Single-channel 1.61 Tb/s Optical Coherent Transmission Enabled by Neural Network-Based Digital Pre-Distortion

Vinod Bajaj[(1,2)], Fred Buchali[(1)], Mathieu Chagnon[(1)], Sander Wahls[(2)], Vahid Aref[(1)]

[(1)] Nokia Bell Labs, Lorenzstr. 10, 70435 Stuttgart, Germany, V.Bajaj-1@tudelft.nl
[(2)] Delft Center for Systems and Control, Delft University of Technology, Delft, The Netherlands

Abstract

We propose a novel digital pre-distortion (DPD) based on neural networks for high-baudrate optical coherent transmitters. We demonstrate experimentally that it outperforms an optimized linear DPD giving a 1.2 dB SNR gain in a 128GBaud PCS-256QAM single-channel transmission over 80km of standard single-mode fiber resulting in a record 1.61 Tb/s net data rate.

Introduction

The next generation coherent optical transceivers target single-channel capacities beyond 600 Gb/s for data-center interconnect and metro-haul applications. These transceivers need to operate at very high symbol rates (>90 GBaud) and use high order modulation formats, e.g. 64-QAM[1]–[4]. Such demanding requirements are very sensitive to linear and nonlinear transmitter impairments limiting the transmit signal to noise power ratio (SNR). Transmitter distortions are usually pre-compensated by linear digital pre-distortion (DPD) mitigating linear effects only. In contrast, a nonlinear DPD can compensate both linear and nonlinear effects resulting larger transmit SNR.

The most studied nonlinear DPDs are based on Volterra series (or its simplified versions such as memory polynomials) which have been investigated intensively – for linearizing radio frequency (RF) amplifiers[5]–[9] and for pre-equalization of optical transmitter components[10]–[15].

Another type of DPD is based on neural networks (NN). Proposed first in 1980s[16],[17], these DPDs have received more attention recently[18]–[25]. To linearize RF amplifier, different structures based on feed-forward NN (FFNN)[20], time-delay NNs (TDNNs)[18],[23],[26], convolutional NNs (CNNs)[24] were proposed. It was recently shown for an RF amplifier that adding a residual NN (ResNet) structure in TDNNs improves the performance and training rate[19].

For optical coherent transmitters, memoryless simple FFNNs were proposed to pre-compensate Mach-Zehnder modulator[27] or a simulated low-resolution digital to analog converter (DAC)[22]. In a recent study[21], a DPD based on recurrent NN (RNN) was proposed in which the performance is evaluated in simulation with 4 samples/symbol input signal. The known problems of RNNs are their complexity of training and slow convergence.

In this paper, we propose a new NN-based DPD with a simple structure that includes CNNs and FFNNs optimized for the responses of an actual transmitter. Since the linear effects are dominant, we find out that a residual bypass of the nonlinear layers (ResNet-like connection) improves the performance significantly. Moreover, we apply a customized activation function named soft-DAC which models the DAC and minimizes the distortion from clipping and quantization.

Our NN-based DPD was trained and tested in lab experiments for single-channel PCS-256-QAM 128 GBaud transmission over standard single mode fiber (SSMF). Our results show that it outperforms optimized linear DPD giving more than 1.2 dB SNR gain. Accordingly, it provides additional 90 Gb/s gain in net rate and enables to achieve a record single-channel net rate beyond 1.61 Tb/s over 80 km SSMF.

The Proposed Neural Network Based DPD

The architecture of the proposed NN-based DPD is depicted in Fig. 1 in which one dimensional (1D) CNNs pre-compensate the memory effects and in the middle part, a fully connected FFNN takes care of the nonlinear effects. The size of each layer is identified in Fig. 1. In the FFNN, the first layer consists of short 1D-CNNs and the following 3 layers have leaky rectified linear unit (Leaky ReLU) activation functions while the last

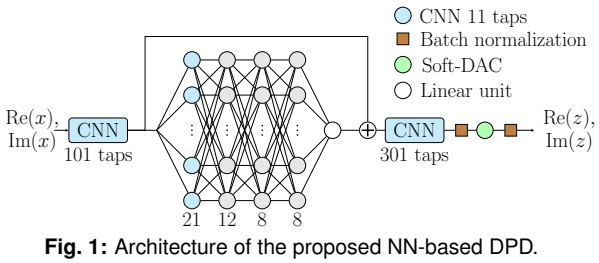

Fig. 1: Architecture of the proposed NN-based DPD.

