A Fiber Nonlinearity Compensation Scheme With Complex-Valued Dimension-Reduced Neural Network

Pinjing He, Feilong Wu, Meng Yang, Aiying Yang, Peng Guo, Yaojun Qiao, and Xiangjun Xin

***Abstract*—A fiber nonlinearity compensation scheme based on a complex-valued dimension-reduced neural network is proposed. The proposed scheme performs all calculations in complex values and employs a dimension-reduced triplet feature vector to reduce the size of the input layer. Simulation and experiment results show that the proposed neural network needed only 20% of computational complexity to reach the saturated performance gain of the real-valued triplet-input neural network, and had a similar saturated gain to the one-step-per-span digital backpropagation. In addition, the proposed scheme was 1.7 dB more robust to the noise from training data and required less bit precision for quantizing trained weights, compared with the real-valued triplet-input neural network.**

***Index Terms*—Kerr effect, fiber nonlinearity compensation, neural network.**

I. Introduction

A DIRECT way to further enhance the capacity of coherent optical fiber communication systems is to compensate for fiber nonlinear impairments [1]. The mainstream fiber nonlinear impairments compensation (NLC) algorithms numerically estimate the solution of the nonlinear Schrödinger equation (NLSE) based on the split-step Fourier method (SSFM) [2]. These methods are named as the digital backpropagation (DBP) [3], [4]. In DBP, linear steps for chromatic dispersion (CD) compensation and nonlinear steps for nonlinear phase derotation are concatenated to simulate the interaction between CD and fiber nonlinearity. The main drawback of DBP is the high computational cost. To alleviate this issue, one kind of variant reduces the required number of steps by low-pass filtering the power terms in nonlinear steps [5]–[8]. DBP requires at least two samples per symbol. Another kind of NLC algorithm is based on the perturbation theory, which posits that the solution of NLSE is made up of linear and nonlinear perturbative components [9]. This kind of algorithm can be performed with a single sample per symbol at a single stage, leading to degraded performance but reduced computational cost compared with DBP [10]. The implementation of perturbation-based NLC can be placed either at the transmitter side [11], [12] or the receiver side [10], [13]–[15].

Recently, NLC algorithms based on machine learning (ML-NLC) received a lot of interest from academics. ML-NLC can adapt to complicated situations and has the potential to even further reduce complexity. A popular ML-NLC approach is fitting the inverse transmission system using the neural network (NN) [16], [17]. To improve the compensation performance, a helpful strategy is designing ML-NLC by utilizing the theory of fiber nonlinearity. One kind of ML-NLC algorithm following this strategy treats linear and nonlinear steps of DBP as layers and activation functions of NN [18]–[22]. Same as DBP, these approaches require at least two samples per symbol. Another type of such ML-NLC works with only one sample per symbol, by feeding an NN with intra-channel cross-phase modulation and intra-channel four-wave mixing triplet feature vector [23], [24]. The triplet feature vector is separated into independent real and imaginary components and is fed into the input layer of an NN. It is difficult for the triplet-input NN (TNN) to learn the correlation between the real part and the imaginary part in this way due to the lack of real-imaginary constraint, which makes the model sensitive to the noise and leads to the potential performance degradation [25]. Besides, the dimension of the triplet feature vector is extremely high for a long-haul transmission system. In this case, the size of the input layer in TNN is huge, leading to a high computational cost.

In this paper, an NLC scheme based on a complex-valued dimension-reduced-triplet-input neural network (CVTNN-DR) is proposed. All calculations of CVTNN-DR are performed in the complex-valued form, which keeps the correlation between real and imaginary components. In addition, according to the inherent symmetry of triplets, a dimension-reduced form of the triplet feature vector is derived, which significantly reduces the input layer size of CVTNN-DR. The proposed method has been investigated via both simulation and experiment. Results show that the proposed scheme needed less computational complexity

Manuscript received August 9, 2021; revised September 30, 2021; accepted October 25, 2021. Date of publication October 29, 2021; date of current version November 11, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 61427813, in part by the Open Fund of IPOC (BUPT) under Grant IPOC2018B003, and in part by the State Key Laboratory of Advanced Optical Communication Systems and Networks, China. *(Corresponding author: Aiying Yang.)*

Pinjing He, Feilong Wu, Meng Yang, Aiying Yang, and Peng Guo are with the Key Laboratory of Photonics Information Technology, Ministry of Industry and Information Technology, School of Optics and Photonics, Beijing Institute of Technology, Beijing 100081, China (e-mail: hepinjing93@gmail.com; 3120185344@bit.edu.cn; 3120205323@bit.edu.cn; yangaiying@bit.edu.cn; guopeng0304@bit.edu.cn).

Yaojun Qiao is with the Beijing University of Posts and Telecommunications, Beijing 100876, China (e-mail: qiao@bupt.edu.cn).

Xiangjun Xin is with the School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China (e-mail: xinxiangjun@bit.edu.cn).

Digital Object Identifier 10.1109/JPHOT.2021.3123624

to reach the saturated gain of TNN. Besides, it showed a similar saturated gain to one-step-per-span DBP. Further, CVTNN-DR showed stronger robustness to the amplified spontaneous emission (ASE) noise in the training data and the noise from quantization on weights, compared with TNN.

