Net 100 Gb/s/λ VCSEL+MMF nonlinear Digital Pre-Distortion using Convolutional Neural Networks

Leonardo Minelli,[1,*] Fabrizio Forghieri,[2] Tong Shao,[3] Ali Shahpari[3], and Roberto Gaudino[1]

[1] *Department of Electronics and Telecommunications (DET), Politecnico di Torino, corso Duca degli Abruzzi 24, 10129 Torino, Italy*
[2] *CISCO Photonics, Via S. M. Molgora, 48/C, 20871 Vimercate MB, Italy*
[3] *CISCO Optical GmbH, Nordostpark 12, 90411 Nürnberg, Germany*
[*] *leonardo.minelli@polito.com*

Abstract: We experimentally demonstrate VCSEL+MMF nonlinear Digital Pre-Distorters, optimized using Convolutional Neural Networks, for fulfilling the IEEE P802.3db$^{\mathrm{TM}}$/D3.2 TDECQ requirements for net 100 Gb/s/λ optical transmitters.

1. Introduction

Today's Internet traffic is continuously growing, and Data Centers Intra-connects (DCI) need to keep up with the increasing traffic demand with cost-effective and power-efficient solutions. In this regard, over 50% of the current DCI are Intensity Modulation-Direct Detection (IM-DD) optical links, leveraging Multi-Mode Fibers (MMF) [1] and Vertical-Cavity Surface-Emitting Lasers (VCSEL). In particular, next generation MMF-VCSEL DCI up to 100 m are expected to provide a net rate of 100 Gbit/s per λ, using simple multilevel Pulse Amplitude Modulation (PAM) formats such as PAM4. However, at such bit rate these optical links are severely bandlimited by optoelectronics and MMF modal dispersion. Moreover, VCSELs introduce significant nonlinear distortions when driven at their typical operating 7-8 mA bias current [2]. Among several DSP equalization solutions proposed at the receiver (RX) and transmitter (TX) side for PAM4 VCSEL-MMF links [3], nonlinear Digital Pre-Distorers (DPD) working at 1 sample-per-symbol (sps) ratio tend to be easier to implement than nonlinear post-equalizers: the TX symbols calibration, performed by the DPD at factory level, can in fact be stored in simplified structures such as Look-up Tables. Moreover, nonlinear DPD allows to improve the Transmitter Dispersion eye Closure Quaternary (TDECQ) test, that has become a relevant key quality measure in the standards for optical MMF links [4]. Currently, one of the most challenging problems is the nonlinear DPD optimization process. The most commonly adopted strategies are based on the Indirect Learning Architecture (ILA) and the Direct Learning Architecture (DLA) [5], focused on compensating TX nonlinearities and bandwidth limitations. Alternatively, system-oriented optimizations are also proposed in literature, such as those based on the End-to-end learning approach. [6] [7]

In this paper, we propose a novel Convolutional Neural Network-based Direct Learning Architecture (CNN-DLA), with which we exploit the statistics of the outputs in the convolutional system to optimize the DPD under an Error Vector Magnitude (EVM) criterion, satisfying dynamics constraints considered in [6] and [7]. We moreover introduce White Gaussian Noise within the system to improve the DPD optimization process. We experimentally evaluate the proposed algorithm on two experimental setups. In the 1st, we optimize different DPDs to transmit +100 net Gb/s/λ PAM4 signals on a 850 nm VCSEL, to then perform TDECQ performance test. In the 2nd, we evaluate the same pre-distorted VCSEL in terms of Bit Error Rate (BER) versus RX outer Optical Modulation Amplitude ($\mathrm{OMA_{outer}}$), using a different RX from the one used for training and 100 m of OM4 fiber. We finally compare the TDECQ performance of nonlinear DPD optimized using EVM-based and TDECQ-oriented approaches. We thus demonstrate that with our DPD optimization approach it is possible to obtain a VCSEL TX module compliant to the standard IEEE P802.3db/D3.2 [4] for 100 Gb/s/λ, with a margin of more than 0.4 dB in terms of TDECQ.

