TDECQ-Based Optimization of Nonlinear Digital Pre-Distorters for VCSEL-MMF Optical Links Using End-to-end Learning

(Article begins on next page)

01 September 2025

TDECQ-Based Optimization of Nonlinear Digital Pre-Distorters for VCSEL-MMF Optical Links Using End-to-End Learning

Leonardo Minelli , Fabrizio Forghieri, *Fellow, IEEE*, Tong Shao, Ali Shahpari , and Roberto Gaudino , *Senior Member, IEEE*

Abstract—We investigate in this article the use of nonlinear digital pre-distorters (DPDs) for improving the performance of optical transmitters (TX) employing vertical-cavity surface-emitting lasers (VCSELs), according to the standard transmitter and dispersion eye closure quaternary (TDECQ) compliance test for short-reach intra data center interconnects (DCI) using PAM4 over multi-mode fibers (MMF). We present a convolutional neural network (CNN) approach for nonlinear DPD optimization, suitable for training the pre-distorters using either a direct learning architecture (DLA) or an end-to-end (E2E) learning system. Then, we focus on a novel E2E architecture based on the reference TDECQ specifications for MMF optical links at net 100 Gbps per wavelength (λ). We experimentally implement the proposed methodology over a VCSEL-MMF setup compliant to the TDECQ test requirements. We evaluate the TDECQ performance of an optical TX employing a commercial 850 nm VCSEL at 107.2 Gbps driven at several nonlinear conditions, comparing nonlinear DPDs optimized using both DLA and TDECQ-based E2E approaches. Experimental results show that nonlinear DPD significantly enhances TDECQ performance, enabling compliance with the IEEE P802.3db™ requirements for net 100 Gbps/λ even in scenarios in which, without nonlinear DPD, the TDECQ test would fail due to VCSEL nonlinear distortions. In particular, nonlinear DPDs trained using the TDECQ-based E2E approach exhibit a consistent 0.8 dB gain in terms of TDECQ with respect to using the DLA.

Index Terms—TDECQ, VCSEL, Nonlinear Digital Pre-Distortion, End-to-end Learning, Multi-Mode Fibers, intra-Data Center Interconnects.

I. INTRODUCTION

IN RECENT years, the use of internet-related services and the corresponding volume of data traffic has experienced exponential growth. This trend is set to persist in the coming years, with the consolidation of 5G, cloud services, Internet of Things (IoT), and the impending arrival of 6G along with the widespread use of Virtual/Augmented Reality (VR/AR) and Machine-to-Machine applications. Consequently, modern data centers, playing a vital role as the backbone of Internet connectivity, are compelled to continually upgrade their infrastructure: specifically, intra-data center interconnects (intra-DCI) are requested to increase their capacity at reasonable upgrade costs. Currently, intra-DCI optical links predominantly utilize Intensity Modulation and Direct Detection (IM-DD) solutions. In 2023, more than 30% of the optical intra-DCI connectivity from 40G to 400G still exploits multi-mode fibers (MMF) rather than single-mode fibers (SMF) [1]. The MMF deployment is the preferred cost-effective solution for optical links up to about one hundred meters, and is typically coupled with the utilization of Vertical Cavity Surface Emitting Lasers (VCSELs) as optical sources. VCSELs are well known for their low-cost chip manufacturing and exceptional power efficiency [2], [3]. Currently, most of intra-DCI implementations rely on VCSELs [4], and their market is projected to further grow in the coming years [5]. Therefore, by exploiting VCSEL-MMF solutions, next-generation of intra-DCI up to 100 m is expected to provide net 100 Gb/s/λ rate using quaternary Pulse Amplitude Modulation (PAM4) format. In line with this objective, MMF transceiver modules are required to fulfill strict specifications both at receiver (RX) and transmitter (TX) side, in order to ensure standardized vendor interoperability [6]. Specifically for the optical TX, among the several proposed metrics, the TDECQ (Transmitter and Dispersion Eye Closure Quaternary) has established itself in recent years as one of the main standard tests for assessing the quality of the TX signal [7], [8]. However, meeting the TDECQ requirements set by each standard, for instance IEEE P802.3db™ for MMF optical links [6], proves to be a significant challenge at such high data rates using commercial devices. The limited bandwidth of electro-optical components and the non-linear distortions caused by VCSELs significantly degrade the quality of the TX signal. As a consequence, several digital signal processing (DSP) solutions have been proposed for PAM4 signal equalization to mitigate for these impairments [9].

In this article, we investigate the use of nonlinear digital pre-distortion (DPD) to improve the TDECQ performance of VCSEL-based optical transmitters. We propose and illustrate an optimized methodology based on convolutional neural networks (CNN) for nonlinear DPD, specifically tailored to the
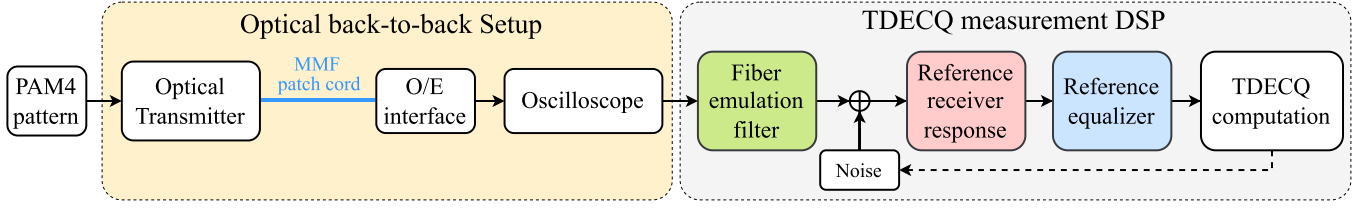
Fig. 1. TDECQ conformance test block diagram, as specified in [6]. O/E: optical-to-electrical.

considered TX system. Our methodology is suitable for both the direct learning architecture (DLA) and the end-to-end (E2E) learning paradigms. Furthermore, we introduce a novel E2E learning architecture, implementing a communication system based on the standard TDECQ test for MMF optical links, which focuses the nonlinear DPD optimization to specifically improve this metric. We experimentally demonstrate the proposed DPD optimization on an optical TX setup compliant to the IEEE requirements for the standard measurement of the TDECQ at net 100 Gbps/λ [6]. We evaluate nonlinear DPDs trained using both DLA and the TDECQ-based E2E approaches, providing, to the best of our knowledge, the first direct comparison between these two optimization over an experimental VCSEL-MMF link. We show that our CNN-based nonlinear DPD optimization is able to remarkably improve the quality of the optical TX signal across a wide range of VCSEL driving conditions. Specifically, the optimized nonlinear DPDs, either based on CNN or Volterra series, experimentally prove to meet the IEEE TDECQ requirements [6] even in scenarios where no DPD or even linear DPD would fail. Moreover, we show that the proposed TDECQ-based E2E learning optimization is able to improve the TDECQ performance by more than 0.8 dB with respect to the DLA approach, proving the capability of the E2E learning to adapt to improve such system-oriented optical TX quality metric.

The article is structured as follows: Section II provides an overview of the TDECQ metric, while Section III explores the context and motivation behind the deployment of nonlinear DPD in VCSEL-MMF optical IM-DD links. Detailed explanation of the proposed DPD optimization methodology is presented in Section IV. In Section V, we showcase the experimental implementation of the nonlinear DPD optimization and deployment, presenting and discussing the results achieved using different pre-distorters. Finally, Section VI concludes the article with final comments on the carried work.

II. OVERVIEW OF THE TDECQ PARAMETER

The TDECQ (Transmitter and Dispersion Eye Closure Quaternary) is a compliance test for PAM4 optical transmitters, which was standardized within the IEEE 802.3bs-2017 [10] Ethernet standard as a replacement for previously used eye-mask and transmitter dispersion penalty (TDP) measures [7]. The TDECQ extends the TDEC metric already deployed for binary NRZ modulations to the PAM4 modulation format, by natively taking into account the presence of a linear adaptive equalizer at the RX side. In particular, the TDECQ evaluates the vertical eye closure of a PAM4-modulated signal when this is propagated

through a reference optical communication link, whose block diagram (specified in [6]) is depicted in Fig. 1. The TDECQ test requires first the transmission of a PAM4 pattern by the optical TX under test, which is then propagated through a back-to-back (B2B) setup and acquired from an oscilloscope after optical-to-electrical (O/E) conversion. The B2B configuration, implemented by connecting the TX and the RX an using a short MMF patch cord (up to 2 meters), allows to attribute the signal distortions exclusively to the optical TX. Then, TDECQ measurement consists of the digital emulation of the filtering effects of a worst-case scenario (WCS) fiber channel, cascaded to the response of a reference RX and a reference adaptive feed-forward equalizer (FFE). In particular, the TDECQ measures, for a target SER, the largest white gaussian noise (WGN) that can be added to the signal at the RX input [10], compared to the maximum tolerable noise when using an ideal distortion-less optical link [8]. The TDECQ can be defined as follows [8]:

$$\text{TDECQ} = 10 \log_{10} \left(C_{eq} \frac{\sigma_{ideal}}{\sigma_{eq}} \right) \quad (1)$$

In (1), σ_{ideal} represents the RX noise root mean square (RMS) value in the ideal scenario (i.e. the WGN standard deviation), and is defined as follows:

$$\sigma_{ideal} = \frac{OMA_{outer}}{6Q_t} \quad (2)$$

where OMA_{outer} is the signal's outer Optical Modulation Amplitude (OMA), and Q_t represents the Q factor for a target SER. The term σ_{eq} refers instead to the RMS value of the maximum tolerable noise at the FFE output. Finally, C_{eq} quantifies the enhancement induced by the cascade of the RX filter and the FFE on the noise injected at the RX input, whose RMS value is given by $\sigma_G = \sigma_{eq}/C_{eq}$[1] Calculating the σ_G value for a specific SER is unfeasible through a closed-form solution. Hence, in the TDECQ measurement process, the tap coefficients of the optimal equalizers are iteratively adjusted by maximizing the σ_G value until the output worst-case SER matches the target value [10]. The SER is measured semi-analytically (i.e. computing error probabilities from histogram windows) over two different sampling instants on the signal eye diagram, spaced by 0.1 in terms of symbol unit interval (UI): the worst-case SER is the highest among the two obtained. As an example, Fig. 2 illustrates the TDECQ measurement on the eye-diagram of a 107.2 Gbps PAM4 signal acquired on the setup we used in our

---

[1]As done in [8], we omit for simplicity the term σ_S, which in the TDECQ definition [10] should be added to σ_G in a root-sum-of-square sense for accounting the actual noise in the RX test equipment. [8].
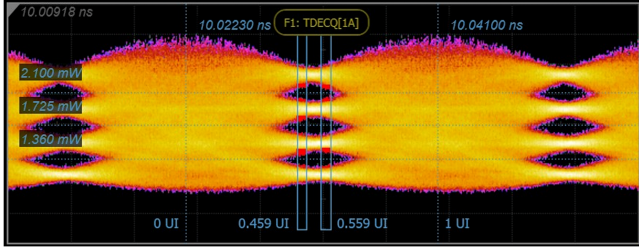
Fig. 2. Experimental TDECQ test on a nonlinearly pre-distorted PAM4 signal at 107.2 Gbps. The eye diagram is obtained after FFE equalization. Vertical blue dotted lines localize the UI, while horizontal blue dotted lines measure the OMA. The SER for the TDECQ test is measured over the two 0.1 UI spaced windows (at 0.459 UI and 0.559 UI, respectively).

experiments (refer to Section V-A). By evaluating the TX signal quality through a comprehensive worst-case emulation of the conditions in which the optical TX will be deployed, the TDECQ test thus provides a performance evaluation from a system-wide perspective. However, its iterative measurement procedure does not allows for a closed-form analytical optimization. Nevertheless, in this article we show that we can specifically improve the optical TX signal with respect to the TDECQ, by implementing nonlinear DPD optimized with a learning architecture modeling the relevant features of this metric.

