Intra-channel Nonlinearity Mitigation in Optical Fiber Transmission Systems Using Perturbation-based Neural Network

Jiazheng Ding, Tiegen Liu, Tongyang Xu, *Member*, *IEEE*, Wenxiu Hu, Sergei Popov, *Member, IEEE, Fellow, Optica,* Mark S. Leeson, *Senior Member*, *IEEE*, Jian Zhao and Tianhua Xu, *Member*, *IEEE*

***Abstract*—In this work, a perturbation-based neural network (P-NN) scheme with an embedded bidirectional long short-term memory (biLSTM) layer is investigated to compensate for the Kerr fiber nonlinearity in optical fiber communication systems. Numerical simulations have been carried out in a 32-Gbaud dual-polarization 16-ary quadrature amplitude modulation (DP-16QAM) transmission system. It is shown that this P-NN equalizer can achieve signal-to-noise ratio improvements of ~1.37 dB and ~0.80 dB, compared to the use of a linear equalizer and a single step per span (StPS) digital back propagation (DBP) scheme, respectively. The P-NN equalizer requires lower computational complexity and can effectively compensate for intra-channel nonlinearity. Meanwhile, the performance of P-NN is more robust to the distortion caused by equalization enhanced phase noise (EEPN). Furthermore, it is also found that there exists a tradeoff between the choice of modulation format and the nonlinear equalization schemes for a given transmission distance.**

***Index Terms*—Optical fiber communication, Fiber nonlinearity, First-order perturbation theory, Neural network, Equalization enhanced phase noise.**

This work is supported by EU Horizon 2020 MSCA Grant 101008280, UK EPSRC Grant EP/S028455/1, Swedish Research Council 2019-05197, RISE SK, and National Key Research and Development Program of China 2022YFE0202100. (Corresponding author: Tianhua Xu and Jian Zhao.)

Jiazheng Ding, Tiegen Liu and Jian Zhao are with School of Precision Instruments and Opto-Electronics Engineering, Tianjin University, Tianjin 300072, China (e-mail: whisperil01_@tju.edu.cn; tgliu@tju.edu.cn; enzhaojian@tju.edu.cn).

Tongyang Xu is with School of Engineering, Newcastle University, Newcastle upon Tyne, NE1 7RU, United Kingdom and with Department of Electronic and Electrical Engineering, University College London, London WC1E7JE, United Kingdom (e-mail: tongyang.xu.@ieee.org).

Wenxiu Hu and Mark S. Leeson are with School of Engineering, University of Warwick, Coventry CV4 7AL, United Kingdom (e-mail: wenxiu.hu@warwick.ac.uk; mark.leeson@warwick.ac.uk).

Sergei Popov is with KTH Royal Institute of Technology, Stockholm 16440, Sweden (sergeip@kth.se).

Tianhua Xu is with School of Engineering, University of Warwick, Coventry CV4 7AL, United Kingdom, with Tianjin University, Tianjin 300072, China, and also with University College London, WC1E 6BT London, United Kingdom (e-mail: tianhua.xu@ieee.org).

I. Introduction

OPTICAL fibers currently underpin global communication infrastructures and services, which carry over 95% of the digital data traffic [1]. The spectral efficiencies (SEs) and capacity of the transmission systems have been promoted through methods such as coherent optical detection, advanced modulation formats, and powerful digital signal processing [2]–[6]. Optical transmission systems suffer from linear impairments, such as chromatic dispersion (CD), polarization mode dispersion (PMD) and laser phase noise (LPN). These impairments can be well compensated via digital signal processing (DSP) [7]–[9]. However, the achievable information rates (AIRs) of optical communication systems are still restricted due to the Kerr effect in the nonlinear regime [10]–[14]. Therefore, fiber nonlinearities are regarded as major detrimental factors limiting the information capacity of modern optical fiber communication systems [4], [15]–[18]. To date, digital back propagation (DBP) is an effective nonlinearity compensation (NLC) technique that has received significant attention, which is modeled by signal backward propagation along an optical fiber via the split-step Fourier method (SSFM) in the digital domain. A variety of studies regarding DBP approaches have been reported [19]–[21]. Unfortunately, a major issue in DBP is that such schemes require high computational resources, which makes the system intractable for real-time operation.

Alternatively, perturbation analysis, a very promising approach, has been widely applied to perform NLC [22]–[25]. Intra-channel nonlinear effects were first analyzed using a perturbational approach in high data-rate single-polarization transmission systems by Mecozzi *et al.* [22]. Tao *et al.* further modified the intra-channel nonlinear perturbative model and investigated the performance under various transmission conditions for dual-polarization quadrature phase shift keying (DP-QPSK) and dual-polarization 16-ary quadrature amplitude modulation (DP-16QAM) systems [25]. The complexity of the perturbation-based nonlinearity compensation model has been reduced using symmetric electronic dispersion compensation (EDC) and pulse shaping [24]. With the rapid development of machine learning (ML)

technology, various effective neural network (NN) topologies and algorithms have been proposed and implemented [26]–[32]. Recently, digital approaches motivated by ML for mitigating fiber nonlinearities have attracted much research attention [33]–[41], harnessing various algorithms, such as nonlinear boundary decisions using a support vector machine (SVM)-based classification nonlinear equalizer (NLE) [36], blind equalization using a dynamic deep neural network, to improve the signal nonlinear tolerance [42]. Moreover, researchers have developed a deep convolutional neural network (DCNN) for mitigating the nonlinear signal distortion in a long-haul fiber communication system [43]. A ML approach based on parameterizing the SSFM for the nonlinear Schrödinger equation (NLSE) was proposed to compensate for fiber nonlinearities [44]. In this scheme, the steps of DBP were implemented in a deep multi-layer neural network, and good performance after the compensation was demonstrated. However, the algorithm required an operation with at least two samples per symbol, which introduced additional complexity.

By contrast, another approach employed perturbation theory and ML algorithms to extract information from the received data to figure out the nonlinear impairments experienced [45]–[47]. This version of fiber nonlinearity compensation algorithms, with the use of the triplets from the perturbation theory as input features, aimed at creating nonlinear function and tensor weights and treated the nonlinear equalization as a regression problem, which is a feature engineering. It operated on one sample per symbol, to improve the transmission performance without sacrificing energy and computation efficiencies. Recently, recurrent neural networks (RNNs) have been widely used in sentiment classification and natural language processing, and researchers have migrated this technology to optical fiber communication systems, achieving good performance [48]–[51]. Bidirectional long short-term memory (biLSTM) has the characteristics of capturing bidirectional dependencies of sequence data, and its output is jointly determined by several previous inputs and several subsequent inputs, so that the features learned from sequence data will be more accurate. The performance and the complexity of bi-Vanilla-RNN was compared with traditional Volterra nonlinear equalizers [50]. The simple recurrent neural network (SRNN) with a time-domain memory was studied to mitigate the intra-channel fiber nonlinearity [48]. The convolution layer (CNN) in combination with biLSTM was experimentally demonstrated but required high computational complexity [51].

In this work, we have developed a first-order perturbation-based neural network (P-NN) scheme with an embedded biLSTM layer, where interactions between neighboring symbols are taken into account, to compensate for the Kerr nonlinearities in a more comprehensive and practical long-distance optical fiber transmission link. The performance of the intra-channel nonlinearity compensation using the P-NN is investigated comprehensively in a DP-16QAM transmission system, where the EDC and the DBP with different number of steps per span (StPS) have been applied as references. The relationship between the peak SNR and the transmission distance, under different compensation schemes, is studied. It is demonstrated that the P-NN equalizer outperforms the linear equalizer, with a ~1.37 dB signal-to-noise ratio (SNR) gain in a single-channel 32-Gbaud DP-16QAM transmission system. Compared to 1-StPS DBP, the P-NN can achieve a ~0.80 dB SNR improvement in a 2000 km standard single mode fiber (SSMF) transmission link due to the optimized nonlinear function in the neural network. Meanwhile, the performance of the P-NN has been analyzed, for the first time to our knowledge, in the presence of considerable equalization enhanced phase noise (EEPN), and the results show that P-NN is more robust to the EEPN distortion than conventional DSP algorithms. The applicability and limitations on the use of the neural network to perform NLC in optical transmission systems with advanced modulation formats have been explored. The AIRs versus transmission distances using the EDC, the DBP and the P-NN for different modulation formats have also been comprehensively evaluated. It is found that the developed P-NN has strong potential and applicability to mitigate fiber nonlinearities especially in the scenarios where computational resources are significantly limited. Finally, the developed P-NN has been tested in a 9-channel 32-Gbaud DP-16QAM Nyquist-spaced superchannel optical transmission system to investigate its NLC performance.

The remainder of the paper is structured as follows. In Section II, we describe perturbation-based nonlinearity mitigation and AIR estimation. In Section III, we present the configuration of the optical transmission system investigated. The numerical results are laid out in Section IV, with discussion of their significance and implications. Finally, we conclude the work in Section V.