2. DPD experimental optimization

In this Section, we describe the optimization for a PAM4 DPD at 107.2 Gb/s (slightly differing from the required 106.25 Gb/s, to ensure instrumental compatibility yet being on the safe side using an higher bit rate). It mainly consists of the two phases illustrated respectively in Fig. 1.a and 1.b. In the first step, we transmit a 2 sps rectangularly shaped waveform over a Back-to-back (B2B) experimental TX system (see Exp. setup #1 in Fig. 1.a). It consists of: an Arbitrary Waveform Generator (sampling rate: 107.2 Gsa/s, bandwidth: 50 GHz, modulation peak-to-peak voltage: 500 mV), an 8 mA DC current supplier, a probed 850 nm VCSEL (bandwidth: 22 GHz) at a 25 °C temperature controlled by a Peltier cell, a fiber collimator, a Variable Optical Attenuator (VOA) and a Digital Communication Analyzer (DCA). The latter integrates a PhotoDiode, an Electrical Amplifier (EA) and a Digital Sampling Oscilloscope (DSO). We used as TX pattern x_{tx} a PAM4 PRBS16 sequence linearly pre-distorted at 1 sps, to better match the statistics of the nonlinearly pre-distorted signal when passing through the VCSEL. The acquired signal, averaged by the DCA, gets resampled to 10 sps, aligned to the TX pattern and normalized to have unit variance. The resulting RX signal y_{rx} is used together with the x_{tx} to form a training dataset for a 1-to-10 sps Digital Twin (DT) of the Transmission system. The latter is the cascade of a derivable resampler (implemented as a polyphase 1D convolutional layer, see *Pytorch*'s *Resample* class) and a 1D-CNN characterized by standard

1D convolutional layers (i.e., with unitary stride and no dilation nor padding, see *Pytorch*'s *Conv1D* class), whose real-time structure can be viewed a FIR Neural Network [8]. The training of the DT consists on several iterations where a mini-batch of consecutive input samples (i.e. a subset of x_{tx}) are convoluted through the DT structure, to then backpropagate the MSE loss computed between the 1D-CNN output $\hat{y}_{rx}$ and the corresponding y_{rx} subset. In our experiments, we obtained a MSE loss equal to -23.5 dB tested on the reference SSPRQ pattern [4], using a 1D-CNN with 3 ReLU layers, 30 channels per inner 1D convolutional layer and kernel sizes (i.e. taps per FIR) equal to 12 (training parameters: 6e4 iterations, 1e2 symbols per batch, Adam optimizer, step size 1e-2).

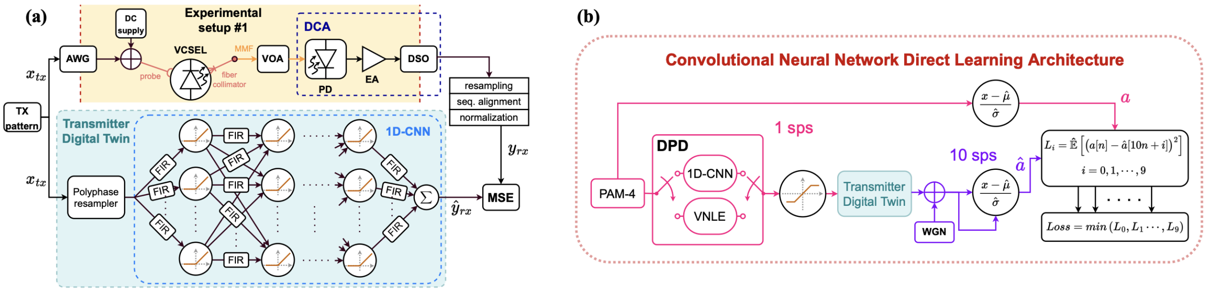
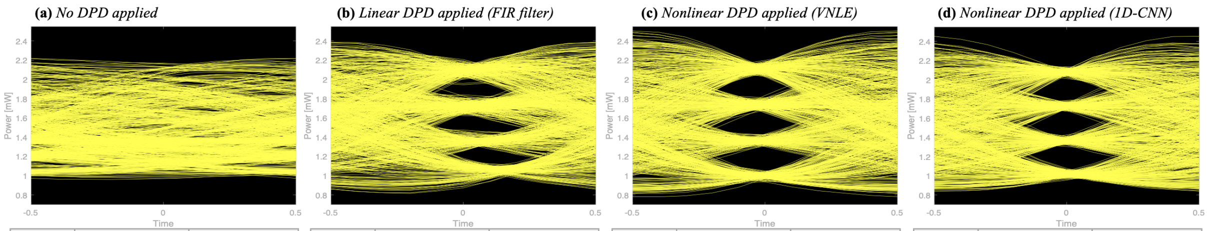


