INSTITUTO
UNIVERSITÁRIO
DE LISBOA

Next generation >200 Gb/s multicore fiber short-reach networks employing machine learning

Derick Augusto Évora Piedade

Master in Telecommunications and Computer Engineering

Supervisor:
Ph.D Tiago Manuel Ferreira Alves, Assistant Professor
Iscte - Instituto Universitário de Lisboa

Co-Superviser:
Ph.D Tomás Gomes Silva Serpa Brandão, Assistant Professor
Iscte - Instituto Universitário de Lisboa

September, 2022

TECNOLOGIAS
E ARQUITETURA

Department of Information Science and Technology

Next generation >200 Gb/s multicore fiber short-reach networks employing machine learning

Derick Augusto Évora Piedade

Master in Telecommunications and Computer Engineering

Supervisor:
Ph.D Tiago Manuel Ferreira Alves, Assistant Professor
Iscte - Instituto Universitário de Lisboa

Co-Superviser:
Ph.D Tomás Gomes Silva Serpa Brandão, Assistant Professor
Iscte - Instituto Universitário de Lisboa

September, 2022

Acknowledgment

First, I would like to acknowledge my supervisors for all the support and advice. I would like to thank Professor Tiago Manuel Ferreira Alves and Professor Tomás Brandão, for all the patience and availability to help me develop the competence for this dissertation. For guide me and helped me acquire new knowledges e made me want to keep learning about the subject.
Also, I would like to thank the support provided by Instituto de Telecomunicações (IT) Portugal under the internal projects DigCore/UIDP/50008/2020, UIDB/50008/2020, and grant BI Nº22-09-04-2021. I also would like to acknowledge ISTAR for the support provided through ISTAR projects UIDB/04466/2020 and UIDP/04466/2020.

Resumo

Este trabalho propõe e avalia o uso de técnicas de machine learning (ML) em sistemas de curto alcance com ritmo binário superior a 200 Gb/s utilizando receptores Kramers-Kronig (KK) e fibras multinúcleo (MCF). Os sistemas de curto alcance usualmente encontrados em conexões intra-data centers (DC) exigem receptores de deteção direta (DD) de baixo custo. Os receptores KK permitem a combinação de sistemas de modulação de maior ordem, tais como o 16-QAM, usado em sistemas coerentes, com o baixo custo dos receptores DD. Portanto, o uso de receptores KK permite melhorar o ritmo binário e eficiência espectral e manter a eficiência de custo dos sistemas DD, o que é importante em DC. O uso de fibras multinúcleo permite o aumento da capacidade do sistema, bem como a densidade de cabos. No entanto, o uso de MCF introduz uma distorção adicional no sistema conhecida como inter-core crosstalk (ICXT). Para mitigar os efeitos do ICXT aleatório, são propostas e avaliadas técnicas de ML de baixa complexidade como *k*-means clustering, *k* nearest neighbor (KNN) e rede neuronais artificiais (ANN).

O desempenho associado à utilização de algoritmos de ML (*k*-means, KNN e duas redes neuronais do tipo *feedforward* (FNN): uma para estimação e outra para classificação), é avaliado e comparado com o desempenho do sistema obtido sem o uso de ML. A utilização de FNN para estimação e classificação conduziram a uma melhoria significativa no desempenho do sistema, mitigando o impacto do ICXT no sinal recebido. Isso é alcançado com o uso de uma rede neuronal com uma arquitetura muito simples contendo quatro entradas e 10 neurónios na camada escondida. Foi demonstrado que os algoritmos *k*-means e KNN não proporcionam melhoria de desempenho em comparação com o sistema sem o uso de ML. Essas conclusões são válidas para sistemas DD de curto alcance baseados em MCF com o produto entre o skew (atraso relativo entre os núcleos) e o ritmo de símbolos muito menor que um ($skew \times symbol\ rate \ll 1$). Com o uso das ANNs, o sistema apresenta uma melhoria de aproximadamente 12 dB na probabilidade de indisponibilidade quando comparado com o sistema sem o uso de ML. Para o limite de BER de $10^{-1.8}$, e comparado com o sistema padrão sem o uso de técnicas de ML, o sistema com o uso de ANN mostra uma melhoria na potência óptica recebida de quase 3 dB e uma melhoria no nível de ICXT de aproximadamente 9 dB em relação ao BER médio.

Palavras-chaves: Redes de curto alcance, fibra multi-núcleo, diafonia entre núcleos, redes neuronais, aprendizagem máquina, receptores Krames-Kronig.

Abstract

This work proposes and evaluates the use of machine learning (ML) techniques on >200 Gb/s short-reach systems employing weakly coupled multicore fiber (MCF) and Kramers-Kronig (KK) receivers. The short-reach systems commonly found in intra data centers (DC) connections demand low cost-efficient direct detection receivers (DD). The KK receivers allow the combination of higher modulation order, such as 16-QAM used in coherent systems, with the low complexity and low cost of DD. Thus, the use of KK receivers allows to increase the bit rate and spectral efficiency while maintaining the cost of DD systems as this is an important requirement in DC. The use of MCF allows to increase the system capacity as well as the system cable density, although the use of MCF induces additional distortion, known as inter-core crosstalk (ICXT), to the system. Thus, low complexity ML techniques such as *k*-means clustering, *k* nearest neighbor (KNN) and artificial neural network (ANN) (estimation feedforward neural network (FNN) and classification feedforward neural network) are proposed to mitigate the effects of random ICXT.

The performance improvement provided by the *k*-means clustering, KNN and the two types of FNN techniques is assessed and compared with the system performance obtained without the use of ML. The use of estimation and classification FNN prove to significantly improve the system performance by mitigating the impact of the ICXT on the received signal. This is achieved by employing only 10 neurons in the hidden layer and four input features. It has been shown that *k*-means or KNN techniques do not provide performance improvement compared to the system without using ML. These conclusions are valid for direct detection MCF-based short-reach systems with the product between the skew (relative time delay between cores) and the symbol rate much lower than one ($skew \times symbol\ rate \ll 1$). By employing the proposed ANNs, the system shows an improvement of approximately 12 dB on the ICXT level, for the same outage probability when comparing with the system without the use of ML. For the BER threshold of $10^{-1.8}$ and compared with the standard system operating without employing ML techniques, the system operating with the proposed ANNs show a received optical power improvement of almost 3 dB and a ICXT level improvement of approximately 9 dB when the mean BER is analized.

Keywords: Short-reach network, multicore fiber, intercore crosstalk, neural networks, machine learning, Kramers-Kronig receivers.

Contents

List of figures

List of tables

List of acronyms

AI Artificial intelligence

ANN Artificial neural network

AR-RNN Auto-regressive recurrent neural network

BER Bit error rate

CD Chromatic dispersion

CSPR Carrier-to-signal power ratio

CW Continuous wave

DAC Digital to analog converter

DC Data center

DCF Dispersion compensating fiber

DD Direct detection

DMD Differential mode delay

DP-DCM Dual polarization discrete change model

DSP Digital signal processing

DWDM Dense wavelength division multiplexing

EDFA Erbium-doped fiber amplifier

EVM Error vector magnitude

FEC Forward error correction

FMF Few modes fiber

FNN Feedforward neural network

GVD Group-velocity dispersion

ICP Internet content providers

ICXT Inter-core crosstalk

IM/DD Intensity modulation/direct detection

IP Internet protocol

KK Kramers-Kronig

KNN *K*-nearest neighbors

LAN Local area network

L-RNN Layer-recurrent neural network

MCF Multicore fiber

MDL Mode dependent loss

MIMO Multiple-input multiple-output

ML Machine learning

MMF Multi-mode fiber

MZM Mach-Zehnder modulator

NEP Noise equivalent power

NN Neural network

NRZ Non return-to-zero

OP Outage probability

PAM Pulse amplitude modulation

PAPR Peak-to-average-power ratio

PIN Positive-Intrinsic-Negative

PDL Polarization-dependent loss

PDM Polarization division multiplexing

PMD Polarization mode dispersion

PMP Phase matching points

PPLN Periodically poled LiNbO3

PSA Phase sensitive amplifiers

PSD Power spectral density

PSM4 Parallel single mode 4 lane

QAM Quadrature amplitude modulation

QPSK Quadrature phase shift keying

RBF-NN Radial basis function neural network

ROP Received optical power

RPS Random phase shifts

RRC Root-raised-cosine

SDM Space division multiplexing

SDN Software defined network

SMF Single-mode fiber

SNR Signal-to-noise ratio

STAXT Short-term average inter-core crosstalk

WDM Wavelength division multiplexing

List of symbols

$\boldsymbol{\alpha}$ Fiber attenuation coefficient

$\boldsymbol{\beta_0}$Propagation constant

$\boldsymbol{\beta_1}$Propagation time delay

$\boldsymbol{\beta_2}$Group velocity dispersion

$\boldsymbol{\beta_3}$ Higher order dispersion

$\Delta\boldsymbol{D_{mn}}$ Difference of dispersion parameters between cores

$\boldsymbol{\zeta}$ Power splitting between polarizations

$\boldsymbol{\eta}$ PIN efficiency

$\boldsymbol{\lambda_0}$ Operating optical wavelength

$\boldsymbol{\nu}$ Optical frequency of the input signal

$\boldsymbol{\rho}$ Roll-off factor

$\boldsymbol{\varphi}\ (\boldsymbol{t})$ Phase of the signal

$\boldsymbol{\omega}$ Angular frequency

$\overline{\boldsymbol{\beta_n}}(\boldsymbol{\omega})$Average of the intrinsic propagation constants of the two polarization directions

$\overline{\Delta\boldsymbol{\beta_{0,mn}}}$ Difference between the averages of the propagation constants at $\omega = 0$

$\overline{\boldsymbol{K}}_{\boldsymbol{nm}}$ Discrete coupling coefficient

$\boldsymbol{\phi}_{\boldsymbol{nm,k}}^{\boldsymbol{a,b}}$ Random phase shifts

$\boldsymbol{A}$ Amplitude of the optical carrier

B Signal bandwidth

c Speed of light in vacuum

$\boldsymbol{D_{\lambda_0}}$ Dispersion parameter at λ_0

$\boldsymbol{d_{mn}}$ Average walkoff parameter between the interfered core and the interfering core

$\boldsymbol{E_{in}}$ Input of the IQ modulator

$\boldsymbol{e(t)}$ Optical signal at the PIN input

$\boldsymbol{e_{icxt}(t)}$ The ICXT induced by the interfering core at the output of the interfered core

$\boldsymbol{e_m(t)}$ Optical signal transmitted at the interfering core

$\boldsymbol{e_n(t)}$ Optical signal transmitted at the interfered core

$\boldsymbol{e_{1,2}(t)}$ Electrical signal at the arms of the modulator

$\boldsymbol{F_{a,b}(\omega)}$Transfer function that models the ICXT from the interfering core

$\boldsymbol{F^{-1}}(\cdot)$ Inverse Fourier transform

$\boldsymbol{f}$ Frequency

$\boldsymbol{H(f)}$ Transfer function of the Nyquist filter

$\boldsymbol{H}(\cdot)$ Hilbert transform of *x*

$\boldsymbol{H_{CD}(\omega)}$ Transfer function of the CD filter

$\boldsymbol{H_f(\omega)}$Transfer function of the MCF propagation

$\boldsymbol{H_{RC}(f)}$ Raised cosine filter

$\boldsymbol{h}$ Planck's constant

$\boldsymbol{i(t)}$Electric current at the PIN output

$\boldsymbol{L}$ Fiber length

$\boldsymbol{M}$ Modulation order

$\boldsymbol{M_{path}}$ **N**umber of spatial paths

$\mathbf{max}(\boldsymbol{x})$ Maximum value of *x*

$\boldsymbol{N}$ Number of phase matching points

$\boldsymbol{N_{sim}}$ Number of the transmitted symbols

$\boldsymbol{n_e(t)}$ Electric noise

$\boldsymbol{q}$ Electron charge

$\boldsymbol{Rs}$ Symbol rate

$\boldsymbol{R_\lambda}$ PIN responsivity

$\boldsymbol{S}$ $\boldsymbol{(t)}$ Information-bearing signal

$\boldsymbol{S_{\lambda_0}}$ **s**lope of the dispersion parameter

$\boldsymbol{s(t)}$ The signal reconstructed by the KK algorithm

$\boldsymbol{s_i(k)}$ Coordinates of the transmitted symbols

$\boldsymbol{s_0(k)}$ Coordinates of the received symbols

$\boldsymbol{T}$ Symbol duration

$\boldsymbol{V_b}$ Bias voltage of the modulator

$\boldsymbol{V_{sv}}$ Switching voltage of the outer modulator

$\boldsymbol{v_{AC\,1,2}}$ $\boldsymbol{(t)}$ In-phase and quadrature components of the QAM signal

$\boldsymbol{v_g}$ Group velocity

$\boldsymbol{z_k}$ Coordinate of the *k*-th RPS

CHAPTER 1

Introduction

1.1 Motivation

Current optical fiber networks are reaching the so-called capacity crunch of 100 Tb/s per single core fiber [1]. Over the last years, the traffic in data centers has been increasing exponentially, demanding new cost-efficient solutions for short-reach optical communications [2]. To solve this problem, different classes of space division multiplexing (SDM) have been proposed, such as multicore fiber (MCF) and few-mode fiber (FMF). These technologies can increase the capacity by increasing the number of spatial paths, where each one can be used to transmit a different signal. Different types of receivers, such as Stokes vector detection, differential coherent detection and Kramers-Kronig (KK) receivers have been proposed to increase spectral efficiency [1]. MCF improves the capacity as well as the interconnects density and KK receivers will allow the use of direct detection (DD) which brings lower power consumption and lower cost for higher bit rates [3]. The performance of these technologies can be highly affected by the inter-core crosstalk (ICXT) caused by the MCF.

SDM has been proposed 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. This dissertation is based on MCFs, where N independent cores provide a N-fold capacity increase when compared with standard single-mode fiber used typically in current networks. 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 data-center communications.

Due to cost purposes, short-reach MCF-based networks should employ direct-detection receivers. However, these receivers lead to nonlinear impairments that can severely limit the achievable capacity and reach. Thus, network solutions based on direct-detection followed by digital signal processing (DSP) based on the KK technique have been proposed for performance improvement and complexity reduction [4]. With the KK technique, linearization of the receiver is

attained, and machine learning techniques can be used to optimize the end-to-end performance of these next generation MCF short-reach networks. Different classes of neural networks (NN), such as radial basis function NNs, recurrent NNs, and convolutional NNs have been already employed and validated in different short-reach scenarios [5, 6]. In this work, DSP equalization based on low-complexity and memoryless machine learning algorithms are implemented to improve the end-to-end performance of next generation short-reach networks employing multicore fibers and direct detection.

1.2 Objectives

The general purpose of this work is to unlock the capacity supported by next generation optical access networks. This is accomplished by proposing a direct-detection SDM short-reach network with a per optical channel capacity of >200 Gb/s using machine learning for end-to-end optimization. Particularly, the following objectives are pursued:

i) To integrate a software platform for simulation of short-reach >200 Gb/s SDM optical fiber networks employing direct-detection and KK receiver linearization.

ii) To identify the main advantages and challenges of different classes of machine learning techniques.

iii) To implement a machine learning technique in the simulator for end-to-end performance optimization in short-reach >200 Gb/s SDM optical fiber networks employing direct-detection.

iv) To assess the impact of the random nature of the inter-core crosstalk (ICXT) along time on the performance of the proposed machine learning equalizer.

1.3 Dissertation structure

The dissertation is structured as follows:

- Chapter 2 presents the literature review in which the fundamentals concepts such as data center, space division multiplexing giving particular attention to homogenous multicore fiber, KK receivers and machine learning are addressed.
- In chapter 3, the system model used for this study is presented and described in detail. The concept of dual-polarization discrete change model is presented to describe the ICXT induced by the MCF. Also, the system performance without the use of ML is assessed and

the parameters used for the system are presented with the corresponding reasoning for their usage.

- Chapter 4 presents assessment of the system performance when impaired by the random ICXT and then an in-depth description of the ML techniques employed is presented. The system performance when employing ML is assessed compared with the system without the use of ML.
- Chapter 5 depicts the main conclusions and suggestions for future work.

1.4 Main contributions

This work presents the following contributions:

- Implementation of ML techniques to mitigate the effects of the random ICXT induced by MCF.
- Demonstration that ANNs can significantly improve the system performance.
- Confirmation that memoryless and simple ANNs are suitable to mitigate the effects of the random ICXT in systems with skew × symbol rate <<1.
- The work performed in the scope of this dissertation resulted in an article published at the Photonics journal (see the appendix). This article summarizes the proposed system and its outcomes, described in chapters 3 and 4 of this dissertation, respectively.
- The work performed in the scope of this dissertation resulted in a presentation at the IX Seminar in Multi-Gigabit Optical Networks, June 3rd, 2022. The presentation was performed by the supervisor (PhD Tiago Manuel Ferreira Alves), where the proposed system and its outcomes were summarized.
- Next generation >200 Gb/s multicore fiber short-reach networks employing machine learning.

CHAPTER 2

Literature review

2.1 Overall review on optical fiber communications

The network traffic grows every year due to the demands of the new technologies and the dependencies of the users on the internet services. Data centers experienced an IP (internet protocol) traffic growth of 26,8% from 2015 to 2020 [1], while the optical interface rates have growth at 20%, and the electronic technology rates had a growth of 10% [2]. This shows an exponential growth of the traffic demand, while the modern technologies are reaching its limits in terms of bandwidth and spectral efficiency [1, 2, 7].

The invention of erbium-doped fiber amplifier (EDFA) promoted the use of wavelength division multiplexing (WDM) systems which became more cost-efficient [8, 9]. Through the years, the use of WDM in optical fiber communications and then the DWDM (dense WDM), managed to keep up the traffic demand, because it allowed the transmission of multiple optical channels in one fiber, which improves the capacity [2]. The use of coherent detection enabled the increase of the spectral efficiency by increasing the number of bits transmitted in one symbol, up to 64 bits on 64-QAM (quadrature amplitude modulation) [10]. Combined with coherent detection polarization division multiplexing (PDM) still enable to further increase the capacity [2]. Coherent detection allows to digitally compensate the chromatic dispersion (CD) and to equalize linear transmission impairments, such as group-velocity dispersion (GVD) and polarization-mode dispersion (PMD) of transmission fibers via digital signal processing [8, 11]. The main disadvantage of this technology is the complexity and the power consumption as it demands a local oscillator at the receiver and several photodetectors to detect the in-phase and quadrature components of the *x* and *y* polarizations.

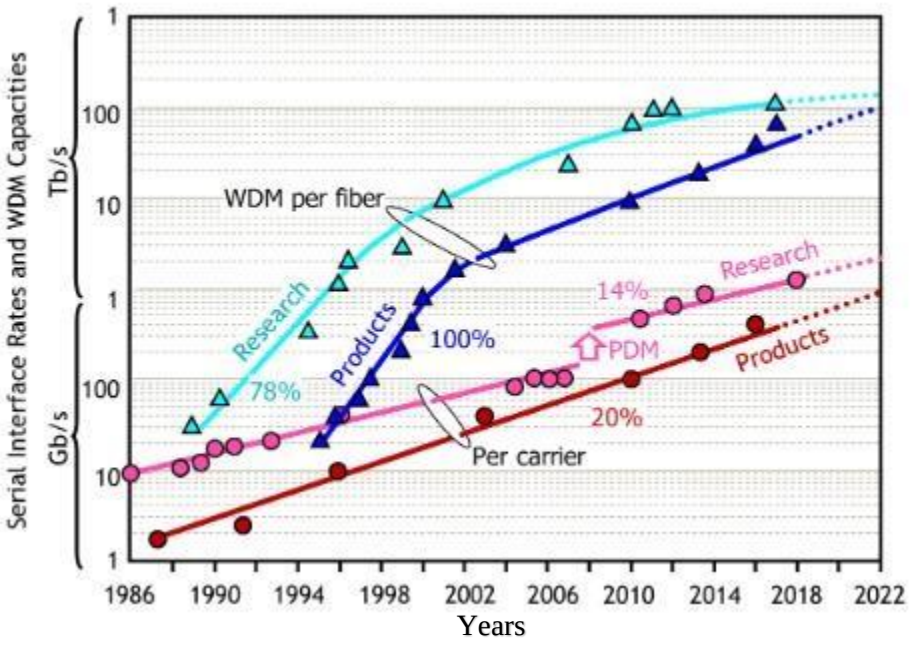


Figure 1 - Comparison of products and research records in terms of per-carrier interface rates and WDM capacities [2].

2.2 Limitation of optical fiber communication systems

As the demand increases, the need of scaling the network capacity also increases. The global traffic is exponentially growing at approximately 45% per year [2]. As shown in figure 1 there is a growth of approximately 20% per year on the optics capacity, and the slow growth is demanding a need of research on optical areas.

To identify how the capacity of the optical systems can be increased, we can analyze the information theory of Shannon [2]. The Shannon's theory relies on the study of five physical properties available for modulation and multiplexing of electromagnetic waves (time, frequency, quadrature, polarization and space), and with these parameters, the maximum capacity reliable to communicate over an equivalent additive white Gaussian noise channel, is given by:

$$C = M_{path} \times B \times 2 \times \log_2(1 + SNR), \tag{2.1}$$

where B is the signal bandwidth, the factor 2 is related to PDM, M_{path} is the number of spatial paths and SNR is the signal-to-noise ratio.

2.2.1 Bandwidth limitations

As shown in (2.1), increasing the bandwidth is a way to increase the capacity. For many years it was believed that optical communications had infinite bandwidth, so the solution was a matter of increasing the number of WDM channels or increasing the bitrate per channel. The increasing of bandwidth starts being problematic due to the complexity of the optical equipment such as lasers, receivers and optical amplifiers. The complexity of optical components grows with their bandwidth, which can be costly and not economically efficient. One of the main limitations on the bandwidth scaling, is the optical amplifiers, mostly EDFAs [12], which their gains is on the C-band (1550 nm to 1565 nm) and L-band (1565 nm to 1625 nm) [2].

Building optical systems with relatively large bandwidth can be a complex and costly task, since increasing the bandwidth relies on the improvement of individual components, such as optical amplifiers, lasers and optical filters. Hence, it becomes clear that the expansion of the bandwidth has limited use, which is not attractive on research point of view.

2.2.2 SNR limitation

Another way of scaling the capacity shown in (2.1) is increasing the SNR. Despite all the efforts that can be made for increasing the SNR, this will lead to a logarithmic return on the capacity increase. Higher SNR can be achieved through lower-noisily amplification or lower-nonlinearity fibers [2]. In [13], ultra-low-noise amplification of an analog optical signal with an inline phase-sensitive frequency-non-degenerate fiber parametric amplifier was demonstrated and achieved a figure noise close to the Raman limit of 0.45 dB. Other types of low noise phase sensitive amplifiers (PSA) have been proposed. For instance, in [14] the authors reported a noise figure below 3 dB with a 20 dB gain in a PSA using directly bonded periodically poled LiNbO3 (PPLN), and in [15], an average noise figure of 1.7 dB at a 23.2 dB gain was achieved with a PSA using an phase-locked pump. On the other hand, in [16] a low-loss and low-nonlinearity pure-silica-core fiber for C and L-band was proposed and achieved a low loss of 0.15 dB/km and an effective area (Aeff) of 130 μm^2.

2.2.3 Space division multiplexing

The last possible way of increasing the capacity of the optical system is by increasing the number of spatial paths. Unlike the previous solutions, SDM has been proposed as a suitable solution for the demand as the capacity increases with the increase of the number of special paths. The simplest use of SDM in optical communications is the use of multiple fiber cables and for the last years is has been the solution for the capacity demand [17]. However, this approach is not advantageous as it increases the power consumption and cost.

Lately new types of SDM such as FMF and MCF have been proposed to increase the capacity of optical system as well as the fiber density. These two technologies allow the transmission of several spatial channels in just one fiber cable [7, 18]. Nevertheless, these technologies present some disadvantages which will be addressed in further sections.

2.3 Data center

A traditional data center (DC) architecture is based on 3 tiers (presentation, application and data). This topology is efficient on the called “north-south” traffic, which is related to client-server applications [19]. Although, in recent years, the internet content providers (ICPs) start hosting a large amount of information, mainly due to video streaming and cloud computing, this traffic growth is mostly internal, between servers or even between data centers. The traditional architecture is not well equipped to handle this type of traffic. Due to virtualization and cloud computing, the solution for handling the increasing traffic is to switch from a three-tier to a two-tier architecture, which is more suitable to deal with “east-west” traffic, related with the communication between servers or closed data centers (less than 100 km) [1, 20].

