Deep learning for BER prediction in optical connections impaired by inter-core crosstalk

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

***Abstract*—Four-level pulse amplitude modulation (PAM4) signals transmission in short-haul intensity modulation-direct detection datacenters connections supported by homogeneous weakly-coupled multicore fibers is seen as a promising technology to meet the future challenge of providing enough bandwidth and achieve high data capacity in datacenter links. However, in multicore fibers, inter-core crosstalk (ICXT) limits significantly the performance of such short-reach connections by causing large bit error rate (BER) fluctuations. 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 by estimation of the root mean square error (RMSE) using a synthetic dataset created with Monte Carlo simulation. Considering PAM4 interdatacenter connections with one interfering core and for different skew-symbol rate products, extinction ratios and crosstalk levels, the obtained results show that the implemented CNN is able to predict the BER without surpassing a RMSE limit of 0.1.**

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

I. Introduction

As the number of network telecommunication users and devices grows exponentially, datacenters are becoming crucial to handle the large amount of data due to their flexibility, scalability in computing and storage resources [1]. The current approach to deal with capacity scalability in datacenter links is using multiple wavelengths (also known as lanes), where each lane carries a 25 Gb/s on-off keying (OOK) signal or a four-level pulse amplitude modulation (PAM4). The PAM4 has already been standardized by the IEEE 802.3bs task force to enable rates of 50 and 100 Gbit/s per lane in datacenters connections [1].

Furthermore, nowadays, datacenter connections typically rely on single core-single mode fibers (SC-SMFs) with intensity modulation-direct detection (IM-DD) transmission. However, it is expected that such fibers will not fulfill the future capacity demand growth, as transmission in SC-SMFs is approaching its limit of 200 Gbit/s per lane with IM-DD [1]. Multicore fibers (MCFs) have been proposed in order to overcome this issue and to reduce fiber per area density in datacenters [2]. In particular, homogeneous weakly-coupled multicore fibers (WC-MCFs) ensure simplicity, low cost and power consumption as required by datacenter links. Even so, transmission in such MCFs is impaired by inter-core crosstalk (ICXT) [3], [4]. This crosstalk in MCFs arises from power coupling between cores and is particularly relevant to the transmission of signals with the same wavelength in neighboring cores [5], [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 cause large bit error rate (BER) fluctuations, that can lead to system outage periods [2], [4], [7]. Hence, ICXT monitoring in such datacenter connections is crucial to predict and maintain acceptable BERs. Furthermore, in IM-DD systems, eye-pattern analysis is of primary importance for monitoring and to extract the optical signal quality, and subsequently predict the BER from the optical signal observation. As this analysis depends on engineering skills and expertise, which is prone to error, automatic prediction of the BER from the eye-pattern, for example, by resorting to machine-learning is a promising way to effectively address this problem and monitor the ICXT in datacenter connections.

Recently, the use of convolutional neural networks (CNN) has been investigated for optical performance monitoring by eye-patterns analysis [8], [9]. In [8], a CNN-based technique is used to estimate the optical signal-to-noise ratio (OSNR) and identify the modulation-format, from eye-pattern images. In [9], an eye-pattern analysis scheme based on a CNN for IM-DD transmissions is presented. In this work, different eye-patterns of OOK and PAM4 signals from back-to-back (B2B) up to 80 km link transmissions are sent to a CNN-based model that outputs eye-pattern characteristics, estimates the fiber link length and Q-factor and performs impairments recognition [9]. Inspired by these previous results, we contribute in this paper with the first application of CNN regression to the problem of eye-pattern analysis and BER prediction in PAM4 inter-datacenter optical connections impaired by ICXT.

This paper is organised as follows. Section II describes the system developed for the CNN-based eye pattern analysis and BER prediction and its performance is assessed in Section III. The main conclusions regarding this work are presented in Section IV.