II. Perturbation-based nonlinearity mitigation and AIR estimation

A. Nonlinear perturbation in optical communication system

In dual-polarization optical transmission systems, the evolution of the optical field envelope is governed by the Manakov equations under the condition that the nonlinear effective length is much larger than the fiber birefringence beating length [25]

$$\frac{\partial u_x}{\partial z}+\frac{\alpha}{2}u_x+i\frac{\beta_2}{2}\frac{\partial^2 u_x}{\partial t^2}=i\frac{8}{9}\gamma[\left|u_x\right|^2+\left|u_y\right|^2]u_x \quad (1)$$

$$\frac{\partial u_y}{\partial z}+\frac{\alpha}{2}u_y+i\frac{\beta_2}{2}\frac{\partial^2 u_y}{\partial t^2}=i\frac{8}{9}\gamma[\left|u_y\right|^2+\left|u_x\right|^2]u_y \quad (2)$$

where u_x and u_y are the complex optical fields in the two mutually orthogonal polarization directions, with time and distance dependence, respectively, α is the attenuation coefficient, β_2 is the group velocity dispersion coefficient and γ denotes the nonlinear coefficient. A signal launched into the optical fiber can be expressed as

$$u_x(z=0,t)=\sum_n \sqrt{P_0}H_n g(0,t-nT) \tag{3}$$

$$u_y(z=0,t)=\sum_n \sqrt{P_0}V_n g(0,t-nT) \tag{4}$$

where H_n and V_n are the complex symbols for *x*- and *y*-polarizations, respectively, the subscript n is the time slot index, T is the symbol interval and $g(t)$ is the waveform of the carrier pulse. In first-order perturbation analysis, the approximate solution to Eq. (1) and Eq. (2) can be written as $u(z,t)\approx u_0(z,t)+\Delta u(z,t)$, where $u_0(z,t)$ is the linear solution with $\gamma=0$, and $\Delta u(z,t)$ is the nonlinear perturbation term. This approximation works well for the regime of weak nonlinearities and strong dispersion. More details of first-order perturbation theory in fiber nonlinearity equalization can be found in [25]. Meanwhile, it is noted that the concept of the fiber nonlinearity model derived based on the perturbation theory is a somewhat similar to the enhanced Gaussian noise (EGN) model [52]. The perturbation method treats the fiber nonlinearity as a deterministic additive distortion, while the EGN model approximates the nonlinear degradation via a sum of Gaussian noise terms, representing the different sources of nonlinearities.

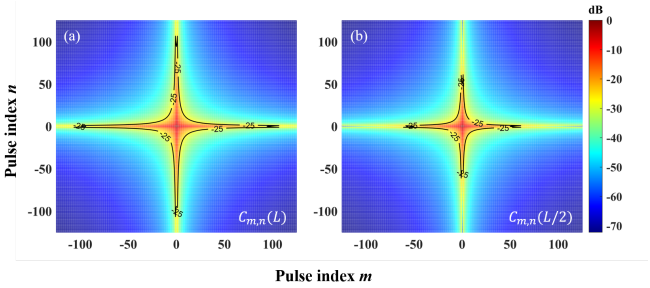

Fig. 1. The perturbation coefficient $C_{m,n}$ (2000 km transmission distance) which is normalized to $C_{0,0}$. (a) Without dispersion pre-compensation $C_{m,n}(L)$, (b) With 50% dispersion pre-distortion $C_{m,n}(L/2)$.

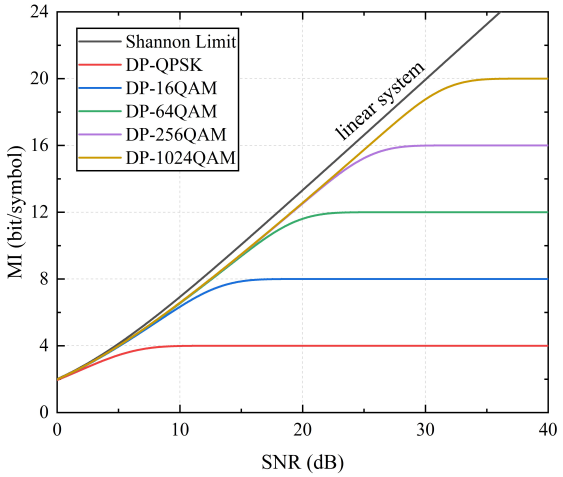

Fig. 2. Theoretical MI versus SNR for dual-polarization transmission systems using different modulation formats over the AWGN channel.

Without loss of generality, we focus on the perturbation analysis of the zero-th symbol in the channel considered. With the assumption that the pulse spreading due to CD is much larger than the symbol duration, the perturbation of this symbol at $t=0$ can be simplified to

$$\Delta u_x(z=l,0)=\sum_{m,n} P_0^{3/2}\left(H_n H_{m+n}^* H_m + V_n V_{m+n}^* H_m\right)C_{m,n} \tag{5}$$

$$\Delta u_y(z=l,0)=\sum_{m,n} P_0^{3/2}\left(V_n V_{m+n}^* V_m + H_n H_{m+n}^* V_m\right)C_{m,n} \tag{6}$$

where P_0 is the signal launch power. The triplets are defined as $H_n H_{m+n}^* H_m$ and $V_n V_{m+n}^* H_m$ for the *x*-polarization, $V_n V_{m+n}^* V_m$ and $H_n H_{m+n}^* V_m$ for the *y*-polarization. The nonlinear perturbation coefficient $C_{m,n}$ is expressed as

$$C_{m,n}(l)=j\frac{8}{9}\gamma\int_0^l dz f(z)\int_{-\infty}^{+\infty} dt g_f^*(z,t)\cdot g(z,t-mT)g(z,t-nT)g^*(z,t-mT-nT) \tag{7}$$

where *T*, *L* and *z* are the symbol interval, the total transmission length, and the propagation distance along the fiber, respectively, while *m* and *n* are the symbol indices; $g_f(z,t)$ denotes the function obtained by applying the dispersion operator to the receiver impulse response; $f(z)$ accounts for the loss/gain profile along the fiber link, and in the case of a periodically amplified multi-span transmission link, it can be expressed as:

$$f(z)=\exp\left[-\alpha \bmod(z,L_s)\right] \tag{8}$$

where L_s is the span length. The analytical calculation of Eq. (7) can be found in [53]. For a 50% dispersion pre-distorted link with a length of L, the signal is dispersion free at the middle of the link. It has been demonstrated that the relationship of $C_{m,n}(-L/2)=-C_{m,n}^*(L/2)$ will lead to a smaller dispersion interaction distance, and this results in a reduction in the number of perturbation terms in truncated approximations [24]. Fig. 1 illustrates the nonlinear perturbation coefficient with (in Fig. 1(b)) and without (in Fig. 1(a)) the 50% dispersion pre-distortion, where the truncation threshold is -25 dB. Here, $C_{m,n}$ is presented as a relative magnitude with respect to the peak value $C_{0,0}$. The decibel (dB) in the Fig. 1 is defined as $20\log_{10}(|C_{m,n}|/|C_{0,0}|)$. It can be seen that the closer to the symbol of interest, the larger the nonlinear contribution of the interference pulse.

B. Mutual information and achievable information rate

The symbol-wised soft-decision mutual information (MI) is an indicator of the achievable channel rate under specific modulation formats using advanced channel decoders [54].

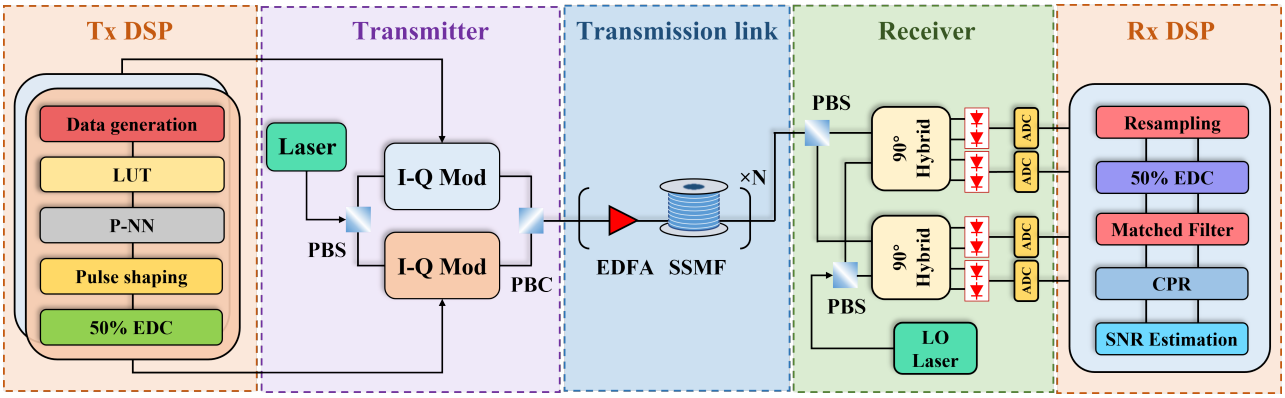

Fig.3. Schematic of the 32-Gbaud DP-16QAM optical fiber communication system using P-NN. (LUT: look up tables, PBS: polarization beam splitter, PBC: polarization beam combiner, LO: local oscillator, ADC: analogue-to-digital convertor, CPR: carrier phase recovery)