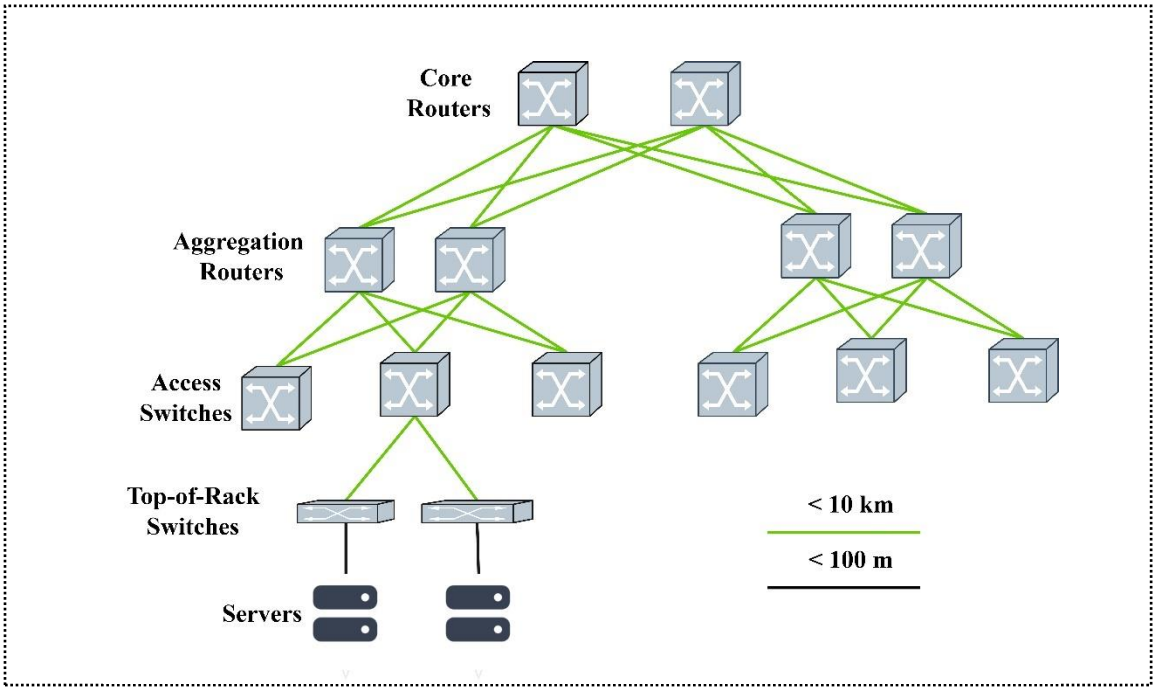


Figure 2 - A traditional three-tier data center architecture.

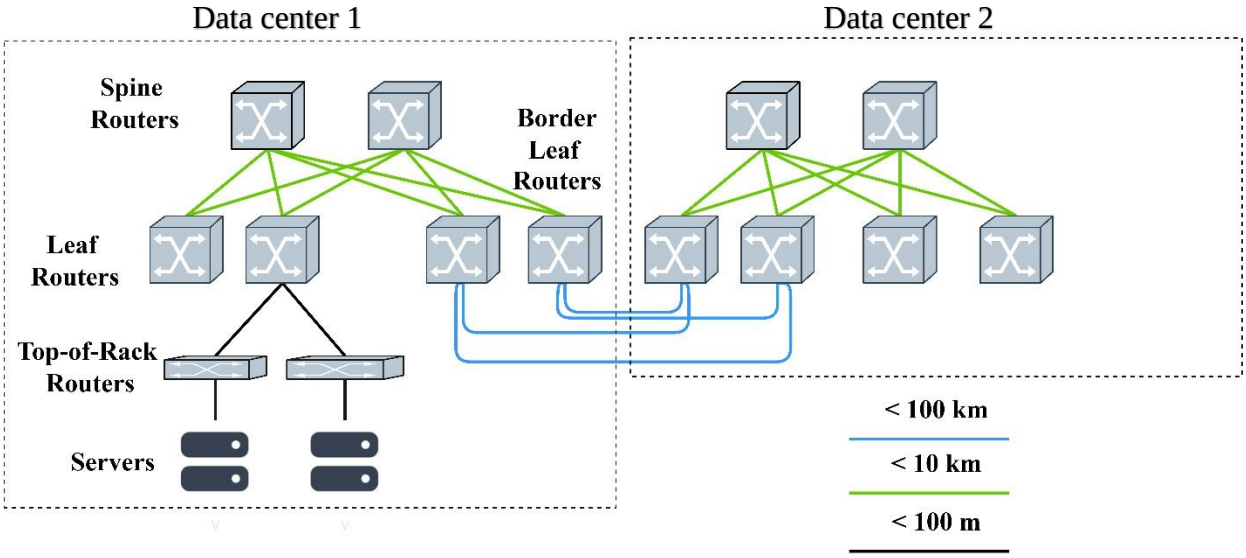


Figure 3 - Two-tier data center topology.

Figure 2 shows a diagram for three-tier architecture. In this scheme, the servers are connected to the access switches which are connected to the aggregation routers and the core routers. The traffic

on this type of architectures goes from the core routers to the servers which is efficient to the north-south traffic, but inefficient to connections between different data centers. A diagram showing the two-tier architecture is depicted in figure 3. As can be observed, the data centers are connected by the border leaf switches and it allows a more efficient solution to transmit data between data centers.

The intra-data center links reach up to 10 km and use wavelength near the second window of optical fiber communications (1310 nm) to minimize the dispersion. With this, there is no need to compensate the dispersion and, due to the short-reach of the link, there is no need of amplification.

Inter-data centers (up to 100 km), use wavelength near 1550 nm due to the band of EDFAs, thus amplification is needed due to the long fiber distance. With this wavelength and higher distance, the residual dispersion needs to be compensated using digital compensation or dispersion compensating fiber (DCF) [1].

Nowadays data centers are using intensity modulation with on/off keying or *N*-Pam (pulse amplitude modulation) with DD due to the simplicity and lower power consumption of the receivers. For the multiplexing, data center links use coarse WDM with wavelength spacing of 20 nm or LAN-WDM with wavelength spacing of 4.5 nm to avoid the power-hungry [1].

2.4 Space division multiplexing

As previously mentioned, the optical systems are reaching the capacity crunch, and, as showed in (2.1), the parameter that remains to be explored is the number of parallel paths. To address this question, space division multiplexing (SDM) was introduced [2]. SDM is the use of multiple spatial paths to transmit information. By increasing the number of paths, the capacity increases in the same ratio.

Nowadays, the data centers use SDM in the form of multiple fiber cables. For example, data centers use 10 parallel links of 10 Gb/s or 4 of 25 Gb/s for 100 Gb/s links [17]. This approach is not scalable for higher bit rates, since it imposes high cost, complexity and power consumption.

To overcome this problem, two main SDM approaches can be followed, multicore fiber (MCF) and multimode fiber (MMF) [7, 18]. MCF consists in one fiber with multiple cores for information transmission. These cores can interfere with each other depending on the structure, which may cause high or low crosstalk [18]. MCF transmission is almost equivalent to single-mode fibers

(SMF), where the difference is the intercore crosstalk (ICXT). This means that each core can be used as a separate WDM system and the capacity increase is proportional to the number of cores. These types of approaches are well suited for data centers with low distance interconnections, because there is no need for amplification, which can be difficult due to multiple cores and dispersion compensation in cases where DD are used. MMF uses multiple modes of propagation for information transmission. In this context, research is nowadays focused on a type of MMF known as few-mode fibers (FMF). FMF allows the transmission of channels using few orthogonal modes. These transmissions require multiple-input multiple-output (MIMO) equalization due to the mode mixing [18]. In FMFs, differential mode delay (DMD) and mode dependent loss (MDL) can be a problem and demand more complex MIMO receivers [21]. DMD can be briefly described as a spread of the signal in time domain, and MDL consists of different loss in each mode, which demands multi-mode optical amplifier to amplify each mode separately [21].

The main goal on SDM is to combine the two proposed technologies (MCF and FMF), thus it can drastically increase the capacity, because each core of the fiber can be used as a multi-mode core, and each mode can transmit a different WDM signal, as shown in [21].

2.4.1 Multicore fiber

In recent years, multicore fibers (MCF) have shown to be a solution to overcome the capacity limitations in short-reach systems [7]. Figure 4 depicts an illustration of the cross-section of a two core MCF.

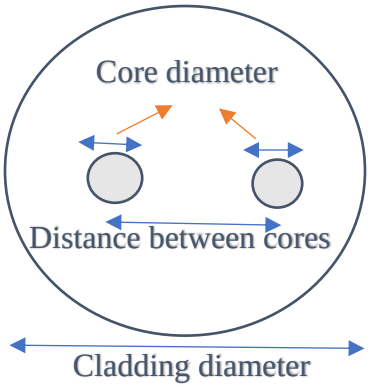


Figure 4 - MCF cross-section illustration.

There are two types of MCF, weakly-coupled MCF and strongly-coupled MCF. The weakly-coupled MCF consists of a fiber with multiple cores distributed in a less dense distribution. The inter core crosstalk (ICXT) decreases by decreasing the distance between the cores, therefore it will not allow as many cores as strongly-coupled MCF. Strongly-coupled MCFs have a higher core density, that can provide more capacity to the fiber, although the lower distance between the cores increase the ICXT and may lead to the need of MIMO DSP for the strong mode mixing [18]. In [22, 23], MIMO-based signal processing was proposed on 56 Gb/s and 112 Gb/s PDM-QPSK for strongly-coupled three cores fiber and both achieved a low penalty in presence of crosstalk as large as -4 dB.

MCFs can be also characterized as homogeneous MCF and heterogeneous MCF. Homogeneous MCFs have cores with the same characteristics, and this can result on simpler systems without the need of high complexity MIMO equalization. In [24], a spectral efficient modulation was demonstrated with a 22-cores homogeneous MCF using PMD-64-QAM at 24.5 Gbaud assuming a 20% forward error correction (FEC) overhead and a BER of 2.7×10^{-2}. Heterogeneous MCF uses different propagation constants between the cores which can be used as a form to suppress the ICXT, although the delay difference between cores demands the use of high complexity MIMO equalization [18].

2.4.2 Advantages and disadvantages of MCF

The most important advantage of MCF is increasing the capacity in the same ratio as the number of cores. Every core can be used as a separate WDM system, which increases drastically the capacity of the data centers links [7, 18]. Another way of increasing the capacity is the combination of MCF with FMF, although it has a disadvantage which is the increase of system complexity. Other advantage of MCF is the reduction of the number of fiber cables. The number of cores can represent a reduction of the same number of cables and keeps the same capacity. For example, fiber transmission links in DCs as PSM4 (Parallel Single Mode 4 lane) are good candidates of MCF applications [25]. PSM4 link can be realized by 1 strand of 8 cores fiber or 2 strands of 4 cores fibers, which means that 8 SMFs ribbon cable and 8 connectors can be replaced by one or two MCFs thinner cable and simplex/duplex single-MCF connectors. These simplifications of the cable and connectors could reduce the cost [7, 26].

One disadvantage of MCFs is the complexity comparing to SMFs, because MCFs demand a restructuring of the systems. For example, it demands a scalable manufacturing of fan-in/fan-out to connect the MCFs to the standards devices [7, 26]. The main disadvantage of MCF is the ICXT [7]. The ICXT is a random process whose statistical properties depends on time, length of the fiber, number cores and distance between cores [27]. The ICXT affects the system performance as one signal transmitted in one core (interfered core) is directly affected by the signal transmitted at the other cores (interfering cores).

2.4.3 Studies on ICXT

In homogeneous weakly-coupled MCF, the cores have similar properties and similar propagation time which increases the coupling between cores. The strong coupling leads to a high ICXT that is more aggravated by the fiber length [27].

The ICXT is the main challenge of MCFs and it can drastically affect the system performance [27]. Research has shown that the ICXT levels have a random longitudinal and time variation [27]. The longitudinal part can be caused by the imperfections of the fiber fabrications and the time variation can be caused by changes of the environment around the fiber. It is a stochastic process and its impact on the performance is significantly dependent on modulation format, data-rate and temporal skew between cores [27]. The effects of ICXT become more significant for higher core count and shorter core-to-core distance [28, 29]. This may lead to significant fluctuations over time, and it can be more relevant as high ICXT are observed over several minutes or even hours leading to service shutdown or outage over large time periods [29] – [31]. As the ICXT impacts the service quality, studies have been made to characterize the ICXT in different aspects. For example, in [32] the theoretical characterization of the decorrelation bandwidth of ICXT in weakly-coupled MCF was performed, and it was validated experimentally that the decorrelation bandwidth corresponding to uncorrelated ICXT is approximately given by the inverse of the skew between cores. It has been established that when the signal bandwidth is broader than the inverse of the skew, the ICXT can be treated as a Gaussian noise, and when the signal bandwidth is narrower than the inverse of the skew, the ICXT should be treated as a static component coupled to the signal [32, 33].

Different techniques for ICXT suppression have been proposed. One of these techniques is the trench-assisted MCF, where each core has a low-index trench layer [18]. This technique aims to decrease the coupling between the cores. Another technique is the heterogeneous MCF which has different intrinsic index between the adjacent cores. The index difference causes different propagation constants, and it contributes to decrease the coupling between cores [18].

Table 1 shows some results of the ICXT studies on the last years. These results prove that MCFs can be used while achieving low levels of crosstalk.

Table 1- Crosstalk achieved employing MCF in recent years [18].

Number of cores	Cladding diameter [μm]	100 km worst crosstalk [dB] at 1550 nm
7	186.5	-31
12	225	-42
19	200	-14.3
19	220	-36.8
30	228	-42

2.4.4 MCF in data centers

The traffic growth caused by the increasing demands from clouds services, is requiring a scalable cost-effective infrastructure for data center network, and simply increasing the number of optical fiber links may be costly. SDM has been a candidate to achieve a cost-efficient scaling. In recent years some studies have been made on MCF in short-reach links mainly on data centers. In [26], a study was made on MCF on an economically point of view to evaluate the possibilities of MCF applications on data center links. In this study, some MCFs compatible with $250\mu m$ diameter coating standards was presented, along with their required fan-in/fan-out dimensions (2D and 3D). In [7], a penalty-free BER down to 10^{-12} was presented from system tests on 200 meters of 8-core MCF and up to 2 km in dual-core MCF. It was shown that the 8-core MCF is compatible with 100 Gb/s PSM4 silicon photonic systems that use surface coupling with four cores allocated to

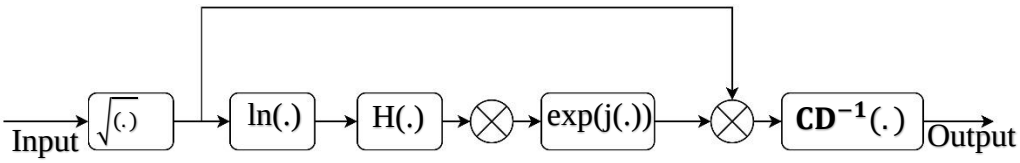
transmit in one direction and four cores allocated to transmit signals on the other direction. The 4-core MCF can be used with surface or edge-coupled 100Gb/s PSM4 systems where the transmitter and receiver are in different chips or in separate location on the same chip. And lastly, the 2-core MCF can be used on WDM system consisting of four 25 Gb/s WDM signals transmitted on one core and four 25 Gb/s signals received on the other core.

It has been demonstrated the suitability of MCF in data centers by achieving relatively high bit rates. For example, in [34], it was possible to demonstrate an achievement of 53.3 Tb/s capacity on a 7 spatial channels PDM-QPSK system and on an 8 spatial channel PDM-8PSK with a capacity of 83.3 Tb/s. In [35], the performance of 24 WDM channels, each with a 320 Gb/s PDM-16-QAM signal has been studied for a 7-core MCF in the presence of ICXT, and it was concluded that a 10 km link can support 50 Tb/s when the ICXT is lower than -18.5dB, and over 10 km when the laser linewidth is below 5 MHz.

2.5 Kramers-Kronig receiver

As mentioned in section 2.1, coherent systems are optimal from a standpoint of spectral efficiency, as they allow the encoding in both quadratures and polarization of the electric field. While coherent systems are a solution to long-reach links, the cost can be a problem to short-reach links [36]. One of the solutions proposed to achieve a similar performance, is the Kramers-Kronig (KK) receiver. They realize a small size, low cost and simple structed systems [3, 4, 37, 38].

These receivers are based on the Kramers-Kronig relation, and they can retrieve the complex signal from the photocurrent, if the minimum phase condition is ensured. The phase of the signal can be reconstructed by applying Hilbert transform to the logarithm of the current [4]. This kind of advanced DD receiver enables the digital compensation of chromatic dispersion on direct detection systems. Another advantage of KK receivers is the cancelation of the signal-to-signal beat interference, which is a nonlinearity induced by the square law detection that may cause significant performance degradation [39].



H(.) – Hilbert transform.

CD^{-1}– Chromatic dispersion compensation.

Figure 5 - KK receiver scheme.

In figure 5, the structure of an ideal KK receiver is shown as this work considers an ideal KK receiver to focus the attention on the study of ICXT effects and to evaluate the performance of the machine learning algorithms to mitigate the impact of the ICXT. KK receivers have been successfully used in optical communications. For example, in [3] it was possible to achieve a single channel of 215 Gb/s over 200 km of fiber using KK receiver. On the same study, it was possible to achieve a 1.72 Tb/s WDM transmission over 200 km of fiber.

The KK receiver has also been studied in DC scenarios as shown in [37] where it was proven to be a feasible option to achieve the cost-efficiency demanded by DC links.

In [40] KK was used with SDM. It was possible to transmit 3663 channels (37 SDM and 99 WDM) with reliable error-free performance achieving 909.5 Tb/s.

2.6 Machine learning

Artificial intelligence (AI) is a field of computer science where computer systems are trained to perform tasks that require intelligence. By employing such cognitive process, the computer is expected to store knowledge, use the acquired knowledge to solve a proposed problem and finally acquire new knowledge during the active phase to improve the performance on solving such problems [41].

Machine learning (ML) is a type of AI that focus on learning patterns by acquiring a large amount of data through different algorithms in which enables it to perform tasks such as prediction, classification and decision. ML learning has been attracting the attention in a wide range of areas such as health, telecommunications and image processing. In optical communications, ML has also been proposed and studied to improve the system performance by mitigating signal distortions or acting directly in the network configuration by performing dynamic resource allocation according

to the real time state of the networks enabling the maximum system capacity through software defined network (SDN) [42]. For example, in optical communications, ML can be used to compensate nonlinearities such as square-law detection in direct detection, for converting the nonlinear effects into linear ones [43]. In this work, we propose the use ML techniques to mitigate the impact of the random ICXT on the system performance. More specifically, ML algorithms will be proposed for performing tasks such as classification of the used QAM signal, reconstruction of the ICXT-impaired signal to perform static standard symbol decision and, lastly, direct symbol decision by learning patterns on the affected signal.

Machine learning can be classified in supervised learning, unsupervised learning and reinforcement learning (subject addressed in the following sections).

2.6.1 Supervised learning

In supervised learning, the algorithm is provided with a large amount of input samples whose outputs are known in advance (ground-truth), and the main goal is to find the relationship between the input and the output. A training data set comprises multiples samples of the input and the corresponding output values and several test and validation stages are performed until the algorithm reach its optimum performance [44]. One way to apply supervised learning on optics communications is for example, to train an algorithm using a dataset that consists the characteristics of the link (path, wavelengths, modulation) as input and the corresponding BER, then it can estimate the approximate BER in correspondence to new inputs [41]. Another way to use supervised learning is to classify and to perform the symbol decision on the signal impaired by the random ICXT.

2.6.1.1 *K*-nearest neighbors

K-nearest neighbors (KNN) algorithm is based on the distance between the new data and its closest *k* neighbors. In the training phase the algorithm stores the vector of the features and the corresponding known classes. On the classification phase, a new input is classified into the group of data with the greatest frequency among its *k* closest neighbors. Thus, in order to get better performance, the *k* parameter has to be well chosen [41, 44]. One application of KNN is the

classification of symbols on decisions, for example, it can decide boundaries in *M*-QAM systems and classify each input as a symbol on the constellations as shown in [45].

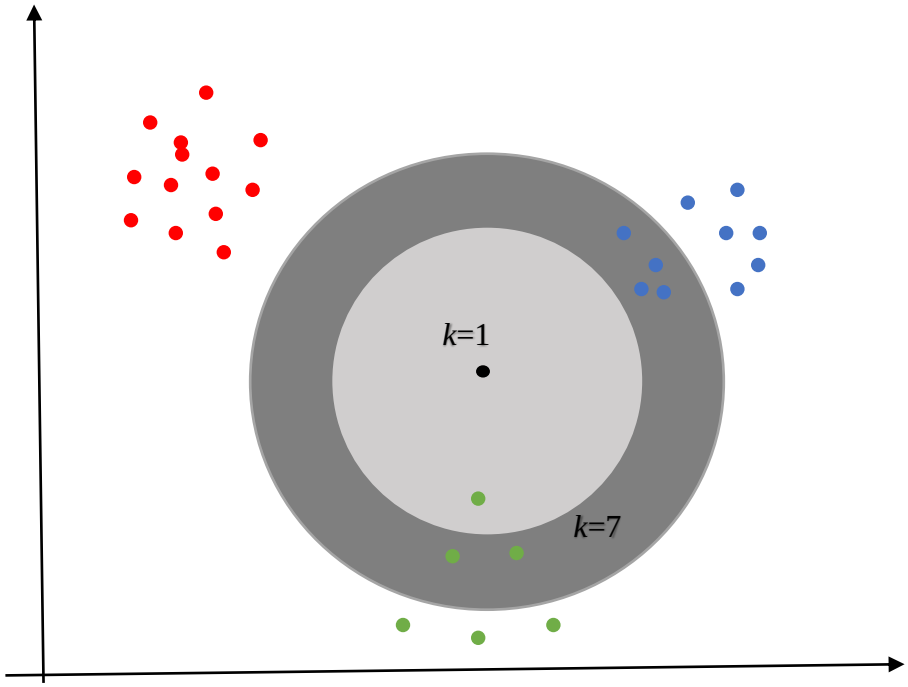


Figure 6 - Example of KNN with *k*=1 and *k*=7.

Figure 6 shows the importance of choosing the proper *k* value, since different values of *k* can lead to different classification outcomes, as illustrated in the figure. In the case of *k*=1 the input would be classified to the green class and with *k*=7, it would be classified as blue class.

2.6.1.2 Artificial neural network

Artificial Neural Network (ANN) is a class of machine learning algorithms which is biologically inspired on the human brain.

An ANN consists of multiple units (or nodes) called neurons, which are organized into layers. A typical Artificial Neural Network comprises an input layer, a set of intermediate layers, called hidden layers, and an output layer, whose result is an estimate or a prediction. Each neuron at a given layer is usually connected to all neurons of the previous layer. Each connection between neurons has an associated weight and each neuron computes a weighted sum of the values

presented at its input connections. The output of a neuron is the result of an activation function applied to the weighted (and biased) sum of its inputs. The weights of the interconnection between neurons are determined during a training process using back propagation algorithm [41, 44]. In the ANNs these connections are simpler and normally designed to solve a single problem. The complexity of the ANN is based on the complexity of the proposed problem to be solved.

In the context of this work, where the objective is to mitigate the effects of the ICXT, simple and memoryless ANN can be used as the ICXT can be considered as a low-complexity problem if certain conditions are assured (topic explored in chapter 4). Since ANN can be used in a wide range of areas to solve numerous problems, there are different types of ANN which configurations can be simple, such as feedforward neural network or more complex ANNs, such deep learning-based networks, which can have several layers and can require allocated memory (*e.g.*, recurrent neural network). These more advanced NN can be used for predicting different strategies for routing and spectrum assignments for elastic optical networks [46].

In optical communications ANNs can be used to perform equalizations to compensate the impairments caused by fiber dispersion, bandwidth limitations, laser chirp, noise and square-law in the case of IM/DD. They are well suited to perform optical performance monitoring due to their ability to learn the complex mapping between samples or extracted features from the symbols and optical fiber channel parameters, such as OSNR, PMD, Polarization-dependent loss (PDL) and CD [41].

For example, in [43], it was demonstrated that a NN-based equalizer was able to accurately reconstruct the received signal affected by inter symbol interference in a 4 km 32 Gbaud IM/DD system with 8-PAM. This demonstration showed that ANNs can be well suited for intra data centers due the short distance and the low complexity.

In SDM systems, machine learning techniques were proposed to support the design of crosstalk-aware schemes used for resources allocation [47] or to mitigate the impact of the crosstalk power between mode groups in mode-multiplexed M-quadrature amplitude modulation (QAM) OFDM-IM-DD systems [48]. NNs were also used to speed up coating loss estimation in heterogeneous trench-assisted MCF design [49].

2.6.2 Unsupervised learning

Unlike supervised learning, unsupervised leaning is based on creating classification of the inputs without knowing the right answer. It aims to classify the data according to the characteristics or relationships between them by clustering into groups. This technique can be well used in cases where the correct outputs are not known in advance.