II. Model description

The system developed for the CNN-based eye pattern analysis and BER prediction has two main components, the Monte Carlo simulator, that creates a dataset and the machine learning

module, that processes this dataset. Both components were implemented with MATLAB. The schematic of the main tasks of the implemented system is illustrated in Fig. 1.

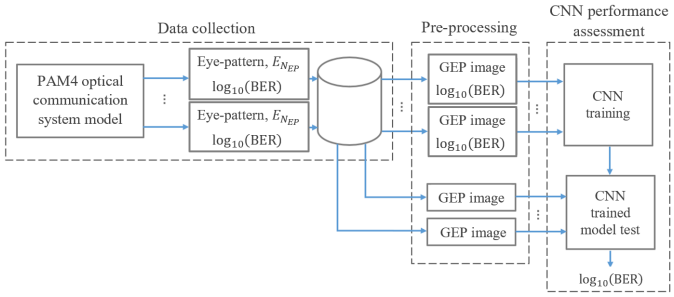

Fig. 1. Main tasks of the developed system for the CNN-based eye-pattern analysis and BER prediction.

The system's first main task corresponds to the generation of eye-patterns and the corresponding BERs calculation from the PAM4 optical communication system model that is detailed in section II-A. Then, data collection is performed as explained in section II.B. Before the CNN training, the eye-patterns are pre-processed to obtain what we denote as grayscale eye-patterns (GEPs) images as described in section II-C. After that, the CNN, whose architecture is described in section II.D, is trained to predict the BER, for a given input GEP. Then, the CNN's prediction ability is evaluated against unseen GEP images.

A. Intercore crosstalk modelling

The impact of ICXT on the performance of optically amplified PAM4 links for inter-datacenter connections has been analyzed in [2] by resorting to the dual polarization-discrete changes model (DP-DCM) that describes with reasonable accuracy the ICXT effect in homogeneous WC-MCFs [10]. In this work, as in [2], only two cores, the interfering core *m* and the interfered core *n* are considered. Two transmitters, one for each core, generate the signal transmitted along core *m*, which is the interfering PAM4 signal, and the signal transmitted along core *n* corresponding to the interfered PAM4 signal.

The PAM4 signal, with the extinction ratio given by the inverse of the ITU-T definition [2], travels along core *n* of the MCF, where linear propagation is assumed, since non-linear effects are usually insignificant in inter-datacenters distances [1], [11]. Fiber loss is also assumed the same in both cores. The evolution of the ICXT along time on the interfered cores is modelled by the DP-DCM 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 MCF in each iteration of the Monte Carlo simulator. A set of N_p RPSs generated in each MC iteration and the corresponding transfer functions are defined as MCF realizations [10]. The RPSs are modelled using an uniform distribution between $[0, 2\pi[$. The temporal dependence of the ICXT is induced by effects such as the walk-off due to different group velocities between cores [11]. In the DP-DCM, the skew between the interfering core *m* and the interfered core *n* is given by $S_{mn} = d_{mn}L$, where L is the MCF length and d_{mn} is the walkoff between cores *m* and *n* [2], [6]. Therefore, in each iteration of the Monte Carlo simulator, a new PAM4 signal with symbols randomly generated is transmitted in core *m* and is passed to the transfer functions corresponding to one MCF realization. The ICXT level is quantified by the ratio between the mean ICXT power and the mean power of the signal both at the output of the interfered core as $X_c = N_p|K_{nm}|^2$, where K_{nm} is the average inter-core coupling coefficient [4], [10].

At the output of the MCF, a chromatic dispersion compensation (CDC) module fully compensates the chromatic dispersion (CD) arising in core *n*, which is modelled considering a dispersion compensating fiber (DCF). 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 MCF and DCF and the ASE noise is modelled as additive white Gaussian noise [2], [4]. The optical filter is modelled by a 4th order super Gaussian filter.