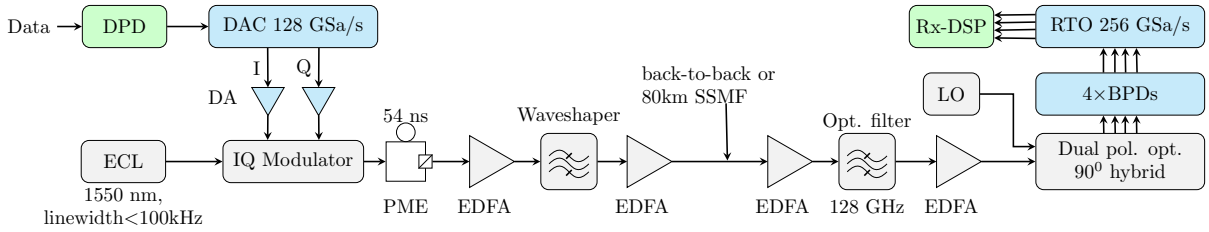

Fig. 2: Experimental setup. ECL: external cavity laser, PME: polarization multiplexing emulator, BPD: balanced photo-diodes

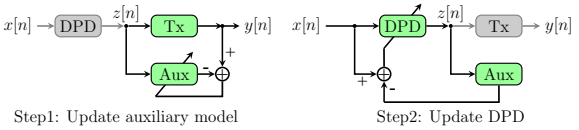

Fig. 3: Direct learning architecture; Green blocks are in use.

layer contains a single linear unit. Since the linear effects are dominant, we find out that an additional shortcut bypassing the FFNN improves the performance and speeds up the training process.

A customized activation unit, named Soft-DAC, is used as the output unit to model the DAC with m bits resolution. This unit should quantize its input uniformly to 2^m discrete levels. However, its input signal should be scaled properly to minimize the total distortions of clipping and quantization. To avoid working with discrete levels directly, the Soft-DAC unit softens the output values using a hyper parameter s(softening factor). In a nutshell, the Soft-DAC unit is implemented as a piece-wise linear function with 2^{m+1} linear pieces. The slope of pieces is alternatively s and $\frac{1}{s}$. By changing s from 1 to 0, the behavior of the Soft-DAC changes from a clipping-only operation to a clipping and $m-$bits quantization operation. During the training, s is changed manually from 1 to 0. This way, the optimal input scaling is trained and the final outputs of Soft-DAC are discrete levels.

As the NNs are designed primarily for real valued data, the complex input sequence $x[n] = \mathrm{Re}(x[n]) + j\mathrm{Im}(x[n])$ is treated as two real-valued sequences $\mathrm{Re}(x[n])$ and $\mathrm{Im}(x[n])$ which are precompensated separately with two disjoint NNs.

To train the DPD, we used direct learning architecture (DLA) depicted in Fig. 3. For this purpose, another NN with similar structure (without Soft-DAC unit and CNNs are swapped) was used to serve as auxiliary model. The training is iterative and the next two steps are repeated sequentially until the trained DPD converges: First the auxiliary NN is trained. The training data includes predistorted signal $\boldsymbol{z}[n] = [\mathrm{Re}(z[n]), \mathrm{Im}(z[n])]$ and received signal $\boldsymbol{y}[n] = [\mathrm{Re}(y[n]), \mathrm{Im}(y[n])]$ as inputs and targets. Then, the DPD is trained to equalize the outputs of the auxiliary model with the desired input of DPD $\boldsymbol{x}[n] = [\mathrm{Re}(x[n]), \mathrm{Im}(x[n])$. For both trainings, we used back-propagation technique with loss function of mean square error and with Adam optimizer[28].

Experimental Setup

The experimental setup is shown in Fig. 2. As the same setup was used and detailed in the recent paper[2], we refer readers to[2] for more details.

We used probabilistic constellation shaping (PCS) 256-QAM signal, shaped using Maxwell-Boltzmann distribution of entropy 7.5 bits/symbol. At the transmitter, a random sequence of 2^{15} PCS 256-QAM symbols was generated. Without pulse shaping (i.e. 1 sample/symbol), it passes to the DPD and is then loaded to the DACs, amplified using driver amplifiers (DAs) and fed to lithium niobate ($LiNbO_3$) IQ modulator. The Micram DAC-5 has 4 effective number of bits (ENOB) at Nyquist frequency, 128 GSa/s sampling rate and 24 GHz 3dB-bandwidth. While, the DAs and the I/Q modulator have about 60 GHz and 41 GHz 3dB-bandwidth respectively.

