UNIVERSITY
INSTITUTE
OF LISBON

CNN-based eye pattern analysis and BER prediction in PAM4 inter-datacenter optical connections impaired by intercore crosstalk

Sofia Pérsio Eugénio Esteves

Master Degree in Telecommunications and Computer Engineering

Supervisor
PhD João Lopes Rebola, Associate Professor
Iscte-IUL

Co-Supervisor
PhD Pedro Figueiredo Santana, Assistant Professor
Iscte-IUL

November, 2021

TECNOLOGIAS
E ARQUITETURA

Department of Information Science and Technology

CNN-based eye pattern analysis and BER prediction in PAM4 inter-datacenter optical connections impaired by intercore crosstalk

Sofia Pérsio Eugénio Esteves

Master Degree in Telecommunications and Computer Engineering

Supervisor
PhD João Lopes Rebola, Associate Professor
Iscte-IUL

Co-Supervisor
PhD Pedro Figueiredo Santana, Assistant Professor
Iscte-IUL

November, 2021

Acknowledgments

I would like to thank to my supervisors Professor João Rebola and Professor Pedro Santana for their guidance and patience during the development of this dissertation. I also want to thank to Iscte-Iul, Instituto de Telecomunicações and Information Sciences and Technologies and Architecture Research Centre for providing access to their resources.

To my friends, thank you for the conscious decision of sharing your time with me.

To my family, words cannot express how deeply and truly grateful I am for all your love, guidance and support.

Resumo

De modo a colmatar a necessidade de fornecer largura de banda suficiente para atingir altas taxas de tráfego de dados em ligações entre centros de dados, foi proposta a transmissão de sinais com modulação de impulsos em amplitude com 4 níveis (PAM4) em ligações de curto alcance entre centro de dados com modulação de intensidade e deteção direta suportadas por fibras homogéneas multinúcleo fracamente acopladas. No entanto, neste tipo de fibras, a diafonia entre núcleos (ICXT) limita significativamente o desempenho das ligações, causando grandes flutuações da taxa de erros binários (BER), o que pode conduzir à indisponibilidade da ligação.

Neste trabalho, através da análise de diagramas de olho usando uma rede neuronal convolucional (CNN) é estimada a BER em ligações ópticas entre centros de dados PAM4 degradadas por ICXT com o objetivo de monitorização do desempenho. Para avaliar o desempenho da CNN é usada como métrica a raiz do erro quadrático médio (RMSE).

Para diferentes atrasos de propagação entre núcleos, razões de extinção e níveis de diafonia, a CNN é capaz de prever BERs sem ultrapassar o limite estabelecido para o RMSE. As CNNs treinadas com diferentes parâmetros ópticos obtiveram o melhor desempenho em termos de generalização em comparação com CNNs treinadas com parâmetros ópticos específicos. Estes resultados confirmam que os modelos baseados em CNN são capazes de extrair informação a partir de imagens de diagramas de olhos, prevendo a BER sem conhecimento prévio dos sinais transmitidos.

Palavras-chave: aprendizagem automática, crosstalk entre núcleos, fibras multinúcleo, rede neuronal convolucional, taxa de erros binários.

Abstract

To meet the required future challenge of providing enough bandwidth to achieve high data traffic rates in datacenter links, four-level pulse amplitude modulation (PAM4) signals transmission in short-haul intensity modulation-direct detection datacenters connections supported by homogeneous weakly-coupled multicore fibers has been proposed. However, in such fibers, a physical effect known as inter-core crosstalk (ICXT) limits significantly the performance of short-reach connections by causing large bit error rate (BER) fluctuations that can lead undesirable system outages.

In this work, a convolutional neural network (CNN) is proposed for eye-pattern analysis and BER prediction in PAM4 inter-datacenter optical connections impaired by ICXT, with the aim of optical performance monitoring. The performance of the CNN is assessed using the root mean square error (RMSE).

Considering PAM4 interdatacenter links with one interfering core and for different skew-symbol rate products, extinction ratios and crosstalk levels, the results show that the implemented CNN is able to predict the BER without surpassing the RMSE limit. The CNNs trained with different optical parameters obtained the best performance in terms of generalization comparing to CNNs trained with specific optical parameters. These results confirm that the CNN-based models can be able to extract features from received eye patterns to predict the BER without prior knowledge of the transmitted signals.

Keywords: bit error rate, convolutional neural network, inter-core crosstalk, machine learning, multicore fiber.

Contents

List of Figures

List of Tables

List of Acronyms

AWGN	Additive White Gaussian Noise
B2B	Back-to-Back
BER	Bit Error Rate
CD	Chromatic Dispersion
CNN	Convolutional Neural Network
DP-DCM	Dual Polarization-Discrete Changes Model
DSP	Digital Signal Processing
FEC	Forward Error Correction
FMF	Few-Mode Fiber
ICXT	Inter-Core Crosstalk
IM-DD	Intensity Modulation-Direct Detection
IP	Internet Protocol
MCF	Multicore Fiber
MF	Modulation Format
MIMO	Multiple Input-Multiple Output
MMF	Multimode Fiber
OOK	On-Off Keying
OP	Outage Probability
OPM	Optical Performance Monitoring
PAM4	Four-level Pulse Amplitude Modulation
PDI	Propagation-Direction Interleaving
PMD	Polarization Mode Dispersion
RL	Reinforcement Learning
RSA	Resource Allocation Algorithms
SC-SMF	Single-Core Single-Mode Fiber
SCC	Spatial Channel Count
SDM	Space Division Multiplexing

SL	Supervised Learning
ToR	Top-of-Rack
UL	Unsupervised Learning
WC-MCF	Weakly-Coupled Multicore Fiber
WDM	Wavelength Division Multiplexing

List of Symbols

α	MCF attenuation coefficient
α_m	MCF attenuation coefficient of core m
α_n	MCF attenuation coefficient of core n
α_{DCF}	DCF attenuation coefficient
$\beta_{0,l}$	Propagation constant at the carrier wavelength in core l
$\beta_{1,l}$	Inverse of the group velocity in core l
$\beta_{2,l}$	Group velocity dispersion in core l
$\beta_{3,l}$	Higher order dispersion in core l
η	Photo-detector efficiency
λ_0	Operating wavelength
ν_0	Nominal optical frequency
ω	Angular frequency
$\overline{\beta}_l$	Average of the propagation constants of polarizations x and y in core l, with l=m or n
$\overline{K}_{nm}$	Average intercore coupling coefficient
$\phi_{nm,k}^{(a,b)}$	RPS of the k-th PMP
ξ_m	Power splitting between polarizations in core m
ξ_n	Power splitting between polarizations in core n
B_0	Receiver optical filter bandwidth
$B_{e,RX}$	Receiver electrical filter bandwidth
c	Speed of light in a vacuum
$c_{F,XT}(t)$	PAM4 signal at the output of the interfered core n with ICXT
$c_F(t)$	PAM4 signal at the output of the interfered core n without ICXT
$c_{m,x}(t)$	PAM4 signal transmitted along the interfering core m in polarization x
$c_{m,y}(t)$	PAM4 signal transmitted along the interfering core m in polarization y
$c_m(t)$	PAM4 signal transmitted along the interfering core m
$c_{n,x}(t)$	PAM4 signal transmitted along the interfered core n in polarization x
$c_{n,y}(t)$	PAM4 signal transmitted along the interfered core n in polarization y

$c_n(t)$	PAM4 signal transmitted along the interfered core n
$c_{PIN}(t)$	Signal at the output of the PIN photo-detector
$c_{XT}(t)$	ICXT signal at the output of the interfering core m
$c_{XT}(t)$	Interfering signals from core m
$D_{\lambda,DCF}$	DCF chromatic dispersion parameter
$D_{\lambda,n}$	MCF chromatic dispersion parameter
d_{mn}	Walkoff between cores m and n
E_i	Collected eye pattern, with $i = 1,...N_{EP}$
f	Frequency
f_0	Optical filter lowpass equivalent center frequency
$F_{a,b}$	ICXT transfer function in polarization directions a and b
F_n	EDFA noise figure
g	Hedges' g
g_{EDFA}	EDFA gain
h	Planck constant
$H_0(f)$	Optical filter transfer function
$H_F(f)$	MCF transfer function in core n
i_l	Optical filter insertion loss in linear units
L	SM-MCF length
L_{DCF}	DCF length
N_0	Number of MCF realizations with BER above the BER limit
N_p	Number of PMPs
N_r	Number of MCF realizations
N_s	Number of samples per symbol
N_{BER}	Number of BER intervals
N_{EP}	Number of collected eye patterns
$N_{PAM4symbols}$	Number of generated PAM4 symbols in each MCF realization
N_{reg}	Number of the bits of the offset register
P_{TX}	Transmission power
$Q(x)$	Q function
q	Electron charge
r	Extinction ratio
R_s	Symbol rate

R_λ	PIN responsivity
s_p	Pooled standard deviation
S_{ASE}	ASE noise power spectral density per polarization mode
S_{mn}	Skew between core m and core n
X_c	Crosstalk level
z_k	Longitudinal coordinate of the k-th PMP
$\mathcal{F}^{-1}$	Inverse Fourier transform

CHAPTER 1

Introduction

This dissertation focuses on addressing physical level issues associated with four-level pulse amplitude modulation (PAM4) signals transmission in multicore fibers (MCF), proposed for short-haul intensity modulation-direct detection (IM-DD) datacenters connections [**1, 2, 3**]. MCFs, from the total spatial channel count (SCC) perspective, are considered as promising technologies to overcome the capacity limits of single core-single mode fiber (SC-SMF) transmissions [**4, 5, 6**]. However, in homogeneous weakly-coupled multicore fibers (WC-MCFs), a physical effect known as inter-core crosstalk (ICXT) limits significantly the performance of such short-reach connections by causing large bit error rate (BER) fluctuations that lead undesirable outage probabilities (OPs) [**7**]. To address this issue, in this work, a convolutional neural network (CNN) is used to perform eye-pattern analysis and extract predictions of the BER of PAM4 inter-datacenter optical connections impaired by ICXT. The CNN developed can be applicable for performance monitoring and, if intended, for system outage detection.

1.1. Motivation

As telecommunications technology evolved, alongside with the number of network users and devices, datacenters became crucial to handle large amounts of data due to their flexibility and scalability in computing and storage resources [**8, 9**]. Even so, the number of devices connected to internet protocol (IP) networks is estimated to be more than three times the global population by the year of 2023 [**10**], which leads to a significant global data center IP traffic growth. This requires answering technologically to the future challenge of providing enough bandwidth to achieve such high data traffic rates in datacenter links.

The current approach to deal with capacity scalability in datacenter links is using multiple wavelengths (also known as lanes, in the datacenters terminology) carrying wavelength division multiplexing (WDM) channels, where each one of these channels carries a conventional on-off keying (OOK) signal [**11**]. 200 GBd OOK per lane IM-DD transmissions has been achieved, however, this strategy is compromised, since it requires complex and expensive components and leads to a high power consumption in order to enable the upcoming data rates requirements [**12, 13**]. For this reason, PAM4 has been proposed

for datacenter connections and has already been standardized by the IEEE 802.3bs task force to enable rates of 50 and 100 Gbit/s per wavelength channel, since it minimizes the spectral efficiency and power consumption issues [**11**]. Datacenter connections are usually categorized as intra-datacenter and inter-datacenter links, in relation to their maximum range, respectively, up to 10 km and 100 km [**11**]. Nowadays, some of these connections typically rely on optical fibers such SC-SMF, however, it is expected that such fibers will no longer fulfill the future capacity demands, as transmission in SC-SMFs is approaching its limit of 100 Tb/s with coherent detection and over 200 Gbit/s per lane with IM-DD [**14, 15**].

Multicore fibers have been proposed in order to overcome this capacity limit issue. In particular, homogeneous WC-MCFs ensure the simplicity, low cost and power consumption design requirements imposed by datacenter links. Even so, transmission in such MCFs is impaired by ICXT, which due to its inherent random nature, can limit significantly the performance of such connections by causing large BER fluctuations due to the high ICXT levels that can occur, leading to undesirable system outage periods [**3, 16**].

Several implementations using machine learning (ML) have been proposed for the physical layer of optical networks, which include fiber nonlinearity mitigation, modulation format (MF) recognition and optical performance monitoring (OPM) [**17, 18**]. Regarding the latter, most recently, the use of convolutional neural networks (CNNs) has been investigated [**19, 20**]. Given the good results obtained with these techniques, in this work, for performance monitoring, a CNN is used for eye-pattern analysis and BER prediction in PAM4 inter-datacenter optical connections impaired by ICXT.

1.2. Goals

The main focus of this dissertation is to apply a CNN for performance monitoring of IM-DD PAM4 datacenter connections impaired by ICXT. The main goals are:

- Study and characterization of short-haul IM-DD datacenter optical links with PAM4 transmission and supported by homogenous WC-MCFs;
- Review of the literature in ML and ML applied to optical performance monitoring;
- Study of the DP-DCM proposed in the literature [**3, 21**] that models accurately the ICXT effect in homogeneous WC-MCFs;

- Implementation and demonstration of the effectiveness of the CNN for performance monitoring of the BER of the PAM4 datacenter connection impaired by ICXT, used as an indicator to decide if ICXT mitigation is required.

1.3. Dissertation organization

This dissertation is organized in the following chapters. A literature review is presented in Chapter 2, which addresses the fundamental concepts related to this work, namely datacenter connections, PAM4 signal transmission in datacenter links supported by MCFs impaired by ICXT and an overview on ML. In Chapter 3, the CNN implemented for eye-pattern analysis and BER prediction of PAM4 inter-datacenter optical connections is presented. The optical telecommunication system equivalent model used for collecting data is also studied, where the DP-DCM developed to characterize the ICXT in MCFs is described in detail. In Chapter 4, the performance of the CNN-based eye-pattern analysis and BER prediction is evaluated for different optical parameters such as the crosstalk level, extinction ratio and skew-symbol rate product based on the root mean square error (RMSE) metric. The final conclusions and future work are presented in Chapter 5.

1.4. Main contributions

We consider that this work has the following main contributions:

- Implementation and validation of the effectiveness of a CNN for performance monitoring of the BER and system outage detection of the PAM4 datacenter connections impaired by ICXT;
- Design and validation of a sampling proposal for eye-pattern and BER balancing;
- Implementation of an eye-pattern to image conversion based on a fixed and a dynamic scale;
- Demonstration that the BER can be extracted (through a proper pre-processing) by analysis of eye-patterns generated from synthetic data (from MATLAB);
- Demonstration that the CNN can predict the BER accurately when trained with several different optical link parameters, instead of being trained for only a specific optical link configuration.

CHAPTER 2

Literature Review

2.1. Datacenters

Datacenters are large-scale computing platforms built to handle high amounts of data by providing flexibility and scalability in computing, networking and storage resources [**8**]. Therefore, as technology evolves, alongside with the number of network users and devices [**10**], datacenters play a fundamental role in connectivity and are expected to ensure high efficiency and reliability on how they manage such significant amounts of traffic.

2.1.1. Datacenter architecture

Traditionally, a datacenter has a three-tier architecture with an access, aggregation and core tiers [**11, 22**]. As shown in Fig.2.1, servers, through Top-of-Rack packet-based switches (ToR), connect to access switches which are usually connected to two aggregation routers. These aggregation routers are then linked to core routers in the upper layer. The connections inside and between the aggregation and core tiers are set in such way to provide redundancy [**11**]. In order to ensure a higher scalability, datacenters may have more than three tiers, by expanding the aggregation tier to more than one stage [**11, 22**].

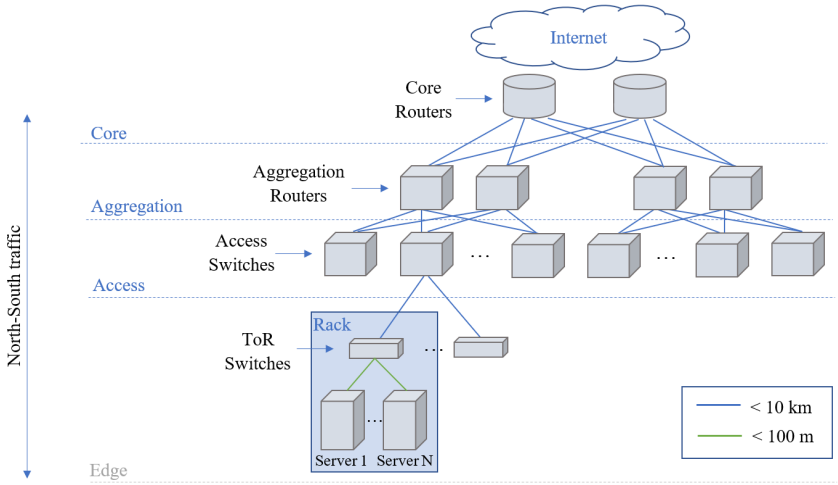


FIGURE 2.1. A three-tier data center architecture, suitable for handling north-south traffic. Blues lines represent connections with less than 10 km and green lines represent connections with less than 100 m. Based on [**8, 11, 22**].

In this architecture, traffic from the Internet passes through the core tier, then to the aggregation tier, access tier and edge layer, where the servers are located [**11**]. Here, the data is channelled to the right server. The same approach is applied in the opposite way. The traffic that travels along the datacenter in this way is named north-south traffic [**11**]. However, when traffic is generated from one server to another, within the same datacenter, in a three-tier architecture, it travels up to the core tier and then back to the edge layer, meaning that it must traverse two access switches, two aggregation routers and a core router, which increases latency and leads high power consumption [**11**]. Moreover, with cloud external providers, such as Microsoft and Amazon, hosting massive amounts of data [**23**], traffic must also travel in such way between different datacenters in big facilities known as hyperscale datacenters. Hence, three-tier datacenter architectures are quite problematic since they cannot scale properly to handle this so-called east-west traffic [**11**].

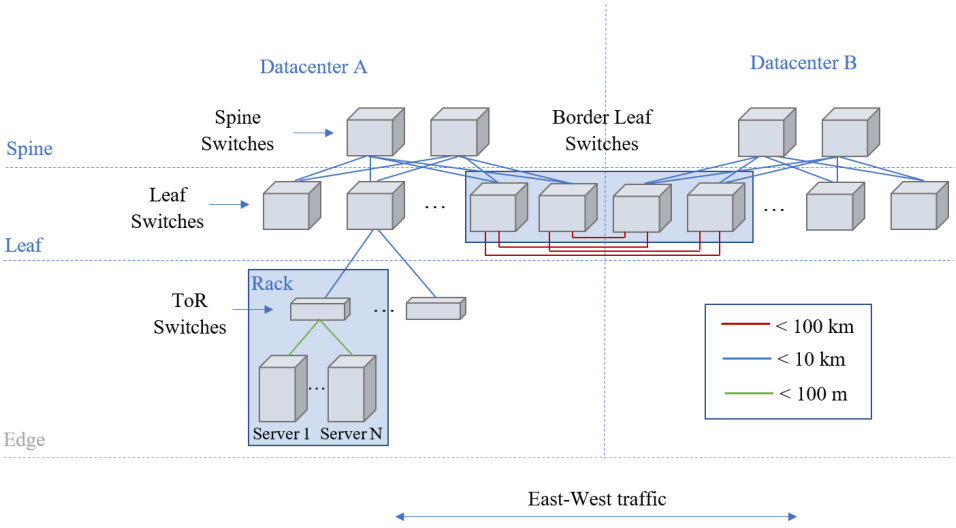


FIGURE 2.2. A two-tier data center architecture, suitable for handling north-south and east-west traffic. Red lines represent connections with less than 100 km, blues lines represent connections with less than 10 km, green lines represent connections with less than 100 m. Based on [**8, 11, 22**].

To manage east-west traffic more efficiently, hyperscale datacenters have switched to a two-tier architecture [**11**], shown in Fig.2.2. ToR switches are now connected directly to leaf switches. These leaf switches are connected to every spine switch inside a datacenter, resulting in several possible paths, that increase the redundancy inside the datacenter [**11**]. In this architecture, east-west traffic is routed to a spine switch before traveling back down to the most suitable leaf switch. Moreover, traffic between different datacenters is handled

by border leaf switches. In order to provide a higher scalability, more leaf switches or spine switches may be added [**11**].

2.1.2. Datacenter connections

Datacenter connections are categorized as short-haul links and can be labed as intra-datacenter and inter-datacenter links by having a range up to 10 km and 100 km, respectively [**11**]. For instance, in the two-tier architecture represented in Fig. 2.2, connections between border leaf switches from different datacenters are considered inter-datacenter links, while all connections within a datacenter are considered intra-datacenter links. Both, above 100 m, are typically supported by optical fiber technology and less propagation impairments in comparison with long haul links are of concern, since polarization mode dispersion (PMD) and nonlinearities are generally low in such short propagation distances [**11**].

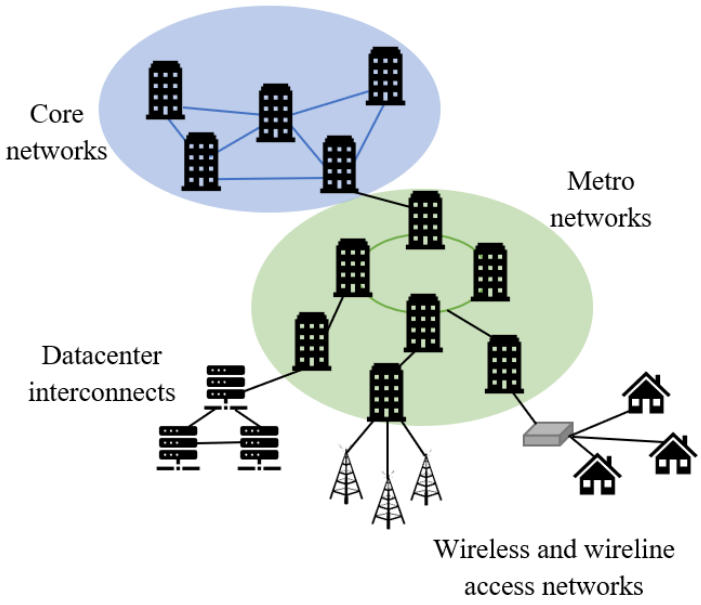


FIGURE 2.3. Schematic of an optical network. Based on [**24**].

Optical networks are essentially divided into core, metro and access networks [**24, 25**], schematically shown in Fig.2.3. Even though coherent detection techniques enabling high spectral efficiency have already been deployed in core and metro haul, short reach optical access networks and the datacenters connections considered in this work, still rely on IM-DD in order to meet the simplicity, low cost and power consumption design requirements.

In Fig. 2.4, an intra-datacenter WDM link is schematically presented. In general, this type of link operates near the 1310 nm wavelength (second window) in order to avoid signal distortion arising from chromatic dispersion (CD) [**11**]. For such short propagation

distances, optical amplification is not required. The performance of intra-datacenter links is usually measured by calculating the receiver sensitivity, which is the received power required to achieve a target BER, determined by a forward error correction (FEC) threshold [**11**].

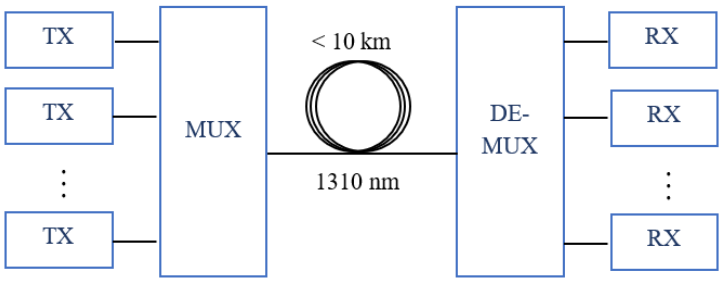


FIGURE 2.4. Schematic of an intra-datacenter link, in which TX stands for transmitter, MUX for multiplexer, DE-MUX for demultiplexer and RX for receiver. Based on [**11**].

Inter-datacenter links, schematically shown in Fig. 2.5, can be up to ten times longer than intra-datacenter links. Such distances require optical amplification which is implemented with erbium-doped fiber amplifiers (EDFAs) that operate near 1550 nm (third window) [**11**]. In this window and for the required data rates, CD is a major impairment, that must be compensated by CD compensation (CDC) modules [**11**]. The performance of inter-datacenter links is typically measured in terms of the optical signal-to-noise ratio (OSNR) required to obtain a target BER specified again by a FEC threshold [**11**].

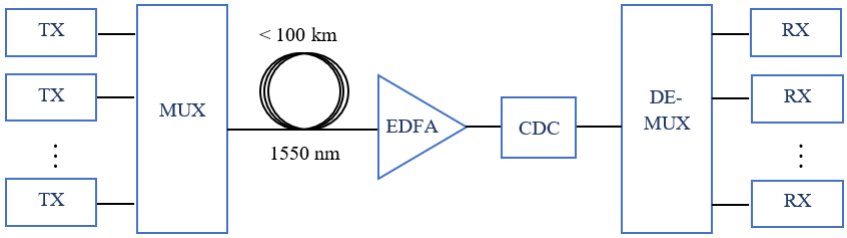


FIGURE 2.5. Schematic of an inter-datacenter link, in which EDFA stands for erbium-doped fiber amplifier and CDC for chromatic dispersion compensation module. Based on [**11**].

Regarding signal transmission, datacenter connections have been mainly using IM signals such as OOK and, most recently, PAM4 have also been considered [**11**]. The latter has been proposed to replace OOK for short reach optical communication using IM-DD and has already been adopted by the IEEE 802.3bs task force to enable 50 and 100 Gbit/s per wavelength, since it minimizes the power consumption and spectral efficiency issues of OOK transmission [**11, 26**]. Current approaches to deal with capacity scalability

in datacenter links rely on increasing the number of WDM channels and increasing the baud rate, which demands higher bandwidth components and adopt a higher number of symbols per bit [**26**]. Scaling the symbol rate is limited by the achievable bandwidth of electrical and optical components [**27**]. Increasing the number of WDM channels has the downside of requiring a higher number of optical and electrical components, as much of the number of new WDM channels, and may severely increase the link cost. Encoding more bits in one symbol allows scaling the capacity without requiring higher bandwidths components, but has the disadvantages of requiring higher OSNRs (7 dB are lost when passing from an OOK signal to a PAM4 format) and enhances the susceptability to intersymbol interference (ISI) [**26, 27**].

2.2. Multicore Fiber

Datacenter links predominantly rely on conventional SC-SMF technology. SC-SMF based transmission systems experiments have been able to reach capacity values up to 100 Tb/s mainly by using coherent detection in long-haul connections [**1, 2, 14**]. Even so, due to the ever-growing predicted traffic rates, these capacity values are expected to no longer fulfil the future demands, which raises the challenge of redesigning optical fibers and associated technology. By the information theory developed by Shannon, the additive white Gaussian noise (AWGN) waveform channel allows to predict the capacity that can be offered by a link [**28**]. Calculations based on this type of channel lead to the conclusion that multiplexing factors, such as increasing the number spatial paths, are promising solutions for capacity scaling demands [**29**]. Therefore, space division multiplexing (SDM) on optical fibers has been proposed and actively researched since 2008, being an object of deep study as a solution to overcome the upcoming capacity critical point of conventional SC-SMF [**8**]. The SDM concept consists of multiple light paths used to transmit independent channels. From the fiber point of view, there are three main approaches to introduce multiple spatial paths [**14, 30**]. The first approach is the intuitive and is the most commonly used and consists of multiple parallel independent SC-SMFs, known as fiber bundles. However, in order to achieve high capacities, a large number of SM-SCFs is required leading to space issues inside datacenters facilities. The second approach is to introduce several different modes in a fiber core, such as in few-mode fiber (FMF) or multimode fiber (MMF), where each mode is assigned to a different channel [**30**]. This approach requires multiple input-multiple output (MIMO) digital signal processing (DSP) to deal with the mode coupling and is not suitable for datacenter link distances, since

it does not meet the simplicity design requirements [**30**]. Finally, the third approach is using MCFs, which are fibers that incorporate multiple cores in the same fiber cladding [**14**]. MCFs, from the total SCC perspective, are considered promising technologies to overcome the capacity limits of SC-SMF transmissions and as a SDM solution for maximizing the capacity and spatial efficiency of datacenter connections, as already claimed in recent works [**4, 5, 6**]. Key enabling technologies for datacenter connections with MCFs, such as SDM MUX/DEMUX modules, SDM switches and SDM transceivers have been assessed in [**6**].