III. NONLINEAR DIGITAL PRE-DISTORTION FOR IMPROVING VCSEL-MMF OPTICAL TRANSMITTERS

The TDECQ has become a key optical TX performance indicator for net 100 Gbps optical MMF links up to 100 meters, with standard Ethernet requirements specifically set in the Short-Range (SR) window of 850 nm [6]. However, it is quite challenging to achieve such high data rate over SR-MMF optical links using commercial hardware. In the first place, the limited bandwidth of the opto-electronic devices severely affect the TX signal. As an example, Fig. 3(a) shows the eye diagram obtained from an experimental VCSEL-based TX setup (VCSEL bias current set to 8 mA), transmitting in back-to-back (B2B) a PAM4 signal at 107.2 Gbps: the PAM4 eye is fully closed in this scenario. The compensation for such a huge bandwidth limitations at the RX side, using post-equalization, would induce a strong enhancement of the RX noise [8]: this strongly penalizes the signal-to-noise ratio, thus the bit error ratio (BER), and ultimately the TDECQ performance. To circumvent this issue, it is preferable to implement pre-compensation, also known as "pre-emphasis", at the transmitter (TX) side. This can be implemented by utilizing a linear Digital Pre-Distorter (DPD), i.e. a linear equalizer applied through DSP at the TX side.

A linear DPD is able to overcome bandwidth limitations, significantly improving the performance of band-limited optical transmitters. However, it cannot compensate for nonlinear effects caused by the use of a commercial VCSEL: Fig. 3(b) shows the eye-diagram obtained in the same aforementioned setup, using a linearly pre-distorted 107.2 Gbps PAM4 signal. As it can be observed, nonlinear eye-skews (see red-dashed line in Fig. 3(b)) affect the PAM4 signals when the laser is

directly modulated at such high speed with a bias current within the typical range for VCSELs (6–8 mA [4]). The resulting asymmetry in the eye diagram significantly hampers symbol decision, leading to a degradation in TDECQ performance. Fig. 3(c) illustrates how the presence of eye skew during the TDECQ test causes the lower PAM4 eye to appear nearly closed within the left histogram window, resulting in a severe penalty on the measurement of worst-case SER. As a consequence, nonlinear pre-compensation of the VCSEL distortions becomes crucial in order to improve the optical TX and its relative TDECQ performance. To this purpose, a nonlinear Digital pre-distorter can be a suitable DSP solution to jointly compensate for bandwidth limitations and nonlinear effects. In particular, nonlinear DPD working at 1 sample-per-symbol ratio provides significant complexity advantages with respect to nonlinear RX post-equalization, as it can be efficiently implemented at the TX side. The pre-distorted TX symbols can be indeed pre-stored at a factory level within the transmitter (e.g., using Look-Up-Table based structures) [11], without the need of an actual real-time deployment of the equalizer implementing the nonlinear DPD. However, the effective optimization of a nonlinear pre-distorter poses significant challenges compared to linear DPD. The latter can be indeed easily achieved by deploying a post-equalizer in place of the pre-distorter: this procedure is also known as "indirect learning" approach (ILA) [12], [13]. However, the ILA relies on the commutation law, which is mathematically appliable only for linear systems [14]. Recent studies have emphasized the effectiveness of Direct Learning Architectures (DLA) and end-to-end (E2E) learning systems in optimizing nonlinear DPD for optical transmission systems [13], [15], [16], [17]. The DLA is an approach specifically designed for optimizing nonlinear DPD, which has consistently shown to outperform the ILA in both wireless and optical communications [12], [13], [17], [18]. E2E learning on the other hand is an approach which envisions the joint co-optimization of various digital signal processing (DSP) components at the transmitter and receiver, not only for nonlinear DPD but also for other modules in VCSEL-based optical interconnects [19]. In the upcoming Section, we will provide a detailed comparison of these two methodologies: we will highlight that E2E learning not only shares the fundamental principles of DLA for optimizing nonlinear DPDs, but also expands the optimization scope to encompass a system-wide perspective.

IV. NONLINEAR DPD OPTIMIZATION METHODOLOGY

In this Section, we illustrate our novel proposed nonlinear DPD optimization algorithm, which implements a TDECQ-based E2E learning approach using a 1D Convolutional Neural Network (CNN) architecture. Our method is based on emulating the target VCSEL-MMF IM-DD optical TX system using a differentiable digital signal processing (DSP) chain, built through a set of experimental measurements (as described later) and implemented as a CNN by interpreting the linear and nonlinear DSP blocks as CNN layers. For the reader who might not be familiar with the relation between nonlinear digital filters and convolution neural network layers, Appendix A illustrates in
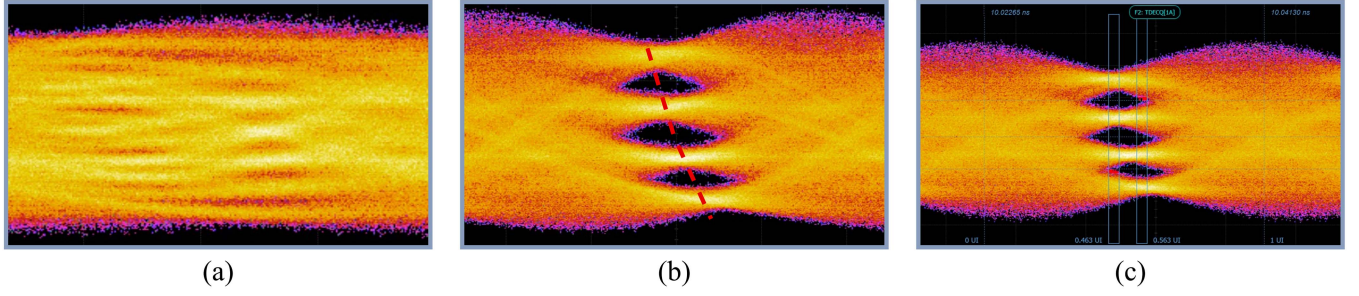
Fig. 3. Experimental RX eye-diagram of a 107.2 Gbps PAM4 signal transmitted over a back-to-back VCSEL-MMF optical IM-DD link: (a) without DPD applied; (b) with linear DPD applied; (c) with linear DPD and RX equalizer applied for TDECQ measurement. VCSEL input bias current is set to 8 mA.
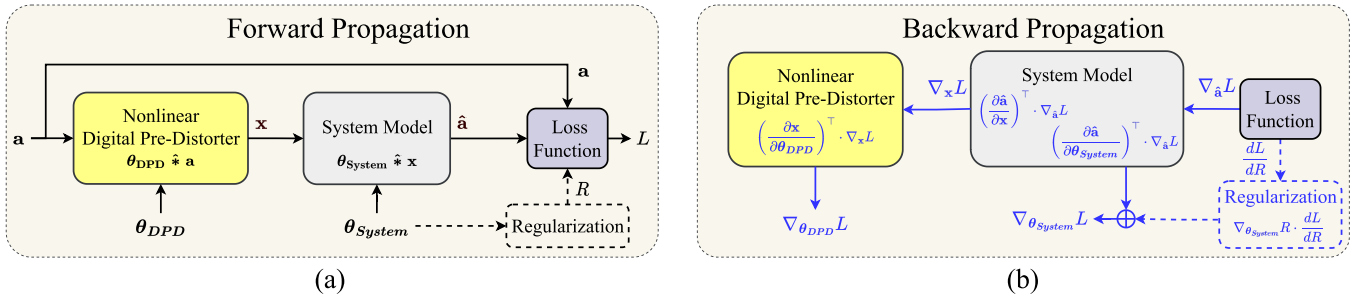

Fig. 4. CNN-based DPD optimization main steps: (a) Forward propagation of the signals; (b) backward propagation of the loss gradients.

detail this strict analogy. We compare the new architecture with an updated version of the DLA we proposed in [16]. Moreover, we illustrate the features of the CNN approach adopted in this work and in [16] compared to the methodology used in our previous works [15], [20], [21]. In the following, we simplify the representation of coefficients (such as weights, bias terms, etc.) that determine the parametric behavior of each illustrated CNN layer (or nonlinear filter) by employing a vectorial notation. To provide an example, we can consider a nonlinear filter, denoted as F, which given the input signal $x[n]$ generates the output $y[n] = (a \cdot x[n-1] + b \cdot x[n-2] \cdot x[n-1]) \cdot c \cdot x[n]^2$: we employ the vector $\boldsymbol{\theta}_F = [a, b, c]$ to refer to the F's coefficients. This notation allows for a more concise and compact representation of the coefficients.

A. CNN-Based Nonlinear DPD Optimization

The proposed nonlinear DPD optimization methodology leverages a baseline scheme shared by both DLA [13] and E2E [22] learning approaches, which is suitable for training any kind of DPD architecture implementable as a differentiable nonlinear filter. The scheme involves emulating the optical transmission of the pre-distorted signals by exploiting an artificial neural network (ANN), built as the cascade of two consecutive layers: one is the nonlinear DPD to be optimized, while the second is a differentiable surrogate model of the system where the pre-distorter will be employed.

The ANN is designed and trained as an *autoencoder* [23]: a learning architecture specifically designed to reconstruct at its output the same signal as provided at its input [24]. This optimization approach enables the *encoder* (in our case, the DPD), to effectively mitigate the distortions introduced within the ANN architecture(i.e. by the surrogate system model). Therefore, the nonlinear DPD is optimized by means of a stochastic gradient descent (SGD) algorithm.

When implementing the autoencoder DPD optimization using the convolutional neural network (CNN) approach that we propose, the training process follows an iterative procedure. This involves transmitting at each iteration a pre-distorted signal through the target communication system, and then to update the DPD parameters according to a given metric in order to improve the TX output. Each iteration involves two main steps, as illustrated in Fig. 4.

In the first step, known as *forward propagation*, the TX signal, defined as the vector $\mathbf{a}$, gets propagated through the CNN autoencoder architecture: first it get pre-distorted by the nonlinear DPD, having coefficients $\boldsymbol{\theta}_{DPD}$, by means of a discrete nonlinear convolution. The pre-distorted signal, defined by the vector $\mathbf{x} = \boldsymbol{\theta}_{DPD} \hat{*} \mathbf{a}$, is then transmitted over the system model, having parameters $\boldsymbol{\theta}_{System}$. The resulting output signal, defined by the vector $\hat{\mathbf{a}} = \boldsymbol{\theta}_{System}, \hat{*}, \mathbf{x}$, is subsequently compared to the input signal $\mathbf{a}$. This comparison is carried out using a loss function, which can be for instance the mean square error (MSE) between the two signal vectors.

