54.5 Tb/s WDM Transmission over Field Deployed Fiber Enabled by Neural Network-Based Digital Pre-Distortion

Bajaj, Vinod; Buchali, Fred; Chagnon, Mathieu; Wahls, Sander; Aref, Vahid

Publication date
2021

Document Version
Accepted author manuscript

Published in
Proceedings of the Optical Fiber Communications Conference and Exhibition, OFC 2021

Citation (APA)
Bajaj, V., Buchali, F., Chagnon, M., Wahls, S., & Aref, V. (2021). 54.5 Tb/s WDM Transmission over Field Deployed Fiber Enabled by Neural Network-Based Digital Pre-Distortion. In *Proceedings of the Optical Fiber Communications Conference and Exhibition, OFC 2021* Article 9489722 IEEE.

54.5 Tb/s WDM Transmission over Field Deployed Fiber Enabled by Neural Network-Based Digital Pre-Distortion

Vinod Bajaj[1,2], **Fred Buchali**[1], **Mathieu Chagnon**[1], **Sander Wahls**[2] **and Vahid Aref**[1]
[1] *Nokia Bell Labs, Lorenzstr. 10, 70435 Stuttgart, Germany*
[2] *Delft Center for Systems and Control, Delft University of Technology, 2628 CN Delft, Netherlands*
v.bajaj-1@tudelft.nl, firstname.lastname@nokia.com and s.wahls@tudelft.nl

Abstract: We demonstrate a record 54.5 Tb/s WDM transmission at 11.35 bit/s/Hz over 48 km of field-deployed SMF connecting business and academic parks enabled by a novel joint I-Q Neural Network-based transmitter digital pre-distortion technique. © 2021 The Author(s)

1. Introduction

The rapid growth in internet traffic is driving the need for high data rate optical fiber transmission systems. To increase the aggregate bit rate carried by a single fiber strand while minimizing the component count, these systems need to be operated at high symbol rates on a tight spectral grid using high-order modulation formats (HOMFs) while maintaining the highest possible signal-to-noise ratio (SNR) [1–4]. Transmitter digital pre-distortion (DPD) techniques enable the generation of signals with high integrity by mitigating undesired component responses. Usually, a linear DPD is used. In such cases, the uncompensated nonlinear response of components still adds distortions and limits the SNR. As HOMFs are sensitive to distortions, addressing these distortions at the transmitter in a way that is tailored to the system and its components is paramount to increase the information rate (IR).

DPD techniques based on neural networks (NN) [5] have recently received more attention [6–10], with notable simulation results in [8] by implementing NN-based DPD using direct learning architecture (DLA) for optical coherent systems. We recently demonstrated a NN-based DPD technique that independently mitigates the distortions of the "I" and "Q" tributaries on a single channel fiber-optic transmission system where an SNR improvement of up to 1.6 dB was achieved [12]. Another recent in-lab demonstration was of a total C-band capacity of 52.1 and 51.6 Tb/s over 80 km of TXF™ fiber and standard single mode fiber (SSMF), respectively [11].

In this paper, we extend our previous work [12] by developing a NN-based DPD technique to *jointly* pre-compensate for the distortions on the I and Q tributaries of the transmitter. The joint DPD is learned from "scratch" on a field-deployed 48 km link connecting industrial, business and academic parks in a metropolitan area. The DPD is applied to a fully populated C-band, consisting of 35 channels spanning a total bandwidth of 4.8 THz. It improves the SNR by about 1 dB compared to a linear DPD, and allows achieving a net capacity of 54.5 Tb/s.

2. Proposed Neural Network-Based Joint DPD

We depict the architecture of the proposed NN-based DPD in Fig. 1. The upper and lower parts belong to the I and the Q tributaries respectively and have identical structure. The DPD comprises three sections. Section (A) contains a 2×2 multiple-input multiple-output (MIMO) composed of linear uni-dimensional convolution filters (1D-CNNs) of 101 taps intended to mitigate the linear response of the IQ modulator (through and cross). Section (C) contains two linear (1D-CNN) filters of 301 taps meant to compensate the linear effects of the digital to analog converters (DACs) and their driver amplifiers (DAs). The two feed-forward NNs (FFNNs) in Section (B) are each composed of five dense fully connected layers. The first layer is a linear 11 tap 1D-CNN layer that fans out the digital signal. The following 3 layers have leaky rectified linear unit activations (ReLUs). The last layer