After CDC and amplification, the PAM4 signal degraded by ICXT and ASE noise reaches the DD receiver dedicated to core *n*, where is converted to an electrical signal by the photo-detector. 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 is calculated by the semi-analytical method known as the exhaustive Gaussian approach [2]. Effects such as electrical noise, signal-ASE, and ASE-ASE beat noises are taken into account analytically, while ICXT and inter-symbol interference are taken into account by the waveform distortion observed in the eye-patterns.

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 3.8×10^{-3}, which is the threshold typically used for datacenters connections with forward-error correction [2]. The electrical and optical receiver filters bandwidth were optimized in B2B operation to maximize the receiver sensitivity [2]. For the MCF length of 80 km and extinction ratios of *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 phase-matching points is set to characterize accurately the RPS mechanism [2]. We consider the crosstalk levels $X_c =$ -16 dB, -14 dB and -12 dB. Two different intercore skew-symbol rate products of $|S_{mn}R_s| = 1000$ and $|S_{mn}R_s| = 0.01$ are also considered. The case of $|S_{mn}R_s| = 1000$ is referred as high skew symbol rate product as $|S_{mn}R_s| >> 1$ [2], [6]. The situation of $|S_{mn}R_s| = 0.01$ is referred as low skew-symbol rate product, since $|S_{mn}R_s| << 1$ [2], [6].

Fig. 2 shows an example of the eye-patterns at the decision circuit input impaired by ICXT, obtained with $X_c =$ -14 dB and *r*=0.1. Fig. 2 a) and b) correspond, respectively, to the best and worst BERs obtained with $|S_{mn}R_s| = 0.01$ and

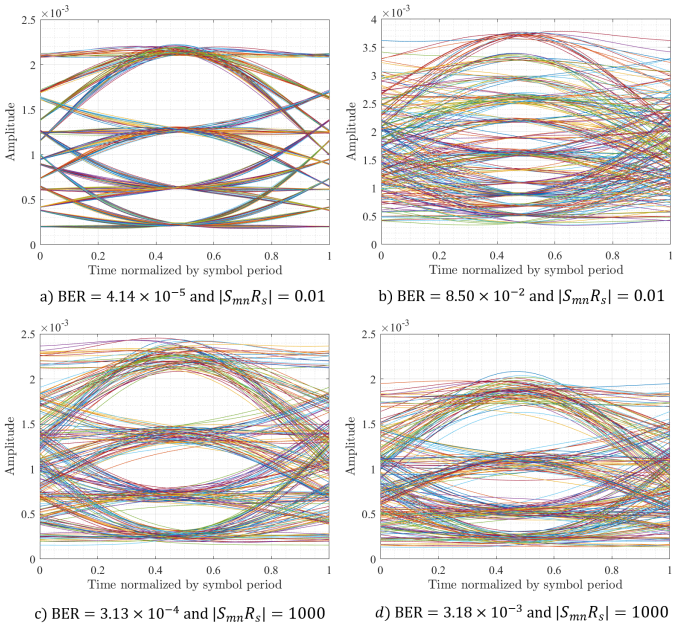

Fig. 2. Eye-patterns at the decision circuit input for X_c = -14 dB and r=0.1 of a) best BER and b) worst BER with $|S_{mn}R_s| = 0.01$ and c) best BER and d) worst BER with $|S_{mn}R_s| = 1000$.

Fig. 2 c) and d) to the best and worst BERs obtained with $|S_{mn}R_s| = 1000$ after 1000 MCF realizations. In Fig. 2 b), the eye-pattern is practically closed, in comparison with Fig. 2 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. 2 a) shows that the amplitude levels are more defined than in the eye-pattern shown in Fig. 2 c) with high $|S_{mn}R_s|$, since, in this last case, more symbols in the interfering core are contributing to ICXT [2]. Fig. 2 confirms that the product $|S_{mn}R_s|$ strongly affects the impact of ICXT on the received eye-patterns.

