Complex Principal Component Analysis-Based Complex-Valued Fully Connected NN Equalizer for DP-64 QAM Coherent Detection

1st Xingyuan Huang
Beijing University of Posts and Telecommunications (BUPT)
School of Electronic Engineering
Beijing, China
yuan0624@bupt.edu.cn

2nd Yongjun Wang*
Beijing University of Posts and Telecommunications (BUPT)
School of Electronic Engineering
Beijing, China
wangyj@bupt.edu.cn

3rd Chao Li
Beijing University of Posts and Telecommunications (BUPT)
School of Electronic Engineering
Beijing, China
lythao@bupt.cn

4th Lu Han
Beijing University of Posts and Telecommunications (BUPT)
School of Electronic Engineering
Beijing, China
clear@bupt.edu.cn

5th Qi Zhang
Beijing University of Posts and Telecommunications (BUPT)
School of Electronic Engineering
Beijing, China
zhangqi@bupt.edu.cn

6th Xiangjun Xin
Beijing University of Posts and Telecommunications (BUPT)
School of Electronic Engineering
Beijing, China
xjxin@bupt.edu.cn

***Abstract*—A novel complex principal component analysis algorithm is applied to coherent optical detection based on complex-valued fully connected NN equalizer. Compared with real-valued equalizer, the time and space complexity are reduced by $\sim 40\%$ and $\sim 70\%$.**

***Index Terms*—Complex principal component analysis, complex-valued neural network**

I. Introduction

A significant problem limiting the capacity of optical fiber channels has always been nonlinear damage [1]. As artificial intelligence is increasingly used in optical fiber communication, more and more machine learning methods are being developed to address various types of nonlinear optical fiber damage. A class of nonlinear models known as neural networks (NNs) only require a large number of optical signals at the receiver side in order to train the networks to establish the mapping relationship between the input and output [2]–[4]. Finally, nonlinear equalization is performed using the learned mapping model. It should be noted that optical signal processing in optical fiber communication systems typically represents digital signals as complex forms. Unfortunately, the link between the real and fictitious components of complex signals is not taken into account by the majority of existing real-valued neural network frameworks. As a result, a class of complex-valued neural networks (CVNNs) that process information using complex-valued parameters and variables has arisen [5], [6].

In this paper, we first apply the combination of perturbation theory-aided and CVNNs in coherent optical communication nonlinear equalization. Furthermore, to reduce space-time complexity, we extend the principal component analysis (PCA) method to the complex domain to fit our designed framework, which we call complex PCA ($\mathbb{C}$PCA). Based on the above, we propose a $\mathbb{C}$PCA-based complex-valued fully connected neural network (P-CFNN) model driven by perturbation theory to mitigate nonlinear effects in optical fibre communication systems.

II. Data-Driven Model

A. Construction of Data Features

According to the well-known Manakov equation representing the propagation of optical pulse, we can obtain the intra-channel cross-phase modulation (IXPM) and intra-channel four-wave mixing (IFWM) triplets from the received symbols aided by first-order perturbation theory are expressed by T_x and T_y in x-polarization and y-polarization, respectively [7]:

$$\begin{cases} T_x = X_m X_{m+n}^* X_n + X_m Y_{m+n}^* Y_n \\ T_y = Y_m Y_{m+n}^* Y_n + Y_m X_{m+n}^* X_n \end{cases}, \tag{1}$$

where m, n, and $m+n$ are symbol indices with respect to the specified symbol. $X_{m/n/m+n}$ and $Y_{m/n/m+n}$ denote the sequence of symbols on the x- and y-polarization, respectively, and * expresses the complex conjugation. The triplets are