II. Principle

A. Dimension-Reduced Triplet Feature Vector

In this subsection, we will introduce the triplet feature vector. Then, we will derive a dimension-reduced triplet feature vector form that utilizes the inherent symmetry of the triplet. Let H_k be the k-th linearly equalized symbol for x polarization and V_k for y polarization. In the following, we will take x polarization as an example and omit the polarization symbol for the sake of simplicity. The triplet feature vector is from the first-order approximation to the nonlinear perturbation term ΔH_k, which is as follows [14]

$$\Delta H_k \approx \sum_{m,n} C_{m,n} \left(H_{k+m} H^*_{k+m+n} + V_{k+m} V^*_{k+m+n} \right) H_{k+n}, \quad (1)$$

where $C_{m,n}$ is the perturbative coefficient, determined by the pulse shape and the distribution of chromatic dispersion, power profile, and fiber nonlinearity. m and n are off-center symbol indexes. Then, a component of the triplet feature vector for x-polarization $T^x_{k,m,n}$ is defined as

$$T^x_{k,m,n} = H_{k+m} \left(H_{k+n} H^*_{k+m+n} + V_{k+n} V^*_{k+m+n} \right). \quad (2)$$

The triplet feature vector $\mathbf{T}_k$ can be derived by selecting and arranging $T_{k,m,n}$ into a sequence. Assuming the number of correlated symbols as $2N+1$, a common efficient way to select $T_{k,m,n}$ is picking tuple (m,n) in a set I, where $I = \{(m,n) | m,n \in \mathbb{Z} \cap |mn| < N \cap |m| < N \cap |n| < N\}$ and $\cap$ denotes the logical conjunction operator [26]. Then, the triplet feature vector can be written as

$$\mathbf{T}_k = \begin{bmatrix} \vdots \\ T_{k,m,n} \\ \vdots \end{bmatrix}, \quad (m,n) \in I. \quad (3)$$

The contribution to the perturbative term ΔH_k from the triplet component $T_{k,m,n}$ has some symmetry, which is determined by perturbative coefficients $C_{m,n}$. The perturbative coefficient has its inherent symmetry, i.e., $C_{m,n} = C_{n,m}$. If the pulse shape is symmetric with respect to the pulse center, e.g., the root-raised cosine pulse, $C_{m,n} = C_{-m,-n}$[14]. The whole symmetry of coefficients can be written as $C_{m,n} = C_{n,m} = C_{-m,-n} = C_{-n,-m}$. A heuristic approach to reduce the dimension of triplet feature vectors is adding up triplet components that correspond to the same coefficients. Arrange these sums into a sequence, we derive a new triplet feature vector $\mathbf{T}'_k$, which has fewer elements than $\mathbf{T}_k$. More concretely, pick $I' \subset I$, where $I' = \{(m,n) | 0 \leq m,n \leq M \cap |mn| < M \cap m \geq n\}$. Then, $\mathbf{T}'_k$ can be written as

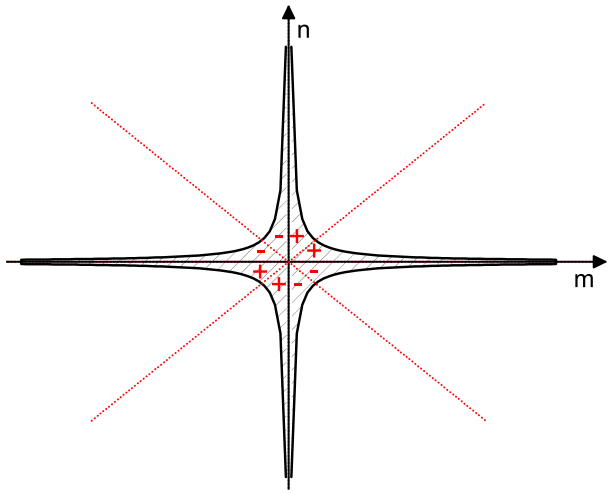

Fig. 1. The scheme for reducing the dimension of the triplet feature vector. The two axes denote the two off-center indexes m and n. The denotation "+" in an area denotes m and n have the same signs and the denotation "-" for different signs. Triplet components with the same denotations are combined to form new components of the dimension-reduced triplet feature vector.

$$\mathbf{T}'_k = \begin{bmatrix} \vdots \\ T_{k,m,n} + T_{k,n,m} + T_{k,-m,-n} + T_{k,-n,-m} \\ \vdots \\ T_{k,-m,n} + T_{k,n,-m} + T_{k,m,-n} + T_{k,-n,m} \\ \vdots \end{bmatrix}, \quad (m,n) \in I'. \quad (4)$$

Fig. 1 depicts how this scheme reduces the dimension. The area filled with sparse slash marks denotes set I. Triplet components in the area marked by the same denotation are added up to form new components. According to the figure, Eq. (3) and Eq. (4), one can conclude that the dimension of $\mathbf{T}'_k$ is only 25% of that of $\mathbf{T}_k$.

