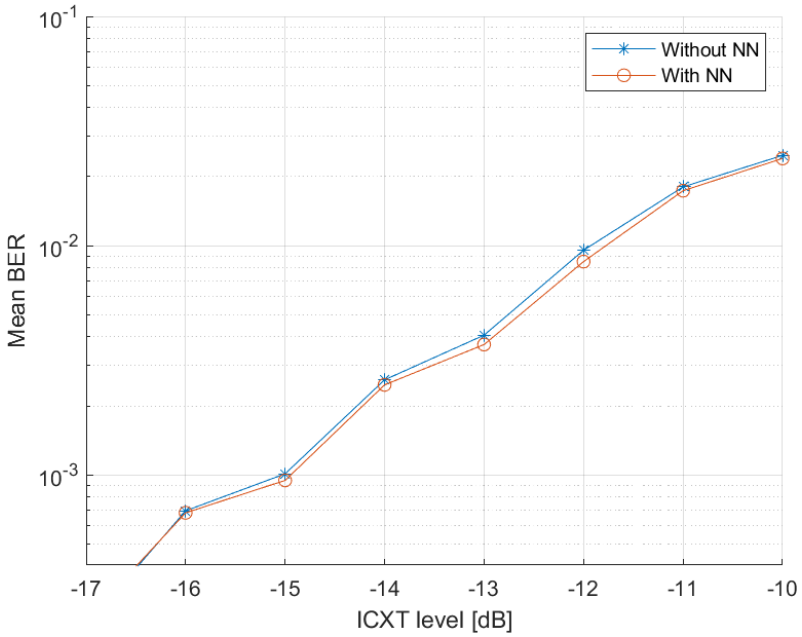


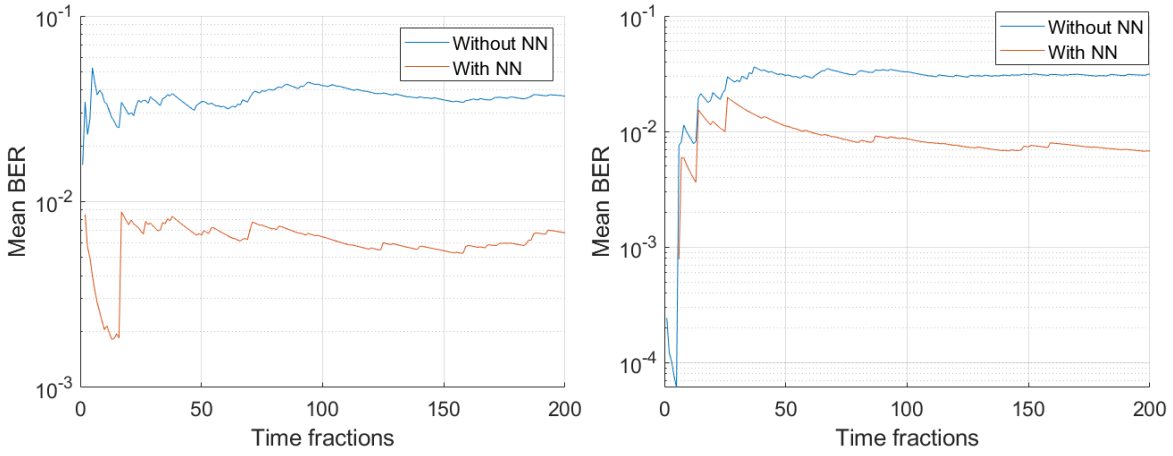
Figure 51: Mean BER vs ICXT level – 2 inputs.

Figure 51 presents the mean BER as a function of the ICXT level. From it, it is finally possible to conclude that the FFNN with only the 2 input features – *I* and *Q* components of the received signal in the interfered core – does not provide any notable improvement on the performance of the 200 Gb/s short-reach self-coherent MCF system. These results were obtained considering a sequence of 2^{18} bits and the BER was averaged over 400 time fractions, for each level.