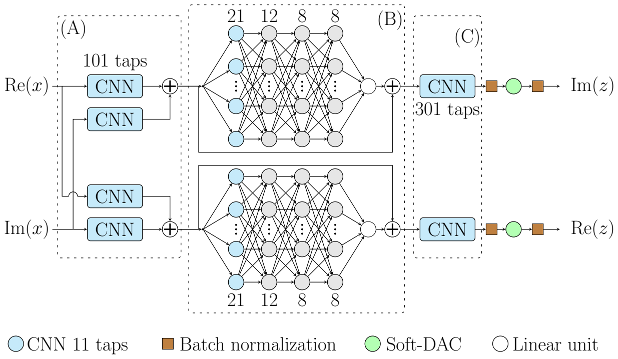
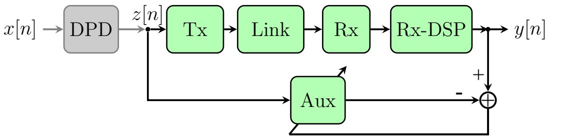
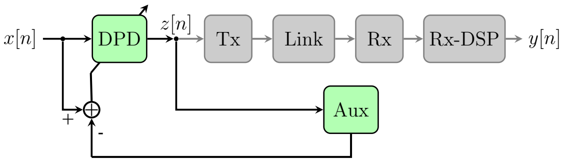


Fig. 1. Architecture of the proposed joint I-Q NN-based DPD.



Step1: Train auxiliary model (Aux)



Step2: Train DPD with the help of auxiliary model

Fig. 2. Direct learning architecture (DLA).

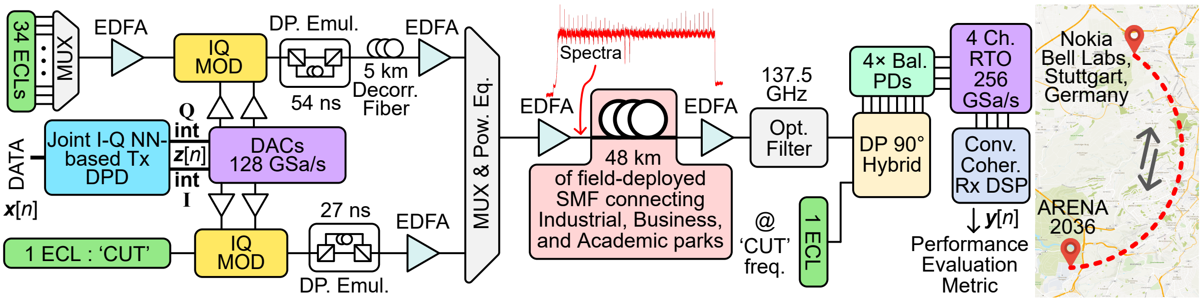


Fig. 3. Experimental setup of the WDM transmission system. Right: map shows the field trial area.

has a single linear output. The sizes of all layers are shown in Fig. 1. As the linear effects are dominant, adding a shortcut bypassing the FFNNs boost the performance and the training speed [12]. We also model the DAC using a customized activation unit, named Soft-DAC, that scales, clips and quantizes the signal in order to minimize the quantization distortion. We refer to [12] for more details.

As depicted in in Fig. 2, we used a DLA to train the DPD. We model first the "communication channel" using a mirror version "(C)→(B)→(A)" of the differentiable DPD structure, serving as an auxiliary channel model (Aux). The DPD and the Aux are trained iteratively until the SNR converges; about 10 iterations in our experiments. The input to the Aux is the DPD output $\boldsymbol{z}[n]$, which is also the input to the channel. A coherent receiver digital signal processing (Rx-DSP) is employed to stabilize the dynamic response of the channel. The Aux is trained to minimize the mean squared error (MSE) between its output and the soft symbol outputs of the Rx-DSP $\boldsymbol{y}[n]$. The DPD is trained by using ideal QAM symbols $\boldsymbol{x}[n]$ as its input and the gradients of the MSE between the Aux output and the desired output of the channel (i.e. ideal QAM symbols $\boldsymbol{x}[n]$), back propagated through the Aux. The both NNs are optimized using Adam algorithm by using data sequences of 2^{18} length at 1 sample per symbol (sps).