One unsupervised learning algorithm is *k*-means clustering. It focuses on splitting the input into *k* clusters according to dissimilarity metrics and it creates groups centered on the mean of all the samples of each one [41, 44]. In case of high bitrates links, high modulations indexes are used and the smallest variations on the system can highly affect the system. *k*-means clustering can be used as an adaptive machine learning-based nonsymmetrical decision technique to mitigate the time varying impairments as demonstrated in [6].

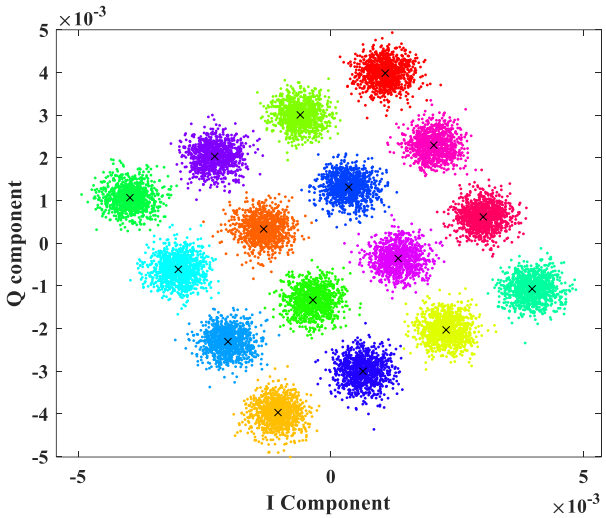


Figure 7- *k*-means clustering classification for a 16-QAM signal.

Figure 7 depicts an example of a *k*-means clustering classification of 16-QAM signal.

2.6.3 Reinforcement learning

In reinforcement leaning, an autonomous entity known as the agent learns by evaluating the feedback of its actions. The feedback is often referred as the reward [41]. The goal is to maximize the performance in long term, thus it evaluates the consequences of its actions on the future. In

reinforcement learning, there is not an explicit learning phase since the algorithm updates are based on the agent's interaction with its own environment [44]. One application of reinforcement learning in optic networks is, for example, self-configuration, including resource allocation and service reconfiguration [44].

2.6.4 Studies with machine learning on IM/DD systems

To minimize the nonlinearity effects, studies have been made using machine leaning on systems with intensity modulation/direct detection (IM/DD) [50]. For example, an experimental study with end-to-end deep learning design proves the great potential of ANN based transceivers. In this experiment, it was possible to transmit 42 Gb/s beyond 40 km with BERs below the 6.7% HD-FEC threshold. The system was based in IM/DD with PAM2/PAM4 modulation and conventional receiver equalization [50].

Other studies have been developed using different types of ANNs, such as feedforward neural network (FNN), radial basis function neural network (RBF-NN), auto-regressive recurrent neural network (AR-RNN) and layer-recurrent neural network (L-RNN). These studies were made on a 50 Gb/s PAM4 direct detection link. Among all these based equalizer techniques with the same number of input and hidden neurons, AR-RNN achieved the best performance while FNN showed the lowest computational complexity. In all of them, only few tens of multiplications are needed to achieve the FEC threshold, showing high potential of use on optical systems [5]. In [51] it was experimentally demonstrated a 100 Gb/s PAM4 short-reach optical link via a 15 km fiber using only post-equalization. A cascade RNN-based equalizer was proposed to effectively compensate signal distortion, showing better BER performance than FNN or RNN-based equalizers. With the aid of cascade RNN, BER lower than 7% FEC threshold is achieved when the receiver sensitivity is above 5 dBm.

In [52], NN and KNN were tested on a 56 Gbaud PAM4 system. In both cases achieved a good performance with the NN equalizer demanding a higher complexity.

Besides NN, other ML techniques have been studied on optical communications. In [53] *k*-means was proposed as a low-cost ML technique to monitor the changes in the fiber environment via monitoring changes in the state of light polarization without the utilization of methods based on back-scattered light.

2.7 Summary

In this review, some concepts related to optical communications have been introduced. In section 2.1 an overall of optical communications has been presented and the exponential growth of the traffic has been addressed. Then the limitations of optical communications have been presented in section 2.2 where Shannon's theory of information has been presented to evaluate the possibilities of capacity scaling. Furthermore, in section 2.3, current data center architecture has been introduced along with its main specifications. In section 2.4 SDM has been introduced as a solution for the capacity scaling according to Shannon's theory. Section 2.5 presented the KK receivers as a solution to solve the nonlinearities of DD to achieve a low-cost system and enabling the digital chromatic compensation, and finally, in section 2.6, machine learning techniques have been introduced to mitigate the nonlinearities as well the ICXT introduced by the MCFs.

CHAPTER 3

System model and performance assessment

In this chapter, the MCF short-reach system employing the KK receiver is presented and the system performance in absence of ICXT is assessed. The system model is described in section 3.1, where the transmitter block is described in detail, the dual polarization discrete change model (DP-DCM) is introduced to describe the behavior of the random ICXT, and the receiver block is detailed including the description of the KK receiver. In section 3.2, the figures of merit used for system performance are presented. In section 3.3, the performance of the baseline system is assessed.

3.1 System model

Figure 8 depicts a block diagram of the system setup considered in this work for assessing the performance of machine learning algorithms on mitigating the impact of the ICXT. The system is composed by the optical transmitters, which generates QAM root-raised-cosine (RRC) shaping. Then, the signals at the output of the transmitter are launched into two different cores of a MCF: i) core *n*, the interfered core, and ii) core *m*, the interfering core which induces ICXT in core *n*. Then, the optical receiver includes a PIN photodetector, an electrical amplifier (which induces thermal noise), the RRC filter, the KK algorithm where the compensation of the dispersion is employed and the ML block to mitigate the impact of the ICXT. Finally, the bit error ratio (BER) is estimated using Monte Carlo simulation.

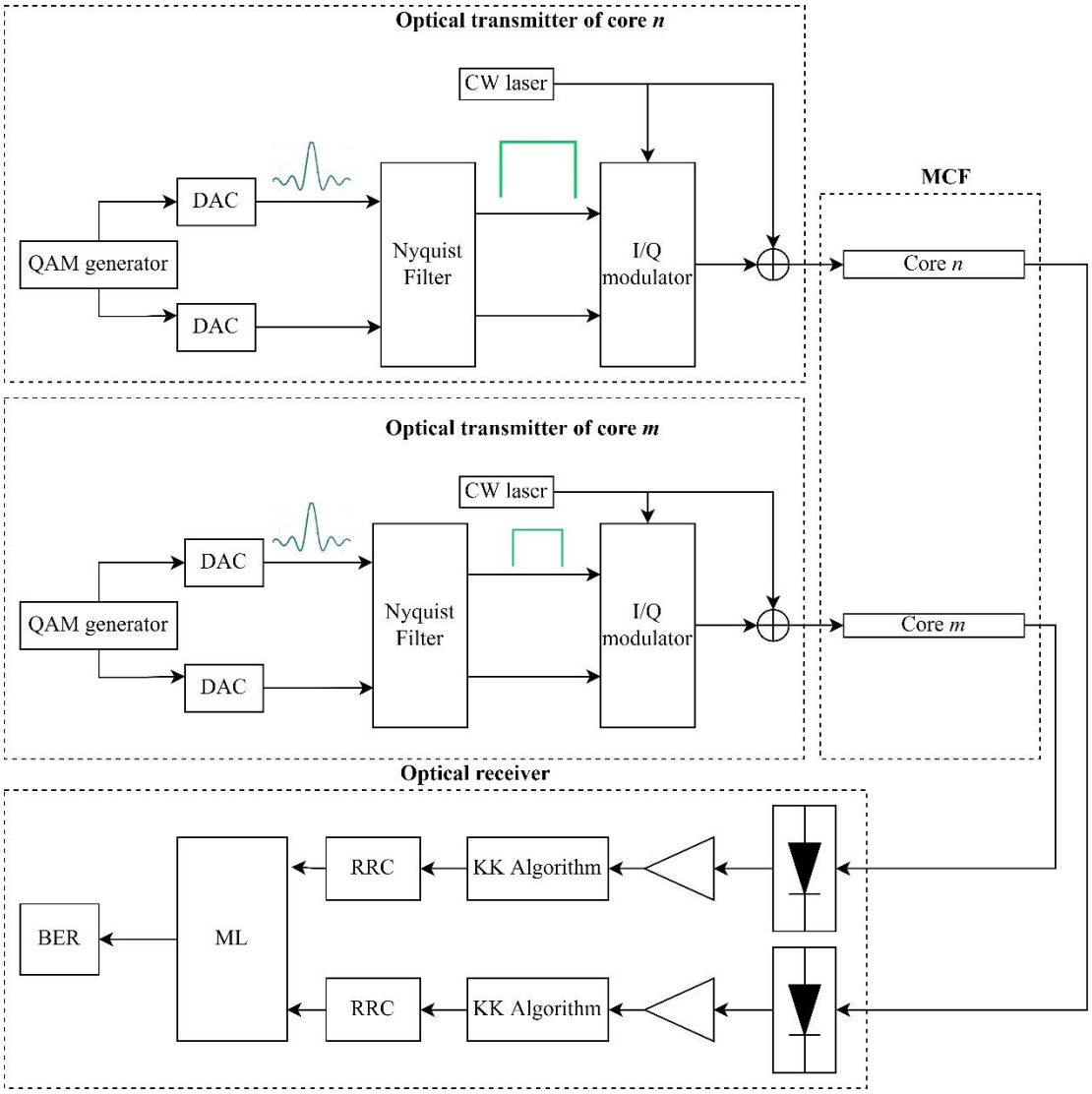


Figure 8 - Schematic diagram of the short-reach MCF based system employing KK receivers.

3.1.1 Optical transmitter

The optical transmitter is responsible by converting the information signal from the electrical to the optical domain. First, a QAM Nyquist signal with a roll-off factor of 5% is generated using distinct random sequences for both components (in-phase and quadrature) of the signal. The modulator is a dual parallel Mach-Zehnder modulator (DPMZM) with the ability to modulate the *I* and *Q* components of the electrical field. This is achieved by biasing the inner MZMs at the null bias point and the outer MZM at the quadrature bias point.

To obtain an almost rectangular spectrum of the transmitted signal, the signal need to be filtered by a Nyquist filter, which transforms a non return-to-zero (NRZ) pulse into a RRC pulse. The transfer function of the Nyquist filter is obtained as

$$H(f) = \sqrt{H_{RC}(f)}\,\frac{\pi f T}{\sin(\pi f T)}, \tag{3.2}$$

where $H_{RC}(f)$ is defined as

$$H_{RC}(f) = \begin{cases} T, & 0 \le |f| \le \frac{1-\rho}{2T} \\ \frac{T}{2}\left\{1 + \cos\left[\frac{\pi T}{\rho}\left(|f| - \frac{1-\rho}{2T}\right)\right]\right\}, & \frac{1-\rho}{2T} \le |f| \le \frac{1+\rho}{2T} \\ 0, & |f| > \frac{1+\rho}{2T} \end{cases}, \tag{3.3}$$

where T is the symbol duration, f is the frequency and ρ is the roll off factor.

The QAM modulation uses two signals modulated in amplitude (I and Q components). The electrical field of the QAM signal after the modulator is defined as

$$e_{out}(t) = \frac{E_{in}}{2}\left[\exp\left(j\frac{\pi}{2V_{sv}}V_{b,3}\right)\frac{e_1(t)}{E_{in,1}} + \exp\left(-j\frac{\pi}{2V_{sv}}V_{b,3}\right)\frac{e_2(t)}{E_{in,2}}\right], \tag{3.4}$$

where V_{sv} is the switching voltage of the outer modulator, $V_{b,3}$ is the bias voltage of the modulator and $e_1(t)$ and $e_2(t)$ are given by

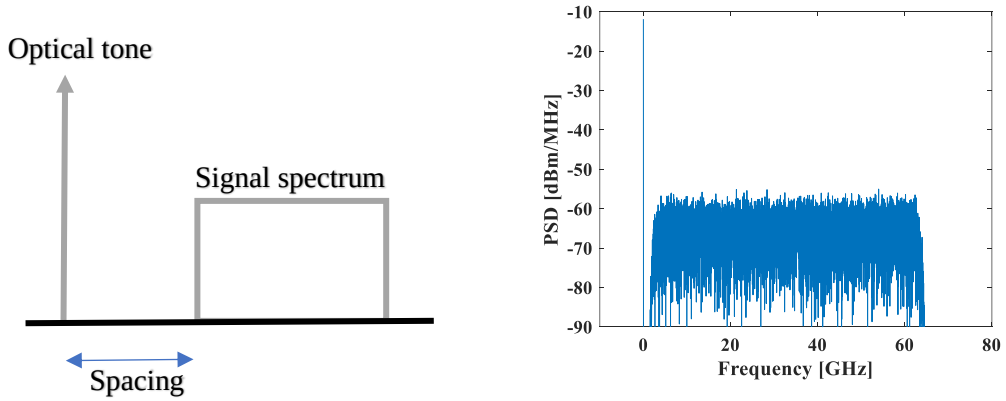
$$e_{1,2}(t) = \frac{E_{in1,2}}{2}\left[\exp\left(j\frac{\pi}{2V_{sv}}v_1(t)\right) + \exp\left(-j\frac{\pi}{2V_{sv}}v_2(t)\right)\right], \tag{3.5}$$

Eq 3.5 corresponds to the electric field at the output of the upper and lower arms of the DPMZM where $v_1(t)$ and $v_2(t)$ are given by

$$v_{1,2}(t) = V_{b\,1,2} + v_{AC\,1,2}(t), \tag{3.6}$$

where $V_{b\,1}$ and $V_{b\,2}$ are the bias voltage of the inner MZMs which are biased at the null bias point ($V_{b\,1,2} = V_{sv}$), and $v_{AC\,1}\,(t)$ and $v_{AC\,2}\,(t)$ are the in-phase and quadrature components of the QAM signal, respectively. The signal at the output of the modulator is a carrier suppressed QAM.

The carrier (an optical tone) is added to the signal to fulfill the minimum phase condition required by the KK receiver. In the simulation, the optical carrier is added into the optical domain. The optical carrier can be added in three different ways: i) by adequately biasing the MZM; ii) by using an external laser; iii) by generating a virtual carrier within the DACs. The drawback of the first approach is the bandwidth. The maximum bandwidth of the transmitted signal achievable is the electrical bandwidth of the information signal, this means that it is not possible to add an additional spacing between the carrier and the information signal. There is no bandwidth limitation with the second approach, hence it is possible to add a spacing between the carrier and the information signal in order to avoid additional distortion. This approach demands more complex implementation as it requires phase synchronization between the carrier and the MZM as well as polarization control. This increase significantly the implementation cost and complexity. As for the virtual carrier-based option, the optimum bandwidth can be achieved at a low complexity implementation.



a) b)

Figure 9 - a) Illustrative spectrum of the signal at the transmitter output. b) Power spectral density obtained by simulation of the signal at the transmitter output.

Figure 9 shows the spectrum of the signal at the output of the transmitter. In figure 9-a), an illustrative spectrum is presented. The optical tone is added to the signal to fulfill the minimum phase condition for the KK receiver [36]. Figure 9-b) shows the power spectral density (PSD) of

the signal at the transmitter output obtained by simulation. The spacing between the carrier and the signal must be chosen to maximize the spectral efficiency without adding distortion to the received signal (this subject will be further developed later on this work).

3.1.2 MCF model

The MCF allows the transmission of multiple optical channels in different cores. In this work, only two cores are considered: one serving as the interfering core, *i.e.*, the core that induces the ICXT, and other acting as the interfered core, *i.e.*, the core impaired by the ICXT. The ICXT induced by the MCF is modeled by the dual polarization discrete change model (DCM) with two cores presented in [27, 54, 55]. Each core operates as a linear single mode fiber (SMF), *i.e.*, it is modeled by the SMF propagation transfer function. Random polarization rotation induced by the fiber birefringence is also included in the transmission model [53].

The dual polarization DCM is a model that generalizes the ICXT in weakly-coupled homogeneous MCF into a dual polarization scheme (*x* and *y*). The amplitude of the field at the input of the interfering core is distributed between the two polarizations. The power distribution is controlled by $\zeta \in \{0,1\}$, where the amplitude of both polarizations is given by $e_{m,x}(t) = \sqrt{\zeta} \cdot e_m(t)$ and $e_{m,y}(t) = \sqrt{1-\zeta} \cdot e_m(t)$ [54].

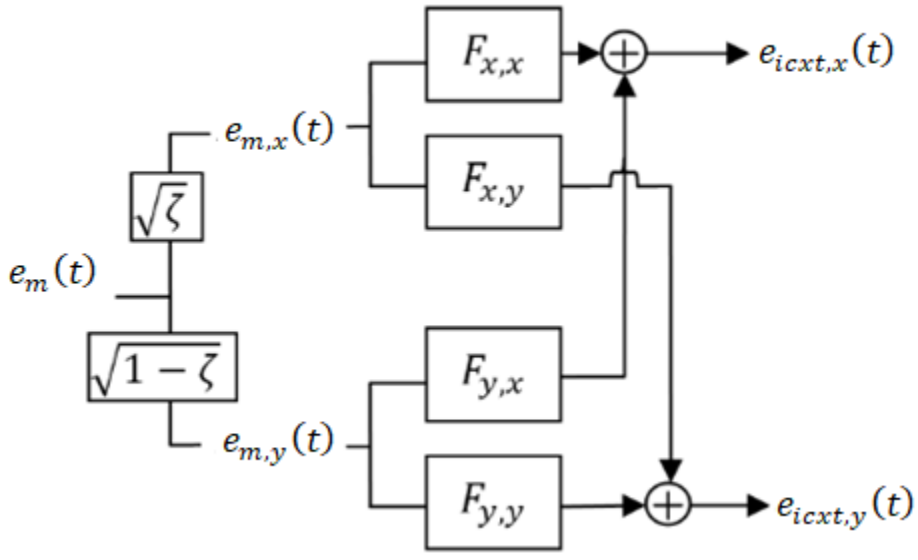


Figure 10 - Conceptual illustration of the DP-DCM (adapted from [54]).

Figure 10 depicts an illustration of the DP-DCM. The two polarizations of the ICXT at the output of the core *n* can be obtained by using the transfer functions $F_{x,x}$, $F_{x,y}$, $F_{y,x}$ and $F_{y,y}$, which are given by

$$F_{a,b}(\omega) = -j\frac{\bar{K}_{nm}}{\sqrt{2}}\exp\left[-j\overline{\beta_n}(\omega)L\right]\sum_{k=1}^{N}\left[\exp\left[-j\overline{\Delta\beta_{mn}}(\omega)z_k\right]\exp\left[-j\phi_{nm,k}^{a,b}\right]\right], \quad (3.7)$$

where L is the fiber length, ω is the angular frequency, $a, b \in \{x, y\}$, $\bar{K}_{nm}$ is the discrete coupling coefficient, which considers the average inter-core coupling coefficient of both polarizations, $\overline{\beta_n}(\omega)$ is the average of the intrinsic propagation constants of the two polarizations directions of core *n* and $\overline{\Delta\beta_{mn}}(\omega)$ is given by

$$\overline{\Delta\beta_{mn}}(\omega) = \overline{\Delta\beta_{0,mn}} + d_{mn}\omega - \Delta D_{mn}\lambda^2\omega^2/(4\pi c), \quad (3.8)$$

where $\overline{\Delta\beta_{0,mn}}$ is the difference between the averages of the propagation constants at $\omega = 0$, d_{mn} is the walkoff parameter between core *n* and *m*, and ΔD_{mn} is the difference between the dispersion parameters of core *m* and *n*. $\phi_{nm,k}^{a,b}$ are the random phase shifts (RPS) that model the random fluctuations of the ICXT induced by structural fiber perturbations in bending radius. The RPSs are random variables with uniform distribution between 0 and 2π, and each RPS is introduced at the *k*-th random coordinate z_k, which is uniformly distributed between $(k-1)L/N$ and kL/N, where *N* is the number of phase matching points (PMP) and *k* is an integer between 1 and *N*.

The ICXT induced by core *m* at the output of the core *n*, is given by

$$e_{icxt}(t) = e_{icxt,x}(t) + e_{icxt,y}(t), \quad (3.9)$$

where $e_{icxt,x}(t)$ and $e_{icxt,y}(t)$ are given by

$$e_{icxt,x}(t) = F^{-1}\left[e_{m,x}(\omega)F_{x,x}(\omega)\right] + F^{-1}\left[e_{m,y}(\omega)F_{y,x}(\omega)\right], \quad (3.10)$$

$$e_{icxt,y}(t) = F^{-1}\left[e_{m,y}(\omega)F_{y,y}(\omega)\right] + F^{-1}\left[e_{m,x}(\omega)F_{x,y}(\omega)\right], \quad (3.11)$$

where $e_{m,x}(\omega)$ and $e_{m,y}(\omega)$ are the QAM signals transmitted in core *m*.

As mentioned, each core operates as a SMF. Thus, the MCF propagation in each core is described by the fallowing transfer function

$$H_f(\omega) = \exp\left(-j\beta_0 L - j\beta_1 \omega L - j\frac{\beta_2}{2}\omega^2 L - j\frac{\beta_3}{6}\omega^3 L\right).\exp\left(-\frac{\alpha}{2}L\right), \quad (3.12)$$

where β_0=$n_{eff}2\pi/\lambda_0$ is the propagation constant, $\beta_1 = 1/v_g$is the propagation time delay, β_2 is the group velocity dispersion, $\beta_3$is the higher order dispersion and α is the fiber attenuation coefficient. These are given, respectively, by

$$\beta_2 = -\frac{{\lambda_0}^2 D_{\lambda_0}}{2\pi c}, \quad (3.13)$$

$$\beta_3 = \left(\frac{{\lambda_0}^2}{2\pi c}\right)^2 S_{\lambda_0} + -\frac{{\lambda_0}^3 D_{\lambda_0}}{2\pi^2 c^2}, \quad (3.14)$$

$$\alpha = \frac{\alpha_{[dB/km]}}{10^4 \log_{10}(e)}, \quad (3.15)$$

where λ_0 is the operating optical wavelength, $D_{\lambda_0}$is the dispersion parameter at λ_0, $S_{\lambda_0}$is the slope of the dispersion parameter and *c* is the light speed in vacuum (*c*=299792458 m/s).

The random polarization rotation induced by the fiber birefringence is also included in the transmission model [55]. Thus, the signal at the output of the interfered core is given by

$$e'_n(t) = \left[e'_{n,x}(t) + e_{icxt,x}(t)\right] + \left[e'_{n,y}(t) + e_{icxt,y}(t)\right], \quad (3.16)$$

where $e'_{n,x}(t)$ and $e'_{n,y}(t)$ are given by

$$e'_{n,x}(t) = \left[\exp(j\theta)\cos(\Gamma).e_{n,x}(t) - \exp(-j\psi)\sin(\Gamma).e_{n,y}(t)\right] * F^{-1}\left[H_f(\omega)\right], \quad (3.17)$$

$$e'_{n,y}(t) = \left[\exp(j\psi)\sin(\Gamma).e_{n,x}(t) + \exp(-j\theta)\cos(\Gamma).e_{n,y}(t)\right] * F^{-1}\left[H_f(\omega)\right], \quad (3.18)$$

where $e_{n(x,y)}(t)$ is the QAM signal transmitted at the core *n* at each polarization (*x* or *y*), θ, ψ and Γ are random processes uniformly distributed between 0 and 2π, and they can be considered constant during a short time duration where the ICXT is considered constant as the decorrelation time of the ICXT is of the order of minutes.

3.1.3 Optical receiver

After passing through the MCF, the QAM signal impaired by the ICXT arrives at the input of the KK receiver, which includes a PIN photodetector, the RRC filter and a DSP block implementing the KK algorithm.

3.1.3.1 PIN photodetector

The PIN is responsible to convert the optical signal back into electric domain. The electric current at the PIN output is given by

$$i(t) = R_\lambda |e(t)|^2 + n_e(t), \tag{3.19}$$

where $|e(t)|^2$ is the instantaneous power of the optical signal at the PIN input, $n_e(t)$ is the electric noise generated by the transimpedance amplifier (TIA) stage of the PIN and R_λ is the PIN responsivity given by

$$R_\lambda = \frac{\eta q}{h\nu}, \tag{3.20}$$

where η is the PIN efficiency, q is the electron charge, ν corresponds to the optical frequency of the input signal and *h* is the Planck's constant. In this work, a PIN responsivity of R_λ =0.7 A/W is considered.

3.1.3.2 KK algorithm

The KK algorithm is based on the Kramers-Kronig relation and, if the minimum phase condition (when the time trajectory of the signal in the complex plane does not include the origin) is ensured,

it enables to retrieve the complex field at the PIN input from the photocurrent detected by the PIN. To retrieve the signal from its intensity, it is necessary that the power of the carrier is larger than the peak power of the information-bearing signal. In other words, the carrier-to-signal power ratio (CSPR),

$$CSPR = |A|^2/\langle|S(t)|^2\rangle, \tag{3.21}$$

should be larger than the peak-to-average-power ratio (PAPR) of the information-bearing signal, given by

$$PAPR = \max(|S(t)|^2)/\langle|S(t)|^2\rangle, \tag{3.22}$$

where $S(t)$ is the information-bearing signal and A is the amplitude of the optical carrier.