For a memoryless channel, X is the discrete complex input, Y is the continuous complex output, the symbols sequence is transmitted over the channel with the distribution $f_{Y|X}\left(y \mid x\right)$, the upper achievable rate can be defined as [55]–[57]

$$MI = \mathbb{E}_{X,Y}\left[\log_2 \frac{f_{Y|X}\left(y \mid x\right)}{f_Y\left(y\right)}\right] \tag{9}$$

where $\mathbb{E}_{X,Y}$ is the expectation operator. However, the real channel distribution probability is unknown, we can obtain a lower bound on Eq. (9) though the mismatched decoding approach [17], [56], [58], and consider an auxiliary channel instead of the memoryless channel $f_{Y|X}\left(y \mid x\right)$. Based on the additive white Gaussian noise (AWGN) channel assumption, the auxiliary channel can be expressed as

$$p_{Y|X}\left(y \mid x\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\left(y-x\right)^2}{2\sigma^2}\right) \tag{10}$$

where σ^2 is the noise variance. For a discrete QAM signal input distribution in a dual polarization optical transmission system, the MI in Eq. (9) can be expressed as

$$MI = \frac{2}{M}\sum_{i=1}^{M}\int_{\mathbb{C}} p_{Y|X}\left(y \mid x_i\right) \cdot \log_2 \frac{p_{Y|X}\left(y \mid x_i\right)}{\frac{1}{M}\sum_{j=1}^{M} p_{Y|X}\left(y \mid x_j\right)} dy \tag{11}$$

where $M = \mid X \mid = 2^p$ denotes the cardinality of the M-QAM constellation with the number of bits per symbol p, $\mathbb{C}$ denotes the set of complex numbers. Fig. 2 shows the MI versus SNR curves for dual-polarization systems using different modulation formats, numerically computed using Gauss-Hermite quadrature over an AWGN channel.

The AIR is the natural figure to measure the overall effective data rate that can be reliably transmitted through the channel under consideration [11]. By using MI, the AIR of the transmission system can be expressed as

$$AIR = N_{ch} \cdot R_s \cdot MI \tag{12}$$

where N_{ch} is the number of channels and R_s denotes the symbol rate of the transmitted signal.

III. Transmission System and Neural Network

To evaluate the performance of the P-NN scheme, numerical simulations regarding the intra-channel nonlinearity mitigation have been performed. Fig. 3 illustrates the configuration of a 32-Gbaud DP-16QAM optical communication system using the P-NN for the nonlinear equalization. In the transmitter, a laser with a linewidth of 100 kHz and a center wavelength of 1550 nm was used as the optical carrier. The data sequence in each polarization from independent random binary sequence generators with the length of 2^{17}-1 were de-correlated with a delay of half the sequence length. The impact of the bit sequence length (used in the NN training) on the P-NN equalization performance is detailed in the Appendix. Before using the P-NN for nonlinear predistortion, the items of the triplets were pre-calculated offline and stored in a look-up table (LUT), rather than computed through many multiplications and integrations, as in the case of the decision directed least mean square (DD-LMS) nonlinear filter equalizer [59]. A root-raised cosine (RRC) filter with a roll-off of 0.1% was used for the Nyquist pulse shaping (NPS). Then, 50% dispersion pre-compensation was used to reduce the dispersion broadening of the optical pulse transmitted in the fiber. The 16-QAM signals modulated by modulators in the two orthogonal polarizations were fed into the SSMF. The numerical simulation of the SSMF was based on the split-step Fourier solution of the Manakov equations with a logarithmic step-size distribution. A periodically amplified multi-span link was applied. An Erbium-doped optical fiber amplifier (EDFA) was employed to compensate for the loss in the fiber loop. At the receiver, ideal coherent

detection with all in-phase and quadrature signal components was realized by mixing the received signals and the carrier from a local oscillator (LO) laser. The compensation of the remaining 50% dispersion and the carrier phase recovery were carried out in the DSP modules. Finally, the SNR and AIR were calculated to access the performance of the system. PMD and the frequency offset of the transmitter and LO lasers were neglected. Detailed parameters of the optical communication system are listed in Table 1.

TABLE I
SYSTEM PARAMETERS

Parameter	Value
Symbol rate (R_s)	32 Gbaud
Central wavelength (λ)	1550 nm
Attenuation coefficient (α)	0.2 dB km^{-1}
Chromatic dispersion coefficient (D)	16.7 ps nm^{-1} km^{-1}
Nonlinear coefficient (γ)	1.3 W^{-1} km^{-1}
Span length (L_s)	100 km
Number of spans (N_s)	20
EDFA noise figure (N_f)	4.5 dB

TABLE II
INFLUENCE OF NEURAL NETWORK PARAMETERS ON SYSTEM PEAK-SNR

mini-Batch Size	Peak SNR (dB)	initial Learning Rate	Peak SNR (dB)
50	16.32	1e-1	15.35
100	16.42	1e-2	15.57
300	16.29	1e-3	16.42
500	16.25	1e-4	16.25

Num of input triplets	Peak SNR (dB)	RMpS
57	15.17	4684
109	15.88	4892
225	16.00	5356
397	16.20	6044
797	16.42	7644
1249	16.49	9452

In conventional perturbation analysis for nonlinearity compensation, prior knowledge of transmission link parameters is an essential prerequisite, and a large amount of computing resources have to be spent on numerical integrations. Now, with the great advantages of machine learning, it is possible to achieve performance improvements under low complexity conditions. Here we treat the nonlinear equalization process as a regression problem. A fully-connected neural network including a biLSTM layer, as shown in Fig. 4, was selected to mitigate the fiber nonlinearities. In order to efficiently train a neural network, we used a small learning rate of 0.001 shown in Table 2, which has been verified without overfitting. To speed up the training process in parallel, we split each training dataset into multiple parallel batches and each one has 100 training samples. The number of epochs was limit to 30 to cut the training time. A dropout layer (DL) with a probability of 0.5 was utilized to avoid overfitting during training. The simplest architecture shown in Fig. 4 was found to simplify the complexity of the neural network without sacrificing performance. After several hidden layers of nonlinear nodes, the output layer had two units corresponding to the real and the imaginary parts of the estimated nonlinear distortion.

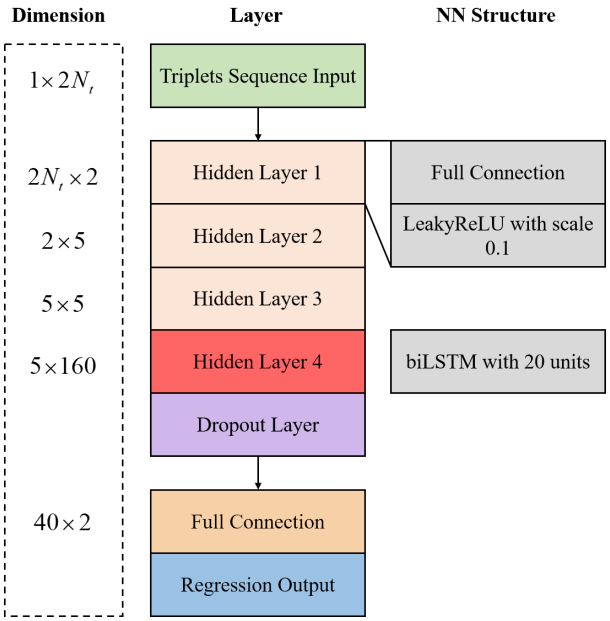

Fig.4. Architecture of the proposed P-NN equalizer.

The principle of the perturbation-based neural network for nonlinear equalization is shown in Fig. 5. The training and the testing stages of the equalization network are shown in Fig. 5(a) and Fig. 5(b), respectively. Before the training of the equalizer, the uncorrelated training, the cross-validation (CV) and testing data sets needed to be prepared. Over 110,000 symbols were used in the training during the transmission, approximately 50,000 in the CV and roughly 80,000 in testing the performance of the neural network. For a fiber link with given parameters, the nonlinear perturbation coefficient only needed to be calculated once under a certain transmission distance. Before training, the coefficient should be truncated according to a predefined threshold (such as the -25 dB shown in Fig. 1), and items which have a significant impact on the symbols of interest retained. During the training stage, the triplets fed into the NN were generated from the received symbols as shown in Fig. 5(a). The P-NN was trained with the Adam optimizer to minimize the mean-squared error which is defined as [60]

