Noisy samples-robust Neural Network Equalizer for Coherent Optical Transceiver Nonlinearity Compensation

Zicai Cao[1, (†)], Shuchang Yao[2, (†)], Longquan Dai[1], Ziheng Zhang[1], Jing Dai[2], Ming Luo[3], Xi Xiao[3,4], Yaqin Wang[2], Qi Yang[1], Deming Liu[1], and Lei Deng[1,*]

[1]*School of Optical and Electronic Information, Huazhong University of Science and Technology, Wuhan 430074, China*
[2]*Fiberhome Telecommunication Technologies Co., LTD, Wuhan 430073, China*
[3]*China Information and Communication Technologies Group Corporation (CICT), 430074 Wuhan, China*
[4]*National Information Optoelectronics Innovation Center, 430074 Wuhan, China*
[*]*denglei_hust@mail.hust.edu.cn* [†] *These authors contributed equally to this work*

Abstract: We experimentally demonstrate a neural network equalizer with robustness for noisy samples on a silicon photonics coherent transceiver, and a complexity reduction is over 50% at the BER of 1.25e-2 compared with a general NNLE.

1. Introduction

High-speed optical transceivers, with degraded linearity properties in broad frequency bands, require high-efficient nonlinear equalizers to improve optical transmission performance. For silicon-based coherent transceivers, the nonlinearity is even greater because of its carrier-dispersion effect. Tremendous works have been published for optical transceiver nonlinearity compensation. Neural network (NN)-based nonlinearity equalizer (NLE), Volterra series-based nonlinearity equalizer (VNLE), and their variants may be the most common ones. They usually locate at the end of the receiver side digital signal processing (DSP) for joint nonlinearity compensation of high baud rate transceivers [1]. They can mitigate the nonlinearity impairments by fitting its inverse transfer function based on the training sequence. In comparison, VNLE has designable higher-order terms and stronger interpretability, while NNLE has relatively low complexity as it has no computation of higher-order terms but uses activation functions to approximate the nonlinear model [2]. However, usually overlooked, the drawback of the NLEs is that they are data-driven, so their performance is highly dependent on the signals with reliable quality, which depends on the optical channel condition and the DSP performance before NLE. Therefore, most of the NLEs reported today can provide a significant gain for high-quality signals but only a weak gain in the presence of poor signal quality [3-5]. The phenomenon becomes nonnegligible for high-speed transceivers, where a high-baud rate and high-order modulation formats are implemented, and signals may suffer from relatively low signal-to-noise ratio (SNR) originating from various impairments, including limited device bandwidth, excessive noise or heavy nonlinearities. As a result, many mislabeled samples are produced in the signals, making it difficult to train the NLEs well and degrading the generalizability of NLEs. The increase in complexity seems to do no significant help for it.

In this work, to resolve the NLE performance issue under low-SNR conditions for high-speed optical transceivers, we propose and experimentally demonstrate a novel NNLE with robustness for noisy samples in 64/40GBaud DP-16/64QAM transmission on silicon photonics 64Gbaud-class coherent optical transceiver which suffers from a relatively low SNR. With a pre-trained process using a simulated database and a subordinate network as an uncorrelated feature filter, the proposed NNLE can achieve more stable generalizability and can reach the FEC threshold of 1.25e-2 under about 50% of the complexity of a general NNLE.

2. Principle

General VNLE is too complex with the existing DSP resources available. NNLE is a powerful approximator for the nonlinear model with lower complexity than VNLE. NNLE consists of multiple layers that perform weighted summation and nonlinear mapping. The output of the first, kth, and last layer $\boldsymbol{h}_{(1)}$, $\boldsymbol{h}_{(k)}$, $\boldsymbol{y}$ of an NNLE and the mean squared error (MSE) as a loss function can be expressed as:

$$\boldsymbol{h}_{(1)} = \boldsymbol{f}_{(1)}\left(\mathbf{W}_1\boldsymbol{x}_t + \mathbf{b}_{(1)}\right),\ \boldsymbol{h}_{(k)} = \boldsymbol{f}_{(k)}\left(\mathbf{W}_k\boldsymbol{h}_{(k-1)} + \mathbf{b}_{(k)}\right),\ \boldsymbol{y} = \mathbf{W}_o\boldsymbol{h}_{(l-1)} + \mathbf{b}_o,\ \mathrm{loss}(\boldsymbol{x}_t, \boldsymbol{y}_t) = \frac{1}{M}\sum_{m=1}^{M}\left|\boldsymbol{y} - \boldsymbol{y}_t\right|^2 \quad (1)$$

where $\boldsymbol{x}_t$, $\boldsymbol{y}_t$ are the received and transmitted symbols and $\mathbf{W}$, $\mathbf{b}$ are the parameters of each layer to be trained while $\boldsymbol{f}$ is the nonlinear activation function and $\tanh(x) = (e^x - e^{-x})/(e^x+e^{-x})$, $\mathrm{sigmoid}(x) = 1/(1+e^{-x})$ are used in this work. It is noted that an NNLE works as an extractor that extracts features from the database from the shallow to the deep. It is obvious that low-quality signals will lead to degraded performance as the actual system characteristics are masked in the numerous noisy samples. Focusing on the low-SNR condition, our proposed NNLE structure with robustness for noisy samples is shown in Fig. 1. The I/Q components of the signals are separated firstly, and each sample consists of a tapped sequence (length = $2L$). The whole structure consists of 3 parts with three training processes. Firstly, a

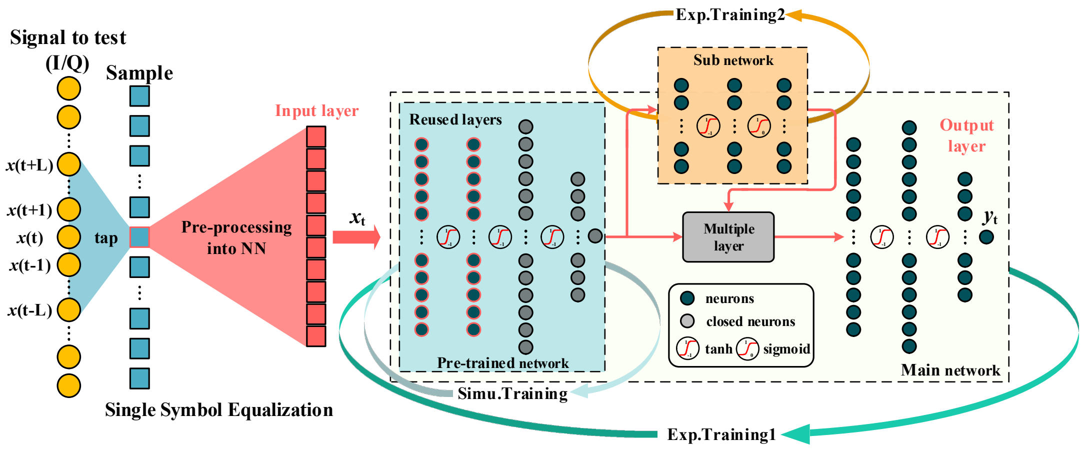

Fig. 1. (a) The schematic diagram of our proposed NNLE structure with noise stability.

database generated by a simulation system designed after the experimental platform is used to pre-train a basic NN, whose shallow layers will be reused to help extract low-level features of the experimental system. This is based on the fact that the two are similar tasks with similar low-level features, such as the overall shape of the constellation.

Subsequent layers are extended on the reused layers, they perform characterizing the deeper features of the experimental signals. The main network will be trained using the experimental signals, and only the new layers will be modified while the shallow layers will be frozen. NN structures with pre-trained models have been proven to have more potential to be well-trained [6]. Besides, to resist the impairment of generalizability by noisy samples, we add a subordinate network to customize a filter that will output a unique weight vector for each sample to remove their noisy features. The sub-network will get a separate training process in which only the layers of it will be trained, and the others perform forward propagation. The training sequences used in the two experimental training processes are intercepted from different experimental signal frames. In the test phase, the NNLE works as a mapper. To compare the performance of a VNLE, a general delayed full-connected network (FCN) NLE, and the proposed one, we calculate the curves of the bit error ratio (BER) versus the complexity evaluated by the required number of multiplier and accumulation (MAC) expressed as:

$$\mathrm{MAC}(Vol)=\sum_{k=1}^{\mathrm{K}}\frac{(\mathrm{L}_k-1+k)!}{(k-1)!(\mathrm{L}_k-1)!},\ \mathrm{MAC(FCN)}=\sum_{l_i=1}^{l}\mathrm{N}_{l,in}*\mathrm{N}_{l,out} \tag{3}$$

where K is the highest order of VNLE and L_k is the number of k-order taps, $\mathrm{N}_{l,in}$ and $\mathrm{N}_{l,out}$ are the input and output neurons number of the l_i th layer of NNLEs. For simplicity, all NNLEs are fixed with 51 taps and their MACs are changed by modifying the hidden layers neurons. It should be mentioned that more complex NNs such as convolutional NN (CNN) and long short-term memory (LSTM) get no discussion here due to their high MACs.

3. Experimental Setup and Discussions

The experimental setup of the 400Gbps+ back-to-back optical coherent transceiver system is shown in Fig. 2. Two external cavity lasers (ECL) with an optical power of 15.5 dBm and 10 dBm are used as the optical carrier and LO, whose wavelength and linewidth are 1550 nm and 100 kHz, respectively. The signals are generated by an arbitrary waveform generator (AWG, Keysight M8194A) with a 3 dB bandwidth of 45 GHz and are then used to drive a Si-based CDM and amplified by an EDFA after modulation. At the receiver, a Si-based ICR is used to reconstruct the

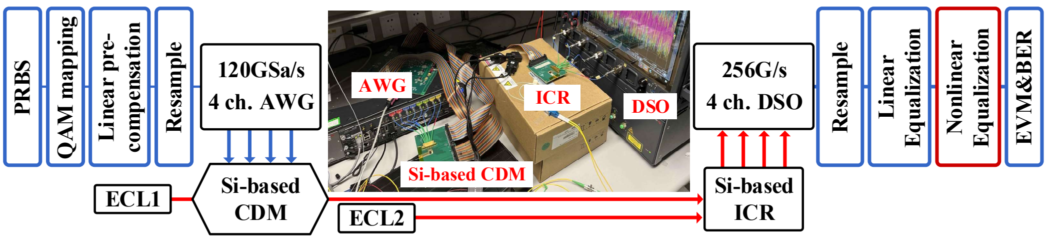

Fig. 2. Experimental setup of 400Gbps+ optical coherent transceiver and profile display of Si-based optical coherent transceiver