With the minimum phase condition ensured, the phase of the signal can be calculated as

$$\varphi(t) = 0.5\, H\big(\ln(i(t))\big), \tag{3.23}$$

where $H(x)$ is the Hilbert transform of x.

Then the signal can be reconstructed as

$$s(t) = \left[\sqrt{i(t)} \times \exp\big(j\varphi(t)\big) - A\right], \tag{3.24}$$

After the reconstruction of the complex signal, the digital compensation of the chromatic dispersion can be performed. For the sake of simplicity, in this work, this compensation is performed using an analog filter with transfer function given by

$$H_{CD}(\omega) = H_f(\omega)^*, \tag{3.25}$$

where $H_f(\omega)^*$ is the complex conjugate of $H_f(\omega)$.

3.2 Figures of merit for performance evaluation

After the description of the system model, the figures of merit employed for performance evaluation are presented in this section. In particular, a brief description of error vector magnitude (EVM) and bit error rate (BER) are presented.

3.2.1 EVM

One method to assess the performance of QAM signals is the EVM, which quantifies the difference between the transmitted and the received symbols.

$$evm = \frac{\sum_{k=1}^{N_{sim}} |s_0(k) - s_i(k)|^2}{\sum_{k=1}^{N_{sim}} |s_i(k)|^2}, \tag{3.26}$$

$$EVM = 10\log_{10}(evm), \tag{3.27}$$

where N_{sim} is the number of the transmitted symbols and $s_0(k)$ and $s_i(k)$ are the coordinates of the received and transmitted symbols, respectively. The EVM can be related with the SNR by

$$EVM = -SNR, \tag{3.28}$$

3.2.2 BER

BER is another figure of merit that is commonly used for evaluating the optical system performance. In the presence of Gaussian noise, the BER can be theoretically computed from the EVM. The relation between these two figures of merit can be used to validate the system model in the presence of Gaussian noise, as is the case of the noise induced by the photodetector.

The BER can be calculated from EVM by

$$BER = 2\frac{1-\frac{1}{\sqrt{M}}}{\log_2(M)}\operatorname{erfc}\left(\sqrt{\frac{3\log_2(M)}{M-1}\frac{1}{\text{evm}\times\log_2(M)}}\right), \tag{3.29}$$

where *M* is the modulation index.

The BER can also be calculated using Monte Carlo simulation, where a direct error count is performed and the BER is the ratio between the received bits with errors and the total received bits. In order to ensure an adequate confidence interval for the BER estimation, the Monte Carlo simulation runs until there is one hundred error occurrences. Monte Carlo simulation can be used in scenarios where there are additional distortions other than Gaussian noise.

3.3 Performance of the baseline system

In this subsection, the performance of the system is evaluated in absence of the ICXT. This assessment is performed without compensation of CD and then employing CD compensation. Furthermore, the BER is assessed in presence of the Gaussian noise induced by the PIN. The baseline system is the system when ICXT is not considered.3.3.1 System performance without dispersion compensation

The system performance is assessed without the compensation of the CD in order to identify the signal degradation as a function of the fiber length for 16-QAM with a symbol rate of 60 Gbaud and 64-QAM signals with a symbol rate of 40 Gbaud.

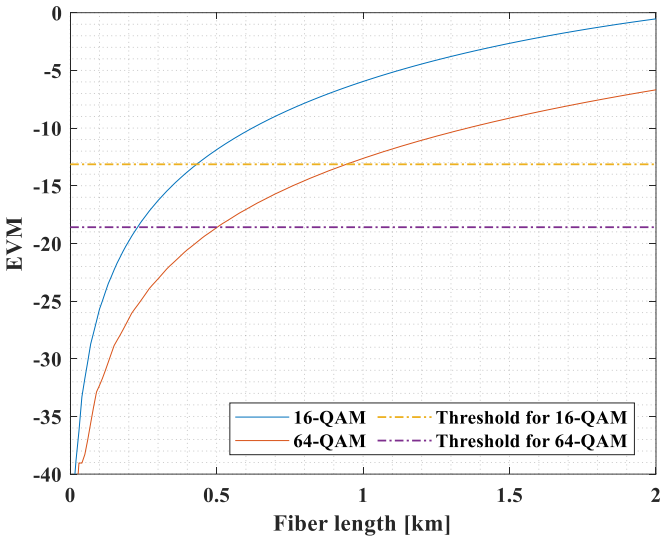
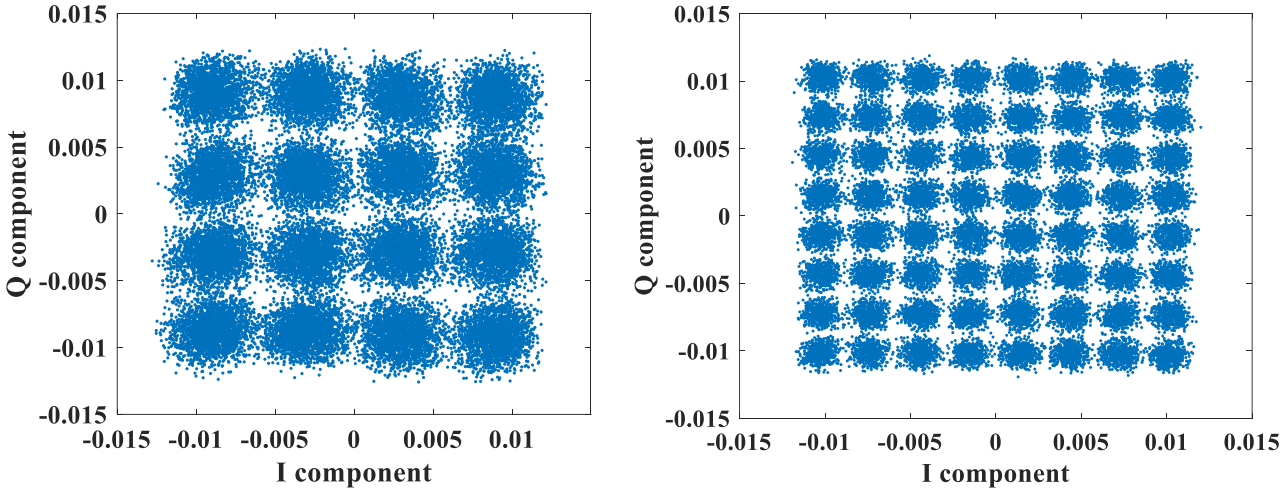


Figure 11 - EMV as function of the fiber length.

Figure 11 shows the EVM as a function of the fiber length considering a CSPR of 13 dB, RRC pulses with a 0.05 roll off and the power of the signal at the input of the fiber is set to 0 dBm. It

shows that the system is highly affected by the chromatic dispersion due to the high symbol rate. The 64-QAM is more tolerant to the CD compared to the 16-QAM as the 64-QAM has a better spectral efficiency. The results show that, for an EVM threshold corresponding to a BER of $10^{-1.8}$, the maximum reach attained with 64-QAM does not exceed 500 m.



a) 16-QAM constellation b) 64-QAM constellation

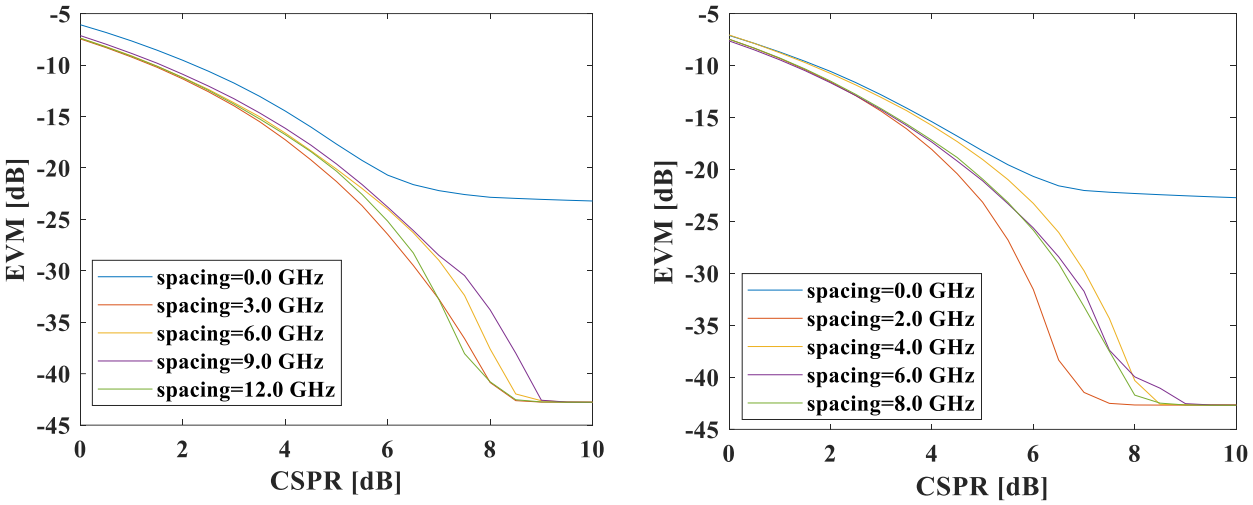
Figure 12 - Constellations with 350 m of fiber.

Figure 12 depicts the constellations obtained for 16-QAM and 64-QAM signals impaired by the chromatic dispersion. With this symbol rate the system presents a lot of distortion in a few meters of fiber length, thus, the compensation of the dispersion should be employed.

3.3.2 System performance with dispersion compensation

As shown in figure 11, the CD highly affects the system and it limits the fiber reach to a few meters which is not ideal for inter data centers links, requiring CD compensation.

The KK receiver allows the reconstruction of the complex signal from its intensity with the distortion depending on the CSPR and the frequency spacing between the optical tone and the signal. With the reconstruction of the complex signal, it is possible to compensate the CD. For this reason it is important to evaluate the KK receiver performance according to CSPR, CD and the frequency spacing between the optical tone and the signal spectrum.



a) 16-QAM b) 64-QAM

Figure 13 - EVM as a function of the CSPR and the spacing between the carrier and the QAM signal.

Figure 13 shows the EVM as a function of the CSPR and the spacing between the tone and the QAM signal for 100 km of fiber. The results show that, beside the case which there is no spacing between the signal and the optical tone, with all the others spacings the EVM stabilizes at a CSPR of approximately 9 dB. For CSPR higher than 9 dB, the EVM remains the same, as in this case the thermal noise is not considered. As most of the spacings achieve the optimum performance at the same level of CSPR, the criteria to choose a spacing is based on spectral efficiency, hence in this work, we choose a spacing between the optical carrier and the signal of 3 GHz in 16-QAM and 2 GHz in 64-QAM.

3.3.3 System performance with thermal noise

In this section, the system performance is assessed in the presence of the thermal noise induced by the PIN. First the parameters used in this work are chosen according to the best performance of the system when it is not impaired by the random ICXT. Furthermore, the main impairment will be the ICXT as the final objective is to assess the performance of a MCF based system.

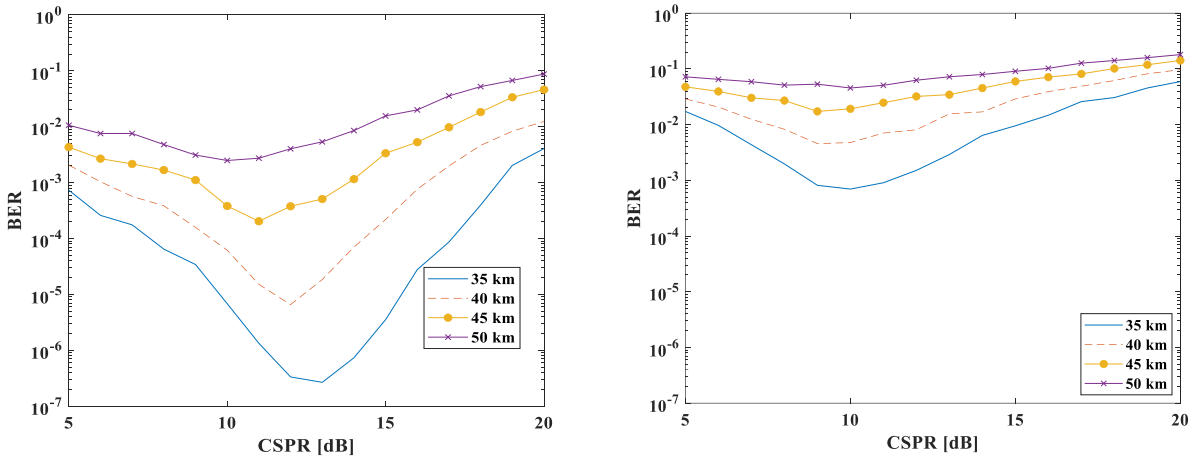
Table 2 - Signal parameters.

Parameters	Value
Symbol rate (Gbaud)	60
Roll-off factor	0.05
Input MCF power [dBm]	0

Table 3 - MCF parameters.

Parameters	Value
Attenuation [dB/km]	0.22
Interfered core effective refractive index	1.4453
Interfering core effective refractive index	1.4455
Wavelength [nm]	1552
Skew×symbol rate	0.001
ICXT level [dB]	-13
Length [km]	35

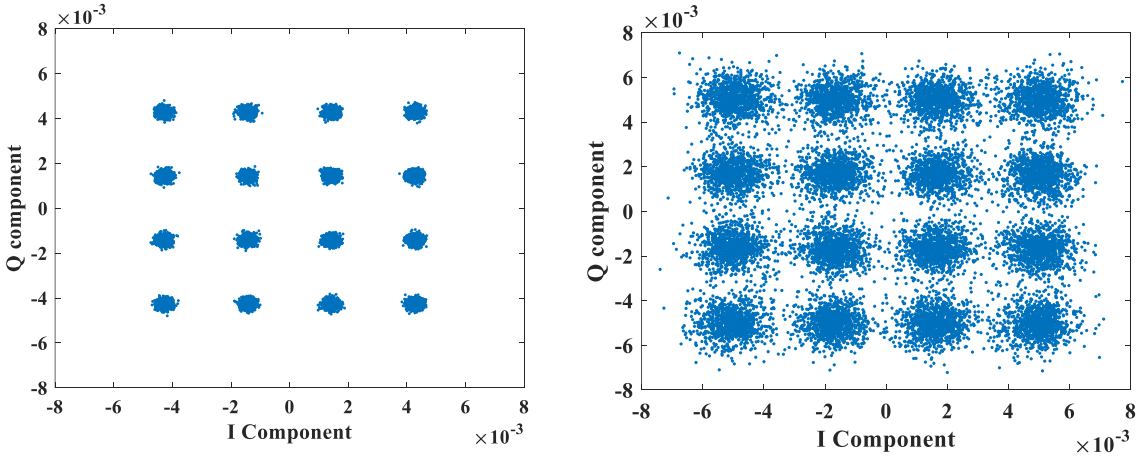
To use a KK receiver, we need to ensure the minimum phase condition. Thus, the CSPR must be optimized. For this we have considered a standard value for the noise equivalent power (NEP) of 10 $\mathrm{pW}/\sqrt{\mathrm{Hz}}$.



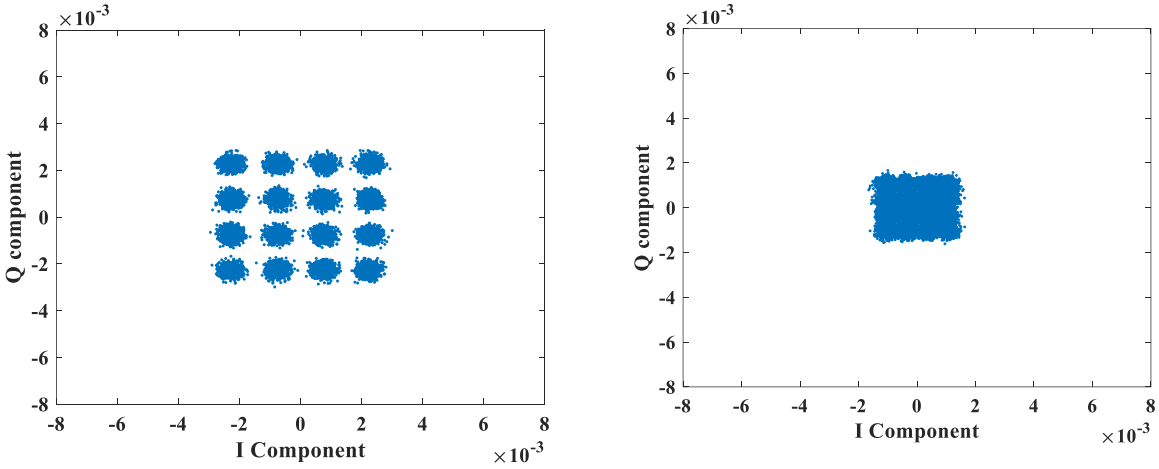
a) 16-QAM b) 64-QAM

Figure 14 - BER as a function of the CSPR and fiber length.

Figure 14 shows the BER as a function of the CSPR for four different fiber lengths. Results are obtained considering the absence of ICXT and perfect dispersion compensation. Figure 14 shows that the optimum CSPR for the 16-QAM signal is approximately 13 dB. This optimum operation point results from a trade-off between SNR and verification of the minimum phase condition. For low CSPR levels, the condition is not satisfied. For high CSPR levels, the SNR degrades. The 64-QAM signal presents a minimum of BER of approximately 10^{-3} while the BER of 16-QAM signal is below 10^{-6}. Thus, in the following, the modulation considered will be the 16-QAM.



a) CSPR=5 dB (w/out KK receiver) b) CSPR=5 dB



c) CSPR= 13 dB d) CSPR=20 dB

Figure 15 - Constellation of a 16-QAM when using: a) CSPR of 5 (w/out KK receiver) dB; b) CSPR of 5 dB; c) CSPR of 12 dB; d) CSPR of 20 dB

Figure 15 depicts the constellation of the 16-QAM signal when using: a) a low CSPR of 5 dB without the distortion of the KK receiver; b) a low CSPR of 5 dB; c) the optimum CSPR of 13 dB and d) a high CSPR of 20 dB. The results show that for a low CSPR, the signal power is higher. With a higher signal power, the SNR is higher, although the signal is degraded due to the low CSPR which affects the minimum phase condition as shown in the comparation between figure 15-a) with figure 15-b). For a CSPR of 13 dB, the signal power decreases. At this operation point, the system achieves the lowest BER. The minimum phase condition is ensured, and the SNR is high enough to achieve a BER below 10^{-6}. For a CSPR of 20 dB, the signal power decreases drastically due to the ratio between the optical carrier power and the signal power. The input power into the MCF is limited to 0 dBm, if we increase the CSPR, the signal power decreases and the SNR also decreases.

After establishing all the parameters, it is important to confirm that the simulation works as expected. For this, it is necessary that the system performance evaluated by Monte Carlo simulation, when impaired only by Gaussian noise induced by PIN, agrees with the system performance calculated theoretically through EVM.

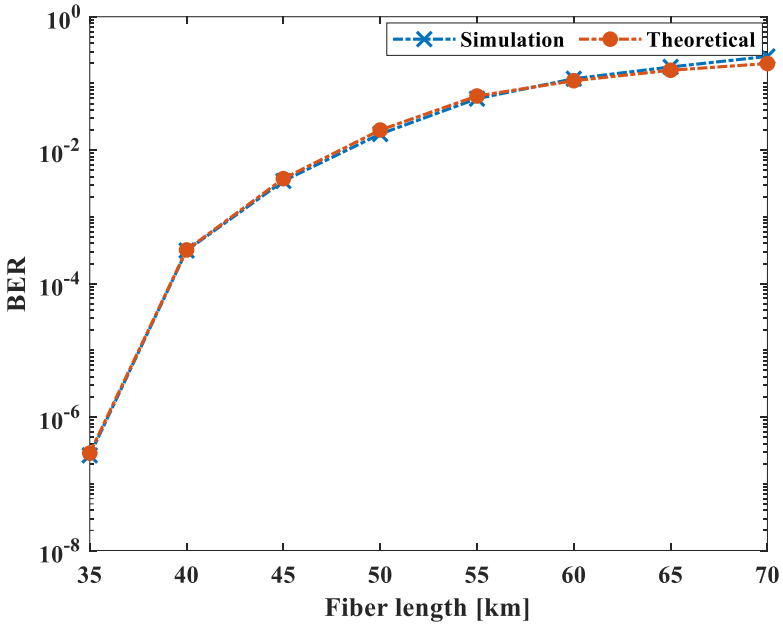


Figure 16 - BER as a function of the fiber length.

Figure 16 depicts the BER as a function of the fiber length considering a CSPR of 13 dB. In simulation, the BER was obtained by Monte Carlo simulation while the theoretical BER was obtained through the theoretical expression shown in equation (3.29). The results show that the system model operates as expected in absence of the ICXT.

3.4 Summary

In this chapter, the system setup used to evaluate the performance of self-coherent >200 Gb/s shorth-reach MCF systems using ML techniques has been presented. The parameters of the different system blocks have been also introduced and discussed. The optical transmitter that generates QAM signals, the DP-DCM used to describe the ICXT, and the KK receiver that reconstructs the complex signal were presented and detailed. Then, the figures of merit for performance evaluation were presented with the focus on EVM and BER. The system performance was evaluated with and without the compensation of the chromatic dispersion. The optimization of the spacing between the optical carrier and the QAM signal, and the of CSPR have been also accomplished. We conclude that, with a spacing of 3 GHz and a CSPR of 13 dB, the system operates at its optimal point.

CHAPTER 4

Performance evaluation of the KK MCF system with ML

In this chapter, the performance of the system in the presence of the random ICXT is assessed and the ML techniques to mitigate the impact of the ICXT are presented. In section 4.1, the statistic of the ICXT is studied and then, in section 4.2, the system performance with the ICXT is assessed. In section 4.3, *k*-means clustering and KNN are presented to perform a dynamic symbol decision. In section 4.4, the ANNs are presented to mitigate the random ICXT and then the optimization of the proposed ANNs is accomplished in section 4.5. Lastly, the outage probability of the proposed system is assessed in section 4.6.

4.1 Validation of the ICXT statistic

The ICXT varies randomly along time. The time scale of the ICXT fluctuations can be of the order of a few minutes or achieve longer periods of few hours [27]. Therefore, the concept of time fraction and short-term average inter-core crosstalk (STAXT) has been introduced to assess the performance of MCF-based networks. A time fraction is a small-time duration, much shorter than the ICXT decorrelation time, where the ICXT is considered constant [54]. STAXT is the average power of the ICXT measured over a time period much shorter than the ICXT decorrelation time. The ICXT varies randomly from time fraction to time fraction, as the time interval between time fractions is much larger than the ICXT decorrelation time.

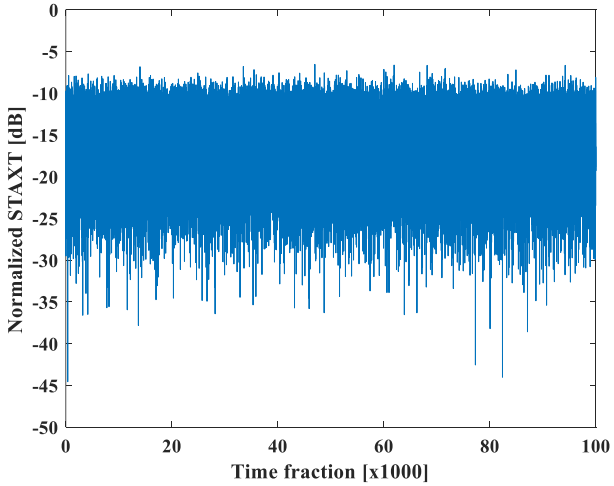
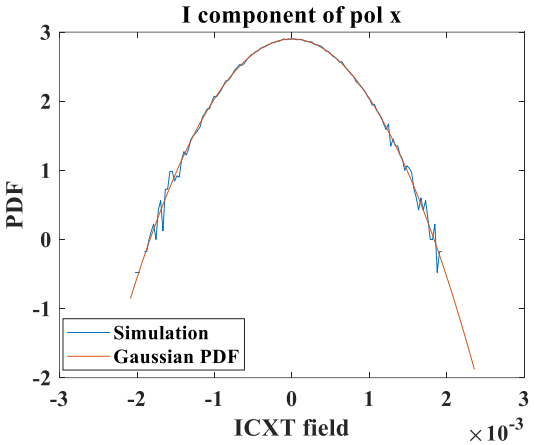


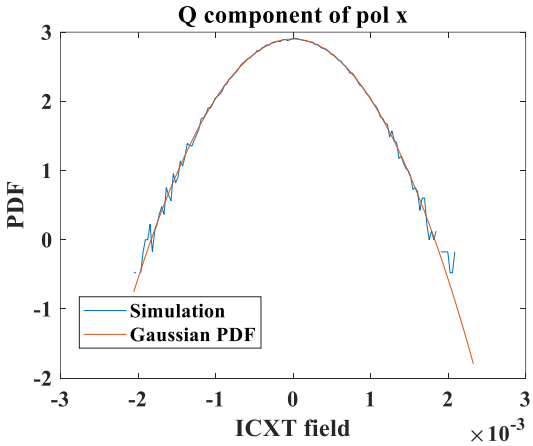
Figure 17 - Normalized STAXT power as a function of the time fraction.