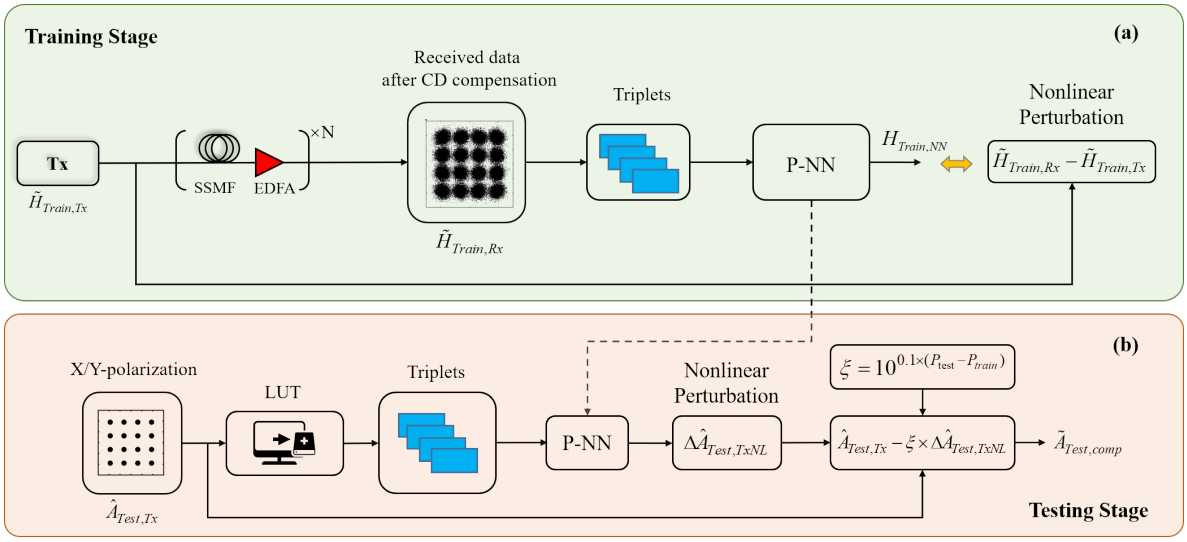

Fig. 5. The block diagram of the perturbation-based neural network for nonlinearity mitigation. (a) Received symbols are applied in the P-NN during the training stage, (b) Transmitted symbols are applied in the P-NN during the testing stage.

$$MSE=\frac{1}{B}\sum_{i=1}^{B}\left|H_{Train,NN}-\left(\tilde{H}_{Train,Rx}-\tilde{H}_{Train,Tx}\right)\right| \quad (13)$$

where B is the batch size, $H_{Train,NN}$ denotes the output of the NN, $\tilde{H}_{Train,Rx}$ and $\tilde{H}_{Train,Tx}$ is the received symbols and the transmitted symbols, respectively. After that, the weight matrix was saved in a table for the subsequent testing stage. Since the P-NN was operated in a data-driven manner and symbols in the two orthogonal polarization states were uncorrelated, the NN only needed to be trained in one polarization and then the nonlinearity compensation could be conducted in both polarizations accordingly. At the testing stage in Fig. 5(b), nonlinear perturbation triplets, which have been calculated and stored in the LUT, formed the input items of the NN.

After the process of the P-NN, the transmitted symbols were pre-distorted by subtracting the nonlinear perturbation items from them. It should be noted that the nonlinearity mitigation was implemented according to the power scaling factor ξ shown in Fig. 5(b), which could be applied for different optical launch powers.

Here, the number of real multiplications per symbol (RMpS) was considered as the indicator of the computational complexity. Since the P-NN equalizer worked on the basis that the CD experienced by the symbols has been compensated, the computational complexity of the EDC needed to be taken into account in the whole P-NN equalization. Therefore, the RMpS of the P-NN was $C_{EDC}+C_{NN}$.

The complexity of the EDC could be expressed as

$$C_{EDC}=4\cdot\frac{N_{FFT}\left(\log_2 N_{FFT}+1\right)\cdot n_s}{N_{FFT}-N_D+1} \quad (14)$$

where n_s is the oversampling ratio in samples per symbol, N_{FFT} is the fast Fourier transform (FFT) size, $N_D=n_s\tau_D/T$, with τ_D being the dispersive channel impulse response and T denotes the symbol duration. The RMpS of the NN part is then be expressed as

$$\begin{aligned}C_{NN}&=2n_{in}n_{h1}+n_{h1}n_{h2}+n_{h2}n_{h3}\\&\quad+2\cdot\left(4n_{h3}n_{hb}+4n_{hb}^2+3n_{hb}+n_{hb}n_{out}\right)\end{aligned} \quad (15)$$

where n_{in} is the number of input triplets, n_{h1}, n_{h2} and n_{h3} are the number of neurons in each layer, n_{hb} is the number of biLSTM hidden cells and n_{out} the number of outputs.

To evaluate the performance of the P-NN equalizer, the basic implementation of a DBP algorithm was utilized as a reference, where the RMpS of the DBP was

$$C_{DBP}=4N_{span}N_{StPS}\left(\frac{N_{FFT}\left(\log_2 N_{FFT}+1\right)\cdot n_s}{N_{FFT}-N_D+1}+n_s\right) \quad (16)$$

where N_{span} is the number of spans during backwards transmission and N_{StPS} the number of back propagation steps per span.

IV. Results and Discussions

To evaluate the performance of P-NN, numerical simulations on the intra-channel nonlinearity compensation were carried out based on a 32-Gbaud DP-16QAM optical fiber communication system. The transmission link comprised 2000 km (20×100 km) of SSMF, unless otherwise specified. Simulation results of SNR versus optical launch power using

the P-NN are illustrated in Fig. 6 with green triangles. As references, the EDC, the NN-NLC [47], the biLSTM, the 1-StPS, the 2-StPS and the 20-StPS DBP were used to evaluate the performance of the P-NN. It is found that the P-NN achieves a peak SNR of ~16.42 dB at the optimum launch power, with the use of ~797 triplets, representing SNR gains of ~1.37 dB and ~0.80 dB, compared to the EDC and the 1-StPS DBP, respectively. In addition, the P-NN outperforms the NN-NLC as the P-NN can capture the bidirectional perturbation impairment suffered by symbols. The biLSTM, driven by transmitted pulses only, shows a similar peak SNR as the NN-NLC. This is because in the biLSTM, the memory effect between pulses was considered only during the mapping process, where no physical layer parameters characterizing fiber nonlinearities were involved. Consequently, the biLSTM shows the inferior NLC performance compared to the P-NN.

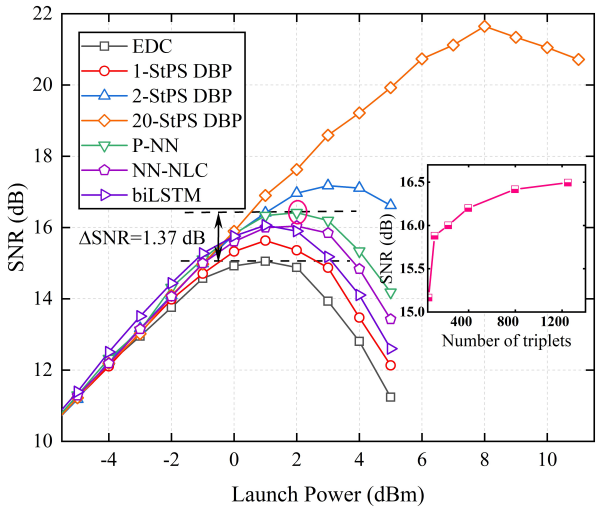

Fig. 6. Simulation results of SNR versus optical launch power using different compensation methods. The inset depicts the number of the input triplets versus the system SNR in 32-Gbaud DP-16QAM transmission system.

It has also been reported that the performance of the SRNN is equivalent to that of the 1-StPS DBP [48]. By contrast, the P-NN employed the pre-compensation via the symmetric EDC and the LUT at the transmitter side, which avoided the cumbersome online calculation of triplets, and shows better performance than the 1-StPS DBP. The inset in Fig. 6 shows the peak SNR value versus the number of neural network input items, at the optimum launch power. It can be observed that the performance of the P-NN improves gradually as the number of triplets increases. It is noted that the increment of the number of neural network inputs will inevitably lead to an increase in the computational complexity, and thus a tradeoff between the performance and the complexity should be taken into account.

Meanwhile, it is noted that the P-NN delivers inferior performance to the DBP for StPS $\geq$ 2. Nonetheless, the improvement in the peak SNR utilizing the multi-step DBP algorithm is also accompanied with a significant increase in the computational complexity. Fig. 7 shows the AIR gain (compared to EDC) and the computational complexity in different NLC schemes for the DP-16QAM transmission system. The histogram indicates that the AIR obtained by the P-NN lies between the NN-NLC and 2-StPS DBP. In terms of the RMpS, the P-NN requires 7644 operations, less than that required in the 2-StPS DBP. The AIR gain in the biLSTM is smaller than that in the P-NN, while its implementation complexity is higher than that in the P-NN. Although the 20-StPS DBP can obtain a larger AIR gain, the RMpS required in the 20-StPS DBP is ten times higher than that of the P-NN. The AIR gain versus the RMpS of the DP-16QAM optical transmission system using the P-NN and the DBP for NLC is shown in the inset of Fig. 7. It is found that the performance of the P-NN is equivalent to the DBP when the computational complexity is ~6400 RMpS. More importantly, the P-NN performs better at the low-complexity region. Under such comprehensive investigations, it is demonstrated that our P-NN scheme behaves more feasible and applicable for real-time operations in next-generation high-speed commercial optical fiber communication systems.