A polarization multiplexing emulator (PME) is used to generate a dual polarization optical signal. The signal is then amplified using an Erbium doped fiber amplifier (EDFA). A Finisar Waveshaper was used to compensate for the low-pass response of the IQ modulator by flattening the optical spectrum of the signal (see[1],[2] for details). In either back-to-back scenario or 80 km SSMF transmission, the received signal was filtered by an optical filter, then amplified by an EDFA and mixed with the local oscillator (LO) in an optical 90° hybrid. Four balanced photo-diodes (BPDs) were used to convert the optical signal into electrical signals which were sampled at 256 GSa/s by a Keysight real-time oscilloscope (RTO). The state-of-the-art DSP, detailed in[1], was applied offline to retrieve the sent data.

Results

We first consider a back-to-back scenario. We optimize a linear DPD using the least square algorithm and observed that the performance stops improving when about 440 taps were used (many of the taps were zero). This long filter length was

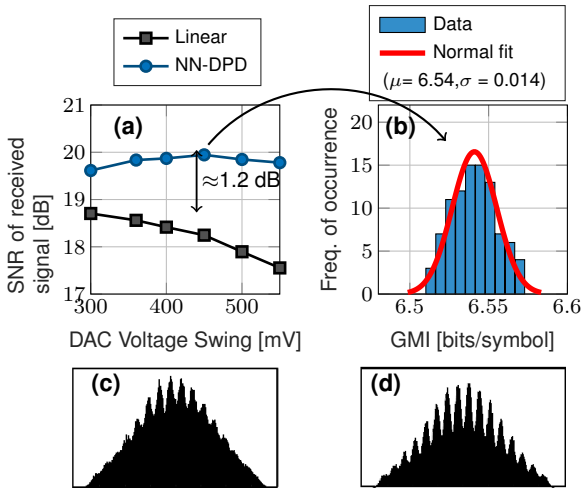

Fig. 4: (a) SNR vs DAC voltage for linear and NN-based DPD, (b) histogram of GMI values for 100 random sequences @$450mV$ with NN-based DPD. Histogram of one tributary of the received signal for (c) linear DPD @ 300 mV and (d) NN-based DPD @ 450 mV.

required due to signal reflections that were generated by rather long RF connections in the lab setup. For fairness of comparison, we set the total number of taps of CNNs in the NN-based DPD to around 440. Accordingly, the NN-based DPD is trained iteravely as explained before. A single sequence of 2^{15} PCS 256-QAM symbols was used for training but upto 100 other independently generated sequences were used for testing.

Both DPDs were first trained separately for different DAC output voltages. Increasing the DAC output voltage increases the driver voltage swing, and thus, the signal power over the constant noise of the transmitter. However, the transmitter nonlinear effects are also enhanced. Fig. 4(a) shows the received SNR of the trained DPDs. We observe that the SNR for the linear DPD degrades by increasing the DAC voltage while the NN-based DPD is quite effective in mitigating nonlinear effects. It gives 1.2 dB SNR gain with respect to the linear DPD at 50% higher DAC voltage. The SNR gain can be explained from Fig. 4(c) and Fig. 4(d) in which we plotted the histogram of a single tributary of the received symbols (PCS 16-ASK). The nonlinear effects are quite visible in Fig. 4(c) for linear DPD whereas they are much less significant in Fig. 4(d) for NN-based DPD.

To assure that the trained NN-based DPD is pattern independent, it's performance is tested with 100 independently generated sequences. For each sequence, we computed the generalized mutual information (GMI) of the received symbols. Fig. 4(b) shows the empirical distribution of GMI of these sequences at 450 mV DAC voltage. We observe that the distribution has very

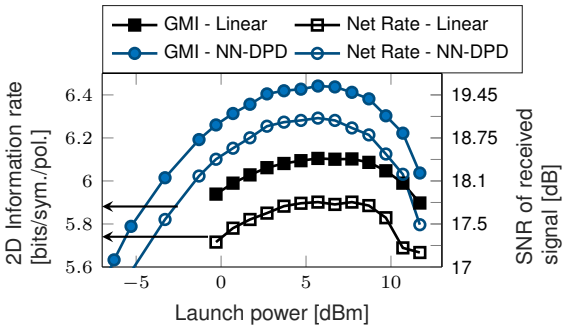

Fig. 5: SNR, GMI and Net information rate at different launch powers for 80 km transmission over SSMF.

small standard deviation confirming that the NN-based DPD is pattern independent. Note that small GMI fluctuation is natural.