B. Complex-Valued Triplet-Input Neural Network

We will introduce the principle of the complex-valued triplet-input neural network (CVTNN) in this subsection. Fig. 2 shows the schematic diagram of CVTNN. In the nonlinear feature mapping part, the triplet feature vector is calculated. Then, it is fed into CVTNN for NLC. The nonlinear activation function is another important part of a neural network. We choose *modReLU* function as the activation function [27]. The CVTNN outputs the estimated NLI term ΔH_k. The equalized k-th symbol $\hat{H}_k$ is derived by $\hat{H}_k = H_k + \xi \Delta H_k$, where ξ is a power scaling factor.

Same as other neural networks, the CVTNN has two stages, i.e., the training stage for training weights and the test stage. At the training stage, ξ is set to 1. The triplet feature vector is calculated from symbols that have been linearly equalized. Weights are optimized by the complex gradient descent algorithm based on Wirtinger calculus [28]. At the test stage, ξ is

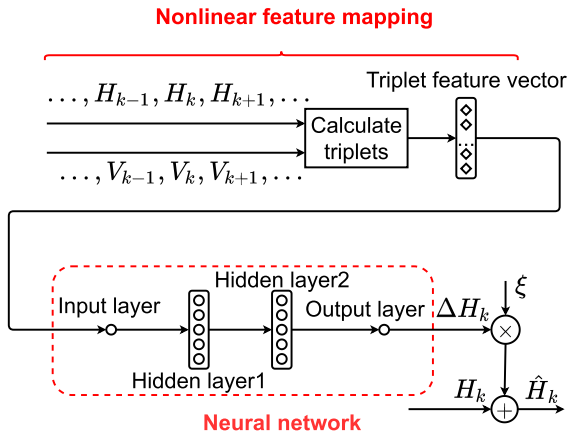

Fig. 2. The schematic diagram of CVTNN.

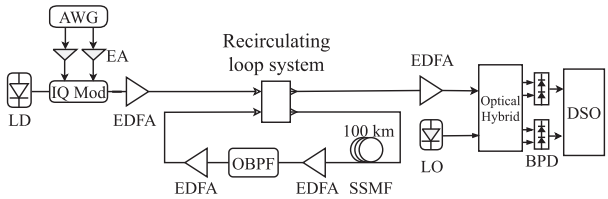

Fig. 3. The experimental platform. LD - laser diode. AWG - arbitrary wave generator. EA - electrical amplifier. EDFA - Erbium-doped fiber amplifier. OBPF - optical band-pass filter. LO - local oscillator. BPD - balanced photodetection. DSO - digital sampling oscilloscope.

optimized for different signal powers. The calculation of the triplet feature vector is different at the training stage. All possible triplet values from three symbols, e.g., 16-QAM symbols, are saved at a pre-defined look-up table. A triplet component is picked from the look-up table according to the decisions of the three linear-equalized symbols.

III. Results

A. System Descriptions and Implementation Details

We used VPItransmissionMaker for implementing the simulation. A single channel 256 Gbps dual-polarization 16-QAM transmission was simulated. The simulation system consisted of 40 spans of 80 km standard single-mode fiber (SSMF), with chromatic dispersion (CD) of 17 ps/nm/km, the nonlinearity of 1.3 /W/km. The noise figure of Erbium-doped fiber amplifiers (EDFA) was set to 4.5 dB. Root-raised-cosine pulses with a roll-off factor of 0.01 were performed. At the transmitter, 50% CD pre-distortion was performed. The step precision of SSFM in the simulation was specified by the maximum tolerable phase change across one step, which was 0.05 degrees.

A single channel 80 Gbps single-polarization 16-QAM transmission experimental system was also set for further investigation. The diagram of the experimental platform is shown in Fig. 3. Signals were generated by an arbitrary wave generator (AWG). The wavelength of lasers was 1550 nm. We used a recirculating loop system to realize an 1800 km transmission, with 100 km SSMF in the loop. An optical band-pass filter (OBPF) was put in the loop to suppress the accumulation of the

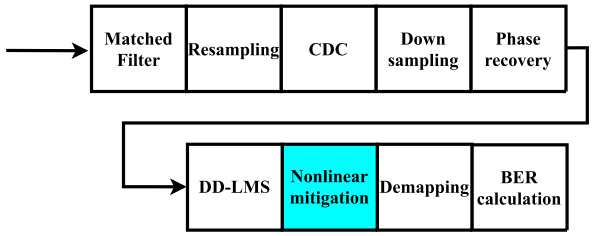

Fig. 4. The digital signal process procedure. CDC: chromatic dispersion compensation. DD-LMS: decision-directed least mean square filter.

ASE noise. After the coherent receiver, signals were sampled by a digital sampling oscilloscope (DSO) with a sampling rate of 50 GSa/s. The DSP procedure was performed offline.

Fig. 4 depicts the digital signal process (DSP) procedure at the receiver side. After the linear equalization procedure, the nonlinear mitigation algorithm was performed. For comparison, the perturbation-based post-compensation (PPC) [10], the TNN [23] and the DBP with 1 step per span were performed. A Q factor was utilized as the performance metric, which was calculated as $Q = \sqrt{2}\mathrm{erfc}^{-1}(2\mathrm{BER})$, where $\mathrm{erfc}^{-1}(\cdot)$ was the inverse error function and BER was the bit error ratio. Weights of neural networks were trained by minimizing a cost function. We chose the mean square error function as the cost function and *Adam* as the optimizer [29]. The learning rate for training was 1×10^{-3} and the size of a mini-batch is 128 feature vectors. The symbols in the training dataset and the test dataset were generated by pseudo-random bit sequence (PRBS). To exclude the impact on the performance from the PRBS pattern, seeds for generating the two datasets were different. There were 131072 symbols in a dataset for simulation and 110000 symbols for the experiment.