B. Evolution of the mean BER considering an ECL and a DFB laser

The results of the mean EVM and mean BER without the FFNN as a function of $\Delta T/T_c$ while employing an ECL and a DFB laser were slightly different. It is known from [50] and [51] that the LPN does not impact the ICXT power, so it was expected that the mean EVM and mean BER results would be the same. These results suggest that the mean BER was not stabilized, since only 200 time fractions were

considered due to time constraints. To assess that, a study of the mean BER over the 200 time fractions considered, was conducted when for both lasers.



a) 100 kHz. *b) 1 MHz.*

Figure 52: Evolution of the mean BER as a function of the time fractions while employing a) an ECL with a 100 kHz linewidth, and b) a DFB laser with a 1 MHz linewidth.

Figure 52 presents the evolution of the mean BER as a function of the time fractions while employing a) an ECL with a 100 kHz linewidth, and b) a DFB laser with a 1 MHz linewidth. As we are able to see through figure 52a), when employing an ECL with a 100 kHz linewidth, the mean BER with and without the NN is not completely stabilized yet. Furthermore, figure 52b) shows that when a DFB laser is employed, the mean BER without NN seems to be stabilized. However, the mean BER when the FFNN is employed, is not stabilized yet.

C. Electrical Noise

The electrical noise present at the receiver can be characterized by the NEP. This parameter is used by the manufacturers to quantify the effect of the circuit noise. Through this parameter, it is possible to obtain the one sided PSD of the electrical circuit noise [43]. This is done by:

$$S_c(f) = (NEP \times R_\lambda)^2 \quad (31)$$

where R_λ is the responsivity.

To calculate the noise power in the simulation bandwidth, the two sided PSD of the electrical noise is considered, which is obtained by:

$$S_{c,2}(f) = \frac{S_c(f)}{2} \quad (32)$$

Then, the noise power distributed uniformly along the simulation bandiwdth can be obtained:

$$p_n = S_{c,2}(f) \times f_s \quad (33)$$

where f_s is the sampling frequency. Finally, the random electrical noise generated for each time sample is obtained with the multiplication of $\sqrt{p_n}$ with a random scalar drawn from the standard normal distribution.

Bibliography

[1] K. Saitoh and S. Matsuo, “Multicore fiber technology,” Journal of Lightwave Technology, vol. 34, no. 1, pp. 55-66, Jan. 2016, doi: 10.1109/JLT.2015.2466444.

[2] W. Klaus, P. J. Winzer, and K. Nakajima, “The role of parallelism in the evolution of optical fiber communication systems,” Proceedings of the IEEE, vol. 110, no. 11, pp. 1619–1654, Nov. 2022, doi: 10.1109/jproc.2022.3207920.

[3] P. Winzer, D. Neilson, and A. Chraplyvy, “Fiber-optic transmission and networking: the previous 20 and the next 20 years [Invited],” Opt. Express, vol. 26, no. 18, pp. 24190-24239, Sep. 2018, doi: 10.1364/OE.26.024190.

[4] D. Butler, M. Li, S. Li, Y. Geng, R. Khrapko, R. Modavis, and V. Nazarov, “Space division multiplexing in short reach optical interconnects,” Journal of Lightwave Technology, vol. 35, no. 4, pp. 677-682, Feb. 2017, doi: 10.1109/JLT.2016.2619981.

[5] K. Saitoh, “Multi-core fiber technology for SDM: coupling mechanisms and design,” Journal of Lightwave Technology, vol. 40, no. 5, pp. 1527–1543, Mar. 2022, doi: 10.1109/jlt.2022.3145052.

[6] T. Hayashi, T. Nagashima, T. Morishima, Y. Saito, and T. Nakanishi, “Multi-core fibers for data center applications,” 45th European Conference on Optical Communication, Dublin, Ireland, 2019, pp. 1-4, doi: 10.1049/cp.2019.0754.

[7] S. Adhikari, S. Jansen, M. Alfiad, B. Inan, V. Sleiffer, A. Lobato, P. Leoni, and W. Rosenkranz, “Self- coherent optical OFDM: an interesting alternative to direct or coherent detection,” 13th International Conference on Transparent Optical Networks, Stockholm, Sweden, 2011, pp. 1-4, doi: 10.1109/ICTON.2011.5971099.

[8] S. T. Le, K. Schuh, M. Chagnon, F. Buchali, R. Dischler, V. Aref, H. Buelow, and K. Engenhardt, “1.72-Tb/s virtual-carrier-assisted direct-detection transmission over 200 km,” Journal of Lightwave Technology, vol. 36, no. 6, pp. 1347–1353, Mar. 2018, doi: 10.1109/jlt.2017.2779331.