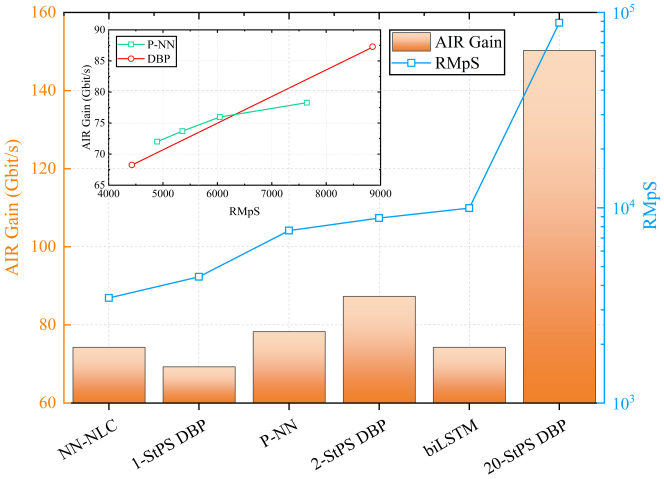

Fig. 7. Comparison on the RMpS and the AIR among different NLC schemes in 32-Gbaud DP-16QAM transmission system with the transmission distance of 2000 km. The inset shows the AIR gain versus the RMpS in 32-Gbaud DP-16QAM transmission system using P-NN and DBP.

In Fig. 8, peak SNR values (at optimum signal powers) obtained in various compensation schemes versus transmission distances are illustrated. It is found that transmission systems using EDC, DBP and the P-NN show similar trends, while the P-NN outperforms EDC and the 1-StPS DBP at all considered transmission distances. Intuitively, the longer the transmission distance, the greater the peak SNR gains in the three NLC schemes, compared to EDC. When transmitting over a small number of spans, the peak SNR of the 1-StPS DBP is only slightly higher than that of the EDC, while the P-NN achieves a SNR gain of ~1.23 dB compared to EDC. When the transmission distance reaches 2000 km, the P-NN can obtain ~1.37 dB SNR gain while there is only ~0.58 dB SNR gain in the case of 1-StPS DBP. Consequently, the P-NN offers greater mitigation of fiber nonlinearities, particularly when the computational resources are

significantly constrained in the transmission scenarios considered.

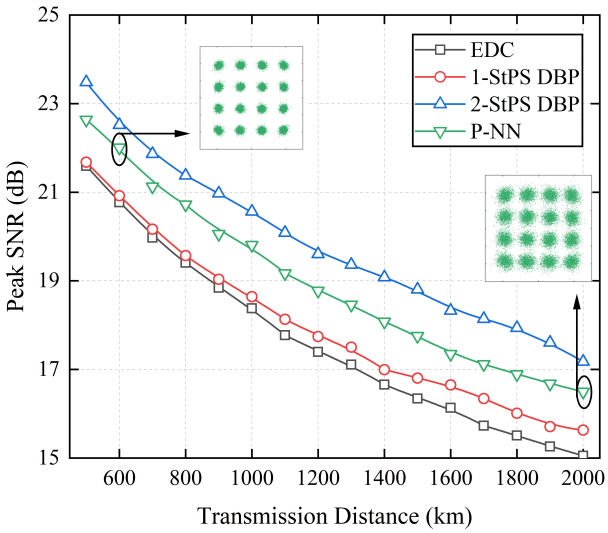

Fig. 8. Performance of peak SNR versus different transmission distances using the EDC, the DBP with different StPS, and the P-NN in the DP-16QAM system. Insets are the *x*-polarization constellation diagrams using the P-NN at two transmission distances.

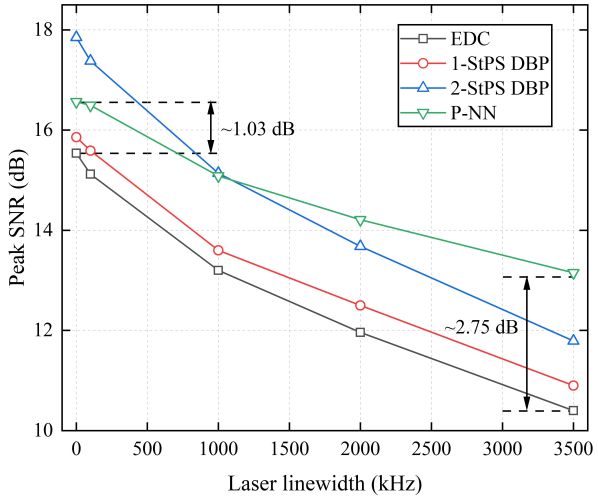

Fig. 9. Impact of laser linewidth on the peak SNR of the DP-16QAM transmission system using the EDC, the different StPS DBP, and the P-NN.

In DSP-assisted high-speed coherent communication systems, EEPN, which originates from the interaction between the LPN and the EDC module or the DBP module will greatly reduce the AIRs of long-haul optical fiber transmission systems [61]–[64]. Fig. 9 depicts the laser linewidth versus the peak SNR for the DP-16QAM transmission system using EDC, the DBP with different StPS values and the P-NN. The intrinsic LPN is removed by using an ideal CPE in the DSP module, therefore the degradation of the system SNR only arises from the distortion caused by EEPN. When there is no EEPN distortion (the laser linewidth is set to 0 Hz), the SNR gain obtained by the P-NN is ~1.03 dB compared to the EDC only case, which is the net gain from the equalization of fiber nonlinearity by the P-NN. Since the data-driven nonlinear equalizer can capture some features of the EEPN impairments (including the EEPN-fiber nonlinearity interaction) during the training process, the P-NN offers greater robustness to EEPN distortions compared with the use of the EDC and the DBP with different StPS values. When the laser linewidth is 1 MHz, the P-NN and 2-StPS DBP show the same performance. The P-NN outperforms all other methods when the linewidth is larger than 1 MHz, and the SNR gain (with respect to EDC) of the P-NN can reach ~2.75 dB when the laser linewidth is 3.5 MHz. This shows that DBP algorithms pose stricter requirements for the laser linewidth in the transmission system, while the P-NN has a stronger tolerance to variations in linewidth.

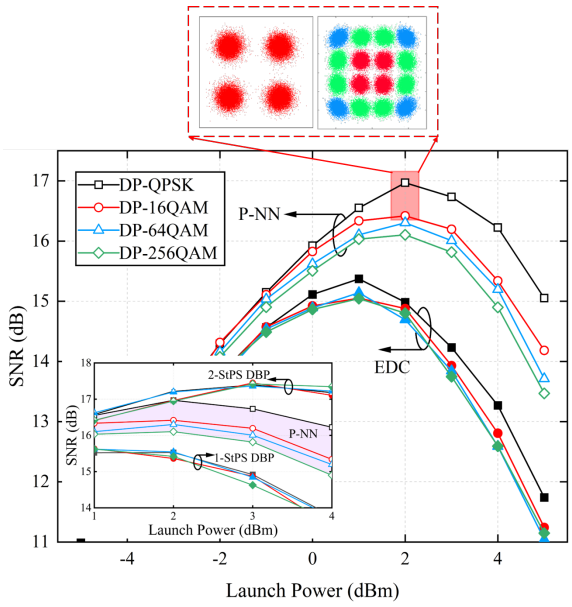

Fig. 10. SNR performance as a function of launch power using EDC, the different StPS DBP, and P-NN for different modulation formats. The inset depicts the SNR versus launch power using different StPS DBP and the P-NN.

Since the P-NN works as a data-driven nonlinear equalizer to learn the perturbation from transmitted symbols, our method can be employed for various modulation formats in practical applications. Results of the SNR versus launch power using EDC, different StPS DBP values and the P-NN for systems with different modulation formats are shown in Fig. 10. Nonlinear distortion in the transmission system is alleviated owing to the nonlinear function produced in the P-NN. The DP-QPSK transmission system outperforms the other three modulation formats under identical system parameters and transmission distances. Compared to the EDC only case at the optimal launch power, the SNR gains obtained in DP-QPSK, DP-16QAM, DP-64QAM and DP-256QAM systems by using the P-NN are ~1.60 dB, ~1.37 dB, ~1.16 dB and ~1.06 dB, respectively. In addition, it is interesting to see that the performance of P-NN shows modulation format dependency in the subplot of Fig. 10. The reason for this phenomenon is that the P-NN models the intra-channel nonlinear distortion as an additive impairment. This generally

works when distortions arise mainly from intra-channel four-wave mixing (IFWM) and significant cumulative dispersion. However, for higher-level QAM modulation formats, intra-channel cross-phase modulation (IXPM) will introduce considerable multiplicative nonlinear phase noise [65], as shown in the outer constellation points of the 16QAM signal in Fig. 10, which cannot be fully compensated using such a P-NN. This results in a decrease of peak SNRs for DP-16QAM, DP-64QAM, DP-256QAM, compared to DP-QPSK.

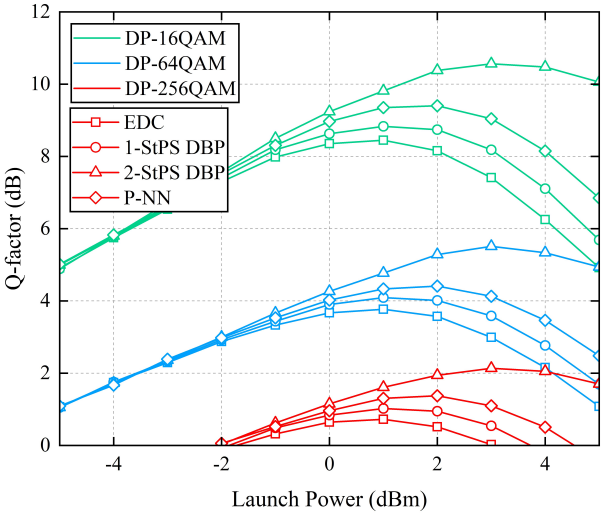

Fig. 11. Simulation results of Q-factor versus optical launch power using the EDC, the different StPS DBP, and the P-NN for different modulation formats.