B. Results of Complex-Valued Neural Network

Fig. 5 shows how the equalization performance varies with the launch power. In each subfigure, figures on the right are constellations for corresponding situations. Fig. 5(a) shows the results in the simulation. The number of correlated symbols, which was 185 in the simulation and 101 in the experiment. The corresponding dimension of the triplet feature vector were 2085 and 1005 in the simulation and experiment, respectively. The CVTNN and the TNN had 4 layers, i.e., the input layer, the first hidden layer, the second hidden layer, and the output layer. The structures of NNs and corresponding number of real-valued multiplications per symbol (RMPS) are shown in Table I and in Table II. The Q factor obtained with the CVTNN scheme was 1.2 dB higher than that of LE at the launch power of 1 dBm, which validates the capability of CVTNN for NLC. An approximately 0.4 dB Q factor improvement was observed compared with PPC, and about 0.2 dB improvement compared with the TNN scheme and DBP. An experimental investigation was also performed. Results are showed in Fig. 5(b). The Q factor of CVTNN was approximately 1.0 dB higher than that of LE at the launch power of 1 dBm, and 0.1 dB higher than the TNN equalization system.

TABLE I
THE STRUCTURE OF NN FOR SIMULATION

Model	Input layer nodes	First hidden layer nodes	Second hidden layer nodes	Output layer nodes	RMPS
CVTNN	2085	1	5	1	6372
CVTNN-DR	624	1	5	1	1627
TNN	4170	2	10	2	8468
TNN-DR	1248	2	10	2	2140

TABLE II
THE STRUCTURE OF NN FOR EXPERIMENT

Model	Input layer nodes	First hidden layer nodes	Second hidden layer nodes	Output layer nodes	RMPS
CVTNN	1005	1	5	1	3132
CVTNN-DR	310	1	5	1	872
TNN	2010	2	10	2	4147
TNN-DR	620	2	10	2	1133

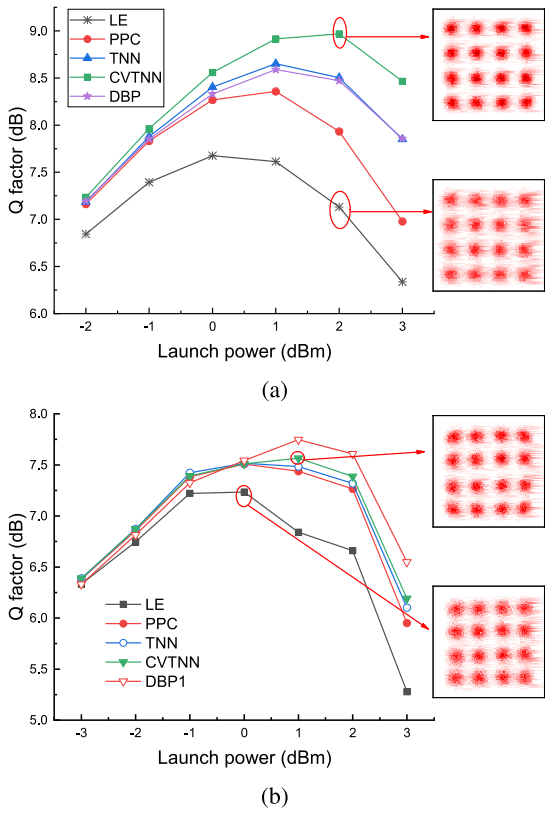

Fig. 5. The signal performance versus the launch power in (a) the simulation and (b) the experiment. The real-valued multiplications of CVTNN was approximately 6372 and that of TNN was 8468.

C. Computational Complexity

The computational complexity is another critical issue for NLC. We choose RMPS as the computational complexity metric. All complex-valued multiplications are based on Karastuba algorithm [30], which is equivalent to 3 real-valued multiplications. Concretely, let $x, w \in \mathbb{C}$, where $\mathbb{C}$ denotes the set of complex numbers. The procedure for calculating $y = wx$ by a Karastuba multiplication is written as

$$y = (k_1 - k_3) + j(k_1 + k_2), \tag{5}$$

where

$$\begin{cases} k_1 &= \Re(x)\left(\Re(w) + \Im(w)\right) \\ k_2 &= \Re(w)\left(\Im(x) - \Re(x)\right) \\ k_3 &= \Im(w)\left(\Re(x) + \Im(x)\right) \end{cases} \tag{6}$$

where $\Re(\cdot)$ and $\Im(\cdot)$ are the real part and the imaginary part, respectively. $j = \sqrt{-1}$.