3. Field Trial Experimental Setup

The experimental setup, shown in Fig. 3, is configured for the WDM transmission of 35 channels equally spaced by 137.5 GHz at 128 Gsymbol/s. The transmitter has two IQ modulators: one for the channel under test (CUT) and one for the bulk modulation of the other 34 WDM channels. All channels employ an external cavity laser (ECL) of 20 kHz linewidth. The inverted and non-inverted outputs of two DAC modules sampling at 128 GSa/s are amplified before driving the two IQ-modulators [2]. Dual-Polarization (DP) is emulated using two separate delay and add interferometers. An SSMF of 5 km in the WDM branch decorrelates the channels. Both branches are amplified before combination via a Waveshaper, also serving as a per-channel power equalizer to compensate for the modulator's response. The 35 channel WDM optical signal is amplified and launched into a SSMF interconnecting an industrial and business area to an academic and research park around Stuttgart, Germany. The link spans 48 km but exhibits a total loss of 16 dB, comparable to DCI-type distances of 80 km. The Maxwell-Boltzmann probabilistic constellation shaping (PCS) of a 256-QAM format with an entropy of 7.5 bits/symbol is employed, as previously determined as a good choice for roughly 19 dB SNR [13].

At the receiver (Rx), the signal is first amplified before the CUT is isolated by a tunable optical filter. A local oscillator (LO), another ECL of 20 kHz linewidth, beats with the CUT through a DP 90° optical hybrid. A 4-channel real-time oscilloscope (RTO) digitizes the 4 balanced photo-detector outputs at 256 GSa/s. The Rx-DSP processes the 4 digitized signals in the following stages. First, the signal is re-sampled at 2 sps. Then chromatic dispersion is removed and timing errors are corrected. The signal polarizations are de-multiplexed by using a 2×2 complex valued MIMO equalizer updated by a multi-modulus algorithm (MMA). Frequency detuning correction and phase noise correction operations are subsequently applied. The residual signal distortions are then compensated by a real valued 4×4 MIMO equalizer [1]. For decoding, we used variable overhead SC-LDPC codes described in [1].

4. Results

Datasets obtained by repeated transmission of a 2^{15} long PCS 256-QAM pre-distorted pattern were used for training the NN-based DPD and the linear DPD. The DACs voltages were set to 300 mV and 470 mV for the linear and the NN-based DPD, respectively, which were found optimal in our previous study [12]. First, both DPDs were trained when the real valued MIMO equalizer was not included in the Rx-DSP. In this case, the I-Q impairments are not compensated by the Rx-DSP and the NN-based joint I-Q DPD learns to mitigate them at the Tx, providing a gain of 3 dB with respect to the disjoint I-Q linear DPD. Next, when the NN-based joint I-Q DPD is employed and the real-valued 4x4 MIMO stage is included in the Rx DSP stack, the SNR after the MIMO stage is 1 dB higher compared to when only a disjoint I-Q linear DPD is employed. This not only indicates the significant gains provided by the NN-based joint I-Q, but also that the joint I-Q DPD can learn to compensate for linear IQ cross-talk that would otherwise need to be compensated at the receiver using noisy receiver inputs.

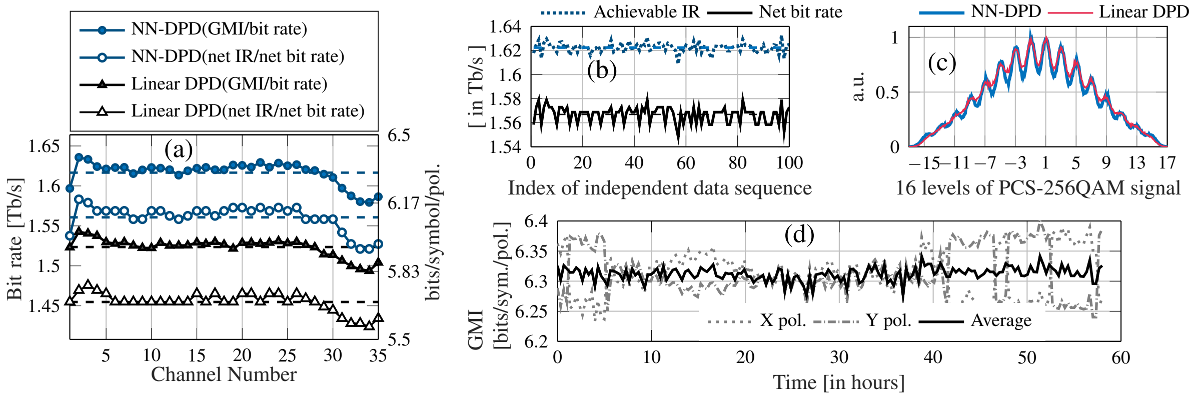


Fig. 4. (a) Data rates and information rates at different WDM channels (dashed lines: average values over the C-band), (b) achievable information rate and net bit rate over a test of 100 statistically independent sequences, (c) histograms of averaged 4-dimensions of the soft symbols output from the Rx-DSP and (d) GMI of the soft symbols output from the Rx-DSP plotted over time, when NN-DPD is employed at Tx.