Figure 17 shows the normalized STAXT power as a function of the time fraction. To obtain the presented results, a ICXT level (XT) of -15 dB was simulated on each time fraction where a constant signal (CW laser) was induced on the interfering core. The results were measured at the output of the interfered core. The MCF has a length of 35 km and the ICXT was generated with 1000 RPS. Figure 17 shows that the variation of STAXT power along time exceeds 20 dB. As reference, the ICXT level is the ratio between the mean ICXT power and signal power, at the output of interfered core [17].

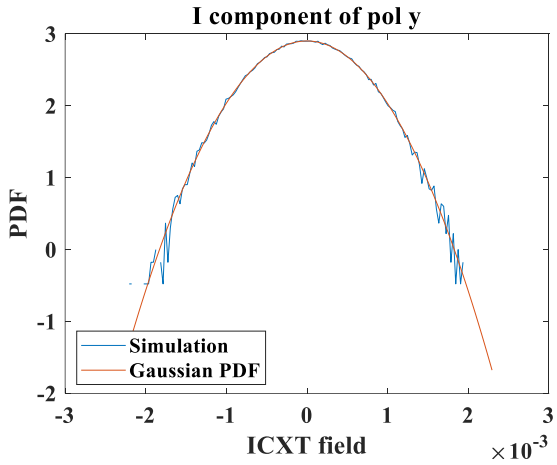
To validate the ICXT generation process implemented in the simulator, the in-phase (I) and quadrature (Q) components of the ICXT field were obtained and compared with the theoretical results. Theoretically, the variation of the ICXT is characterized by a Gaussian distribution [54].



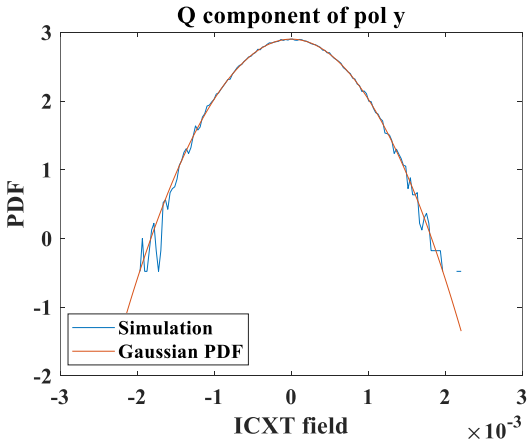
a) In phase component of the ICXT amplitude at polarization *x*.



b) Quadrature component of the ICXT amplitude at polarization *x*.



c) In phase component of the ICXT amplitude at polarization *y*.



d) Quadrature component of the ICXT amplitude at polarization *y*.

Figure 18 - PDF of the quadrature components of the ICXT field.

Figure 18 shows the probability density function (PDF) of the in-phase and quadrature components of the ICXT field in both polarizations (*x* and *y*) directions. The simulation results of the ICXT are obtained with 1000 random phase shifts and for a ICXT level of -15 dB. A Gaussian PDF, obtained from the mean and variance of the simulation results, is also shown in figure 18, as reference. The results show that the ICXT components are well described by a Gaussian PDF, as predicted theoretically in [54].

4.2 Impact of the ICXT on the BER

In this subsection, the impact of the ICXT on the system performance is studied. The main goal is to assess the BER behavior when the system is impaired by the random ICXT. This study is performed considering the ICXT as the main source of the system distortion.

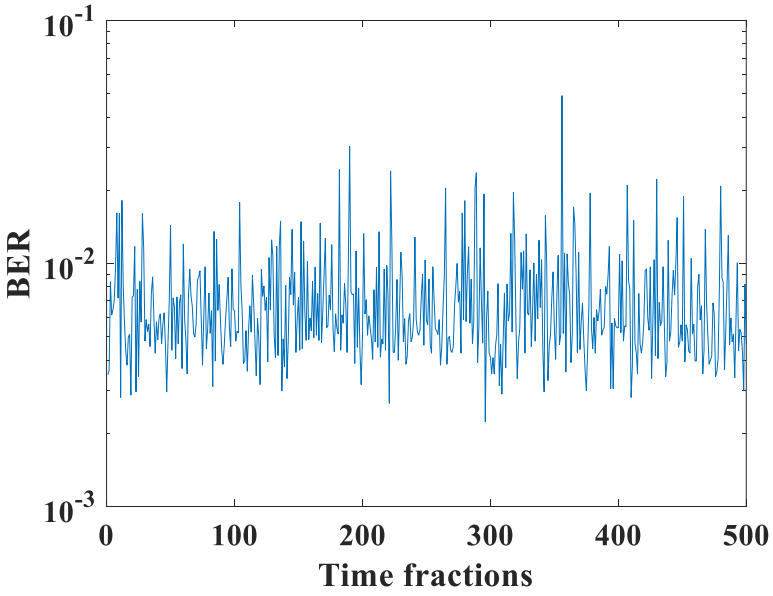


Figure 19 - BER as a function of time fraction.

Figure 19 shows the BER along 500 different time fractions. The ICXT level used is -15 dB and the system parameters are those presented in tables 2 and 3. The results show that the ICXT can cause high BER variations along time which can remain for several minutes or even hours. This is an unwanted effect as it may lead to variation of the service performance provided to the end users.

4.3 Machine learning algorithm for performance optimization

In this work, three types of ML techniques are proposed and implemented: *k*-means clustering, KNN and ANN. The main goal of these techniques is to mitigate the impact of the random ICXT by performing different symbol decision other than the traditional decision with static boundaries, reconstruct the impaired signal or even perform the symbol decision. It has been demonstrated that the ICXT varies. Hence, a static symbol decision may not perform as expected. Thus, dynamic decision according to the ICXT or mitigate the impact of the ICXT for a static symbol decision may be alternative solutions with potential to provide system performance improvement.

4.3.1 *k*-means clustering and KNN classification

The first approach to improve the performance of the system impaired by the ICXT is to perform dynamic decision boundaries with *k*-means clustering or KNN. The proposed techniques can learn patterns on a two-dimensional space and then create different classes to represent each symbol.

- KNN algorithm is based on the distance between the new data and the neighbors counted by the *k* parameter. The neighbor symbols are attributed by the training process where every input symbol has a correspondent output that matches with a class with the same characteristics. On the active phase, a new input symbol is classified into the group of data with the greatest number of nearest neighbor symbols according to the number of neighbors defined by the *k* parameter [44]. The classification used in the algorithm is based on the Euclidean distance. In this work, we divide the symbols into two group of samples (*I* and *Q* components of the QAM signal at the KK receiver output of the interfered core) for the training and active phase.
- *k*-means clustering focuses on dividing the input into *k* clusters, which are automatically determined during the training process according to dissimilarity metrics. It creates groups centered on the mean of all the samples of each cluster (centroids). The algorithm uses Euclidean distance to perform the symbol decision [44]. As features for the classification, the *k*-means algorithm receives the in-phase and the quadrature components of the QAM signal received in the interfered core at the output of the KK receiver.

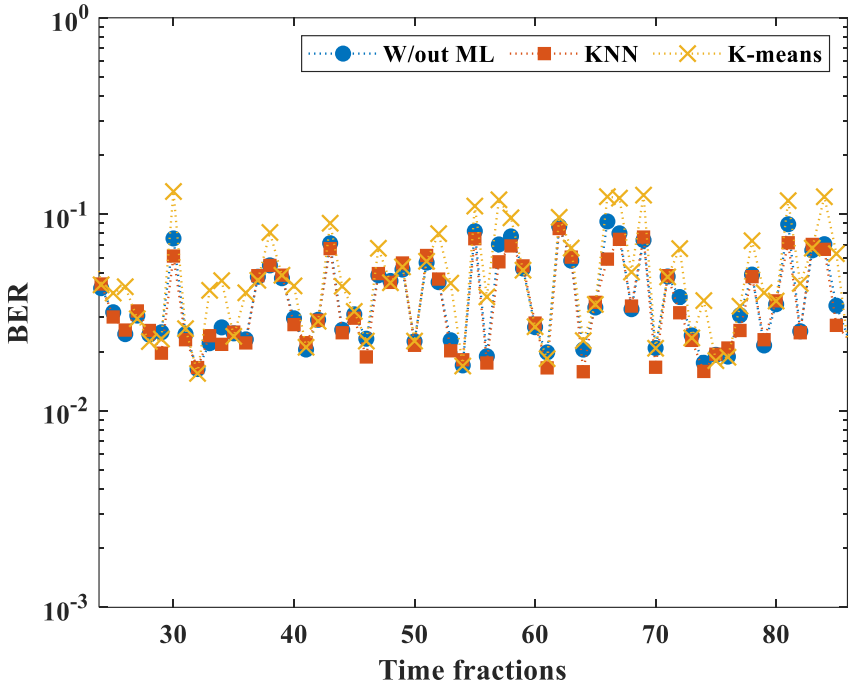
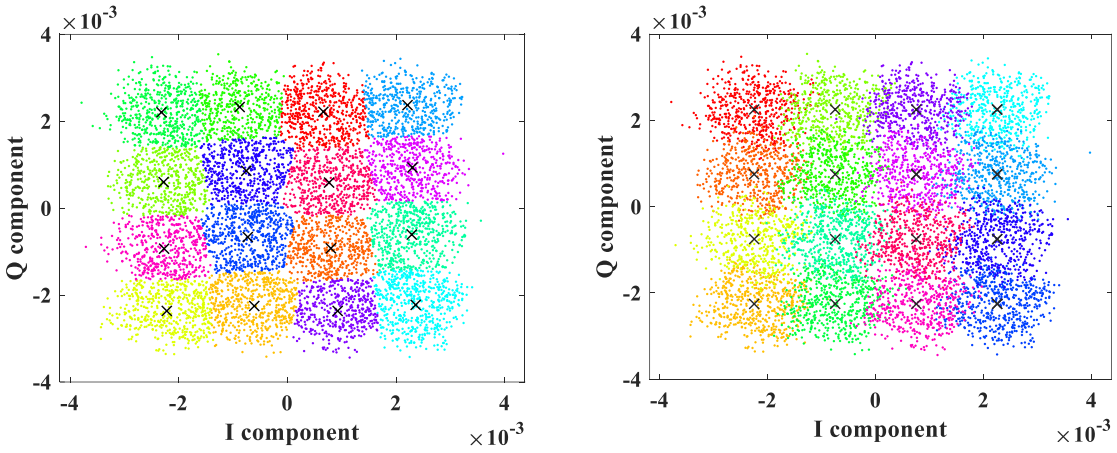


Figure 20 - BER as a function of the time fraction for: absence of ML algorithm (circles), KNN (squares) and *k*-means clustering (crosses).

Figure 20 shows the BER of the received signal before and after using ML algorithms, considering an ICXT level of -13 dB. For each time fraction, the same samples of the random ICXT are used to ensure the same conditions for all algorithms. The results obtained show that *k*-means and KNN algorithms do not provide a significant performance improvement when compared with the performance obtained without using ML and in some cases the ML techniques perform worse than the case without using ML, as shown in table 4. This occurs because the ICXT is random and the simplest redesign of the decision boundaries does not allow to better identify the different clusters.

Table 4 - BER

	Mean	Maximum	Minimum
k-means	0.0553	0.1441	0.0135
KNN	0.0395	0.1113	0.0125
W/out ML	0.0407	0.1001	0.0142



a) *k*-means clustering classification b) KNN classification

Figure 21 - Constellation of the 16-QAM after *k*-means clustering and KNN classification.

Figure 21 depicts the constellation of the 16-QAM signal after the classification performed by the *k*-means clustering and KNN.

4.3.2 Artificial neural networks

The ANN consists of an input layer, fed by the feature values, hidden layers and an output layer, which provides the estimation or classification result. In this work one hidden layer is considered since the use of multiple layers increases the complexity, hence the training duration.

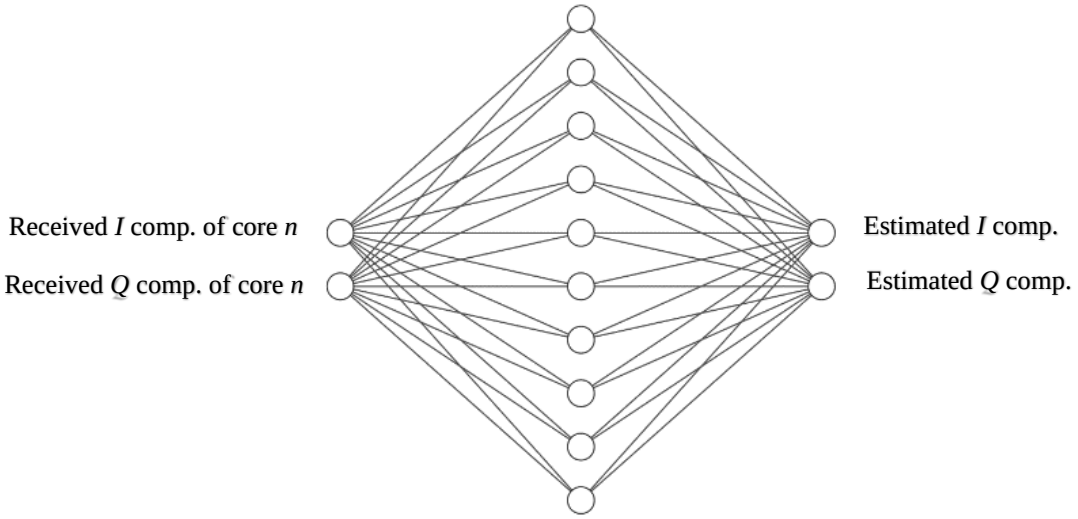
As the simpler ML techniques presented in previous section do not provide performance improvement, ANN is considered as a new method that aims to improve the system performance. The objective is to learn the behavior of the ICXT to mitigate its impact on the system performance. This is performed by testing two neural networks: i) the estimation feedforward neural network (FNN), and ii) the classification FNN. The estimation FNN can be used to solve input to output problems. It is normally used to compensate nonlinearities and predict the output according to the input. The classification FNN is used to perform classification where a set of classes are defined in which its objective is to classify the input into a certain class. In this work, the implemented approaches are as follows:

- An estimation FNN that aims to improve the performance of the system impaired by ICXT. To identify the required features employed by the FNN in the training phase to learn how

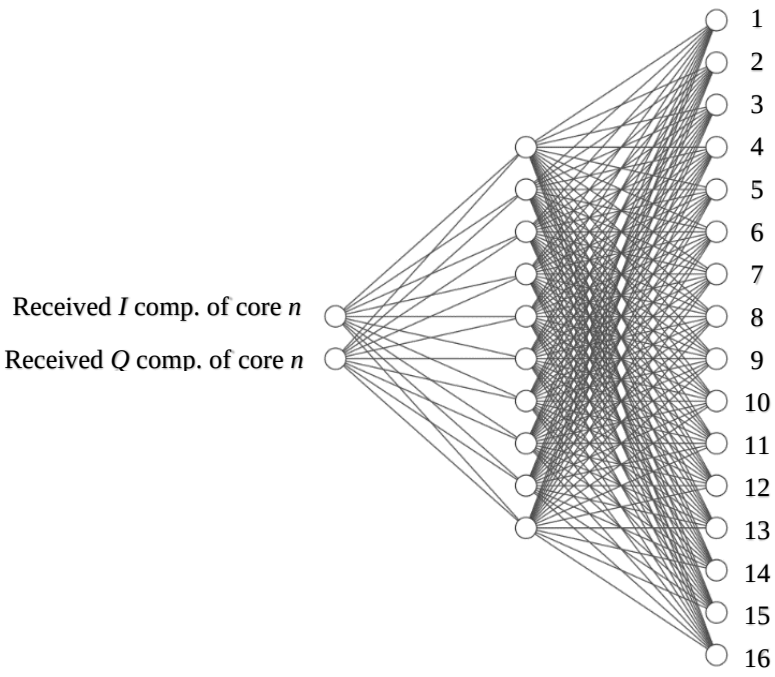
to mitigate the ICXT impairment. The main goal of the FNN is to estimate the transmitted signal launched into interfered core, and then employ a static symbol decision approach. In this work the FNN is chosen as it is expected to mitigate the impact of the ICXT and predict the output according to the information feeded to the input.

- A classification FNN to identify the required features employed by the FNN in the training phase to learn how to classify the symbols of the received signal at the interfered core. The objective is to perform the decision of the symbols, as it can learn the pattern and directly perform the decision. As the final objective is to retrieve the correct symbols, the estimation FNN can also be used to classify the received symbols as a QAM symbol. It is expected that it can perform the correct symbol decision without needing to reconstruct the ICXT impaired signal.

4.3.2.1 ANN with two features



a) Estimation FNN configuration.



b) Classification FNN configuration.

Figure 22 - Schematic diagram of the estimation and classification FNN used to mitigate the impact of the ICXT in DD-MCF short-reach networks.

Figure 22 depicts the scheme of the ANN employed with two inputs. The ANNs are feeded with two inputs (features): the in-phase and the quadrature components of the signal transmitted in the interfered core at the output of the KK receivers. The output of the estimation FNN is the *I* and *Q* components of the estimated signal, while the output of the classification FNN is the QAM symbols values.

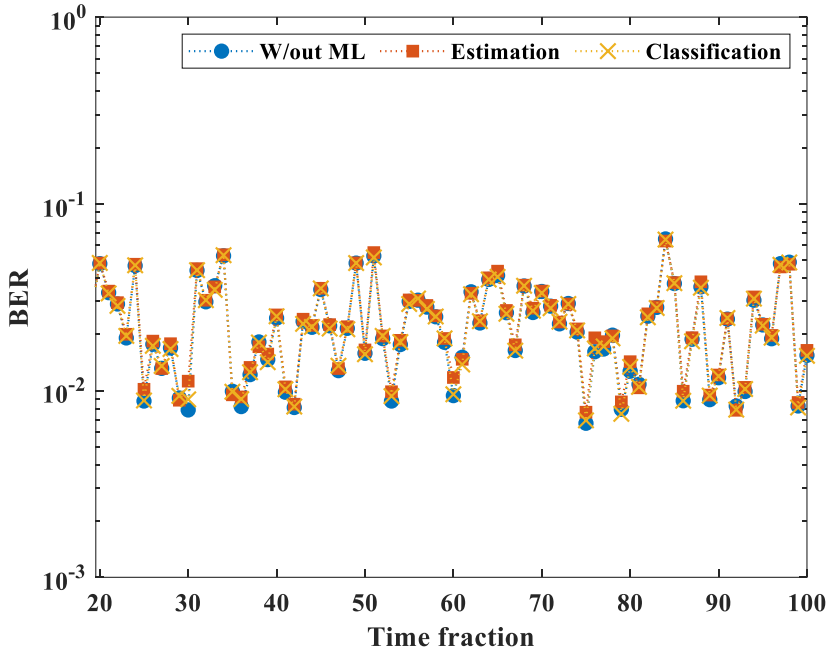


Figure 23 - BER as a function of the time fraction for: absence of ML (circles), estimation FNN (squares), classification FNN (crosses).

Figure 23 shows the performance of the estimation and classification FNN. The case in which the system is operating without the ML block is also shown, as reference. For this simulation a ICXT level of -13 dB is used. The results show that simple FNN configurations with only two inputs did not perform better than the system without the use of ML. The FNNs with two inputs do not present enough information to recognize patterns to mitigate the effects of the ICXT. As the ICXT depends on the signal injected into the interfering core *m* [29], [56], we may need to provide the signal detected at the output of core *m* as an input feature.

4.3.2.2 ANNs with four features

ANNs with two inputs cannot perform better due to the lack of ICXT information. The solution is to increase the number of inputs. In point-to-point links with skew × symbol rate << 1, it is possible to collect the information of the signal received at the interfering core.

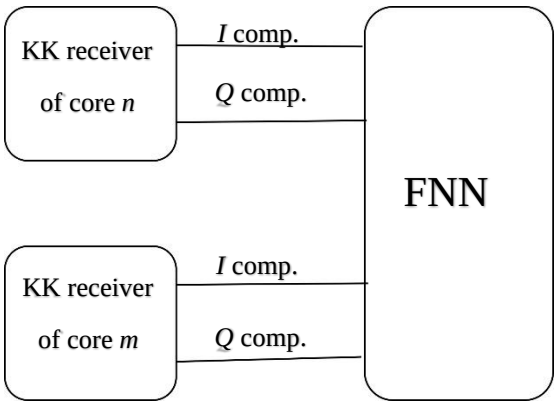


Figure 24 - Scheme of the inputs of the FNNs.

Figure 24 depicts the scheme of the inputs of the FNNs. The NNs are feeded with the *I* component and *Q* component at the output of the interfered and interfering core.

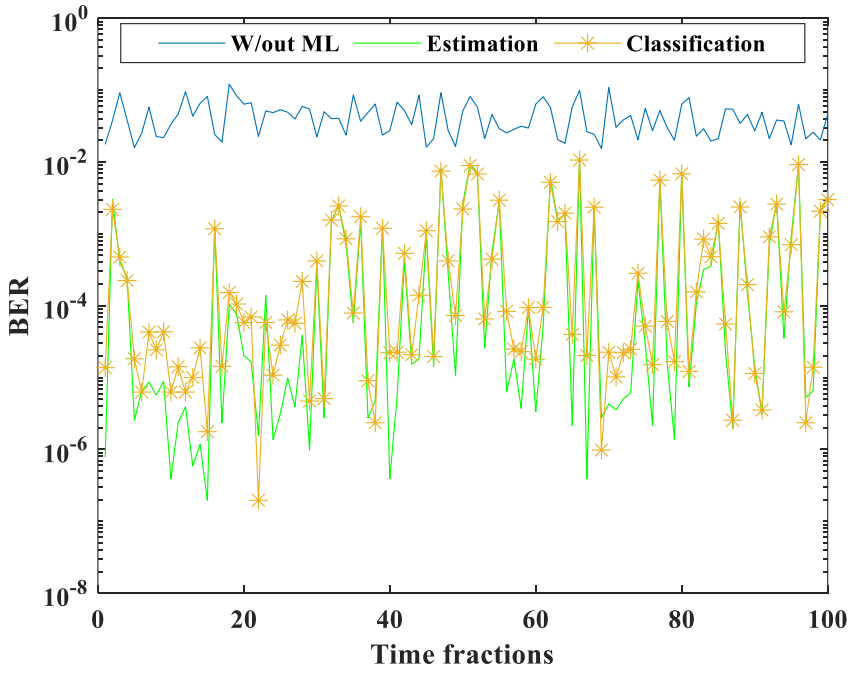
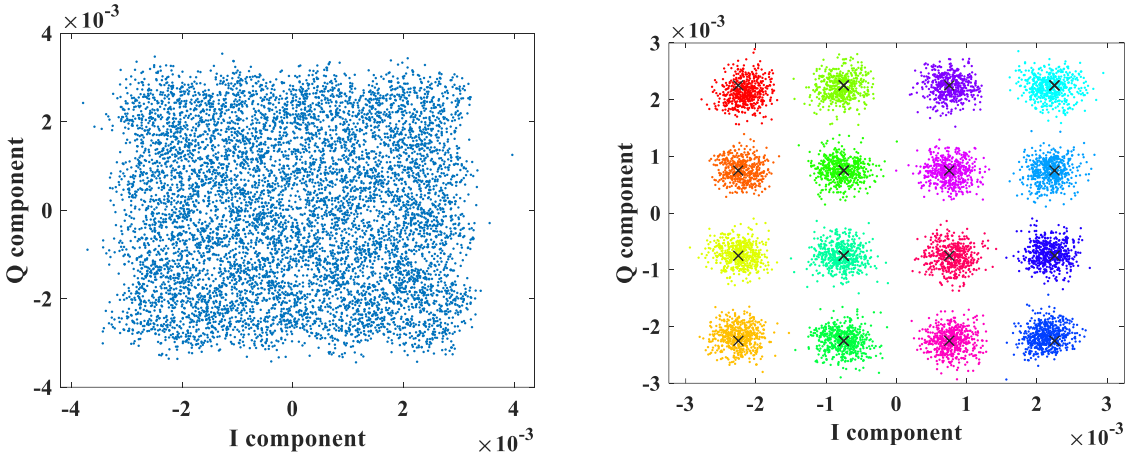


Figure 25 - BER as a function of the time fraction for: the absence of ML, with the estimation and classification FNN.