Fig. 1: a) Overall schematic of the Digital Twin modeling phase b) CNN DLA architecture adopted to optimize the DPD.

In the second step, the DT is used to train a nonlinear DPD, either a 1D-CNN or a Volterra NonLinear Equalizer (VNLE), in the architecture illustrated in Fig. 1.b. For 1000 iterations, a batch of 3000 random PAM4 symbols is convoluted through the DPD, being then clipped by an Hard-Limiter (HL) function (as in [6]) to pass through the DT. The resulting 10 sps output is then scaled to zero mean and unit variance on each possible decimation instant by a derivable normalization layer, as well as the input PAM-4 sequence (aligned according to the system delay): this prevents the DPD from synthesize uncontrolled gains, that with the HL function would tend to vanish the backpropagated gradients. Moreover, to avoid that the DPD drives the signal so that the VCSEL acts as linear at expense of an excessive $\text{OMA}_{\text{outer}}$ reduction, WGN is added before normalization to stimulate an increase in the VCSEL's output Signal-to-Noise ratio. Finally, the loss function is defined as the minimum among the the MSE computed between the normalized DLA input a and each possible 1 sps decimation within the symbol Unit Interval (UI) of the output $\hat{a}$. The DPD therefore is optimized according to an equivalent minimum EVM criterion.

3. Experimental results

In this Section we compare the VCSEL performance using two nonlinear DPD obtained with the proposed optimization method (i.e., a 5th order VNLE and a 1D-CNN with 2 hidden Relu Layer and 30 channels per inner convolution), a linear DPD (i.e., a FIR filter obtained through ILA) and a benchmark case which doesn't use pre-distortion. The total memory (linear and nonlinear) of all considered DPDs is 7 symbols. In Fig. 2 we show the eyediagrams acquired by the DCA on the SSPRQ pattern together with the related measures of TDECQ (compliant to [4]), undershoot and overshoot (defined as in [9] but with an hit ratio of 3e-3 according to [4]).



TDECQ	undershoot (3e-3)	overshoot (3e-3)
---	5.90%	6.60%

TDECQ	undershoot (3e-3)	overshoot (3e-3)
7.50 dB	15.5%	22.25%

TDECQ	undershoot (3e-3)	overshoot (3e-3)
4.10 dB	11.20%	24.20%

TDECQ	undershoot (3e-3)	overshoot (3e-3)
4.35 dB	11.35%	23.85%

Fig. 2: SSPRQ eyediagrams acquired on the Experimental setup #1 (see Fig.1): a) no DPD applied ($\text{OMA}_{\text{outer}}$=0.35 dBm) b) FIR DPD applied ($\text{OMA}_{\text{outer}}$=0.3 dBm) dB) c) VNLE DPD applied ($\text{OMA}_{\text{outer}}$=0.8 dBm) d) 1D-CNN DPD applied ($\text{OMA}_{\text{outer}}$=0.5 dBm). Symbol Rate R_s: 53.6 GBaud, avg. power $\bar{P}$: 2.05 dBm, OPL:5 dB. 26.8 GHz 4-poles Bessel filter digitally applied by the DCA. Measurements compliant to [4]

