Nonlinear Pre-distortion through a Multi-rate End-to-end Learning Approach over VCSEL-MMF IM-DD Optical Links

Nonlinear Pre-distortion through a Multi-rate End-to-end Learning Approach over VCSEL-MMF IM-DD Optical Links

Leonardo Minelli[(1)], Fabrizio Forghieri[(2)], Roberto Gaudino[(1)]

[(1)] Dipartimento di Elettronica e Telecomunicazioni (DET), Politecnico di Torino, Italy, leonardo.minelli@polito.it
[(2)] CISCO Photonics Italy, Vimercate, Italy

Abstract *We experimentally demonstrate a nonlinear digital pre-distorter for PAM-M shaping in VCSEL+MMF IM-DD links able to operate at a generic baud rate using a fractional sample-per-symbol Neural Network. We focus on efficient and practical multi-rate operation, signal amplitude constraints, and linear equalizer at the receiver.*

Introduction

Data Centers Interconnects (DCI) internal traffic is continuously growing, requesting a continuous increase in short-reach optical links data rates. Marketwise, about 50% of these DCI internal optical links are still using Multi-Mode Fibers (MMF) [1] and Vertical-Cavity Surface-Emitting Lasers (VCSEL) due to their high-power efficiency and significantly low manufacturing chip cost, and they exploit Intensity Modulation-Direct Detection (IM-DD) On-Off Keying (OOK). Focusing on MMF lengths around 100 m, there is today a growing interest to find solutions targeting 100+ Gbit/s per λ using multilevel Pulse Amplitude Modulation (PAM) formats. To counteract the resulting bandwidth limitations and nonlinear distortions affecting MMF-VCSEL links at high data rates, linear and nonlinear equalization technologies have been widely investigated in recent years based on different Digital Signal Processing (DSP) techniques [2]. While part of the research is currently focused on nonlinear post-equalizers at the receiver (RX) [3,4], significant attention was also put on Digital Pre-Distorters (DPD), able to pre-equalize before transmission (TX), as nonlinear equalization DSP algorithms are usually easier to implement at the TX (on noiseless PAM-M) than at the RX.

Besides the choice of the DPD typology, implementable either using Look-Up Tables [5,6], Volterra-series [7], or Neural Networks (NN) [8], another important issue is choosing the strategy used to optimize/train them. While the Indirect Learning (ILA) and Direct Learning Architecture (DLA) are the two mainly adopted strategies for training [7,9-11], several alternatives were also investigated, such as DPDs optimization based on reinforcement learning [12] or End-to-End (E2E) learning [13-16]. The latter approach, where the DPD is jointly optimized with the post-equalizer as a unified autoencoder [17], theoretically leads to the "absolute" optimal performance under specific physical constraints, such as bounded VCSEL input and different Baud Rate, DAC and ADC sampling frequency. In fact, in this paper we propose a DPD optimization method whose novelty resides in natively taking into account different sampling rates for different parts of the DSP. In particular, our proposal can handle generic non-integer (but rationale) ratios among baud rate D, DAC sampling rate f_{DAC} and ADC sampling rate f_{ADC} (i.e., dealing with a non-integer number of samples per symbol in both the TX and RX DSP).

We thus show a novel E2E Neural Network (NN) architecture able to be optimized through forward and backward propagation at the aforementioned multiple sampling frequencies. Our method is capable to synthesize a nonlinear NN DPD that encodes PAM-M symbols into signals with an arbitrarily fractional sample-per-symbol (sps) ratio, natively fulfilling the amplitude constraints imposed at the DAC output and VCSEL input. We demonstrate the proposed DSP on an experimental VCSEL-MMF IM-DD setup at 100+ Gbit/s bit rate per λ, showing that this DPD approach can provide a significant BER performance gain compared to a system using only feed-forward equalizer (FFE) at the RX side. We use a VCSEL with $B_{3dB}^{opt.}$=20 GHz and P_{out}=5 mW and OM4 MMF fiber.