Figure 25 shows the BER of the received signal before and after using ML algorithms, considering an ICXT level of -13 dB. Both FNNs have a single hidden layer of 10 neurons. The optimization algorithm used in the training phase was the scaled conjugate gradient, and the activation function used was the tangent sigmoid. The only difference between these FNNs and the previous FNNs

configuration is the number of inputs. Thus, with four inputs, the FNNs are expected to learn how a core (interfering core) can affect the other (the interfered). In contrast to previous ML techniques, figure 25 shows that the 4-input estimation and classification FNN enable the mitigation of the ICXT-induced BER degradation. This is confirmed by the comparison between the constellations obtained without ML and after employing the estimation FNN, shown in figure 26 a) and b), respectively.



a) Constellation of the received signal

b) Reconstructed signal by the estimation FNN

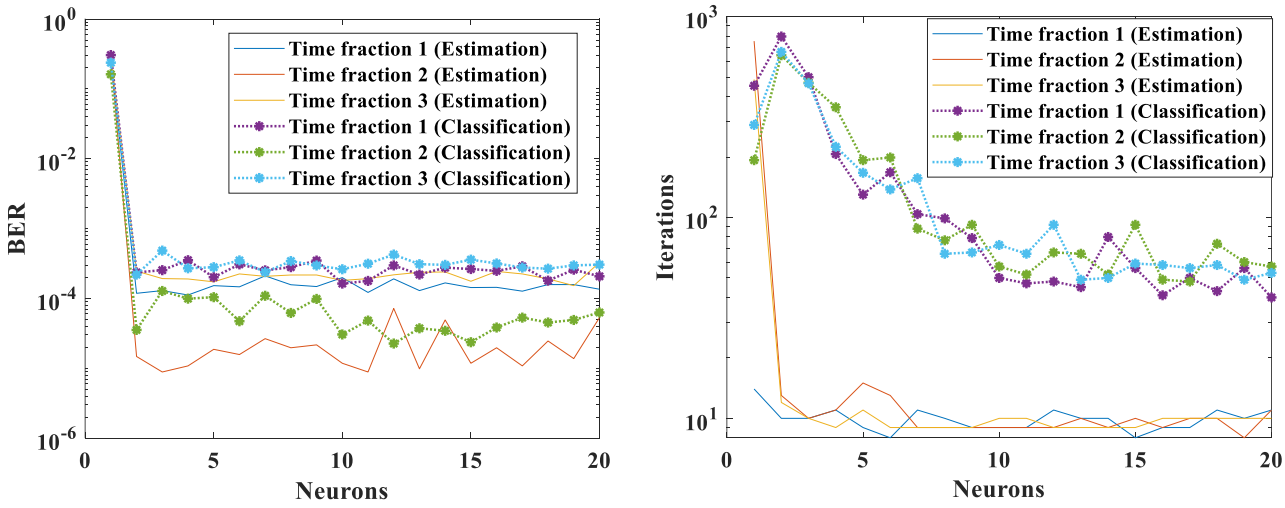
Figure 26 - Constellation of the 16-QAM signal before and after using the FNN.

4.4 Optimization of the ANNs

The previous section has proven that the ANNs can improve the performance of the system by reconstructing the distorted signal or by performing the symbol decision. Now, it is important to optimize the NNs configuration according to best performance while maintaining a low complexity. To perform the optimization, the number of neurons used and the training symbols are studied.

4.4.1 Number of neurons

The number neurons in the hidden layer are studied in this subsection. The BER is used to evaluate the system performance when employing the FNN techniques according to the number of neurons. The number of iterations as a function of the number of neurons on the hidden layers is used to evaluate the training phase complexity.



a) BER vs Neurons

b) Iterations vs Neurons

Figure 27 - Performance of the NNs as function of a) number of neurons and b) number of training iterations as a function of the number of neurons.

Figure 27 shows the performance of the FNNs as function of number of neurons in the hidden layer. Figure 27 a) represents the BER as a function of number of neurons in three different time fractions. It shows that in FNNs with number of the neurons higher than two, the performance is basically the same. According to the theory, if the relation between the inputs and the output is nonlinear, it is necessary at least two neurons at the hidden layer [57]. Figure 27 b) shows the number of iterations required for each training process. The higher the number of iterations needed to train the FNN, the higher the duration of the training process. Thus, the number of neurons can be chosen according to the number of iterations as long as it is higher than one neuron. The number of iterations stabilizes at two neurons for the estimation FNN while the number of iterations starts to stabilize at 10 neurons on the classification FNN. In this work 10 neurons are used, as it is

relatively a low number of neurons and achieve the optimum number of iterations for training process of the classification FNN.

It is important to point out that estimation FNN takes significantly less iterations in the training phase than the classification FNN. This is due the loss function used for error optimization in each case, as the estimation FNN uses mean square error loss function and the classification FNN used the crossentropy loss function, which is a more sophisticated function that requires more iterations to achieve the optimum performance. Thus, the training duration on estimation FNN is much smaller than the classification FNN training duration.

4.4.2 Number of training samples

Another important parameter is the number of samples used to train the network. In this section, the BER is assessed according to the number of samples used in the training set.

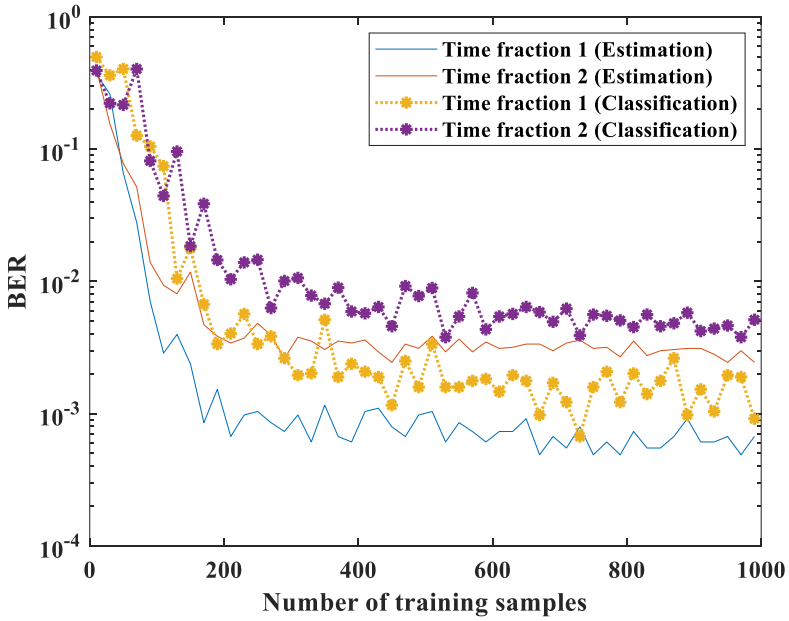


Figure 28 - BER as a function of the number of training samples.

Figure 28 shows the evolution of the BER as a function of the number of training samples. The results show that the higher the number of training samples, the better the NNs performs. Nevertheless, it is not necessary a high number of training samples to achieve the desired performance. In DD systems with skew × symbols rate << 1, the ICXT can be treated as a static

distortion component coupled into the signal. Then, if we use 16-QAM mapping, each QAM symbol of the interfered core will be impaired by 16 different possibilities. In this case there are 16^2 combinations of symbols transmitted into the two cores. Thus, it would be expected about 256 samples required for training, as one sample corresponds in two QAM signals and the FNNs requires four inputs (*Q* and *I* components of each QAM signal). The results show that the BER starts to stabilize when using approximately 256 samples to train the NNs, which is aligned with the theory. In this work, 10000 samples corresponding to 10000 QAM symbols in each core were used to train the FNNs, as it is higher than the minimum required number. It is computationally fast to process and does not have a significant impact on the training duration.

4.4.3 Product between the skew and symbol rate

One of the most important parameters for the system performance with the use of ANNs is related to the MCF characteristics. As mentioned in chapter 2, the skew can determine the behavior of the ICXT. The skew is the relative delay between cores and the product between the skew and the symbol rate (skew×Rs) affects the behavior of ICXT [29]. If the skew×symbol rate is much bigger than one, several QAM signals at the interfering core may affect a symbol at the interfered core and the ICXT can be treated as Gaussian noise. On the other hand, if it is much smaller than one, the ICXT can be treated as a static component coupled to the signal, *i.e.*, the ICXT induced in a given symbol of the interfered core only depends on the symbol transmitted in the interfering core at the same time instant. In this work, we consider a point-to-point link as it is possible to retrieve the information of the signal at the output of the KK receiver at the interfered and interfering cores. In a network scenario, it would be necessary to consider different strategies to retrieve all the information required for the NNs to perform the same as in a point-to-point configuration.

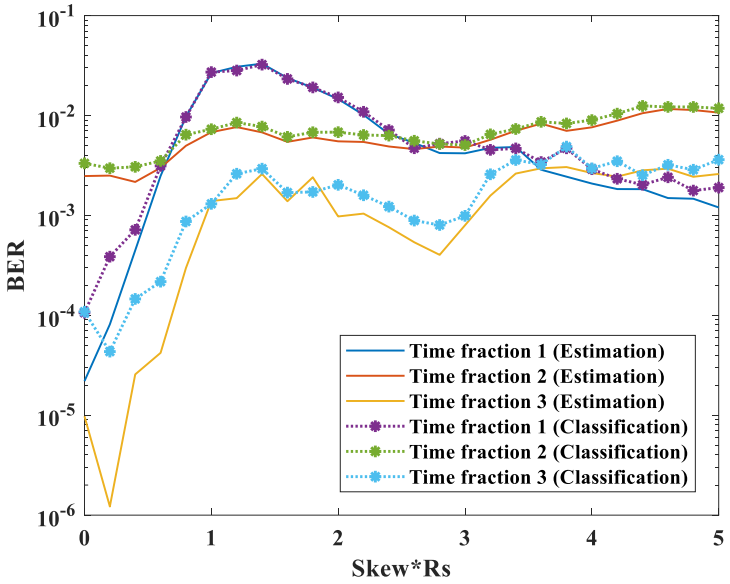


Figure 29 - BER as a functions of $skew \times Rs$.

Figure 29 depicts the BER as a function of the product between the skew and the symbols rate. For each skew, the same RPSs are used to generate the ICXT in order to guarantee the same conditions for every skew in a given time fraction. The results show that the NNs perform better when skew×symbol rate is lower than one. Thus, when skew×symbol rate is higher than one, the ICXT can be seen as additive Gaussian noise, and the ANNs is unable to learn the ICXT behavior. In the following, we will focus our attention on the case skew×symbol rate<<1.

Table 5- FNN parameters.

Parameters	**Value**
Number of inputs	4
Number of hidden layers	1
Number of neurons per hidden layer	10
Number of outputs	2 (estimation) or 16 (classification)
Training symbols	10000
Symbols to estimate the BER	10^6

At this point, the configuration of the FNNs is fully established. Table 4 outlines the parameters used by the FNNs.

4.5 Outage probability

The ICXT can cause the system to operate at a high BER during several minutes or even hours. This means that the BER can be higher than the FEC BER threshold along a significant time interval. When this happens, the system operates with an intolerant number of errors which may lead to system shoutdown or outage, as the FEC can not correct the errors. Due to this situation, the outage probability (OP) is introduced to evaluate the probability of the system being unavailable. In this work, an outage is considered when the BER in a given time fraction is higher than $10^{-1.8}$, which corresponds to the BER threshold of a 20% FEC [29]. To estimate the OP, the simulation runs until the number of outage occurrences (BER>$10^{-1.8}$) is one hundred, then it divides by the total number of time fractions simulated. In other words, the OP is given by

$$OP = \frac{100}{N}, \tag{4.30}$$

where N is the total number of simulated time fractions before reaching 100 outage occurrences.

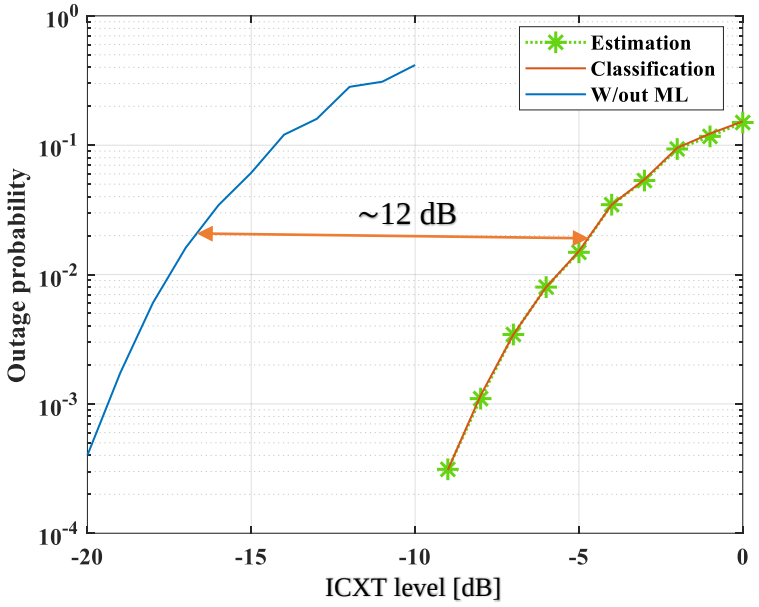
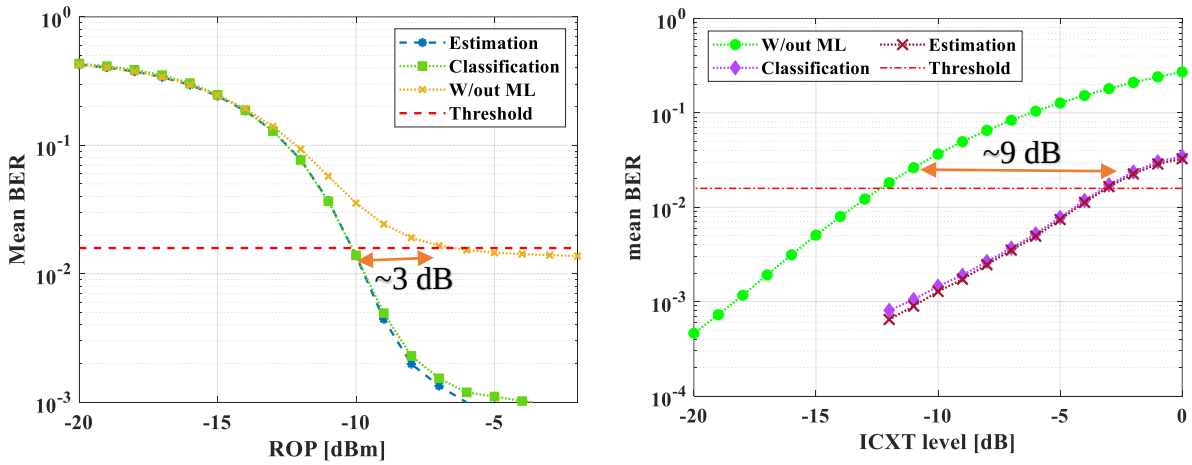


Figure 30 - Outage probability as a function of the ICXT level. The absence of ML techniques and the use of estimation and classification FNN are considered.

Figures 30 depicts the outage probability as a function of the ICXT level (XT) when employing the two ML algorithms and without the use of ML. The simulation considered a 35 km link and all the link design is the same for the three presented scenarios (see in tables 2, 3 and 4). The results only present a minimum OP value of the order of 10^{-3} because the simulation for OP is computationally heavy and it can take several days or even months to calculate the OP when the XT decreases. This scenario is more aggravated when the ML techniques are used as for each time fraction, two ICXT must be generated and each FNN has its own training process. For instance, it takes 5 times more to obtain the outage when using ML techniques than to obtain the outage in absence of ML. Figure 30 shows that, compared with the case in which ML is not employed, an improvement of approximately 12 dB is achieved by using estimation or classification FNN. With these results it is reasonable to conclude that it is possible to employ a >200Gb/s short-reach MCF based system with almost no outage occurrences if the XT is kept low enough (below -20 dB).



a) b)

Figure 31 - Mean BER as a function of a) the received optical power and b) mean BER as a function of the ICXT level.

Figure 31 a) shows the mean BER as a function of the received optical power (ROP) considering an ICXT level of −15 dB. The BER is averaged over 1000 time fractions to obtain stabilized mean BER estimates. The results show that, for the BER = $10^{-1.8}$ (corresponding to the 20% FEC threshold), the FNNs provide an additional ROP tolerance of almost 3 dB compared with the case in which ML is not employed. It is also shown that the BER = $10^{-1.8}$ is attained for

ROP=−10 dBm. If we consider the typical power levels launched into the optical fiber (between 0 and 10 dBm), this ROP level means a link budget between 10 and 20 dB. This link budget enables us to use recently fabricated MCFs where fiber loss does not exceed 0.2 dB/km and fan-in/fan-out insertion losses are typically below 1 dB [58, 59]. For ROPs below −14 dBm, the FNNs does not provide performance improvement as the system is limited by the thermal noise. Figure 31 b) shows the mean BER as a function of ICXT field and as expected, the FNNs provide a significantly improvement when comparing to the system without the use of ML. The results show that when employing ML, the system present a tolerance of approximately 9 dB when compared to the system without the use of ML.

4.6 Summary

In this chapter the self-coherent >200 Gb/s MCF system employing ML for performance improvement has been implemented, and the system performance was evaluated with and without the use of ML. In section 4.1 the validation of the ICXT model has been made. Section 4.2 presented the impact of the ICXT on the performance of a 35 km MCF based system. In section 4.3, machine learning techniques have been presented as a solution to mitigate the effects of the random ICXT. First, simpler algorithms such as *k*-means clustering and KNN have been proposed and implemented. Then, in section 4.4 more advanced ML techniques have been proposed to mitigate the impact of the ICXT. In this section, two different FNNs algorithms have been presented (estimation and classification) and they show a significant improvement on the performance of the system when using 4 features for training. In section 4.5, the optimization of the proposed FNNs has been performed according to number of neurons, training symbols and the skew × symbol rate.

Finally, the system outage has been studied in section 4.6. In this section the outage probability is presented as a function of ICXT level, and it has been proved that the FNNs can significantly improve the system outage. It was possible to achieve an outage probability of approximately 12 dB when employing FNN. The system operating with the FNN achieves and ROP improvement of 3 dB and ICXT level improvement of 9 dB.

CHAPTER 5

Conclusion and future work

In this chapter, the conclusions of this dissertation are presented and then some future works are proposed.

5.1 Final conclusion

In this work, a >200Gb/s MCF short-reach network employing KK receiver and ML was studied, and the impact of the ICXT induced by the MCF on the system performance was analyzed. The proposed MCF based system employs an ideal KK receiver that takes into account the thermal noise, chromatic dispersion, random rotation of polarization and ICXT. Furthermore, the system was designed so that the main impairment is the ICXT, and then the ML techniques were implemented as a solution to mitigate the effects of ICXT.

In chapter 2, the theoretical concepts used in this work were presented. First, an overall outline of the optical communications was presented for a better understanding of the current situation these networks. Then the MCF was presented as a solution to overcome the capacity crunch on inter data center links. Data centers use DD based receiver as cost-efficient solution. These receivers lead to nonlinear limitations, thus KK receiver was proposed as a solution for linearization, capacity increase and spectral efficiency. Furthermore, ML was proposed as a solution to mitigate the impact of the ICXT on the system performance and some previous works using ML in optical communications systems were detailed.

Chapter 3 presented the system's model and then a detailed description of the different system blocks was performed. The system comprises an optical transmitter capable of transmitting QAM signal, as the KK receiver allows the reception of QAM signals employing DD. To describe the ICXT induced by the MCF, the DP-DCM was presented, and then the KK receiver was fully described. Furthermore, the performance of the system without ICXT was assessed in order to choose the better parameters for the best performance on the system base model.

In chapter 4, the statistic of the ICXT was analyzed in order to validate the DP-DCM implemented in the simulator. Then the impact of the ICXT on the BER was evaluated, and high

BER variations was highlighted as a serious problem on the link. The ICXT varies from time fractions to time fractions, and in some time fractions, the BER can be higher than the BER limit permitted by the FEC used.

In this work, three ML techniques were proposed as a solution to mitigate the impact of the ICXT on the system performance. Simpler techniques such as *k*-means clustering and KNN were implemented. It was concluded that these techniques do not provide additional performance improvement. ANNs were also proposed to mitigate the impact of the ICXT on the system performance. First, two simpler ANN with two inputs, namely the *I* and *Q* components of the signal at the output of the KK receiver of the interfered core, were implemented. As for the previous techniques, the performance was the same as the system without ML. Then, the next step was to increase the number of inputs, of the ANN. The ANN was fedded with four features (*I* and *Q* components of the signal received at the interfering and interfered core) to mitigate the effects of the random ICXT. The ANNs only could mitigate the impact of the ICXT when skew × symbol rate<<1.The proposed ANNs present a better result in every time fraction comparing to the system without ML. With the confirmation of the improvement on the system performance caused by the NNs, the next step was to optimize these ANNs to achieve the lowest complexity possible without compromising the performance. The estimation FNN achieves the optimum performance with 2 neurons at the hidden layer, while the classification FNN achieves the optimum performance with 10 neurons at the hidden layer. A brief study shows that the FNN techniques achieve the optimum performance using approximately 1000 samples for training.

Lastly the system OP was studied, and it was concluded that, with FNNs, the system can achieve the same OP with higher XT. It was possible to achieve an ICXT level improvement of 12 dB comparing to the case without the use of ML.

5.2 Future work

Based on the work performed in this dissertation, the following subjects are suggested for future work:

- To assess the performance of the ML in long-haul amplified links.
- To assess the performance of the ML with higher order modulations index, such as 64-QAM or even 256-QAM.
- To study the possibilities to increase the number of inputs of the ANNs to stabilize the BER at a certain value on every time fractions.

- To deepen the study of ICXT in scenarios characterized by skew×symbol rate>>1 in order to find ANN techniques able to mitigate the impact of the ICXT.

Appendix

photonics

MDPI

Article

Short-Reach MCF-Based Systems Employing KK Receivers and Feedforward Neural Networks for ICXT Mitigation

Derick Piedade [1,2,*], Tiago Alves [1,2] and Tomás Brandão [1,3]

[1] Department of Information Science and Technologies, ISCTE-Instituto Universitário de Lisboa, 1649-026 Lisboa, Portugal; tiago.manuel.alves@iscte-iul.pt (T.A.); tomas.brandao@iscte-iul.pt (T.B.)
[2] Instituto de Telecomunicacões, 1049-001 Lisboa, Portugal
[3] ISTAR-IUL, 1649-026 Lisboa, Portugal
* Correspondence: daepe@iscte-iul.pt

Abstract: This paper proposes and evaluates the use of machine learning (ML) techniques for mitigating the effect of the random inter-core crosstalk (ICXT) on 256 Gb/s short-reach systems employing weakly coupled multicore fiber (MCF) and Kramers–Kronig (KK) receivers. The performance improvement provided by the *k*-means clustering, *k* nearest neighbor (KNN) and feedforward neural network (FNN) techniques are assessed and compared with the system performance obtained without employing ML. The FNN proves to significantly improve the system performance by mitigating the impact of the ICXT on the received signal. This is achieved by employing only 10 neurons in the hidden layer and four input features for the training phase. It has been shown that *k*-means or KNN techniques do not provide performance improvement compared to the system without using ML. These conclusions are valid for direct detection MCF-based short-reach systems with the product between the skew (relative time delay between cores) and the symbol rate much lower than one ($skew \times symbol\ rate \ll 1$). By employing the proposed FNN, the bit error rate (BER) always stood below $10^{-1.8}$ on all the time fractions under analysis (compared with 100 out of 626 occurrences above the BER threshold when ML was not used). For the BER threshold of $10^{-1.8}$ and compared with the standard system operating without employing ML techniques, the system operating with the proposed FNN shows a received optical power improvement of almost 3 dB.

Keywords: short-reach systems; multicore fiber; machine learning; Kramers–Kronig receiver

Citation: Piedade, D.; Alves, T.; Brandão, T. Short-Reach MCF-Based Systems Employing KK Receivers and Feedforward Neural Networks for ICXT Mitigation. *Photonics* **2022**, *9*, 286. https://doi.org/10.3390/photonics9050286

Received: 16 February 2022
Accepted: 20 April 2022
Published: 22 April 2022

1. Introduction

Current optical fiber networks are reaching the so-called capacity crunch of 100 Tb/s per single core fiber [1]. Over the last years, the traffic in data centers has been increasing exponentially, demanding new cost-efficient solutions for short-reach optical communications [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].