Before discussing neural networks, the computational complexity of DBP is reviewed first, which is calculated as

$$RMPS_{\text{DBP}} = \frac{N_{\text{step}} f_{\text{sam}}}{f_{\text{sym}}} \left(RMPS_{\text{FDE}} + 7\right), \tag{7}$$

where N_{step} denotes the number of steps. f_{sam} denotes the sampling frequency. f_{sym} is the symbol rate. In this paper, $f_{\text{sam}}/f_{\text{sym}} = 2$. $RMPS_{\text{FDE}}$ denotes the computational complexity of the frequency-domain chromatic dispersion equalization, which is calculated as

$$RMPS_{\text{FDE}} = \frac{3N_{\text{FFT}}(\log_2 N_{\text{FFT}} + 1)}{N_{\text{FFT}} - K + 1}, \tag{8}$$

where N_{FFT} is the block size for implementing the fast Fourier transform algorithm. K denotes the number of filter taps, which determined by the considered amount of chromatic dispersion [31]. The factor 3 comes from the usage of Karastuba algorithm.

For the CVTNN based scheme, the test stage is considered, similar to [32]. The RMPS of CVTNN is calculated as follows

$$RMPS_{\text{CVTNN}} = RMPS_{\text{FDE}} + 3 \times (N_1 D_{\text{tri}} + N_1 N_2 + N_2 N_{\text{O}}), \tag{9}$$

where D_{tri} denotes the dimension of the triplet feature vector, which equals the number of nodes in the input layer. N_1 and N_2 denote the number of nodes of the first and the second hidden layer of CVTNN, respectively. N_{O} denotes the number of nodes in the output layer, which is constantly 1. Terms in parentheses of Eq.(9) are the number of complex-valued multiplications and the factor 3 comes from three real-valued multiplications for a complex-valued multiplication. The computational complexity of calculating the triplet feature vector is ignored because of the usage of the look-up table.

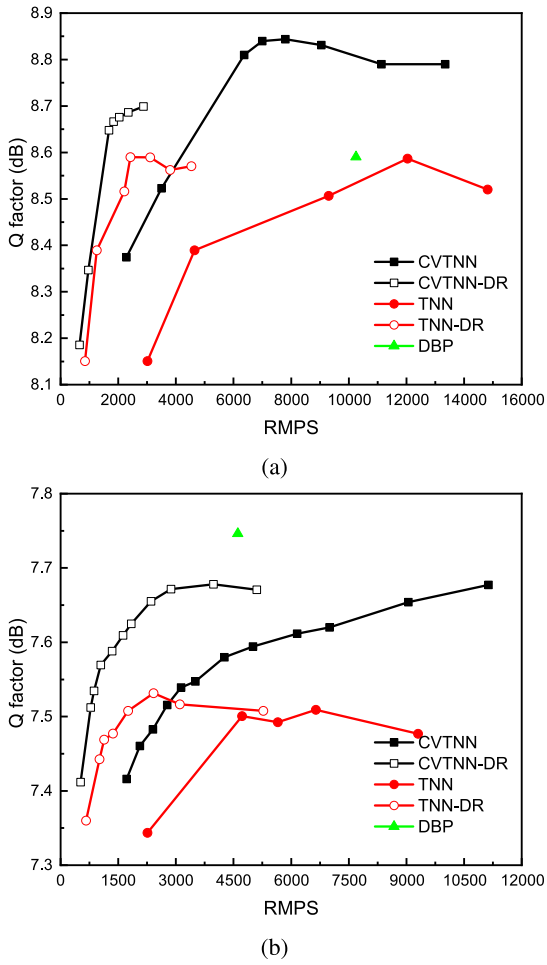

Fig. 6. The signal performance versus the real-valued multiplication per symbol in (a) the simulation and (b) the experiment.

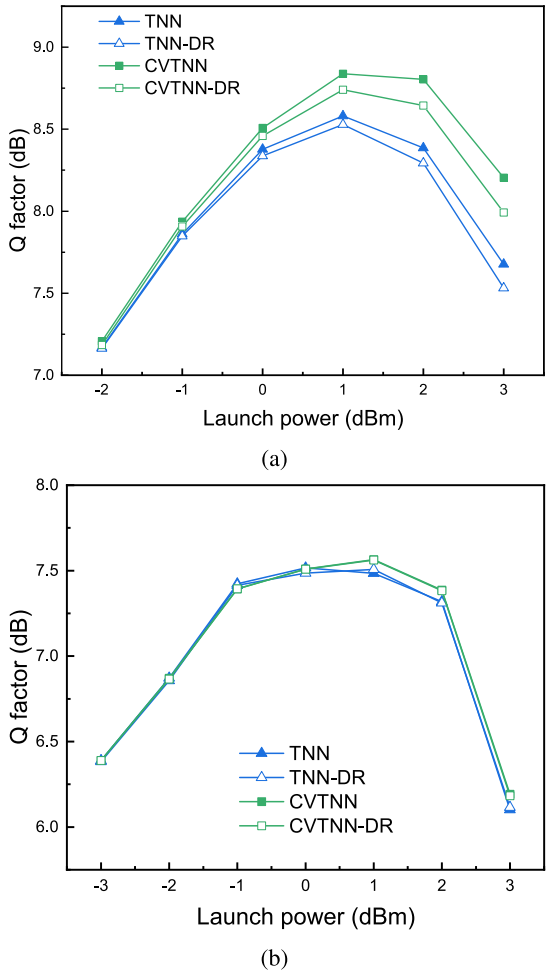

Fig. 7. The equalization performance versus the launch power in (a) the simulation and (b) the experiment. "DR" denotes the dimension-reduced scheme.