Fig. 3 shows the BERs histograms and the corresponding average BER obtained after 1000 MCF realizations with X_c = -14 dB, r=0.1 and r=0, for a) $|S_{mn}R_s| = 1000$ and b) $|S_{mn}R_s| = 0.01$. 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 not exceeded and the system is never in outage. Fig. 3 also shows that the effect of ICXT on the BER distribution is less detrimental with $r = 0.1$. For $r = 0$, a higher spreading of the BER is observed, and, hence, more MCF realizations lead to outage. The results in both histograms confirm again 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 than in comparison with the one obtained for $|S_{mn}R_s| = 0.01$.

B. Data collection

The data used to train and test the CNN corresponds to synthetic data collected from different Monte Carlo simulator iterations. Each pair of collected data consists of a received

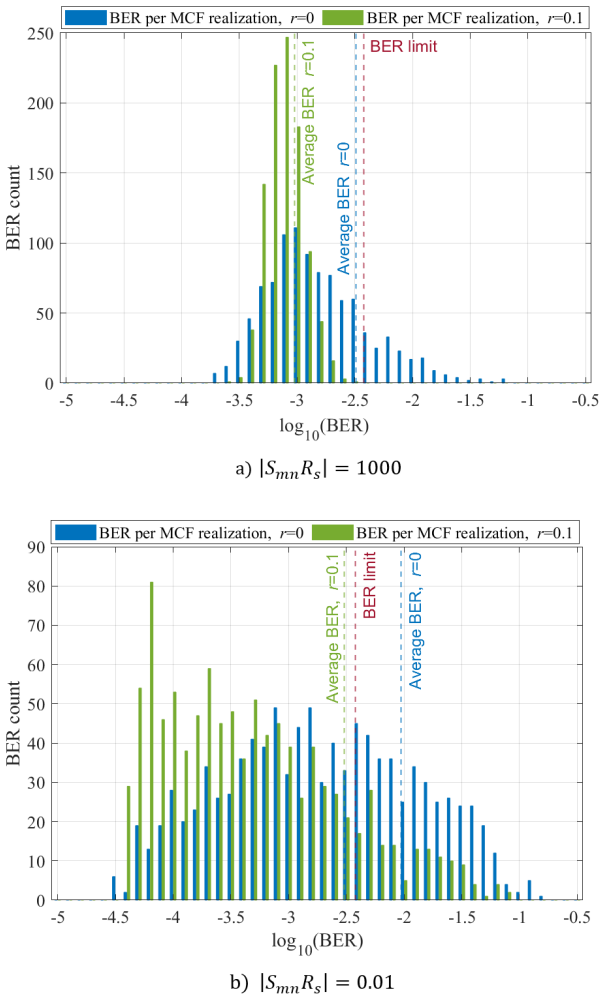

Fig. 3. Histogram of the $\log_{10}(BER)$ for 1000 MCF realizations with X_c = -14 dB, r=0.1 and r=0, for a) $|S_{mn}R_s| = 1000$ and b) $|S_{mn}R_s| = 0.01$.

and synthetically generated eye-pattern, E_i (with i=1,...,N_{EP}), where N_{EP} is the number of eye-patterns generated, such as the ones shown in Fig. 2, and the corresponding BER calculated logarithmically, i.e., $\log_{10}$(BER). Each eye-pattern corresponds to a 32×256 matrix, with 32 amplitude samples per symbol and 256 generated PAM4 symbols in each Monte Carlo iteration. The eye-patterns are obtained by varying several optical link parameters, such as the crosstalk level, skew-symbol rate product and extinction ratio.

The CNN performance is highly dependent on the diversity of the training data. Hence, we apply an active sampling approach to collect a balanced number of BERs and corresponding eye-patterns. For the different optical links (defined as the IM-DD datacenter links with different parameters), the data generated 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 are saved in every 0.1 interval of the $\log_{10}$(BER). Within these N_{BER} intervals, where $N_{BER} = \log_{10}(\text{BER}_{max}) - \log_{10}(\text{BER}_{min}) \times 0.1$, a second balancing step is performed with 20 subintervals, 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 BERs distribution is much dependent on the link parameters, as seen in section II-A.