In modern high-speed long-distance optical communication systems, the fiber nonlinearity is very strong, which has a decisive influence on the final bit error rate (BER) of the system. The Q-factor can be a good measure of system performance, which is $Q = 20\log_{10}\left[\sqrt{2}erfc^{-1}\left(2 \cdot BER\right)\right]$. Fig. 11. depicts simulation results of Q-factor versus optical launch power by using EDC, DBP with different StPS values and the proposed P-NN for different modulation formats. At a transmission distance of 2000 km, the P-NN equalizer can improve the Q factor of 32-Gbuad DP-16QAM, DP-64QAM and DP-256QAM transmission systems by ~0.95 dB, ~0.65 dB, and ~0.6 dB, respectively. Meanwhile, it can be found that the compensation performance of the P-NN equalizer is modulation format dependent.

Fig. 12 compares AIRs as function of the transmission distances in 32-Gbaud transmission system using EDC, DBP with different StPS values and the P-NN for different modulation formats. It can be seen that DP-256QAM using P-NN outperforms DP-64QAM using 20-StPS DBP at transmission distances of less than 1000 km, and AIRs obtained in DP-256QAM and DP-64QAM systems using the P-NN are basically the same after 2000 km transmission. This means that, to obtain a target information rate, there exists a tradeoff between the choice of modulation format and the nonlinear equalization method for a given transmission distance.

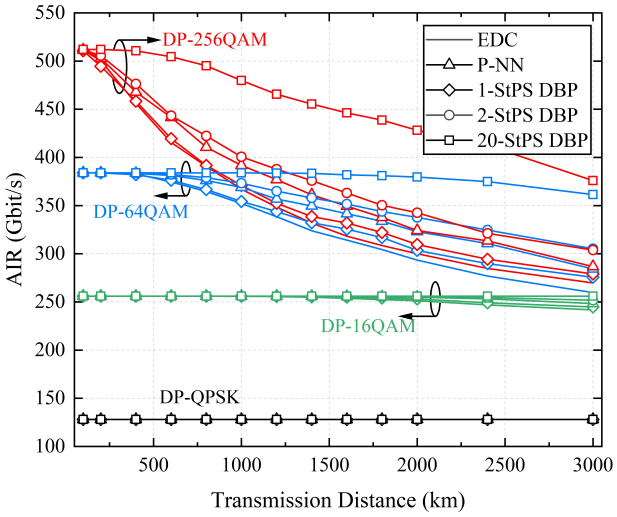

Fig. 12. Dependence of AIRs on the transmission distances using the EDC, DBP with different StPS, and P-NN for different modulation formats

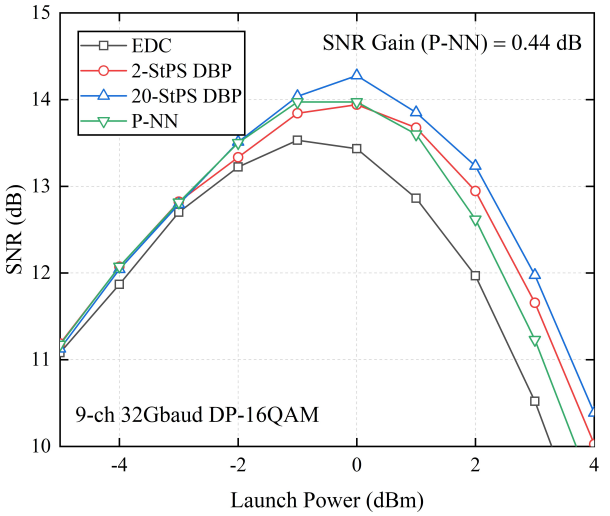

Fig. 13. Simulation results of SNR and optical launch power using different compensation methods in a 9-channel 32-Gbaud DP-16QAM Nyquist-spaced superchannel transmission system.

In addition, the P-NN nonlinearity equalizer has also been investigated in a WDM multi-channel optical transmission system. Fig. 13 depicts the SNR versus launch power per channel in a 9-channel 32-Gbaud DP-16QAM Nyquist-spaced superchannel optical fiber transmission system using the P-NN equalizer and DBP with different StPS, where the EDC is used as a reference. Since the data-driven nonlinear equalizer partially captures the inter-channel interference during training stage, a peak SNR gain of ~0.44 dB is obtained compared to the EDC, and the performance of the P-NN is slightly higher than the single-channel 2-StPS DBP. It can be found that 20-StPS DBP achieves a higher SNR gain than the P-NN, however, the number of RMpS required will be ten times higher than the P-NN, which limits its practical application. In addition, the input characteristics of the P-NN equalizer used

here only contain the intra-channel nonlinearity, therefore, the performance is suboptimal in a multi-channel transmission scheme. In future work, we will extend the perturbation-based machine learning technique to the WDM transmission systems to develop more cost- and performance-effective nonlinear equalizers.

V. Conclusions

In this paper, a first-order nonlinear P-NN, embedded with a bidirectional long short-term memory layer, is developed to mitigate fiber nonlinearities. By applying deep learning technology, this P-NN equalizer outperforms a linear equalizer, with a SNR gain of ~1.37 dB in a 2000 km 32-Gbaud DP-16QAM transmission system. Compared to 1-StPS DBP, the P-NN can achieve a ~0.80 dB SNR improvement in a 2000 km SSMF transmission link due to the nonlinear function in the NN. It is shown that the implementation complexity is lower than the DBP algorithm, therefore, the proposed P-NN has potential and applicability to mitigate fiber nonlinearities especially when computational resources are strictly limited, for example DSP hardware in high-capacity optical fiber communication systems. The relationship between the peak SNR of the communication system and the transmission distance under different compensation methods has also been investigated. Furthermore, the performance of P-NN shows that it is more robust to the distortion caused by EEPN. The dependence of AIRs on the transmission distances using different compensation methods indicates that, to obtain a target AIR, there exists a tradeoff between the choice of modulation format and the nonlinear equalization method for a given transmission distance. Finally, the P-NN has also been investigated in a WDM multi-channel optical transmission system.

Appendix

A. Influence of the training sequence length on the equalization performance of P-NN

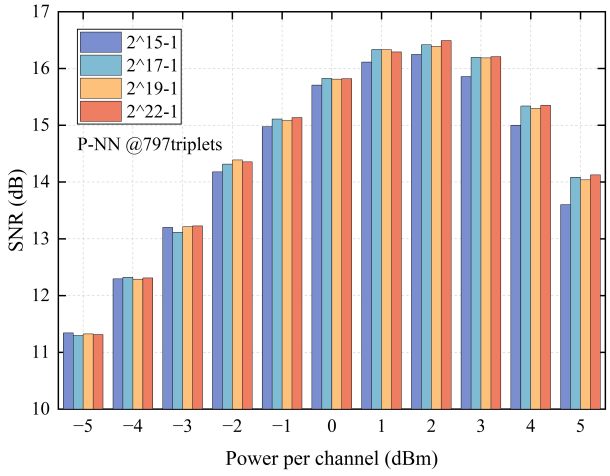

Fig. 14. SNR versus power per channel using the P-NN with different lengths of training bit sequence for the DP-16QAM transmission of 2000km.

Neural networks have the fitting and prediction capability, therefore, the NN nonlinear equalization using the pseudo-random bit sequence (PRBS) as the training data may show an overfitting effect [66]–[68]. In this work, independent random binary sequences were utilized to assess the performance of the NN. In order to explore the influence of the bit sequence length on the equalization performance, SNRs at the receiver, in the scenarios of training data with different bit sequence lengths (2^{15}-1, 2^{17}-1, 2^{19}-1 and 2^{22}-1), have been evaluated, as shown in Fig. 14.

It is found that the performance of P-NN was suboptimal, when the bit sequence length is less than 2^{17}-1, especially in the nonlinear regime, due to the insufficient amount of training data. When the sequence length is over 2^{17}-1, the SNR gain by increasing the training data length is marginal. In this case, the performance improvement in the use of P-NN arises from the equalization of intra-channel nonlinear impairments.

References

[1] Cisco, “Cisco Annual Internet Report (2018–2023),” Mar. 2020.

[2] G. P. Agrawal, *Fiber-optic communication systems*, 4th ed. United States: John Wiley & Sons, Inc., Hoboken, New Jersey, 2010.

[3] E. Ip, A. P. Lau, D. J. Barros, and J. M. Kahn, “Coherent detection in optical fiber systems,” *Opt. Exp.*, vol. 16, no. 26, pp. 753–791, 2008, doi: 10.1364/oe.16.021943.

[4] O. Vassilieva, I. Kim, and T. Ikeuchi, “Enabling technologies for fiber nonlinearity mitigation in high capacity transmission systems,” *J. Lightw. Technol.*, vol. 37, no. 1, pp. 50–60, 2019, doi: 10.1109/JLT.2018.2877310.