Similarly, the RMPS of TNN can be calculated layer by layer as follows

$$RMPS_{\mathrm{TNN}} = RMPS_{\mathrm{FDE}} + N_1' D_{\mathrm{tri}} + N_1' N_2' + N_2' N_{\mathrm{O}}', \tag{10}$$

where N_1' and N_2' are the number of nodes of the first and second hidden layer in TNN, respectively. In the case of 185 symbols correlated with each other, the dimension of the triplet feature vector D_{tri} is 2085.

Fig. 6 shows how the Q factor varies with the computational complexity. The computational complexity changed with the number of correlated symbols and the architecture of the two kinds of NNs, i.e., the number of nodes in the input layer and hidden layers. The number of nodes in the output layer was kept constant. Both in the simulation and in the experiment, CVTNN-DR outperformed TNN-DR with the same RMPS. Comparing TNN, CVTNN-DR only needed approximately 10% and 20% RMPS to reach the saturated gain of TNN in the simulation and experiment, respectively. These results prove the efficiency of CVTNN-DR. Comparing CVTNN, CVTNN-DR had a slightly saturated performance degradation, i.e., approximately 0.2 dB in the simulation and less than 0.1 dB in the experiment. However, CVTNN-DR needed significantly fewer RMPS to reach the saturated performance gain, which validates that the proposed dimension-reduced scheme can efficiently reduce the complexity. The saturated performance gains of CVTNN-DR were similar to DBP both in the simulation and experiment. The reason why DBP slightly outperformed CVTNN-DR in the experiment may be the corruption on the dataset caused by the device impairments.

D. Results of Dimension-Reduced Scheme

In this subsection, we will investigate the impact of the dimension-reduced scheme on equalization performance. The performance with and without the dimension-reduced scheme is shown in Fig. 7. Fig. 7(a) is the simulation result. The number of correlated symbols were 185 and 101 in the simulation and experiment, respectively. NN structures were shown in Table I and Table II. There was a slight performance difference between CVTNN and CVTNN-DR. A gap also appeared between TNN and TNN-DR. The reason may be that the symmetric property of pulses did not hold because of the ASE noise. Fig. 7(b) shows results in the experiment. Gaps between with and without the dimension-reduced scheme were

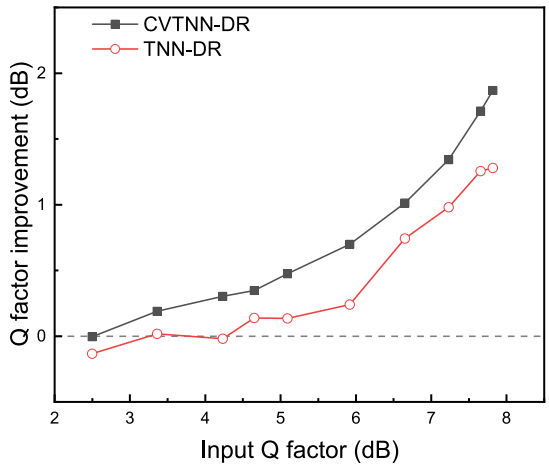

Fig. 8. Q factor improvement versus the Q factor of input data.

negligible. The discrepancy between the two groups of results might come from reduced fiber nonlinearity due to the shorter transmission distances and the single polarization transmission in the experiment. Overall, these results validate the efficiency of the dimension-reduced scheme.

E. Robustness to the Noise

In this subsection, the impact of the ASE noise in the training data and the quantization noise on the equalization performance will be investigated. Fig. 8 shows that the Q factor improvement of different NLC schemes changes with the quality of the input signal. The quality metric we used was the Q factor of the input signal. The signal was derived by a single transmission at the optimal launch power. Then, the signal was corrupted with different amounts of Gaussian noise to generate different training datasets, with different qualities but the same amount of fiber nonlinearity. Both CVTNN-DR and TNN-DR failed to compensate for fiber nonlinearity when the Q factor of the input signal was lower than 2.5 dB. When the input Q factor was greater than 2.5 dB, CVTNN-DR based NLC scheme improves the signal quality, while TNN-DR still failed to improve the signal quality until the input Q factor reached 4.2 dB. The quality threshold of CVTNN-DR was 1.7 dB lower than that of TNN-DR, which indicates that CVTNN-DR is more robust to the ASE noise.