Without DPD, the eyediagram was too closed to let the DCA to measure the TDECQ. In the Linear DPD case instead, the TDECQ=7.50 dB is far from the maximum reference value of 4.4 dB in [4], due to the residual nonlinear eye-skew. In the case of the nonlinear DPDs instead the TDECQ is lower than the reference: 4.35 dB for the 1D-CNN and 4.10 dB for the VNLE, with the Volterra thus outperforming the Neural Network by 0.25 dB. In all cases the under/overshoot requirements are met, being lower than the reference maximum of 29% in [4].

With the same 107.2 Gbps DPDs we then evaluated the performance of the pre-distorted VCSEL on a transmission system different from the one used for the DPD training: an overall schematic is illustrated in Fig.3.a. We indeed modified the 1st experimental setup by introducing 100 m of OM4 fiber, and we used a PIN+TIA RX (PD Bandwidth:25 GHz) together with a 256 Gsa/s Real-Time Oscilloscope (RTO, Bandwidth:80 GHz), also in order to measure the BER. For different Optical Path Losses (OPL), a DSP post-processing was applied to the acquired signals, that were resampled to 2 sps, aligned, normalized and equalized either with a 2 sps FFE (memory:15 symbols) or a 2 sps DFE (20 pre-cursors, 10 post-cursors). We finally decoded by hard decision the PAM-4 signal (grey coding assumed) and evaluated the BER for about 2.5e6 symbols. The BER vs OMA_{outer} performance related to all the 8 cases are reported in Fig3.b (overall comparison) and 3.c (zooming around the target BER:2.4e-4).

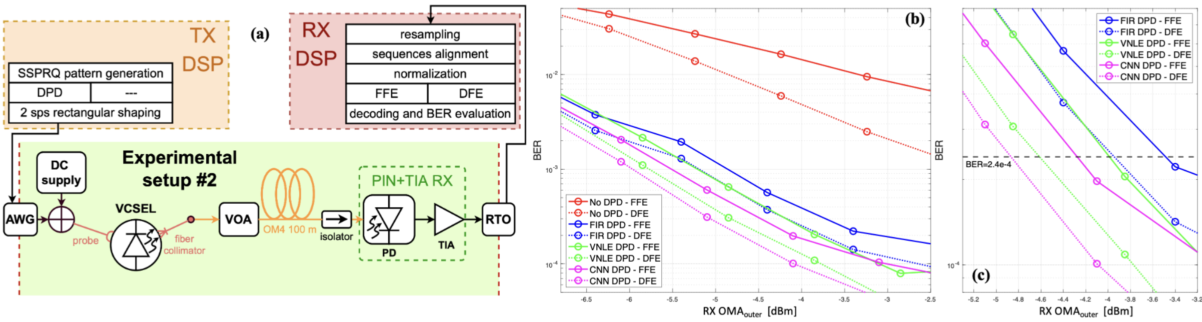


Fig. 3: a) Overall schematics of the experimental setup used for the BER vs RX OMA_{outer} performance evaluation b) BER vs RX OMA_{outer} performance evolution (overall comparison) c) BER vs OMA_{outer} performance evolution (zoomed around target BER=2.4e-4)

