

Neural-Network-Based Nonlinearity Equalizer for 128 GBaud Coherent Transceivers

Vladislav Neskorniuk^{1,2}, Fred Buchali¹, Vinod Bajaj^{1,3}, Sergei K. Turitsyn²,
Jaroslaw E. Prilepsky², Vahid Aref¹

¹Nokia Bell Labs, Lorenzstr. 10, 70435 Stuttgart, Germany

²Aston Institute of Photonic Technologies, Aston University, B4 7ET Birmingham, UK

³Delft Center for Systems and Control, Delft University of Technology, 2628 CD Delft, The Netherlands

v.neskorniuk@aston.ac.uk

Abstract: We propose an efficient neural-network-based equalization jointly compensating fiber and transceiver nonlinearities for high-symbol-rate coherent short-reach links. Providing about 0.9 dB extra SNR gain, it allows achieving experimentally the record single-channel 1.48 Tbps net rate over 240 km G.652 fiber. © 2021 The Author(s)

1. Introduction

High-symbol-rate coherent optical transceivers can be attractive for short-reach applications varying from data-center interconnect to metro systems [1]. Application of higher-order modulation formats can increase the information throughput of these systems, however, it requires higher signal-to-noise ratio (SNR) which causes detrimental nonlinear effects originating from both the fiber and transceiver components. Semi-analytical approaches to nonlinearity compensation (NLC), notably, Volterra series (e.g. [2]) and perturbation-based equalization (PBE) [3] are based on the simplified link models and, thus, have inherent limitations when used in the real-world systems. In the presence of mixed and complex impairments, application of neural networks has shown promising performance in modeling, compensation or optimization of optical systems (see, e.g. [4–7]). Although complex equalization can be learned by sophisticated black-box neural-network (NN)-based approaches [5], using “domain knowledge” helps in developing efficient NN-based equalizers (see, e.g. [6]).

In this paper, we propose an algorithm for the joint equalization of both transmitter (TX) hardware and fiber nonlinear distortions at the receiver (RX) with the focus on high baud-rate systems. The proposed method is based on a sequence of an adaptive PBE, intended for the channel nonlinearities, followed by a deep neural network (DNN) that equalizes both the TX nonlinearities and their interaction with the channel ones. Compared to mainstream TX-based digital pre-distortion (DPD) approaches to mitigate the transceiver distortions (reviewed in [7]), the proposed equalizer does not require an iterative feedback loop for training and so can be easily integrated with the existing systems. Moreover, RX based equalization doesn’t affect the peak-to-average power ratio (PAPR) of the transmitted signal, thus easing the transmission. The other published RX-based nonlinearity equalizers [5] are mostly black-box approaches, having relatively low performance-to-complexity ratio, or are aimed only at fiber nonlinearity compensation and, therefore, not optimal in the presence of considerable device impairments (e.g. [6]). Furthermore, to the best of our knowledge, none of these equalizers was successfully implemented in high-symbol-rate links, having specific transmission features and impairments.

2. Proposed architecture

The proposed algorithm (Fig. 1) aims to compensate both Kerr and transceiver nonlinearity and is applied in baseband to a symbol sequence already processed by a conventional receiver-based DSP (NLC block in Fig. 2). It is formed by two jointly optimized sequential stages. The first stage is more dedicated for fiber nonlinearity, while the second stage takes care of residual nonlinear effects, including transceiver nonlinearity.

The first stage is based on the perturbation-based equalization (PBE) [3], a well-studied Kerr nonlinearity compensation technique. It approximates fiber nonlinear distortions by a weighted sum of cubic terms. Let

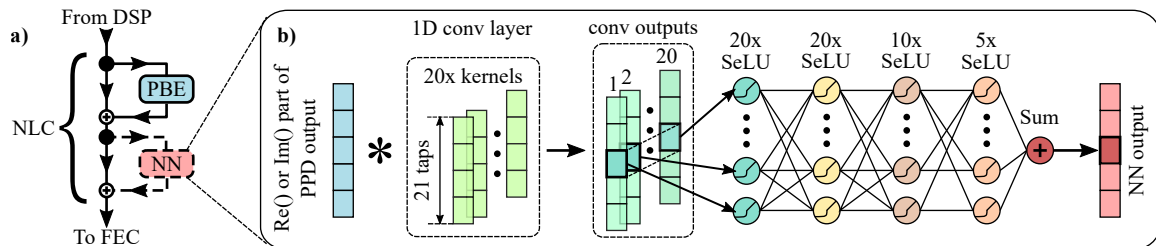


Fig. 1. Scheme of the NLC algorithms considered in this article: conventional PBE (solid lines in (a)) and the joint algorithm formed by the PBE being followed by the NN (b) (solid and dashed lines in (a)).