2.2.1. ICXT in multicore fibers

One of the main issues to be dealt in MCF signal transmission is the ICXT [**14**]. This interference in MCFs occurs from coupling between cores and is particularly relevant with the transmission of signals with the same wavelength in neighboring cores [**4**]. The ICXT has been experimentally shown to have a stochastic time evolution, which can result in high levels of ICXT in short or long time periods [**7, 16**]. This leads to poor signal transmission quality and even system outages.

The effect of ICXT also depends of the type of MCFs, which can be categorized as weakly-coupled or strongly-coupled (SC-MCF) in terms of signal coupling between cores [**14**]. In WC-MCFs, the core-to-core distance, i.e, core pitch, is set to guarantee a low level of interference between neighbouring cores for a ICXT level lower than -30 dB. Hence, the core pitch must be higher than 30 µm to ensure a coupling coefficient lower than 0.01 m^{-1} [**14**]. In strongly-coupled MCFs, in order to increase the core density and fiber capacity, the core pitch is decreased resulting in a high level of interference between neighbouring cores and a higher ICXT [**14**]. In this case, the core pitch is lower than 30 µm, which leads to a coupling coefficient higher than 0.1 m^{-1} [**14**]. Transmission in SC-MCF can be theoretically considered similar to the transmission in MMFs, and, as so, it requires MIMO-DSP to separate the signals in the different cores and reduce the ICXT at the receiver [**14**]. Due to the complex receiver setup, SC-MCFs are not considered suitable datacenter links transmission due to the lower cost requirement. Therefore, WC-MCFs have been considered a promising technology to improve the transmission capacity per fiber in datacenter connections [**31**].

MCFs can also be categorized as homogeneous or heterogeneous [**14, 31**]. In homogeneous MCFs, all cores are made of the same material, which leads to equal core

propagation constants [**14**]. In heterogeneous MCFs, adjacent cores have a different geometry and effective refractive index, leading to different propagation constants between neighbouring cores [**14**]. The latter can be a good strategy to mitigate ICXT, however, it also requires DSP to compensate the delays between the signals transmitted in the fiber cores and recover properly the signal received at the end of each core [**14**]. Thus, heterogeneous MCFs links are more complex to implement in datacenter connections.

The ICXT stochastic behaviour is problematic for datacenters connections, since they are expected to ensure high efficiency and reliability on how they manage significant amounts of traffic. For decreasing the ICXT in homogeneous WC-MCFs, the coupling coefficient between cores must be reduced [**14**]. Trench-assisted and hole-assisted MCFs have been proposed for that matter [**14**]. Another approach to ease ICXT in MCFs is to employ propagation-direction interleaving (PDI) techniques resorting to resource allocation algorithms (RSA) where adjacent cores are assigned to opposite transmission directions [**4, 6, 14**]. Low crosstalk and high core count MCFs can be designed using trench-assisted MCFs with PDI [**14**]. The core pitch and layout, the refractive index, operating wavelength, as well as the bending and twist of the fiber (typical of short links) should be considered when evaluating techniques for decreasing the ICXT [**14**].

2.2.2. PAM4 signal transmission in datacenter links supported by MCFs

PAM4 has been proposed to replace OOK for short reach optical communication using IM-DD, since it minimizes the power consumption and enhances spectral efficiency compared to OOK transmissions [**11, 26**]. In addition, PAM4 signals transmission over MCFs has also been considered and studied for scaling the capacity of short haul links and, particularly, of datacenter connections [**4, 5, 6**].

Several works have experimentally demonstrated PAM4 signals transmission supported by MCFs and have been reported in [**5**]. The standard 125 µm cladding diameter MCF has been considered a preferred choice for applications in short-hauls to avoid bending ICXT-related impairments [**5**]. It has been also stated that a four or eight core-count are a more realistic and deployable MCF-based solution in datacenters [**5**]. Fiber cable density in datacenters can be improved by implementing 400 Gb/s PAM4 transmission based on 4-core SM-MCF using 100 Gb/s per wavelength transceiver technology, which was demonstrated over 2 km [**5**]. In [**32**], a PAM4 transmission with a directly modulated 1.5µm single-mode vertical cavity surface emitting laser (VCSEL) supported by a 7-core MCF was achieved. PAM4 signals up to 70 Gbaud were generated with a VCSEL

in optical back-to-back (B2B) and 50 Gbaud PAM4 signals were successfully transmitted in each core through 1 km without dispersion compensation and 10 km with dispersion compensation [**32**]. In [**33**], a 80 Gbaud/wavelength/core PAM4 signal transmission is demonstrated using an integrated externally modulated laser, which enabled 7×149 Gbit/s transmission speed over 1 km supported by SM-MCF [**33**].

2.3. Machine learning

Machine Learning is a branch of Artificial Intelligence and it is, essentially, a modeling technique that finds a model, by itself, from a given training data [**17, 34**]. These models are later applied to the actual field data and are generally assigned to detection, classification, recognition and decision-making purposes. ML algorithms can be divided in three main categories, depending on the training method, which are supervised learning (SL), unsupervised learning (UL) and reinforcement learning (RL) [**17**]. In optical fiber communications, the first two are the most addressed techniques according to [**17, 18**]. SL consists of assigning to each training dataset an input and the output which the model is supposed to produce for that particular input [**34**]. After several trainings, the ML algorithm is finally able to obtain a model that predicts the output with sufficient accuracy from input datasets different from the training dataset. The two most common applications of supervised learning are classification and regression [**17, 34**]. Discrete variables as outputs are associated with classification problems, which focus on determining the classes to which the input data belongs [**17, 34**]. Continuous variables are related to regression problems, which do not infer classes but, instead, estimate output values [**17, 34**]. In contrast, the training dataset in UL contains only inputs without outputs. Therefore, the models are generated by examining common patterns in the input data. This type of learning is most often associated with clustering techniques, which are useful for grouping similar data.

One of the most well known ML algorithms is the Artificial Neural Network (ANN), which can behave as an universal approximation function mediator between the input data and output data after the training phase [**35, 36**]. The simplest architecture of an ANN is composed of an input layer, an hidden layer and an output layer. These layers contain nodes, or also called neurons, modeled as an activation function, with weights and biases, which are continuously updated during through a back propagation training process, until the output layer can produce the desired output [**36**].

With the growth of data and computation capacity, ML has expanded into deep learning (DL) and ANNs developed into deep neural networks (DNN) composed by multiple hidden layers and a high number of nodes [**37**]. This allowed to address a wider variety of problems and data types resorting to new algorithms with specific structures such as convolutional neural networks (CNNs), which allow to process images, by adding convolution, activation and pooling layers for feature extraction [**37, 38**].

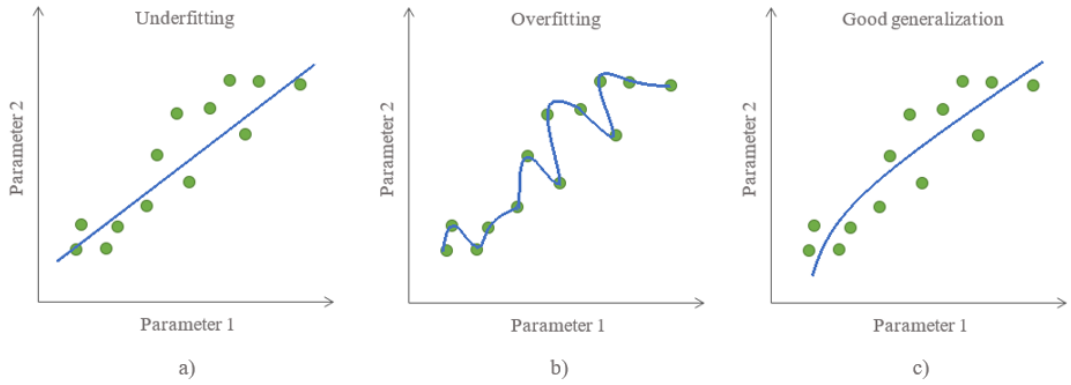


FIGURE 2.6. Schematic of underfitting, overfitting and good generalization for an uni-dimensional regression problem. Based on [**17**].

Generalization is an important aspect of a consistent model [**34**]. Typically, overfitting and underfitting phenomenons occur when a model is unable to generalize. Fig.2.6 a), shows an example of underfitting, in which relevant features of the data are not captured due to a model unable to yield new data [**17, 34**]. In overfitting, schematically represented in Fig.2.6 b), the model yields precisely the training data, which results in a low generalization and inaccurate predictions for new input data [**17, 34**]. Therefore, it is important to choose unbiased data, in order to decrease the disparity between the training data and the data that will be later applied to the obtained model. However, one of the main challenges DL still faces is the shortage of available data and unbalanced data within the training data [**39**]. To address this issue, some techniques such as data augmentation and domain randomization have been studied [**39, 40**]. The basic principle of data augmentation is obtaining new data by transforming the already existing data [**39**]. Domain randomization generates synthetic data similar enough to emulate the data under study [**40**]. Training with synthetic data generated through simulation can be cheaper, diverse, which prevents from training with unbiased data, and less time consuming compared to collecting real data [**40, 41, 42**].

As an addition to collecting unbiased and large amounts of data, validation is also considered as a method to confront generalization. This method consists of splitting the available data into a training set and a validation set. Then, during training, the model performance is evaluated using the validation set [**17**]. If the model presents an acceptable performance in terms of generalization or a stabilized error, the training can be finished. If not, the model must be modified and a new training and validation process must be started. A slight variation of validation is cross-validation that splits training data into groups for the training and validation, but systematically reorganises the sets [**17**].

2.4. ML in optical performance monitoring

The application of ML techniques to the physical layer of optical communication networks has already been addressed in several works and has gained a lot of interest by the research community in the last few years [**17, 18, 43**]. Hence, typical areas of application of ML in the physical layer of optical communications networks are fiber nonlinearity mitigation, modulation format (MF) recognition and optical performance monitoring (OPM). Even so, ML approaches to deal with the ICXT in WC-MCFs have not been yet addressed in the literature to the best of the authors' knowledge, which we aim to address in this work.

Regarding OPM, the use of CNNs have been recently addressed . In [**19**], a CNN-based technique is assessed to perform OSNR estimation and MF recognition, from eye-pattern images of several modulation formats, PAM4 being one of them. In [**20**] an eye-pattern analysis scheme based on a CNN for IM-DD transmissions is also presented. Different eye-patterns of OOK and PAM4 signals from B2B and up to 80 km link transmissions are sent to a CNN-based model that outputs eye diagram characteristics, fiber link length, Q-factor and impairments recognition [**20**].

As the previous studies using CNNs presented good results, in this work, we focus on contributing with the study of the performance of a CNN in predicting BERs through eye-patterns from PAM4 datacenter optical connections impaired by ICXT and suported by WC-MCFs, which to the best of our knowledge, has not been yet addressed in the literature.

CHAPTER 3

System model description

In this chapter, the CNN for eye-pattern analysis and BER prediction of PAM4 inter-datacenter optical connections is presented. The optical telecommunication system equivalent model is described in section 3.1, where the DP-DCM developed to characterize the ICXT in MCFs [**21**] is described in detail. The study of the BER distribution and respective eye-patterns in an optically amplified PAM4 link impaired by ICXT is performed in section 3.2. The main tasks of the CNN-based eye-pattern analysis and BER prediction are presented in section 3.3.

3.1. Optical communication system model

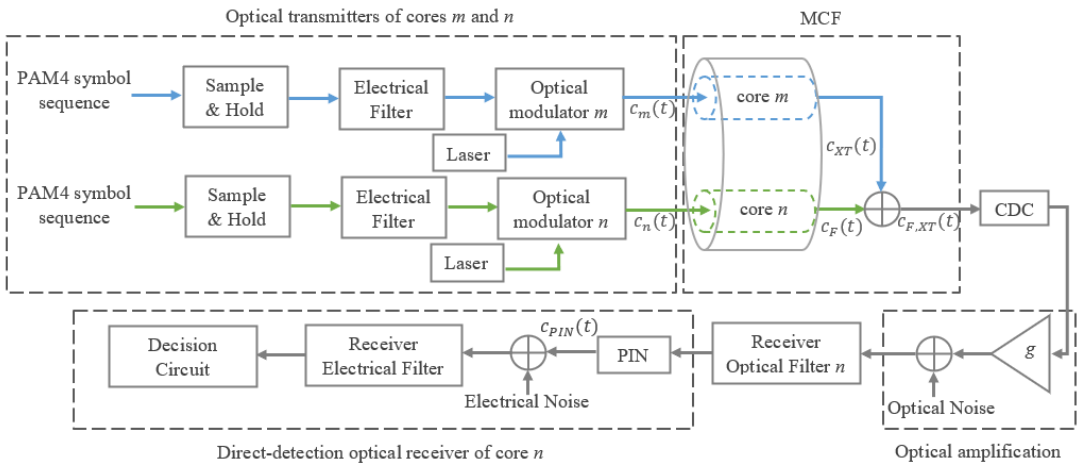


FIGURE 3.1. System equivalent model to assess the impact of ICXT on a PAM4 optical communication link with DD and MCFs. Based on [**2, 3**]

The optical telecommunication system equivalent model developed in [**2, 3**] for a PAM4 inter-datacenter optical link supported by MCF is shown in Fig. 3.1. In this work, by analysis of the results provided by this model, namely the BER and received eye-patterns, a CNN will be tested and trained to predict the BER.

The impact of ICXT on the performance of optically amplified PAM4 links for inter-datacenter connections has been analyzed in [**2, 3**] by resorting to the DP-DCM that describes accurately the ICXT effect in homogeneous WC-MCFs [**21**]. The DP-DCM of the MCF used in [**2, 3**], considers firstly only two cores, the interfering core m and the

interfered core n as shown in Fig. 3.1. Hence, two transmitters, one for each core, generate different signals where the signal transmitted along core m, $c_m(t)$, is the interfering PAM4 signal and the signal transmitted along core n, $c_n(t)$, is the interfered PAM4 signal. After symbols generation, the PAM4 symbols with non-ideal extinction ratio are sampled and passed through an electrical filter that models the frequency limitations of the electrical part of the transmitter. After electrical filtering, the PAM4 signal is converted to the optical domain by an optical modulator with a finite extinction ratio and without chirp. The extinction ratio is given by the inverse of the ITU-T definition as $r = \frac{P_0}{P_3}$, where P_0 and P_3 are the powers of the PAM4 symbols '0' and '3', respectively.

The optical signals $c_m(t)$ and $c_n(t)$ are transmitted in two perpendicular polarization directions, x and y, which are represented by the power splitting of the transmitted PAM4 signal by both polarization directions at the input of the MCF. Hence, in the interfering core m, the power splitting in the PAM4 transmitted signal in polarization x, $c_{m,x}(t)$, and in the PAM4 transmitted signal in polarization y, $c_{m,y}(t)$, is given by

$$c_{m,x}(t) = c_m(t) \times \sqrt{\xi_m} \tag{3.1}$$

$$c_{m,y}(t) = c_m(t) \times \sqrt{1 - \xi_m} \tag{3.2}$$

where ξ_m determines the power distribution between the two polarization directions in core m and can vary between 0 and 1.

In the interfered core n, the PAM4 transmitted signal in polarization x, $c_{n,x}(t)$, and in polarization y, $c_{n,y}(t)$ is given by

$$c_{n,x}(t) = c_n(t) \times \sqrt{\xi_n} \tag{3.3}$$

$$c_{n,y}(t) = c_n(t) \times \sqrt{1 - \xi_n} \tag{3.4}$$

The results in [**21**] show that the variance of the ICXT field amplitude is similar for all polarization directions and is independent of ξ_m and ξ_n.

After splitting, the PAM4 signal travels along core n of the MCF, which is modelled by the linear propagation transfer function $H_F(\omega)$. Linear propagation through the MCF is assumed, since non-linear effects are usually insignificant in inter-datacenters distances

[**11, 44**]. The signal at the output of core n without ICXT is represented as

$$c_F(t) = c_n(t) * \mathcal{F}^{-1}[H_F(\omega)] \tag{3.5}$$

where $*$ stands for convolution, $\mathcal{F}^{-1}$ stands for the inverse Fourier Transform and $H_F(\omega)$ is given by

$$H_F(\omega) = \exp(-j\bar{\beta}_n(\omega)L).\exp(-\frac{\alpha_n}{2}L) \tag{3.6}$$

where ω is the angular frequency, $\bar{\beta}_n(\omega)$ is the average of the propagation constants in core n, which is given by the sum of the contributions in both polarization directions in n as in $\bar{\beta}_n(\omega) = (\beta_n^x(\omega) + \beta_n^y(\omega))/2$; where α_n is the attenuation coefficient of the core n and L is the SM-MCF length. In the absence of the ICXT effect, linear propagation is also considered in core m with the average of the propagation constants in core m being given by $\bar{\beta_m}(\omega) = (\beta_m^x(\omega) + \beta_m^y(\omega))/2$. The $\bar{\beta}_l(\omega)$ is represented as [**45**]

$$\bar{\beta}_l(\omega) = \beta_{0,l} + \beta_{1,l}\omega + \frac{\beta_{2,l}}{2}\omega^2 + \frac{\beta_{3,l}}{6}\omega^3 \tag{3.7}$$

where l can be m or n, when referring to core m or m, respectively. In (3.7), $\beta_{0,l}$ is the propagation constant at the carrier wavelength, $\beta_{1,l}$ is the inverse of the group velocity, $\beta_{2,l}$ is the group velocity dispersion and $\beta_{3,l}$ is the higher-order dispersion [**45**].

The effect of ICXT on the interfered cores is modelled by the DP-DCM simply by using transfer functions that change randomly along time, being this dependence introduced by applying random phase shifts (RPSs) along the longitudinal direction of the fiber. The transfer functions $F_{a,b}(\omega)$ model the frequency response of the ICXT from the polarization a, with $a = x$ or y, at the input of core m to the polarization b, with $b = x$ or y, at the output of core n and are represented as [**21**]

$$F_{a,b}(\omega) = -\frac{j}{\sqrt{2}}\bar{K}_{nm}\exp(-j\bar{\beta}_n(\omega)L)\cdot\exp(-\frac{\alpha_m}{2}L)\sum_{k=1}^{N_p}\exp[-j\Delta\beta_{mn}z_k]\exp[-j\phi_{nm,k}^{(a,b)}] \tag{3.8}$$

where α_m is the attenuation coefficient of the core m. In the model proposed by [**2, 3**], $\bar{K}_{nm}$ is the average inter-core coupling coefficient [**21**], N_p is the number of phase-matching points (PMPs), $\Delta\beta_{mn}$ is given by $\bar{\beta_m}(\omega) - \bar{\beta}_n(\omega)$ and $\phi_{nm,k}^{(a,b)}$ represent the RPSs associated with the k-th PMP, which are modelled using an uniform distribution between $[0, 2\pi[$ and z_k is the longitudinal coordinate of the k-th PMP randomly distributed between two

consecutive PMPs and is given by

$$z_k = \frac{L}{N_p}(r_k + k - 1) \tag{3.9}$$

where $r_k (1 \leq k \leq N_p)$ are independent random variables generated with an uniform distribution between [0, 1[. In this work, we also assume $\alpha_m \approx \alpha_n$.

The DP-DCM models the evolution of the ICXT impact on the system performance in different MCF realizations by generating randomly different N_p sets of RPSs in each MCF realization. Therefore, in each iteration of the Monte Carlo (MC) simulator, a new PAM4 signal with symbols randomly generated is transmitted in core m and one MCF realization corresponding to the transfer functions given by (3.8) is generated. The transfer functions $F_{x,x}(\omega)$ and $F_{y,x}(\omega)$ model the ICXT generated from polarization x and y of the core m that is going to interfere with the electrical field in the polarization x of the core n. The transfer functions $F_{x,y}(\omega)$ and $F_{y,y}(\omega)$ model the ICXT generated from both polarizations of core m that interfere with the field of core n in the polarization y. The ICXT level quantifies the ammount of ICXT power that will affect the interfered core and is defined by the ratio between the mean ICXT power and the mean power of the signal both at the output of the interfered core n and is calculated as in $X_c = N_p|K_{nm}|^2$ [**21**].

The temporal dependence of the ICXT is induced by effects such as the walk-off due to different group velocities between cores [**44**]. In this model, the skew between the interfering core m and the interfered core n is given by $S_{mn} = d_{mn}L$, where d_{mn} is the walkoff between cores m and n defined by $d_{mn} = \beta_{1,m} - \beta_{1,n}$.

In the DP-DCM, the effect of ICXT in the PAM4 signal after propagation in core n, is obtained by $c_{F,XT}(t) = c_F(t) + c_{XT}(t)$, where the interfering signals from core m, $c_{XT,x}(t)$ (with a=x and b=y) and $c_{XT,y}(t)$ (with a=y and b=x), are represented as

$$c_{XT,a}(t) = c_{m,a}(t) * \mathcal{F}^{-1}[F_{a,a}(\omega)] + c_{m,b}(t) * \mathcal{F}^{-1}[F_{a,b}(\omega)] \tag{3.10}$$

$$c_{XT}(t) = c_{XT,a}(t)\hat{a} + c_{XT,b}(t)\hat{b} \tag{3.11}$$

In Fig. 3.1, at the output of the MCF, a CDC module compensates the CD arising from the transmission through core n, which can be a major source of performance degradation due to the required data rates. The CDC is modelled considering a DCF in linear

propagation with a length that fully compensates the distortion due to CD on the signal $c_{F,XT}(t)$. The ideal DCF length is given by

$$L_{DCF} = \frac{-D_{\lambda,n}L}{D_{\lambda,DCF}} \tag{3.12}$$

where $D_{\lambda,n}$ is the core n dispersion parameter and $D_{\lambda,DCF}$ is the DCF dispersion parameter.

After the CDC module, there is an EDFA to compensate the inter-datacenter link losses and an optical filter to reduce the amplified spontaneous emission (ASE) noise power generated by the EDFA.

The amplifier gain is set to compensate all the losses from the SM-MCF and DCF and the ASE noise is modelled as additive white Gaussian noise with power spectral density, per polarization mode, given by [**45**]

$$S_{ASE} = \frac{F_n}{2}(g_{EDFA} - 1)hv_0 \tag{3.13}$$

where g_{EDFA} is the EDFA gain in linear units, hv_0 is the photon energy and F_n is the EDFA noise figure.

The optical filter, modelled by a 4th order super Gaussian filter, reduces the power of the optical noise and introduces negligible inter-symbol interference (ISI) [**45**]. The transfer function of the i-th order super Gaussian filter is given by [**46**]

$$H_0(f) = \frac{1}{\sqrt{i_L}} \exp\left[-\left(\frac{2|f - f_0|}{B_0}\right)^{2i} \ln(\sqrt{2})\right] \tag{3.14}$$

where f_0 is the optical filter lowpass equivalent center frequency, i_L is the insertion loss in linear units, and B_0 is the optical filter bandwidth at -3 dB.

After CDC and optical amplification, the PAM4 signal degraded by ICXT and ASE noise passes the DD receiver dedicated to core n, where is converted to an electrical signal, $c_{PIN}(t)$, by the PIN photo-detector with a responsivity of $R_\lambda = \frac{\eta q}{h\nu_0}$. Electrical noise is added after photodetection and an electrical filter, modelled as a 3^{rd} order Bessel filter, is used to reduce the noise power. In the decision circuit, the BER of each MCF realization, the average BER and the OP are assessed. The BER of each MCF realization is calculated by the semi-analytical method known as the exhaustive Gaussian approach. For a PAM4 signal, the BER is given by [**47**]

$$BER = \frac{1}{2 \cdot 4^{N_{reg}}} \left\{ \sum_{\substack{k=1 \\ a_k=0}}^{4^{N_{reg}}} Q\left(\frac{F_1 - i_{0,k}}{\sigma_{0,k}}\right) + \sum_{\substack{k=1 \\ a_k=1}}^{4^{N_{reg}}} \left[Q\left(\frac{i_{1,k} - F_1}{\sigma_{1,k}}\right) + Q\left(\frac{F_2 - i_{1,k}}{\sigma_{1,k}}\right) \right] + \right.$$
$$\left. \sum_{\substack{k=1 \\ a_k=2}}^{4^{N_{reg}}} \left[Q\left(\frac{i_{2,k} - F_2}{\sigma_{2,k}}\right) + Q\left(\frac{F_3 - i_{2,k}}{\sigma_{2,k}}\right) \right] + \sum_{\substack{k=1 \\ a_k=3}}^{4^{N_{reg}}} Q\left(\frac{i_{3,k} - F_3}{\sigma_{3,k}}\right) \right\} \tag{3.15}$$

where $4^{N_{reg}}$ is the maximum length of PAM4 symbols optical sequence generated using deBruijn sequences, with N_{reg} as the length of the offset register used to generate the sequence; $i_{0,k}$, $i_{1,k}$, $i_{2,k}$ and $i_{3,k}$ correspond to the means of the currents at the input of the decision circuit for the symbols a_k at the time sampling instants $t_k = t_0 + T_s(k-1)$, where t_0 is extracted from the received eye-pattern at the decision circuit input, $k \in 1, ..., 4^{N_{reg}}$, $\sigma_{0,k}$, $\sigma_{1,k}$, $\sigma_{2,k}$ and $\sigma_{3,k}$ are the noise standard deviations for the different time sampling instants [**47**] and the function $Q(x)$ is given by [**3**]

$$Q(x) = \int_x^{\infty} \frac{1}{\sqrt{2}} e^{-\frac{\xi^2}{2}} d\xi \tag{3.16}$$

In the simulation, the decision thresholds F_1, F_2 and F_3 are optimized in each time-fraction by applying the bisection method to minimize the BER [**2, 3**]. Effects such as ICXT, ISI, electrical noise, signal-ASE, and ASE-ASE beat noises are taken into account using this semi-analytical method.

After several MCF realizations, the average BER is obtained by averaging the BERs obtained in each MCF realization. The OP has been appointed as a key system performance indicator on weakly-coupled DD MCF-based systems and evaluates the probability of a system outage [**16**]. To assess the performance of optical links impaired by ICXT, the OP is defined as the probability of the BER in the presence of ICXT, to exceed a given BER limit [**3**]. The BER limit determines the minimum BER value in which the system becomes unavailable. In the simulation, the OP is estimated by [**2, 3**]

$$OP = \frac{N_o}{N_r} \tag{3.17}$$

where N_0 is the number of MCF realizations with BER above the BER limit and N_r is the number of simulated MCF realizations necessary to reach N_0 occurrences of BER above the BER limit.

3.2. Study of the inter-datacenter PAM4 link performance with ICXT

In this section, the impact of ICXT on the optically amplified PAM4 links performance is assessed by analyzing the received eye-patterns and the distribution of the obtained BER at the decision circuit. The system and simulation parameters used throughout this section are presented in Table 3.1.

Parameters	**Value**
Number of generated PAM4 symbols in each MCF realization	$N_{PAM4symbols} = 4^4$
Number of samples per symbol	$N_s = 32$
Symbol rate	$R_s = 56$ Gbaud
Operating wavelength	$\lambda_0 = 1550$ nm
Receiver electrical filter bandwidth	$B_{e,RX} = 0.85 \times R_s$
Receiver optical filter bandwidth	$B_0 = 1.6 \times R_s$
SM-MCF length	$L = 80$ km
Optimized transmission power for r=0 and r=0.1	$P_{TX,r=0} =$ -1.912 dBm, $P_{TX,r=0.1} =$ 0.26 dBm
Number of PMPs	$N_p = 1000$
EDFA noise figure	4.77 dB
Skew-symbol rate product	$\lvert S_{mn}R_s\rvert = 1000$, $\lvert S_{mn}R_s\rvert = 0.01$
Number of interfering cores with PAM4 signaling	$N_i = 1$
MCF chromatic dispersion parameter	$D_{\lambda,n} = 17$ ps/(nm·km)
MCF attenuation coefficient	$\alpha = 0.2$ dB/km
DCF chromatic dispersion parameter	$D_{\lambda,DCF} = 100$ ps/(nm·km)
DCF attenuation coefficient	$\alpha_{DCF} = 0.5$ dB/km
PIN responsivity	$R_\lambda = 1$ A/W
ICXT levels	$X_c =$ -20 dB, -14 dB
BER limit with ICXT	3.8×10^{-3} ($\log_{10}(BER_{limit}) \approx -2.4202$)
Target BER without ICXT	3.8×10^{-5} ($\log_{10}(BER_{limit}) \approx -4.4202$)

TABLE 3.1. System and simulation parameters

In this work, we consider that the inter-datacenter link, impaired by ICXT, is in outage when the BER is above a pre-defined limit of 3.8×10^{-3}, which is the threshold typically used for datacenters connections with forward-error correction [**3, 48, 49**]. The electrical and optical receiver filters bandwidth were optimized in B2B operation to maximize the receiver sensitivity [**2, 3**]. For the MCF length of 80 km and r=0 and r=0.1, the signal power at the transmitter output has also been optimized to achieve the BER of 3.8×10^{-5} without ICXT. The number of PMPs is set to characterize accurately the RPS mechanism [**3, 48, 49**]. Two different intercore skews with skew-symbol rate product of $|S_{mn}R_s| = 1000$ and $|S_{mn}R_s| = 0.01$ are also chosen to perform these studies. The case of $|S_{mn}R_s| = 1000$ is referred as high skew symbol rate product as $|S_{mn}R_s| >> 1$

[**3, 50, 51**]. The situation of $|S_{mn}R_s| = 0.01$ is referred as low skew-symbol rate product, since $|S_{mn}R_s| << 1$ [**3, 50, 51**].