After training both DPDs at their optimal DAC voltage in back-to-back setting, these predistortions were used in transmission experiments over 80 km of SSMF. Fig. 5 shows the received SNR (right axis) and the corresponding GMI (left axis) in terms of launch power. The SNR is maximized for both DPDs at about 6 dBm launch power. We observe again that the NN-based DPD outperforms the linear DPD with about 1.2 dB SNR gain. Accordingly, the GMI (in bits/symbol/polarization) is increased from 6.10 for the linear DPD to 6.44 for the NN-based DPD. Note that a larger SNR gain can be obtained by training during fiber transmission.

Fig. 5 shows the net information rate (IR) after applying high performance soft-decision forward error correction codes. We used a family of 130 optimized spatially coupled LDPC codes[1],[29] with variable overheads ranging from 3% to 100%. At each launch power, a code with the smallest overhead was chosen which could decode the experimental traces error free. As shown in Fig. 5, the net IR is increased from 5.94 to 6.29 bits/symbol/polarization. This result corresponds to the increase of net rate from the previous record[2] of 1.52 Tb/s to the new record of 1.61 Tb/s for single-channel transmission.

Conclusion

We proposed a novel neural network-based DPD which was trained and verified experimentally with a state-of-the-art 128 GBuad optical coherent transmission setup. We showed that our novel DPD outperforms linear DPDs with 1.2 dB SNR gain resulting a record single channel net data rate of 1.61 Tb/s over 80 km of SSMF.

Acknowledgements

This work was carried out under the EU Marie Skłodowska-Curie project FONTE (No. 766115).

References

[1] F. Buchali, K. Schuh, R. Dischler, M. Chagnon, V. Aref, *et al.*, "DCI field trial demonstrating 1.3-Tb/s single-channel and 50.8-Tb/s WDM transmission capacity", *J. Lightw. Technol.*, vol. 38, no. 9, pp. 2710–2718, 2020.

[2] F. Buchali, V. Aref, M. Chagnon, K. Schuh, H. Hettrich, *et al.*, "1.52 tb/s single carrier transmission supported by a 128 gsa/s sige dac", in *2020 Optical Fiber Communications Conference (OFC)*, 2020, Th4C–2.

[3] H. Sun, M. Torbatian, M. Karimi, R. Maher, S. Thomson, *et al.*, "800g dsp asic design using probabilistic shaping and digital sub-carrier multiplexing", *J. Lightw. Technol.*, 2020.

[4] *Ciena wavelogic 5, 800g*. [Online]. Available: `https://www.ciena.com/insights/articles/Ciena-unveils-WaveLogic-5-800G-and-so-much-more.html`.

[5] C. Eun and E. J. Powers, "A new Volterra predistorter based on the indirect learning architecture", *IEEE Trans. Signal Process.*, vol. 45, no. 1, pp. 223–227, 1997.

[6] Y. H. Lim, Y. S. Cho, I. W. Cha, and D. H. Youn, "An adaptive nonlinear prefilter for compensation of distortion in nonlinear systems", *IEEE Trans. Signal Process.*, vol. 46, no. 6, pp. 1726–1730, 1998.

[7] J. Kim and K. Konstantinou, "Digital predistortion of wideband signals based on power amplifier model with memory", *Electronics Letters*, vol. 37, no. 23, pp. 1417–1418, 2001.

[8] D. R. Morgan, Z. Ma, J. Kim, M. G. Zierdt, and J. Pastalan, "A generalized memory polynomial model for digital predistortion of rf power amplifiers", *IEEE Trans. Signal Process.*, vol. 54, no. 10, pp. 3852–3860, 2006.

[9] D. Zhou and V. E. DeBrunner, "Novel adaptive nonlinear predistorters based on the direct learning algorithm", *IEEE Trans. Signal Process.*, vol. 55, no. 1, pp. 120–133, 2006.

[10] P. W. Berenguer, M. Nölle, L. Molle, T. Raman, A. Napoli, C. Schubert, and J. K. Fischer, "Nonlinear digital pre-distortion of transmitter components", *J. Lightw. Technol.*, vol. 34, no. 8, pp. 1739–1745, 2015.

[11] H. Faig, Y. Yoffe, E. Wohlgemuth, and D. Sadot, "Dimensions-reduced volterra digital pre-distortion based on orthogonal basis for band-limited nonlinear opto-electronic components", *IEEE Photon. J.*, vol. 11, no. 1, pp. 1–13, 2019.