In the second step, known as *backward propagation*, the gradients with respect to the previously calculated loss, denoted as L, are computed in order to perform the SGD update of the CNN coefficients. The computation is done recursively, by exploiting the chain rule of calculus [24]. Fig. 4(b) illustrates the
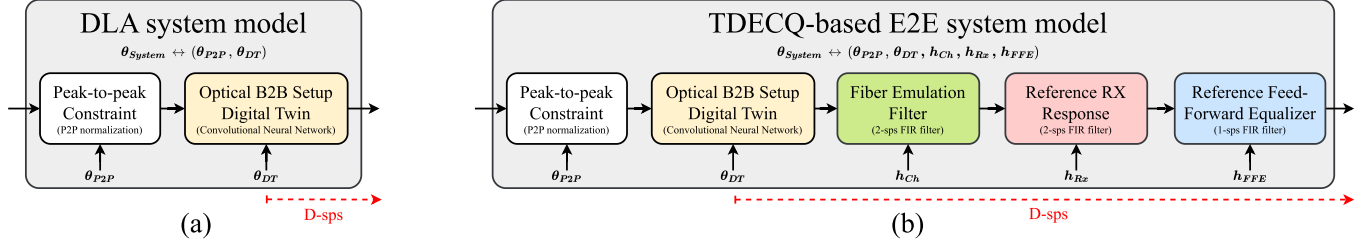
Fig. 5. System model architectures for nonlinear DPD optimization: (a) system model for the direct learning architecture; (b) end-to-end learning system model based on TDECQ specifications [6]. Red-dashed lines indicate a sps ratio D greater than 1 when discrete signals are propagated through the system, starting from the digital twin. FIR filters with a sps ratio $K < D$ are implemented accordingly using dilated $Conv1d()$ layers with dilation factor $d = D/K$.

back-propagation process: this involves computing and accumulating by means of multiplication the input-to-output gradients of all the autoencoder layers from the loss to the CNN input parameters. As a result, the DPD can be updated through SGD by exploiting the loss gradient with respect to its coefficients $\nabla_{\boldsymbol{\theta}_{\mathrm{DPD}}} L$, computed as follows:

$$\nabla_{\boldsymbol{\theta}_{\mathrm{DPD}}} L = \left(\frac{\partial \mathbf{x}}{\partial \boldsymbol{\theta}_{DPD}}\right)^{\top} \cdot \left(\frac{\partial \hat{\mathbf{a}}}{\partial \mathbf{x}}\right)^{\top} \cdot \nabla_{\hat{\mathbf{a}}} L \qquad (3)$$

where $\frac{\partial \hat{\mathbf{a}}}{\partial \mathbf{x}}$ and $\frac{\partial \mathbf{x}}{\partial \boldsymbol{\theta}_{\mathrm{DPD}}}$ are the *system's output-w.r.t.-input* and *DPD's output-w.r.t.-parameters* Jacobian matrices, respectively. The gradient loss with respect to the system model parameters $\nabla_{\boldsymbol{System}} L$ instead is computed as follows:

$$\nabla_{\boldsymbol{\theta}_{\mathrm{System}}} L = \left(\frac{\partial \hat{\mathbf{a}}}{\partial \boldsymbol{\theta}_{System}}\right)^{\top} \cdot \nabla_{\hat{\mathbf{a}}} L + \nabla_{\boldsymbol{\theta}_{System}} R \cdot \frac{dL}{dR}$$
$$(4)$$

where $\frac{\partial \hat{\mathbf{a}}}{\partial \boldsymbol{\theta}_{System}}$ is the *system's output-w.r.t.-parameters* Jacobian matrix, $\frac{dL}{dR}$ is the (scalar) derivative of L with respect to the regularization term in the loss function (see Fig. 4) and $\nabla_{\boldsymbol{\theta}_{System}} R$ is the regularization term gradient with respect to the system parameters. In Equation (4), the $\nabla_{\boldsymbol{\theta}_{System}} R \cdot \frac{dL}{dR}$ term is computed in case the loss function regularization term is depending on the system model coefficients (as for the E2E loss depicted in Section IV-C).

Once the gradient computation has been completed, the iteration concludes with a SGD update of the DPD coefficients, as follows:

$$\boldsymbol{\theta}_{DPD} \leftarrow \boldsymbol{\theta}_{DPD} - \epsilon \cdot \nabla_{\boldsymbol{\theta}_{\mathrm{DPD}}} L \qquad (5)$$

where ϵ is the *learning rate* (or *step size*). While in the case of a DLA approach only the DPD coefficients are updated, keeping fixed the other parameters of the autoencoder, in the E2E learning approach also some of the system model coefficients are subject to the SGD update, i.e. those relative to the *decoder* (which in our case is a linear RX equalizer, as explained below).

B. DLA and TDECQ-Based E2E Optimization Architectures

Despite the common baseline optimization scheme, the DLA and the E2E learning approach differ significantly in the way the system model is conceived.

In the *direct learning architecture (DLA)*, the system model mainly consists of a digital twin of the optical transmitter, whose

impairments have to be compensated by the DPD [13]. In Section V-B, we illustrate the procedure for retrieving a CNN digital twin (DT) of an experimental VCSEL-based TX from optical back-to-back (B2B) setup. Additionally, in order to account for the limitations of the transmitter's hardware [23], it becomes necessary to incorporate an initial layer into the DLA system model, which enforces constraints on the pre-distorted digital signal. In our specific scenario, we thus introduce a maximum peak-to-peak (P2P) amplitude constraint at the output of the DPD, considering both the limited dynamic range of the Digital-to-Analog Converter (DAC) and the restricted input dynamic range of the VCSEL [15]. Fig. 5(a) presents the block scheme of the adopted DLA system model architecture in this article. It comprises two main components: a digital twin of an optical B2B setup (which includes the optical TX where to apply DPD) and a peak-to-peak (P2P) normalization layer. For simplicity, the introduction of a resample layer at the model's input is not depicted, although it has be incorporated in case of sampling rate discrepancy between the DPD and the DAC (embedded in the digital twin) [15]. The digital twin is implemented as a convolutional neural network (see Section V-B for details), with coefficients defined by the vector $\boldsymbol{\theta}_{DT}$. The P2P constraint instead is implemented by the $\mathrm{P2P}_{\mathrm{norm}}()$ layer function, defined as follows:

$$\mathrm{P2P}_{\mathrm{norm}}(\mathbf{x}) = \frac{\mathbf{x} - min(\mathbf{x})}{max(\mathbf{x}) - min(\mathbf{x})} \left(x_{new}^{max} - x_{new}^{min}\right) + x_{new}^{min}$$
$$(6)$$

where x_{new}^{min} and x_{new}^{max}, represented in Fig. 5 by the vector $\boldsymbol{\theta}_{P2P} = [x_{new}^{min}, x_{new}^{max}]$, define the desired maximum and minimum value of the pre-distorted vector signal vector $\mathbf{x}$ at the input of the digital twin. During the DPD optimization, the DLA parameters $\boldsymbol{\theta}_{System} = [\boldsymbol{\theta}_{P2P}, \boldsymbol{\theta}_{DT}]$ are not updated: the DPD is thus trained to to encode the TX signal in order to mitigate the DT impairments, yet fulfilling the P2P constraints. While the DLA focused solely on optimizing the transmitter impairments, the *End-to-end (E2E) learning architecture* takes a more comprehensive approach, by implementing an autoencoder which encompasses the entire chain of transmitter, channel, and receiver [22]. Specifically, the E2E learning approach consists of the joint training of a TX *encoder* (in our case, the DPD) together with a RX *decoder* (in our case, the linear RX post-equalizer), in order to compensate for the channel distortions [23]. In the considered scenario, these can incorporate the impairments

caused by several components along the link, i.e. the optical transmitter (i.e. the VCSEL), the multi-mode fiber channel and the noise introduced at the RX side. As a consequence, the E2E learning appears to be the most suitable approach to optimize TX metrics such as the TDECQ using nonlinear DPD: in fact, it is possible to implement an E2E autoencoder compliant to the TDECQ conformance test (see Fig. 1), by designing a proper surrogate model of the system. Fig. 5(b) illustrates our proposed implementation of a TDECQ-based E2E system model. The architecture extends the system model adopted for the DLA (see Fig. 5(a)), by incorporating the DSP blocks required for the TDECQ measurement, according to the IEEE specifications for net 100 Gbps/λ transmission over MMF [6]: in addition to the P2P normalization layer and the digital twin, three further CNN layers are thus introduced:

1) a FIR filter with 31 discrete tap coefficients (also named *taps*) at 2-sps ratio, having the equivalent response of a fourth-order Bessel-Thomson filter with 3-dB bandwidth equal to 18 GHz [10]. This filter models the effects of a WCS optical channel: in the considered scenario, the filter emulates the modal and chromatic dispersion of a MMF with length up to 100 meters [6].

2) a 31 taps 2-sps FIR filter having the equivalent response of a fourth-order Bessel-Thomson filter with 3-dB bandwidth equal to half of the TX Baud Rate [10]. This filter emulates the response of a reference RX filter for PAM4 signals.

3) a 9 taps 1-sps Feed-forward Equalizer (FFE), which is the TDECQ reference RX equalizer specified by the IEEE standard [6].

These three components, designed as FIR filters, are implemented in the E2E system model as three discrete linear convolution ($Conv1d()$[2]) layers, having a *dilation* factor [25] set in order to implement the loss function detailed in Section IV-C. Specifically, the discrete sequence at output of the system model must have a sps ratio D greater than 1. Consequently, FIR filters designed with an sps ratio of $K < D$ are implemented using $Conv1d()$ layers with a dilation factor of $d = D/K$. For instance, the 1-sps FFE (refer to Fig. 5(b)) is mapped into a $Conv1d()$ layer with dilation factor set to D: from a DSP perspective, this is equivalent to introducing $D - 1$ zeros between consecutive samples in the FFE impulse response $\mathbf{h_{FFE}}$. This allows a faithful simulation of the filtering effects expected by the TDECQ test, while preserving an high sps ratio on the propagated TX signal (as required by the loss function detailed below).

Using the TDECQ-based E2E system model, during the DPD optimization also the parameters $\mathbf{h_{FFE}}$ of the FFE are updated (i.e. the FIR filter tap coefficients). The parameters of the RX filter ($\mathbf{h_{Rx}}$) and of the WCS optical channel ($\mathbf{h_{Ch}}$) are instead kept fixed, as well as those of the P2P normalization layer ($\boldsymbol{\theta_{P2P}}$) and of the digital twin ($\boldsymbol{\theta_{DT}}$). We use $\mathbf{h}$ in place of $\boldsymbol{\theta}$ to indicate that the coefficients of the introduced linear convolution layers, since these also represent the discrete impulse response

of the FIR filter. Using the proposed system model, the E2E architecture does not directly introduce the impairments of the RX noise: its effect is rather introduced analytically as an additive regularization term in the loss function [15], as explained in the following Subsection.