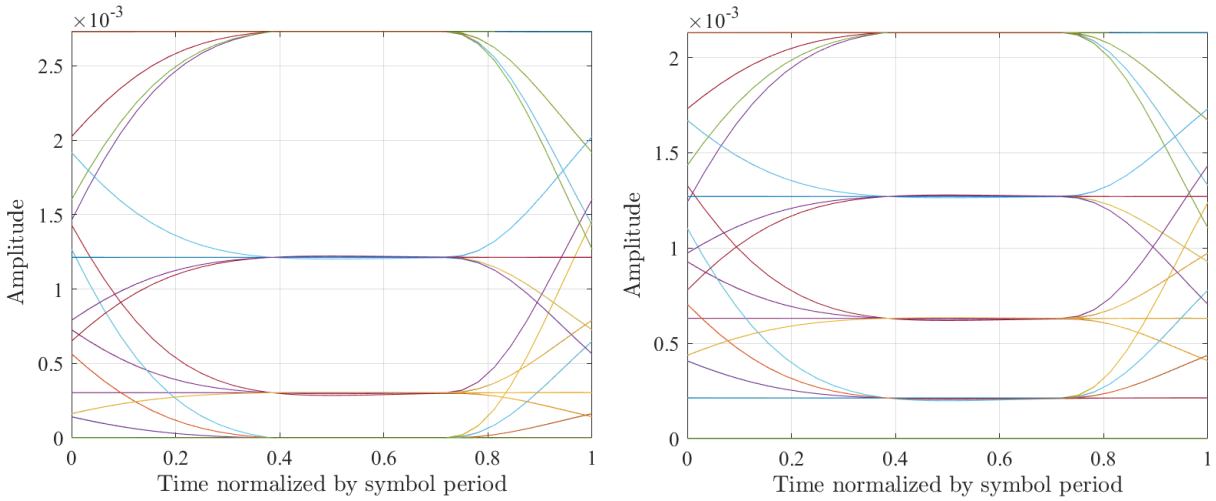


FIGURE 3.2. Eye-patterns of the PAM4 signal at the MCF input for a) r=0 and b) r=0.1.

The eye-patterns in Fig. 3.2 represent an example of the PAM4 transmitted signals $c_m(t)$ and $c_n(t)$ at the output of the optical transmitter for $r = 0$ and $r = 0.1$. In contrast, Fig. 3.3 and Fig. 3.4 show the impact of the ICXT on the eye-patterns at the decision circuit input of the optically amplified PAM4 IM-DD system with full loss and chromatic dispersion compensation. Fig. 3.3 shows the received eye-patterns for X_c = -14 dB and $|S_{mn}R_s|$ = 1000 of a) best BER and b) worst BER with r=0 and c) best BER and d) worst BER with r=0.1 after 1000 MCF realizations, which was shown to be a number high enough to obtain a stabilized average BER [**3**]. Fig. 3.4 shows the received eye-patterns for X_c = -14 dB and $|S_{mn}R_s| = 0.01$ of a) best BER and b) worst BER with r=0 and c) best BER and d) worst BER with r=0.1 after 1000 MCF realizations.

In Fig. 3.3 b), for the worst BER with r=0, the lowest eye is fully closed due to the strong ICXT. However, as seen in Fig. 3.3 d), for the worst BER with r=0.1, the ICXT degrades less the received eye-pattern (the lowest eye is not fully closed), which leads to a lower BER. The product $|S_{mn}R_s|$ has been shown to be relevant on the impact of ICXT on the performance of optical links with DD [**16, 50, 52**]. The same conclusions can be drawn through the analysis of Fig. 3.4 b) and d) where the eye-patterns are fully closed for both extinction ratios, in comparison with Fig. 3.3 b) and d), which shows that, for optical links with low $|S_{mn}R_s|$, the ICXT is more detrimental than for high $|S_{mn}R_s|$. However, for the best BERs obtained with low $|S_{mn}R_s|$, Fig. 3.4 a) c) show that the

amplitude levels are more defined than in the eye-patterns shown in Fig. 3.3 a) and c) with high $|S_{mn}R_s|$, where more symbols in the interfering core are contributing to ICXT [**3**].

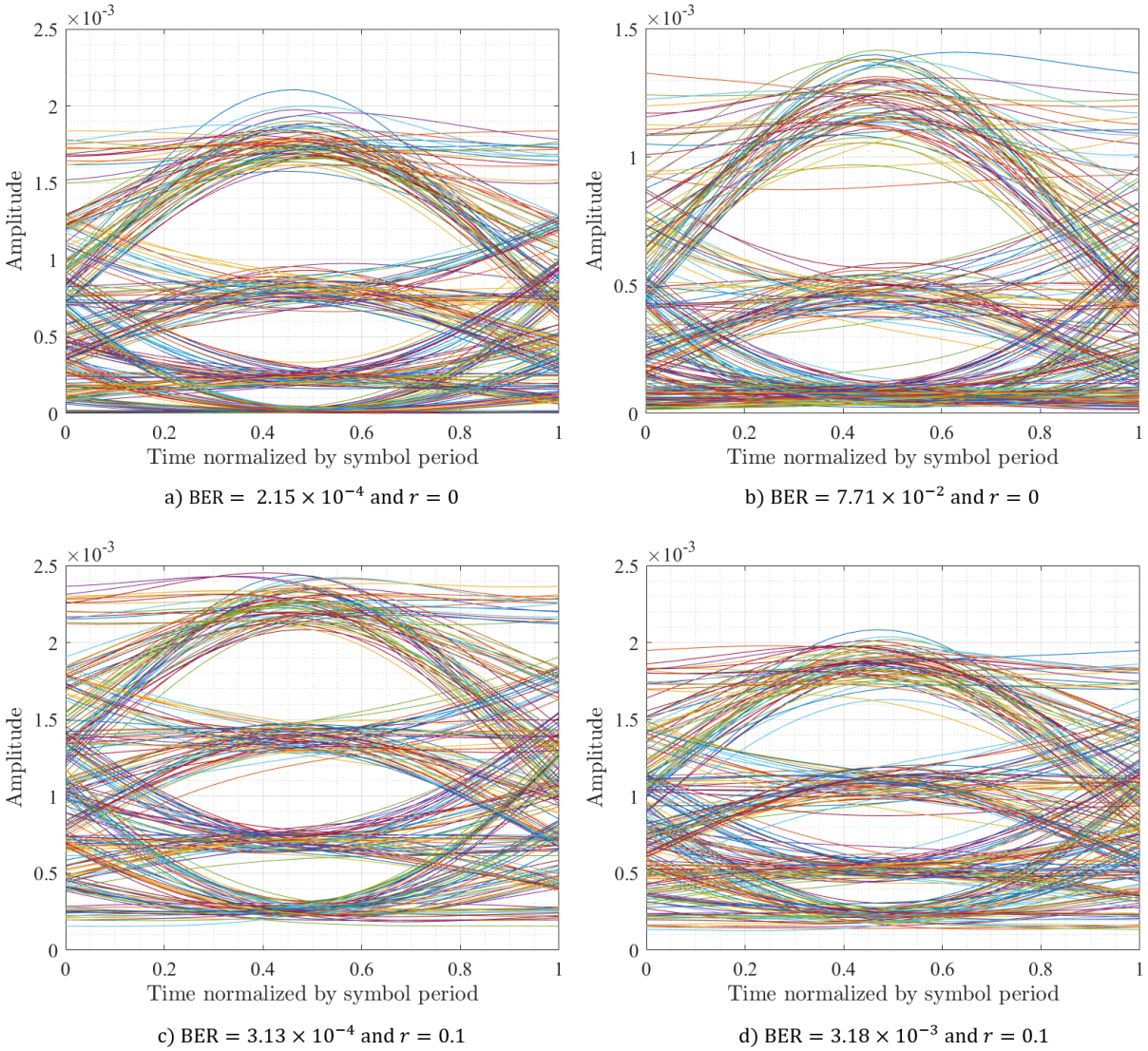


FIGURE 3.3. Eye-patterns at the decision circuit input for X_c = -14 dB, $|S_{mn}R_s| = 1000$ for a) best BER and b) worst BER with r=0 and a) best BER and b) worst BER with r=0.1.

Figs. 3.5 and 3.6 show the histograms of the BERs and the corresponding average BER obtained in 1000 MCF realizations, with, respectively X_c = -14 dB and X_c = -20 dB, r=0.1, r=0, for two different inter-core skews, a) $|S_{mn}R_s| = 1000$ and b) $|S_{mn}R_s| = 0.01$.

Fig. 3.5 shows that several MCF realizations experience system outage, especially for $|S_{mn}R_s| = 0.01$, since the corresponding BERs surpass the BER limit, as can be confirmed by the OPs presented in Table 3.2, estimated from Figs. 3.5 and 3.6, using $N_r = 1000$

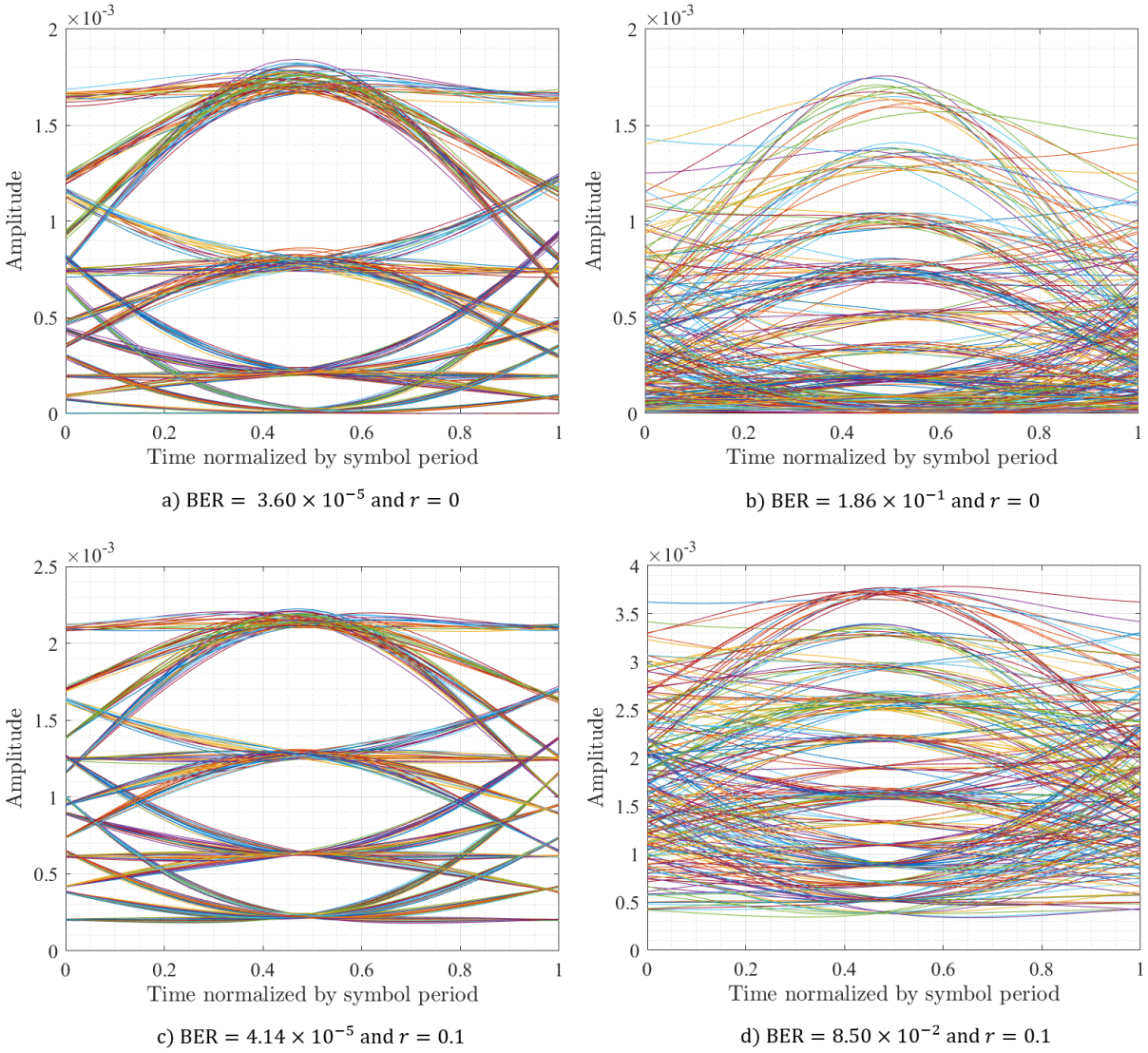


FIGURE 3.4. Eye-patterns at the decision circuit input for X_c = -14 dB, $|S_{mn}R_s| = 0.01$ for a) best BER and b) worst BER with r=0 and a) best BER and b) worst BER with r=0.1.

occurrences in Eq. (3.17). Only for $|S_{mn}R_s| = 1000$ and r=0, the BER limit is never exceeded and there is no system outage.

In Fig. 3.6, as the ICXT level is 6 dB lower, the influence of ICXT on the performance is reduced, which leads to a higher number of MCF realizations that have lower BERs and are lower than the BER limit. Therefore, the only system outages occurs for $|S_{mn}R_s| = 0.01$ and r=0 and for $|S_{mn}R_s| = 0.01$ and r=0.1, where the corresponding OPs are 0.007 and 0.001, much lower than the ones considered in Fig. 3.5.

For both crosstalk levels, the effect of ICXT on the BER distribution is less detrimental with $r = 0.1$. This influence of the extinction ratio on the ICXT impact has been already

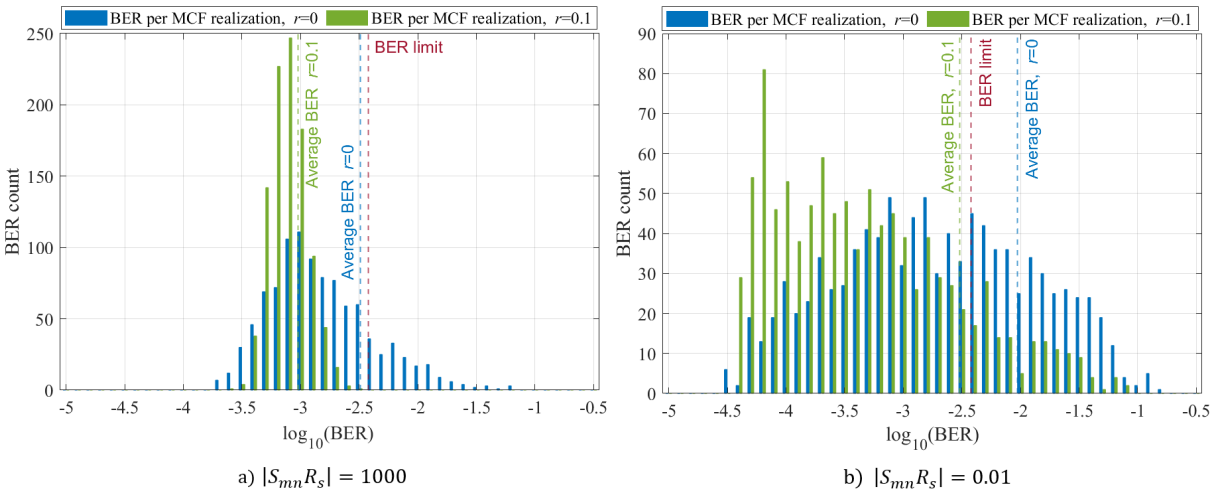


FIGURE 3.5. Histogram of the $\log_{10}(BER)$ for 1000 MCF realizations with X_c = -14 dB, r=0.1 and r=0, for a) $|S_{mn}R_s| = 1000$ and b) $|S_{mn}R_s| = 0.01$.

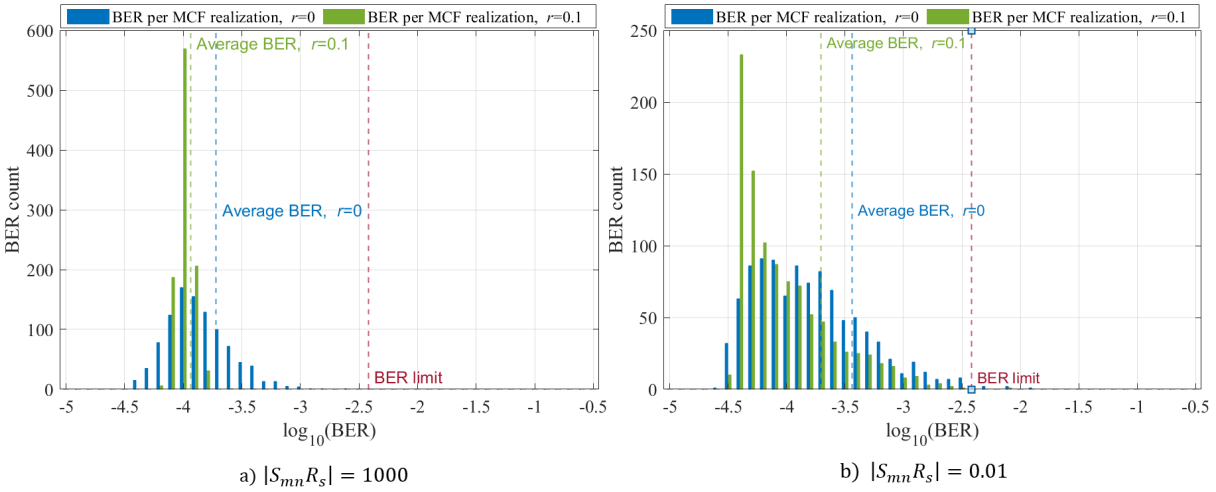


FIGURE 3.6. Histogram of the $\log_{10}(BER)$ for 1000 MCF realizations with X_c = -20 dB, r=0.1 and r=0 for a) $|S_{mn}R_s| = 1000$ and b) $|S_{mn}R_s| = 0.01$.

observed in OOK systems [**47**]. For $r = 0$, a higher spreading of the BER values is observed, and, hence, more MCF realizations have lower BERs and more lead to a BER that surpasses the BER limit. The results in both figures show also that the product $|S_{mn}R_s|$ has a significant influence on the BERs distribution, since for $|S_{mn}R_s| = 1000$, the BER range is significant lower in comparison with the BER range obtained for $|S_{mn}R_s| = 0.01$. Therefore, as seen is Table 3.2, system outage is more likely to occur for $|S_{mn}R_s| = 0.01$ and r=0 and less likely to occur for $|S_{mn}R_s| = 1000$ and $r = 0.1$.

3.3. CNN for eye-pattern analysis and BER prediction

In this work, we study the use of a MATLAB-based CNN for eye-pattern image analysis and BER prediction in PAM4 datacenter optical connections supported by homogeneous

Simulation		OP
X_c = -14 dB	$\lvert S_{mn}R_s \rvert = 1000$, $r = 0$	0.1860
	$\lvert S_{mn}R_s \rvert = 1000$, $r = 0.1$	0
	$\lvert S_{mn}R_s \rvert = 0.01$, $r = 0$	0.397
	$\lvert S_{mn}R_s \rvert = 0.01$, $r = 0.1$	0.149
X_c = -20 dB	$\lvert S_{mn}R_s \rvert = 1000$, $r = 0$	0
	$\lvert S_{mn}R_s \rvert = 1000$, $r = 0.1$	0
	$\lvert S_{mn}R_s \rvert = 0.01$, $r = 0$	0.007
	$\lvert S_{mn}R_s \rvert = 0.01$, $r = 0.1$	0.001

TABLE 3.2. OP obtained from Figs. 3.5 and 3.6, for 1000 MCF realizations, X_c = -14 dB, X_c = -20 dB, $r = 0$, $r = 0.1$, L =80 km, $|S_{mn}R_s| = 1000$ and $|S_{mn}R_s| = 0.01$.

WC-MCFs impaired by ICXT. The schematic of the main tasks of the MATLAB simulator is illustrated in Fig. 3.7.

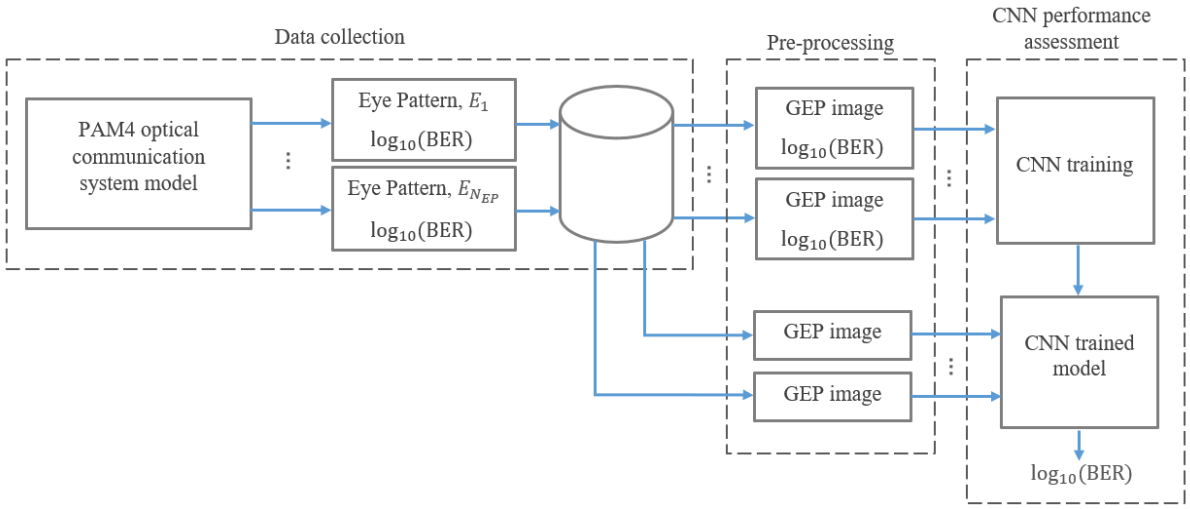


FIGURE 3.7. Main tasks of the MATLAB-based CNN for eye-pattern analysis and BER prediction model.

The first main task corresponds to data collection, where eye-patterns and the corresponding BERs are collected from the PAM4 optical communication system model described in section 3.1 for different types of optical links. The latter are obtained by varying several optical link parameters, such as crosstalk level, skew-symbol rate product and extinction ratio. Before a CNN training, the eye-patterns are pre-processed, according to the different optical links under study, to obtain what we denote as grayscale eye-patterns (GEPs) images. After that, the CNN is trained, following a SL approach, with GEPs as inputs and BERs as outputs. Then, the CNN is tested with unseen GEP images, where the prediction performance is evaluated through the root mean square error (RMSE) metric [**53**].

3.3.1. Data collection

The data used to train and test the CNN corresponds to the data collected from the optical telecommunication system equivalent model developed in [**3**] and described in section 3.1 that models with a good accuracy the ICXT effect in PAM4 inter-datacenter optical links supported by homogeneous WC-MCFs. Each pair of collected data consists of a received eye-pattern, E_i (with i=1,...,N_{EP}), where N_{EP} is the number of eye-patterns collected, such as the ones shown in Figs. 3.3 and 3.4, and the corresponding BER calculated logarithmically, i.e., $\log_{10}$(BER), at the decision circuit. Each eye-pattern is a 32×256 matrix, with N_s=32 amplitude samples per symbol and 256 is the number of generated PAM4 symbols in each MCF realization.

A good performance of the CNN is highly dependent on the training data. So, it is important to choose a high number of unbiased data to achieve a consistent model by decreasing the disparity between the training data and testing data. Therefore, a guided sampling is performed to collect a balanced number of BERs and corresponding eye-patterns. For that purpose, for the different optical links, the data generated by the MATLAB optical telecommunication system equivalent model is collected following a two-step balancing scheme, schematically shown in Fig. 3.8, within a [$\log_{10}(\text{BER})_{min}$, $\log_{10}(\text{BER})_{max}$] range, where 1000 pairs of data (eye-patterns) are saved in every 0.1 interval of the $\log_{10}$(BER). Within these N_{BER} intervals, where $N_{BER} = \frac{\log_{10}(\text{BER}_{max}) - \log_{10}(\text{BER}_{min})}{0.1}$, a second balancing step is performed with 20 sub intervals, where 50 pairs of data are saved in every 0.005 interval of the $\log_{10}$(BER). Notice that the BER range is not the same for all optical links studied, since the BER values distribution is much dependent on the link parameters as seen in section 3.2.

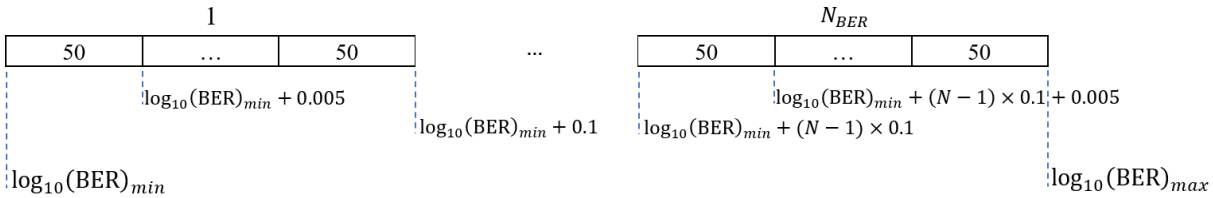


FIGURE 3.8. Two-step balancing scheme of the collected data generated by the optical telecommunication system equivalent model described in 3.1, where N_{BER} is the total number of intervals with a sub-division of 0.1 in logarithmic scale in [$\log_{10}(\text{BER})_{min}$, $\log_{10}(\text{BER})_{max}$].

As the ICXT has a stochastic behaviour with a random time evolution and frequency dependence, it leads to random fluctuations of the BER and different BER distributions, depending on the parameters of the optical link under study. Some BERs are less frequent

than others, particularly for higher BERs associated with system outages. Especially for low outage probabilities, these BERs become extremely rare to occur. Therefore, to collect a good amount of balanced data inside each logarithmic BER interval and to maintain the time of simulation at acceptable levels, particularly for BERs less likely to occur, a slight change of the DP-DCM model has been implemented following a domain randomization-based approach.

As mentioned in section 3.1, in each iteration of the MC simulator, one MCF realization corresponding to the transfer functions given by eq. (3.8) is generated with random different N_p RPSs. This randomness models the unpredictability and affects the ICXT on the received PAM4 signal and the corresponding BER. Firstly, the MC simulator starts with a random set of N_p RPSs using the uniform distribution between $[0, 2\pi[$ and is stored and used in the first MCF realization. After that, based on a single uniformly distributed random number in the interval $[0, 1]$, it is determined how the set of N_p RPSs for the next MCF realization is generated. If the random number is above or equal to 0.5, a new set of N_p RPSs is generated following the previously described procedure and, again, it is stored. If not, a new set is obtained by adding normally distributed random noise with zero mean and unitary variance to the set of N_p RPSs previously stored that is known to lead to a BER inside a BER interval. The set of RPSs generated with this random perturbation is not stored, and, therefore, it is never used again in the following MCF realizations. At the end of each MC simulator iteration, the eye-pattern and the corresponding $\log_{10}$(BER) are saved if the corresponding N_{BER} interval is not full. As the MC simulator iterations continue, following the previously described procedures, the number of N_p RPSs sets stored increases and the random perturbation approach is performed by randomly selecting one of the stored sets of N_p RPSs.