As the ICXT has a stochastic behaviour, some BERs are less frequent than others, particularly for higher BERs associated with low outage probabilities. 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 that are less likely to occur, the DP-DCM model has been implemented with a slight change following a domain randomization-based approach.

Firstly, the Monte Carlo simulator starts with a random set of N_p RPSs, which is stored and used in the first MCF realization. After that, based on a single uniformly distributed random number in the interval $[0, 1]$, the simulator determines 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 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. This new set of RPSs generated with this random perturbation is not stored, and, therefore, it is not used for the following MCF realizations. Only the eye-pattern and $\log_{10}$(BER) are saved if the corresponding N_{BER} interval is not filled.

C. Eye-pattern pre-processing

Before training and testing the CNN, the eye-patterns are pre-processed to obtain what we denote as GEP images. Typically, oscilloscopes sample the received signal and generate a two-dimensional plot, 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 [12]. 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. In this work, a similar approach is used, where the plot density is represented by the grayscale value of each pixel. Furthermore, two different approaches regarding the eye-pattern representation are considered: fixed scale GEP and dynamic scale GEP. The process to obtain GEP images is explained as follows.

First, for a $x \times y$ GEP image, the amplitudes of a synthetic eye-pattern obtained from the simulator 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 is performed by dividing the eye-pattern amplitudes by the maximum amplitude that occurred in all eye-patterns obtained during a CNN training. For the case of a dynamic scale GEP, each eye-pattern is normalized taking into account its maximum amplitude, without resorting to any information from other eye-patterns. After normalization of both approaches, the eye-pattern to GEP conversion process continues by generating a $y \times x$ GEP matrix with zero as elements. The plot density is modeled by incrementing one unit in all elements of the $x \times y$ 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]. In this work, we use 32×32 GEP images.

After this transformation process, the GEP images are used as input data for the CNN.

D. Convolutional neural network

In this work, the CNN model is developed on MATLAB using the Deep Learning Toolbox and Deep Network Designer.

The proposed CNN architecture is composed of a sequential stack of layers based on [8], [9], and is schematically shown in Fig. 4. 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. In Fig. 4, a fixed scale GEP is presented. The CNN 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, with both kernel sizes of 5×5, produce, respectively, 32 and 64 feature maps, whereas the layers C3, C4 and C5, with kernels sizes of 3×3, produce 128, 256 and 512 feature maps, respectively. 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 output of each convolutional layer is normalized using a batch normalization (BN) layer followed by a rectified linear unit (ReLU) layer for an effective and faster training [13]. After each of the first four Conv+BN+ReLU operations, a down-sampling is performed by an average pooling layer. We set the four pooling layers, P1, P2, P3 and P4, with 2×2 subsampling regions and stride (2, 2). After feature extraction at the last C5+BN+ReLU operation, a dropout layer is placed to prevent overfitting, followed by a fully connected layer and a regression layer used to predict the BER.

Regarding computational complexity and resources, using the proposed method, the CNN must be trained firstly offline in a specific PAM4 datacenter connection to extract the features of the ICXT impairing that connection and to learn how to predict the BER in an accurate way. Nowadays, such offline approach does not imply any highly demanding computational resources. Then, the trained CNN must be implemented in the digital signal processors (DSPs) of the receiving equipment. Nowadays, DSPs at optical receivers are trivial and sufficiently powerful to perform complicated tasks, such as forward-error correction in IM-DD systems [1]. Furthermore, the ICXT decorrelation time is in the order of minutes and within a subminute scale, the ICXT can be highly correlated [14]. As deep neural networks in real-world applications have an inference time of few ms to one second [15], the proposed approach seems perfectly capable to cope with the ICXT non-stationarity.

