Joint Pre- and Post-Learned Perturbation Nonlinearity Compensation Optimization for Long-haul Optical Fiber Transmission Based on End-to-end Deep Learning

Lyu Li, Zekun Niu, Junzhe Xiao, Weisheng Hu, and Lilin Yi*
State Key Lab of Advanced Communication Systems and Networks, School of Electronic Information and Electrical Engineering Shanghai Jiao Tong University, Shanghai, China
e-mail: lilinyi@sjtu.edu.cn

Abstract: We propose a joint pre- and post-learned perturbation nonlinearity compensation (LPNC) by using end-to-end deep learning. The experiment shows a 0.6 dB gain is validated in the 21-channel 64QAM 60 GBaud 800km transmission fiber link. © 2025 The Author(s)

1. Introduction

Fiber nonlinearity limits the transmission performance of the high baud-rate long-haul optical communication system [1]. Recently, a number of powerful approaches have been considered for the fiber nonlinearity compensation (NLC), for example, the digital back-propagation (DBP) technologies [2], perturbative nonlinearity compensation (PNC) methods [3]. In particular, PNC is a powerful tool for nonlinearity compensation that could be as a benchmark to evaluate the performance of other emerging techniques.

In PNC methods, the intra-channel fiber nonlinear distortion is modeled as the sum of three-symbol interactions in two polarizations according to the first-order perturbation analysis of the Manakov equation. When applying PNC on a receiver side, an error occurs in the calculation of the three-symbol products due to the absence of transmitted symbols [4]. Additionally, when PNC is utilized for pre-distortion compensation, the performance is influenced by uncertainties in transmission link parameters [3]. The aforementioned problems result in limitations on the ability of the PNC to compensate for nonlinearities.

In this paper, a novel approach is proposed to enhance the performance of fiber nonlinearity compensation, by utilizing a joint pre- and post- learned PNC (LPNC) in both the transmitter and receiver, combined with constellation shaping. Based on our previous work [5], the pre- and post-PNC can be optimized using end-to-end learning. The post-distortion LPNC at the receiver could be trained directly, while the pre-distortion LPNC at the transmitter requires back-propagation gradients estimated by a differentiable surrogate channel (DSC) for updates. Concurrently, we also combine the constellation shaping into the end-to-end learning framework to optimize the transmission performance. The experimental results demonstrate that joint LPNC achieves best performance and provides 0.6 dB Q factor improvement in a 21-channel 64QAM 60 GBaud WDM transmission over a distance of 800 km. Furthermore, we also show that the joint pre- and post-LPNC can also be pruned without losing performance, with 17% and 37.5% reductions compared to pre-LPNC and post-LPNC, respectively.

2. The end-to-end learning for proposed framework

The structure of the end-to-end learning for joint pre- and post- LPNC method is illustrated in Fig. 1 (a). The proposed scheme consists of five modules: Encoder-NN, pre-LPNC, DSC, post-LPNC and Decoder-NN. The Encoder-NN is applied to convert bit sequence to a symbol sequence, with the goal of combatting channel distortion. The Decoder-NN then recovers the symbol sequence back into a bit sequence. Both of the Encoder-NN and Decoder-NN are established on the full connect neural network (FCNN), as detailed in Fig. 1. The pre- and post-LPNC are employed to mitigate the fiber nonlinearity. The nonlinear term is described as (1),

$$\Delta x_0 = \sum P_0^{3/2}\left(x_{k_1} x^*_{k_1+k_2} x_{k_2} + y_{k_1} y^*_{k_1+k_2} x_{k_2}\right) C_{k_1,k_2}, \qquad |k_1| \text{and} |k_2| \le q/2, \tag{1}$$

where the $C_{k1,k2}$ represents the nonlinear perturbation coefficients, is calculated as [4]. k_1 and k_2 are the time-index of symbol complex amplitude which is lower than the matric length q. P_0 denotes the pulse peak power at launch point. The structure of the LPNC is illustrated on the left side of Fig. 1 (b). To enhance LPNC performance, the $C_{k1,k2}$ should be updated to fit the actual transmission link parameters. In our approach, the end-to-end learning is applied to jointly optimize the pre- and post-LPNC. The DSC, rather than regular perturbation (RP) model and Split-Step Fourier method (SSFM), is used to estimate gradient, due to its low complexity and fast computation in long-distance transmission. The DSC is also implemented using a fully connected neural network (FCNN), as shown on the right side of Fig. 1(b). Through the DSC, the gradient computed from the Decoder-NN can be backpropagated to optimize both the pre-LPNC and Encoder-NN. The mean square error (MSE) is used as the loss function, and the Adam algorithm with a learning rate of 0.0001 is employed for learning optimization.