C. DPD Optimization Loss Function

As illustrated at the beginning of this this Section, using either a DLA or an E2E learning approach, the nonlinear DPD is trained by minimizing a given loss function according to a SGD optimization algorithm. However, as discussed in [15], a simple mean square error (MSE) is inadequate for optimizing the pre-distorter effectively. This inadequacy arises due to the intrinsic nonlinearities within the considered communication system, which encompass not only the VCSEL nonlinear distortions but also the P2P constraint at the system input: this can lead to an unbalanced compensation of the TX symbols, e.g. penalizing the outer PAM4 levels with respect to the internal ones [15]. To better address these challenges, we thus propose an novel loss function, suitable for both DLA and E2E approaches, which focuses the DPD optimization on improving the correct RX symbols detection rather than solely minimizing the MSE between the autoencoder input and output. The computational scheme of the proposed loss function is illustrated in Fig. 6(a). The loss function assumes the utilization of a *multi-rate* CNN system model, i.e. given in input the 1-sps pre-distorted sequence, it musts output a sequence with a sps ratio $D > 1$. The design of this loss function draws inspiration from the Error Vector Magnitude (EVM), a widely-used performance metric in the assessment of advanced modulation formats [26]. In particular, it evaluates the normalized MSE between the input (TX) symbol sequence $\mathbf{a}$ and the output (RX) symbol sequence in the CNN-based DPD optimization architecture, considering the optimal decimation instant of the D-sps sequence $\hat{\mathbf{a}}$ (refer to Fig. 4). The computational steps involved by the loss are the following:

1) The sequence $\hat{\mathbf{a}}$ is downsampled to a 1-sps ratio for all possible D symbol decision instants, resulting in decimated sequences $\hat{\mathbf{a}}_1, \hat{\mathbf{a}}_2,..., \hat{\mathbf{a}}_D$.

2) The TX symbol sequence $\mathbf{a}$ and the decimated RX sequences $\hat{\mathbf{a}}_1, \hat{\mathbf{a}}_2,..., \hat{\mathbf{a}}_D$ are normalized to have zero mean and unit variance.

3) The MSE is computed between the TX and RX sequences for all the decimation instants.

4) The final loss L is determined as the optimal value (i.e. the minimum) among the losses $L_1, L_2,..., L_D$.

This procedure enables the DPD optimization to specifically improve the symbols decoding at the optimal decision instant: the latter indeed needs to be found numerically (i.e. looking for the best MSE), due to the time-domain eye-skew caused by the VCSEL (refer to Fig. 3(b)). Additionally, the normalization of the symbols focuses the DPD to optimize the relative positions of the PAM4 constellation points rather than tracking specific absolute values. This prevents excessive amplification of the TX PAM4 levels by the DPD, which, in combination with the peak-to-peak (P2P) constraint, could penalize the quality of the external PAM4 symbols [15]. Furthermore, to analytically

---

[2]We use in this article the notation of the *PyTorch* library to refer to the conventional CNN layers.
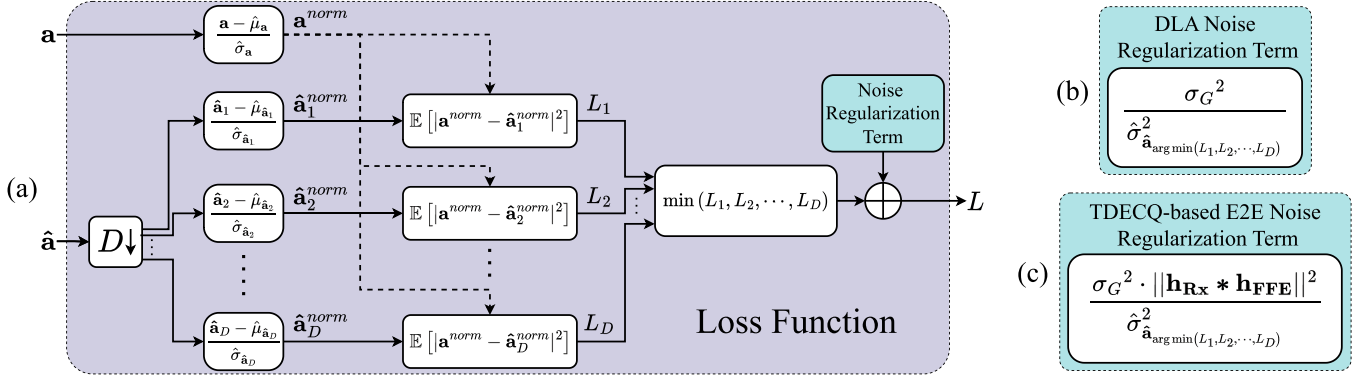
Fig. 6. (a) Custom Loss function for optimizing nonlinear DPD using the CNN-based approach. The noise regularization term depend whether the system architecture is (b) adopting the DLA system model (refer to Fig. 5(a)); (c) adopting the TDECQ-based E2E system model (refer to Fig. 5(b)).

account for the presence of RX noise that can impair the TX signal, a noise regularization term is incorporated into the loss, similarly to what done in [15]. In this work, the regularization term represents the equivalent contribution to the loss function of an additive white Gaussian noise sequence having zero mean and variance σ_G^2, combined to the output signal of the digital twin.

When adopting the DLA system model, the noise regularization term R_{DLA} (depicted in Figure 6(b)) is defined as:

$$R_{DLA} = \frac{\sigma_G^2}{\hat{\sigma}^2_{\hat{\mathbf{a}}_{\arg\min(L_1, L_2, \ldots, L_D)}}} \quad (7)$$

where $\hat{\sigma}^2_{\hat{\mathbf{a}}_{\arg\min(L_1, L_2, \ldots, L_D)}}$ represents the variance used to normalize the optimal decimated RX sequence, i.e. the one that yields the minimum MSE (as shown in Fig. 6(a)).

In the case instead of the TDECQ-based E2E system, the noise regularization term R_{E2E} (depicted in Fig. 6(c)) also models the propagation of the noise through the reference RX response filter and the following equalizer, as follows:

$$R_{E2E} = \frac{\sigma_G^2 \cdot ||\mathbf{h_{Rx}} * \mathbf{h_{FFE}}||^2}{\hat{\sigma}^2_{\hat{\mathbf{a}}_{\arg\min(L_1, L_2, \ldots, L_D)}}} \quad (8)$$

where $||\mathbf{h_{Rx}} * \mathbf{h_{FFE}}||$ is the Euclidean norm of the discrete linear convolution between the impulse response of reference RX filter $\mathbf{h_{Rx}}$ and the impulse response of the feed-forward equalizer $\mathbf{h_{FFE}}$. The use of the noise regularization term provides several benefits to the nonlinear DPD optimization. In both the DLA and E2E cases, it induces the maximization of the $\hat{\sigma}^2_{\hat{\mathbf{a}}_{\arg\min(L_1, L_2, \ldots, L_D)}}$ term, thus stimulating the DPD to amplify the power of the TX signal as much as possible while adhering to the P2P constraints. This prevents the DPD from driving the VCSEL with an excessively low optical modulation amplitude (OMA), as this parameters must fulfill minimum requirements [6]. Moreover, in the E2E case, it prevents the RX noise enhancement induced by the FFE [8], by inducing the minimization of the $||\mathbf{h_{Rx}} * \mathbf{h_{FFE}}||^2$ term, which is equivalent to the C_{eq} term in the TDECQ formula (refer to (1)). Consequently, the DPD optimization considers the fact that the RX equalizer can only partially compensate for bandwidth

limitations. Moreover, the proposed E2E regularization term enables analytical consideration of the WGN noise introduced at the RX input as specified by the reference TDECQ measurement scheme [6].

The proposed loss function therefore tackles multiple challenges in optimizing nonlinear DPDs. Combined with an appropriate initialization of the DPD coefficients (refer to Section V), it has demonstrated significant success in training different nonlinear DPD architectures in the considered scenario. The outcomes of employing this approach will be illustrated in Section V-D.

D. CNN-Based Versus FIRNN-Based Optimization

In this section, we have presented a novel CNN-based approach for nonlinear DPD optimization, which represents a significant departure from the methodology proposed in our previous works [15], [20], [21] based on finite impulse response neural networks (FIRNN). Despite this paradigm shift, the underlying system emulation remains unchanged from a DSP perspective. In fact, he parametric structure of a CNN layer can be effectively mapped to a FIRNN layer, ensuring equivalent functionality. In terms of deep learning terminology [24], a 1D convolutional ($Conv1d()$) layer with a kernel size of k, an input channel size of ch_{in}, and an output channel size of ch_{out} performs the same filtering operation on ch_{in}-channel signals as a FIRNN layer with a FIR memory of k, ch_{in} input neurons, and ch_{out} output neurons. However, the major distinction lies in the approach to forward and backward signal and gradient propagation. FIRNNs implement them simultaneously throughout the architecture, emulating real-time digital systems with continuous adaptation, such as practical RX post-equalizers [15]. On the other hand, CNNs perform signal propagation block-by-block through the layers, similar to an offline DSP chain. As a result, the derivatives for stochastic gradient descent update (SGD) are computed only after the entire signal has propagated through the system.

The transition from the FIRNN approach to the CNN approach thus unlocks additional features for nonlinear DPD optimization. Firstly, the CNN-based DPD learning architecture

can be implemented using well-established and advanced Deep Learning software libraries such as *TensorFlow* and *PyTorch*. These frameworks provide sophisticated and optimized implementations of convolutional layers, supporting their *automatic differentiation* [27]. This significantly simplifies code development and facilitates future upgrades. In contrast, automatic differentiation for online optimization algorithms like in FIRNN is currently not implemented by these libraries: consequently, FIRNN requires manual software development.

Moreover, the CNN approach offers the advantage of computing *differentiable signal statistics* by virtue of its block-by-block processing procedure. By leveraging the vector signals generated at the output of various DSP blocks within the CNN autoencoder, it becomes feasible to estimate multiple statistics of the propagated signal at each stage of the emulated communication system: these can be mean values or variances, as well as maximum and minimum values. This capability enables the use of differentiable signal normalizations, which can facilitate optimization convergence [28], as demonstrated in the case of symbol normalization. Signal normalizations also enable the application of differentiable constraints on signal dynamics, such as the P2P normalization. Unlike the previously adopted P2P hard-limiting function [15], [16], [20], [21], this new implementation of the constraint avoids the vanishing gradient issues caused by the hard-limiter clipping of the DPD output. Furthermore, signals statistic estimation allows for better tuning of the optimization loss function during the training, such as the MSE-based selection optimum decimation instant (refer to Section IV-C).

Therefore, by adopting the CNN-based approach, both DLA and E2E approaches benefit from a more comprehensive emulation of the TX system behavior, leading to a refined DPD optimization. Experimental results have demonstrated its effectiveness in several TX conditions, as it will be illustrated in the upcoming Sections.

V. EXPERIMENTAL OPTIMIZATION OF NONLINEAR DPD OVER A VCSEL-MMF OPTICAL LINK

In this Section, we illustrate the experimental implementation of methodology depicted in the previous Section on a real VCSEL-MMF optical transmission link. We first describe the experimental modeling of an optical B2B setup using a CNN digital twin. We then show the experimental procedure for deploying the CNN-based nonlinear DPD optimization. Finally, we compare and discuss the optical TX performance of both DLA and TDECQ-based E2E approaches.