The synthetically generated dataset is randomly split before each CNN training, where 70% is assigned as training data, 15% as validation data and 15% as test data. The CNN is trained with a stochastic gradient descent with momentum

optimizer. The maximum number of epochs is set to 30, since lower values resulted in worse performances and higher values did not enhance the performance and led to a much higher computation time. A mini-batch size with 8 observations at each iteration is used, since it performed better when compared to 16, 32 and 64 mini-batch sizes. The initial learning rate has been empirically set to 1×10^{-4} and reduced by a factor of 0.1 after 20 epochs.

III. CNN PERFORMANCE ASSESSMENT

In this section, the performance of the CNN developed for BER prediction from eye-pattern analysis in PAM4 inter-datacenter optically amplified short IM-DD connections impaired by ICXT is assessed by estimation of the root mean square error (RMSE). This performance evaluation metric has been chosen because it has been widely used in the literature as a key CNN regression performance indicator [13]. In this work, we consider a RMSE below 0.1 as an acceptable prediction of the $\log_{10}$ (BER).

A. CNN performance for specific optical link parameters

Firstly, the accuracy of the BER prediction is assessed by training the CNN and testing its regression capabilities for each type of optical link. Table I shows the RMSEs obtained for the CNN when trained with fixed and dynamic scale GEPs.

When using fixed scale GEPs, the CNN model is able to predict the $\log_{10}$ (BER) without surpassing the RMSE limit of 0.1, with exception for X_c = -16 dB and X_c = -12 dB with $|S_{mn}R_s| = 1000$ and *r*=0. From all training scenarios, 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 in general for the case of $|S_{mn}R_s| = 0.01$ and *r*=0.1, being the RMSE lower than 0.05, for all crosstalk levels. When using a dynamic scale GEPs, the CNN is also able to predict the $\log_{10}$ (BER) without surpassing the RMSE limit of 0.1, except for X_c = -16 dB and -12 dB with $|S_{mn}R_s| = 1000$ and *r*=0.

Table I shows that the CNN predicts the BER from the eye-patterns impaired by ICXT, with a reduced RMSE, when the CNN is trained for specific optical link parameters.

B. CNN generalization

In this subsection, the generalization capabilities of the CNN to other optical link parameters is evaluated by verifying its generalization capabilities to situations for which it has not been trained. A first study is performed with a CNN trained with 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 maximum amplitude obtained in all eye-patterns generated for both crosstalk levels of -16 and -14 dB. A RMSE of 0.654 was obtained, showing that the CNN is unable to predict correctly the BER from the test GEPs obtained with crosstalk levels different from the ones used during training. The CNN was also unable to predict correctly the BER of dynamic scale GEPs obtained from optical links with X_c = -16 dB. However, as the RMSE is lower than in the case of fixed GEPs (RMSE=0.285), it seems more able to perform the generalization.

TABLE I
RMSE OF THE CNN TRAINED AND TESTED WITH FIXED AND DYNAMIC SCALE GEPS, RESPECTIVELY DENOTED AS FS-GEPS AND DS-GEPS, AS A FUNCTION OF X_c.

Link type		FS-GEPs	DS-GEPs
X_c = -16 dB	$\|S_{mn}R_s\| = 0.01, r = 0$	0.083	0.075
	$\|S_{mn}R_s\| = 0.01, r = 0.1$	0.037	0.037
	$\|S_{mn}R_s\| = 1000, r = 0$	0.113	0.105
	$\|S_{mn}R_s\| = 1000, r = 0.1$	0.074	0.080
X_c = -14 dB	$\|S_{mn}R_s\| = 0.01, r = 0$	0.041	0.091
	$\|S_{mn}R_s\| = 0.01, r = 0.1$	0.049	0.037
	$\|S_{mn}R_s\| = 1000, r = 0$	0.099	0.093
	$\|S_{mn}R_s\| = 1000, r = 0.1$	0.066	0.069
X_c = -12 dB	$\|S_{mn}R_s\| = 0.01, r = 0$	0.091	0.076
	$\|S_{mn}R_s\| = 0.01, r = 0.1$	0.047	0.054
	$\|S_{mn}R_s\| = 1000, r = 0$	0.112	0.107
	$\|S_{mn}R_s\| = 1000, r = 0.1$	0.087	0.090