3.3.2. Eye-pattern pre-processing

Before training and testing the CNN, the eye-patterns are pre-processed to obtain what we denote as grayscale eye-patterns (GEPs). Typically, oscilloscopes sample the received signal and generate a two-dimensional database, that statistically represents the time, where the unit interval or bit period of the eye-pattern is defined by the data clock, and the amplitude of the digital signal [**54**]. As the number of samples increases, a third dimension is considered, denoted as plot density, which represents the number of pixels that are located in the same position on the oscilloscope display [**54**]. In this work, a similar approach is used to obtain GEP images.

First, for a $x \times y$ GEP image, the amplitudes of a synthetic eye-pattern obtained from MATLAB are normalized into $[1, y] \in \mathbb{N}$ amplitudes to obtain the vertical position of the GEP image pixel assigned to the corresponding amplitude. For the case of a fixed scale GEP, this normalization takes into account the maximum amplitude that occurred in all eye-patterns obtained for the corresponding type of optical link under study, i.e., the eye-patterns used during a CNN training. For the case of a dynamic scale GEP, each eye-pattern is normalized taking into account its maximum amplitude. After that, the eye-pattern to GEP conversion process continues with a $y \times x$ GEP matrix with zero elements. The plot density is modeled by incrementing one unit in all elements of the GEP matrix with the rows given by the eye-pattern normalized amplitudes and the corresponding columns given by the rows of the original eye-pattern normalized into $[1,\ x]$.

Fig. 3.9 shows the process to transform a synthetic 32×256 eye-pattern matrix obtained from MATLAB into a 32×32 fixed scale GEP image. For instance, the amplitude in the $(1, 1)^{th}$ element of the eye-pattern matrix, E, is normalized into $[1,\ 32] \in \mathbb{N}$ amplitudes, which gives 4. This means that this amplitude falls in $(4, 1)^{th}$ element of the GEP matrix, G. The amplitude is considered in the GEP matrix by incrementing one unit in that given element. If more amplitudes of the first row of the eye-pattern have the normalized amplitude 4, the $(4, 1)^{th}$ element of the GEP matrix is incremented accordingly to these number of amplitudes. The case of $E(6, 1)$ amplitude follows the same logic, where the normalized amplitude is 2 and one unit is incremented in $G(2, 6)$.

Notice that for a different GEP size, since each eye-pattern is a 32×256 matrix, an additional normalization needs to be performed during the GEP transformation process. For instance, to obtain a 64×64 GEP, the amplitudes of the eye-pattern are normalized into $[1,\ 64] \in \mathbb{N}$ amplitudes to obtain the vertical position of the GEP image pixel assigned to the corresponding amplitude. Then, the transformation process follows the previously described procedure, however, the plot density is modeled by incrementing one unit in all elements with the rows given by the eye-pattern normalized amplitudes and the corresponding columns given by the rows of the original eye-pattern normalized into $[1,\ 64]$.

After this transformation process, the GEP images, obtained from the GEP matrices, are used as input data in the CNN described in the next subsection.

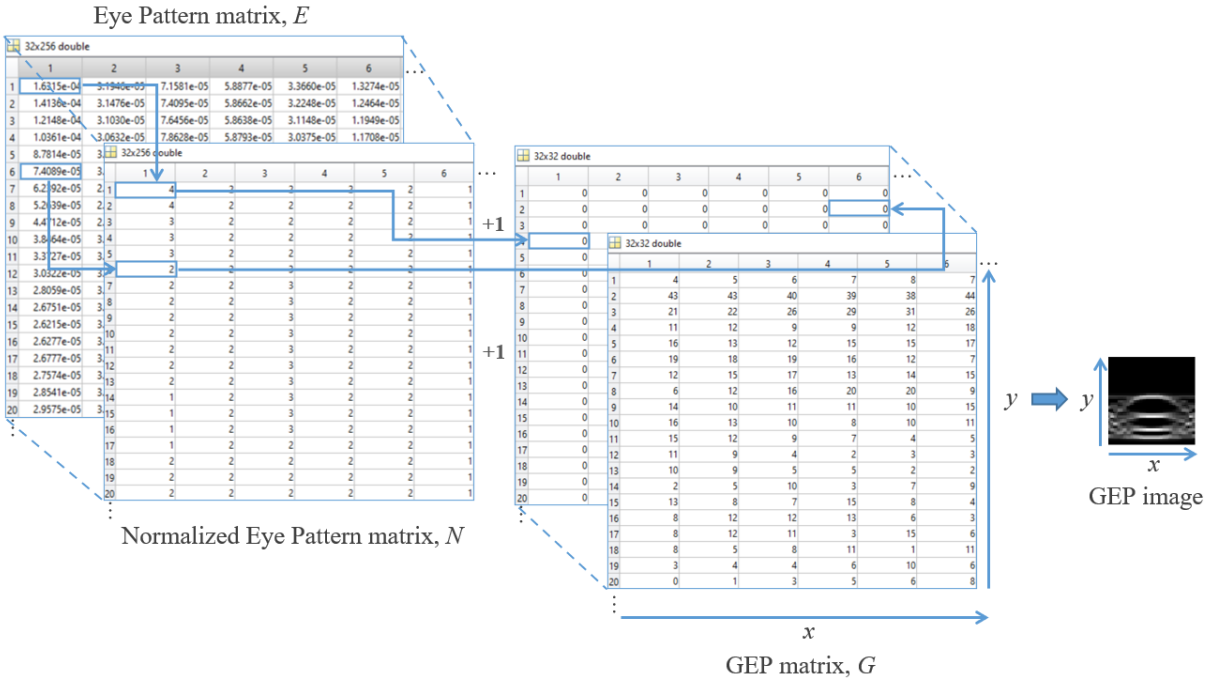


FIGURE 3.9. Example of the process to transform eye-patterns into 32×32 GEP images.

3.3.3. CNN architecture

In this work, the CNN model is developed on the MATLAB platform using the Deep Learning Toolbox and Deep Network Designer. The proposed CNN architecture is based on [**19, 20**] and is schematically represented in Fig. 3.10. First, normalization is applied by re-scaling the data in the range $[-1, 1]$ every time a GEP image is forward propagated through the input layer of the CNN. This architecture has five convolutional (Conv) layers, C1 to C5, that pass the GEP images through a set of convolutional kernels with stride (1,1). The layers C1 and C2, where both kernel sizes are 5×5, produce, respectively, 32 and 64 feature maps and the layers C3, C4 and C5, where the corresponding kernels sizes are 3×3, produce, respectively, 128, 256 and 512 feature maps. All convolution layers add the required padding to the input, either a GEP image or feature map, to ensure that its border pixels are completely exposed to the filter and the resulting feature map has the same size as the input. The outputs of each convolutional layers are normalized using a batch normalization (BN) layer followed by a Rectified Linear Unit (ReLU) layer for a more effective and faster training [**38**]. After the first four Conv+BN+ReLU operations, a down-sampling is performed by an average pooling layer. In our CNN, we set the four pooling layers, P1, P2, P3 and P4, with 2×2 subsampling regions and stride (2,2). After feature extraction (C5+BN+ReLU), a dropout layer is placed to prevent overfitting and a fully connected layer followed by a regression layer are used to predict the BER.

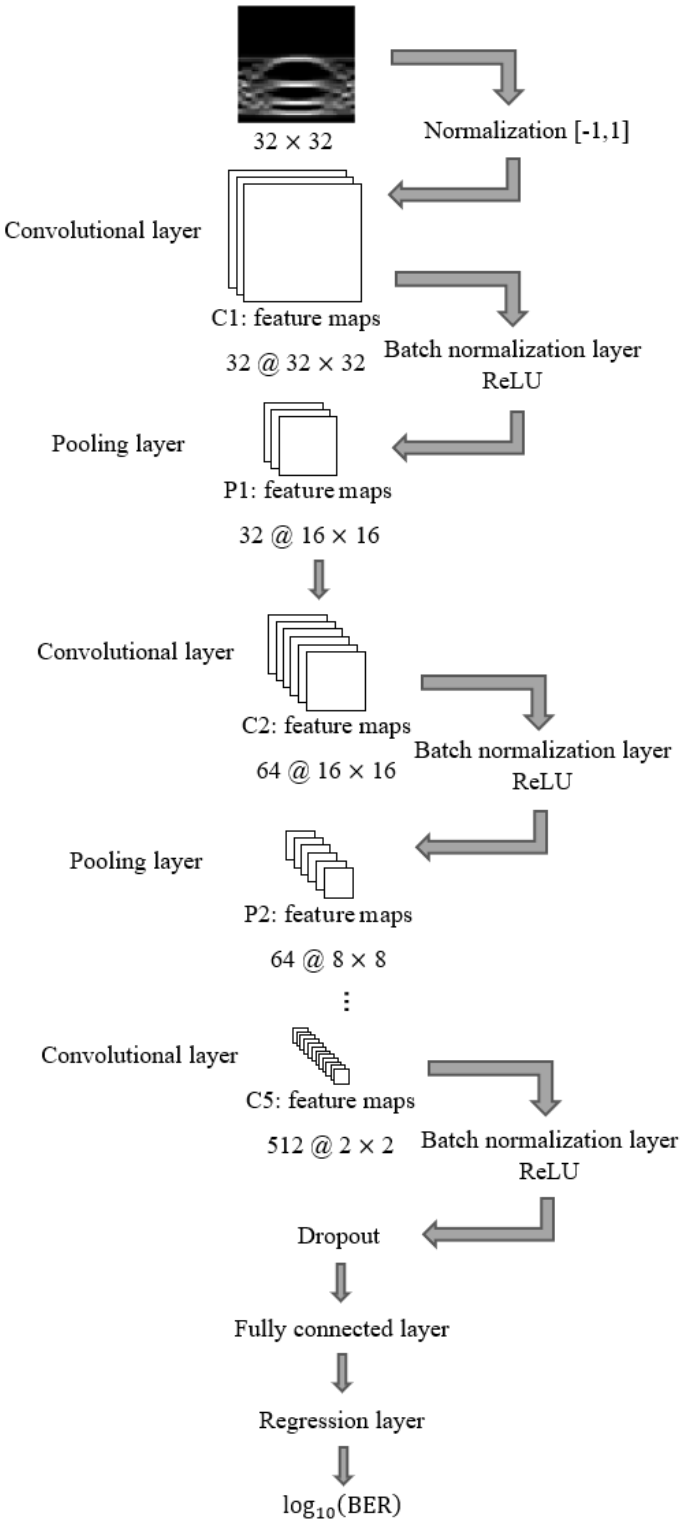


FIGURE 3.10. CNN architecture considered in this work to learn the BER from the GEP images.

3.3.4. Performance evaluation metric

In this work, the performance evaluation metric chosen to evaluate the CNN-based eye-pattern image analysis and BER prediction is the Root Mean Square Error (RMSE), which

has been widely used in the literature as a key CNN regression performance indicator [**53, 55, 56**], and is given by

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} e_i^2} \tag{3.18}$$

where n is the number of eye-patterns considered either in the training, test or validation sets and e_i is the difference between the expected values and the predicted values of the BER. In this work, we consider a RMSE below 0.1 as an acceptable prediction of the $\log_{10}$ (BER).

3.4. Conclusions

In this chapter, the CNN for eye-pattern analysis and the prediction of the BER from PAM4 inter-datacenter optical connections was presented. The optical telecommunication system equivalent model was described in section 3.1, where the DP-DCM developed to characterize the ICXT in MCFs was described in detail. The study of the BER distribution and respective eye-patterns in an optically amplified PAM4 link impaired by ICXT is performed in section 3.2, which showed that the product $|S_{mn}R_s|$ has a significant influence on the BERs distribution, since for $|S_{mn}R_s|$= 1000, the BER ranges and the distribution across this range are significant lower in comparison with the BER range and distribution obtained for $|S_{mn}R_s|$ =0.01. These different BER distributions and ranges may influence significantly the CNN training. It was also shown that system outage is more likely to occur for $|S_{mn}R_s|$ =0.01 and r=0 and less likely to occur for $|S_{mn}R_s|$= 1000 and r= 0.1, which is related to the BER distributions obtained for the two different $|S_{mn}R_s|$.

The main tasks performed by the CNN-based eye-pattern analysis and BER prediction was presented in section 3.3, with an emphasis on the data collection process from the optical telecommunication system equivalent model, eye-patterns pre-processing based on oscilloscopes' eye-patterns representation and the CNN architecture.

CHAPTER 4

CNN performance assessment and discussion

In this chapter, the performance of the CNN-based eye-pattern analysis and BER prediction in PAM4 inter-datacenter optically amplified short IM-DD connections impaired by ICXT is assessed. In section 4.1, the simulation parameters are introduced. The GEP size influence on the CNN performance is evaluated in section 4.2 and an assessment regarding the CNN non-deterministic behaviour is presented in section 4.3. The performance of the CNN-based eye-pattern analysis and BER prediction is evaluated in section 4.4 for fixed scale GEPs and in section 4.5 for dynamic scale GEPs. The main conclusions drawn from this chapter results are presented in section 4.6.

4.1. Simulation parameters

Throught this chapter, the CNN developed for eye-pattern analysis and BER prediction in PAM4 inter-datacenter optical connections impaired by intercore crosstalk with the architecture presented in section 3.3 is trained and tested with the system simulation and training CNN parameters shown in Table 4.1. The remaining simulation parameters of the optical communication system model that generates the data provided to the CNN models are presented in Table 3.1. As studied in section 3.2, the skew-symbol rate products $|S_{mn}R_s| = 1000$ and $|S_{mn}R_s| = 0.01$ and the extinction ratios $r = 0$ and $r = 0.1$ are relevant to the impact of ICXT, as they affect the BER range and distribution. Therefore, these parameters are also chosen to perform these studies. We also consider the crosstalk levels X_c = -16 dB, -14 dB and -12 dB.

In this work, the data used to obtain a CNN regression model is randomly splitted before each CNN training, where 70% is assigned as training data, 15% as validation data and 15% as test data. The main performance metric chosen to evaluate the CNN performance is the RMSE and we consider an acceptable error margin limit of 0.1, in logarithmic scale. The hyperparameters of the CNN were tuned and empirically set. The network is trained with a stochastic gradient descent with momentum (SGDM) optimizer that updates the weights and biases of the CNN. The maximum number of epochs, i.e., the number of times the training data is given to the CNN during training, is set to 30, since lower values resulted in worse performances and higher values did not enhance

the performance and resulted in a much higher computation time. A mini-batch size, that gives the number of training data samples sent to the CNN after each parameter update, with 8 observations at each iteration is used, since from the CNN training tests, it performed better when comparing to 16, 32 and 64 mini-batch sizes. The initial learning rate, which defines the frequency the optimizer updates the CNN parameters, is has been empirically set to 1×10^{-4} and reduced by a factor of 0.1 after 20 epochs.

Parameters	**Value**
Skew-symbol rate product	$\|S_{mn}R_s\| = 1000, \|S_{mn}R_s\| = 0.01$
ICXT levels	X_c = -16 dB, X_c = -14 dB, X_c = -12 dB
Extinction ratio	$r = 0$, $r = 0.1$
Data splitting	Training data = 70%, validation data = 15% and test data = 15%
Optimizer	SGDM
Mini batch size	8
Epochs	30
Initial learning rate	1×10^{-4}
Learning rate drop factor	0.1
Learning rate drop period	20
Error margin limit (logarithmic scale)	0.1

TABLE 4.1. System simulation parameters and CNN hyperparameters.

The total ammount of data collected from the PAM4 optical telecommunication link equivalent model considering different system parameters is presented in Table 4.2. Notice that the BER range is not the same for all optical links studied, since the BER values distribution is much dependent on the link parameters as seen in section 3.2. However, inside each logarithmic BER interval of 0.1, it is guaranteed that there are 1000 eye-patterns, for a good data balacing.

4.2. Grayscale eye-pattern size influence on the CNN performance

In this section, the impact of the GEP images size, 32×32, 64×64, 128×128 and 256×256, on the CNN performance is studied. To perform this study, an optical link with X_c = -14 dB, r=0.1 and $|S_{mn}R_s| = 1000$ is considered.

Fig. 4.1 shows the RMSE values obtained for the BER predictions of each GEP size under study after ten CNN trainings. After splitting, the training, validation and test data remained the same during the ten trainings. By doing this, the variability of the RMSE results for each GEP size is only influenced by the CNN training algorithm. It can be observed that, for all GEP sizes, the RMSEs differ for each training, which confirms the non-deterministic behaviour of the CNN algorithm. There is no significant difference

Link type		N_{EP}	$\log_{10}$(BER) range
X_c = -16 dB	$\lvert S_{mn}R_s\rvert = 1000$, $r = 0$	16000	[-3.8, -2.2]
	$\lvert S_{mn}R_s\rvert = 1000$, $r = 0.1$	6000	[-3.7, -3.1]
	$\lvert S_{mn}R_s\rvert = 0.01$, $r = 0$	26000	[-4.5, -1.9]
	$\lvert S_{mn}R_s\rvert = 0.01$, $r = 0.1$	21000	[-4.3, -2.3]
X_c = -14 dB	$\lvert S_{mn}R_s\rvert = 1000$, $r = 0$	25000	[-3.7, -1.2]
	$\lvert S_{mn}R_s\rvert = 1000$, $r = 0.1$	14000	[-3.3, -1.9]
	$\lvert S_{mn}R_s\rvert = 0.01$, $r = 0$	34000	[-4.5, -1.1]
	$\lvert S_{mn}R_s\rvert = 0.01$, $r = 0.1$	33000	[-4.4, -1.1]
X_c = -12 dB	$\lvert S_{mn}R_s\rvert = 1000$, $r = 0$	20000	[-3.1, -1.2]
	$\lvert S_{mn}R_s\rvert = 1000$, $r = 0.1$	10000	[-3, -2]
	$\lvert S_{mn}R_s\rvert = 0.01$, $r = 0$	35000	[-4.3, -0.8]
	$\lvert S_{mn}R_s\rvert = 0.01$, $r = 0.1$	34000	[-4.3, -1]

TABLE 4.2. Data collected from the optical telecommunication system equivalent model described in section 3.1, for different crosstalk levels, extinction ratio and skew-symbol rate products.

between the training and test RMSEs for each training attempt, from which we conclude that the model does not present overfitting or underfitting. Table 4.3 presents the average RMSEs and respective standard deviations of the ten CNN trainings for each GEP size corresponding to the results presented in Fig. 4.1.

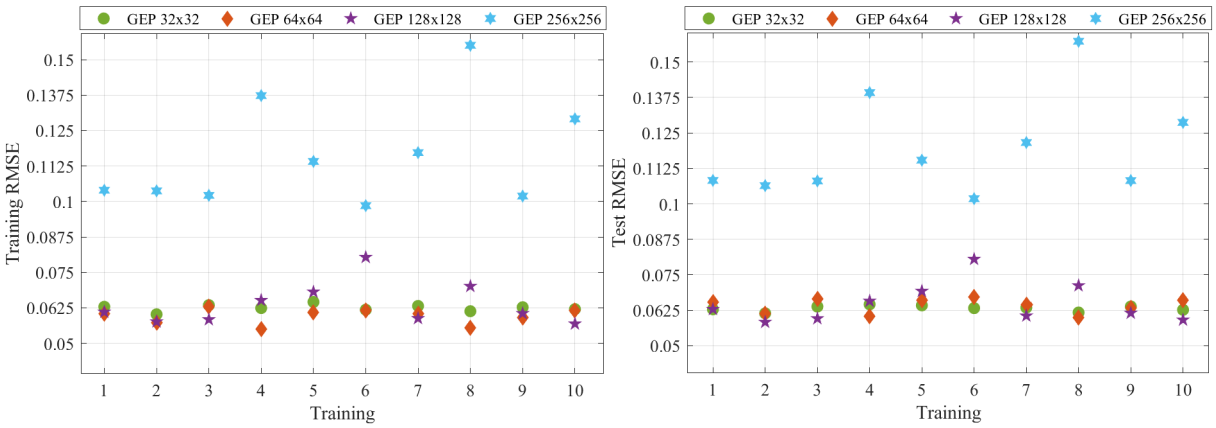


FIGURE 4.1. RMSE of the BER prediction obtained in ten trainings of the CNN for different GEP sizes for an optical link with X_c = -14 dB, r=0.1 and $|S_{mn}R_s| = 1000$.

The 256×256 GEP has the worst performance, since it leads to a higher average RMSE, of 1.16×10^{-1} for the training set and 1.20×10^{-1} for the test set, for all ten trainings in comparison with the other GEP sizes. For 256×256 GEPs, a more variability of the RMSE results is observed for the ten trainings, as can be seen in Fig. 4.1, and confirmed by its highest standard deviation. Fig. 4.1 and Table 4.3 also show that the CNN trained with 32×32, 64×64 and 128×128 GEPs present similar training and test RMSEs, from which we can conclude that the current CNN archicteture is more adequate for smaller

GEP sizes. As the RMSEs are very similar, there is the advantage in training the CNN with 32×32 GEP images, due to lower computation time.

GEP size	**Avg training RMSE**	**Std training RMSE**	**Avg test RMSE**	**Std test RMSE**
32×32	6.25×10^{-2}	1.2×10^{-3}	6.32×10^{-2}	1.0×10^{-3}
64×64	5.95×10^{-2}	2.7×10^{-3}	6.41×10^{-2}	2.7×10^{-3}
128×128	6.37×10^{-2}	7.4×10^{-3}	6.49×10^{-2}	7.0×10^{-3}
256×256	1.16×10^{-1}	1.87×10^{-2}	1.20×10^{-1}	1.76×10^{-2}

TABLE 4.3. Average (Avg) RMSE and standard deviation (Std) of the BER predictions obtained in ten trainings of the CNN for different GEP sizes for an optical link with X_c = -14 dB, r=0.1 and $|S_{mn}R_s| = 1000$.

4.3. CNN non-deterministic behaviour assessment

As mentioned in the section 4.2, the CNN algorithm presents a non-deterministic behaviour, which means that different CNN trainings using the same data splitting can perform differently, i.e., the predicted BERs are not the same. This is directly reflected in the RMSE and, therefore, in the CNN performance evaluation. To study this effect and determine the reliability of the CNN regarding the variability of results, a further study is performed, where training the CNN with the same percentage of data splitting, but with the data points distributed differently is also assessed. This situation is denoted as "different data splitting" and the corresponding RMSE results from the training, validation and test data are presented in Fig. 4.2. The results with the same splitting along the ten trainings shown in Fig. 4.2 correspond to the ones presented in Fig. 4.1. As seen in Fig. 4.2 a), for the training data points, the highest RMSE is obtained in the fifth training using the same splitting and, in general, the different splitting approach present lower RMSEs. However, Figs. 4.2 b) and c) do not present the same behaviour, since with the validation data points obtained similar RMSEs have been obtained and the test data points give, in general, lower RMSEs with the same splitting during the ten trainings.

To compare the RMSEs of the two training approaches, we resort to statistical hypothesis testing, using the parameters shown in Table 4.4, where the null hypothesis under test is the nonexistence of a significant statistical difference between the RMSEs of the two training approaches and whether this difference is due to the different splitting and not due to chance. By performing the Shapiro-Wilk test, using a right-tailed normal distribution, both training approaches do not reject the null hypothesis that their RMSEs

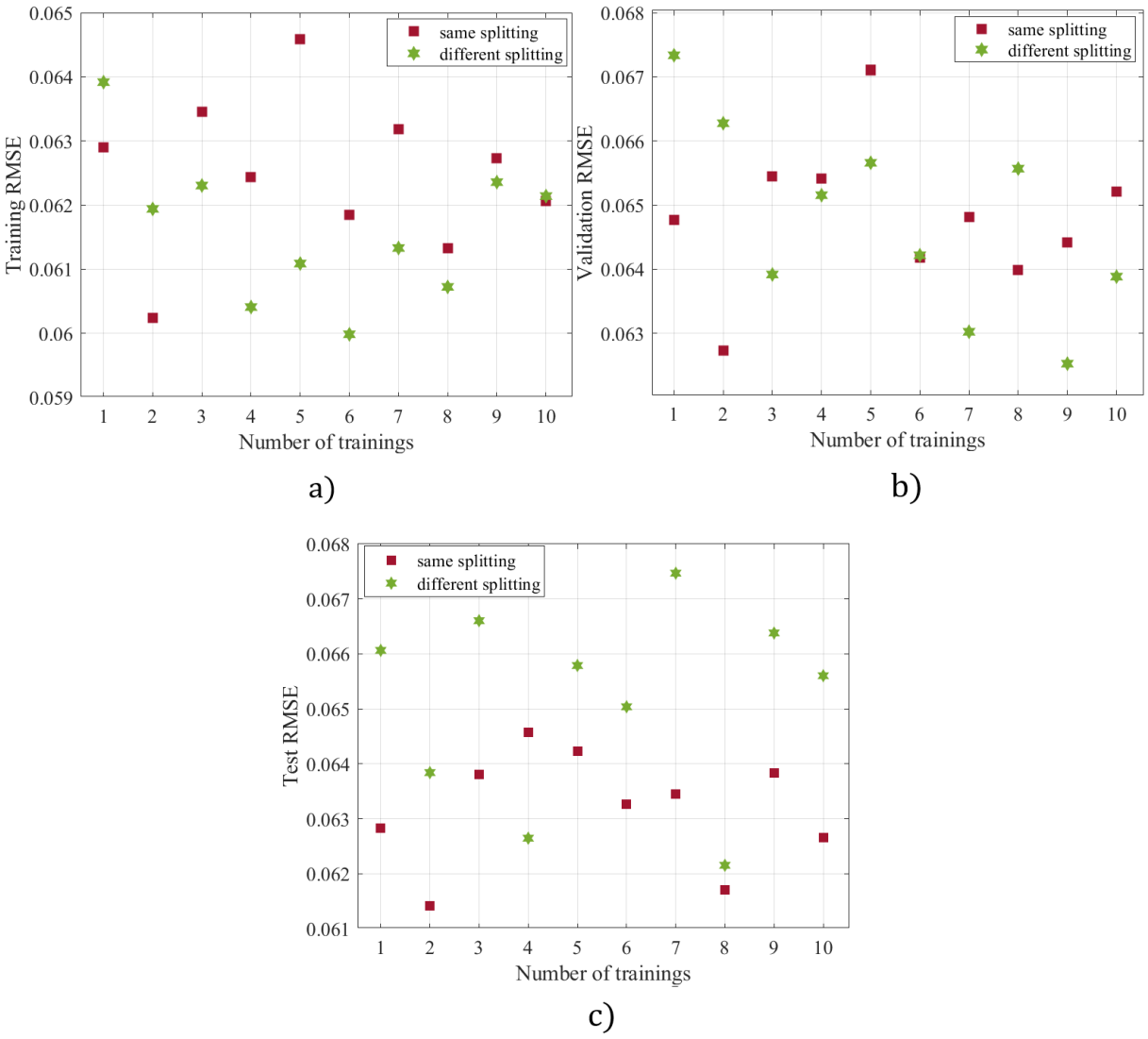


FIGURE 4.2. RMSE of the BER prediction obtained with different training, validation and test data splitting before each one of the ten CNN trainings performed, labeled as different splitting, and the same data splitting during the ten trainings, labeled as same splitting.

are normally distributed, since at the 5% significance level, the p-value is 0.392039 for different data splitting and without data splitting the p-value is 0.693304, which are higher than 5%. One of the most used techniques in statistical hypothesis testing is the Student's t-test, used when the difference between samples follows a normal distribution with mean equal to zero and unknown variance. A paired-sample t-test is performed [**57**] between the two training approaches, where at the 5% significance level, the t-value with a degree of freedom $D_f = 9$ and the p-value are, respectively, $t_{(9)} = 3.7166$; p=0.0048. This means that the null hypothesis is to be rejected, i.e., a significant statistical difference between the RMSEs of the two training approaches.

Even so, as the number of samples used to perform these tests can be considered statistically small, as an alternative to the t-test, a Wilcoxon signed rank test, which is non-parametric, is performed. The p-value obtained at the 5% significance level is p=0.0137, which means that the null hypothesis is also rejected. Thus, a conclusion that can be drawn regarding the results of both tests is that, even with a balanced range of synthetic data, different data splitting per each independent CNN training combined with the non-deterministic behaviour of the CNN algorithm leads to a little higher variability of the RMSE results as well as higher RMSEs than with the same data splitting.