A well-trained NN is supposed to be performed on specific hardware, e.g., application-specific integrated circuit (ASIC). The quantization of trained weights is essential on ASIC because it can significantly reduce the power dissipation and the calculating latency. However, the quantization noise of a given bit precision will cause degradation in the equalization performance. We will carry out an investigation on the robustness to quantization noise. We trained CVTNN-DR and TNN-DR with floating-point precision both in the simulation and in the experiment at the optimal launch power. Then, trained weights were transformed to fixed-point representation for a given bit precision. Concretely, real or imaginary components of weights in a layer were normalized by the maximum absolute value at first. The normalized weights were then quantized according

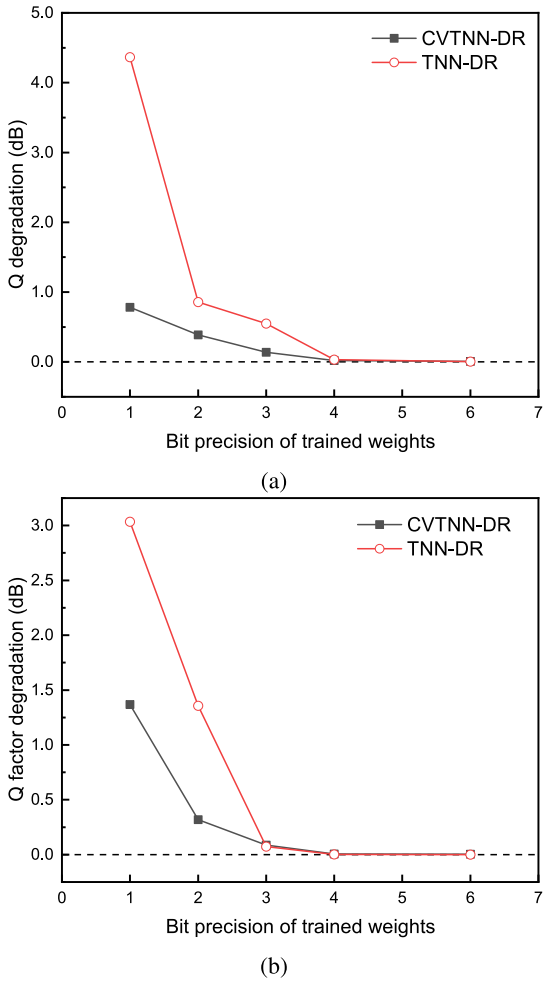

Fig. 9. The degradation of equalization performance caused by quantization on trained weights varies with the bit precision in (a) the simulation and (b) the experiment. "DR" denotes the dimension-reduced scheme.

to the given bit precision and scaled up by the same factor. We chose the performance with floating-point weights as the baseline and treated it as the full precision. Fig. 9 depicts the performance degradation variation with the bit precision of the trained weights. It is worth noting that we did not quantize the signal samples. In the simulation, CVTNN-DR achieved the same equalization performance as the baseline if the bit precision was 3. Such saturate bit precision for TNN-DR was 4. With 3 b, the Q degradation of CVTNN-TR was about 0.5 dB lower than that of TNN-DR. In the experiment, the saturated bit precision of CVTNN-DR became 2 and that of TNN-DR was 3. With 2 bits, the Q factor degradation of CVTNN-DR was approximately 1 dB lower than that of TNN-DR. These results validate that CVTNN-DR based scheme is more robust to the quantization noise and requires fewer bits for implementation on hardware.

IV. Summary

A complex-valued dimension-reduced-triplet-input neural network is proposed for equalizing the fiber nonlinearity of the long-haul transmission system. All calculations in the proposed scheme are performed in complex values. Besides, a

dimension-reduced form of the triplet feature vector is employed to reduce the size of the input layer thus reduce the computing complexity. The proposed scheme has been investigated both in simulation and experiment. In the simulation, results showed that the Q factor of CVTNN was 0.2 dB higher than that of TNN. In the experiment, the Q factor of CVTNN was about 0.1 dB higher than that of TNN. CVTNN-DR also showed a similar performance gain to one-step-per-span DBP. Because of the usage of Karastuba algorithm and the dimension-reduced feature vector, CVTNN-DR only needed less than 20% RMPS to reach the saturated gain of TNN. We also performed the investigation for the robustness to the noise. Results showed that the proposed CVTNN-DR based scheme was 1.7 dB more robust to the ASE noise and required less bit precision for quantizing trained weights. We believe that with the help of quantization on trained weights, the proposed scheme is a promising candidate for implementation on DSP hardware such as ASIC.

REFERENCES

[1] A. Mecozzi and R. J. Essiambre, "Nonlinear shannon limit in pseudolinear coherent systems," *J. Lightw. Technol.*, vol. 30, no. 12, pp. 2011–2024, 2012.

[2] G. P. Agrawal, *Fiber-Optic Communication Systems*, vol. 222. Amsterdam, Netherlands: Wiley, 2012.

[3] E. Ip and J. M. Kahn, "Compensation of dispersion and nonlinear impairments using digital backpropagation," *J. Lightw. Technol.*, vol. 26, no. 20, pp. 3416–3425, 2008.

[4] E. Ip, "Nonlinear compensation using backpropagation for polarization-multiplexed transmission," *J. Lightw. Technol.*, vol. 28, no. 6, pp. 939–951, 2010.

[5] L. B. Du and A. J. Lowery, "Improved single channel backpropagation for intra-channel fiber nonlinearity compensation in long-haul optical communication systems," *Opt. Exp.*, vol. 18, no. 16, 2010, Art. no. 17075.

[6] D. Rafique, M. Mussolin, M. Forzati, J. Mårtensson, M. N. Chugtai, and A. D. Ellis, "Compensation of intra-channel nonlinear fibre impairments using simplified digital back-propagation algorithm," *Opt. Exp.*, vol. 19, no. 10, pp. 9453–9460, 2011.