As the CNN did not perform well in the previous domain shift studies, two training approaches with mixed data, i.e., with different optical link parameters, have also been assessed. Fig. 5 shows the true BER (obtained from the Monte Carlo simulator) as a function of the predicted BER for a CNN trained and tested with fixed scale GEPs obtained with all the crosstalk levels shown in Table 1, $|S_{mn}R_s| = 0.01$, *r*=0 and $\log_{10}$ (BER) $\in [-3.1, -1.9[$. The results show that the CNN trained with different crosstalk levels reached a RMSE of 0.082, considerably lower than the one obtained when training the CNN with a single crosstalk level.

Fig. 6 shows the BERs distribution of the test set from a CNN trained with dynamic scale GEPs using 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 was tested with the full BER range considering eye-patterns obtained for the three crosstalk levels and Fig. 6 shows that the CNN is able to predict the BER for the three crosstalk levels without surpassing the RMSE limit.

IV. CONCLUSIONS

A deep learning approach, based on a CNN, for BER prediction in PAM4 inter-datacenter amplified IM-DD connections impaired by ICXT was proposed. For validation, a synthetic dataset was generated with Monte Carlo simulation employing an active sampling strategy for a proper dataset balancing.

It is shown that the CNN provides good BER predictions with RMSEs below 0.1, for different optical link parameters, skew-symbol rate product, extinction ratio and crosstalk level. Regarding its implementation, a CNN based on dynamic GEPs is a better solution, since it does not require prior knowledge of the amplitudes obtained for other eye-patterns.

In future work, we propose to test the CNN with experimental data and fine-tune the CNN for those experiments.

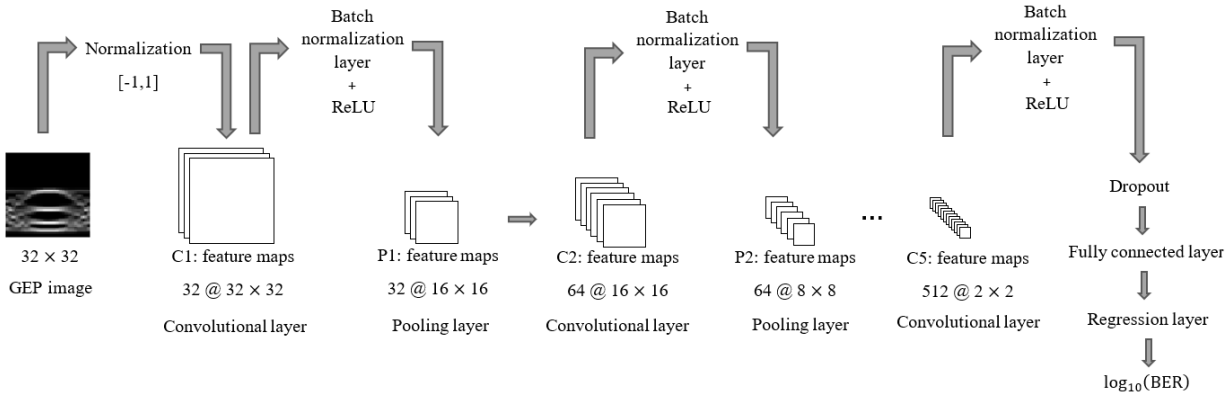

Fig. 4. CNN architecture considered in this work to learn the BER from the GEP images.