To study the effect of the variability of results, the effect size between the RMSEs of the two types of training is measured through the Hedges' g formula given by [**58**]

$$g = \frac{\overline{x_1} - \overline{x_2}}{s_p} \times \frac{n-3}{n-2.25}\sqrt{\frac{n-2}{n}} \tag{4.1}$$

where $\overline{x_1}$ is the average RMSE of the test set of the CNN trained with different splitting, $\overline{x_2}$ is the average RMSE of the test set of the CNN trained with the same splitting, n_1 and n_2 are the number of independent CNN trainings performed, s_p is the pooled standard deviation given by $s_p = \sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2)}{(n_1-1)+(n_2-1)}}$, $n = n_1 + n_2$ and $\frac{n-3}{n-2.25}\sqrt{\frac{n-2}{n}}$ is a bias correction typically recommended when $n < 50$ [**59**].

From (4.1), the obtained $g = 6.06 \times 10^{-4}$, which means that a really small effect between the two training approaches is observed. Therefore, we assume that the CNN performance studies presented in this chapter, following a random data splitting, are reliable regarding the variability of results.

Type of training	**Avg training RMSE**	**Std training RMSE**	**Avg test RMSE**	**Std test RMSE**
Different splitting	6.16×10^{-2}	1.2×10^{-3}	$\overline{x_1} = 6.52 \times 10^{-2}$	1.7×10^{-3}
Same splitting	6.25×10^{-2}	1.2×10^{-3}	$\overline{x_2} = 6.32 \times 10^{-2}$	1.0×10^{-3}

TABLE 4.4. Average (Avg) training and test RMSE and the corresponding standard deviations (Std) obtained with different data splitting and the same data splitting considering ten CNN trainings using 32×32 GEPs.

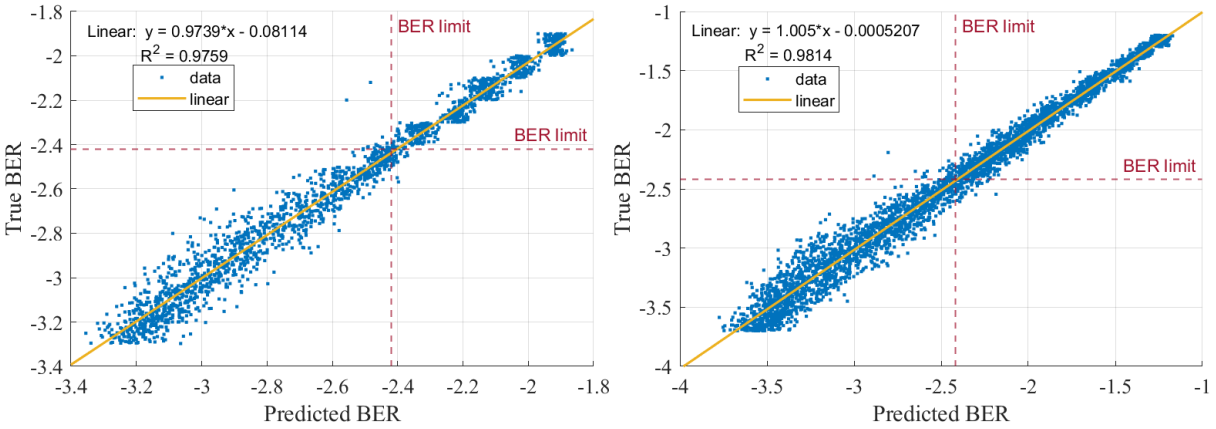
4.4. CNN performance with fixed scale GEP images

Firstly, the accuracy of the BER prediction is assessed by training the considered CNN from the eye-pattern analysis and test the corresponding regression model for each type of optical link. The data collected from the optical equivalent model used to perform this and subsequent studies is presented in Table 4.2. In this first study, regarding the

pre-processing step of the CNN-based model, each eye-pattern is normalized taking into account the maximum amplitude that occurred in all eye-patterns obtained for the corresponding type of optical link under study, in order to obtain proper GEP images. This normalization is denoted has having a fixed scale and is explained with detail in section 3.3.2.

Fig. 4.3 shows the BER prediction distribution of the test set from the CNN trained for the case of an optical link with X_c = -14 dB, $|S_{mn}R_s|$ = 1000, for a) r=0.1 and b) r=0. The "linear" curve represents a linear regression of the data points (generated with the MATLAB Basic fitting tool) and the coefficient of determination, R^2, indicates that, for the case of Fig. 4.3 a) the linear fit of the data explains 97.6% of its variance. In both cases, the CNN shows a smaller dispersion of results in relation to the "linear" curve in predicting BERs above the BER limit. A possible reason for this behaviour can be drawn by comparing these results with the BER distributions presented in Fig. 3.5 a), where, for both cases of extinction ratio, less diversity of amplitudes with BERs above the BER limit is obtained, making it easier for the CNN to learn and predict the BER based on the GEPs, because probably the GEPs are more distinguishable between them than the ones with lower BERs. In Fig. 4.3 a), a "clustering" type of behaviour is seen above the BER limit, which can be possibly caused by a higher similarity between GEPs obtained for a particular BER above the BER limit, since high BERs are less likely to occur. Another possible reason is that the GEPs that lead to less likely higher BERs for r=0.1, have been obtained with more eye-patterns generated using the random perturbation described in section 3.3.1. This random perturbation may not lead to meaningful differences between the eye-patterns obtained, leading to this clustering effect. For both cases, each CNN presents a RMSE lower than the error margin limit of 0.1. Even so, the CNN trained with $r = 0.1$ led to a RMSE of 0.066, which outperformed the RMSE of 0.099 of the one obtained with $r = 0$.

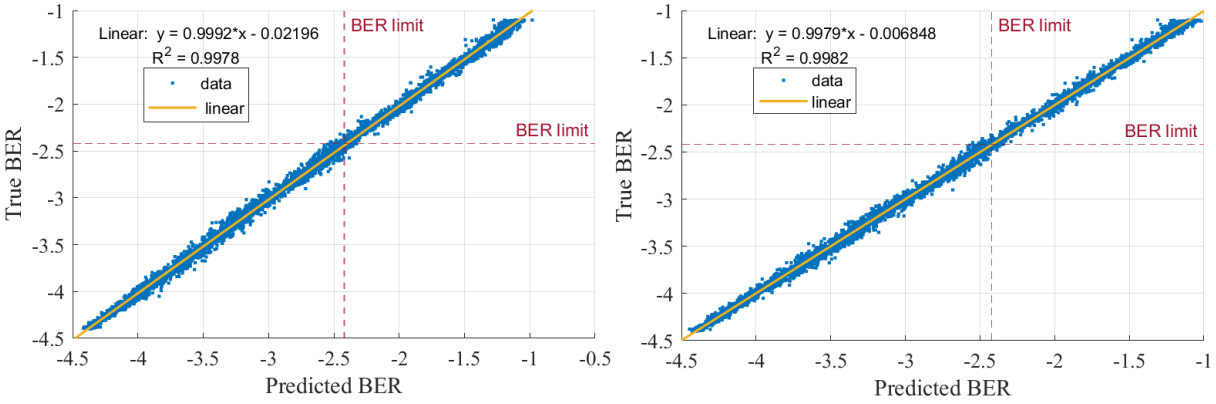
Fig. 4.4 shows the BER prediction distribution of the test set from the CNN trained for the case of an optical link with X_c = -14 dB, $|S_{mn}R_s|$ = 0.01 for a) r=0.1 and b) r=0. In comparison with the results presented in Fig. 4.3, lower RMSE values and less dispersion of results in relation to the "linear" curve are obtained. For both cases, each CNN presents a RMSE lower than the error margin limit and the RMSEs are very similar for the two extinction ratios. As seen in Fig. 3.5 b), for $|S_{mn}R_s| = 0.01$, a wider range of amplitudes with BERs either below or above the BER limit is obtained, which leads to



a) r = 0.1, RMSE = 0.066 b) r = 0, RMSE = 0.099

FIGURE 4.3. BER prediction distribution of the test set from the CNN trained for the case of an optical link with X_c = -14 dB, $|S_{mn}R_s|$ = 1000 for a) r=0.1 and b) r=0. The BER limit that leads to an outage probability is also depicted.

a superior performance of the CNN in comparison with the case represented in Fig. 4.3. This can possibly be explained due to a much higher representation of eye-patterns with distinct features obtained without using the random perturbation, leading to a higher number of GEPs with useful information for the CNN training.

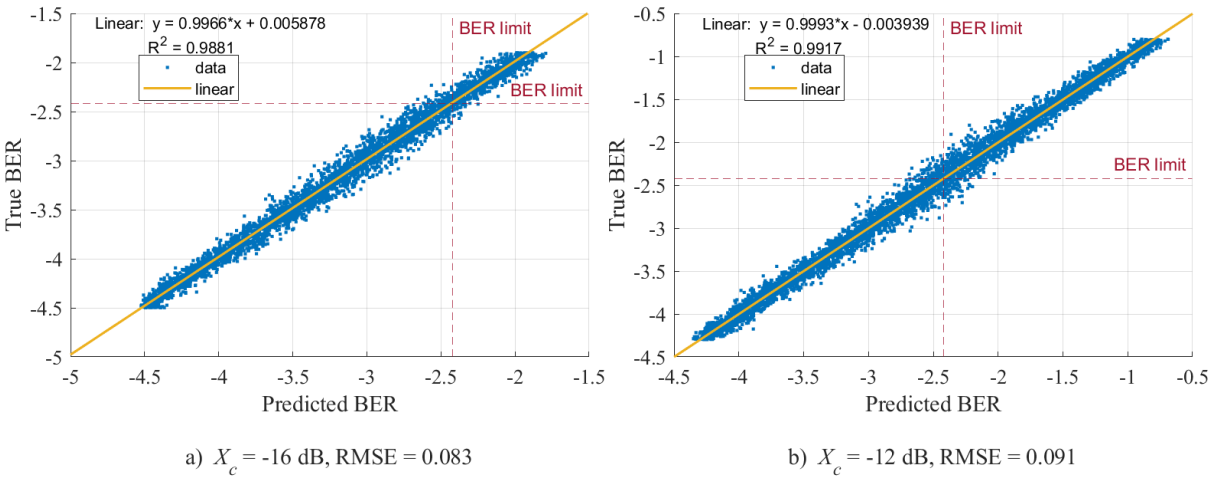


a) r = 0.1, RMSE = 0.049 b) r = 0, RMSE = 0.041

FIGURE 4.4. BER prediction distribution of the test set from the CNN trained for the case of an optical link with X_c = -14 dB, $|S_{mn}R_s|$ = 0.01 with a) r=0.1 and b) r=0.

Considering different levels of crosstalk, Fig. 4.5 shows the BER prediction distribution of the test data from the CNN trained for the case of an optical link with $|S_{mn}R_s|$ = 0.01 and r=0 for a) X_c = -16 dB and b) X_c = -12 dB. Figs. 4.3 a) and b), show a higher dispersion of results compared to Fig. 4.4 and the obtained RMSEs are

close to the margin error, being 0.083 and 0.091 obtained, respectively, for $X_c = -16$ dB and $X_c = -12$ dB.



a) X_c = -16 dB, RMSE = 0.083 b) X_c = -12 dB, RMSE = 0.091

FIGURE 4.5. BER prediction distribution of the test set from the CNN trained for the case of an optical link with $|S_{mn}R_s| = 0.01$ and r=0 for a) X_c = -16 dB and b) X_c = -12 dB.

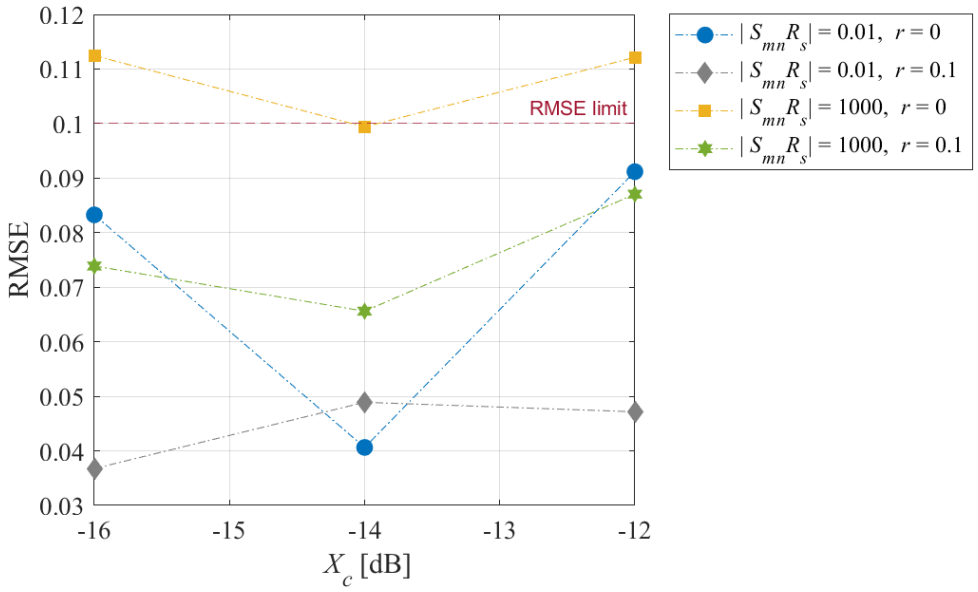


FIGURE 4.6. RMSE of the CNN models trained and tested with fixed scale GEPs as a function of X_c, for $|S_{mn}R_s| = 0.01$ and r=0; $|S_{mn}R_s| = 0.01$ and r=0.1; $|S_{mn}R_s| = 1000$ and r=0 and $|S_{mn}R_s| = 1000$ and r=0.1.

In Fig. 4.6, the summary of the RMSEs obtained in the previous studies is presented. Overall, each CNN model was able to predict the $\log_{10}$ (BER) without surpassing the RMSE limit of 0.1, except for the cases of X_c = -16 dB and X_c = -12 dB with $|S_{mn}R_s| =$ 1000 and r=0. For all CNNs trained, this case, r=0 and $|S_{mn}R_s| = 1000$, is the one that it is more difficult to train, leading to a worst BER prediction. The best predictions (with

lower RMSE) are obtained for the case of $|S_{mn}R_s| = 0.01$ and r=0.1, being the RMSE lower than 0.05, for all the crosstalk levels. In fact, with r=0.1, the RMSEs obtained with the different crosstalk levels are more similar, than in the case with r=0.

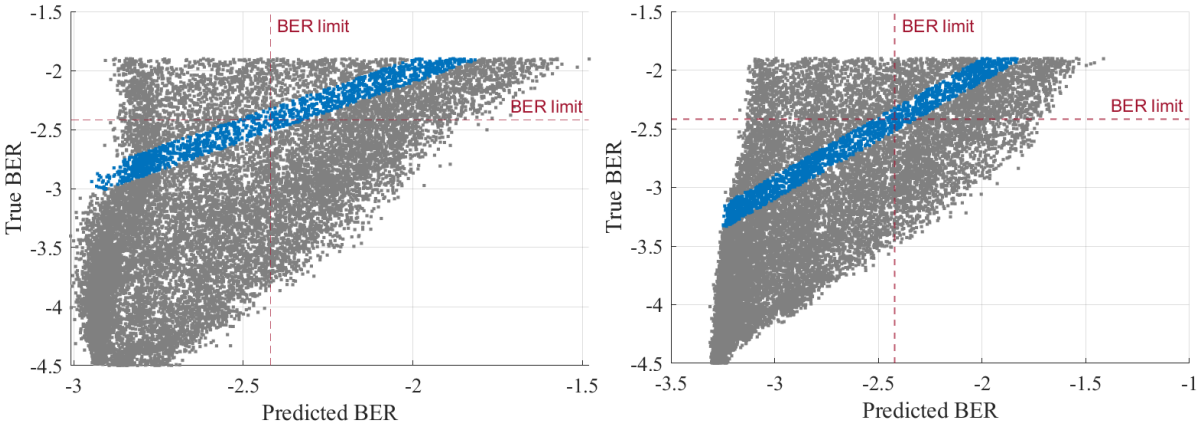


FIGURE 4.7. BER prediction distribution of optical links with X_c = -16 dB, $|S_{mn}R_s| = 0.01$, r=0 and GEPs normalized with the eye-pattern maximum amplitude obtained with X_c = -16 dB and b) GEPs normalized with the eye-pattern maximum amplitude obtained with X_c = -16 dB and X_c = -14 dB from a CNN trained for optical links with X_c = -14 dB, $|S_{mn}R_s| = 0.01$ and r=0. The blue data points represent the predictions with a margin error below 0.1.

Fig. 4.7, shows the results obtained for a CNN trained for the case of optical links with X_c = -14 dB, $|S_{mn}R_s| = 0.01$ and r=0 and tested with GEPs from optical links with X_c = -16 dB, $|S_{mn}R_s| = 0.01$ and r=0. In Fig. 4.7 a), the GEPs were normalized with the eye-pattern maximum amplitude obtained in all eye-patterns simulated with X_c = -16 dB and b), the GEPs were normalized with the eye-pattern maximum amplitude obtained in all eye-patterns generated for both crosstalk levels of -16 dB and -14 dB. As the GEPs's true BER values increase, both CNN models predictions are more inconsistent, since a wider range of predicted BERs is obtained. In both figures, the linear behaviour expected of the regression model is not observed. Fig. 4.7 a) presents a higher RMSE than Fig. 4.7 b). In Fig. 4.7 b), for $\log_{10}$ (BER) ≈ -2, the predicted BERs vary between ≈ -3.13 and ≈ -1.42 and for $\log_{10}$ (BER) between -4.5 and -1.9, the CNN predicts $\log_{10}$ (BER) around -3.26. These results show that the CNN is unable to predict correctly the BER of test GEPs obtained with other crosstalk level, meaning that the CNN is not capable of generalization.

However, Fig. 4.8 shows a CNN trained and tested with eye-patterns obtained with all the crosstalk levels, with X_c = -16 dB, -14 dB and -12 dB , $|S_{mn}R_s| = 0.01$, r=0

and $\log_{10}$ (BER) $\in [-3.1, -1.9[$, which is able to predict more correctly the BER from GEPs with different crosstalk levels, leading to a RMSE of 0.082 is obtained. This indicates that the CNN needs to train with different optical parameters to perform a better generalization in predicting the BERs.

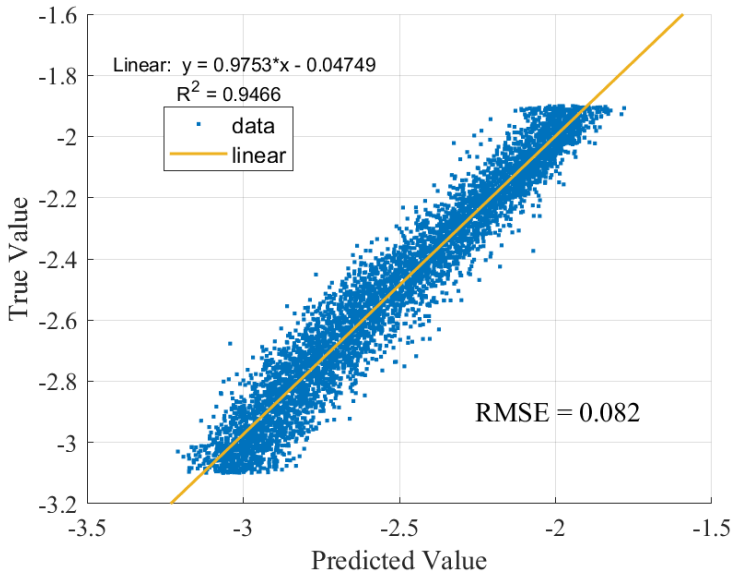


FIGURE 4.8. BER prediction distribution of the test set from the CNN trained for the case of optical links with X_c = -12 dB, X_c = -14 dB, X_c = -16 dB, $|S_{mn}R_s| = 0.01$, r=0.

Fig. 4.9 shows another situation where the CNN is completely unable to predict correctly the $\log_{10}$ (BER) from the GEPs. In this case, we are assessing if a CNN trained for a specific skew-symbol rate product can generalize its BER predictions to a different skew-symbol rate product. Fig. 4.9 a) presents the results, considering X_c = -14 dB and r=0, of the CNN model trained for the case of an optical links with X_c = -14 dB, $|S_{mn}R_s| = 0.01$ and tested for the case of an optical link with $|S_{mn}R_s| = 1000$. Fig. 4.9 b) presents the results, considering X_c = -14 dB and r=0.1, of the CNN trained for the case of an optical links with $|S_{mn}R_s| = 0.01$ and tested for the case of an optical link with $|S_{mn}R_s| = 1000$.

4.5. CNN performance with dynamic scale GEP images

The performance of the CNN presented in section 3.3 is also assessed regarding a different pre-processing step, where each eye-pattern is normalized taking into account its maximum amplitude to obtain the GEP image, which we denote as having a dynamic scale. As previously performed, a CNN is trained and tested considering different optical link parameters. The RMSEs obtained for the dynamic scale GEPs as a function of the

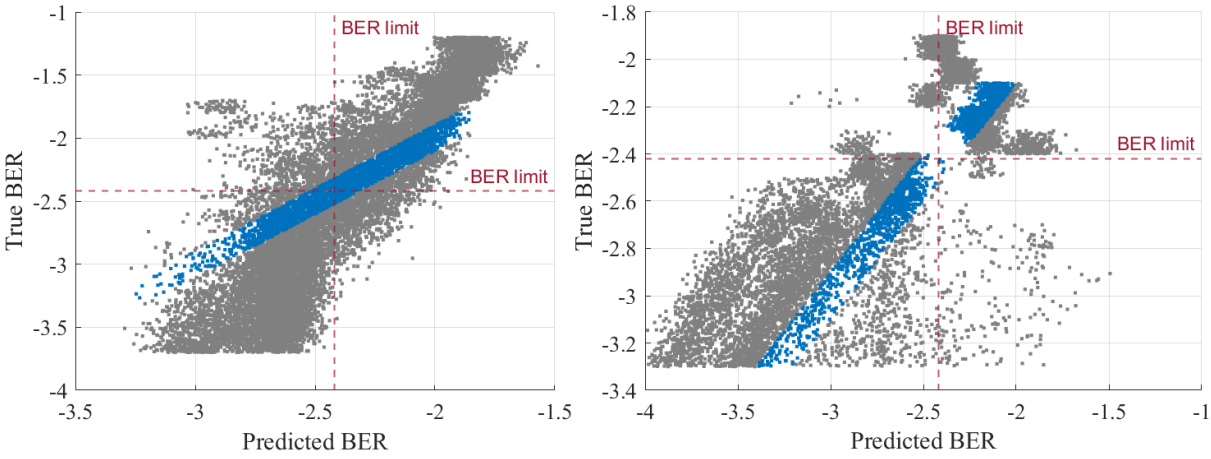


FIGURE 4.9. BER prediction distribution of optical links with a) X_c = -14 dB, $|S_{mn}R_s| = 1000$, r=0 tested in a CNN trained for the case of optical links with X_c = -14 dB, $|S_{mn}R_s| = 0.01$ and r=0 and with b) X_c = -14 dB, $|S_{mn}R_s| = 1000$, r=0.1 tested in a CNN trained for the case of optical links with X_c = -14 dB, $|S_{mn}R_s| = 0.01$ and r=0.1.

crosstalk level are shown in Fig. 4.10 and show a similar behavior to the one obtained for the fixed scale GEPs in Fig. 4.6. Each CNN model is able to predict the $\log_{10}$ (BER) without surpassing the RMSE limit of 0.1, except for the cases of X_c = -16 dB and -12 dB with $|S_{mn}R_s| = 1000$ and r=0.

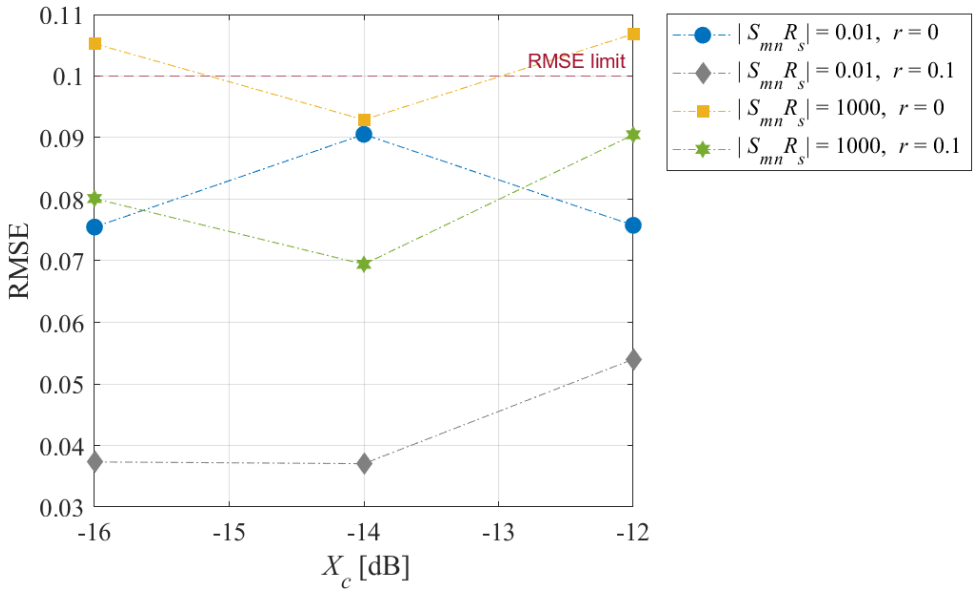


FIGURE 4.10. RMSE of the CNN models trained and tested with dynamic scale GEPs as a function of X_c, for $|S_{mn}R_s| = 0.01$ and r=0; $|S_{mn}R_s| = 0.01$ and r=0.1; $|S_{mn}R_s| = 1000$ and r=0 and $|S_{mn}R_s| = 1000$ and r=0.1 for dynamic scale GEPs.

The results presented in Fig. 4.11, show that the CNN regression model obtained for the case of optical links with X_c = -14 dB, $|S_{mn}R_s|$ = 0.01 and r=0 is also unable to predict correctly the BER of dynamic scale GEPs obtained from optical links with X_c = -16 dB, $|S_{mn}R_s|$ = 0.01 and r=0, where a RMSE = 0.285 is obtained. However, it seems more able to generalize, giving BERs are closer to have a linear behaviour comparing to the case presented in Fig. 4.9.

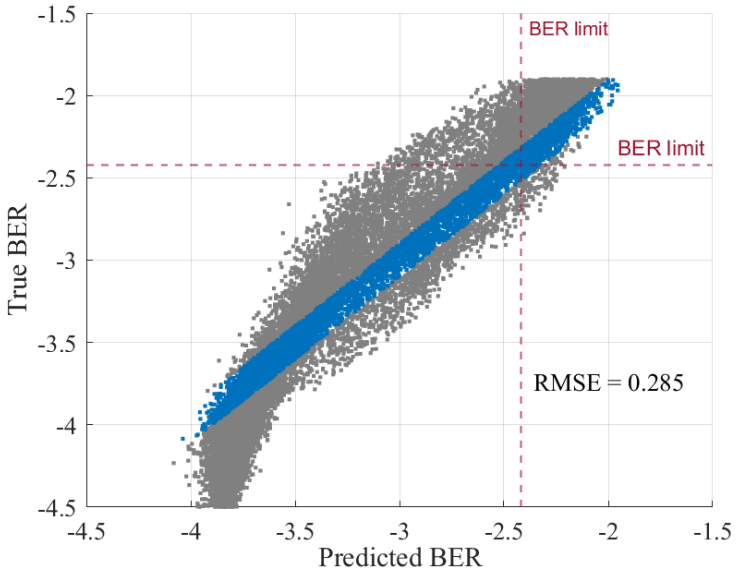


FIGURE 4.11. BER prediction distribution of GEPs from optical links with X_c = -16 dB, $|S_{mn}R_s|$ = 0.01 and r=0 from a CNN regression model obtained for the case of an optical link with X_c = -14 dB, $|S_{mn}R_s|$ = 0.01 and r=0.

Considering GEPs with dynamic scale, Fig. 4.12 a) presents the results of the CNN model trained for the case of optical links with X_c = -14 dB, $|S_{mn}R_s|$ = 0.01, r=0 and tested for the case of an optical link with $|S_{mn}R_s|$ = 1000. Fig. 4.12 b) presents the results, considering X_c = -14 dB, r=0.1, of the CNN model trained for the case of an optical links $|S_{mn}R_s|$ = 0.01 and tested for the case of optical link with $|S_{mn}R_s|$ = 1000. Comparing with Fig. 4.9, using GEP images with dynamic scale gives a better generalization to another skew-symbol rate product, being the results improved particularly for the case of Fig. 4.12 a), which gives a RMSE below the limit.