A. Adopted VCSEL-MMF Experimental Setup

For our experiments, we utilized an optical back-to-back (B2B) setup that adheres to the specifications provided for measuring the TDECQ over 100 Gbps/λ SR-MMF links [6]. The experimental setup is depicted in Fig. 7(a), and it comprises the following components:
- An Arbitrary Waveform Generator (AWG) with a sampling rate of $f_{DAC} = 107.2$ GSa/s and a bandwidth of 50 GHz.
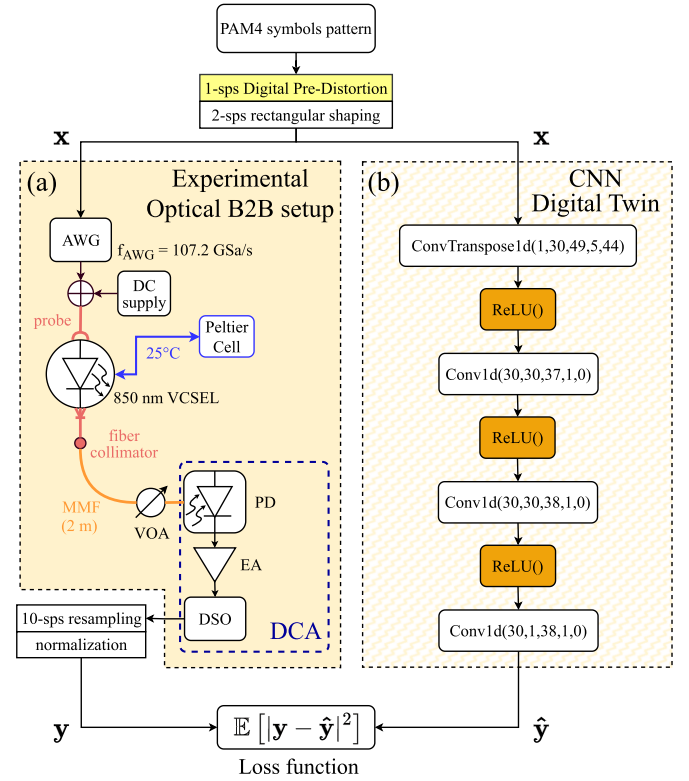

Fig. 7. Overall Schematics for the CNN experimental modeling of an optical TX setup. (a) Experimental setup for the DPD optimization and performance evaluation; (b) CNN Digital Twin of the experimental setup.

- A Direct Current (DC) generator for providing the VCSEL bias current.
- An 850 nm VCSEL with a bandwidth of 22 GHz, serving as the main component of the optical transmitter to be pre-distorted.
- A Peltier cell attached to the VCSEL to maintain its temperature at 25 °C.
- A 2-meter OM4 multi-mode fiber patch cord used to propagate the VCSEL's collimated optical waveform.
- A MMF Variable Optical Attenuator (VOA) for adjusting the optical power of the received signal.
- A Keysight DCA-M N1092 A Digital Communication Analyzer, which consists of a PIN Photo-Diode (PD), an Electrical Amplifier (EA), and a Digital Sampling Oscilloscope (DSO). The device is equipped with the official software for TDECQ measurement, compliant with the standards described in [6].

We employed this setup to characterize the digital twin for the DPD optimization and subsequently evaluate the performance of the VCSEL-based optical TX using DPD. To ensure instrumental compatibility, all experiments were conducted using PAM4 sequences transmitted at a Symbol Rate of 53.6 GBaud, which aligns with the sampling rate of the AWG by up-sampling the 1-sps pre-distorted sequences through rectangular shaping (i.e. repeating for 2 consecutive times each TX symbol). The resulting bit rate ($R_b = 107.2$ Gbps) slightly surpasses the standard gross rate required for achieving a net 100 Gbps PAM4

transmission (106.25 Gbps, as stated in [6]). However, this approach allows for a conservative assessment of the optical TX performance, being on the safe side since adopting an higher bit rate than required (as done in [16]).

B. Digital Twin Modeling of the Experimental Setup

The optimization of nonlinear DPDs using either DLA or E2E approaches necessitates a surrogate model of the optical TX setup, as we mentioned in Section IV. Specifically, in the context of a TDECQ-based optimization, the digital twin primarily needs to model the response of the VCSEL-based optical TX. This entails the adoption of a B2B setup, aimed at mitigating the propagation effects stemming from the MMF link. These effects can be later emulated through DSP using a WCS fiber channel, consistently with the TDECQ test's principle (see Section II). In this work, we implement the digital twin using a CNN whose structure is depicted in Fig. 7(b). This is composed of multiple $Conv1d(ch_{in}, ch_{out}, k_{size}, s, p)$ layers. Here, ch_{in} denotes the input channel size, ch_{out} represents the output channel size, k_{size} denotes the kernel size, s is the stride used for correlation (i.e. the downsampling ratio of the input signal), and p indicates the padding applied to the input signal. The nonlinearity of the CNN is implemented by alternating the convolution layers with $ReLU()$ activation functions. Furthermore, the first layer is implemented as *1D transposed convolution layer*, known as *ConvTranspose1d*($ch_{in}, ch_{out}, k_{size}, s, p$) [29] with stride $s > 1$, to up-sample the input signal at the desired output sps ratio.[3] The purpose of the digital twin is to accurately predict the output, denoted as $\mathbf{y}$, of a given experimental setup when provided with the corresponding input TX signal, $\mathbf{x}$. To achieve this, an iterative SGD procedure [24] is employed to optimize the neural network-based model. To ensure an integer output samples-per-symbol (sps) ratio suitable for the proposed loss function (see Section IV-C), the acquired experimental output $\mathbf{y}$ is resampled to a sps rate of $D = 10$. Additionally, the experimental signal is normalized to have a zero mean and unit variance, as illustrated in Fig. 7, in order to facilitate model convergence. During each training iteration, a random subset[4] of $\mathbf{x}$ consecutive samples is fed into the CNN. Then, the MSE is computed by comparing the CNN output signal vector, denoted as $\hat{\mathbf{y}}$, with the related subset of the experimental output $\mathbf{y}$. Finally, the CNN coefficients are updated through SGD by computing the MSE loss gradients.

The CNN digital twin is obtained by using two different experimental acquisitions. The first one is employed as *training dataset* to optimize the model. It involves transmitting a linearly pre-distorted PRBS sequence of 2^{16} symbols (obtained using ILA, as explained later) over the transmission setup. Utilizing a pre-emphasized sequence rather than a standard plain PAM4 signal emerges as a favorable choice, as its symbols levels distribution aligns more effectively with the distribution of the target nonlinearly pre-distorted sequence. Consequently, the digital twin trained using a linearly pre-distorted sequence is

able to accurately model the nonlinear effects of pre-distorted VCSEL driving conditions. The second acquisition, defined as *test dataset*, is instead used to assess the accuracy of the trained digital twin and is obtained by transmitting a nonlinearly pre-distorted PAM4 sequence, i.e. the same used to assess also the nonlinear DPD performance (details in Section V-D). This allows to validate the reliability of the digital twin in emulating the optical TX setup at the critical moment when the nonlinear DPD optimization has ended, and the pre-distorter will be actually deployed in the real scenario.

The CNN model adopted in our experiments involved a $convTranspose1d()$ layer followed by three $ReLU()$ and $Conv1d()$ layers. The detailed layer parameters utilized in the digital twin for these experiments can be found in Fig. 7. Each hidden convolution layer was configured to have 30 input and output channels, while the kernel sizes were chosen to ensure a minimum *memory*[5] equivalent to 15 symbols [25]. The model was updated through SGD for 700 iterations using minibatches o 1000 consecutive TX symbols. With the chosen hyperparameters, the CNN digital twin achieved on the test dataset an optimal normalized[6] Mean Squared Error (MSE) performance of -24 dB in each VCSEL driving condition. Further improvement in the MSE performance was not observed by increasing either the memory (i.e. the CNN layers kernel sizes) or the number of hidden channels.

C. Experimental Nonlinear DPD Optimization

The effectiveness of the nonlinear DPD optimization methodology, discussed in Section IV, was evaluated using the experimental setup illustrated in Fig. 7(a). To ensure its efficacy across different DPD structures, two distinct nonlinear DPD architectures were deployed at a 1 sps ratio: a CNN DPD implementation (with similar structure to the CNN digital twin) and a discrete-time Volterra-series nonlinear equalizer (VNLE) DPD implementation [11], [30]. The VNLE equalization can be defined as the filtering operation on the input signal $x[n]$ to produce the output signal $y[n]$, as follows:

$$y[n] = h_0 + \sum_{k=1}^{K} \sum_{t_1=0}^{M_k} \ldots \sum_{t_k=0}^{M_k} h_k(t_1, \ldots t_k) \prod_{j=1}^{k} x$$
$$\times \left[n - t_j - \frac{M_1 - M_k}{2} \right] \qquad (9)$$

where $h_k(t_1, \ldots t_k)$ represents the k-th order Volterra kernels, and h_0 denotes the bias term. The memory lengths for the linear ($k = 1$) and nonlinear terms are denoted as M_1 and M_2 to M_k, respectively. The term $\frac{M_1 - M_k}{2}$ accounts for the difference in the memory lengths between the linear kernel terms (corresponding to the taps of a FIR filter) and the nonlinear kernel terms, assuming M_1 to be the largest memory and $\lfloor M_1/2 \rfloor$ to be the delay in symbols introduced by the VNLE. The schematics of

---

[3]It is equivalent to applying on a given input signal a baseline $Conv1d()$ layer after the addition of $s - 1$ zeros between the consecutive input samples.

[4]This approach mimics the use of conventional *mini-batches* [24].

[5]The *memory* is the minimum number of consecutive symbols to give in input to a nonlinear FIR filter to produce an output sample (see Appendix A).

[6]The MSE is normalized with respect to the power of the experimental output $\mathbf{y}$ in the test dataset, defined by its variance $\sigma_{\mathbf{y}}^2$.
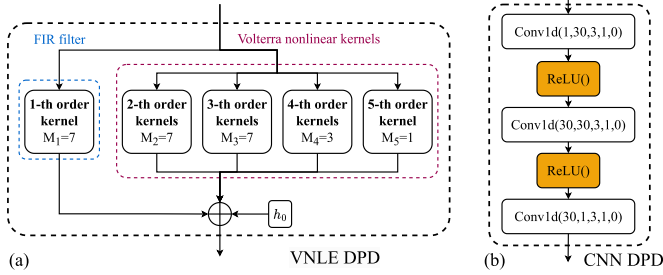
Fig. 8. Nonlinead DPDs deployed in the experiments (a) Volterra-series nonlinear equalizer; (b) convolutional neural network.

the VNLE and CNN DPDs adopted during the experiments are depicted in Fig. 8.

The CNN DPD involves 3 Conv1d() layers, alternated by $ReLU()$ activation functions, with each hidden convolution layer having 30 input and output channels (similarly to the CNN digital twin). Each convolution layer was set to have a kernel size $k_{size} = 3$, in order to have a total memory of the DPD equal to 7 symbols. The VNLE DPD also adopts a total memory length of 7 symbols, having a linear memory length $M_1 = 7$, and integrates nonlinear Volterra kernels up to the 5-th order, with $M_2 = 7$, $M_3 = 7$, $M_4 = 3$ and $M_5 = 1$. Both DPD architectures were optimized using the CNN-based Direct Learning Architecture (DLA) and End-to-End (E2E) approaches discussed in this work. The experimental nonlinear DPD optimization procedure adopted for each considered VCSEL driving condition involved the following steps:

1) *Initial characterization:* A first characterization of the optical transmitter is achieved by transmitting a plain PRBS PAM4 signal of 2^{16} symbols without applying DPD. The sequence is averaged through the PRBS period by the DCA, in order to minimize the noise impairments [15].

2) *Linear DPD estimation:* The acquired non-pre-distorted experimental data is exploited to estimate a Linear 1-sample per symbol (1-sps) FIR DPD using an indirect learning approach (ILA) [12]: this involves deploying as linear DPD a 1-sps feed-forward equalizer optimized using a Least Mean Square (LMS) algorithm. The linear DPD was set to have the same memory length as for the nonlinear DPD to be optimized: specifically, the FIR number of tap coefficients was set to 7 symbols.