[5] A. Redyuk, E. Averyanov, O. Sidelnikov, M. Fedoruk, and S. Turitsyn, “Compensation of nonlinear impairments using inverse perturbation theory with reduced complexity,” *J. Lightw. Technol.*, vol. 38, no. 6, pp. 1250–1257, 2020, doi: 10.1109/JLT.2020.2971768.

[6] D. Zibar, P. Molly, and C. G. Schaeffer, “Machine learning techniques in optical communication,” *J. Lightw. Technol.*, vol. 34, no. 6, pp. 1442–1452, 2016, doi: 10.1109/ISSCC.2008.4523135.

[7] B. Karanov *et al.*, “Experimental investigation of deep learning for digital signal processing in short reach optical fiber communications,” in *IEEE Workshop on Signal Processing Systems, SiPS: Design and Implementation*, 2020, pp. 1–6. doi: 10.1109/SiPS50750.2020.9195215.

[8] X. Zhou and L. Nelson, “Advanced DSP for 400 Gb/s and beyond optical networks,” *J. Lightw. Technol.*, vol. 32, no. 16, pp. 2716–2725, 2014, doi: 10.1109/JLT.2014.2321135.

[9] M. S. Faruk and S. J. Savory, “Digital signal processing for coherent transceivers employing multilevel formats,” *J. Lightw. Technol.*, vol. 35, no. 5, pp. 1125–1141, 2017, doi: 10.1109/JLT.2017.2662319.

[10] A. Mecozzi and R. J. Essiambre, “Nonlinear shannon limit in pseudolinear coherent systems,” *J. Lightw. Technol.*, vol. 30, no. 12, pp. 2011–2024, 2012, doi: 10.1109/JLT.2012.2190582.

[11] A. Alvarado, T. Fehenberger, B. Chen, and F. M. J. Willems, “Achievable information rates for fiber optics: applications and computations,” *J. Lightw. Technol.*, vol. 36, no. 2, pp. 424–439, 2018, doi: 10.1109/JLT.2017.2786351.

[12] P. Bayvel *et al.*, “Maximizing the optical network capacity,” *Philos. Trans. R. Soc. A*, vol. 374, no. 20140440, 2016, doi: 10.1098/rsta.2014.0440.

[13] T. Xu, N. A. Shevchenko, Y. Zhang, C. Jin, J. Zhao, and T. Liu, “Information rates in Kerr nonlinearity limited optical fiber communication systems,” *Opt. Exp.*, vol. 29, no. 11, pp. 17428–17439, 2021, doi: 10.1364/oe.415753.

[14] A. D. Ellis, M. E. McCarthy, M. A. Z. Al Khateeb, M. Sorokina, and N. J. Doran, “Performance limits in optical communications due to fiber nonlinearity,” *Adv. Opt. Photonics*, vol. 9, no. 3, pp. 429–503, 2017, doi: 10.1364/aop.9.000429.

[15] Y. Gao, J. C. Cartledge, J. D. Downie, J. E. Hurley, D. Pikula, and S. S. H. Yam, “Nonlinearity compensation of 224 Gb/s dual-

polarization 16-QAM transmission over 2700 km," *IEEE Photon. Technol. Lett.*, vol. 25, no. 1, pp. 14–17, 2013, doi: 10.1109/LPT.2012.2227111.

[16] R. Dar, M. Feder, A. Mecozzi, M. Shtaif, and P. J. Winzer, "Nonlinear phase and polarization rotation noise in fully loaded WDM systems," in *2015 European Conference on Optical Communication, ECOC*, 2015, pp. 1–3. doi: 10.1109/ECOC.2015.7341790.

[17] T. Xu *et al.*, "Modulation format dependence of digital nonlinearity compensation performance in optical fibre communication systems," *Opt. Exp.*, vol. 25, no. 4, pp. 3311–3326, 2017, doi: 10.1364/oe.25.003311.

[18] S. Gaiarin, F. Da Ros, R. T. Jones, and D. Zibar, "End-to-end optimization of coherent optical communications over the split-step fourier method guided by the nonlinear fourier transform theory," *J. Lightw. Technol.*, vol. 39, no. 2, pp. 418–428, 2021, doi: 10.1109/JLT.2020.3033624.

[19] Z. Zheng *et al.*, "Fiber nonlinearity mitigation in 32-Gbaud 16QAM Nyquist-WDM systems," *J. Lightw. Technol.*, vol. 34, no. 9, pp. 2182–2187, 2016, doi: 10.1109/JLT.2016.2535408.

[20] C. Li, F. Zhang, Y. Zhu, M. Jiang, Z. Chen, and C. Yang, "Fiber nonlinearity mitigation in single carrier 400 G and 800 G Nyquist-WDM systems," *J. Lightw. Technol.*, vol. 36, no. 17, pp. 3707–3715, 2018, doi: 10.1109/JLT.2018.2844467.

[21] B. Karanov, T. Xu, N. A. Shevchenko, D. Lavery, R. I. Killey, and P. Bayvel, "Span length and information rate optimisation in optical transmission systems using single-channel digital backpropagation," *Opt. Exp.*, vol. 25, no. 21, pp. 25353–25362, 2017, doi: 10.1364/oe.25.025353.

[22] A. Mecozzi, C. B. Clausen, and M. Shtaif, "Analysis of intrachannel nonlinear effects in highly dispersed optical pulse transmission," *IEEE Photon. Technol. Lett.*, vol. 12, no. 12, pp. 1633–1635, 2000, doi: 10.1109/68.896331.

[23] E. P. da Silva, M. P. Yankov, F. da Ros, T. Morioka, L. K. Oxenlowe, and L. K. Oxenløwe, "Perturbation-based FEC-Assisted iterative nonlinearity compensation for WDM systems," *J. Lightw. Technol.*, vol. 37, no. 3, pp. 875–881, 2019, doi: 10.1109/JLT.2018.2882638.

[24] Y. Gao *et al.*, "Reducing the complexity of perturbation based nonlinearity pre-compensation using symmetric EDC and pulse shaping," *Opt. Exp.*, vol. 22, no. 2, pp. 1209–1219, 2014, doi: 10.1364/oe.22.001209.

[25] Z. Tao, L. Dou, W. Yan, L. Li, T. Hoshida, and J. C. Rasmussen, "Multiplier-free intrachannel nonlinearity compensating algorithm operating at symbol rate," *J. Lightw. Technol.*, vol. 29, no. 17, pp. 2570–2576, 2011.

[26] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015, doi: 10.1038/nature14539.

[27] H. Qin, R. Gong, X. Liu, X. Bai, J. Song, and N. Sebe, "Binary neural networks: a survey," *Pattern Recognit.*, vol. 105, 2020, doi: 10.1016/j.patcog.2020.107281.

[28] S. Alford, R. Robinett, L. Milechin, and J. Kepner, "Pruned and structurally sparse neural networks," in *2018 IEEE MIT Undergraduate Research Technology Conference, URTC 2018*, 2018, pp. 3–6. doi: 10.1109/URTC45901.2018.9244787.

[29] M. Rastegari, V. Ordonez, J. Redmon, and A. Farhadi, "XNOR-Net: ImageNet classification using binary convolutional neural networks," in *ECCV*, 2016, vol. 9908. doi: 10.1007/978-3-319-46493-0.

[30] F. M. Nardini, C. Rulli, S. Trani, and R. Venturini, "Distilled neural networks for efficient learning to rank," *IEEE Trans. Knowl. Data Eng.*, 2022, doi: 10.1109/TKDE.2022.3152585.

[31] L. Huang *et al.*, "Low power keyword recognition accelerator based on approximate calculation of deep-shift neural network," in *IEEE 15th International Conference on Solid-State & Integrated Circuit Technology (ICSICT)*, 2020, pp. 1–3. doi: 10.1109/ICSICT49897.2020.9278185.

[32] Y. Li, X. Dong, and W. Wang, "Additive powers-of-two quantization: an efficient non-uniform discretization for neural networks," in *2020 International Conference on Learning Representations*, 2020, pp. 1–15. [Online]. Available: http://arxiv.org/abs/1909.13144

[33] P. J. Freire *et al.*, "Complex-valued neural network design for mitigation of signal distortions in optical links," *J. Lightw. Technol.*, vol. 39, no. 6, pp. 1696–1705, 2020, doi: 10.1109/JLT.2020.3042414.

[34] Q. Fan, G. Zhou, T. Gui, C. Lu, and A. P. T. Lau, "Advancing theoretical understanding and practical performance of signal processing for nonlinear optical communications through machine learning," *Nat. Commun.*, vol. 11, no. 1, pp. 1–11, 2020, doi: 10.1038/s41467-020-17516-7.

[35] C. Hager and H. D. Pfister, "Physics-based deep learning for fiber-optic communication systems," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 1, pp. 280–294, 2020, doi: 10.1109/JSAC.2020.3036950.

[36] T. Nguyen, S. Mhatli, E. Giacoumidis, L. Van Compernolle, M. Wuilpart, and P. Megret, "Fiber nonlinearity equalizer based on support vector classification for coherent optical OFDM," *IEEE Photonics Journal*, vol. 8, no. 2, 2016, doi: 10.1109/JPHOT.2016.2528886.