The proposed approach for DPD optimization

In this section, we outline a practical sequence of steps toward the optimization of the proposed fractionally spaced NN DPD in an experimental setup. In our notation in the rest of the paper, as shown in Fig. 1, for each discrete sequence, we add a subscript on its discrete-time index *n* to recall its related sampling frequency f_s (i.e., $f_s = D \rightarrow n_D$, $f_s = f_{DAC} \rightarrow n_{DAC}$, $f_s = f_{ADC} \rightarrow n_{ADC}$). Moreover, the ratio between the involved sampling rates must be rationale. For instance, as in one of our experiments, D=50 GBaud, f_{DAC}=92 *GSa/s,* and f_{ADC}=*200 GSa/s.*

1) received signal acquisition, denoising and noise spectral estimation. We start our DPD optimization procedure by measuring several periods (~1000) of pseudo-random PAM-M signal at the RX. The obtained sequence $y[n_{ADC}]$,

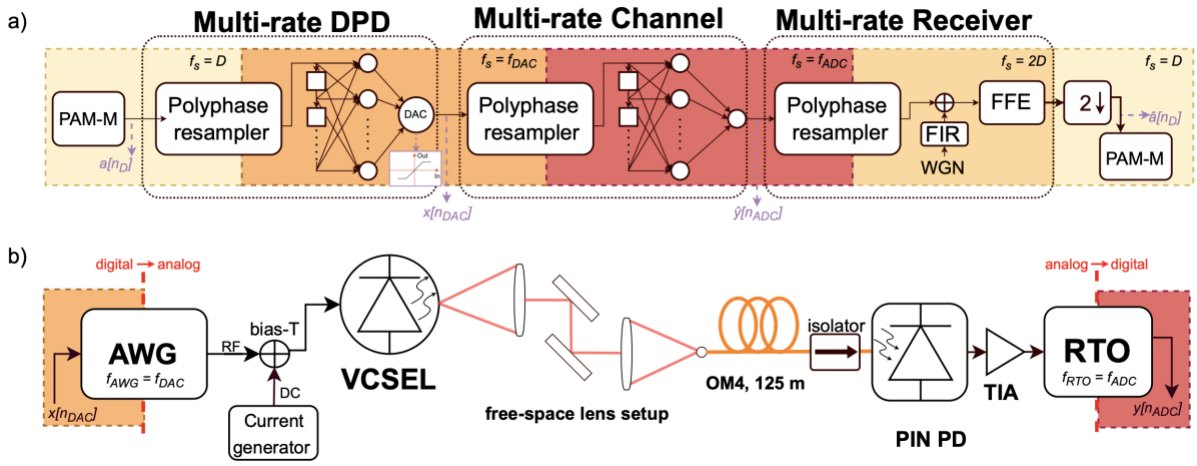

Figure 1: An overall schematic of the nonlinear predistortion experiment**.** On the upper part (a) the End-to-end multi-rate system. On the lower part (b) the experimental VCSEL-MMF IM-DD link. After DPD optimization, the pre-distorted PAM-M signal can be directly transmitted from the AWG. RF: Radio Frequency; DC: Direct Current; PD: Photo-Diode; TIA: Trans-Impedance Amplifier.

is averaged over the PRBS repetitions, to obtain a denoised version $\bar{y}[n_{ADC}]$, which is subtracted to each period of the original RX signal, to obtain an estimation of the receiver noise as the residual signal $r[n_{ADC}] = y[n_{ADC}] - \bar{y}[n_{ADC}]$.

2) Multi-rate Channel modelization and colored noise filter design: The channel deterministic transformation (i.e., linear and nonlinear VCSEL+MMF induced distortions) is modeled through the optimization of a multi-rate neural network (see Fig. 1), that taking as input the transmitted sequence $x[n_{DAC}]$, is trained to output the signal $\hat{y}[n_{ADC}]$ as a mimimum mean-square error (MSE) estimate of $\bar{y}[n_{ADC}]$ (equivalently to the DLA first optimization step[11]). The VCSEL+MMF model is composed (in DSP) by a polyphase resampler cascaded to a Feed-Forward Neural Network (FFNN), whose inputs are the entries of a digital delay line, and with a linear output activation. For receiver noise modeling, we start from a White Gaussian random generator injected into an FIR filter whose transfer function fits the Power Spectral Density (PSD) estimated on the previously described signal $r[n_{ADC}]$.