3) *Training dataset acquisition:* To train the digital twin model, the PRBS PAM4 signal, now pre-distorted by the linear DPD is again transmitted over the experimental setup, to be then acquired and denoised as in step 1;

4) *CNN digital twin optimization:* Using the training dataset, the CNN digital twin is optimized to model the behavior of the optical transmitter, as illustrated in Section V-B.

5) *Nonlinear DPD initialization:* Before optimizing the nonlinear DPD, the latter is pre-initialized using the response of the linear DPD as a starting point. For the VNLE DPD, all nonlinear parameters beyond the first order were set to zero, while the linear Finite Impulse Response (FIR) coefficients (first-order kernel) were set equal to those of

the linear DPD. In the case of the CNN DPD, we trained the model to identify the linear DPD response by minimizing the Mean Squared Error (MSE) between its pre-distorted output and the output of the linear pre-distorter, using the Stochastic Gradient Descent (SGD) optimization.[7]

6) *Nonlinear DPD optimization:* As final step, the nonlinear DPD optimization is done by using either the Direct Learning Architecture (DLA) or the End-to-End (E2E) learning approaches, adopting the CNN-based system models and loss functions explained in Section IV.

The nonlinear DPD optimization using both DLA and E2E optimization approaches involved utilizing the same set of training *hyperparameters* [24]. This choice was made to ensure a fair comparison between the two techniques, employing a conservative approach to guarantee optimization convergence in both cases. Specifically, the number of SGD training iterations (refer to Section IV-A) was set to 2000 times. For each iteration, 3000 randomly PAM4 symbols were generated to form the TX signal vector **a**. In both the DLA and E2E DPD optimization architectures, the P2P normalization layer was set as $\boldsymbol{\theta_{P2P}} = [-1, 1]$, to align with the AWG P2P normalized bounds[8] on the input discrete sequences. Moreover, a 2-sps rectangular upsampling was integrated as a resampling layer at the beginning of the system model. This integration ensured consistent matching of the DPD and DAC sampling rates with the real TX system (as explained in Section IV-B). SGD updates were implemented using an Adam optimizer [24] with a learning rate of 0.001. Regarding the regularization term, the noise power was set as $\sigma_G{}^2 = 0.001$, which was approximately 30 dB lower than the power of the signal at the digital twin output during the DPD optimization. To avoid a significant increase in computational complexity, a numerical $\sigma_G{}^2$ search emulating the standard TDECQ measurement [10] was not pursued. However, our tests showed that using $\sigma_G{}^2$ values within the range of $[0.0005, 0.005]$ yielded similar TDECQ performance results.

D. Experimental Results

We report in this Subsection the experimental performance of the nonlinear DPDs optimized using the experimental procedure illustrated above: we compare the performance of both the CNN and VNLE DPD architectures (see Fig. 8), trained using either the DLA and TDECQ-based E2E approaches for 1-sps digital pre-distortion of PAM4 symbols at 107.2 Gbps. We also include in the comparison the results achieved using linear DPD obtained using the ILA approach (i.e. the linear pre-distorter obtained during the second step in the nonlinear DPD optimization procedure). The optimization and validation of the DPD were carried out by transmitting the signal under various driving conditions, which allowed to assess the performance in relation to the intensity of the nonlinear distortions induced

---

[7]This step exploits the same SGD-based optimization adopted for training the CNN digital twin (refer to Section V-B). In this case, the input signal **x** is replaced by the plain PRBS PAM4 signal (used in Step 1) and the output signal **y** is replaced by the PAM4 sequence after linear DPD.

[8]The $[-1, 1]$ P2P dynamics is normalized with respect to the AWG P2P modulation voltage.
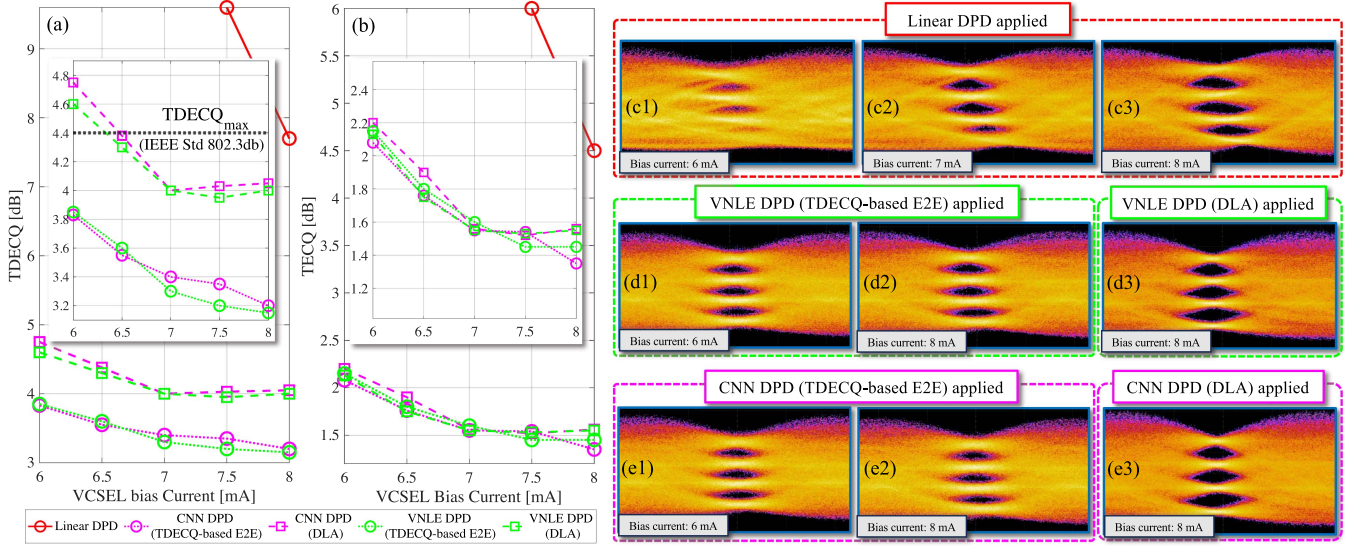
Fig. 9. Left - Experimental performance comparison of different PAM4 DPDs working at 107.2 Gbps, as a function of the VCSEL input bias current: (a) TDECQ test; (b) TECQ test. Inset pictures in (a, b) zoom in specifically on the nonlinear DPD performances. Right - Experimental 107.2 Gbps PAM4 signal eye-diagram acquired by the DSO: (c) with linear DPD applied; (d) with VNLE DPD applied; (e) with CNN DPD applied. DPDs are obtained (c1, c2, c3) using ILA; (d1, d2, e1, e2) using TDECQ-based E2E learning; (d3, e3) using DLA.

by the VCSEL. Specifically, we employed different input bias currents for the VCSEL and set different peak-to-peak modulation voltage swings on the AWG. The experimental validation measurements were conducted using a PAM4 signal that differed from the PRBS16 sequence used in the training dataset. Instead, the standard PAM4 Short Stress Pattern Random Quaternary (SSPRQ) was employed, which is the designated sequence for TDECQ measurements according to IEEE specifications for net 100 Gbps PAM4 transmission [6]. For each considered case, we evaluate the performance of the pre-distorted VCSEL-based TX using standard TDECQ and the TECQ test according specifications for SR-MMF optical TX at net 100 Gbps/λ [6]. The TECQ (Transmission Eye-Closure for PAM4) is an optical TX compliance metric alternative to the TDECQ, which assesses the same performance criteria without considering the effects of the dispersion introduced by the fiber (i.e. omitting the fiber emulation filter in Fig. 1) [6]. Both TDECQ and TECQ were measured using the official Keysight software integrated by the DCA. Fig. 9 illustrates the experimental results when driving the VCSEL with the AWG P2P modulation voltage set at 500 mV, as a function of the VCSEL input bias current. The bias current was varied within the typical range of 6–8 mA for 850 nm VCSELs [4]. The average optical power at the input of the DCA receiver was set to 2 dBm. Fig. 9(a) and (b) depict the TDECQ and TECQ performance, while Fig. 9(c)–(e) show the relative experimental eye-diagrams acquired by the the DSO. Experimental results without applying any DPD are not reported due to the impossibility in all the situations to met the target BER = 2.4e-4 value for TDECQ/TECQ standard test [6]: in this case, the performance is affected not only by the nonlinear distortions, but also by the RX noise enhancement induced by the equalizer, in compensation for the bandwidth limitations previously shown in Fig. 3(a). Furthermore, when using linear DPD, both TDECQ and TECQ measurements were not feasible

for low VCSEL bias currents. The relative performance (red lines in Fig. 9(a) and (b)) is only reported for 7.5 mA and 8 mA: using linear DPD, as the VCSEL bias current decreases, there is a progressive increase in nonlinear eye-skew (see Fig. 9(c)), with a consequent shift and closure of the lower part of the PAM4 eye diagram severely degrading the BER performance. This clearly demonstrates the necessity for a nonlinear DPD, which is instead capable of fully compensating for the VCSEL distortions either optimized using DLA or E2E approaches (refer to Fig. 9(d) and (e)). Nonlinear DPD significantly improves the optical TX performance, making possible the TDECQ and TECQ test for low bias currents, while gaining more than 3.5 dB in terms of TDECQ (Fig. 9(a)) and 3.0 dB gain in terms of TECQ (Fig. 9(b)) compared to linear DPD at higher bias currents. It must be noticed that the nonlinear DPDs meet the IEEE requirements for a maximum TDECQ and TECQ of 4.4 dB [6] in almost all cases, except for the nonlinear DPDs trained using the DLA approach when the bias current is set to 6 mA (as depicted in the inset picture of Fig. 9(a)). By making a close comparison of the nonlinear DPD performance (see inset pictures in Fig. 9(a) and (b)) clearly shows that the TDECQ-based E2E system optimization surpasses the DLA-based approach in terms of TDECQ performance. Notably, the experimental eye-diagrams acquired by the DCA exhibit poorer (i.e they are more closed) in the E2E case (Fig. 9(d1, d2, e1, e2)) compared to the DLA (Fig. 9(d3, e3)). However, this behavior is consistent to the differing optimization objectives of the two approaches. The DLA indeed focuses on optimizing the DPD specifically to improve the signal at the output of the optical TX, without necessarily guaranteeing an optimal performance after propagation through the reference TDECQ system. The proposed E2E system instead targets the DPD toward the optimal performance right at the output of the reference TDECQ DSP scheme (i.e. after the emulated WCS fiber link, the reference RX and the FFE, see Fig. 1), consistently
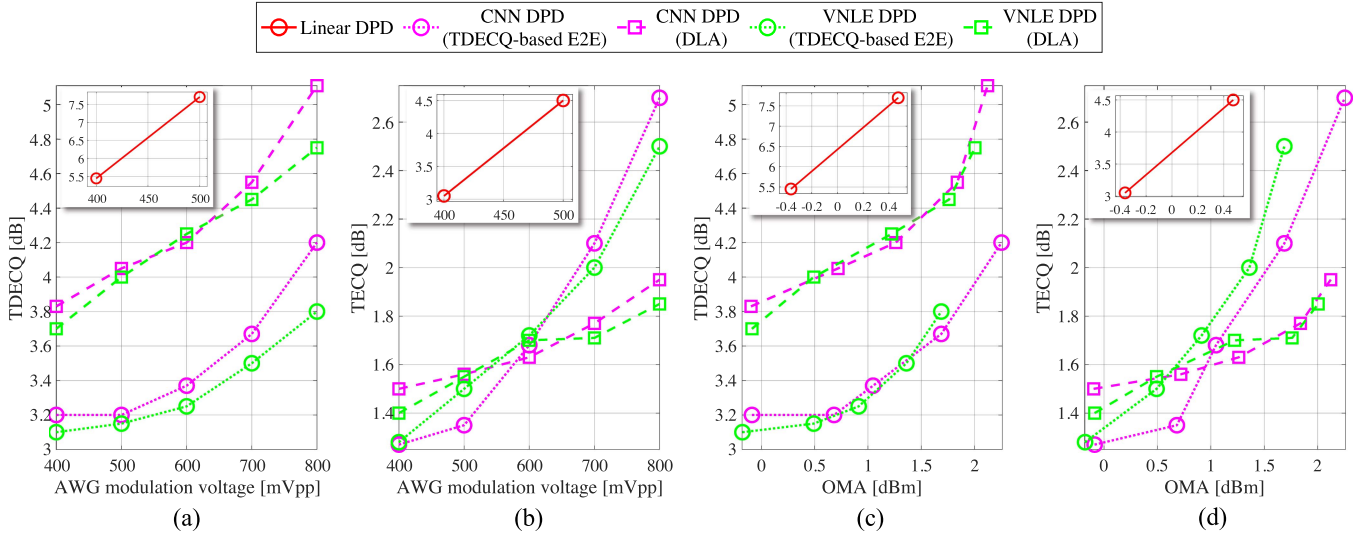
Fig. 10. Performance comparison for different pre-distorters at 107.2 Gbps on the experimental setup (a) TDECQ as function of the AWG P2P modulation voltage; (b) TECQ as function of the AWG P2P modulation voltage; (c) TDECQ as function of the OMA at the DCA input (d) TECQ as function of the OMA at the DCA input. Inset pictures show the TDECQ/TECQ performance of the linear DPD in the same scenarios.