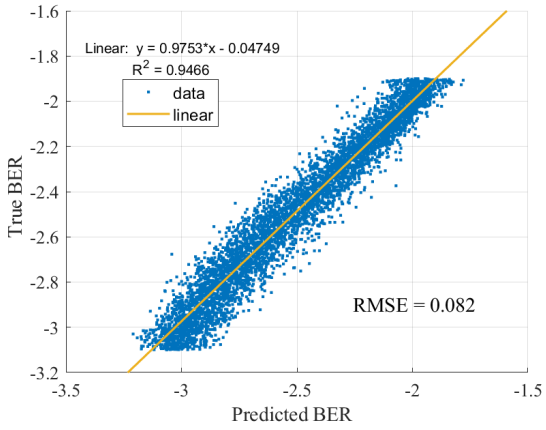

Fig. 5. BER predictions of the test set from the CNN trained for the case of optical links with X_c = -12, −14 and X_c = −16 dB, with $|S_{mn}R_s|$ = 0.01 and r=0. The line represents the linear regression of the data.

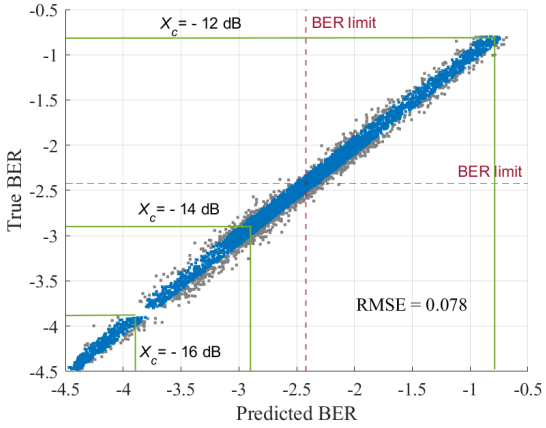

Fig. 6. BER predictions of the test set from a CNN trained for optical links with X_c = -16 dB and $\log_{10}$ (BER) $\in [-4.5, -3.9]$, X_c = -14 dB and $\log_{10}$ (BER) $\in [-3.8, -2.9[$, X_c = -12 dB and $\log_{10}$ (BER) $\in [-2.9, -0.8]$, $|S_{mn}R_s|$ = 0.01 and r=0. The blue data points represent the predictions with a margin error below 0.1.

Acknowledgments

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

References

[1] 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, Aug. 2018.

[2] R. Dias, J. Rebola, and A. Cartaxo, "Outage probability due to crosstalk from multiple interfering cores in PAM4 inter-datacenter connections," *MDPI Photonics*, vol. 8, no. 1, pp. 1–15, Jan. 2021.

[3] K. Saitoh and S. Matsuo, "Multicore fiber technology," *Journal Lightwave Technology*, vol. 34, no. 1, pp. 55–66, Jan. 2016.

[4] T. Hayashi, T. Sasaki, and E. Sasaoka, "Behavior of inter-core crosstalk as a noise and its effect on Q-factor in multi-core fiber," *IEICE Transactions on Communications*, vol. E97.B, pp. 936–944, May 2014.

[5] 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.

[6] G. Rademacher, R. Luís, B. 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.

[7] J. Rebola, A. Cartaxo, T. Alves, and A. 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, Jul. 2019.

[8] 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, Oct. 2017.

[9] 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, Dec. 2019.

[10] R. Soeiro, T. Alves, and A. Cartaxo, "Dual polarization discrete changes model of inter-core crosstalk in multi-core fibers," *IEEE Photonics Technology Letters*, vol. 29, no. 16, pp. 1395–1398, Aug. 2017.

[11] 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.

[12] ANRITSU, "Understanding eye pattern measurements," application note no. 11410-00533 (revision A), 2010.

[13] 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.

[14] T. Alves, A. Cartaxo, and J. Rebola, "Stochastic properties and outage in crosstalk-impaired OOK-DD weakly-coupled MCF applications with low and high skew×bit-rate," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 26, no. 4, pp. 1–8, Aug. 2020.

[15] A. Geifman, "The correct way to measure inference time of deep neural networks," link: https://towardsdatascience.com/the-correct-way-to-measure-inference-time-of-deep-neural-networks-304a54e5187f, May 2020.