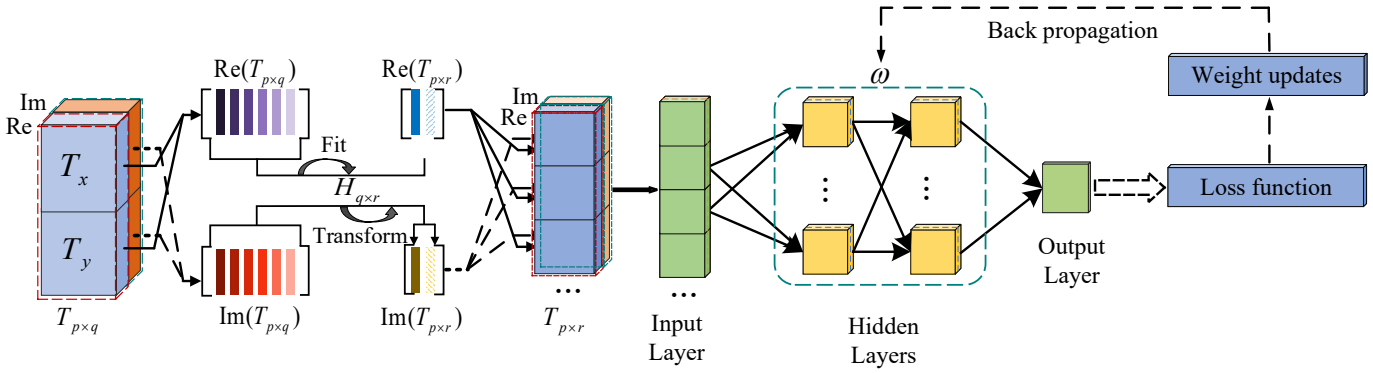

Fig. 1. The P-CFNN architecture consists of input features after dimensionality reduction by ℂPCA, the input layer, two fully connected hidden layers, the output layer, and back propagation constantly updating the weights. ℂPCA diagram.

primarily dominated by the factors m and n. In this paper, we adopt the following rules to define m and n [8]:

$$\begin{cases} |m| + |n| \le K \\ |m \times n| \le K \end{cases}, \tag{2}$$

where K is a hyperparameter that controls the number of triplets.

The PCA method is one of the most widely used dimension reduction algorithms for data. As shown in Fig. 1, we assume that T_x and T_y analysed through perturbation theory have p samples and q dimensional features for each sample, which form a complex matrix $T_{p\times q}$. We believe that a better r dimension feature should satisfy the maximum variance of the projection point of the q dimension sample. We refer the real part data of the triplets as $s_i (1 \le i \le p)$, and the ℂPCA we propose first projects data point s_i onto base $h_j (1 \le j \le r)$. Its projection distance is $s_i^{\mathrm{T}} h_j$, where T denotes the transpose of the matrix. The variance D_j of all data projected on this base is expressed as:

$$D_j = \frac{1}{p} \sum_{i=1}^{p} (s_i^{\mathrm{T}} h_j - s_{center}^{\mathrm{T}} h_j)^2. \tag{3}$$

Finally, the first r maximum eigenvectors of the maximum variance can be obtained after a series of calculations and transformations to create the dimensionality reduction matrix $H_{q\times r}$, as shown in Fig. 1:

$$H_{q\times r} = \{h_1, h_2, ...h_r\}_{q\times r}. \tag{4}$$

$T_{p\times q}$ is divided into two parts: the real part and the imaginary part. The real part of $T_{p\times q}$ is extracted using ℂPCA to fit the dimension reduction matrix $H_{q\times r}$. Then, the fitted dimension reduction matrix is applied to the imaginary part of $T_{p\times q}$ to transform the dimensions. Consequently, the real and imaginary components of the reduced dimension $T_{p\times r}$ can be expressed as:

$$\begin{cases} Re(T_{p\times r}) = Re(T_{p\times q}) \cdot H_{q\times r} \\ Im(T_{p\times r}) = Im(T_{p\times q}) \cdot H_{q\times r} \end{cases}. \tag{5}$$

Based on the above theories, we use first-order perturbation theory to analyse received symbols to obtain nonlinear impairment features. Then we specifically proposed a ℂPCA technique to reduce the overall dimension of the perturbation eigenvalues in complex forms, which preserves the complex property and can directly drive complex-valued fully connected neural network (CFNN).

B. Complex-Valued Neural Network Design

In this section, we present the design of our framework for CFNNs. Fig. 1 shows the CFNN architecture for the training process, in which the activation function, weight initialization and loss function are all complex values. We extended exponential linear unit (ELU) activation function to the complex ELU (ℂELU) in our CFNN [9]. ELU is an evolutionary version of a rectified linear unit (ReLU) activation function, which can maintain the activation function in a noise-robust state, and can be expressed as:

$$ELU = \begin{cases} a, & a > 0 \\ \mu\left(e^a - 1\right), & a \le 0 \end{cases}, \tag{6}$$

where μ is a hyperparameter that represents the slope of the negative section, typically set to 0.1. Consequently, the ℂELU, which applies separate ELUs on both the real and imaginary components of a neuron, is employed as CFNN hidden layer activation functions, i.e.,

$$\mathbb{C}ELU = ELU\left(Re\left(z\right)\right) + jELU\left(Im\left(z\right)\right). \tag{7}$$

In addition, we use the $\mathbb{C}SoftMax$ function as the output activation function to complete the label classification task of the received signals, which is expressed as:

$$\mathbb{C}SoftMax(z_i) = \frac{e^{|z_i|}}{\sum_{c=1}^{C} e^{|z_c|}}, \tag{8}$$

where z_i is the output value of the i-th ($i = 1, 2, 3, ..., C$) neuron and C is the number of output nodes (i.e., the number of classifications) in the output layer. To overcome the problem of local optimal solutions as much as possible, we use adaptive moment estimation (Adam) as our optimizer.