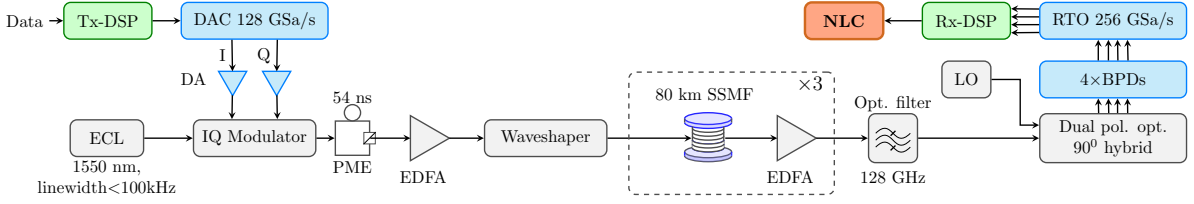


Fig. 2. Experimental setup of the studied transmission system. The detailed description of the setup is given in [1].

$U_k = u_{h/v}(t_0 + kT)$ be the k -th adjacent symbol of $u_{h/v}(t_0)$ in the received symbol sequence of the filtered horizontal (h) or vertical (v) polarization. T denotes the symbol period. For a given symbol $u_{h/v}(t_0)$, the PBE computes the perturbative correction $\Delta u_{h/v}(t_0) = \sum_{m,n} C_{m,n} (U_m U_{m+n}^* U_n)$ with the trained coefficients $C_{m,n} \in \mathbb{C}$. Here, m and n represent the symbol indices with respect to the symbol of interest U_0 . The number of terms makes a tradeoff between performance and complexity. Following the analytical insights in [3] and trying various ranges of m and n on experimental data, we constrained $|m| \leq 30$, $|n| \leq \min\{48/|m|, 30\}$, resulting in 434 terms. Note that the conventional PBE [3] also considers cross-polarization correction terms. However, we neglected them as they brought no significant performance gain on our experimental data.

While the first stage is based on analytical perturbative models, the second stage is a hand-crafted neural network (NN) (Fig. 1b). Inspired by the NN in [7], it aims to compensate both transceiver impairments and the residual nonlinear effects of the channel. The NN has 5 layers. The first layer is a uni-dimensional (1D) convolution with $20 \times$ kernels of 21 taps each and the nonlinear activation function of scaled exponential linear unit (SeLU) [8]. It is followed by 3 dense layers of size 20, 10, and 5 units with SeLU activations too. The output layer is a simple linear one. Like in [7], we added a bypass connection between input and output layers improving the total performance.

Each in-phase (I) or quadrature-phase (Q) tributary of a signal polarization uses a separately trained equalizer, resulting in four equalizers for a dual-polarized (DP) signal. Each equalizer was trained in two steps to improve the convergence. First, both stages, i.e., PBE and NN, were trained independently by minimizing the mean-squared error (MSE) loss using ordinary least-squares and Adam [9] (with a mini-batch size of 10^3) optimizers, respectively. Next, both stages were connected in series and jointly optimized by the same Adam algorithm.

3. Experimental setup

We considered a single-channel dual-polarized (DP) transmission at 128 Gbaud using either a probabilistically shaped (PCS)-256QAM or uniform 64QAM signal. The PCS-256QAM format was shaped with Maxwell-Boltzmann distribution of entropy 7.5 bits/symbol. The experimental setup is shown in Fig. 2. We briefly explain the setup and refer the readers for further details to [1], where the same setup was used. After applying a linear pre-emphasis filter, the signal is loaded to the DACs operating at 128 GSa/s with an effective number of bits (ENOB) of 4 at 64 GHz and 24 GHz 3-dB bandwidth. The 'I' and 'Q' tributary channels from the DACs are amplified using driver amplifiers (DAs) and fed to a single-polarization IQ-modulator. The DAs and the modulator have 60 GHz and 41 GHz 3dB-bandwidths, respectively. A polarization multiplexing emulator (PME) generates a DP signal by split and add with ≈ 54 ns delay. This DP signal is then amplified by an erbium-doped fiber amplifier (EDFA) and filtered using a Finisar wave-shaper to flatten the optical spectrum. We considered the transmission over 3 spans of 80 km G.652 standard single mode fiber (SSMF) with each span followed by an EDFA.