3) E2E Multi-rate optimization: based on the FFNN channel model found in the previous point, we then move to the E2E multi-rate system optimization (see again Fig. 1(a)), where a multi-rate DPD is trained jointly with a 2 sps RX FFE equalizer. The DPD outputs a signal whose sampling rate is the same as the DAC rate and has a limited peak-to-peak swing compatible with the VCSEL allowable input. This signal amplitude limitation is key to our NN optimization, and it is natively taken into account in the back-propagation algorithm. This last optimization is performed using a Stochastic Gradient Descent algorithm, minimizing the following MSE:

$$\mathrm{MSE} = \mathrm{E}[|a[n_D - L] - \hat{a}[n_D]|^2] \qquad (2)$$

where D is the Baud rate, $a[n_D - L]$ is the multi-rate DPD input delayed by the E2E system latency L (in symbols) and $\hat{a}[n_D]$ is the FFE output decimated by a factor 2.

Our E2E optimization uses a specific backpropagation algorithm, which extends the one originally proposed in [18] to work at several sampling rates. While in the forward direction the signal propagates through a polyphase resampler and gets up-sampled, then passes through the antialiasing FIR filter and is then decimated, the relative loss gradient is back-propagated in a symmetric fashion. In fact, the time-dependent loss derivative can be mathematically interpreted as a signal that gets up-sampled, backward filtered and decimated changing its sampling rate reciprocally to the forward propagated signal.

As another novelty of our approach, aimed at efficient convergence, the receiver noise is not added as a time-domain signal at the FFE input, but it is added semi-analytically as an additive regularization term [19] in the FFE's Stochastic Gradient Descent (SGD) update.

Finally, since DPD's amplitude constraint tends to penalize outer PAM-M levels, we introduce an optimization heuristic that gives them higher weights in the error function compared to the inner PAM-M levels.

Experimental setup

Tab. 1: E2E system parameters

TX DPD FFNN depth:	1 hidden layer (ReLU)
TX DPD FFNN input/hidden size	21 taps/neurons
Channel FFNN depth:	1 hidden layer (ReLU)
Channel FFNN input/hidden size	61 taps/neurons
RX FFE length:	31 taps
RX Noise FIR length:	61 taps

A schematic of the experimental setup is illustrated in Fig. 1(b). In the training phase, a PAM-M sequence of period equal to 2^{15} symbols is generated using an Arbitrary Waveform Generator (AWG) with sampling rate f_{DAC}=92 GSample/s and DAC peak-to-peak voltage $DAC\text{-}vpp$ = 700 mV (i.e., the maximum allowed). After 9 mA of bias current addition (i.e. the optimal value for BER performances without DPD), the signal is injected into an 850 nm VCSEL ($\sim 14\ mApp$ modulated current). The emitted light