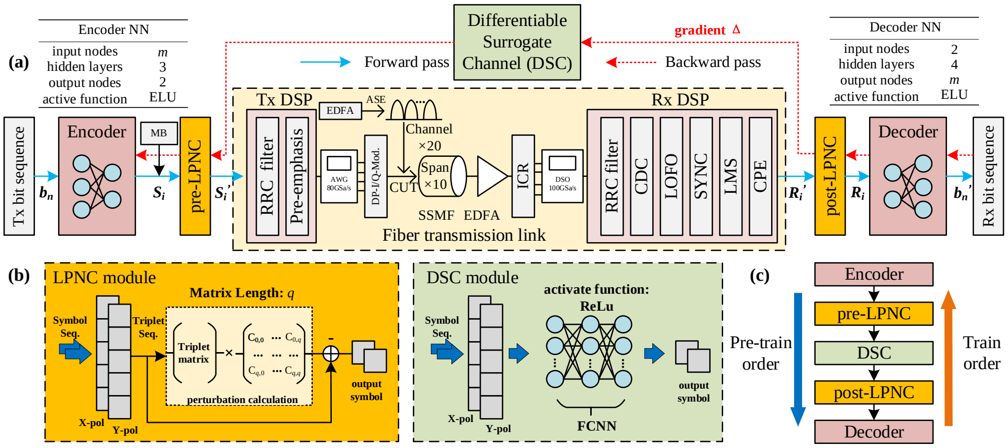

Fig. 1. (a) The proposed scheme. (b) The structure of LPNC module (left) and DSC module (right). (c) The pretraining and train order.

Assuming that the modulation format order is M, the number of bits per symbol is $m=\log_2 M$. At the transmitter side, the Tx bit sequence $[b_1, b_2, \ldots, b_n]$, where $b_n \in \{0, 1\}$, is input into the Encoder-NN and mapped to the symbol sequence $[S_1, S_2, \ldots, S_i]$. The different classes of transmitter symbols are subsequently sampled such that each symbol probability conforms to $P(S_i)$ according to the Maxwell-Boltzmann (MB) distribution [3]. Then, the transmitter symbols are compensated by the pre-LPNC module and then sent to the transmitter. At the receiver side, the received symbols, denoted as $[R'_1, R'_2, \ldots, R'_i]$, are fed into the post-LPNC module and the recovered symbol sequence $[R_1, R_2, \ldots, R_i]$ could be obtained. The Decoder-NN would make the symbol decision. The Rx bit sequence $[b'_1, b'_2, \ldots, b'_n]$ would be compared with Tx bit sequence to calculate the MSE and gradient. Subsequently, the gradient would be backward from Decoder-NN to Encoder-NN as indicated by the red line in Fig. 1 (a). All modules will be optimized according to the training order shown in Fig. 1 (c). In the actual training process, the gradient would be back-propagated from Decoder-NN to Encoder-NN, thus allowing for continuous iterative training for each module. In order to obtain accurate gradient, the final gradient is propagated further only after the loss is relatedly smooth. Since neural networks can be difficult to train and are prone to local optima when initialized randomly, pre-training is required [6]. The procedure for pre-training is shown in Fig. 1(c). The Encoder-NN is first initialized with the Gray-labeled constellations. Subsequently, the perturbation coefficient of pre-LPNC and post-PNC could be calculated according to [3]. The DSC can then be trained using the symbols obtained before the Tx DSP and after the Rx DSP. Finally, the MSE between transmitted and received bits would optimize the Decoder-NN.

3. Experimental results and Discussion

To evaluate the performance of the proposed scheme, a 60 GBaud 21-channel dual-polarization 64QAM transmission experiment was conducted, the results of which are presented in Fig. 1 (a). At the transmitter, the probabilistically shaped symbols were generated and then passed through an RRC filter with a roll-off factor of 0.1. The symbols were then generated using an 80 GSa/s arbitrary waveform generator (AWG) with a bandwidth of 17 GHz. After nonlinearity pre-compensation, the electrical waveforms were modulated by a dual-polarization I/Q modulator with a bandwidth of 40 GHz. The modulated optical signals were amplified using erbium-doped fiber amplifiers (EDFA) with 5.0 dB noise figure and subsequently transmitted. The total transmission distance was approximately 800 km which contains 10 spans of fiber. In WDM long-haul transmission, the channel under test is generated by the AWG, while additional interference channels originate from amplified spontaneous emission (ASE) noise. Following fiber channel transmission, the optical waveforms are coherently detected using 40 GHz bandwidth integrated coherent receiver (ICR). The signals are then sampled by a 100 GSa/s digital oscilloscope (DSO) with 30 GHz electrical bandwidth. The Q-factors are employed to assess the transmission performance.

Firstly, we compared the performance with without LPNC (Wo. LPNC), pre-LPNC, post-LPNC, and proposed method after end-to-end training. From the Fig. 2 (a), it can be discovered that the joint pre-and post- LPNC yields highest compensation performance, with an approximately 0.6 dB improvement to Wo. LPNC. It is noteworthy that pre-LPNC has higher performance than post-LPNC, including the pre-LPNC could mitigate more nonlinearity if LPNC could obtain transmitted symbols. The results from the joint pre-and post-LPNC demonstrate that the residual nonlinearity has been compensated in comparison to the pre-LPNC and post-LPNC methods, with a 0.25 dB and

0.48 dB Q-factor improvement, respectively. The Fig. 2 (b) displayed the constellations applied at the optimal power level. The shaped constellations at the transmitter side are similar to standard square QAM while the probability is different. It also can be seen that the after applying the joint pre-and post-LPNC, the appearance of constellation became more distinct, indicating an enhanced potential for compensation. The Fig. 2 (c) display the perturbation coefficient of pre-LPNC and post-LPNC after training and the former is more clearly.