At the receiver, the optical signal was amplified, filtered, and applied to an optical 90° -hybrid followed by balanced photodiodes (BPDs) of 100 GHz bandwidth. An 80 GHz real-time oscilloscope (RTO) samples and acquires the four resulting waveforms at 256 GSa/s. The DSP was applied offline including the DSP blocks of chromatic-dispersion compensation, 2×2 MIMO polarization demultiplexing, and data-aided carrier phase recovery. The residual signal distortions are then compensated by a real valued 4×4 MIMO equalizer.

4. Performance assessment

We compared the performance of the proposed NN-based equalizer (joint NLC) with the two other cases: when PBE is applied after the linear RX-DSP (PBE only) and the benchmark case with no NLC being applied after linear RX-DSP (Linear DSP). Two random sequences of 500k PCS-256QAM and uniform 64QAM symbols were independently generated, transmitted, and processed by the RX-DSP to generate the training and testing datasets. PCG64 [10] was used as a pseudorandom number generator. We transmitted these sequences at different power levels to find the optimal ones for the considered algorithms.

Fig. 3 shows the resulting performance. We can see that applying PBE only led to a small SNR gain ≈ 0.1 dB. Nonetheless, PBE gain was higher for larger launch powers, where the fiber behaves more nonlinear. By applying the joint NLC, the SNR gain was increased considerably to 0.94 dB and 1 dB for PCS-256QAM and 64QAM, respectively. We attribute the difference in gain between PBE and the joint equalizer to the following facts: (i) in short-reach transmission data is more contaminated from transceiver impairments than fiber nonlinearity [11]. This

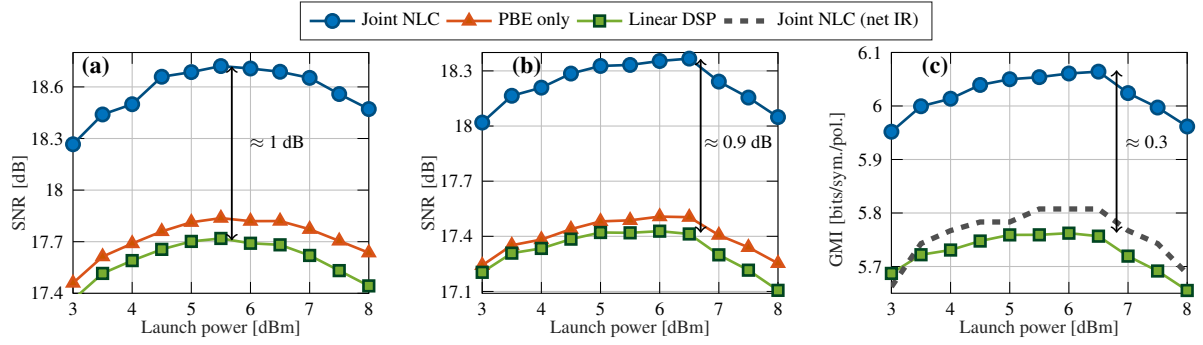


Fig. 3. SNR of received signal over different launch powers for (a) 64QAM and (b) PCS-256QAM. (c) shows GMI and net IR (dashed line) for PCS-256QAM over different launch powers.