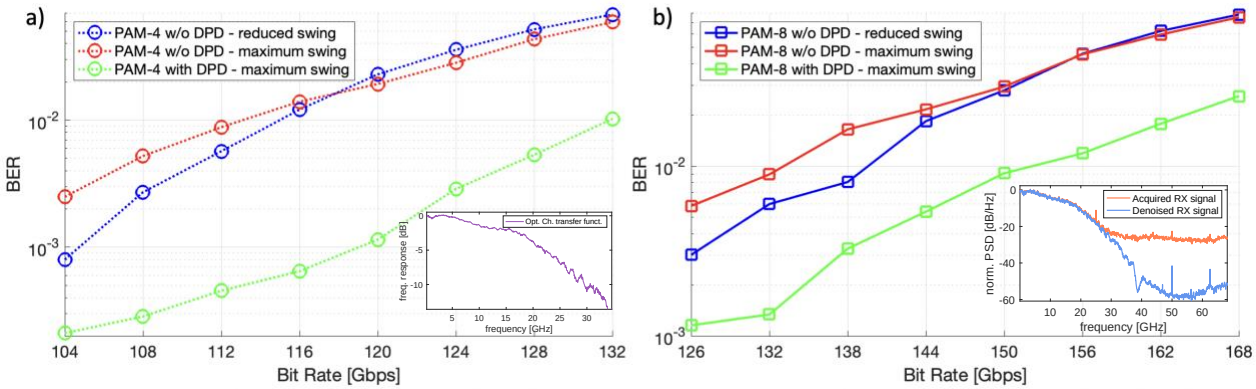

Figure 2: BER versus Bit Rate curves after 125m OM4 MMF, transmitting (a) PAM-4 and (b) PAM-8 signals over the experimental transmission system. Red curves: no DPD, maximum voltage swing (700 mV); Blue curves: no DPD, reduced voltage swing (500 mV); Green curves: DPD applied, maximum voltage swing (700 mV). Inset figures: (a) Optical Channel transfer function; (b) Power Spectral Density (PSD) normalized w.r.t DC component of RX signal before and after denoising.

is then collimated using a free-space lens setup into a 125 m long OM4 fiber and then converted to electric current by a PIN photodiode. Finally, the signal was acquired by a Real-Time Oscilloscope (RTO) with a sampling frequency equal to f_{DAC}=200 GSample/s. The DSP internal parameters are reported in Table 1. The multi-rate channel has been trained using an entire period of the transmitted sequence (2^{15} symbols). The E2E optimization was then performed using 1e5 PAM-M symbols, generated from a second random sequence different and decorrelated from the transmitted one.

Results and discussion

After training different DPDs to be used for several bit rates and for both PAM-4 and PAM-8 modulations, we experimentally measure the resulting BER on a third PAM-M sequence with a period equal to 2^{16} symbols (completely decorrelated from the other sequences to avoid NN overfitting issues). For comparison, we also measured the performances when the DPD is not applied, shaping the symbols with a Gaussian filter (with $f_{3dB}= 0.75 \cdot D$). We use the maximum allowed $DAC\text{-}vpp$=700 mV in both cases, training the FFE for 2e5 symbols and testing 5e5 symbols over 5 different measurements. Moreover, to get the best performances without DPD, we measured BER also when reducing the swing to 500 mV in the non-pre-distorted scenario. In fact, we experimentally observed that for higher baud rates, the nonlinear distortions cause the main impairments to the signal: reducing *DAC-Vpp* thus gives advantages when DPD is not applied. In Fig. 2, the resulting BER versus Bit Rate is reported for the three considered scenarios, showing that our proposed nonlinear DPD method outperforms in all circumstances the best performances achieved without DPD. Using PAM-4, the DPD gives a performance gain of nearly 14 Gbps for a BER=1e-3 and more than 17 Gbps for a BER=1e-2. Moreover, using PAM-8, it is possible to transmit at more than 150 Gbps for a BER=1e-2, where the best non-pre-distorted scenario can reach up to 140 Gbps. To better observe the effects of the DPD on the transmitted signals, in Fig. 3 we show the distribution of the

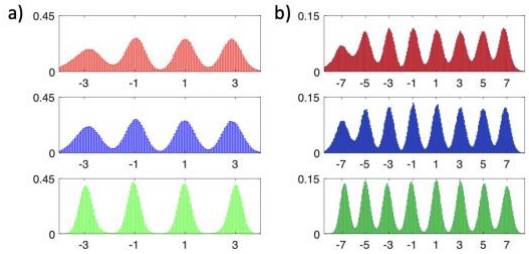

Figure 3: Distributions of the received symbols after FFE for: (a) PAM-4; (b) PAM-8. Red histograms: no DPD, maximum voltage swing; Blue histograms: no DPD, reduced swing; Green histograms: DPD applied.

received symbols after the FFE in the three scenarios for both PAM-4 and PAM-8 at D=56 GBaud and D=46 GBaud, respectively. It can be seen that the distortions nonlinearly broaden the Gaussian distributions around their nominal PAM-M levels when the DPD is not applied, and this especially affects the two lower PAM-M levels. Reducing the $DAC\text{-}vpp$ only mitigates this nonlinear effect. Applying instead the nonlinear DPD effectively overcomes this issue, diminishing the overall noise+Inter-symbol Interference impairments: the distributions around nominal levels in Fig. 3 are in fact significantly reduced when applying the DPD.