to where the metric is evaluated. Consequently, TDECQ-based E2E optimization consistently achieves a gain of more than 0.8 dB compared to DLA, increasing to 1.0 dB for a bias current of 6 mA. Conversely, TECQ performance remains nearly equivalent using both optimization methods. This shows that the TDECQ-based E2E architecture specifically focuses on optimizing the DPD for the desired metric, which aligns with the purpose for which the E2E system was designed. Furthermore, both CNN and VNLE DPD implementations consistently demonstrate equivalent performance across all the scenarios evaluated: this highlights the effectiveness of the proposed methodology in optimizing different nonlinear DPD architectures, and suggests that further increasing the complexity of the equalizers may not yield additional performance improvements. Fig. 10 then illustrates the experimental results achieved when the VCSEL is driven with a fixed bias current of 8 mA while varying the AWG P2P modulation voltage from 400 mV to 800 mV. The average RX optical power remains set at 2 dBm, as in the previous analysis (see Fig. 9). Fig. 10(a) and (b) illustrate the TDECQ and TECQ performance as a function of the AWG modulation voltage, while Fig. 10(c) and (d) show the same curves as a function of the optical modulation amplitude (OMA) measured by the DCA.

Similarly to the observations made for low VCSEL bias currents, the TDECQ and TECQ performance of the linear DPD (depicted in the inset pictures of Fig. 10) were not measurable for modulation voltages exceeding 500 mV. This limitation arises again from the inability to meet the target maximum BER due to the increasing nonlinear distortions. In this scenario, the TDECQ-based E2E approach outperforms the DLA method in terms of TDECQ (Fig. 10(a) and (c)). Conversely, for increasing OMAs and modulation amplitudes DLA shows better performance in terms of TECQ (Fig. 10(b) and (d)) than E2E learning (yet fulfilling with margin the IEEE requirements [6]). This trend underscores the specialized nature of the TDECQ-based E2E

optimization, which proves more adept at enhancing TDECQ performance (its intended target) rather than other associated metrics. Additionally, when employing TDECQ-based E2E learning, the VNLE DPD appears to outperform the CNN DPD. However, this discrepancy can be attributed to a reduction in the OMA induced by the VNLE in response to the heightened VCSEL nonlinearity caused by the increased modulation voltage. When comparing the performance in terms of OMA, both CNN and VNLE exhibit equivalent results. Nonetheless, whether comparing signals with the same modulation P2P voltage (Fig. 10(a)) or considering them in relation to the OMA (Fig. 10(b)), the E2E approach consistently demonstrates a gain in TDECQ of more than 1 dB compared to the DLA. Experimental results thus indicate the TDECQ-based E2E approach the best DPD optimization strategy to improve this specific metric.

E. Discussion on the TDECQ-Based Optimization

The analysis of the experimental results suggests that the TDECQ-based E2E architecture, due to its significant improvement in TDECQ performance, may be the optimal approach for VCSEL-MMF nonlinear DPD optimization. However, it must be pointed out that conformance standards for SR-MMF optical TX not only consider quality metrics like TDECQ and TECQ but also specify limits on the overshoots and undershoots within the output signal of the optical TX. In particular, according to IEEE requirements [6], the maximum value of overshoot/undershoot defined as in [31] is specified as 29% for a hitting ratio equal to 3e-3. Fig. 11 illustrates the relative amount of undershoots and overshoots measured by the DCA for each scenario considered in the experiments. It can be observed that pre-distorted signals always met in the experiments the undershoot requirements with a margin (see Fig. 11(a) and (c)). However, nonlinear DPDs optimized using the TDECQ-based E2E approach exceeded the overshoot limits in nearly every
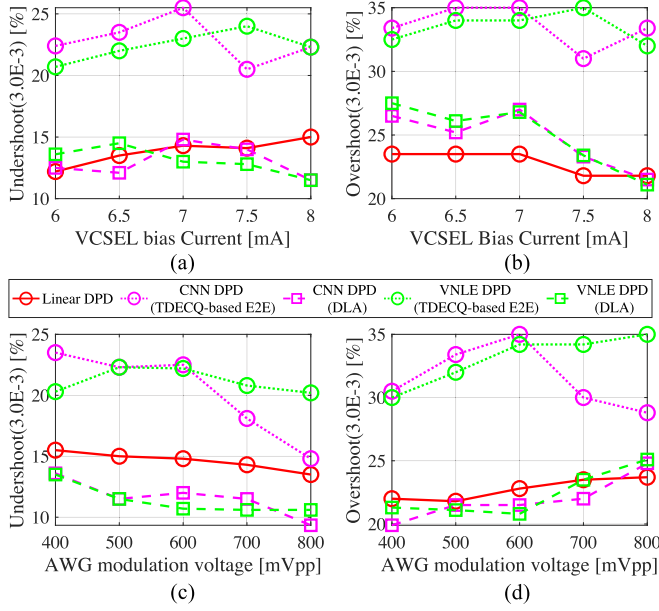
Fig. 11. Evolution of overshoots and undershoots with a hitting ratio of 3e-03 using linear and nonlinear DPD, as (a) and (b) function of VCSEL input bias current and (c) and (d) AWG modulation voltages.

scenario. This can be attributed to the focus of the E2E optimization architecture on compensating for the fiber emulation filter. The limitations of the relative bandwidth necessitate the DPD to increase the overshoots/undershoots in the TX signal to pre-compensate for the optical fiber channel, as post-mitigation using the FFE would introduce RX noise enhancement. Therefore, the experimental results indicate the presence of a trade-off between achieving optimal TDECQ performance and meeting the constraints imposed by the TX dynamics. In order to address this trade-off, it may be necessary to develop an optimization DPD strategy that specifically prioritizes the fulfillment of overshoot constraints while simultaneously improving the TDECQ performance. This can be accomplished by designing a proper CNN-based optimization architecture in future works, capable of enhancing the quality metrics of the optical TX while also adhering to some conformance constraints of the optical TX signal.

VI. CONCLUSION

In this article, we proposed a novel nonlinear DPD optimization methodology focused on enhancing the TDECQ performance of VCSEL-based optical transmitters in net 100 Gbps/λ SR-MMF optical links. We conducted a comprehensive comparison between the DLA and E2E learning optimization approaches, utilizing a specifically designed CNN framework for training nonlinear DPD tailored to meet the requirements of the considered transmission scenario. Additionally, we proposed a novel E2E architecture, based on the TDECQ standard specifications outlined in [6].

We presented a detailed experimental implementation of the proposed methodology by employing nonlinear DPD within an actual optical transmitter setup that conforms to the TDECQ measurement IEEE standards [6]. We provided comprehensive experimental results by applying nonlinear DPD to a commercial VCSEL operating under various conditions, encompassing different levels of nonlinear distortions.

The experimental results demonstrated the effectiveness of the CNN-based DPD optimization approach for enhancing VCSEL-based optical transmitters operating at a net 100 Gbps rate using PAM4 modulation. In almost all the tested cases, both the DLA and E2E methods successfully met the IEEE requirements for TDECQ and TECQ, even in conditions where linear DPD would be insufficient to achieve the target BER for measuring those metrics. Furthermore, the proposed novel TDECQ-based E2E architecture, specifically designed to optimize the TDECQ metric, consistently demonstrated a significant TDECQ gain of nearly 1.0 dB compared to the DLA-based approach in the majority of cases. This advancement enables better adherence to the IEEE TDECQ requirements across a wider range of nonlinear conditions, such as lower bias current or higher modulation swings.

Additionally, we discussed the implications of using this TDECQ-based optimization architecture and provided additional experimental results that suggest the need for further work in designing a DPD optimization architecture that leads to VCSEL-based transmitters fully compliant with all IEEE requirements for net 100 Gbps transmission over SR-MMF optical links. Future works might address also slow effects with a potential impact on the TX performance, such as temperature fluctuations [19] and VCSEL self-heating. In conclusion, both the proposed methodology and the experimental results provide valuable insights into the nonlinear DPD optimization for VCSEL-MMF optical links, indicating system-oriented optimization approaches such as End-to-end learning as a promising solution for definitively enhancing the performance of VCSEL-based optical transmitters.

APPENDIX A

NONLINEAR FINITE IMPULSE RESPONSE FILTERS AS CONVOLUTIONAL NEURAL NETWORKS LAYERS

By considering a generic linear digital filter with impulse response $h[n]$, its application over a discrete-time input sequence $x[n]$ can be mathematically described by the one-dimensional (1D) *discrete linear convolution* operation:

$$y[n] = h[n] * x[n] = \sum_{m=-\infty}^{\infty} h[m] \cdot x[n-m] \quad (10)$$

where the signal $y[n]$ is the filter output.