Fig. 4.13 shows the BER prediction distribution of the test set from a CNN trained with disjoint BER intervals, i. e., for the case of optical links with X_c = -16 dB and $\log_{10}$ (BER) $\in [-4.5, -3.9]$, X_c = -14 dB and $\log_{10}$ (BER) $\in [-3.8, -2.9[$, X_c = -12 dB and $\log_{10}$ (BER) $\in [-2.9, -0.8]$, for $|S_{mn}R_s|$ = 0.01 and r=0. This CNN is tested with the full $\log_{10}$ (BER) range considering eye-patterns obtained for the three crosstalk levels

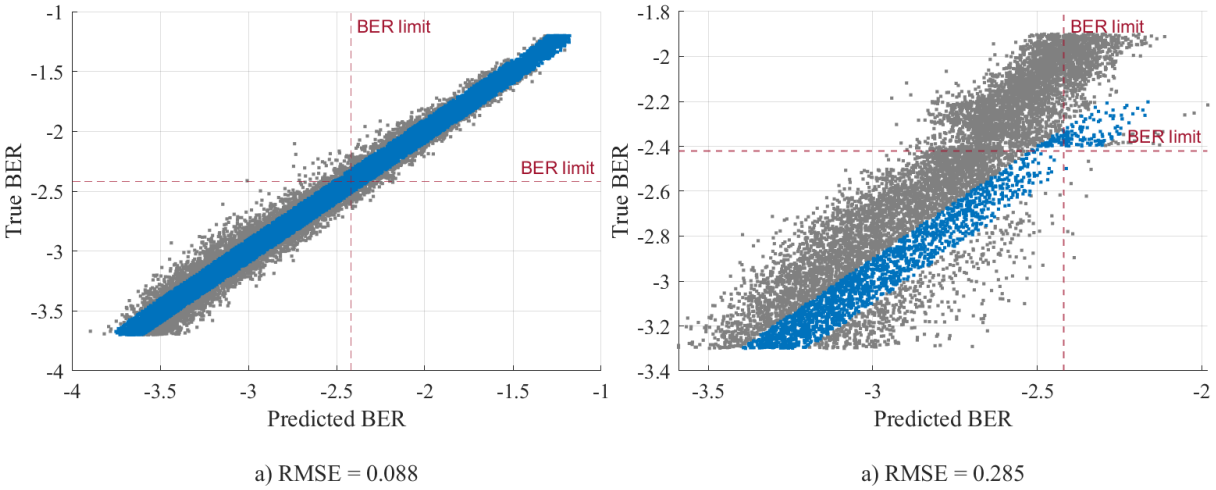


FIGURE 4.12. BER prediction distribution of GEPs from optical links with a) X_c = -14 dB, $|S_{mn}R_s| = 1000$, r=0 tested in the CNN regression model obtained for the case of optical links with X_c = -14 dB, $|S_{mn}R_s| = 0.01$ and r=0 and with b) X_c = -14 dB, $|S_{mn}R_s| = 1000$, r=0.1 tested in the CNN regression model obtained for the case of optical links with X_c = -14 dB, $|S_{mn}R_s| = 0.01$ and r=0.1.

and the results are shown in Fig. 4.14, which shows that, in this case, the CNN is able to predict the BER for the three crosstalk levels considered without surpassing the RMSE limit of 0.1.

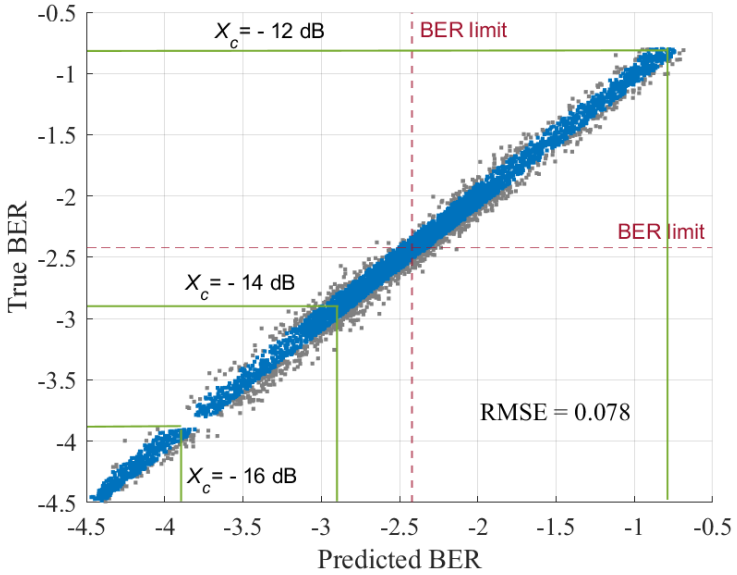


FIGURE 4.13. BER prediction distribution of the test set from a CNN trained for the case of optical links with X_c = -16 dB and $\log_{10}$ (BER) $\in [-4.5, -3.9]$, X_c = -14 dB and $\log_{10}$ (BER) $\in [-3.8, -2.9[$, X_c = -12 dB and $\log_{10}$ (BER) $\in [-2.9, -0.8]$, $|S_{mn}R_s| = 0.01$ and r=0.

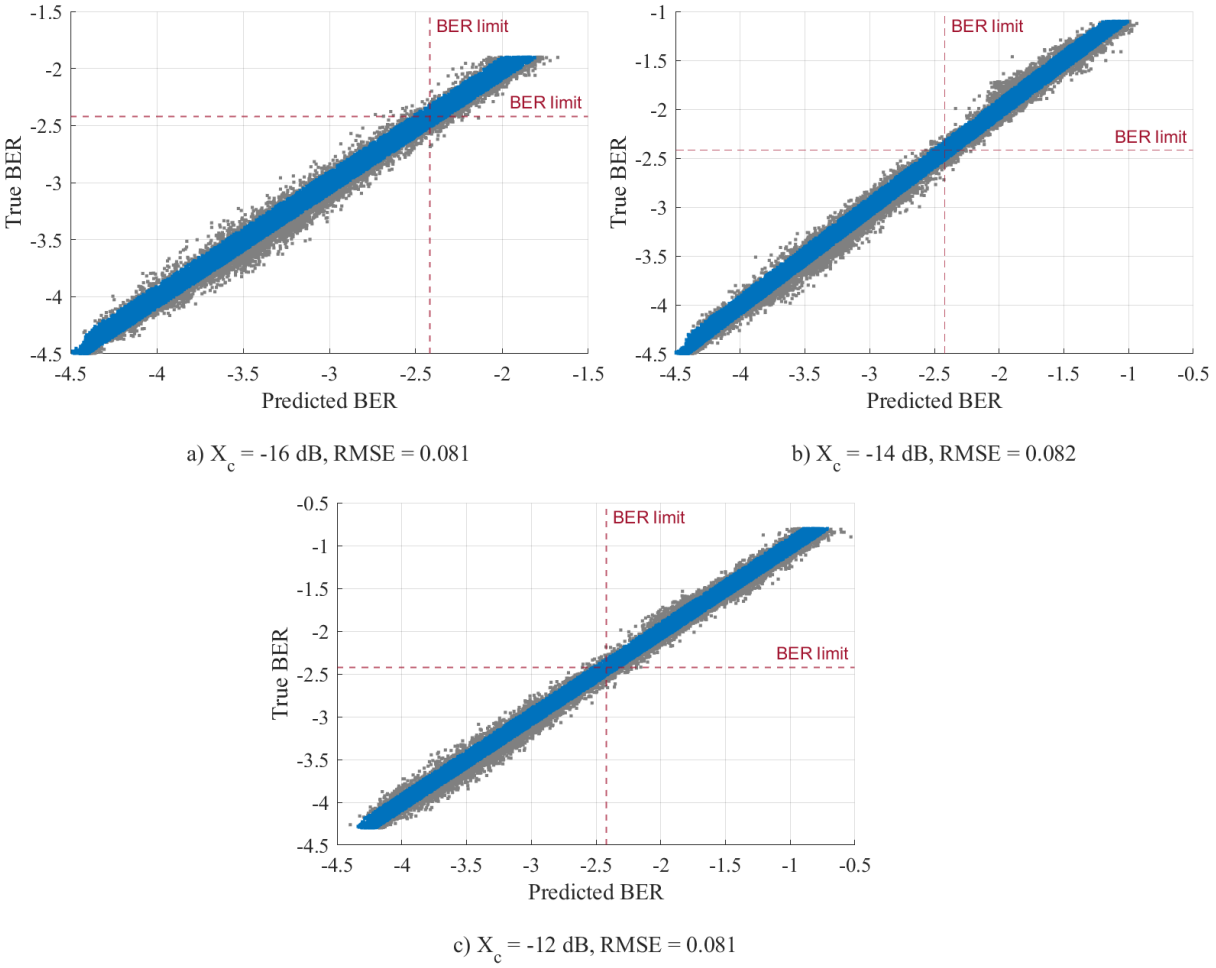


FIGURE 4.14. BER prediction distribution of GEPs from optical links with $|S_{mn}R_s| = 0.01$ and r=0, for a) X_c = -16 dB, b)X_c = -14 dB and c) X_c = -12 dB, trained for the case of optical links with X_c = -16 dB and $\log_{10}$ (BER) $\in [-4.5, -3.9]$, X_c = -14 dB and $\log_{10}$ (BER) $\in [-3.8, -2.9[$, X_c = -12 dB and $\log_{10}$ (BER) $\in [-2.9, -0.8]$.

4.6. Conclusions

In this chapter, the performance of the CNN-based eye-pattern analysis and BER prediction in PAM4 inter-datacenter optically amplified short IM-DD connections impaired by ICXT has been studied by assesssing the RMSE of the obtained BER predictions.

The studies performed took into consideration one interfering core, skew-symbol rate products $|S_{mn}R_s|$ = 1000 and 0.01, extinction ratios r = 0 and 0.1, crosstalk levels X_c = -16 dB, -14 dB and -12 dB and a link length of 80 km. Two different approaches regarding the GEPs have been assessed: fixed scale GEPs and dynamic scale GEPs. With the fixed scale GEP images, by training a CNN and test the corresponding regression model for each of the optical link parameters presented above, each CNN model was able to predict the $\log_{10}$ (BER) without surpassing the RMSE limit of 0.1, except for the cases

of X_c = -16 dB and -12 dB with $|S_{mn}R_s|$ = 1000 and r=0. The best performance was obtained with X_c = -16 dB, $|S_{mn}R_s|$ = 0.01 and r=0. We have also shown that training a CNN with a certain crosstalk level, $|S_{mn}R_s|$ and extinction ratio and trying to generalize the CNN to test data obtained from other crosstalk levels or other skew-symbol rate products, leads to innacurate BER predictions. The BER predictions can be improved by training the CNN with eye-patterns generated for all the different crosstalk levels studied, giving RMSEs below the limit.

The same tests have been performed with dynamic scale GEPs and the results obtained are very similar to the ones obtained with a fixed scale GEP images. However, when trying to generalize the CNN behavior to parameters values other than those trained, the dynamic situation leads to a better generalization. For both fixed and dynamic scale GEPs, the CNNs models trained and tested with $|S_{mn}R_s| = 1000$ and $r = 0$, have presented the worst performances.

CHAPTER 5

Conclusions and future work

This chapter summarizes the dissertation final conclusions and presents some proposals for future work.

5.1. Conclusions

In this work, the use of a CNN for eye-pattern image analysis and BER prediction in PAM4 datacenter optical connections supported by homogeneous WC-MCFs impaired by ICXT has been studied. A literature review has been presented in Chapter 2, which addresses the fundamental concepts related to this work, namely datacenter connections, PAM4 signal transmission in datacenter links supported by MCFs impaired by ICXT and an ML overview.

In Chapter 3, the equivalent model of the PAM4 optical telecommunication system equivalent model supported by MCFs was presented in section 3.1, being the DP-DCM developed to characterize the ICXT in MCFs described in detail. The study of the BER distribution and respective eye-patterns in the optically amplified PAM4 link impaired by ICXT has been performed in section 3.2, which showed that the product $|S_{mn}R_s|$ has a significant influence on the BERs distribution, since for $|S_{mn}R_s|$= 1000, the BER range and the distribution across this range is significant lower in comparison with the range obtained for $|S_{mn}R_s|$ =0.01. It was also shown that system outage is more likely to occur for $|S_{mn}R_s|$ =0.01 and r=0 and less likely to occur for $|S_{mn}R_s|$= 1000 and r= 0.1, which is related to the BER distributions obtained for the two different $|S_{mn}R_s|$. Then, in section 3.3, the main tasks of the CNN implemented for eye-pattern analysis and BER prediction has been presented, with an emphasis on the data collection process from the system equivalent model, eye-patterns pre-processing and the considered CNN architecture.

In chapter 4, the performance of the CNN-based eye-pattern analysis and BER prediction in PAM4 inter-datacenter optically amplified short IM/DD connections impaired by ICXT has been studied by assessing the RMSE of the obtained BER predictions. Two different approaches regarding the GEPs have been assessed: fixed scale GEPs and dynamic scale GEPs. With the fixed scale GEPs and by training a CNN and test the corresponding regression model for each of the optical link parameters considered, each

CNN model was able to predict the $\log_{10}$ (BER) without surpassing the RMSE limit of 0.1, except for the cases of X_c = -16 dB and X_c = -12 dB with $|S_{mn}R_s|$ = 1000 and r=0, being the RMSE only around 0.01 higher than the limit established.

Training the CNN with a certain crosstalk level, $|S_{mn}R_s|$ and r and testing with data with different optical link parameters, the CNN is unable to predict correctly the $\log_{10}$ (BER). These situations can be improved by training the CNN with eye-patterns collected for the different crosstalk levels or with dynamic scale GEPs. From an implementation point of view, a CNN based on dynamic GEPs seems a better solution, since it does not require a prior knowledge in terms of amplitudes from other eye-patterns. Each GEP is generated based only on the corresponding eye-pattern maximum amplitude. Considering dynamic scale GEPs, when the CNN was trained with data with a certain crosstalk level, $|S_{mn}R_s|$ and extinction ratio and tested with a different level of crosstalk, a good performance was obtained for r=0 while, for r=0.1, the RMSE surpassed the limit.

An additional test has been performed, by training the CNN with different $\log_{10}$ (BER) ranges for the different crosstalk levels, same $|S_{mn}R_s|$ and r and the CNN was able to predict the BERs for the three crosstalk levels along the total BERs range.

Furthermore, we have statistically studied that the CNN results presented, which follow a random data splitting, can be considered reliable in terms of variability of results.

As a final remark, the results obtained in this work confirm that CNN-based models can extract features from eye-patterns to predict the BER without prior knowledge of the transmitted signals, since only the eye-pattern image is used to extract information and predict the BER. Such ML algorithms can be seen as a potencial side-tool to provide additional monitoring information to existing optical systems.

5.2. Future work

Regarding the results obtained in this work, we suggest for future work:

- Generalize the CNN architecture to predict the BER with different crosstalk levels, extinction ratios, skew-symbol rate products and other optical link parameters, instead of having a CNN trained for each parameters case.
- Assess the performance of the CNNs trained with synthetic data for monitoring experimental data and generalize the CNN for these experiments.
- Train the CNN to predict the OP from the eye-patterns.
- Study of ICXT mitigation techniques using ML.

Appendices

APPENDIX A

Conference paper

Convolutional neural network for BER prediction in inter-datacenter optical connections impaired by inter-core crosstalk

Sofia Esteves[1], João Rebola[1,2] and Pedro Santana[1,3]
[1]*Instituto Universitário de Lisboa (ISCTE-IUL), Lisboa, Portugal*
[2]*Optical Communications and Photonics Group, Instituto de Telecomunicações, Lisboa, Portugal*
[3]*Information Sciences and Technologies and Architecture Research Center (ISTAR-IUL), Lisboa, Portugal*
E-mails: spees@iscte-iul.pt, joao.rebola@iscte-iul.pt, pedro.santana@iscte-iul.pt

***Abstract*—To meet the required future challenge of providing enough bandwidth to achieve high data traffic rates in datacenter links, four-level pulse amplitude modulation (PAM4) signals transmission in short-haul intensity modulation-direct detection (IM-DD) datacenters connections supported by homogeneous weakly-coupled multicore fibers has already been proposed. However, in such fibers, a physical effect known as inter-core crosstalk (ICXT) limits significantly the performance of short-reach connections by causing large bit error rate (BER) fluctuations that can lead to undesirable system outages. In this work, a convolutional neural network (CNN) is proposed for eye-pattern analysis and BER prediction in PAM4 inter-datacenter optical connections impaired by ICXT, with the aim of optical performance monitoring. The performance of the CNN is assessed using the root mean square error (RMSE). Considering PAM4 interdatacenter links with one interfering core and for different skew-symbol rate products, extinction ratios and crosstalk levels, the results show that the implemented CNN is able to predict the BER without surpassing the RMSE limit. The CNNs trained with different optical parameters obtained the best performance in terms of generalization comparing to CNNs trained with specific optical parameters. These results confirm that the CNN-based models can be able to extract features from received eye patterns to predict the BER without prior knowledge of the transmitted signals.**

***Index Terms*—bit error rate, convolutional neural network, inter-core crosstalk, machine learning, multicore fiber**

I. Introduction

As the number of network users and devices evolved, datacenters became crucial to handle large amounts of data due to their flexibility and scalability in computing and storage resources [1]. Even so, the number of devices connected to internet protocol (IP) networks is estimated to be more than three times the global population by the year of 2023 [2], which leads to a significant global data center IP traffic growth. This requires answering technologically to the future challenge of providing enough bandwidth to achieve such high data traffic rates in datacenter links.

The current approach to deal with capacity scalability in datacenter links is using multiple wavelengths (also known as lanes) carrying wavelength division multiplexing (WDM) channels, where each one of these channels carries a conventional on-off keying (OOK) signal [3]. Four-level pulse amplitude modulation (PAM4) has been proposed for datacenter connections and has already been standardized by the IEEE 802.3bs task force to enable rates of 50 and 100 Gbit/s per wavelength channel in datacenters connections [3]. Datacenter connections are usually categorized as intra-datacenter and inter-datacenter links, in relation to their maximum range, respectively, up to 10 km and 100 km [3].

Nowadays, datacenter connections typically rely on single core-single mode fibers (SC-SMFs), however, it is expected that such fibers will no longer fulfill the future capacity demands, as transmission in SC-SMFs is approaching its limit of 100 Tb/s with coherent detection and over 200 Gbit/s per lane with IM-DD [4], [5]. Multicore fibers (MCFs) have been proposed in order to overcome this capacity limit issue. In particular, homogeneous weakly-coupled multicore fibers (WC-MCFs) ensure the simplicity, low cost and power consumption design required by datacenter links. Even so, transmission in such MCFs is impaired by inter-core crosstalk (ICXT). This interference in MCFs occurs from power coupling between cores and is particularly relevant with the transmission of signals with the same wavelength in neighboring cores [6]. The ICXT has been experimentally shown to have a stochastic time evolution, which can result in high levels of ICXT in short time periods and can limit significantly the performance of such connections by causing large bit-error-rate (BER) fluctuations leading to undesirable system outage periods [7], [8]. Hence, ICXT monitoring in such connections is crucial to predict and maintain acceptable BERs.

Regarding predictive models, for either regression or classification, machine learning (ML) has been widely addressed. ML is a branch of artificial intelligence (AI) and it is, essentially, a modeling technique that finds a model, by itself, from a given training data [9], [10]. One of the most well known ML algorithms is the artificial neural network (ANN), which can behave as an universal approximation function mediator between the input data and output data after the training phase [11], [12]. The simplest architecture of an ANN is composed of an input layer, an hidden layer and an output layer. These layers contain nodes, or also called neurons, modeled as an activation function, with weights and biases, which are con-

tinuously updated during through a back propagation training process, until the output layer can produce the desired output [12]. With the growth of data and computation capacity, ML has expanded into deep learning (DL) and ANNs developed into deep neural networks (DNN) composed by multiple hidden layers and a high number of nodes [13]. This allowed to address a wider variety of problems and data types resorting to new algorithms with specific structures such as convolutional neural networks (CNNs), which allow to process images, by adding convolution, activation and pooling layers for feature extraction [13], [14]. However, one of the main challenges DL still faces is the shortage of available data and unbalanced data within the training data [15]. To address this issue, some techniques such as domain randomization have been studied [15], [16]. Domain randomization generates synthetic data similar enough to emulate the data under study [16]. Training with synthetic data generated through simulation can be cheaper, diverse, which prevents from training with unbiased data, and less time consuming compared to collecting real data [16]–[18]. The application of ML techniques to the physical layer of optical communication networks has already been addressed in several works and has gained a lot of interest by the research community in the last few years [9], [19], [20]. Hence, typical areas of application of ML in the physical layer of optical communications networks are fiber nonlinearity mitigation, modulation format (MF) recognition and optical performance monitoring (OPM). Regarding the latter, most recently, the use of CNNs has been investigated in [21], [22] for optical performance monitoring by analysing eye-patterns. In [21], a CNN-based technique is assessed to perform optical signal-to-noise ratio (OSNR) estimation and MF recognition, from eye-pattern images of several modulation formats, PAM4 being one of them. In [22] an eye-pattern analysis scheme based on a CNN for intensity modulation-direct detection (IM-DD) transmissions is also presented. Different eye-patterns of OOK and PAM4 signals from back-to-back (B2B) and up to 80 km link transmissions are sent to a CNN-based model that outputs eye-pattern characteristics, fiber link length, Q-factor and impairments recognition [22]. Given the good results obtained with these techniques, in this work, for performance monitoring, a CNN is used for eye-pattern analysis and BER prediction in PAM4 inter-datacenter optical connections impaired by ICXT.

II. Model Description

The schematic of the main tasks of the MATLAB simulator for the CNN-based eye pattern analysis and BER prediction in PAM4 inter-datacenter optical connections impaired by ICXT is illustrated in Fig. 1.

The first main task corresponds to data collection, where eye-patterns and the corresponding BERs are collected from the PAM4 optical communication system model described in section II-A. The eye-patterns are obtained by varying several optical link parameters, such as crosstalk level, skew-symbol rate product and extinction ratio. Before a CNN training, the eye-patterns are pre-processed, according to the

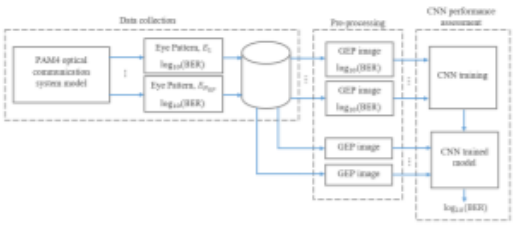


Fig. 1. Main tasks of the MATLAB-based CNN for eye-pattern analysis and BER prediction model.

different optical links under study, to obtain what we denote as grayscale eye-patterns (GEPs) images. After that, the CNN is trained with GEPs as inputs and BERs as outputs. Then, the CNN is tested with unseen GEP images.

A. Intercore crosstalk modelling

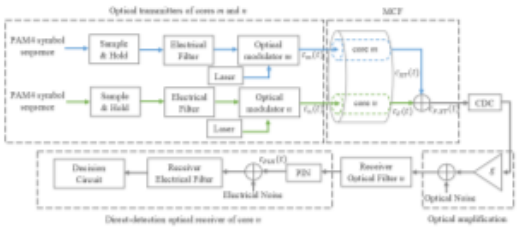


Fig. 2. System equivalent model to assess the impact of ICXT on a PAM4 optical communication link with DD and MCFs. Based on [7].

The optical telecommunication system equivalent model developed in [7] for a PAM4 inter-datacenter optical link supported by MCF is shown in Fig. 2.

The impact of ICXT on the performance of optically amplified PAM4 links for inter-datacenter connections has been analyzed in [7] by resorting to the dual polarization-discrete changes model (DP-DCM) that describes accurately the ICXT effect in homogeneous WC-MCFs [23]. The DP-DCM of the MCF used in [7], considers firstly only two cores, the interfering core m and the interfered core n as shown in Fig. 2. Two transmitters, one for each core, generate different signals where the signal transmitted along core m, $c_m(t)$, is the interfering PAM4 signal and the signal transmitted along core n, $c_n(t)$, is the interfered PAM4 signal. The optical signals $c_m(t)$ and $c_n(t)$ are transmitted in two perpendicular polarization directions, x and y, which are represented by the power splitting of the transmitted PAM4 signal by both polarization directions at the input of the MCF.

The PAM4 signal travels along core n of the MCF, where linear propagation is assumed, since non-linear effects are usually insignificant in inter-datacenters distances [3], [24]. The effect of ICXT on the interfered cores is modelled by the DP-DCM simply by using transfer functions that change randomly along time, being this dependence introduced by

applying random phase shifts (RPSs) along the longitudinal direction of the fiber. The transfer functions $F_{a,b}(\omega)$ model the frequency response of the ICXT from the polarization a, with $a = x$ or y, at the input of core m to the polarization b, with $b = x$ or y, at the output of core n and are represented as [23]

$$F_{a,b}(\omega) = -\frac{j}{\sqrt{2}}\bar{K}_{nm}\exp(-j\bar{\beta}_n(\omega)L)\cdot\exp(-\frac{\alpha_m}{2}L) \sum_{k=1}^{N_p}\exp[-j\Delta\beta_{mn}z_k]\exp[-j\phi_{nm,k}^{(a,b)}] \quad (1)$$

where α_m is the attenuation coefficient of the core m. In the model proposed by [7], [25], $\bar{K}_{nm}$ is the average inter-core coupling coefficient [23], N_p is the number of phase-matching points (PMPs), $\Delta\beta_{mn}$ is given by $\bar{\beta}_m(\omega) - \bar{\beta}_n(\omega)$ and $\phi_{nm,k}^{(a,b)}$ represent the RPSs associated with the k-th PMP, which are modelled using an uniform distribution between $[0, 2\pi[$ and z_k is the longitudinal coordinate of the k-th PMP randomly distributed between two consecutive PMPs and is given by

$$z_k = \frac{L}{N_p}(r_k + k - 1) \quad (2)$$

where $r_k (1 \leq k \leq N_p)$ are independent random variables generated with an uniform distribution between [0, 1[. In this work, we also assume $\alpha_m \approx \alpha_n$.

The DP-DCM models the evolution of the ICXT impact on the system performance in different MCF realizations by generating randomly different N_p sets of RPSs in each MCF realization. Therefore, in each iteration of the Monte Carlo (MC) simulator, a new PAM4 signal with symbols randomly generated is transmitted in core m and one MCF realization corresponding to the transfer functions given by (1) is generated. The transfer functions $F_{x,x}(\omega)$ and $F_{y,x}(\omega)$ model the ICXT generated from polarization x and y of the core m that is going to interfere with the electrical field in the polarization x of the core n. The transfer functions $F_{x,y}(\omega)$ and $F_{y,y}(\omega)$ model the ICXT generated from both polarizations of core m that interfere with the field of core n in the polarization y. The ICXT level quantifies the ammount of ICXT power that will affect the interfered core and is defined by the ratio between the mean ICXT power and the mean power of the signal both at the output of the interfered core n and is calculated as in $X_c = N_p|K_{nm}|^2$ [23].

The temporal dependence of the ICXT is induced by effects such as the walk-off due to different group velocities between cores [24]. In this model, the skew between the interfering core m and the interfered core n is given by $S_{mn} = d_{mn}L$, where d_{mn} is the walkoff between cores m and n defined by $d_{mn} = \beta_{1,m} - \beta_{1,n}$.

In Fig. 2, at the output of the MCF, a chromatic dispersion compensation (CDC) module compensates the chromatic dispersion (CD) arising from the transmission through core n, which can be a major source of performance degradation due to the required data rates. The CDC is modelled considering a dispersion compensating fiber (DCF) in linear propagation with a length that fully compensates the distortion due to CD.

After the CDC module, an erbium-doped fiber amplifier (EDFA) compensates the inter-datacenter link losses and an optical filter reduces the amplified spontaneous emission (ASE) noise power generated by the EDFA. The amplifier gain is set to compensate all losses from the single mode-multicore fiber (SM-MCF) and DCF and the ASE noise is modelled as additive white Gaussian noise [7]. The optical filter, modelled by a 4th order super Gaussian filter, reduces ASE noise power and introduces negligible inter-symbol interference (ISI) [26].