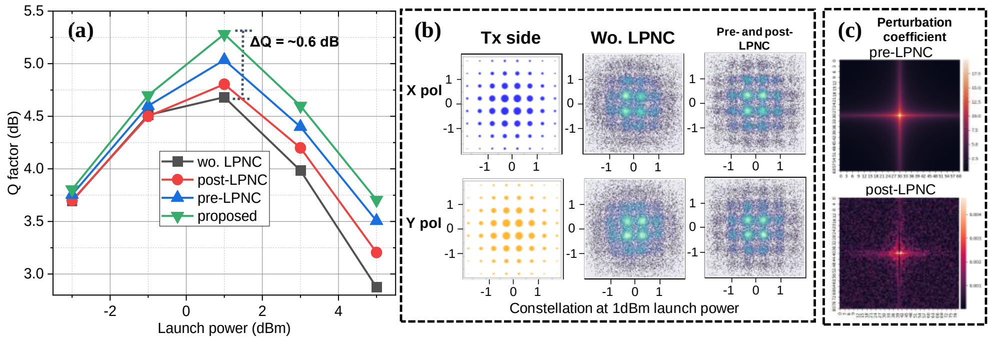

Fig. 2. (a) The Q factor performance at the launch power range of -3 to 5 dBm. (b) The constellation at 1 dBm launch power. (c) The perturbation coefficients of pre-LPNC and post-LPNC

The q_1, q_2, which represent the perturbation coefficients length of pre-LPNC and post-LPNC respectively, would be adjusted to reduce the complexity. We obtain the Q factor of three kinds of LPNC with different perturbation matrix length. The results are shown in Table 1. As q_1 and q_2 decrease, the performance of all LPNC methods initially remains unchanged and then exhibits a significantly decline. Furthermore, it has been demonstrated that the pre- and post-LPNC exhibits superior performance compared to others under the same complexity conditions. The performance is the same when q_1 is equal to 60 for pre-LPNC, q_2 is equal to 80 for post-LPNC, and q_1 and q_2 are equal to 25 and 25, respectively, for proposed method. It is shown that pre- and post-LPNC complexity decreases by 17% and 37.5% compared to pre-LPNC and post-LPNC, respectively. Thus, the proposed method is able to reduce the complexity with the same performance.

Table 1. The Q factor (dB) of LPNC with different perturbation matrix length

q_1	pre-LPNC	q_2	post-LPNC	q_1	q_2	Proposed
120	5.035	120	4.804	100	100	5.325
100	5.022	100	4.792	70	70	5.341
80	4.858	80	4.781	70	50	5.315
60	4.791	60	4.703	50	50	5.227
20	4.732	20	4.680	25	25	4.780

4. Conclusion

In this paper, we propose the end-to-end learning to train a joint pre-and post-LPNC method in both the transmitter and receiver. The proposed method performs better than using LPNC on the transmitter and receiver side alone. Meanwhile, we also preliminarily explore the pruning method to reduce the complexity of the method, and the results show that the LPNC complexity can be effectively reduced by reducing the length of the perturbation coefficient matrix without loss of performance. This method provides a new idea for simultaneous training at both the transmitter and receiver DSP and limiting of nonlinear impairments.

References

[1] T. Gui et al., “Real-time Demonstration of Homodyne Coherent Bidirectional Transmission for Next-generation Data Center Interconnects,” J. Lightw. Technol, vol. 39, no. 4, pp. 1231-1238 (2021).
[2] Q. Fan et al., “Advancing Theoretical Understanding and Practical Performance of Signal Processing for Nonlinear Optical Communications through Machine Learning,” Nat. Commun, vol. 11, no. 1, pp. 3694 (2020).
[3] V. Neskorniuk et al., “Memory-aware End-to-end Learning of Channel Distortions in Optical Coherent Communications,” Opt. Express, vol. 31, no. 1, pp. 1-20 (2022).
[4] M. Fan et al., “Low-complexity Triplet-correlative Perturbative Fiber Nonlinearity Compensation for Long-haul Optical Transmission,” J. Lightw. Technol, vol. 40, no. 16, pp. 5416-5425 (2022).
[5] Z. Niu et al., “End-to-end Deep Learning for Long-haul Fiber Transmission Using Differentiable Surrogate Channel,” J. Lightw. Technol, vol. 40, no. 9, pp. 2808-2822 (2022).
[6] K, Gümüş et al., “End-to-end Learning of Geometrical Shaping Maximizing Generalized Mutual Information,” Proc. OFC (2020).