The above results clearly demonstrate that without DPD the performance of a VCSEL-MMF optical link cannot meet the BER requirements for the 100 Gb/s links, either using DFE or FFE. Regarding instead the performance of the different DPDs, the 1D-CNN DPD outperforms the other predistorters: at BER=2.4e-4, it gains in terms of OMA_{outer} up to 0.8 dB w.r.t the FIR DPD and 0.3 dB w.r.t the VNLE DPD using the FFE, while using the DFE the gain is up to 1.0 dB w.r.t the FIR DPD and 0.3 dB w.r.t the VNLE DPD. It is noticeable that, differently than in the TDECQ test, in this case is the Neural Network to give the best performance, that with a TX OMA_{outer}=3.5 dBm [4] brings to a Power Budget (PB) in terms of Optical Path Loss (OPL) equal to 7.8 dB (using FFE) and 8.4 dB (using DFE) at target BER=2.4e-4: with respect to the reference PB=6.4 dB for max TDECQ [4], we thus achieved respectively a margin of 1.4 dB and 2.0 dB using the 1D-CNN DPD.

4. Results using a TDECQ-oriented optimization

In addition to the previous results, we evaluated the performance of the same two nonlinear DPDs after training them using an optimization more oriented to the TDECQ: this consists on modifiyng the Loss function illustrated in Sec.2 by backpropagating the EVM related to the 2 decimation instants giving the best performance (that at 10 sps are spaced by 0.1 UI according to the TDECQ in [4]). With this modified optimization, the TDECQ of both VNLE and 1D-CNN was lowered to 3.95 dB (and under/overshoot below 11.5/22.5% in both cases): the margin w.r.t the maximum reference TDECQ=4.4 dB could thus be increased from 0.05/0.30 dB to 0.45 dB. BER vs OMA_{outer} preliminary performance evaluation showed however equivalent results to those illustrated in Fig.3.

5. Conclusion

In this paper, we experimentally demonstrated an effective VCSEL DPD optimization algorithm, based on Convolutional Neural Networks to train the DPD under EVM and TDECQ oriented criterions. The optimized nonlinear DPDs, either based on Volterra series or ANN demonstrated to achieve TX performance compliant to IEEE P802.3db^{TM}/D3.2, with a margin up to 0.45 dB in terms of TDECQ and, moreover, a margin up to 2.0 dB in terms of OPL for a target BER=2.4e-4.

Acknowledgment: *This work was carried out under a research contract with Cisco Photonics. We also acknowledge the PhotoNext Center at Politecnico di Torino (http://www.photonext.polito.it/) and Cisco Optical GmbH at Nürnberg.*

References

1. Jonathan King. In Support of 200G MMF Ethernet PMDs. IEEE 802.3 Next-generation 200 Gb/s and 400 Gb/s MMF PHYs Study Group.
2. Jim A. Tatum et al. VCSEL-Based Interconnects for Current and Future Data Centers. *Journal of Lightwave Technology*, 33(4):727–732, 2015.
3. Honghang Zhou et al. Recent Advances in Equalization Technologies for Short-Reach Optical Links Based on PAM4 Modulation: A Review. *Applied Sciences*, 9(11), 2019.
4. IEEE Approved Draft 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 P802.3db/D3.2, July 2022*.
5. Gil Paryanti, Hananel Faig, Lior Rokach, and Dan Sadot. A Direct Learning Approach for Neural Network Based Pre-Distortion for Coherent Nonlinear Optical Transmitter. *Journal of Lightwave Technology*, 38(15):3883–3896, 2020.
6. Leonardo Minelli et al. Nonlinear Pre-distortion through a Multi-rate End-to-end Learning Approach over VCSEL-MMF IM-DD Optical Links. In *2022 European Conference on Optical Communication (ECOC)*, 2022.
7. Leonardo Minelli et al. A Multi-Rate approach for Nonlinear Pre-distortion using End-to-end Deep Learning in IM-DD systems. *Journal of Lightwave Technology*, Accepted for publication.
8. E.A. Wan. Temporal backpropagation for FIR neural networks. In *1990 IJCNN International Joint Conference on Neural Networks*, pages 575–580 vol.1, 1990.
9. 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 Std 802.3cu-2021 (Amendment to IEEE Std 802.3-2018 and its approved amendments)*, pages 1–87, 2021.