SDM can be based on MCFs, where *N* independent cores provide a capacity increase of *N*-fold compared with standard single-mode fiber used in current networks. The simultaneous transmission in multiple cores of the MCF leads to ICXT, which is usually considered as the main impairment of MCF systems [5]. The ICXT varies randomly along the fiber length, time, and frequency, which may affect the system's performance. High ICXT levels have been observed over several minutes or even hours, which leads to service shutdown or outage over large time periods [6]. For minimizing the impact of the ICXT on MCF-based systems, the techniques proposed so far include adaptive modulation [7], MIMO techniques [8], or optical code division multiple access spreading technique [9]. However, these techniques only provide incremental improvements of the system performance.

Due to cost purposes, short-reach MCF-based networks should employ direct-detection (DD) receivers. However, these receivers lead to nonlinear impairments that can severely

limit the achievable capacity and reach. Thus, advanced DD receivers based on Kramers–Kronig (KK) technique have been proposed for performance improvement and complexity reduction compared with coherent detection counterpart [10]. With the KK technique, linearization of the receiver is attained, and the signal phase information can be recovered.

Recently, machine learning has been employed in optical communications to recover from nonlinear distortions, including non-Gaussian additive noise, non-white laser phase noise, and fiber nonlinearities in both IM/DD and coherent systems [11,12]. Simplest machine learning techniques include *k*-means and *k* nearest neighbor (KNN), which represent an unsupervised technique used for clustering and a supervised learning technique used for classification, respectively [13]. Feedforward neural networks (FNNs) are one of the simplest machine learning techniques where a set of perceptrons are organized into layers to form a fully connected neural network (NN). FNNs are suitable for memoryless systems and have been employed in optical fiber communications to perform equalizations, as shown in [11]. Beside FNN, more advanced NN architectures have been studied in optical communications, such as deep learning [12,14]. These more advanced NN are used to predict different strategies for routing and spectrum assignments for elastic optical networks [14]. They can also be used to mitigate the non-linearities, such as signal-to-signal beat interference caused by the square-law detection [12]. In SDM systems, machine learning techniques were proposed to support the design of crosstalk-aware schemes used for resources allocation [15] or to mitigate the impact of the crosstalk power between mode groups in mode-multiplexed M-quadrature amplitude modulation (QAM) OFDM-IM-DD systems [16]. NNs were also used to speed up coating loss estimation in heterogeneous trench-assisted MCF design [17].

In this work, *k*-means clustering and KNN, as well as a low-complexity FNN, are proposed to mitigate the effects of the random variation of the ICXT along time, induced by weakly coupled MCF in short-reach systems employing KK receivers.

2. System Modelling

Figure 1 depicts the system model considered in this work. The system is composed of the optical transmitter which transmits root-raised-cosine (RRC) pulses. Then, the signals at the output of the transmitter are launched into two different cores of an MCF: (i) core n, that is the interfered core, and (ii) core m, that is the interfering core which induces ICXT in core n. Then, the optical receiver includes a PIN photodetector, an electrical amplifier (which induces thermal noise), the RRC filter, the KK algorithm, and the ML block for ICXT mitigation. Finally, the bit error ratio (BER) is estimated using Monte Carlo simulation.

2.1. Optical Transmitter

The optical transmitter is responsible for converting the information signal from the electrical to the optical domain. First, a 16-QAM Nyquist signal with a roll-off factor of 5% and symbol rate of 64 Gbaud is generated using distinct random sequences for both components (in-phase and quadrature) of the signal. The modulator is a dual parallel Mach-Zehnder modulator (DPMZM) with the ability to modulate the I and Q components of the electrical field. This is achieved by biasing the inner MZMs in the null bias point and the outer MZM in the quadrature bias point.

Figure 2 shows the spectrum of the signal at the output of the transmitter. In Figure 2a, an illustrative spectrum is presented. The optical tone is added to the signal to fulfill the minimum phase condition for the KK receiver [10]. Figure 2b shows the power spectral density (PSD) of the signal at the transmitter output obtained by simulation. The spacing between the carrier and the signal is 5% of the signal bandwidth. This spacing was chosen to maximize the spectral efficiency without adding distortion to the received signal.

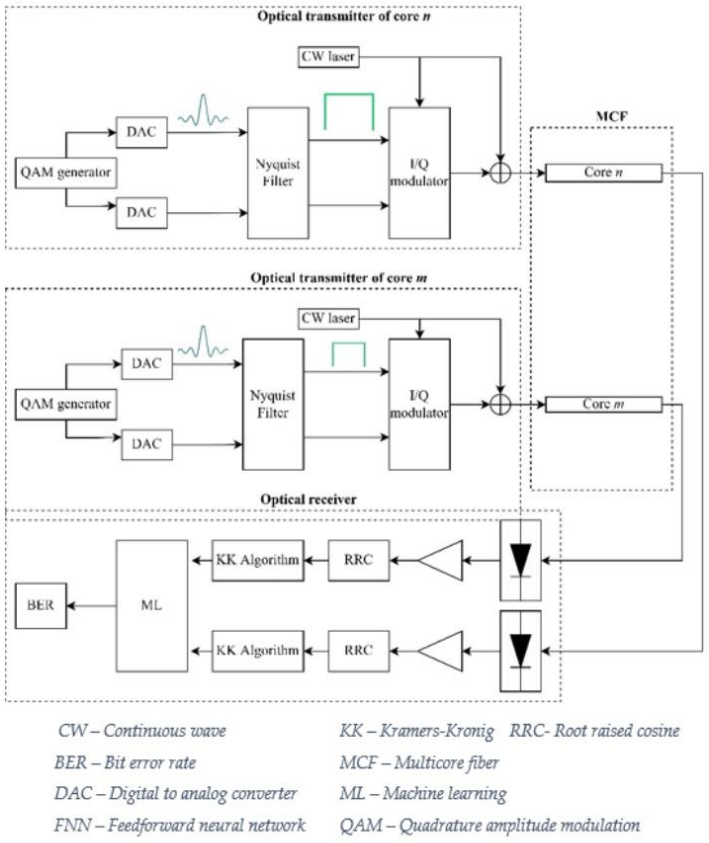


Figure 1. Schematic diagram of the short-reach MCF based system employing KK receivers.

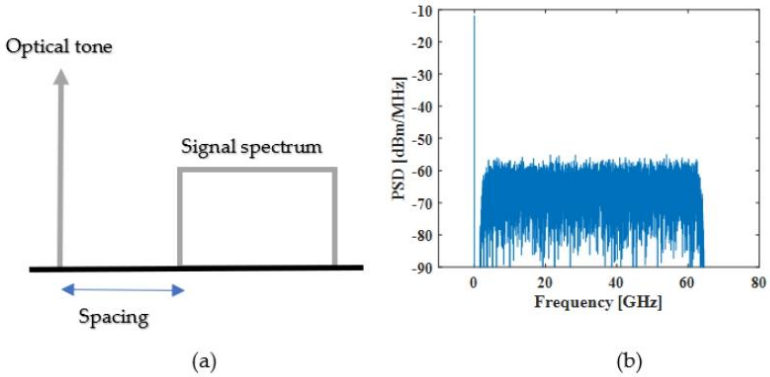


Figure 2. (**a**) Illustrative spectrum of the signal at the transmitter output. (**b**) Power spectral density obtained by simulation of the signal at the transmitter output.

2.2. MCF Model

The MCF allows the transmission of multiples optical channels in different cores. In this work, only two cores are considered: one serving as the interfering core, i.e., the core that induces the ICXT, and the other acting as the interfered core, i.e., the core impaired by the ICXT. The ICXT induced by the MCF is modeled by the dual polarization discrete changes model (DCM) [18–20]. Each core operates as a linear single mode fiber (SMF), i.e., it is modulated by the SMF propagation transfer function. The random polarization rotation induced by the fiber birefringence is also included in the transmission model [19].

To validate the ICXT simulation model, the statistical properties of the in−phase (I) and quadrature (Q) components of the ICXT obtained by simulation must agree with the theoretical analysis.

Figure 3 shows the probability density function (PDF) of the in-phase and quadrature components of the ICXT field in both polarizations (x and y) directions. The simulation results of the ICXT are obtained with 1000 random phase shifts and for a ICXT level of −15 dB. The ICXT level is the ratio between the mean ICXT power and signal power, at the output of interfered core [6]. A Gaussian PDF obtained from the mean and variance of the simulation results is also shown in Figure 3 as reference. The results of Figure 3 show that the ICXT components are well described by a Gaussian PDF, as predicted theoretically in [18], which validates the ICXT simulation model.

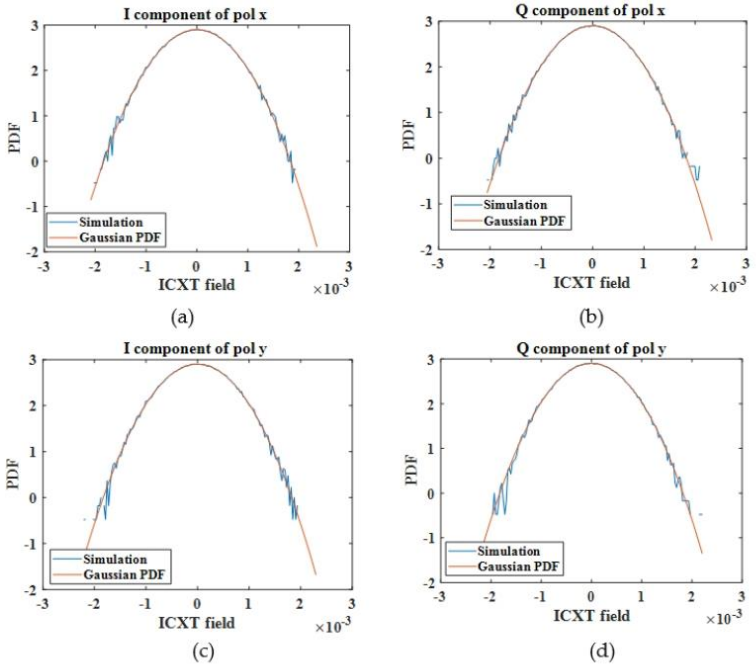


Figure 3. PDF of the quadrature components of the ICXT in x and y polarization. (**a**,**b**) ICXT fields of polarization x. (**c**,**d**) ICXT fields of polarization y.

2.3. Optical Receiver

The receiver includes a PIN photodetector, an electrical amplifier, the RRC filter, and the symbol decision circuit.

The KK algorithm is based on the Kramers−Kronig relation, and it enables to retrieve the complex field at the PIN input from the photocurrent detected by the PIN if the minimum phase condition is ensured [10,21].

Figure 4 depicts the structure of an ideal KK receiver. The main goal of this paper is to propose a simple ML algorithm to mitigate the impact of the ICXT on the system performance. For this reason, limitations due to the practical implementation of KK receivers are not addressed. After reconstructing the complex QAM signal at the KK receiver output, fiber dispersion is fully compensated using an analog filter.

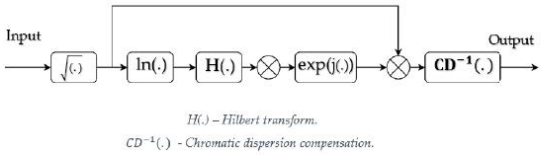


Figure 4. Schematic diagram of the KK receiver scheme.

2.4. Machine Learning

In this work, three different ML techniques are implemented and assessed: a FNN, *k*-means clustering, and KNN. These algorithms are chosen due to their low complexity and memoryless configurations, which are indicated for short-reach networks.

2.4.1. *K*-Means Clustering

In this work, the *k*-means clustering algorithm is implemented and the performance of the system is evaluated.

K-means clustering focuses on dividing the input into *k* clusters (attributed during training process) according to dissimilarity metrics and it creates groups centered on the mean of all the samples of each cluster (centroids). The algorithm uses Euclidean distance to perform the symbol decision [13]. As features for the classification, the *k*-means algorithm receives the in-phase and the quadrature components of the QAM signal transmitted in the interfered core at the end of the KK receiver.

2.4.2. *K* Nearest Neighbor

KNN algorithm is based on the distance between the new data and the neighbors counted by the *k* parameter. The neighbor symbols are attributed to the training process where every input symbol has a correspondent output that matches with a class with the same characteristics. On the active phase, a new input symbol is classified into the group of data with the greatest number of nearest neighbor symbols according to the number of neighbors defined by the *k* parameter [13]. The classification used in the algorithm is also based on the Euclidean distance.

2.4.3. Feedforward Neural Network

In this work, a shallow feedforward neural network is employed to mitigate the impact of the random ICXT on the system performance. This FNN is implemented to learn the behavior of the ICXT and then transform the ICXT-impaired received signal into a new output showing higher ICXT tolerance.

One of the main constraints of machine learning algorithms is the complexity. In this work, we are addressing MCF-based short-reach systems with a product between the skew and the symbol rate much lower than one ($skew \times symbol\ rate \ll 1$) [6]. This means that the ICXT induced in a given symbol of the interfered core only depends on the symbol transmitted in the interfering core at the same time instant [6].

Figure 5 depicts the scheme of the FNN employed in this work. The FNN is fed with four inputs (features): the in-phase and the quadrature components of the signal transmitted in the interfered and interfering cores at the end of the KK receivers. The main

target of the FNN is to predict the I and Q components of the transmitted signal at core n without the effect of the transmission impairments. As the ICXT depends on the signal injected into the interfering core m [6,22], we need to provide the signal detected at the output of core m as a training feature.

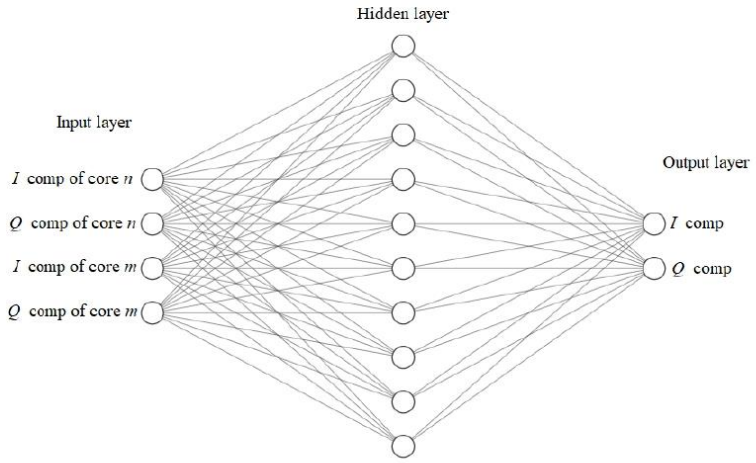


Figure 5. Schematic diagram of the FNN used to mitigate the impact of the ICXT in DD-MCF short-reach networks.

2.5. Simulation Conditions and Parameters

This subsection presents the simulation conditions used to evaluate the effectiveness of the ML techniques to mitigate the impact of the random ICXT on the system performance. This work is focused on short-reach connections, and thus, the length of the MCF does not exceed 50 km.

For the simulation, a thermal noise with a noise equivalent power (NEP) of 10 $pW/Hz^{1/2}$ is considered. To estimate the BER, Monte Carlo simulation using a bit stream with 2^{14} bits is performed. To evaluate each BER value, 100 errors are considered. All parameters used are indicated in Tables 1–3.

Table 1. Signal parameters.

Parameters	Value
Symbol rate (Gbaud)	64
Modulation	16-QAM
Roll-off factor	0.05
Input MCF power (dBm)	0
CSPR (dB)	13

Table 2. MCF parameters.

Parameters	Value
Attenuation (dB/km)	0.22
Interfered core effective refractive index	1.4453
Interfering core effective refractive index	1.4455
Wavelength (nm)	1552
Skew × bitrate	0.001
ICXT level (dB)	−13
Length (km)	35

Table 3. FNN parameters.

Parameters	Value
Number of inputs	4
Number of hidden layers	1
Number of neurons per hidden layer	10
Number of outputs	2
Training symbols	20,000
Symbols to estimate the BER	10^6

The FNN is trained using 20,000 symbols, as indicated in Table 3. These 20,000 symbols correspond to a transmission time of 0.3 µs. The complexity and online requirements of the algorithm depend on the time required by the training phase of the FNN and the update rate. The rate at which the network must be trained depends on the variation of the ICXT along time. In particular, the neural network should be trained in time intervals over which the variation of the ICXT is almost negligible. Previous works have been shown that the decorrelation time of the ICXT is of the order of a few minutes or higher [20]. This means that, if we choose to train the network once per second, to guarantee that the ICXT is constant during the active phase of the network, the training overhead is negligible. The training phase will also require some processing time to optimize the weights and biases of the FNN. Although this processing time is dependent on the real time implementation employed, we expect that it will not affect the symbol rate of the system.

To use a KK receiver, we need to ensure the minimum phase condition. Thus, the carrier-to-signal power ratio (CSPR) must be optimized according to a noise equivalent power (NEP) of 10 $pW/\sqrt{Hz}$.

Figure 6 shows the BER as a function of the CSPR for four different fiber lengths. Results are obtained considering the absence of ICXT and perfect dispersion compensation. Figure 6 shows that the optimum CSPR is approximately 13 dB. This optimum operation point results from a trade-off between signal-to-noise ratio and verification of the minimum phase condition. For low CSPR levels, the condition is not satisfied. For high CSPR levels, the SNR degrades.

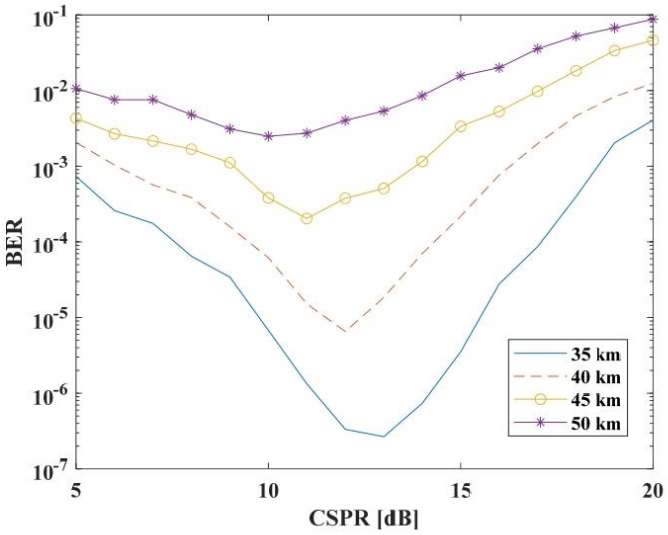


Figure 6. BER as a function of the CSPR and fiber length for 16−QAM without ICXT.

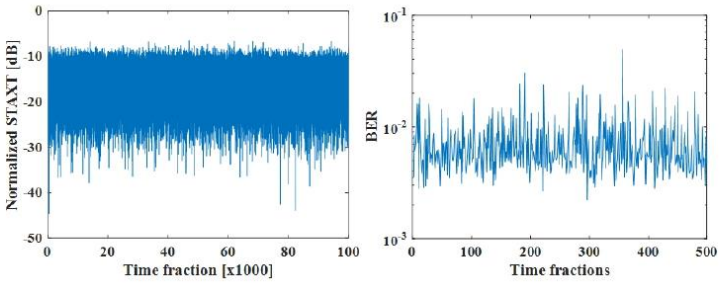
The FNN has one single hidden layer with 10 neurons. Further investigation showed that the FNN has a similar performance with number of neurons higher than 2, although the lower the number of neurons, the longer the training time. The algorithm used for training is the scaled conjugate gradient and the activation function used is the tangent sigmoid.

3. Results and Discussion

3.1. Impact of the ICXT on the System Performance

The ICXT varies randomly over time. The time variations can be of the order of a few minutes or even hours [20]. Therefore, the concept of time fraction and short-term average inter-core crosstalk (STAXT) has been introduced to assess the performance of MCF-based systems [20]. A time fraction is a small-time duration much shorter than the ICXT decorrelation time, where the ICXT is considered constant [18]. STAXT is the average power of the ICXT measured during a time period much shorter than the ICXT decorrelation time. If the time interval between time fractions is much larger than the ICXT decorrelation time, then the ICXT varies from time fraction to time fraction. This random variation of the ICXT can cause high performance changes over time.

Figure 7 shows the impact of the ICXT on the system. In Figure 7a, the STAXT is presented as a function of the time fractions. The results show the high variation of the ICXT power at the output of the interfered core, such as reported in [5,18,20,22]. In Figure 7a,b, the ICXT level is −15 dB. Figure 7b shows the calculation of the BER along 500 different time fractions. The results show that the ICXT can cause high BER variations along time which can remain for several minutes or even hours, leading to system outage.



(**a**) (**b**)

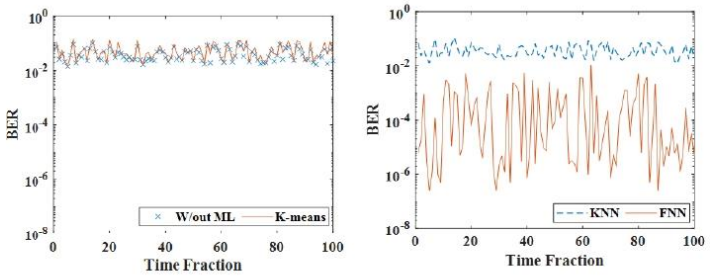
Figure 7. (**a**) Normalized STAXT as a function of time fractions. (**b**) BER as a function of time fraction.

3.2. Performance of Short-Reach MCF-Based System Employing ML Techniques

In this subsection, the BER obtained with and without using ML techniques is evaluated by simulation. The performance improvement enabled by shallow FNNs, $k-$means or KNN techniques are identified and insight about the design of ML-assisted short-reach MCF−based systems is provided.

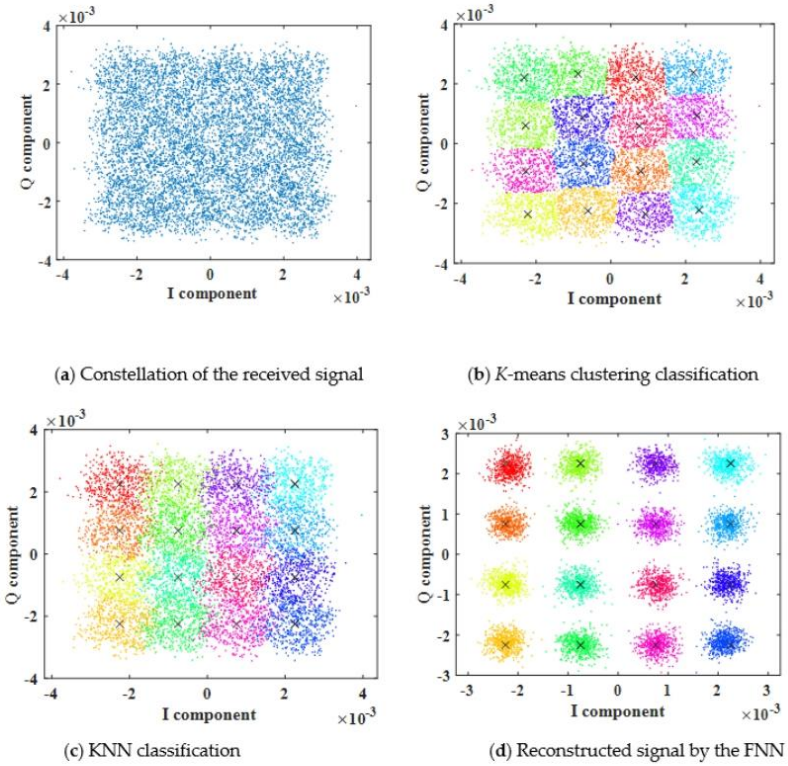
Figure 8 shows the BER of the received signal before and after using ML algorithms, considering a ICXT level of −13 dB. The results obtained show that $k-$means and KNN algorithms do not provide performance improvement when compared with the performance obtained without using ML. This occurs because the ICXT is random and the simplest redesign of the decision boundaries does not allow to better identify the different clusters, as inferred from the constellations shown in Figure 9a–c. In contrast, Figure 8 shows that the FNN enables the mitigation of the ICXT-induced BER degradation. This is confirmed by the comparison between the constellations obtained without ML and after employing the FNN, shown in Figure 9a,d, respectively. For instance, if we consider a BER threshold of $10^{-1.8}$ to define the system outage [6], then the BER before applying ML techniques presents 100 occurrences (out of 626) above the threshold (the ones shown in Figure 8). By applying the proposed FNN, the BER always stood below the threshold. We chose to represent only

the BER occurrences above the threshold in Figure 8, as the goal is to evaluate the system performance improvement provided by FNN in extreme degradation scenarios.



(**a**) (**b**)

Figure 8. BER as a function of time fractions for: (**a**) without the use of ML (crosses), $k-$means clustering (line); (**b**) KNN (dashed line), FNN (line).



(**a**) Constellation of the received signal (**b**) *K*-means clustering classification

(**c**) KNN classification (**d**) Reconstructed signal by the FNN

Figure 9. Constellation of the 16−QAM signal before and after using ML techniques.