Discussion and Conclusion

In this paper we illustrated a general optimization method, based on an E2E system working at several sampling rates, to train a fractional sps nonlinear DPD natively fulfilling the amplitude constraints of an optical VCSEL-MMF IM-DD link and the multiple sampling rates. The proposed method enables in our experiments to reach 150Gbit/s (PAM-8 after 125m OM4 at BER=10^{-2}) using a 20-GHz VCSEL. In the latter condition, the DPD FFNN natively works at 1.84 sps ratio, performing ~850 multiplications per symbol (mps): as a comparison, using a more conventional 2-sps Volterra Nonlinear DPD with the same memory and order equal to 2 or 3 would require respectively 1150 mps or 13800 mps[20].

Acknowledgements: *This work was carried out under a research contract with Cisco Photonics. We also acknowledge the PhotoNext initiative at Politecnico di Torino (http://www.photonext.polito.it/) and its laboratory, where all experiments have been performed.*

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, https://www.ieee802.org/3/NGMMF/public/Jan18/young_NGMMF_01a_jan18.pdf, last accessed on 5 May 2022.

[2] H. Zhou, Y. Li, Y. Liu, L. Yue, C. Gao, W. Li, J. Qiu, H. Guo, X. Hong, Y. Zuo, J. Wu, “Recent Advances in Equalization Technologies for Short-Reach Optical Links Based on PAM4 Modulation: A Review,” *Applied Sciences*. 2019; 9(11):2342. DOI: 10.3390/app9112342

[3] Y. Yu, T. Bo, Y. Che, D. Kim, and H. Kim, "Low-Complexity Equalizer Based on Volterra Series and Piecewise Linear Function for DML-Based IM/DD System," *2020 Optical Fiber Communications Conference and Exhibition (OFC)*, 2020, pp. 1-3.

[4] Y. Gao, C. Yang, J. Wang, X.Qin, H. Guo, X. Zhang, C. Shen, H. Li, Z. Chen, and C. J. Chang-Hasnain, "288 Gb/s 850 nm VCSEL-based Interconnect over 100 m MMF based on Feature-enhanced Recurrent Neural Network," *2022 Optical Fiber Communications Conference and Exhibition (OFC)*, 2022, pp. 01-03.

[5] J. Zhang, P. Gou, M. Kong, K. Fang, J. Xiao, Q. Zhang, X. Xin, and J.Yu, "PAM-8 IM/DD Transmission Based on Modified Lookup Table Nonlinear Predistortion," in *IEEE Photonics Journal*, vol. 10, no. 3, pp. 1-9, June 2018, Art no. 7903709, DOI: 10.1109/JPHOT.2018.2828869.

[6] Z. He, K. Vijayan, M. Mazur, M. Karlsson, and J. Schröder, "Look-up Table based Pre-distortion for Transmitters Employing High-Spectral-Efficiency Modulation Formats," *2020 European Conference on Optical Communications (ECOC)*, 2020, pp. 1-4, DOI: 10.1109/ECOC48923.2020.9333231.

[7] V. Shivashankar, C. Kottke, V. Jungnickel and R. Freund, "Investigation of Linear and Nonlinear Pre-Equalization of VCSEL," *Broadband Coverage in Germany; 11. ITG-Symposium*, 2017, pp. 1-5.

[8] Z. He, J. Song, C. Häger, K. Vijayan, P. Andrekson, M. Karlsson, A.G. i Amat, H. Wymeersch, and J. Schröder, "Symbol-Based Supervised Learning Predistortion for Compensating Transmitter Nonlinearity," *2021 European Conference on Optical Communication (ECOC)*, 2021, pp. 1-4, DOI: 10.1109/ECOC52684.2021.9605892.