[9] V. Kamalov, L. Jovanovski, V. Vusirikala, S. Zhang, F. Yaman, K. Nakamura, T. Inoue, E. Mateo, and Y. Inada, “Evolution from 8QAM live traffic to PS 64-QAM with neural-network based nonlinearity compensation on 11000 km open subsea cable,” Optical Fiber Communications Conference and Exposition, San Diego, CA, USA, 2018, pp. 1-3, Paper Th4D.5.

[10] Z. Xu, C. Sun, T. Ji, J. H. Manton, and W. Shieh, “Computational complexity comparison of feedforward/radial basis function/recurrent neural network-based equalizer for a 50-Gb/s PAM4 direct-detection optical link,” Optics Express, vol. 27, no. 25, pp. 36953-36964, Dec. 2019, doi: 10.1364/oe.27.036953.

[11] D. Piedade, T. Alves, and T. Brandão, “Short-reach MCF-based systems employing KK receivers and feedforward neural networks for ICXT mitigation,” Photonics, vol. 9, no. 5, pp. 286-286, Apr. 2022, doi: 10.3390/photonics9050286.

[12] K. Peffers, T. Tuunanen, C. Gengler, M. Rossi, W. Hui, V. Virtanen, and J. Bragge, “Design science research process: a model for producing and presenting information systems research,” Journal of Management Information Systems, vol. 24, no. 3, Dec. 2007, doi: 10.2753/MIS0742-1222240302.

[13] D. Uzunidis, M. Logothetis, A. Stavdas, D. Hillerkuss, and I. Tomkos, “Fifty years of fixed optical networks evolution: a survey of architectural and technological developments in a layered approach,” Telecom, vol. 3, no. 4. MDPI AG, pp. 619–674, Nov. 2022. doi: 10.3390/telecom3040035.

[14] J. Pires, “Sistemas e Redes de Telecomunicações”. Lisboa, PT: IST, 2006.

[15] X. Pang, O. Ozolins, R. Lin, L. Zhang, A. Udalcovs, L. Xue, R. Schatz, U. Westergren, S. Xiao, W. Hu, G. Jacobsen, S. Popov, and J. Chen, “200 Gb/s/lane IM-DD technologies for short-reach optical interconnects,” Journal of Lightwave Technology, vol. 38, no. 2, pp. 492-503, Jan. 2020, doi: 10.1109/JLT.2019.2962322.

[16] J. M. Kahn, J. Krause Perin, and A. Shastri, “Data center links beyond 100 Gb/s per wavelength,” Optical Fiber Technology, vol. 44., pp. 69–85, Aug. 2018. doi: 10.1016/j.yofte.2017.12.006.

[17] M. Rapisarda, J. A. Hernández, A. Gatto, P. Parolari, P. Boffi, M. Svaluto Moreolo, J. M. Fábrega, L. Nadal, R. Martínez, V. López, J.-P. Fernández-Palacios, G. Otero, and D. Larrabeiti, “All-optical aggregation and distribution of traffic in large metropolitan area networks using multi-Tb/s S-BVTs,” Journal of Optical Communications and Networking, vol. 14, no. 5, pp. 316-326, May 2022, doi: 10.1364/JOCN.448115.

[18] E. Agrell, M. Karlsson, A. R. Chraplyvy, D. J. Richardson, P. M. Krummrich, P. Winzer, K. Roberts, J. K. Fischer, S. J. Savory, B. J. Eggleton, M. Secondini, F. R. Kschischang, A. Lord, J. Prat, I. Tomkos, J. E. Bowers, S. Srinivasan, M. Brandt-Pearce, and N. Gisin, “Roadmap of optical communications,” Journal of Optics, vol. 18, no. 6. IOP Publishing, p. 063002, May 2016, doi: 10.1088/2040-8978/18/6/063002.

[19] A. Lord, P. Wright, and A. Mitra, “Core networks in the flexgrid era,” Journal of Lightwave Technology, vol. 33, no. 5, pp. 1126–1135, Mar. 2015, doi: 10.1109/jlt.2015.2396685.