[37] R. Gu, Z. Yang, and Y. Ji, "Machine learning for intelligent optical networks: A comprehensive survey," *J. Netw. Comput. Appl.*, vol. 157, no. 102576, 2020, doi: 10.1016/j.jnca.2020.102576.

[38] T. Kamiyama, H. Kobayashi, and K. Iwashita, "Neural network nonlinear equalizer in long-distance coherent optical transmission systems," *IEEE Photon. Technol. Lett.*, vol. 33, no. 9, pp. 421–424, 2021, doi: 10.1109/LPT.2021.3067341.

[39] E. Giacoumidis, Y. Lin, M. Blott, and L. P. Barry, "Real-time machine learning based fiber-induced nonlinearity compensation in energy-efficient coherent optical networks," *APL Photonics*, vol. 5, no. 041301, pp. 1–5, 2020, doi: 10.1063/1.5140609.

[40] V. Neskorniuk, F. Buchali, V. Bajaj, S. K. Turitsyn, J. E. Prilepsky, and V. Aref, "Neural-network-based nonlinearity equalizer for 128 GBaud coherent transcievers," in *2021 Optical Fiber Communications Conference and Exhibition*, 2021, pp. Th1A.30. doi: 10.1364/ofc.2021.th1a.30.

[41] P. J. Freire *et al.*, "Domain adaptation: the key enabler of neural network equalizers in coherent optical systems," in *2022 Optical Fiber Communications Conference and Exhibition*, 2022, pp. Th2A.35. doi: 10.1364/ofc.2022.th2a.35.

[42] O. Sidelnikov, A. Redyuk, and S. Sygletos, "Equalization performance and complexity analysis of dynamic deep neural networks in long haul transmission systems," *Opt. Exp.*, vol. 26, no. 25, pp. 32765–32776, 2018, doi: 10.1364/oe.26.032765.

[43] O. Sidelnikov, A. Redyuk, S. Sygletos, M. Fedoruk, and S. Turitsyn, "Advanced convolutional neural networks for nonlinearity mitigation in long-haul WDM transmission systems," *J. Lightw. Technol.*, vol. 39, no. 8, pp. 2397–2406, 2021, doi: 10.1109/JLT.2021.3051609.

[44] C. Hager and H. D. Pfister, "Deep learning of the nonlinear schrödinger equation in fiber-optic communications," *IEEE Int. Symp. Inf. Theory*, pp. 1590–1594, 2018, doi: 10.1109/ISIT.2018.8437734.

[45] P. He *et al.*, "A fiber nonlinearity compensation scheme with complex-valued dimension-reduced neural network," *IEEE Photonics Journal*, vol. 13, no. 6, pp. 1–7, 2021, doi: 10.1109/JPHOT.2021.3123624.

[46] V. Kamalov *et al.*, "Evolution from 8QAM live traffic to PS 64-QAM with neural-network based nonlinearity compensation on 11000 km open subsea cable," in *2018 Optical Fiber Communications Conference and Exhibition*, 2018, pp. Th4D.5.

[47] S. Zhang *et al.*, "Field and lab experimental demonstration of nonlinear impairment compensation using neural networks.," *Nat. Commun.*, vol. 10, no. 1, Dec. 2019, doi: 10.1038/s41467-019-10911-9.

[48] Y. Zhao *et al.*, "Low-complexity fiber nonlinearity impairments compensation enabled by simple recurrent neural network with time memory," *IEEE Access*, vol. 8, pp. 160995–161004, 2020, doi: 10.1109/ACCESS.2020.3021146.

[49] D. Wang *et al.*, "Data-driven optical fiber channel modeling: A deep learning approach," *J. Lightw. Technol.*, vol. 38, no. 17, pp. 4730–4743, 2020, doi: 10.1109/JLT.2020.2993271.

[50] S. Deligiannidis, C. Mesaritakis, and A. Bogris, "Performance and complexity analysis of bi-directional recurrent neural network models versus volterra nonlinear equalizers in digital coherent systems," *J. Lightw. Technol.*, vol. 39, no. 18, pp. 5791–5798, 2021, doi: 10.1109/JLT.2021.3092415.

[51] P. J. Freire *et al.*, "Performance versus complexity study of neural network equalizers in coherent optical systems," *J. Lightw. Technol.*,

vol. 39, no. 19, pp. 6085–6096, 2021, doi: 10.1109/JLT.2021.3096286.

[52] A. Carena, G. Bosco, V. Curri, Y. Jiang, P. Poggiolini, and F. Forghieri, "EGN model of non-linear fiber propagation," *Opt. Exp.*, vol. 22, no. 13, pp. 16335–16362, 2014, doi: 10.1364/oe.22.016335.

[53] A. Ghazisaeidi and R. J. Essiambre, "Calculation of coefficients of perturbative nonlinear pre-compensation for Nyquist pulses," in *2014 European Conference on Optical Communication, ECOC*, 2014, pp. 1–3. doi: 10.1109/ECOC.2014.6964065.

[54] R. J. Essiambre, G. J. Foschini, P. J. Winzer, G. Kramer, and B. Goebel, "Capacity limits of optical fiber networks," *J. Lightw. Technol.*, vol. 28, no. 4, pp. 662–701, 2010, doi: 10.1109/JLT.2009.2039464.

[55] T. Fehenberger, A. Alvarado, P. Bayvel, and N. Hanik, "On achievable rates for long-haul fiber-optic communications," *Opt. Exp.*, vol. 23, no. 7, pp. 9183–9191, 2015, doi: 10.1364/oe.23.009183.

[56] M. Secondini, E. Forestieri, and G. Prati, "Achievable information rate in nonlinear WDM fiber-optic systems with arbitrary modulation formats and dispersion maps," *J. Lightw. Technol.*, vol. 31, no. 23, pp. 3839–3852, 2013, doi: 10.1109/JLT.2013.2288677.

[57] L. Szczecinski and A. Alvarado, *Bit-Interleaved Coded Modulation: Fundamentals, Analysis and Design*. John Wiley & Sons, Inc., 2015. doi: 10.1561/0100000019.

[58] D. M. Arnold, H. A. Loeliger, P. O. Vontobel, A. Kavčić, and W. Zeng, "Simulation-based computation of information rates for channels with memory," *IEEE Trans. Inf. Theory*, vol. 52, no. 8, pp. 3498–3508, 2006, doi: 10.1109/TIT.2006.878110.

[59] M. Malekiha, I. Tselniker, and D. V. Plant, "Efficient nonlinear equalizer for intra-channel nonlinearity compensation for next generation agile and dynamically reconfigurable optical networks," *Opt. Exp.*, vol. 24, no. 4, pp. 4097–4108, 2016, doi: 10.1364/oe.24.004097.

[60] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," in *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 2015, pp. 1–15.

[61] W. Shieh and K.-P. Ho, "Equalization-enhanced phase noise for coherent-detection systems using electronic digital signal processing," *Opt. Exp.*, vol. 16, no. 20, pp. 15718–15727, 2008, doi: 10.1364/oe.16.015718.

[62] A. P. T. Lau, W. Shieh, and K. P. Ho, "Equalization-enhanced phase noise for 100Gb/s transmission and beyond with coherent detection," *Opt. Exp.*, vol. 18, no. 16, pp. 17239–17251, 2010, doi: 10.1109/ICCS.2010.5686612.

[63] C. Jin, N. A. Shevchenko, Z. Li, S. Popov, Y. Chen, and T. Xu, "Nonlinear coherent optical systems in the presence of equalization enhanced phase noise," *J. Lightw. Technol.*, vol. 39, no. 14, pp. 4646–4653, 2021, doi: 10.1109/JLT.2021.3076067.

[64] Y. Mori, C. Zhang, and K. Kikuchi, "Novel configuration of finite-impulse-response filters tolerant to carrier-phase fluctuations in digital coherent optical receivers for higher-order quadrature amplitude modulation signals," *Opt. Exp.*, vol. 20, no. 24, pp. 26236–26251, 2012, doi: 10.1364/oe.20.026236.

[65] Y. Fan *et al.*, "Modulation format dependent phase noise caused by intra-channel nonlinearity," in *2012 European Conference on Optical Communication, ECOC*, 2012, pp. We.2.C.3.

[66] P. J. Freire, A. Napoli, B. Spinnler, N. Costa, S. K. Turitsyn, and J. E. Prilepsky, "Neural Networks-Based Equalizers for Coherent Optical Transmission: Caveats and Pitfalls," *IEEE J. Sel. Top. Quantum Electron.*, vol. 28, no. 4, pp. 1–23, 2022, doi: 10.1109/JSTQE.2022.3174268.

[67] T. A. Eriksson, H. Bulow, and A. Leven, "Applying Neural Networks in Optical Communication Systems: Possible Pitfalls," *IEEE Photon. Technol. Lett.*, vol. 29, no. 23, pp. 2091–2094, 2017, doi: 10.1109/LPT.2017.2755663.

[68] Z. Yang, F. Gao, S. Fu, M. Tang, and D. Liu, "Overfitting effect of artificial neural network based nonlinear equalizer: from mathematical origin to transmission evolution," *Sci. China Inf. Sci.*, vol. 63, no. 6, p. 160305, 2020, doi: 10.1007/s11432-020-2873-x.