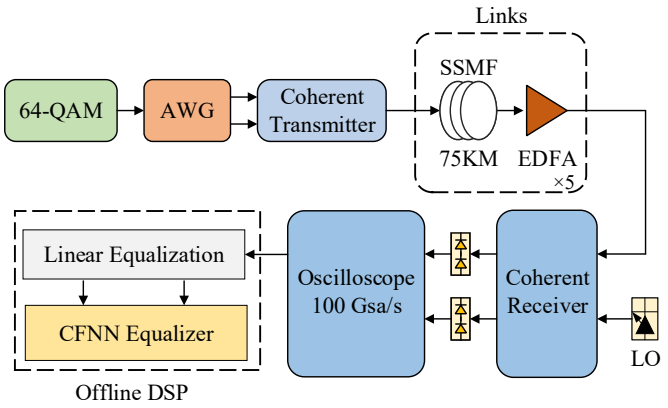

Fig. 2. Optical communication experimental platform for 120-Gb/s DP-64 QAM. AWG: arbitrary waveform generator; SSMF: standard single-mode fibre; EDFA: Erdium-doped fibre amplifier; LO: local oscillator.

III. Experimental Results Analysis

In this section, we validated the performance of the perturbation theory-aided P-CFNN classifier for DP-64 QAM symbolic decisions in terms of nonlinear compensation. We verified the feasibility of the proposed algorithm by transmitting 64-QAM signals in 120-Gbit/s coherent optical communication experimental platform as shown in Fig. 2. The symbol sequence of the pulse shaping 64-QAM signals generated by MATLAB is uploaded to an arbitrary waveform generator (AWG) at a sampling rate of 25 GSa/s on the transmitter side. The transmission chain is a multispan G.652D standard single-mode fibre (SSMF) with a length of 5×75 km. The chromatic dispersion and nonlinear index of the SSMF are $21.6676 \times 10^{-24} s^2/km$ and $1.3/(W \cdot km)$, respectively. At the receiver side, the offline DSP setup includes the linear equalization to restore useful signals generated by the demodulator and the CFNN equalizer to determine the symbols.

When training the neural network, we use two hidden layers with 32 and 64 neurons in the CFNN when designing our overall framework, taking both performance and computing speed into account. Considering that in CVNNs, one complex parameter ($P_{\mathbb{C}}$) is equivalent to two real parameters ($P_{\mathbb{R}}$) so that $P_{\mathbb{C}} = 2P_{\mathbb{R}}$, and one complex multiplication ($M_{\mathbb{C}}$) requires four multiplications ($M_{\mathbb{R}}$) of real numbers, i.e., $M_{\mathbb{C}} = 4M_{\mathbb{R}}$. In order to compare the performance of CFNN at an equivalent level of complexity in practical optical communication engineering, we set the number of neurons in the two layers of real-valued fully connected neural network (RFNN) to twice that of CFNN, i.e. 64 and 128. The datasets for each launched optical powers (LOP) contain approximately 2^{20} symbols, and we divided them into 50% for training and 50% for testing. To ensure data independence, the data pattern used in the training and test datasets has a maximum normalized cross-correlation of 0.5%. Furthermore, to avoid the negative effects of the sequence of data inputs on network training, we shuffle the data before each epoch to improve the generalization performance of the network.