[20] Z. Li, M. S. Erkilinc, S. Pachnicke, H. Griesser, B. C. Thomsen, P. Bayvel, and R. I. Killey, “Direct-detection 16-QAM nyquist-shaped subcarrier modulation with SSBI mitigation,” IEEE International Conference on Communications, London, UK, 2015, pp. 5204-5209, doi: 10.1109/ICC.2015.7249150.

[21] S. Kanazawa, H. Yamazaki, Y. Nakanishi, T. Fujisawa, K. Takahata, Y. Ueda, W. Kobayashi, Y. Muramoto, H. Ishii, and H. Sanjoh, “Transmission of 214-Gbit/s 4-PAM signal using an ultra-broadband lumped-electrode EADFB laser module,” Optical Fiber Communications Conference Postdeadline Papers, Anaheim, CA, USA, 2016, doi: 10.1364/OFC.2016.Th5B.3.

[22] K. Kikuchi, “Fundamentals of coherent optical fiber communications,” Journal of Lightwave Technology, vol. 34, no. 1, pp. 157–179, Jan. 2016, doi: 10.1109/jlt.2015.2463719.

[23] M. Torbatian, D. Lavery, M. Osman, D. Yao, D. Millar, Y. Gao, A. Kakkar, Z. El-Sahn, C. Doggart, A. E. Morra, N. Abughalieh, S. Yang, X. Chen, R. Maher, H. Sun, K.-T. Wu, and P. Kandappan, "Performance oriented DSP for flexible long haul coherent transmission," Journal of Lightwave Technology, vol. 40, no. 5, pp. 1256–1272, Mar. 2022, doi: 10.1109/jlt.2021.3134155.

[24] J. Yu, M. Kong, H. Chien, K. Wang, J. Shi, X. Li, X. Pan, X. Xin, Y. Xia, B. Ye, X. Wei, T. Wang, and Y. Chen., "400G/channel 50-GHz WDM coherent transmission: PS 64QAM versus hybrid 32/64QAM," Optical Fiber Communications Conference and Exhibition, San Diego, CA, USA, 2019, pp. 1-3, paper Th3G.3.

[25] M. Kong, K. Wang, J. Ding, J. Zhang, W. Li, J. Shi, F. Wang, L. Zhao, C. Liu, Y. Wang, W. Zhou, and J. Yu, "640-Gb/s/carrier WDM transmission over 6,400 km based on PS-16QAM at 106 Gbaud employing advanced DSP," Journal of Lightwave Technology, vol. 39, no. 1, pp. 55–63, Jan. 2021, doi: 10.1109/jlt.2020.3024771.

[26] M. Kong, J. Shi, B. Sang, J. Ding, K. Wang, W. Li, F. Wang, C. Liu, Y. Wang, Y. Wei, B. Zhu, L. Zhao, W. Zhou, and J. Yu, "800-Gb/s/carrier WDM coherent transmission over 2000 km based on truncated PS-64QAM utilizing MIMO Volterra equalizer," Journal of Lightwave Technology, vol. 40, no. 9, pp. 2830–2839, May 2022, doi: 10.1109/jlt.2022.3148336.

[27] M. Noormohammadpour and C. S. Raghavendra, "Datacenter traffic control: understanding techniques and tradeoffs," IEEE Communications Surveys & Tutorials, vol. 20, no. 2, pp. 1492–1525, Dec. 2017, doi: 10.1109/comst.2017.2782753.

[28] Cisco, "Global cloud index: forecast and methodology", 2016–2021, White Paper, 2018, pp. 1–46.

[29] L. Zhang, J. Chen, E. Agrell, R. Lin, and L. Wosinska, "Enabling technologies for optical data center networks: spatial division multiplexing," Journal of Lightwave Technology, vol. 38, no. 1, pp. 18–30, Jan. 2020, doi: 10.1109/jlt.2019.2941765.

[30] C. Xie and J. Cheng, "Coherent optics for data center networks," IEEE Photonics Society Summer Topicals Meeting Series, Cabo San Lucas, Mexico, 2020, pp. 1-2, doi: 10.1109/SUM48678.2020.9161052.

[31] X. Zhou, R. Urata, and H. Liu, "Beyond 1Tb/s datacenter interconnect technology: challenges and solutions," Optical Fiber Communications Conference and Exhibition, San Diego, CA, USA, 2019, pp. 1-3, paper Tu2F.5.