Fig. 4(a) shows the achievable bit rate and the corresponding generalized mutual information (GMI) obtained when the CUT is swept through the 35 channels. We observed that the NN-DPD outperforms the linear DPD by giving a considerable gain in GMI. The average GMI (in bits/symbol/pol./carrier) is increased from 5.95 for the linear DPD to 6.31 for the NN-DPD. For the linear DPD, we achieved an average net bitrate of 1.45 Tb/s, which is slightly lower than what was obtained in [2]. We attribute this to the greater impairments from the 48 km of fibers criss-crossing the city compared to 80 km of spooled fiber in a lab. To verify that the NN did not learn the transmit pattern, we applied 100 different random sequences of PCS-256QAM symbols to the NN-DPD and measured their performance (Fig. 4(b)). The variance in achievable bitrate is very small, $<1\%$ (± 10 Gb/s) at most. The variations in GMI and net bitrate versus channel number are attributed to the EDFAs' gain profile. Moreover, the average decoding loss (in bits/symbol/pol./carrier) when employing the NN-DPD is reduced to 0.22 compared to 0.27 for the linear DPD, thanks to the more Gaussian-like distribution of the soft symbols at the Rx-DSP output ($y[n]$ in Fig. 2) when using the NN-DPD instead of the linear DPD (see Fig. 4(c)), since the FEC decoding algorithm conventionally assumes a Gaussian likelihood for the received symbols. The average net IR increases from 5.68 to 6.09 bits/symbol/pol./carrier, which corresponds to an average net bit rate of 1.55 Tb/s per channel. Overall, the 35 WDM channels result in a net transmission capacity record of 54.5 Tb/s, which is 7.2% higher than the capacity achieved by the linear DPD. Finally, we assessed the stability of the system performance when using a trained and fixed NN-based DPD. Fig. 4(d) shows the performance of 1 channel (out of 35) over the course of 3 days, with measurements taken every 20 minutes. The fluctuations are very small, even at a high GMI of 6.31 bits/symbol/pol., indicating that the learned Tx NN-based nonlinear DPD is temporally stable.

5. Conclusion

We demonstrated a field trial of 35 channel WDM transmission employing a novel joint I-Q neural networks-based DPD (NN-DPD). The NN-DPD adds a gain of 0.41 bits/symbol/pol. in net information rate and enabled us to achieve a record net rate of 54.5 Tb/s at 11.35 bit/s/Hz spectral efficiency. The performance of the NN-DPD evaluated over time shows its good stability.

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

References

1. F. Buchali 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 et al., "1.52 Tb/s single carrier transmission supported by a 128 GSa/s SiGe DAC", in OFC 2020, p. Th4C2, 2020.
3. "Ciena WaveLogic 5, 800G,", "https://www.ciena.com/insights/articles/Ciena-unveils-WaveLogic-5-800G-and-so-much-more.html".
4. H. Sun et al., "800G DSP ASIC Design Using Probabilistic Shaping and Digital Sub-Carrier Multiplexing," J. Lightw. Technol., 2020.
5. A. Bernardini et al., "The use of a neural net for copeing with nonlinear distortions," in Eur. Micr. Conf., vol. 2, pp. 1718–1723, 1990.
6. Y. Wu et al., "Residual Neural Networks for Digital Predistortion," arXiv: 2005.05655, 2020.
7. X. Hu et al., "Convolutional neural network for behavioral modeling and predistortion of wideband power amplifiers," arXiv, 2020.
8. G. Paryanti et al., "A direct learning approach for neural network based pre-distortion for coherent nonlinear optical transmitter," J. Lightw. Technol., vol. 38, no. 15, pp. 3883-3896, 2020.
9. J. Sun et al., "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 10348–10356, 2019.
10. M. Schaedler et al., "AI-based digital predistortion for IQ Mach-Zehnder modulators," Asia Comm. and Phot. Conf., ACP 2019.
11. F. Buchali et al., "52.1 Tb/s C-band DCI transmission over DCI distances at 1.49 Tb/s/lambda," ECOC 2020, Mo1E-4.
12. V. Bajaj et al., "Single-channel 1.61 Tb/s Optical Coherent Transmission Enabled by Neural Network-Based Digital Pre-Distortion," ECOC 2020, Tu1D-5.
13. F. Buchali et al., "128 GSa/s SiGe DAC implementation enabling 1.52 Tb/s single carrier transmission," J. Lightw. Technol., 2020.