As previously stated, K is a hyperparameter that governs the number of triplets and the feature dimension of the input signals. The number of triplets is positively correlated with K from the limit in Eq. (2). We tested the effects of real PCA on RFNNs and $\mathbb{C}$PCA on CFNNs, and plotted the Q-factor curves when K was set to 20 and gradually increased to 80 under 1 dBm LOP in Fig. 3(a). The relation between Q and the BER is expressed as $\mathrm{Q} = 20log_{10}[\sqrt{2}erfc^{-1}(2\mathrm{BER})]$. For the P-RFNN and P-CFNN equalizers, both the real PCA and $\mathbb{C}$PCA start dimensionality reduction from K=80. As shown in Fig. 3(a), when the selection dimensions of $\mathbb{C}$PCA are reduced from K=80 to K=50, ΔQ-factor obtains a maximum increase of approximately 0.57 dB. It can be seen that K from 80 to 50 is the optimal dimensionality reduction range. Hence we uniformly input the number of feature dimensions with K=50, in which both the P-CFNN and P-RFNN adopted a dimension reduction treatment with K from 80 to 50. We compared the Q-factor performance contrast of the RFNN, CFNN, P-CFNN, P-RFNN and the without (W/O) employing the NLC algorithm in Fig. 3(b), all of which were executed in the same experimental environment. Fig. 3(c) shows that the CFNN classifier can achieve 7% hard decision-forward error correction (HD-FEC) whose Q-factor is 8.52 dB when the LOP is from 0 dBm to 2 dBm, and the P-CFNN classifier reaches above 7% HD-FEC when the LOP is from -4 dBm to 4 dBm. Meanwhile, the Q-factor of P-CFNN is about 0.51 dB higher than that of P-RFNN at 1dBm. Therefore, CFNN and P-CFNN equalizers can achieve better compensation effects than RFNN and P-RFNN, respectively.

Our discussion of the overall algorithm complexity is divided into time complexity and space complexity. The multipliers can express time complexity because their operation complexity is several times that of adders, which is typically measured by the number of floating-point operations (FLOPs), and the number of parameters is commonly used to indicate the space complexity. Besides, the calculation times of the flow of the perturbation feature triplets in the network model contribute significantly to the complexity of equalizers. In this case, we adjust the number of input triplets to calculate the total FLOPs for each algorithm separately. For the sake of simplicity, we take the RFNN equalization performance at an LOP of 1 dBm as the benchmark value, where K is set to 80, N_T is the number of feature triplets corresponding to the K value, and the Q-factor is 8.77 dB. Next, we need to calculate the K value required by other NNs to reach the benchmark and calculated the corresponding space-time complexity. Based on the analysis above, the hidden layer of RFNN is twice that of CFNN, while their output layers of Eq. (8) for classification tasks are both 64 categories. As shown in TABLE I, we calculated when the Q-factor reached 8.77 dB, the whole number of real multiplications for the RFNN with $N_T = 1753$ is $2 \times 1753 \times 64 + 64 \times 128 + 128 \times 64 = 240768$, and the number of parameters indicating the space complexity is also 240768. But the N_T of the CFNN only needs to be set to 1005, in which case the corresponding number of FLOPs and parameters are $4{\times}1005{\times}32{+}4{\times}32{\times}64{+}4{\times}64{\times}64 = 153216$

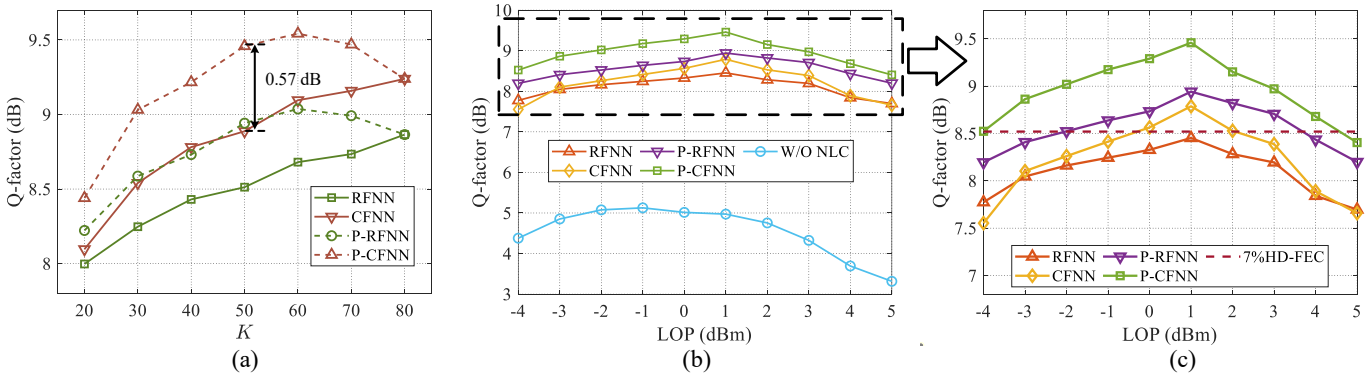

Fig. 3. Performance analysis of equalizers: (a) Curve of the Q-factor with K from 80 via PCA dimension reduction for when LOP is 1 dBm; (b) Q-Factor vs. different LOPs. RFNN: red; CFNN: yellow; P-RFNN: purple and P-CFNN: green; (c) The partial magnification of (b).