[32] R. Nagarajan, M. Filer, Y. Fu, M. Kato, T. Rope, and J. Stewart, "Silicon photonics-based 100 Gbit/s, PAM4, DWDM data center interconnects," Journal of Optical Communications and Networking, vol. 10, no. 7, pp. 25-36, July 2018, doi: 10.1364/JOCN.10.000B25.

[33] S. Beppu, D. Soma, S. Sumita, Y. Wakayama, H. Takahashi, T. Tsuritani, I. Morita and M. Suzuki, "402.7-Tb/s MDM-WDM transmission over weakly coupled 10-mode fiber using rate-adaptive PS-16-QAM signals," Journal of Lightwave Technology, vol. 38, no. 10, pp. 2835-2841, doi: 10.1109/JLT.2020.2979195.

[34] G. Rademacher, R. S. Luís, B. J. Puttnam, N. K. Fontaine, M. Mazur, H. Chen, R. Ryf, D. T. Neilson, D. Dahl, J. Carpenter, P. Sillard, F. Achten, M. Bigot, J. Sakaguchi, and H. Furukawa, "1.53 Petabit/s C-band transmission in 55-mode fiber," European Conference on Optical Communications (ECOC), Basel, Switzerland, 2022, paper Th3C.3.

[35] P. Mishra, S. Saxena, S. Munige, A. Pandey, and A. Mishra, "Successful 200G transmission over 45 km of 4-core single mode MCF," Workshop on Recent Advances in Photonic, Mumbai, India, 2022, pp. 1-2, doi: 10.1109/WRAP54064.2022.9758272.

[36] Y. Sun, R. Lingle, B. Holland, R. Shubochkin, K. Bansal and D. DiGiovanni, "System transmission over multicore fiber for datacom optical interconnect applications," IEEE CPMT Symposium Japan, Kyoto, Japan, 2021, pp. 142-145, doi: 10.1109/ICSJ52620.2021.9648899.

[37] G. Rademacher, R. S. Luis, B. J. Puttnam, R. Ryf, S. van der Heide, T. A. Eriksson, N. K. Fontaine, H. Chen, R.-J. Essiambre, Y. Awaji, H. Furukawa, and N. Wada, "High-capacity transmission in a coupled-core three-core multi-core fiber," Journal of Lightwave Technology, vol. 39, no. 3, pp. 757–762, Feb. 2021, doi: 10.1109/jlt.2020.3013966.

[38] D. Soma, S. Beppu, Y. Wakayama, S. Sumita, H. Takahashi, N. Yoshikane, I. Morita, T. Tsuritani, and M. Suzuki, "50.47-Tbit/s standard cladding coupled 4-core fiber transmission over 9,150 km," Journal of Lightwave Technology, vol. 39, no. 22, pp. 7099–7105, Nov. 2021, doi: 10.1109/jlt.2021.3109890.

[39] R. S. Luis, G. Rademacher, B. J. Puttnam, T. A. Eriksson, H. Furukawa, A. Ross-Adams, S. Gross, M. Withford, N. Riesen, Y. Sasaki, K. Saitoh, K. Aikawa, Y. Awaji, and N. Wada, "1.2 Pb/s throughput transmission using a 160 μm cladding, 4-core, 3-mode fiber," Journal of Lightwave Technology, vol. 37, no. 8, pp. 1798–1804, Apr. 2019, doi: 10.1109/jlt.2019.2902601.

[40] G. Rademacher, B. J. Puttnam, R. S. Luís, J. Sakaguchi, W. Klaus, T. A. Eriksson, Y. Awaji, T. Hayashi, T. Nagashima, T. Nakanishi, T. Taru, T. Takahata, T. Kobayashi, H. Furukawa, and N. Wada, "10.66 Peta-bit/s transmission over a 38-core-three-mode fiber," Optical Fiber Communications Conference and Exhibition, San Diego, CA, USA, 2020, paper Th3H.1.

[41] T. M. F. Alves and A. V. T. Cartaxo, "Characterization of the stochastic time evolution of short-term average intercore crosstalk in multicore fibers with multiple interfering cores," Optics Express, vol. 26, no. 4, pp. 4605-4620, Feb. 2018, doi: 10.1364/OE.26.004605.