[12] R. Elschner, R. Emmerich, C. Schmidt-Langhorst, F. Frey, P. W. Berenguer, *et al.*, "Improving achievable information rates of 64-gbd pdm-64qam by nonlinear transmitter predistortion", in *2018 Optical Fiber Communications Conference (OFC)*, 2018, pp. 1–3.

[13] G. Khanna, B. Spinnler, S. Calabrò, E. De Man, and N. Hanik, "A robust adaptive pre-distortion method for optical communication transmitters", *IEEE Photon. Technol. Lett.*, vol. 28, no. 7, pp. 752–755, 2015.

[14] Y. Yoffe, G. Khanna, E. Wohlgemuth, E. de Man, B. Spinnler, *et al.*, "Low-resolution digital pre-compensation enabled by digital resolution enhancer", *J. Lightw. Technol.*, vol. 37, no. 6, pp. 1543–1551, 2019.

[15] A. Napoli, P. W. Berenguer, T. Rahman, G. Khanna, M. M. Mezghanni, *et al.*, "Digital pre-compensation techniques enabling high-capacity bandwidth variable transponders", *Opt. Comm.*, vol. 409, pp. 52–65, 2018.

[16] D. Psaltis, A. Sideris, and A. A. Yamamura, "A multilayered neural network controller", *IEEE control systems magazine*, vol. 8, no. 2, pp. 17–21, 1988.

[17] A. Bernardini, M. Carrarini, and S. De Fina, "The use of a neural net for copeing with nonlinear distortions", in *20th Europ. Microw. Conf.*, vol. 2, 1990, pp. 1718–1723.

[18] T. Gotthans, G. Baudoin, and A. Mbaye, "Digital predistortion with advance/delay neural network and comparison with volterra derived models", in *IEEE 25th Inter. Symp. Personal, Indoor, and Mobile Radio Comm. (PIMRC)*, 2014, pp. 811–815.

[19] Y. Wu, U. Gustavsson, A. Graell i Amat, and H. Wymeersch, *Residual neural networks for digital predistortion*, 2020. arXiv: `2005.05655`.

[20] C. Tarver, A. Balatsoukas-Stimming, and J. R. Cavallaro, "Design and implementation of a neural network based predistorter for enhanced mobile broadband", in *2019 IEEE Inter. Workshop Signal Process. Sys. (SiPS)*, 2019, pp. 296–301.

[21] G. Paryanti, H. Faig, S. L. Rokach, and D. Sadot, "A direct learning approach for neural network based predistortion for coherent nonlinear optical transmitter", *J. Lightw. Technol.*, 2020.

[22] M. Abu-Romoh, S. Sygletos, I. D. Phillips, and W. Forysiak, "Neural-network-based pre-distortion method to compensate for low resolution DAC nonlinearity", in *45th Europ. Conf. on Optical Comm. (ECOC 2019)*, 2019, pp. 1–4.

[23] R. Hongyo, Y. Egashira, T. M. Hone, and K. Yamaguchi, "Deep neural network-based digital predistorter for doherty power amplifiers", *IEEE Microw. Wireless Compon. Lett.*, vol. 29, no. 2, pp. 146–148, 2019.

[24] X. Hu, Z. Liu, X. Yu, Y. Zhao, W. Chen, *et al.*, "Convolutional neural network for behavioral modeling and predistortion of wideband power amplifiers", *arXiv preprint arXiv:2005.09848*, 2020.

[25] J. Sun, W. Shi, Z. Yang, J. Yang, and G. Gui, "Behavioral modeling and linearization of wideband rf power amplifiers using bilstm networks for 5g wireless systems", *IEEE Trans. Veh. Technol.*, vol. 68, no. 11, pp. 10 348–10 356, 2019.

[26] S. Boumaiza and F. Mkadem, "Wideband rf power amplifier predistortion using real-valued time-delay neural networks", in *2009 Europ. Microw. Conf. (EuMC)*, IEEE, 2009, pp. 1449–1452.

[27] M. Schaedler, M. Kuschnerov, S. Calabrò, F. Pittalà, C. Bluemm, and S. Pachnicke, "AI-based digital predistortion for IQ Mach-Zehnder modulators", in *2019 Asia Comm. and Photon. Conf. (ACP)*, 2019, pp. 1–3.

[28] D. P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, 2014. arXiv: `1412.6980`.

[29] L. Schmalen, V. Aref, J. Cho, D. Suikat, D. Rösener, and A. Leven, "Spatially coupled soft-decision error correction for future lightwave systems", *Journal of Lightwave Technology*, vol. 33, no. 5, pp. 1109–1116, 2014.