After CDC and optical amplification, the PAM4 signal degraded by ICXT and ASE noise passes the DD receiver dedicated to core n, where is converted to an electrical signal, $c_{PIN}(t)$, by the PIN photo-detector with a responsivity of $R_\lambda = \frac{\eta q}{h\nu_0}$. Electrical noise is added after photodetection and an electrical filter, modelled as a 3^{rd} order Bessel filter, is used to reduce the noise power. In the decision circuit, the BER of each MCF realization, the average BER is assessed. The BER of each MCF realization is calculated by the semi-analytical method known as the exhaustive Gaussian approach. For a PAM4 signal, the BER is given by [27]

$$BER = \frac{1}{2\cdot 4^{N_{reg}}}\left\{\sum_{\substack{k=1\\a_k=0}}^{4^{N_{reg}}} Q\left(\frac{F_1 - i_{0,k}}{\sigma_{0,k}}\right) + \sum_{\substack{k=1\\a_k=1}}^{4^{N_{reg}}}\left[Q\left(\frac{i_{1,k} - F_1}{\sigma_{1,k}}\right) + Q\left(\frac{F_2 - i_{1,k}}{\sigma_{1,k}}\right)\right] + \sum_{\substack{k=1\\a_k=2}}^{4^{N_{reg}}}\left[Q\left(\frac{i_{2,k} - F_2}{\sigma_{2,k}}\right) + Q\left(\frac{F_3 - i_{2,k}}{\sigma_{2,k}}\right)\right] + \sum_{\substack{k=1\\a_k=3}}^{4^{N_{reg}}} Q\left(\frac{i_{3,k} - F_3}{\sigma_{3,k}}\right)\right\} \quad (3)$$

where $4^{N_{reg}}$ is the maximum length of PAM4 symbols optical sequence generated using deBruijn sequences, with N_{reg} as the length of the offset register used to generate the sequence; $i_{0,k}$, $i_{1,k}$, $i_{2,k}$ and $i_{3,k}$ correspond to the means of the currents at the input of the decision circuit for the symbols a_k at the time sampling instants $t_k = t_0 + T_s(k-1)$, where t_0 is extracted from the received eye-pattern at the decision circuit input, $k \in 1, ..., 4^{N_{reg}}$, $\sigma_{0,k}$, $\sigma_{1,k}$, $\sigma_{2,k}$ and $\sigma_{3,k}$ are the noise standard deviations for the different time sampling instants [27] and the function $Q(x)$ is given by [7]

$$Q(x) = \int_x^\infty \frac{1}{\sqrt{2}} e^{-\frac{\xi^2}{2}} d\xi \quad (4)$$

In the simulation, the decision thresholds F_1, F_2 and F_3 are optimized in each time-fraction by applying the bisection method to minimize the BER [7], [25]. Effects such as ICXT, ISI, electrical noise, signal-ASE, and ASE-ASE beat noises are taken into account using this semi-analytical method.

After several MCF realizations, the average BER is obtained by averaging the BERs obtained in each MCF realization.

In this work, we consider that the inter-datacenter link, impaired by ICXT, is in outage when the BER is above a pre-defined limit of 3.8×10^{-3}, which is the threshold typically used for datacenters connections with forward-error correction [7], [28]. The electrical and optical receiver filters bandwidth were optimized in B2B operation to maximize the receiver sensitivity [7]. For the MCF length of 80 km and r=0 and r=0.1, the signal power at the transmitter output has also been optimized to achieve the BER of 3.8×10^{-5} without ICXT. The number of PMPs is set to characterize accurately the RPS mechanism [7], [28]. Two different inter-core skews with skew-symbol rate product of $|S_{mn}R_s| = 1000$ and $|S_{mn}R_s| = 0.01$ are also chosen to perform these studies. The case of $|S_{mn}R_s| = 1000$ is referred as high skew symbol rate product as $|S_{mn}R_s| >> 1$ [7]. The situation of $|S_{mn}R_s| = 0.01$ is referred as low skew-symbol rate product, since $|S_{mn}R_s| << 1$ [7].

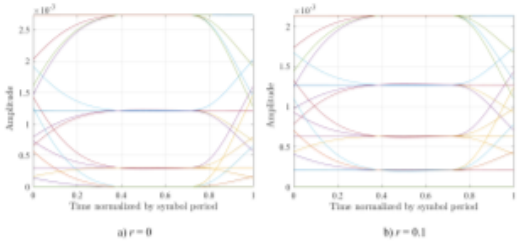


Fig. 3. Eye-patterns of the PAM4 signal at the MCF input for a) r=0 and b) r=0.1.

The eye-patterns in Fig. 3 represent an example of the PAM4 transmitted signals $c_m(t)$ and $c_n(t)$ at the output of the optical transmitter for $r = 0$ and $r = 0.1$. In contrast, Fig. 4 and Fig. 5 show the impact of the ICXT on the eye-patterns at the decision circuit input of the optically amplified PAM4 IM-DD system with full loss and chromatic dispersion compensation. Fig. 4 shows the received eye-patterns for X_c = -14 dB and $|S_{mn}R_s| = 1000$ of a) best BER and b) worst BER with r=0 and c) best BER and d) worst BER with r=0.1 after 1000 MCF realizations, which is high enough to obtain a stabilized average BER [7]. Fig. 5 shows the received eye-patterns for X_c = -14 dB and $|S_{mn}R_s| = 0.01$ of a) best BER and b) worst BER with r=0 and c) best BER and d) worst BER with r=0.1 after 1000 MCF realizations.

In Fig. 4 b), for the worst BER with r=0, the lowest eye is fully closed due to the strong ICXT. However, as seen in Fig. 4 d), for the worst BER with r=0.1, the ICXT degrades less the received eye-pattern (the lowest eye is not fully closed), which leads to a lower BER. The product $|S_{mn}R_s|$ has been shown to be relevant on the impact of ICXT on the performance of optical links with DD [8], [29], [30]. The same conclusions can be drawn through the analysis of Fig. 5 b) and d) where the eye-patterns are fully closed for both extinction ratios, in comparison with Fig. 4 b) and d), which shows that, for optical links with low $|S_{mn}R_s|$, the ICXT is more detrimental than for high $|S_{mn}R_s|$. However, for the best BERs obtained with low $|S_{mn}R_s|$, Fig. 5 a) and c) show that the amplitude levels are more defined than in the eye-patterns shown in Fig. 4 a) and c) with high $|S_{mn}R_s|$, where more symbols in the interfering core are contributing to ICXT [7].

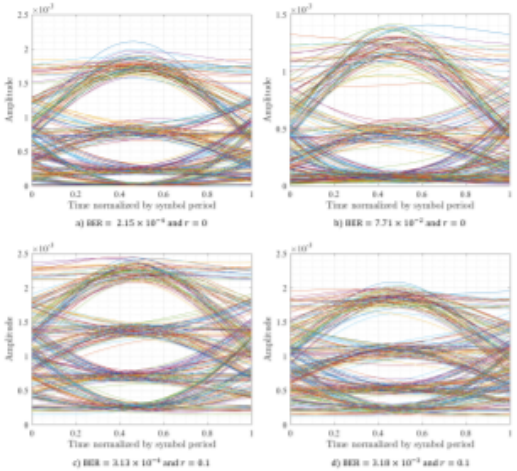


Fig. 4. Eye-patterns at the decision circuit input for X_c = -14 dB, $|S_{mn}R_s| = 1000$ for a) best BER and b) worst BER with r=0 and a) best BER and b) worst BER with r=0.1.

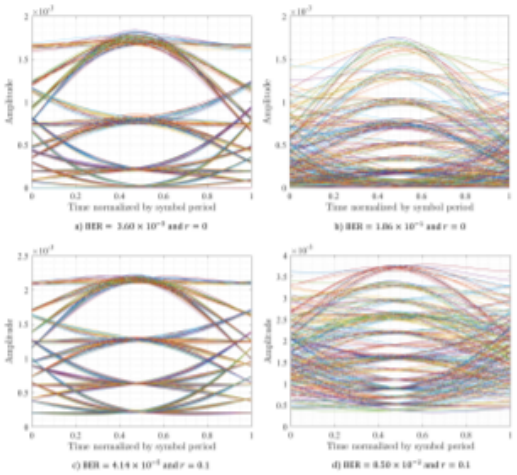


Fig. 5. Eye-patterns at the decision circuit input for X_c = -14 dB, $|S_{mn}R_s| = 0.01$ for a) best BER and b) worst BER with r=0 and a) best BER and b) worst BER with r=0.1.

Fig. 6 shows the histograms of the BERs and the corresponding average BER obtained in 1000 MCF realizations with X_c = -14 dB, *r*=0.1 and *r*=0, for two different inter-core skews, a) $|S_{mn}R_s| = 1000$ and b) $|S_{mn}R_s| = 0.01$. Several MCF realizations experience system outage, especially for $|S_{mn}R_s| = 0.01$, since the corresponding BERs surpass the BER limit. Only for $|S_{mn}R_s| = 1000$ and *r*=0.1, the BER limit is never exceeded and there is no system outage. The effect of ICXT on the BER distribution is less detrimental with $r = 0.1$. This influence of the extinction ratio on the ICXT impact has been already observed in OOK systems [27]. For $r = 0$, a higher spreading of the BER values is observed, and, hence, more MCF realizations have lower BERs and more lead to a BER that surpasses the BER limit. The results in both histograms show also that the product $|S_{mn}R_s|$ has a significant influence on the BERs distribution, since for $|S_{mn}R_s| = 1000$, the BER range is significant lower in comparison with the BER range obtained for $|S_{mn}R_s| = 0.01$. Therefore, system outage is more likely to occur for $|S_{mn}R_s| = 0.01$ and *r*=0 and less likely to occur for $|S_{mn}R_s| = 1000$ and $r = 0.1$.

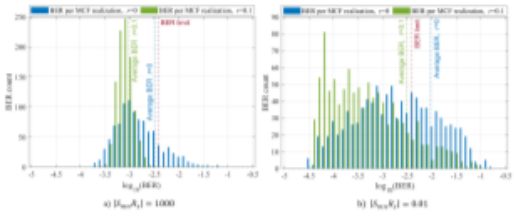


Fig. 6. Histogram of the $\log_{10}(BER)$ for 1000 MCF realizations with X_c = -14 dB, *r*=0.1 and *r*=0, for a) $|S_{mn}R_s| = 1000$ and b) $|S_{mn}R_s| = 0.01$.

B. Data collection and eye-pattern pre-processing

The data used to train and test the CNN corresponds to the data collected from the optical telecommunication system equivalent model developed in [7] and described in section II-A that models with a good accuracy the ICXT effect in PAM4 inter-datacenter optical links supported by homogeneous WC-MCFs. Each pair of collected data consists of a received eye-pattern, E_i (with *i*=1,...,N_{EP}), where N_{EP} is the number of eye-patterns collected, such as the ones shown in Figs. 4 and 5, and the corresponding BER calculated logarithmically, i.e., $\log_{10}$(BER), at the decision circuit. Each eye-pattern is a 32×256 matrix, with N_s=32 amplitude samples per symbol and 256 is the number of generated PAM4 symbols in each MCF realization.

A good performance of the CNN is highly dependent on the training data. So, it is important to choose a high number of unbiased data to achieve a consistent model by decreasing the disparity between the training data and testing data. Therefore, a guided sampling is performed to collect a balanced number of BERs and corresponding eye-patterns. For that purpose, for the different optical links, the data generated by the MATLAB optical telecommunication system equivalent model is collected following a two-step balancing scheme within a $[\log_{10}(\text{BER})_{min}, \log_{10}(\text{BER})_{max}]$ range, where 1000 pairs of data (eye-patterns) are saved in every 0.1 interval of the $\log_{10}$(BER). Within these N_{BER} intervals, where $N_{BER} = \frac{\log_{10}(\text{BER}_{max}) - \log_{10}(\text{BER}_{min})}{0.1}$, a second balancing step is performed with 20 sub intervals, where 50 pairs of data are saved in every 0.005 interval of the $\log_{10}$(BER). Notice that the BER range is not the same for all optical links studied, since the BER values distribution is much dependent on the link parameters as seen in section II-A.

As the ICXT has a stochastic behaviour with a random time evolution and frequency dependence, it leads to random fluctuations of the BER and different BER distributions, depending on the parameters of the optical link under study. Some BERs are less frequent than others, particularly for higher BERs associated with system outages. Especially for low outage probabilities, these BERs become extremely rare to occur. Therefore, to collect a good amount of balanced data inside each logarithmic BER interval and to maintain the time of simulation at acceptable levels, particularly for BERs less likely to occur, a slight change of the DP-DCM model has been implemented following a domain randomization-based approach.

As mentioned in section II-A, in each iteration of the MC simulator, one MCF realization corresponding to the transfer functions given by eq. (1) is generated with random different N_p RPSs. This randomness models the unpredictability and affects the ICXT on the received PAM4 signal and the corresponding BER. Firstly, the MC simulator starts with a random set of N_p RPSs using the uniform distribution between $[0, 2\pi[$ and is stored and used in the first MCF realization. After that, based on a single uniformly distributed random number in the interval $[0, 1]$, it is determined how the set of N_p RPSs for the next MCF realization is generated. If the random number is above or equal to 0.5, a new set of N_p RPSs is generated following the previously described procedure and, again, it is stored. If not, a new set is obtained by adding normally distributed random noise with zero mean and unitary variance to the set of N_p RPSs previously stored that is known to lead to a BER inside a BER interval. The set of RPSs generated with this random perturbation is not stored, and, therefore, it is never used again in the following MCF realizations. At the end of each MC simulator iteration, the eye-pattern and the corresponding $\log_{10}$(BER) are saved if the corresponding N_{BER} interval is not full. As the MC simulator iterations continue, following the previously described procedures, the number of N_p RPSs sets stored increases and the random perturbation approach is performed by randomly selecting one of the stored sets of N_p RPSs.

Before training and testing the CNN, the eye-patterns are pre-processed to obtain what we denote as GEPs. Typically, oscilloscopes sample the received signal and generate a two-dimensional database, that statistically represents the time, where the unit interval or bit period of the eye-pattern is defined by the data clock, and the amplitude of the digital signal [31]. As the number of samples increases, a third dimension

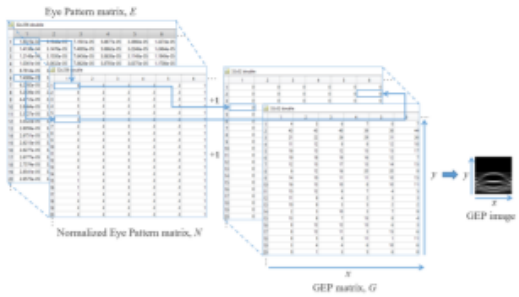


Fig. 7. Example of the process to transform eye-patterns into 32×32 GEP images.

is considered, denoted as plot density, which represents the number of pixels that are located in the same position on the oscilloscope display [31]. In this work, a similar approach is used to obtain GEP images.

First, for a $x \times y$ GEP image, the amplitudes of a synthetic eye-pattern obtained from MATLAB are normalized into $[1, y] \in \mathbb{N}$ amplitudes to obtain the vertical position of the GEP image pixel assigned to the corresponding amplitude. For the case of a fixed scale GEP, this normalization takes into account the maximum amplitude that occurred in all eye-patterns obtained for the corresponding type of optical link under study, i.e., the eye-patterns used during a CNN training. For the case of a dynamic scale GEP, each eye-pattern is normalized taking into account its maximum amplitude. After that, the eye-pattern to GEP conversion process continues with a $y \times x$ GEP matrix with zero elements. The plot density is modeled by incrementing one unit in all elements of the GEP matrix with the rows given by the eye-pattern normalized amplitudes and the corresponding columns given by the rows of the original eye-pattern normalized into $[1, x]$.

Fig. 7 shows the process to transform a synthetic 32×256 eye-pattern matrix obtained from MATLAB into a 32×32 fixed scale GEP image. For instance, the amplitude in the $(1,1)^{th}$ element of the eye-pattern matrix, *E*, is normalized into [1, 32] $\in \mathbb{N}$ amplitudes, which gives 4. This means that this amplitude falls in $(4,1)^{th}$ element of the GEP matrix, *G*. The amplitude is considered in the GEP matrix by incrementing one unit in that given element. If more amplitudes of the first row of the eye-pattern have the normalized amplitude 4, the $(4,1)^{th}$ element of the GEP matrix is incremented accordingly to these number of amplitudes. The case of $E(6,1)$ amplitude follows the same logic, where the normalized amplitude is 2 and one unit is incremented in $G(2,6)$.

Notice that for a different GEP size, since each eye-pattern is a 32×256 matrix, an additional normalization needs to be performed during the GEP transformation process. For instance, to obtain a 64×64 GEP, the amplitudes of the eye-pattern are normalized into [1, 64] $\in \mathbb{N}$ amplitudes to obtain the vertical position of the GEP image pixel assigned to the corresponding amplitude. Then, the transformation process follows the previously described procedure, however, the plot density is modeled by incrementing one unit in all elements with the rows given by the eye-pattern normalized amplitudes and the corresponding columns given by the rows of the original eye-pattern normalized into [1, 64].

After this transformation process, the GEP images, obtained from the GEP matrices, are used as input data in the CNN described in the next subsection.

C. Convolutional neural network

In this work, the CNN model is developed on the MATLAB platform using the Deep Learning Toolbox and Deep Network Designer. The proposed CNN architecture is based on [21], [22] and is schematically represented in Fig. 8. First, normalization is applied by re-scaling the data in the range $[-1, 1]$ every time a GEP image is forward propagated through the input layer of the CNN. This architecture has five convolutional (Conv) layers, C1 to C5, that pass the GEP images through a set of convolutional kernels with stride (1,1). The layers C1 and C2, where both kernel sizes are 5×5, produce, respectively, 32 and 64 feature maps and the layers C3, C4 and C5, where the corresponding kernels sizes are 3×3, produce, respectively, 128, 256 and 512 feature maps. All convolution layers add the required padding to the input, either a GEP image or feature map, to ensure that its border pixels are completely exposed to the filter and the resulting feature map has the same size as the input. The outputs of each convolutional layers are normalized using a batch normalization (BN) layer followed by a Rectified Linear Unit (ReLU) layer for a more effective and faster training [14]. After the first four Conv+BN+ReLU operations, a down-sampling is performed by an average pooling layer. In our CNN, we set the four pooling layers, P1, P2, P3 and P4, with 2×2 subsampling regions and stride (2,2). After feature extraction (C5+BN+ReLU), a dropout layer is placed to prevent overfitting and a fully connected layer followed by a regression layer are used to predict the BER.

Fig. 8. CNN architecture considered in this work to learn the BER from the GEP images.

The performance evaluation metric chosen to evaluate the CNN-based eye-pattern image analysis and BER prediction is the root mean square error (RMSE), which has been widely used in the literature as a key CNN regression performance indicator [32]–[34], and is given by

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} e_i^2} \tag{5}$$

where n is the number of eye-patterns considered either in the training, test or validation sets and e_i is the difference between the expected values and the predicted values of the BER. In this work, we consider a RMSE below 0.1 as an acceptable prediction of the $\log_{10}$ (BER).

III. CNN PERFORMANCE ASSESSMENT

In this section, the performance of the CNN-based eye-pattern analysis and BER prediction in PAM4 inter-datacenter optically amplified short IM-DD connections impaired by ICXT is assessed. The simulation parameters are presented in Table I.

TABLE I
SIMULATION PARAMETERS

Parameter	Value
Skew-symbol rate product	$\lvert S_{mn}R_s\rvert = 1000$, $\lvert S_{mn}R_s\rvert = 0.01$
ICXT level	X_c = -16 dB, X_c = -14 dB, X_c = -12 dB
Extinction ratio	$r = 0$, $r = 0.1$

A. CNN performance for specific optical link parameters

Firstly, the accuracy of the BER prediction is assessed by training the considered CNN from the eye-pattern analysis and testing the corresponding regression model for each type of optical link. Fig. 9 shows the summary of the RMSEs obtained for the CNNs trained with fixed scale GEPs. Each CNN model is able to predict the $\log_{10}$ (BER) without surpassing the RMSE limit of 0.1, except for the cases of X_c = -16 dB and X_c = -12 dB with $|S_{mn}R_s| = 1000$ and r=0. For all CNNs trained, this case, r=0 and $|S_{mn}R_s| = 1000$, is the one that it is more difficult to train, leading to a worst BER prediction. The best predictions (with lower RMSE) are obtained for the case of $|S_{mn}R_s| = 0.01$ and r=0.1, being the RMSE lower than 0.05, for all the crosstalk levels. In fact, with r=0.1, the RMSEs obtained with the different crosstalk levels are more similar, than in the case with r=0.

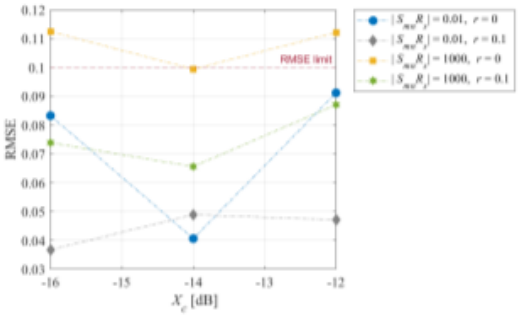


Fig. 9. RMSE of the CNN models trained and tested with fixed scale GEPs as a function of X_c, for $|S_{mn}R_s| = 0.01$ and r=0; $|S_{mn}R_s| = 0.01$ and r=0.1; $|S_{mn}R_s| = 1000$ and r=0 and $|S_{mn}R_s| = 1000$ and r=0.1.

The RMSEs obtained for the dynamic scale GEPs as a function of the crosstalk level are shown in Fig. 10 and show a similar behavior to the one obtained for the fixed scale GEPs in Fig. 9. Each CNN model is able to predict the $\log_{10}$ (BER) without surpassing the RMSE limit of 0.1, except for the cases of X_c = -16 dB and -12 dB with $|S_{mn}R_s| = 1000$ and r=0.

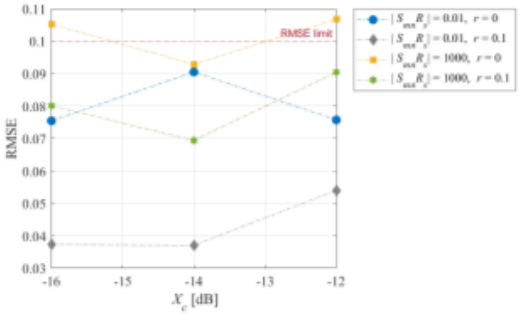


Fig. 10. RMSE of the CNN models trained and tested with dynamic scale GEPs as a function of X_c, for $|S_{mn}R_s| = 0.01$ and r=0; $|S_{mn}R_s| = 0.01$ and r=0.1; $|S_{mn}R_s| = 1000$ and r=0 and $|S_{mn}R_s| = 1000$ and r=0.1 for dynamic scale GEPs.

B. CNN generalization

In this subsection, the performance of the CNN generalization is assessed. Fig. 11, shows the results obtained for a CNN trained fixed scale GEPs for the case of optical links with X_c = -14 dB, $|S_{mn}R_s| = 0.01$ and r=0 and tested with fixed scale GEPs from optical links with X_c = -16 dB, $|S_{mn}R_s| = 0.01$ and r=0. The GEPs were normalized with the eye-pattern maximum amplitude obtained in all eye-patterns generated for both crosstalk levels of -16 dB and -14 dB. As the GEPs' true BER values increase, the CNN predictions are more inconsistent, since a wider range of predicted BERs is obtained. The linear behaviour expected of the regression model is not observed. Fig. 11 presents a much higher RMSE than the RMSE limit of 0.1. These results show that the CNN is unable to predict correctly the BER of test GEPs obtained with other crosstalk level, meaning that the CNN is not capable of generalization.

Fig. 12 shows a CNN trained and tested with eye-patterns obtained with X_c = -16 dB, -14 dB and -12 dB , $|S_{mn}R_s| = 0.01$, r=0 and $\log_{10}$ (BER) $\in [-3.1, -1.9]$, which is able to predict more correctly the BER from GEPs with different crosstalk levels, leading to a RMSE of 0.082. This indicates that the CNN needs to train with different optical parameters to perform a better generalization for predicting the BERs.

Regarding dynamic scale GEPs, the results presented in Fig. 13, show that the CNN regression model obtained for the case of optical links with X_c = -14 dB, $|S_{mn}R_s| = 0.01$ and r=0 is also unable to predict correctly the BER of dynamic scale GEPs obtained from optical links with X_c = -16 dB, $|S_{mn}R_s| = 0.01$ and r=0, where a RMSE = 0.285 is obtained.

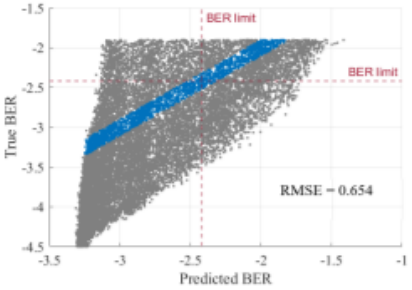


Fig. 11. BER predictions of optical links with X_c = -16 dB, $|S_{mn}R_s| = 0.01$, r=0 and GEPs normalized with the eye-pattern maximum amplitude obtained with X_c = -16 dB and X_c = -14 dB from a CNN trained for optical links with X_c = -14 dB, $|S_{mn}R_s| = 0.01$ and r=0. The blue data points represent the predictions with a margin error below 0.1.

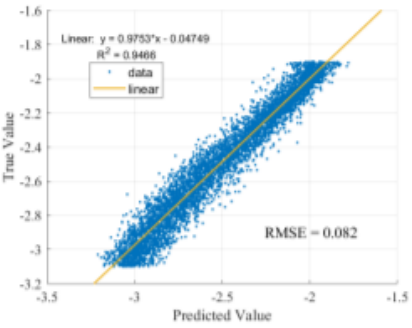


Fig. 12. BER predictions of the test set from the CNN trained for the case of optical links with X_c = -12 dB, X_c = -14 dB, X_c = −16 dB, $|S_{mn}R_s| = 0.01$, r=0.

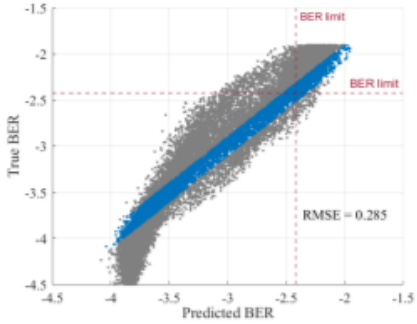


Fig. 13. BER predictions of GEPs from optical links with X_c = -16 dB, $|S_{mn}R_s| = 0.01$ and r=0 from a CNN regression model obtained for an optical link with X_c = -14 dB, $|S_{mn}R_s| = 0.01$ and r=0.

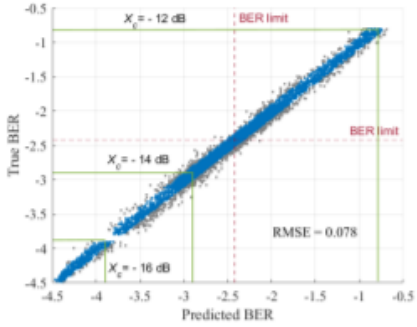


Fig. 14. BER predictions of the test set from a CNN trained for optical links with X_c = -16 dB and $\log_{10}$ (BER) $\in [-4.5, -3.9]$, X_c = -14 dB and $\log_{10}$ (BER) $\in [-3.8, -2.9[$, X_c = -12 dB and $\log_{10}$ (BER) $\in [-2.9, -0.8]$, $|S_{mn}R_s| = 0.01$ and r=0.

However, it seems more able to generalize, predicting BERs closer to a linear behaviour comparing to the case presented in Fig. 11, for the fixed scale.

Fig. 14 shows the BER prediction distribution of the test set from a CNN trained with dynamic scale GEPs of disjoint BER intervals, i. e., for the case of optical links with X_c = -16 dB and $\log_{10}$ (BER) $\in [-4.5, -3.9]$, X_c = -14 dB and $\log_{10}$ (BER) $\in [-3.8, -2.9[$, X_c = -12 dB and $\log_{10}$ (BER) $\in [-2.9, -0.8]$, for $|S_{mn}R_s| = 0.01$ and r=0. This CNN is tested with the full $\log_{10}$ (BER) range considering eye-patterns obtained for the three crosstalk levels and the results are shown in Fig. 15, which shows that, in this case, the CNN is able to predict the BER for the three crosstalk levels considered without surpassing the RMSE limit of 0.1.