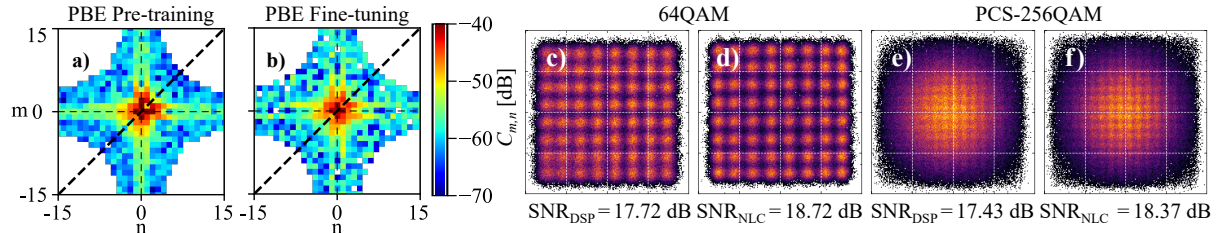


Fig. 4. (a-b) The distributions of PBE coefficients $C_{m,n}$ learnt (a) during pre-training and (b) joint fine-tuning for H-pol of PCS-256QAM at 6.5 dBm launch power. Dashed line is the PBE coefficients' $C_{m,n}$ symmetry axis $m = n$. (c-f) Constellations before and after the joint NLC for H-pol of 64QAM (c) and (d) and PCS-256QAM (e) and (f) at their optimal power levels.

is the main reason why the NN-based equalizer is so effective. (ii) When the symbol rate is beyond 100 GBaud, the channel becomes very dispersive, which makes the PBE with a limited number of terms less efficient.

As shown in Fig. 3c, the SNR gain of joint NLC corresponds to a considerable increase in the generalized mutual information (GMI), which rose from 5.76 bits/symbol/polarization (bits/sym./pol.) for the linear DSP case to 6.06 bits/sym./pol. when the joint NLC was applied. After forward error correction (FEC) we obtained the net information rate (IR) of 5.81 bits/sym./pol. as shown in Fig. 3c. For FEC decoding, we used optimized SC-LDPC codes with variable overheads with scaled min-sum decoder [12]. For each launch power, the FEC with the smallest overhead capable of decoding the bits error-free was chosen. To the best of our knowledge, this result sets a new record of 1.48 Tbps for single-channel 240 km transmission over G.652 SSMF.

5. Conclusion

We proposed a joint neural-network-based equalizer for both channel and transceiver nonlinearity and verified it experimentally in a single-channel 128 GBaud dual-polarization transmission system over 3×80 km of G.652 SSMF. Compared to the standard linear receiver DSP, this additional equalizer provided an extra SNR gain of 0.9 dB and a GMI rise of 0.3 bits/sym./pol. for transmitting PCS-256QAM signals. It enabled a record single-channel net IR of 1.48 Tbps over 240 km SSMF.

Acknowledgements: VN would like to thank Karsten Schuh for insightful discussions. This work has received funding from EU Horizon 2020 program under the Marie Skłodowska-Curie grant agreement No. 766115 (FONTE). JEP is supported by Leverhulme Project RPG-2018-063. SKT acknowledges the support of EPSRC project TRANSNET.

References

1. F. Buchali et al., "128 Gsa/s SiGe DAC implementation enabling 1.52 Tb/s single carrier transmission", IEEE JLT, **39**, p. 763, (2020).
2. G. Khanna et al., "A robust adaptive pre-distortion method for optical communication transmitter", IEEE PTL, **28**, p. 752, (2015).
3. M. Malekiha et al., "Efficient nonlinear equalizer for intra-channel nonlinearity compensation for next generation agile and dynamically reconfigurable optical networks", Opt. Express, **24**, p. 4097, (2016).
4. B. Karanov et al., "Deep learning for communication over dispersive nonlinear channels: performance and comparison with classical digital signal processing", IEEE Allerton Conf., (2019).
5. F. N. Khan et al., "An optical communication's perspective on machine learning and its applications", IEEE JLT, **37**, p. 493, (2019).
6. V. Oliari et al., "Revisiting Efficient Multi-Step Nonlinearity Compensation with Machine Learning: An Experimental Demonstration", IEEE JLT, **38**, p. 3114, (2020).
7. 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.
8. G. Klambauer et al., "Self-normalizing neural networks", NIPS 2017, **30**, p. 971.
9. D. P. Kingma et al., "Adam: A method for stochastic optimization", arXiv preprint, arXiv:1412.6980, (2014).
10. M. E. O'Neill, "PCG: A Family of Simple Fast Space-Efficient Statistically Good Algorithms for Random Number Gen.", (2014).
11. L. Galdino et al., "On the limits of digital back-propagation in the presence of transceiver noise", Opt. Express, **25**, p. 4564, (2017).
12. L. Schmalen et al., "Spatially Coupled Soft-Decision Error Correction for Future Lightwave Systems", IEEE JLT, **33**, p. 1109, (2014).