[7] Y. Gao, J. H. Ke, K. P. Zhong, J. C. Cartledge, and S. S. Yam, "Assessment of intrachannel nonlinear compensation for 112 Gb/s dual-polarization 16QAM systems," *J. Lightw. Technol.*, vol. 30, no. 24, pp. 3902–3910, 2012.

[8] X. Liang and S. Kumar, "Multi-stage perturbation theory for compensating intra-channel nonlinear impairments in fiber-optic links," *Opt. Exp.*, vol. 22, no. 24, 2014, Art. no. 29733.

[9] A. Vannucci, P. Serena, and A. Bononi, "The RP method: A. new tool for the iterative solution of the linear Schrödinger equation," *J. Lightw. Technol.*, vol. 20, no. 7, pp. 1102–1112, 2002.

[10] A. Ghazisaeidi *et al.*, "Submarine transmission systems using digital nonlinear compensation and adaptive rate forward error correction," *J. Lightw. Technol.*, vol. 34, no. 8, pp. 1886–1895, 2016.

[11] 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.

[12] L. Dou *et al.*, "Real-time 112 Gb/s DWDM coherent transmission with 40 extended reach by transmitter-side low-complexity nonlinear mitigation," in *Proc. Eur. Conf. Opt. Commun.*, 2012, pp. 12–14.

[13] T. Oyama *et al.*, "Robust and efficient receiver-side compensation method for intra-channel nonlinear effects," *Opt. InfoBase Conf. Papers*, vol. 1, no. 3, pp. 8–10, 2014.

[14] 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, 2016, Art. no. 4097.

[15] E. Averyanov, A. Redyuk, O. Sidelnikov, M. Soroklna, M. Fedoruk, and S. Turitsyn, "Perturbative machine learning technique for nonlinear impairments compensation in WDM systems," in *Proc. Eur. Conf. Opt. Commun.*, 2018, pp. 1–3.

[16] S. T. Ahmad and K. P. Kumar, "Radial basis function neural network nonlinear equalizer for 16-QAM coherent optical OFDM," *IEEE Photon. Technol. Lett.*, vol. 28, no. 22, pp. 2507–2510, Nov. 2016.

[17] 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. 32 765–32 776, 2018.

[18] C. Häger, H. D. Pfister, R. M. Bütler, G. Liga, and A. Alvarado, "Revisiting multi-step nonlinearity compensation with machine learning," in *Proc. 45th Eur. Conf. Opt. Commun.*, 2019, pp. 1–4.

[19] 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," *Nature Commun.*, vol. 11, no. 1, pp. 1–11, 2020.

[20] V. Oliari *et al.*, "Revisiting efficient multi-step nonlinearity compensation with machine learning: An experimental demonstration," *J. Lightw. Technol.*, vol. 38, no. 12, pp. 3114–3124, 2020.

[21] B. I. Bitachon, A. Ghazisaeidi, M. Eppenberger, B. Baeuerle, M. Ayata, and J. Leuthold, "Deep learning based digital backpropagation demonstrating SNR gain at low complexity in a 1200 km transmission link," *Opt. Exp.*, vol. 28, no. 20, 2020, Art. no. 29318.

[22] P. He, A. Yang, P. Guo, Y. Qiao, and X. Xin, "A layer-reduced neural network based digital backpropagation algorithm for fiber nonlinearity mitigation," *IEEE Photon. J.*, vol. 13, no. 3, Jun. 2021, Art. no. 7200512.

[23] 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 *Proc. Opt. Fiber Commun. Conf.*, Washington, D.C., USA, 2018, Art. no. Th4D.5.

[24] S. Zhang *et al.*, "Field and lab experimental demonstration of nonlinear impairment compensation using neural networks," *Nature Commun.*, vol. 10, no. 1, pp. 1–8, 2019.

[25] A. Hirose and S. Yoshida, "Generalization characteristics of complex-valued feedforward neural networks in relation to signal coherence," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 4, pp. 541–551, Apr. 2012.

[26] 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.

[27] M. Arjovsky, A. Shah, and Y. Bengio, "Unitary evolution recurrent neural networks," in *Proc. Int. Conf. Mach. Learn.*, 2016, pp. 1120–1128.

[28] A. Hirose, *Complex-Valued Neural Networks: Advances and Applications*, vol. 18. Hoboken, NJ, USA: Wiley, 2013.

[29] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Representations, - Conf. Track Proc.*, 2015, pp. 1–15.

[30] Y. O. A. Karatsuba, "Multiplication of many-digital numbers by automatic computers," *Doklady Akademii Nauk SSSR*, vol. 145, pp. 293–294, 1962.

[31] B. Spinnler, "Equalizer design and complexity for digital coherent receivers," *IEEE J. Sel. Topics Quantum Electron.*, vol. 16, no. 5, pp. 1180–1192, Sep./Oct. 2010.

[32] P. J. Freire *et al.*, "Performance versus complexity study of neural network equalizers in coherent optical systems," *J. Lightw. Technol.*, vol. 23, no. 19, pp. 6085–6096, 2021.