IV. Conclusions

The performance of the CNN-based eye-pattern analysis and BER prediction in PAM4 inter-datacenter optically amplified IM-DD connections impaired by ICXT has been studied by assessing the RMSE of the obtained BER predictions. Two different approaches regarding the GEPs have been assessed: fixed scale GEPs and dynamic scale GEPs. With the fixed scale GEPs, training a CNN and test the corresponding regression model for each of the optical link parameters considered, each CNN model was able to predict the $\log_{10}$ (BER) without surpassing the RMSE limit of 0.1, except for the cases of X_c = -16 dB and X_c = -12 dB with $|S_{mn}R_s| = 1000$ and r=0, being the RMSE only around 0.01 higher than the limit established.

By training the CNN with a certain crosstalk level, $|S_{mn}R_s|$ and r and testing with data extracted for different optical link parameters, we have shown that the CNN is unable to predict correctly the $\log_{10}$ (BER). These situations can be improved by training the CNN with eye-patterns collected for the different crosstalk levels or with dynamic scale GEPs. From an implementation point of view, a CNN based on dynamic GEPs seems a better solution, since it does not require a prior knowledge of the signal amplitudes obtained for other eye-patterns. Each GEP is generated based only on the

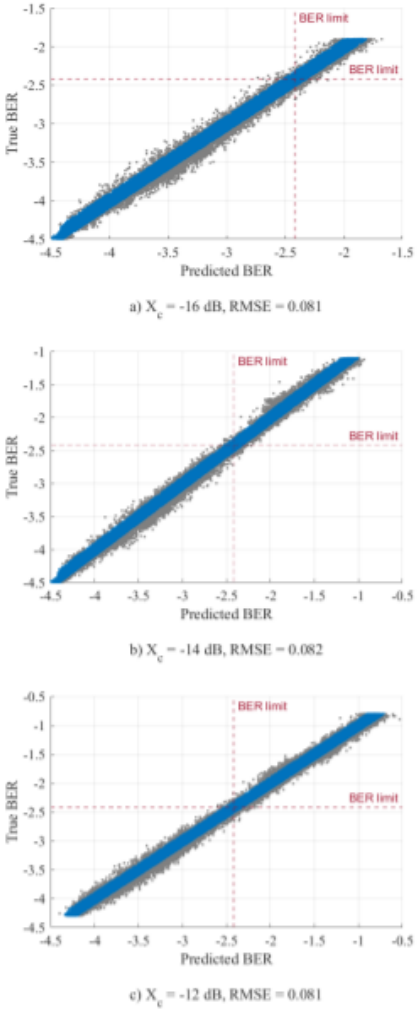


Fig. 15. BER predictions of GEPs from optical links with $|S_{mn}R_s| = 0.01$ and r=0, for a) X_c = -16 dB, b)X_c = -14 dB and c) X_c = -12 dB, trained for optical links with X_c = -16 dB and $\log_{10}$ (BER) $\in [-4.5, -3.9]$, X_c = -14 dB and $\log_{10}$ (BER) $\in [-3.8, -2.9[$, X_c = -12 dB and $\log_{10}$ (BER) $\in [-2.9, -0.8]$.

corresponding eye-pattern maximum amplitude. Considering dynamic scale GEPs, when the CNN was trained with data with a certain crosstalk level, $|S_{mn}R_s|$ and extinction ratio and tested with a different level of crosstalk, a good performance was obtained for r=0 while, for r=0.1, the RMSE surpassed the limit.

By training the CNN with different $\log_{10}$ (BER) ranges for the different crosstalk levels, same $|S_{mn}R_s|$ and r, the CNN was able to predict the BERs for the three crosstalk levels along the total BERs range. As a final remark, the results obtained in this work confirm that CNN-based models can extract features from eye-patterns to predict the BER without prior knowledge of the transmitted signals, since only the eye-pattern image is used to extract information and predict the BER.

Acknowledgments

This work was supported under the project of Instituto de Telecomunicações UIDB/50008/2020.

References

[1] M. Noormohammadpour and C. S. Raghavendra, "Datacenter traffic control: Understanding techniques and tradeoffs," *IEEE Communications Surveys Tutorials*, vol. 20, no. 2, pp. 1492–1525, 2018.

[2] Cisco. (2020) Cisco annual internet report - cisco annual internet report (2018–2023) white paper. [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html

[3] J. Krause Perin, A. Shastri, and J. M. Kahn, "Data center links beyond 100 Gbit/s per wavelength," *Optical Fiber Technology*, vol. 44, pp. 69 – 85, 2018, special Issue on Data Center Communications. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1068520017302183

[4] K. Saitoh and S. Matsuo, "Multicore fiber technology," *Journal Lightwave Technology*, vol. 34, no. 1, pp. 55–66, Jan 2016. [Online]. Available: http://jlt.osa.org/abstract.cfm?URI=jlt-34-1-55

[5] J. Zhang, K. Wang, Y. Wei, L. Zhao, W. Zhou, J. Xiao, B. Liu, X. Xin, F. Zhao, Z. Dong, and J. Yu, "280 Gb/s IM/DD PS-PAM-8 transmission over 10 km SSMf at O-Band for optical interconnects," in *2020 Optical Fiber Communications Conference and Exhibition (OFC)*, 2020, pp. 1–3.

[6] H. Yuan, M. Furdek, A. Muhammad, A. Saljoghei, L. Wosinska, and G. Zervas, "Space-division multiplexing in data center networks: On multi-core fiber solutions and crosstalk-suppressed resource allocation," *Journal of Optical Communications and Networking*, vol. 10, no. 4, pp. 272–288, Apr 2018. [Online]. Available: http://jocn.osa.org/abstract.cfm?URI=jocn-10-4-272

[7] R. Dias, J. Rebola, and A. Cartaxo, "Outage probability due to crosstalk from multiple interfering cores in PAM4 inter-datacenter connections," *Photonics*, vol. 8, no. 1, 2021. [Online]. Available: https://www.mdpi.com/2304-6732/8/1/9

[8] J. L. Rebola, A. V. T. Cartaxo, T. M. F. Alves, and A. S. Marques, "Outage probability due to intercore crosstalk in dual-core fiber links with direct-detection," *IEEE Photonics Technology Letters*, vol. 31, no. 14, pp. 1195–1198, 2019.

[9] F. Musumeci, C. Rottondi, A. Nag, I. Macaluso, D. Zibar, M. Ruffini, and M. Tornatore, "An overview on application of machine learning techniques in optical networks," *IEEE Communications Surveys Tutorials*, vol. 21, no. 2, pp. 1383–1408, 2019.

[10] P. Kim, *MATLAB Deep Learning: With Machine Learning, Neural Networks and Artificial Intelligence*, 1st ed. USA: Apress, 2017.

[11] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, http://www.deeplearningbook.org.

[12] A. Sifaoui, A. Abdelkrim, and M. Benrejeb, "On the use of neural network as a universal approximator," *International Journal of Sciences*, vol. 2, pp. 386–399, 08 2008.

[13] D. Wang and M. Zhang, "Artificial intelligence in optical communications: From machine learning to deep learning," *Frontiers in Communications and Networks*, vol. 2, p. 9, 2021. [Online]. Available: https://www.frontiersin.org/article/10.3389/frcmn.2021.656786

[14] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai, and T. Chen, "Recent advances in convolutional neural networks," *Pattern Recognition.*, vol. 77, no. C, p. 354–377, May 2018. [Online]. Available: https://doi.org/10.1016/j.patcog.2017.10.013

[15] A. Mikołajczyk and M. Grochowski, "Data augmentation for improving deep learning in image classification problem," in *2018 International Interdisciplinary PhD Workshop (IIPhDW)*, 2018, pp. 117–122.

[16] J. Tobin, R. Fong, A. Ray, J. Schneider, W. Zaremba, and P. Abbeel, "Domain randomization for transferring deep neural networks from simulation to the real world," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2017, pp. 23–30.

[17] M. Ani, H. Basevi, and A. Leonardis, "Quantifying the use of domain randomization," in *2020 25th International Conference on Pattern Recognition (ICPR)*, 2021, pp. 6128–6135.

[18] X. Yue, Y. Zhang, S. Zhao, A. Sangiovanni-Vincentelli, K. Keutzer, and B. Gong, "Domain randomization and pyramid consistency: Simulation-to-real generalization without accessing target domain data," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019, pp. 2100–2110.

[19] J. Mata, I. de Miguel, R. J. Durán, N. Merayo, S. K. Singh, A. Jukan, and M. Chamania, "Artificial intelligence (AI) methods in optical networks: A comprehensive survey," *Optical Switching and Networking*, vol. 28, pp. 43 – 57, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S157342771730231X

[20] W. S. Saif, M. A. Esmail, A. M. Ragheb, T. A. Alshawi, and S. A. Alshebeili, "Machine learning techniques for optical performance monitoring and modulation format identification: A survey," *IEEE Communications Surveys Tutorials*, vol. 22, no. 4, pp. 2839–2882, 2020.

[21] D. Wang, M. Zhang, Z. Li, J. Li, M. Fu, Y. Cui, and X. Chen, "Modulation format recognition and OSNR estimation using CNN-based deep learning," *IEEE Photonics Technology Letters*, vol. 29, no. 19, pp. 1667–1670, 2017.

[22] D. Wang, Y. Xu, J. Li, M. Zhang, J. Li, J. Qin, C. Ju, Z. Zhang, and X. Chen, "Comprehensive eye diagram analysis: A transfer learning approach," *IEEE Photonics Journal*, vol. 11, no. 6, pp. 1–19, 2019.

[23] 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, 2017.

[24] A. Macho, C. García-Meca, F. J. Fraile-Peláez, M. Morant, and R. Llorente, "Birefringence effects in multi-core fiber: coupled local-mode theory," *Optics Express*, vol. 24, no. 19, pp. 21 415–21 434, Sep 2016. [Online]. Available: http://www.opticsexpress.org/abstract.cfm?URI=oe-24-19-21415

[25] R. Dias, J. Rebola, and A. Cartaxo, "Transmission of PAM4 signals in amplified inter-datacenters connections with direct-detection and multicore fibers limited by inter-core crosstalk," Master's thesis, Iscte–Instituto Universitário de Lisboa, 2020.

[26] G. Agrawal, *Fiber-Optic Communication Systems*, 4th ed. Hoboken, NJ, USA: John Wiley & Sons, 2010.

[27] J. L. Rebola and A. V. T. Cartaxo, "On the quaternary level spacing signalling optimisation for increasing the transmission distance in optical communication systems," 2000.

[28] Z. Xing, A. Samani, M. Xiang, E. El-Fiky, T. M. Hoang, D. Patel, R. Li, M. Qiu, M. G. Saber, M. Morsy-Osman, and D. V. Plant, "100 Gb/s PAM4 transmission system for datacenter interconnects using a SiP ME-MZM based DAC-less transmitter and a VSB self-coherent receiver," *Optics Express*, vol. 26, no. 18, pp. 23 969–23 979, Sep 2018. [Online]. Available: http://www.osapublishing.org/oe/abstract.cfm?URI=oe-26-18-23969

[29] R. S. Luís, B. J. Puttnam, A. V. T. Cartaxo, W. Klaus, J. M. D. Mendinueta, Y. Awaji, N. Wada, T. Nakanishi, T. Hayashi, and T. Sasaki, "Time and modulation frequency dependence of crosstalk in homogeneous multi-core fibers," *Journal of Lightwave Technology*, vol. 34, no. 2, pp. 441–447, 2016.

[30] G. Rademacher, R. S. Luís, B. J. Puttnam, Y. Awaji, and N. Wada, "Crosstalk dynamics in multi-core fibers," *Optics Express*, vol. 25, no. 10, pp. 12 020–12 028, May 2017. [Online]. Available: http://www.osapublishing.org/oe/abstract.cfm?URI=oe-25-10-12020

[31] "Understanding eye pattern measurements application note," 2010.

[32] S. Mehtab, J. Sen, and S. Dasgupta, "Robust analysis of stock price time series using cnn and lstm-based deep learning models," in *2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, Nov 2020, pp. 1481–1486.

[33] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)? – arguments against avoiding RMSE in the literature," *Geoscientific Model Development*, vol. 7, no. 3, pp. 1247–1250, 2014. [Online]. Available: https://gmd.copernicus.org/articles/7/1247/2014/

[34] S. Miao, Z. J. Wang, and R. Liao, "A CNN regression approach for real-time 2D/3D registration," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1352–1363, 2016.

References

[1] I. Jorge, J. Rebola, and A. Cartaxo, "Transmission of PAM4 signals in intra-datacenters connections with direct-detection and multicore fibers limited by inter-core crosstalk," Master's thesis, Iscte–Instituto Universitário de Lisboa, 2019.

[2] R. Dias, J. Rebola, and A. Cartaxo, "Transmission of PAM4 signals in amplified inter-datacenters connections with direct-detection and multicore fibers limited by inter-core crosstalk," Master's thesis, Iscte–Instituto Universitário de Lisboa, 2020.

[3] R. Dias, J. Rebola, and A. Cartaxo, "Outage probability due to crosstalk from multiple interfering cores in PAM4 inter-datacenter connections," *Photonics*, vol. 8, no. 1, 2021. [Online]. Available: https://www.mdpi.com/2304-6732/8/1/9

[4] H. Yuan, M. Furdek, A. Muhammad, A. Saljoghei, L. Wosinska, and G. Zervas, "Space-division multiplexing in data center networks: On multi-core fiber solutions and crosstalk-suppressed resource allocation," *Journal of Optical Communications and Networking*, vol. 10, no. 4, pp. 272–288, Apr 2018. [Online]. Available: http://jocn.osa.org/abstract.cfm?URI=jocn-10-4-272

[5] B. Zhu, "SDM fibers for data center applications," in *Optical Fiber Communication Conference (OFC) 2019*. Optical Society of America, 2019, p. M1F.4. [Online]. Available: http://www.osapublishing.org/abstract.cfm?URI=OFC-2019-M1F.4

[6] L. Zhang, J. Chen, E. Agrell, R. Lin, and L. Wosinska, "Enabling technologies for optical data center networks: Spatial division multiplexing," *Journal of Lightwave Technology*, vol. 38, no. 1, pp. 18–30, 2020.

[7] T. M. F. Alves and A. V. T. Cartaxo, "Intercore crosstalk in homogeneous multicore fibers: Theoretical characterization of stochastic time evolution," *Journal of Lightwave Technology*, vol. 35, no. 21, pp. 4613–4623, 2017.

[8] M. Noormohammadpour and C. S. Raghavendra, "Datacenter traffic control: Understanding techniques and tradeoffs," *IEEE Communications Surveys Tutorials*, vol. 20, no. 2, pp. 1492–1525, 2018.

[9] T. Benson, A. Akella, and D. A. Maltz, "Network traffic characteristics of data centers in the wild," in *Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement*, ser. IMC '10. New York, NY, USA: Association for Computing Machinery, 2010, p. 267–280. [Online]. Available: https://doi.org/10.1145/1879141.1879175

[10] Cisco. (2020) Cisco annual internet report - cisco annual internet report (2018–2023) white paper. [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html

[11] J. Krause Perin, A. Shastri, and J. M. Kahn, "Data center links beyond 100 Gbit/s per wavelength," *Optical Fiber Technology*, vol. 44, pp. 69 – 85, 2018, special Issue on Data Center Communications. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1068520017302183

[12] X. Pang, O. Ozolins, L. Zhang, A. Udalcovs, R. Lin, R. Schatz, U. Westergren, S. Xiao, W. Hu, G. Jacobsen, S. Popov, and J. Chen, "Beyond 200 Gbps per lane intensity modulation direct detection (IM/DD) transmissions for optical interconnects: Challenges and recent developments," in *2019 Optical Fiber Communications Conference and Exhibition (OFC)*, 2019, pp. 1–3.

[13] H. Mardoyan, F. Jorge, O. Ozolins, J. M. Estaran, A. Udalcovs, A. Konczykowska, M. Riet, B. Duval, V. Nodjiadjim, J. Dupuy, X. Pang, U. Westergren, J. Chen, S. Popov, and S. Bigo, "204-GBaud on-off keying transmitter for inter-data center communications," in *2018 Optical Fiber Communications Conference and Exposition (OFC)*, 2018, pp. 1–3.

[14] K. Saitoh and S. Matsuo, "Multicore fiber technology," *Journal Lightwave Technology*, vol. 34, no. 1, pp. 55–66, Jan 2016. [Online]. Available: http://jlt.osa.org/abstract.cfm?URI=jlt-34-1-55

[15] J. Zhang, K. Wang, Y. Wei, L. Zhao, W. Zhou, J. Xiao, B. Liu, X. Xin, F. Zhao, Z. Dong, and J. Yu, "280 Gb/s IM/DD PS-PAM-8 transmission over 10 km SSMf at O-Band for optical interconnects," in *2020 Optical Fiber Communications Conference and Exhibition (OFC)*, 2020, pp. 1–3.

[16] J. L. Rebola, A. V. T. Cartaxo, T. M. F. Alves, and A. S. Marques, "Outage probability due to intercore crosstalk in dual-core fiber links with direct-detection," *IEEE Photonics Technology Letters*, vol. 31, no. 14, pp. 1195–1198, 2019.

[17] F. Musumeci, C. Rottondi, A. Nag, I. Macaluso, D. Zibar, M. Ruffini, and M. Tornatore, "An overview on application of machine learning techniques in optical networks," *IEEE Communications Surveys Tutorials*, vol. 21, no. 2, pp. 1383–1408, 2019.

[18] J. Mata, I. de Miguel, R. J. Durán, N. Merayo, S. K. Singh, A. Jukan, and M. Chamania, "Artificial intelligence (AI) methods in optical networks: A comprehensive survey," *Optical Switching and Networking*, vol. 28, pp. 43 – 57, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S157342771730231X

[19] D. Wang, M. Zhang, Z. Li, J. Li, M. Fu, Y. Cui, and X. Chen, "Modulation format recognition and OSNR estimation using CNN-based deep learning," *IEEE Photonics Technology Letters*, vol. 29, no. 19, pp. 1667–1670, 2017.

[20] D. Wang, Y. Xu, J. Li, M. Zhang, J. Li, J. Qin, C. Ju, Z. Zhang, and X. Chen, "Comprehensive eye diagram analysis: A transfer learning approach," *IEEE Photonics Journal*, vol. 11, no. 6, pp. 1–19, 2019.

[21] 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, 2017.

[22] J. Chen, Y. Gong, M. Fiorani, and S. Aleksic, "Optical interconnects at the top of the rack for energy-efficient data centers," *IEEE Communications Magazine*, vol. 53, no. 8, pp. 140–148, 2015.

[23] K. Bilal, S. U. R. Malik, S. U. Khan, and A. Y. Zomaya, "Trends and challenges in cloud datacenters," *IEEE Cloud Computing*, vol. 1, no. 1, pp. 10–20, 2014.

[24] M. Chagnon, "Optical communications for short reach," *Journal of Lightwave Technology*, vol. 37, no. 8, pp. 1779–1797, 2019.

[25] Y. Miyamoto and R. Kawamura, "Space division multiplexing optical transmission technology to support the evolution of high-capacity optical transport networks," *NTT Technical Review*, vol. 15, 06 2017.

[26] X. Zhou, R. Urata, and H. Liu, "Beyond 1Tb/s datacenter interconnect technology: Challenges and solutions (invited)," in *Optical Fiber Communication Conference (OFC) 2019*. Optical Society of America, 2019, p. Tu2F.5. [Online]. Available: http://www.osapublishing.org/abstract.cfm?URI=OFC-2019-Tu2F.5

[27] X. Zhou, H. Liu, R. Urata, and S. Zebian, "Scaling large data center interconnects: Challenges and solutions," *Optical Fiber Technology*, vol. 44, pp. 61 – 68, 2018, special Issue on Data Center Communications. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1068520017302432

[28] E. Agrell, A. Alvarado, and F. Kschischang, "Implications of information theory in optical fibre communications," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 374, p. 20140438, 03 2016.

[29] 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. 24 190–24 239, Sep 2018. [Online]. Available: http://www.opticsexpress.org/abstract.cfm?URI=oe-26-18-24190

[30] D. Richardson, J. Fini, and L. Nelson, "Space division multiplexing in optical fibres," *Nature Photonics*, vol. 7, pp. 354–362, 05 2013.

[31] W. Klaus, B. J. Puttnam, R. S. Luís, J. Sakaguchi, J.-M. D. Mendinueta, Y. Awaji, and N. Wada, "Advanced space division multiplexing technologies for optical networks [invited]," *Journal of Optical Communications and Networking*, vol. 9, no. 4, pp. C1–C11, Apr 2017. [Online]. Available: http://jocn.osa.org/abstract.cfm?URI=jocn-9-4-C1

[32] J. Van Kerrebrouck, X. Pang, O. Ozolins, R. Lin, A. Udalcovs, L. Zhang, H. Li, S. Spiga, M. Amann, L. Gan, M. Tang, S. Fu, R. Schatz, G. Jacobsen, S. Popov, D. Liu, W. Tong, G. Torfs, J. Bauwelinck, J. Chen, and X. Yin, "High-speed PAM4-based optical SDM interconnects with directly modulated long-wavelength VCSEL," *Journal of Lightwave Technology*, vol. 37, no. 2, pp. 356–362, 2019.

[33] O. Ozolins, X. Pang, A. Udalcovs, R. Lin, J. V. Kerrebrouck, L. Gan, L. Zhang, M. Tang, S. Fu, R. Schatz, U. Westergren, G. Jacobsen, D. Liu, W. Tong, G. Torfs, J. Bauwelinck, J. Chen, S. Popov, and X. Yin, "7×149 gbit/s pam4 transmission over 1 km multicore fiber for short-reach optical interconnects," in *Conference on Lasers and Electro-Optics*. Optical Society of America, 2018, p. SM4C.4. [Online]. Available: http://www.osapublishing.org/abstract.cfm?URI=CLEO_SI-2018-SM4C.4

[34] P. Kim, *MATLAB Deep Learning: With Machine Learning, Neural Networks and Artificial Intelligence*, 1st ed. USA: Apress, 2017.

[35] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, http://www.deeplearningbook.org.

[36] A. Sifaoui, A. Abdelkrim, and M. Benrejeb, "On the use of neural network as a universal approximator," *International Journal of Sciences*, vol. 2, pp. 386–399, 08 2008.

[37] D. Wang and M. Zhang, "Artificial intelligence in optical communications: From machine learning to deep learning," *Frontiers in Communications and Networks*, vol. 2, p. 9, 2021. [Online]. Available: https://www.frontiersin.org/article/10.3389/frcmn.2021.656786

[38] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai, and T. Chen, "Recent advances in convolutional neural networks," *Pattern Recognition.*, vol. 77, no. C, p. 354–377, May 2018. [Online]. Available: https://doi.org/10.1016/j.patcog.2017.10.013

[39] A. Mikołajczyk and M. Grochowski, "Data augmentation for improving deep learning in image classification problem," in *2018 International Interdisciplinary PhD Workshop (IIPhDW)*, 2018, pp. 117–122.

[40] J. Tobin, R. Fong, A. Ray, J. Schneider, W. Zaremba, and P. Abbeel, "Domain randomization for transferring deep neural networks from simulation to the real world," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2017, pp. 23–30.

[41] M. Ani, H. Basevi, and A. Leonardis, "Quantifying the use of domain randomization," in *2020 25th International Conference on Pattern Recognition (ICPR)*, 2021, pp. 6128–6135.

[42] X. Yue, Y. Zhang, S. Zhao, A. Sangiovanni-Vincentelli, K. Keutzer, and B. Gong, "Domain randomization and pyramid consistency: Simulation-to-real generalization without accessing target domain data," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019, pp. 2100–2110.

[43] W. S. Saif, M. A. Esmail, A. M. Ragheb, T. A. Alshawi, and S. A. Alshebeili, "Machine learning techniques for optical performance monitoring and modulation format identification: A survey," *IEEE Communications Surveys Tutorials*, vol. 22, no. 4, pp. 2839–2882, 2020.

[44] A. Macho, C. García-Meca, F. J. Fraile-Peláez, M. Morant, and R. Llorente, "Birefringence effects in multi-core fiber: coupled local-mode theory," *Optics Express*, vol. 24, no. 19, pp. 21 415–21 434, Sep 2016. [Online]. Available: http://www.opticsexpress.org/abstract.cfm?URI=oe-24-19-21415

[45] G. Agrawal, *Fiber-Optic Communication Systems*, 4th ed. Hoboken, NJ, USA: John Wiley & Sons, 2010.

[46] C. Pulikkaseril, L. A. Stewart, M. A. F. Roelens, G. W. Baxter, S. Poole, and S. Frisken, "Spectral modeling of channel band shapes in wavelength selective switches," *Opt. Express*, vol. 19, no. 9, pp. 8458–8470, Apr 2011. [Online]. Available: http://www.osapublishing.org/oe/abstract.cfm?URI=oe-19-9-8458

[47] J. L. Rebola and A. V. T. Cartaxo, "On the quaternary level spacing signalling optimisation for increasing the transmission distance in optical communication systems," 2000.

[48] S. M. Ranzini, S. M. Rossi, R. C. Figueiredo, M. M. Rodrigues, and A. Chiuchiarelli, "1-Tb/s WDM PAM-4 transmission over 80 km of uncompensated SSMF link with direct detection and reduced complexity DSP," in *2017 SBMO/IEEE MTT-S International Microwave and Optoelectronics Conference (IMOC)*, 2017, pp. 1–3.

[49] Z. Xing, A. Samani, M. Xiang, E. El-Fiky, T. M. Hoang, D. Patel, R. Li, M. Qiu, M. G. Saber, M. Morsy-Osman, and D. V. Plant, "100 Gb/s PAM4 transmission system for datacenter interconnects using a SiP ME-MZM based DAC-less transmitter and a VSB self-coherent receiver," *Optics Express*, vol. 26, no. 18, pp. 23 969–23 979, Sep 2018. [Online]. Available: http://www.osapublishing.org/oe/abstract.cfm?URI=oe-26-18-23969

[50] R. S. Luís, B. J. Puttnam, A. V. T. Cartaxo, W. Klaus, J. M. D. Mendinueta, Y. Awaji, N. Wada, T. Nakanishi, T. Hayashi, and T. Sasaki, "Time and modulation frequency dependence of crosstalk in homogeneous multi-core fibers," *Journal of Lightwave Technology*, vol. 34, no. 2, pp. 441–447, 2016.

[51] 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, 2019.

[52] G. Rademacher, R. S. Luís, B. J. Puttnam, Y. Awaji, and N. Wada, "Crosstalk dynamics in multi-core fibers," *Optics Express*, vol. 25, no. 10, pp. 12 020–12 028, May 2017. [Online]. Available: http://www.osapublishing.org/oe/abstract.cfm?URI=oe-25-10-12020

[53] S. Mehtab, J. Sen, and S. Dasgupta, "Robust analysis of stock price time series using cnn and lstm-based deep learning models," in *2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, Nov 2020, pp. 1481–1486.

[54] "Understanding eye pattern measurements application note," 2010.

[55] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)? – arguments against avoiding RMSE in the literature," *Geoscientific Model Development*, vol. 7, no. 3, pp. 1247–1250, 2014. [Online]. Available: https://gmd.copernicus.org/articles/7/1247/2014/

[56] S. Miao, Z. J. Wang, and R. Liao, "A CNN regression approach for real-time 2D/3D registration," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1352–1363, 2016.

[57] Student, "The probable error of a mean," *Biometrika*, pp. 1–25, 1908.

[58] L. V. Hedges, "Distribution theory for glass's estimator of effect size and related estimators," *Journal of Educational and Behavioral Statistics*, vol. 6, no. 2, pp. 107–128, 1981. [Online]. Available: https://EconPapers.repec.org/RePEc:sae:jedbes:v:6:y:1981:i:2:p:107-128

[59] J. A. Durlak, "How to Select, Calculate, and Interpret Effect Sizes," *Journal of Pediatric Psychology*, vol. 34, no. 9, pp. 917–928, 02 2009. [Online]. Available: https://doi.org/10.1093/jpepsy/jsp004