[42] B. R. P. Pinheiro, J. L. Rebola and A. V. T. Cartaxo, "Analysis of inter-core crosstalk in weakly-coupled multi-core fiber coherent systems," Journal of Lightwave Technology, vol. 39, no. 1, pp. 42-54, Jan. 2021, doi: 10.1109/JLT.2020.3024609.

[43] A. Cartaxo, "Transmissão por fibra óptica", Departamento de Engenharia Electrotécnica e de Computadores, Instituto Superior Técnico, Lisbon, Portugal, 2005.

[44] M. Pereira, "Performance limitations of 40 Gb/s SSB MB-OFDM metropolitan networks induced by phase-to-intensity conversion of laser phase noise", MSc Dissertation, Instituto Superior Técnico, Lisbon, Portugal, 2015.

[45] F. N. Khan, Q. Fan, C. Lu, and A. P. T. Lau, "An optical communication's perspective on machine learning and its applications," Journal of Lightwave Technology, vol. 37, no. 2, pp. 493–516, Jan. 2019, doi: 10.1109/jlt.2019.2897313.

[46] J. W. Nevin, S. Nallaperuma, N. A. Shevchenko, X. Li, Md. S. Faruk, and S. J. Savory, "Machine learning for optical fiber communication systems: an introduction and overview," APL Photonics, vol. 6, no. 12, p. 121101, Dec. 2021, doi: 10.1063/5.0070838.

[47] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA, USA: MIT Press, 2016. Accessed: Dec. 5, 2022. [Online]. Available: https://www.deeplearningbook.org/

[48] T. S. R. Shen and A. P. T. Lau, "Fiber nonlinearity compensation using extreme learning machine for DSP-based coherent communication systems," 16th Opto-Electronics and Communications Conference, Kaohsiung, Taiwan, 2011, pp. 816–817.

[49] T. M. F. Alves, A. V. T. Cartaxo and J. L. Rebola, "Stochastic properties and outage in crosstalk-impaired OOK-DD weakly-coupled MCF applications with low and high skew×bit-rate," IEEE Journal of Selected Topics in Quantum Electronics, vol. 26, no. 4, p. 4300208, Aug. 2020, doi: 10.1109/JSTQE.2020.2995306.

[50] T. M. F. Alves, A. V. T. Cartaxo and J. L. Rebola, "DD-OOK multi-core fiber systems impaired by intercore crosstalk and laser phase noise," Journal of Lightwave Technology, vol. 40, no. 5, pp. 1544-1551, Mar. 2022, doi: 10.1109/JLT.2021.3138186.

[51] F. M. Saraiva, "Laser phase noise impaired next generation short-reach networks employing MCFs and Kramers Kronig receivers", MSc Dissertation, Iscte - Instituto Universitário de Lisboa, Lisbon, Portugal, 2022.

[52] R. O. J. Soeiro, T. M. F. Alves and A. V. T. Cartaxo, "Dual polarization discrete changes model of inter-core crosstalk in multi-core fibers," IEEE Photonics Technology Letters, vol. 29, no. 16, pp. 1395-1398, Aug. 2017, doi: 10.1109/LPT.2017.2723662.

[53] M. Azizoglu and P. A. Humblet, "Optical DPSK with generalized phase noise model and narrowband reception," Proceedings of ICC '93 - IEEE International Conference on Communications, vol. 3, pp. 1591-1596, 1993, doi: 10.1109/ICC.1993.397552.

[54] D. Piedade, "Next generation >200 Gb/s multicore fiber short-reach networks employing machine learning", MSc Dissertation, Iscte – Instituto Universitário de Lisboa, Lisbon, Portugal, 2022.

[55] YOFC, "Few-mode fiber" https://en.yofc.com/view/2351.html https://myphotos2020.oss-cn-beijing.aliyuncs.com/en/upload/20190321/1d6fird0ol9jn4cr.pdf (accessed Oct. 10, 2023).

[56] G. Qi, J. Yao, J. Seregelyi, S. Paquet, C. Bélisle, X. Zhang, K. Wu, R. Kashyap, "Phase-noise analysis of optically generated millimeter-wave signals with external optical modulation techniques," Journal of Lightwave Technology, vol. 24, no. 12, pp. 4861-4875, Dec. 2006, doi: 10.1109/JLT.2006.884990.