In order to conclude the analysis of the performance improvement provided by the FNN, we also evaluated the mean BER as a function of the received optical power (ROP). Figure 10 shows the mean BER as a function of the ROP considering an ICXT level of −15 dB. The BER is averaged over 500 time fractions to obtain stabilized mean BER estimates. The results show that, for the BER = $10^{-1.8}$ (corresponding to the 20% FEC

threshold), the FNN provides an additional ROP tolerance of almost 3 dB compared with the case in which ML is not employed. It is also shown that the BER = $10^{-1.8}$ is attained for ROP = −10 dBm. If we consider the typical power levels launched into the optical fiber (between 0 and 10 dBm), this ROP level means a link budget between 10 and 20 dB. This link budget enables us to use recently fabricated MCFs where fiber loss does not exceed 0.2 dB/km and fan/in fan/out insertion losses are typically bellow 1 dB [23,24]. For ROPs below −14 dBm, the FNN does not provide performance improvement as the system is limited by the thermal noise.

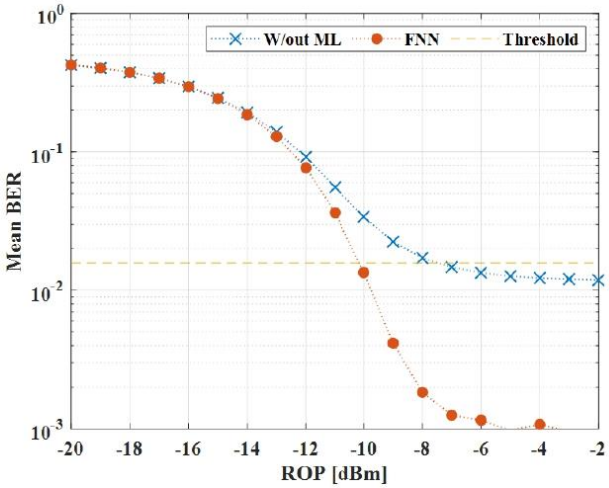


Figure 10. Mean BER as a function of the received optical power using FNN (circles) and without the use of ML (crosses).

4. Conclusions

In this paper, *k*-means, KNN, and FNN techniques are proposed and optimized to improve the tolerance of 256 Gb/s DD short−reach systems to the ICXT induced by weakly coupled MCF. It has been shown that memoryless FNNs provide significant improvement of the system performance and may represent a simple and effective solution to mitigate the impact of the ICXT induced in MCFs. This conclusion is valid for short-reach systems with $skew \times symbol\ rate \ll 1$, where the ICXT induced in the interfered core at a given time instant only depends on the signal transmitted in the interfering core at the same time instant. The *k*-means clustering and KNN techniques have shown to be useless to mitigate the ICXT. With the proposed FNN, the ICXT−impaired MCF system recovered from 100 occurrences (out of 626) with BER above the BER threshold, which is used to declare system outage, to a full operation with no outage. We also confirmed that, for a BER = $10^{-1.8}$ (20% FEC threshold), the FNN provides an additional ROP tolerance of almost 3 dB when compared with the system without the use of ML.

Author Contributions: Conceptualization, T.A., T.B. and D.P.; methodology, T.A., T.B. and D.P.; software, D.P.; validation, T.A. and T.B.; formal analysis, T.A. and T.B.; investigation, T.A., T.B. and D.P.; resources, T.A. and D.P.; data curation, D.P.; writing—original draft preparation, D.P.; writing—review and editing, D.P., T.A. and T.B.; visualization, D.P.; supervision, T.A. and T.B.; project administration, T.A.; funding acquisition, T.A. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by Fundação para a Ciência e a Tecnologia (FCT) from Portugal under the internal projects of instituto de Telecomunicações DigCore/UIDP/50008/2020, UIDB/EEA/50008/2020, and grant BI Nº22-09-04-2021. ISTAR projects UIDB/04466/2020 and UIDP/04466/2020, and SELFIE projects are also acknowledged.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Richardson, D.J.; Fini, J.M.; Nelson, L.E. Space-division multiplexing in optical fibres. *Nat. Photonics* **2013**, *7*, 354–362. [CrossRef]
2. Perin, J.K.; Shastri, A.; Kahn, J.M. Data center links beyond 100 Gbit/s per wavelength. *Opt. Fiber Technol.* **2018**, *44*, 69–85. [CrossRef]
3. Winzer, P.J.; Neilson, D.T.; Chraplyvy, A.R. Fiber optic transmission and networking: The previous 20 and the next 20 years [Invited]. *Opt. Express* **2018**, *26*, 24190–24239. [CrossRef] [PubMed]
4. Butler, D.L.; Li, M.-J.; Li, S.; Geng, Y.; Khrapko, R.R.; Modavis, R.A.; Nazarov, V.N.; Koklyushkin, A.V. Space division multiplexing in short reach optical interconnects. *J. Light. Technol.* **2016**, *35*, 677–682. [CrossRef]
5. Alves, T.M.F.; Cartaxo, A.V.T. Theoretical modelling of random time nature of inter-core crosstalk in multicore fibers. In Proceedings of the 2016 IEEE Photonics Conference (IPC), Waikoloa, HI, USA, 2–6 October 2016; pp. 521–522. [CrossRef]
6. Alves, T.F.; Cartaxo, A.V.T.; Rebola, J.L. Stochastic properties and outage in crosstalk-impaired OOK-DD weakly-coupled MCF applications with low and high skew×bit-rate. *IEEE J. Sel. Top. Quantum Electron.* **2020**, *26*, 1–8. [CrossRef]
7. Alves, T.M.F.; Luís, R.S.; Puttnam, B.J.; Cartaxo, A.; Awaji, Y.; Wada, N. Performance of adaptive DD-OFDM multicore fiber links and its relation with intercore crosstalk. *Opt. Express* **2017**, *25*, 16017–16027. [CrossRef] [PubMed]
8. Luís, R.S.; Rademacher, G.; Puttnam, B.J.; Awaji, Y.; Wada, N. Long distance crosstalk-supported transmission using homogeneous multicore fibers and SDM-MIMO demultiplexing. *Opt. Express* **2018**, *26*, 24044–24053. [CrossRef] [PubMed]
9. Wang, Y.; Saitoh, K. Inter-core crosstalk mitigation in multicore fibers by optical CDMA spreading technique with referenced channels. In Proceedings of the Asia Communications and Photonics Conference (ACPC), OSA Technical Digest (Optica Publishing Group), Chengdu, China, 2–5 November 2019; p. M4A.177.
10. Le, S.T.; Schuh, K.; Chagnon, M.; Buchali, F.; Dischler, R.; Aref, V.; Buelow, H.; Engenhardt, K.M. 1.72-Tb/s Virtual-carrier-assisted direct-detection transmission over 200 km. *J. Light. Technol.* **2018**, *36*, 1347–1353. [CrossRef]
11. Gaiarin, S.; Pang, X.; Ozolins, O.; Jones, R.T.; Da Silva, E.P.; Schatz, R.; Westergren, U.; Popov, S.; Jacobsen, G.; Zibar, D. High Speed PAM-8 optical interconnects with digital equalization based on neural network. In Proceedings of the Asia Communications and Photonics Conference, Wuhan, China, 2–5 November 2016; pp. 1–3.
12. Xu, Z.; Sun, C.; Ji, T.; Manton, J.H.; Shieh, W. Computational complexity comparison of feedforward/radial basis function/recurrent neural network-based equalizer for a 50-Gb/s PAM4 direct-detection optical link. *Opt. Express* **2019**, *27*, 36953–36964. [CrossRef] [PubMed]
13. Rafique, D.; Velasco, L. Machine learning for network automation: Overview, architecture, and applications. *J. Opt. Commun. Netw.* **2018**, *10*, D126–D143. [CrossRef]
14. Yu, J.; Cheng, B.; Hang, C.; Hu, Y.; Liu, S.; Wang, Y.; Shen, J. A deep learning based RSA strategy for elastic optical networks. In Proceedings of the 18th International Conference on Optical Communications and Networks (ICOCN), Huangshan, China, 5–8 August 2019; pp. 1–3. [CrossRef]
15. Yao, Q.; Yang, H.; Zhu, R.; Yu, A.; Bai, W.; Tan, Y.; Zhang, J.; Xiao, H. Core, mode, and spectrum assignment based on machine learning in space division multiplexing elastic optical networks. *IEEE Access* **2018**, *6*, 15898–15907. [CrossRef]
16. Zhang, Q.W.; Liu, M.; Zhou, H.; Wang, F.; Chen, J.; Cao, B.Y.; Song, Y.X.; Zhang, J.J.; Li, Y.C.; Wang, M. A machine learning enabled optical IMDD SDM OFDM system. In Proceedings of the 23rd Opto-Electronics and Communications Conference (OECC), Jeju, Korea, 2–6 July 2018; pp. 1–2. [CrossRef]
17. Mu, X.; Ottino, A.; Ferreira, F.M.; Zervas, G. Optimization of 125-µm Heterogeneous multi-core fibre design using artificial intelligence. *IEEE J. Sel. Top. Quantum Electron.* **2021**, *28*, 1–13. [CrossRef]
18. Soeiro, R.O.J.; Alves, T.M.F.; Cartaxo, A.V.T. Dual polarization discrete changes model of inter-core crosstalk in multi-core fibers. *IEEE Photonics Technol. Lett.* **2017**, *29*, 1395–1398. [CrossRef]
19. Pinheiro, B.R.P.; Rebola, J.L.L.; Cartaxo, A.V.T. Analysis of inter-core crosstalk in weakly-coupled multi-core fiber coherent systems. *J. Light. Technol.* **2021**, *39*, 42–54. [CrossRef]
20. Alves, T.M.F.; Cartaxo, A.V.T. Characterization of the stochastic time evolution of short-term average intercore crosstalk in multicore fibers with multiple interfering cores. *Opt. Express* **2018**, *26*, 4605–4620. [CrossRef] [PubMed]

21. Honda, E.; Mori, Y.; Hasegawa, H.; Sato, K.-I. Intra- and inter-datacentre converged networks utilising Kramers-Kronig receivers. In Proceedings of the 45th European Conference on Optical Communication (ECOC), Dublin, Ireland, 22–26 September 2019; pp. 1–4. [CrossRef]
22. Luis, R.S.; Puttnam, B.J.; Cartaxo, A.; Klaus, W.; Mendinueta, J.M.D.; Awaji, Y.; Wada, N.; Nakanishi, T.; Hayashi, T.; Sasaki, T. Time and modulation frequency dependence of crosstalk in homogeneous multi-core fibers. *J. Light. Technol.* **2016**, *34*, 441–447. [CrossRef]
23. Matsui, T.; Sagae, Y.; Sakamoto, T.; Nakajima, K. Design and applicability of multi-core fibers with standard cladding diameter. *J. Light. Technol.* **2020**, *38*, 6065–6070. [CrossRef]
24. Kopp, V.; Park, J.; Singer, J.; Neugroschl, D.; Gillooly, A. Low return loss multicore fiber-fanout assembly for SDM and sensing applications. In Proceedings of the Optical Fiber Communications Conference and Exhibition (OFC), San Diego, CA, USA, 8–12 March 2020; pp. 1–3.

Bibliography

[1] J. K. Perin, A. Shastri, and J. M. Kahn, "Data center links beyond 100 Gbit/s per wavelength," *Optical Fiber Technology*, vol. 44, pp. 69–85, December 2018, doi: 10.1016/j.yofte.2017.12.006.

[2] P. J. Winzer, D. T. Neilson, and A. R. Chraplyvy, "Fiber-optic transmission and networking: The previous 20 and the next 20 years [Invited]," *Optics Express*, vol. 26, no. 18, pp. 24190, 2018, doi: 10.1364/oe.26.024190.

[3] S. T. Le *et al.*, "1.72-Tb/s virtual-carrier-assisted direct-detection transmission over 200 km," *Journal of Lightwave Technology*, vol. 36, no. 6, pp. 1347–1353, 2018, doi: 10.1109/JLT.2017.2779331.

[4] R. S. Luís, G. Rademacher, B. J. Puttnam, C. Antonelli, S. Shinada, and H. Furukawa, "Simple method for optimizing the DC bias of Kramers-Kronig receivers based on AC-coupled photodetectors," *Optics Express*, vol. 28, no. 3, pp. 4067, 2020, doi: 10.1364/oe.383369.

[5] 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, 2019, doi: 10.1364/oe.27.036953.

[6] J. J. G. Torres *et al.*, "Adaptive nonsymmetrical demodulation based on machine learning to mitigate time-varying impairments," *IEEE Avionics and Vehicle Fiber-Optics and Photonics Conference* (AVFOP), vol. 3, pp. 289–290, 2016, doi: 10.1109/AVFOP.2016.7789906.

[7] D. L. Butler *et al.*, "Space division multiplexing in short reach optical interconnects," *Journal of Lightwave Technology*, vol. 35, no. 4, pp. 677–682, 2017, doi: 10.1109/JLT.2016.2619981.

[8] K. Kikuchi, "Fundamentals of coherent optical fiber communications," *Journal of Lightwave Technology*, vol. 34, no. 1, pp. 157-179, January 2016, doi: 10.1109/JLT.2015.2463719.

[9] B. Bouzid, “Analysis and review of erbium doped fiber amplifier,” *Saudi International Electronics and Communications and Photonics Conference*, pp. 1-5, 2013, doi: 10.1109/SIECPC.2013.6551005.

[10] S. Bigo, “Coherent detection: A key enabler for next-generation optical transmission systems?”, *9th International Conference on Transparent Optical Networks*, pp. 332-335, 2007, doi: 10.1109/ICTON.2007.4296102.

[11] K. Kikuchi, “Coherent optical communications —History, state-of-the-art technologies, and challenges for the future—,” *Joint Conference of the OptoElectronics and Communications Conference and the Australian Conference on Optical Fibre Technology*, pp. 1-4, 2008, doi: 10.1109/OECCACOFT.2008.4610574.

[12] J. Renaudier *et al.*, “First 100-nm continuous-band WDM transmission system with 115 Tb/s transport over 100 km using novel ultra-wideband semiconductor optical amplifiers,” *European Conference on Optical Communication* (ECOC), pp. 1-3, 2017, doi: 10.1109/ECOC.2017.8346084.

[13] O. Lim, V. Grigoryan, M. Shin and P. Kumar, “Ultra-low-noise inline fiber-optic phasesensitive amplifier for analog optical signals,” *Conference on Optical Fiber Communication/National Fiber Optic Engineers Conference*, pp. 1-3, 2008, doi: 10.1109/OFC.2008.4528359.

[14] K. Enbutsu, T. Umeki, O. Tadanaga, H. Takenouchi and M. Asobe, "PPLN-based low-noise in-line phase sensitive amplifier with highly sensitive carrier-recovery system," *21st Asia-Pacific Conference on Communications (APCC)*, pp. 666-669, 2015, doi: 10.1109/APCC.2015.7412591.

[15] T. Kazama *et al*., “Low-noise phase sensitive amplification using an optical phase-locked pump,” *24th OptoElectronics and Communications Conference* (OECC) and *International Conference on Photonics in Switching and Computing* (PSC), pp. 1-3, 2019, doi: 10.23919/PS.2019.8818030.

[16] Y. Yamamoto, Y. Kawaguchi and M. Hirano, “Low-loss and low-nonlinearity pure-silicacore fiber for C- and L-band broadband transmission,” *Journal of Lightwave Technology*, vol. 34, no. 2, pp. 321-326, January 2016, doi: 10.1109/JLT.2015.2476837.

[17] M. Sharif, J. K. Perin, and J. M. Kahn, “Modulation schemes for single-laser 100 Gb / s links: Single-carrier,” vol. 33, no. 20, pp. 4268–4277, October 2015.

[18] K. Saitoh, “Multicore fiber technology,” *Optical Fiber Communications Conference* (OFC), vol. 34, no. 1, pp. 55–66, 2015, doi: 10.1364/ofc.2015.th4c.1.

[19] H. R. H. Andersson, A. De Blanche, and T. Lundqvist, “Flipping the data center network: Increasing east-west capacity using existing hardware,” *Conference on Local Computer Networks*, (LCN), pp. 211–214, October 2017, doi: 10.1109/LCN.2017.92.

[20] Y. Zong *et al*., “Location-aware energy efficient virtual network embedding in software defined optical data center networks,” *Journal of Optical Communications and Networking*, vol. 10, no. 7, pp. B58–B70, 2018, doi: 10.1364/JOCN.10.000B58.

[21] K. Shibahara, T. Mizuno, D. Lee and Y. Miyamoto, “Advanced MIMO signal processing for dense SDM transmission using multi-core few-mode fibers,” *21st OptoElectronics and Communications Conference* (OECC) held jointly with *International Conference on Photonics in Switching* (PS), pp. 1-3, 2016.

[22] S. Randel *et al*., “MIMO-based signal processing of spatially multiplexed 112-Gb/s PDMQPSK signals using strongly-coupled 3-core fiber,” *37th European Conference and Exhibition on Optical Communication*, pp. 1-3, 2011.

[23] R. Ryf *et al*., “MIMO-based crosstalk suppression in spatially multiplexed 3X56-Gb/s PDM-QPSK signals for strongly coupled three-core fiber,” *IEEE Photonics Technology Letters*, vol. 23, no. 20, pp. 1469-1471, October 2011, doi: 10.1109/LPT.2011.2162826.

[24] B. J. Puttnam *et al.*, “Homogeneous, single-mode MCF transmission,” *IEEE Photonics Society Summer Topical Meeting Series* (SUM), vol. 1, pp. 224–225, 2016, doi: 10.1109/PHOSST.2016.7548811.

[25] T. Shi, T. Su, N. Zhang, C. Hong and D. Pan, “Silicon photonics platform for 400G data center applications,” *Optical Fiber Communications Conference and Exposition* (OFC), pp. 1-3, 2018.

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

[27] 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, 2018, doi: 10.1364/oe.26.004605.

[28] A. V. T. Cartaxo and T. M. F. Alves, "Analytical model of inter-core crosstalk of real homogeneous multi-core fibers," *IEEE Photonics Conference* (IPC), pp. 519-520, 2016, doi: 10.1109/IPCon.2016.7831209.

[29] 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, pp. 1-8, July-August 2020, doi: 10.1109/JSTQE.2020.2995306.

[30] J. L. Rebola, T. M. F. Alves and A. V. T. Cartaxo, "Outage probability due to intercore crosstalk from multiple cores in short-reach networks," *IEEE Photonics Technology Letters*, vol. 33, no. 6, pp. 281-284, March 2021, doi: 10.1109/LPT.2021.3056956.

[31] T. M. F. Alves, J. L. Rebola and A. V. T. Cartaxo, "Outage probability due to intercore crosstalk in weakly-coupled MCF systems with OOK signaling," *Optical Fiber Communications Conference and Exhibition* (OFC), pp. 1-3, 2019.

[32] A. V. T. Cartaxo and T. M. F. Alves, "Theoretical characterization of the decorrelation bandwidth of intercore crosstalk in weakly-coupled multicore fibers*," European Conference on Optical Communication* (ECOC), pp. 1-3, 2018, doi: 10.1109/ECOC.2018.8535417.

[33] T. M. F. Alves and A. V. T. Cartaxo, "Decorrelation bandwidth of intercore crosstalk in weakly coupled multicore fibers with multiple interfering cores*," Journal of Lightwave Technology*, vol. 37, no. 3, pp. 744-754, February 2019, doi: 10.1109/JLT.2018.2880076.

[34] J. M. D. Mendinueta *et al.*, "High-capacity super-channel-enabled multi-core fiber optical switching system for converged inter/intra data center and edge optical networks," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 26, no. 4, pp. 1-13, July-August 2020, doi: 10.1109/JSTQE.2020.2969558.

[35] S. K. Salim and R. S. Fyath, "Performance investigation of multicore optical interconnects for data centers," *International Conference on Computer Science and Software Engineering* (CSASE), pp. 20-25, 2020, doi: 10.1109/CSASE48920.2020.9142110.

[36] A. Mecozzi and C. Antonelli, "Kramers – Kronig coherent receiver", *Optica 3*, vol. 3, no. 11, 2016.

[37] E. Honda, Y. Mori, H. Hasegawa, and K. I. Sato, “Intra- and inter-data center converged networks utilizing Kramers-Kronig receivers,” *45th European Conference on Optical Communication* (ECOC), pp. 1–4, 2019, doi: 10.1049/cp.2019.0966.

[38] C. Antonelli, A. Mecozzi, and M. Shtaif, “Kramers Kronig PAM transceiver and two-sided polarization-multiplexed Kramers Kronig transceiver,” *Journal of Lightwave Technology*, vol. 36, no. 2, pp. 468–475, 2017.

[39] Y. Zhou *et al.*, “160 Gb/s 256QAM transmission in a 25 GHz grid using Kramers-Kronig detection,” *Optical Fiber Communications Conference and Exhibition* (OFC), pp. 1-3. 2019.

[40] D. Kong *et al.*, “909.5 Tbit/s dense SDM and WDM transmission based on a single source optical frequency comb and Kramers-Kronig detection,” *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 27, no. 2, pp. 1-8, March-April 2021, no. 2100108, doi: 10.1109/JSTQE.2020.3024004.

[41] F. Musumeci *et al.*, “An overview on application of machine learning techniques in optical networks,” *IEEE Communications Surveys & Tutorials*, vol. 21, no. 2, pp. 1383–1408, 2019.

[42] 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, 2019, doi: 10.1109/JLT.2019.2897313.

[43] S. Gaiarin *et al.*, “High speed PAM-8 optical interconnects with digital equalization based on neural network,” *Asia Communications and Photonics Conference* (ACP), pp. 1-3, 2016.

[44] D. Rafique and L. Velasco, “Machine learning for network automation: overview, architecture, and applications,” *Journal of Optical Communications and Networking*, vol. 10, no. 10, pp. D126–D143, 2018, doi: 10.1364/JOCN.10.00D126.

[45] D. Wang *et al.*, “KNN-based detector for coherent optical systems in presence of nonlinear phase noise,” *21st Optoelectronics and Communications Conference* (OECC) held Jointly with *International Conference Photonics Switch* (PS), pp. 18–20, 2016.

[46] J. Yu *et al.*, “A deep learning based RSA strategy for elastic optical networks,” *18th International Conference on Optical Communications and Networks* (ICOCN), pp. 1-3, 2019, doi: 10.1109/ICOCN.2019.8934862.

[47] Q. Yao *et al.*, "Core, mode, and spectrum assignment based on machine learning in space division multiplexing elastic optical networks," *IEEE Access*, vol. 6, pp. 15898-15907, 2018, doi: 10.1109/ACCESS.2018.2811724.

[48] Q. W. Zhang *et al.*, "A machine learning enabled optical IMDD SDM OFDM system," *23rd OptoElectronics and Communications Conference* (OECC), pp. 1-2, 2018, doi: 10.1109/OECC.2018.8729701.

[49] X. Mu *et al.*, "Optimization of 125-μm heterogeneous multi-core fibre design using artificial intelligence," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 28, no. 4, pp. 1-13, July-August 2022, doi: 10.1109/JSTQE.2021.3104821.

[50] B. Karanov *et al.*, "End-to-end deep learning of optical fiber communications," *Journal of Lightwave Technology*, vol. 36, no. 20, pp. 4843–4855, 2018, doi: 10.1109/JLT.2018.2865109.

[51] Z. Xu, C. Sun, T. Ji, H. Ji and W. Shieh, "Cascade recurrent neural network enabled 100-Gb/s PAM4 short-reach optical link based on DML," *Optical Fiber Communications Conference and Exhibition* (OFC), pp. 1-3, 2020.

[52] I. Lyubomirsky, "Machine learning equalization techniques for highspeed PAM4 fiber optic communication systems", *CS229 Final Project Report, Stanford University*, 2015. Available: http://cs229.stanford.edu/proj2015/232_report.pdf.

[53] M. Šlapák, J. Vojtěch, O. Havliš and R. Slavík, "Monitoring of fibre optic links with a machine learning-assisted low-cost polarimeter," *IEEE Access*, vol. 8, pp. 183965-183971, 2020, doi: 10.1109/ACCESS.2020.3009524.

[54] 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, August 2017, doi: 10.1109/LPT.2017.2723662.

[55] 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, January 2021, doi: 10.1109/JLT.2020.3024609.

[56] R. S. Luís *et al*., "Time and modulation frequency dependence of crosstalk in homogeneous multi-core fibers," *Journal of Lightwave Technology*, vol. 34, no. 2, pp. 441-447, January 2016, doi: 10.1109/JLT.2015.2474128.

[57] G. Bebis and M. Georgiopoulos, “Feed-forward neural networks,” in *IEEE Potentials*, vol. 13, no. 4, pp. 27-31, October-November 1994, doi: 10.1109/45.329294.

[58] T. Matsui *et al.*, “Design and applicability of multi-core fibers with standard cladding diameter,” *Journal of Lightwave Technology*, vol. 38, no. 21, pp. 6065-6070, November 2020.

[59] V. I. Kopp *et al.*, “Low return loss multicore fiber-fanout assembly for SDM and sensing applications,” *Optical Fiber Communications Conference and Exhibition* (OFC), pp. 1-3, 2020.