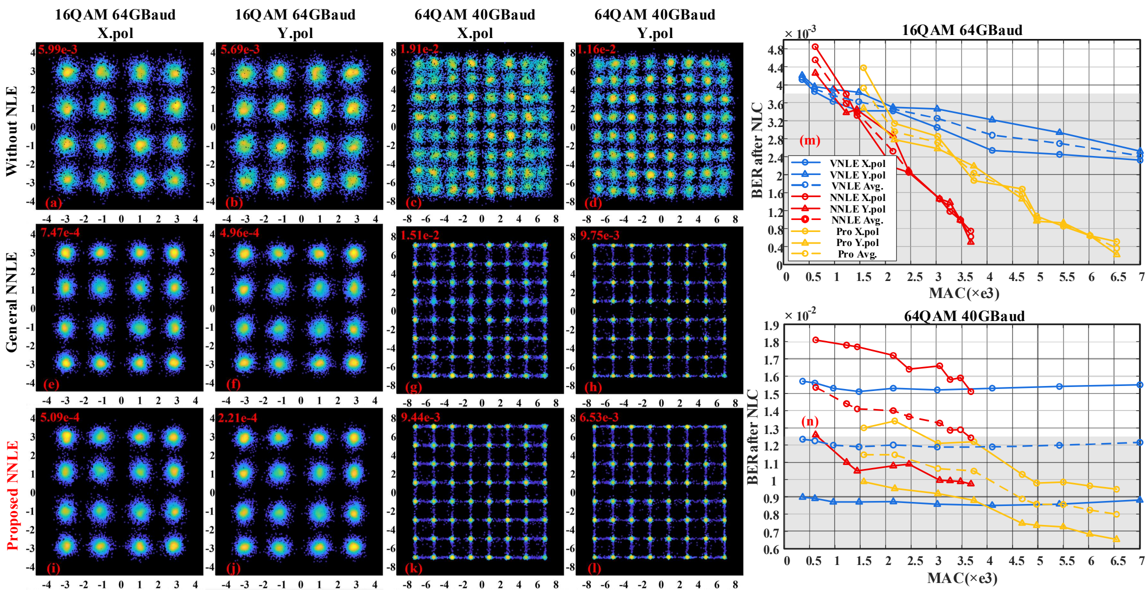

Fig. 3 Received constellation of the X polarization and Y polarization of the 16QAM 64GBaud signals and the 64QAM 40GBaud signals without NLE (a-d), with a general NNLE (e-h), with our proposed NNLE (i-l). (m) The BER vs. MAC curves of the VNLE, general NNLE and our proposed NNLE of the 16QAM 64GBaud signals and (n) of the 64QAM 40GBaud signals.

optical field. The signals are captured and converted to electrical signals by a 256 GSa/s real-time oscilloscope (Keysight, UXR0704A) with a bandwidth >70 GHz. The NLEs are added between the linear equalization and demapping. We test the NLEs using DP-16QAM 64GBaud and DP-64QAM 40GBaud signals, all of the NLEs are trained and tested using approximately 15,000 symbols, respectively. The displayed results are obtained by averaging the performance of multiple frames measured at different times.

Figs. 3(a-l) show the received constellations of the test signals without NLE, with a general NNLE and our proposed NNLE. Fig. 3(m) and Fig. 3(n) show the BER vs. MAC curves of the three NLEs for the 64GBaud 16QAM signals and the 40GBaud 64QAM signals, respectively. It can be noticed that these NLEs can all obtain significant gains with the complexity increases. The NNLE can usually obtain comparable performance to VNLE at a lower complexity, while the proposed NNLE has an unworthy complexity cost. It is concluded that common NLEs can already meet the requirements under relatively high signal quality, and there is no need for additional structures. It can be seen from Fig. 3(n) that with low-quality signals, the increase of complexity seems to do no help for the performance of the VNLE, and the NNLE can only provide a weak gain. The proposed NNLE can achieve considerable performance gain as the complexity increase and can reach the FEC threshold of 1.25e-2 using the 15% overhead with a complexity reduction of over 50% compared to the general one. In summary, the proposed NNLE can achieve more stable generalizability than the general NNLEs under poor signal quality.

4. Conclusions

We propose and experimentally demonstrate a novel NNLE with robustness for noisy samples in 64/40GBaud DP-16/64QAM transmission on silicon photonics 64Gbaud-class coherent optical transceivers. The proposed NNLE can achieve more stable generalizability and can reach the FEC threshold of 1.25e-2 under about 50% of the complexity of a general NNLE. The results also indicate that novel NLE techniques will be necessary for novel high-speed coherent optical transceivers in the future.

5. Acknowledgement

We thank the support of the National Key Research and Development Program of China (2022YFB2903200) and the National Nature Science Foundation of China (NSFC) (62171190).

6. References

[1] L. Huang et al., J. Lightw. Technol., vol. 40, no. 14, pp. 4528 - 4538, Jul. 2022.
[2] M. Schädler et al., J. Lightw. Technol., vol. 39, no. 10, pp. 3095 - 3105, May. 2021.
[3] M. Schaedler et al., Appl. Sci., vol. 9, no. 21, pp. 4675, Oct. 2019.
[4] S. Deligiannidis et al., J. Lightw. Technol., vol. 39, no. 18, pp. 5791 - 5798, Sep. 2021.
[5] K.Wang et al., J. Lightw. Technol., vol. 40, no. 9, pp. 2791 - 2798, May. 2022.
[6] W. Kong et al., Photon. Res., vol. 9, no. 8, pp. 1607-1615, Jul. 2021.