[9] V. Bajaj, M. Chagnon, S. Wahls and V. Aref, "Efficient Training of Volterra Series-Based Pre-distortion Filter Using Neural Networks," *2022 Optical Fiber Communications Conference and Exhibition (OFC)*, 2022, pp. 1-3.

[10] V. Bajaj, F. Buchali, M. Chagnon, S. Wahls and V. Aref, "Deep Neural Network-Based Digital Pre-Distortion for High Baudrate Optical Coherent Transmission," in *Journal of Lightwave Technology*, vol. 40, no. 3, pp. 597-606, 1 Feb.1, 2022, DOI: 10.1109/JLT.2021.3122161.

[11] G. Paryanti, H. Faig, L. Rokach and D. Sadot, "A Direct Learning Approach for Neural Network Based Pre-Distortion for Coherent Nonlinear Optical Transmitter," in *Journal of Lightwave Technology*, vol. 38, no. 15, pp. 3883-3896, 1 Aug.1, 2020, DOI: 10.1109/JLT.2020.2983229.

[12] J. Song, Z. He, C. Häger, M. Karlsson, A.G. i Amat, H. Wymeersch, and J. Schröder, "Over-the-fiber Digital Predistortion Using Reinforcement Learning," *2021 European Conference on Optical Communication (ECOC)*, 2021, pp. 1-4, DOI: 10.1109/ECOC52684.2021.9605972.

[13] B. Karanov, M. Chagnon, F. Thouin, T. A. Eriksson, H. Bu¨low, D. Lavery, P. Bayvel, and L. Schmalen, "End-to-End Deep Learning of Optical Fiber Communications," *Journal of Lightwave Technology*, vol. 36, no. 20, pp. 4843-4855, 15 Oct.15, 2018, DOI: 10.1109/JLT.2018.2865109.

[14] Boris Karanov, Domaniç Lavery, Polina Bayvel, and Laurent Schmalen, "End-to-end optimized transmission over dispersive intensity-modulated channels using bidirectional recurrent neural networks," Opt. Express 27, 19650-19663 (2019), DOI: 10.1364/OE.27.019650.

[15] V. Neskorniuk, A. Carnio, V. Bajaj, D.Marsella, S. K. Turitsyn, J. E. Prilepsky, and V. Aref, "End-to-End Deep Learning of Long-Haul Coherent Optical Fiber Communications via Regular Perturbation Model," *2021 European Conference on Optical Communication (ECOC)*, 2021, pp. 1-4, DOI: 10.1109/ECOC52684.2021.9605928.

[16] V. Aref and M. Chagnon, "End-to-End Learning of Joint Geometric and Probabilistic Constellation Shaping," *2022 Optical Fiber Communications Conference and Exhibition (OFC)*, 2022, pp. 1-3.

[17] T. O’Shea and J. Hoydis, "An Introduction to Deep Learning for the Physical Layer," *IEEE Transactions on Cognitive Communications and Networking*, vol. 3, no. 4, pp. 563-575, Dec. 2017, DOI: 10.1109/TCCN.2017.2758370.

[18] E. A. Wan, "Temporal backpropagation for FIR neural networks," *1990 IJCNN International Joint Conference on Neural Networks*, 1990, pp. 575-580 vol.1, DOI: 10.1109/IJCNN.1990.137629.

[19] C. M. Bishop, "Training with Noise is Equivalent to Tikhonov Regularization," in *Neural Computation*, vol. 7, no. 1, pp. 108-116, Jan. 1995, DOI: 10.1162/neco.1995.7.1.108.

[20] J. Tsimbinos and K. V. Lever, "The computational complexity of nonlinear compensators based on the Volterra inverse," in *Proceedings of 8th Workshop on Statistical Signal and Array Processing*, 1996, pp. 387-390, DOI: 10.1109/SSAP.1996.534897.