In practice, when dealing with Finite Impulse Response (FIR) filters and finite-length discrete sequences, the convolutional operation can be represented in the vector form as $\mathbf{y} = \mathbf{h} * \mathbf{x}$. Here, $\mathbf{h}$ is a T-dimensional vector defining the FIR filter tap coefficients, while the vectors $\mathbf{x}$ and $\mathbf{y}$ refer to the input and
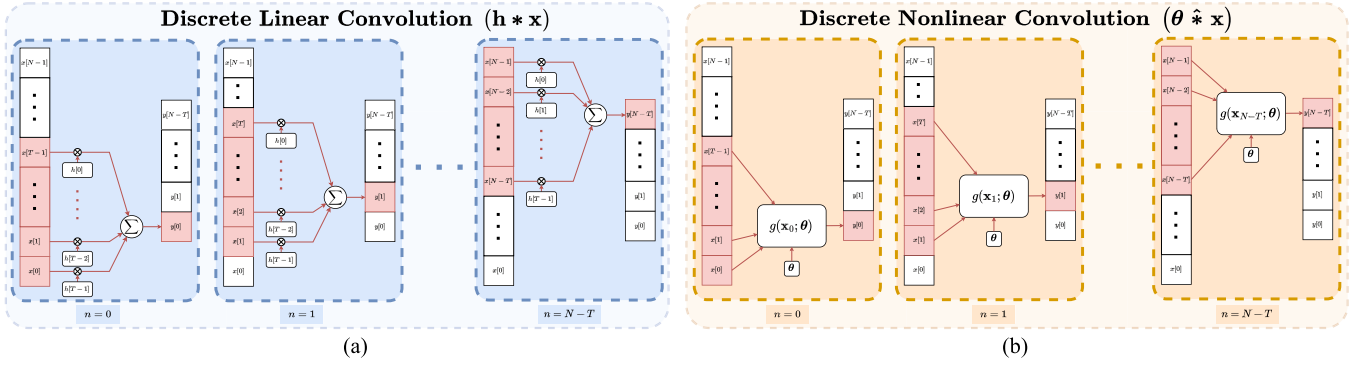
Fig. 12.   Computational schematics for (a) Discrete Linear Convolution; (b) Discrete Nonlinear Convolution.

output discrete sequences (with length N and $N - T + 1$, respectively). Each sample of $\mathbf{y}$ is computed by the linear time-invariant function $f(\mathbf{x}_n) = f(\mathbf{x}_n; \mathbf{h})$ depending on parameters $\mathbf{h}$, as follows :

$$f: \ \mathbf{x}_n \in \mathbb{R}^T \rightarrow y[n] = \mathbf{h}^\top \cdot \mathbf{x}_n \in \mathbb{R} \qquad (11)$$

where $\mathbf{x}_n = [x[n], x[n+1], \ldots, x[n+T-1]]^\top$. A graphical representation is provided in Fig. 12(a).

The described 1D linear discrete convolutions of linear FIR filters are differentiable operations, and represent a baseline type of *convolution layers*, which are the building blocks of convolutional neural networks (CNN). In order to generalize this differentiable transformation to nonlinear time-invariant filters, in this article we define the *discrete nonlinear convolution* operation $\mathbf{y} = \boldsymbol{\theta} \hat{*} \mathbf{x}$: by preserving the above definition of $\mathbf{x}$ and $\mathbf{y}$, the input-output relation of a nonlinear time-invariant filter can be modeled as the function $g(\mathbf{x}_n) = g(\mathbf{x}_n; \boldsymbol{\theta})$ depending on parameters $\boldsymbol{\theta}$, as follows:

$$g: \ \mathbf{x}_n \in \mathbb{R}^T \rightarrow y[n] = g(\mathbf{x}_n; \boldsymbol{\theta}) \in \mathbb{R} \qquad (12)$$

where the vector $\boldsymbol{\theta}$ contains all the parameters (or "coefficients") of the nonlinear filter. The operation is represented in Fig. 12(b). It can be observed that, as in the linear case, each output samples depends only on T consecutive input samples (where T represents the *memory* of the nonlinear filter). Moreover, as a consequence of the time-invariance, it is straight forward to see that shifted samples of the output $y[n+i]$ depend on analog shifted input sequences $\mathbf{x}_{n+i}$.

Therefore, under the assumption that $\theta(\mathbf{x}_n, \boldsymbol{\theta})$ is differentiable, a nonlinear time-invariant FIR filter can be abstracted as simple CNN layer, and used with a black box approach as differentiable nonlinear operator to represent the several components of an E2E communication systems.

ACKNOWLEDGMENT

REFERENCES

[1] J. King, "In support of 200 G MMF Ethernet PMDs," IEEE 802.3 Next-generation 200 Gb/s and 400 Gb/s MMF PHYs Study Group. [Online]. Available: //www.ieee802.org/3/NGMMF/public/Jan18/young_NGMMF_01a_jan18.pdf

[2] A. Maharry et al., "A 50 gbps 9.5 pJ/bit VCSEL-based optical link," in *Proc. IEEE Photon. Conf.*, 2021, pp. 1–2.

[3] J. A. Tatum, G. D. Landry, D. Gazula, J. K. Wade, and P. Westbergh, "VCSEL-Based optical transceivers for future data center applications," in *Opt. Fiber Commun. Conf.*, 2018, pp. 1–3. [Online]. Available: https://opg.optica.org/abstract.cfm?URI=OFC-2018-M3F.6

[4] J. A. Tatum et al., "VCSEL-Based interconnects for current and future data centers," *IEEE/OPTICA J. Lightw. Technol.*, vol. 33, no. 4, pp. 727–732, Feb. 2015.

[5] "VCSEL Market to Grow at 19.2% CAGR from 1. 6bn in 2022 to 3.9bn in 2027," 2022. [Online]. Available: https://www.semiconductor-today.com/news_items/2022/sep/yole-210922.shtml

[6] *IEEE Standard for Ethernet - Amendment 3 : Physical Layer Specifications and Management Parameters for 100 Gb/s, 200 Gb/s, and 400 Gb/s Operation over Optical Fiber using 100 Gb/s Signaling,*" IEEE Standard 802.3db-2022, 2022.

[7] J. King, D. Leyba, and G. D. LeCheminant, "TDECQ (transmitter dispersion eye closure quaternary) replaces historic eye-mask and TDP test for 400 Gb/s PAM4 optical transmitters," in *Proc. Opt. Fiber Commun. Conf. Exhib.*, 2017, pp. 1–3.

[8] S. Echeverri-Chacón et al., "Transmitter and dispersion eye closure quaternary (TDECQ) and its sensitivity to impairments in PAM4 waveforms," *IEEE/OPTICA J. Lightw. Technol.*, vol. 37, no. 3, pp. 852–860, Feb. 2019.

[9] H. Zhou et al., "Recent advances in equalization technologies for short-reach optical links based on PAM4 modulation: A. review," *Appl. Sci.*, vol. 9, no. 11, 2019, Art. no. 2342. [Online]. Available: https://www.mdpi.com/2076-3417/9/11/2342

[10] *IEEE Standard for Ethernet - Amendment 10 Media Access Control Parameters, Physical Layers, and Management Parameters for 200 Gb/s and 400 Gb/s Operation*, IEEE Standard 802.3bs-2017, 2017.

[11] A. Rezania, J. C. Cartledge, A. Bakhshali, and W.-Y. Chan, "Compensation schemes for transmitter- and receiver-based pattern-dependent distortion," *IEEE Photon. Technol. Lett.*, vol. 28, no. 22, pp. 2641–2644, Nov. 2016.

[12] H. Paaso and A. Mammela, "Comparison of direct learning and indirect learning predistortion architectures," in *Proc. IEEE Int. Symp. Wireless Commun. Syst.*, 2008, pp. 309–313.

[13] G. Paryanti, H. Faig, L. Rokach, and D. Sadot, "A direct learning approach for neural network based pre-distortion for coherent nonlinear optical transmitter," *IEEE/OPTICA J. Lightw. Technol.*, vol. 38, no. 15, pp. 3883–3896, Aug. 2020.

[14] A. Mämmelä, "Commutation in linear and nonlinear systems," *Frequenz*, vol. 60, no. 5/6, pp. 92–94, 2006.

[15] L. Minelli, F. Forghieri, A. Nespola, S. Straullu, and R. Gaudino, "A multi-rate approach for nonlinear pre-distortion using end-to-end deep learning in IM-DD systems," *IEEE/OPTICA J. Lightw. Technol.*, vol. 41, no. 2, pp. 420–431, Jan. 2023.

[16] L. Minelli, F. Forghieri, T. Shao, A. Shahpari, and R. Gaudino, "Net 100 Gb/s/λ VCSEL MMF nonlinear digital pre-distortion using convolutional neural networks," in *Proc. Opt. Fiber Commun. Conf. Exhib.*, 2023, pp. 1–3.

[17] H. Jiang, M. Fu, Y. Zhu, L. Yi, W. Hu, and Q. Zhuge, "Digital pre-distortion coefficients identification using gauss-newton based direct learning architecture," in *Proc. Opt. Fiber Commun. Conf. Exhib.*, 2023, pp. 1–3.

[18] V. Bajaj, F. Buchali, M. Chagnon, S. Wahls, and V. Aref, "Deep neural network-based digital pre-distortion for high baudrate optical coherent transmission," *IEEE/OPTICA J. Lightw. Technol.*, vol. 40, no. 3, pp. 597–606, Feb. 2022.

[19] M. Srinivasan et al., "End-to-end learning for VCSEL-Based optical interconnects: State-of-the-art, challenges, and opportunities," *IEEE/OPTICA J. Lightw. Technol.*, vol. 41, no. 11, pp. 3261–3277, Jun. 2023.

[20] L. Minelli F. Forghieri, and R. Gaudino, "End-to-end deep learning for VCSEL's nonlinear digital pre-distortion," in *Proc. IEEE Italian Conf. Opt. Photon.*, 2022, pp. 1–4.

[21] L. Minelli, F. Forghieri, and R. Gaudino, "Nonlinear pre-distortion through a multi-rate end-to-end learning approach over VCSEL-MMF IM-DD optical links," in *Proc. Eur. Conf. Opt. Commun.*, 2022, pp. 1–4.

[22] B. Karanov et al., "End-to-end deep learning of optical fiber communications," *IEEE/OPTICA J. Lightw. Technol.*, vol. 36, no. 20, pp. 4843–4855, Oct. 2018.

[23] T. O'Shea and J. Hoydis, "An introduction to deep learning for the physical layer," *IEEE Trans. Cogn. Commun. Netw.*, vol. 3, no. 4, pp. 563–575, Dec. 2017.

[24] I. J. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016. [Online]. Available: http://www.deeplearningbook.org

[25] V. Dumoulin and F. Visin, "A guide to convolution arithmetic for deep learning," 2016, *arXiv:1603.07285*.

[26] R. Schmogrow et al., "Error vector magnitude as a performance measure for advanced modulation formats," *IEEE Photon. Technol. Lett.*, vol. 24, no. 1, pp. 61–63, Jan. 2012.

[27] A. G. Baydin, B. A. Pearlmutter, A. A. Radul, and J. M. Siskind, "Automatic differentiation in machine learning: A survey," *J. Marchine Learn. Res.*, vol. 18, pp. 1–43, 2018.

[28] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. 32nd Int. Conf. Mach. Learn.*, 2015, pp. 448–456. [Online]. Available: https://proceedings.mlr.press/v37/ioffe15.html

[29] M. D. Zeiler, D. Krishnan, G. W. Taylor, and R. Fergus, "Deconvolutional networks," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2010, pp. 2528–2535.

[30] M. Schädler, G. Böcherer, F. Diedolo, and S. Calabró, "Nonlinear component equalization: A. comparison of deep neural networks and volterra series," in *Proc. Eur. Conf. Opt. Commun.*, 2022, pp. 1–4.

[31] *IEEE Standard for Ethernet - Amendment 11: Physical Layers and Management Parameters for 100 Gb/s and 400 Gb/s Operation over Single-Mode Fiber at 100 Gb/s per Wavelength*, IEEE Standard 802.3cu-2021, 2021.