and $2 \times 1005 \times 32 + 2 \times 32 \times 64 + 2 \times 64 \times 64 = 76608$, respectively. Hence, the time complexity of the CFNN is reduced by 36% and the space complexity is reduced by 68% compared with the RFNN. Furthermore, after the real PCA algorithm and our designed $\mathbb{C}$PCA are applied to the RFNN and CFNN, we can calculate the FLOPs and parameters for the P-CFNN with N_T = 453 is 82560 and 41280 with the above method. Compared with the P-RFNN whose time and space complexity are both 137344, P-CFNN is reduced by 40% and 70%, respectively. According to the comprehensive analysis of Q-factor improvement under the same complexity shown in Fig. 3 and the complexity reduction under the same performance shown in TABLE I, the P-CFNN has absolute advantages over the P-RFNN. Consequently, we concluded that the CFNN fully preserves the connections between real and imaginary parts of the input signal features, and it is more advantageous as an equalizer than RFNN. Meanwhile, the proposed $\mathbb{C}$PCA algorithm is more effective on CFNN than the conventional PCA algorithm on RFNN, which can further explore the very large CVNNs potential.

TABLE I
COMPLEXITY OF NEURAL NETWORKS

	K	N_T	FLOPs	Parameters
FLOPs	80	1753	240768	240768
FLOPs	50	1005	153216	76608
P-RFNN	48	945	137344	137344
P-CFNN	26	453	82560	41280

IV. CONCLUSION

In this paper, we proposed a model of P-CFNN driven by perturbation theory and demonstrated it experimentally on a 375 km 120-Gbit/s DP-64 QAM coherent optical communication system. To further reduce the overall computational complexity of the model, we designed a novel $\mathbb{C}$PCA technique applied to CFNN. Under the constraint of the same Q-factor, we confirmed that the proposed P-CFNN we obtained a 40% reduction in time complexity and a 70% reduction in space complexity compared with the P-RFNN, which also proved the very large application prospect of the proposed $\mathbb{C}$PCA for CVNNs.

ACKNOWLEDGMENT

This work was supported by the the National Key Research and Development Program of China (2021YFB2900703) and the National Natural Science Foundation of China (62075014).

REFERENCES

[1] I. T. Lima, T. D. DeMenezes, V. S. Grigoryan, M. O'sullivan, and C. R. Menyuk, "Nonlinear compensation in optical communications systems with normal dispersion fibers using the nonlinear fourier transform," Journal of Lightwave Technology, vol. 35, no. 23, pp. 5056–5068, Oct. 2017.

[2] W. Xiong, B. Redding, S. Gertler, Y. Bromberg, H. D. Tagare, and H. Cao, "Deep learning of ultrafast pulses with a multimode fiber," APL Photonics, vol. 5, no. 9, p. 096106, Sep. 2020.

[3] Nan, Chi, Yiheng, Zhao, Meng, Shi, Peng, Zou, and Xingyu, "Gaussian kernel-aided deep neural network equalizer utilized in underwater PAM-8 visible light communication system." Optics express, vol. 26, no. 20, pp. 26 700–26 712, Oct. 2018.

[4] 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 Communications, vol. 11, no. 1, pp. 1–11, Jul. 2020.

[5] S. A. Bogdanov and O. S. Sidelnikov, "Use of complex fully connected neural networks to compensate for nonlinear effects in fibre-optic communication lines," Quantum Electronics, vol. 51, no. 5, p. 459, Feb. 2021.

[6] W. Zhou, J. Shi, L. Zhao, K. Wang, C. Wang, Y. Wang, M. Kong, F. Wang, C. Liu, J. Ding et al., "Comparison of real-and complex-valued nn equalizers for photonics-aided 90-Gbps D-band PAM-4 coherent detection," Journal of Lightwave Technology, vol. 39, no. 21, pp. 6858–6868, Nov. 2021.

[7] R. Alexey, A. Evgeny, S. Oleg, F. Mikhail, and T. Sergei, "Compensation of nonlinear impairments using inverse perturbation theory with reduced complexity," Journal of Lightwave Technology, vol. 38, no. 6, pp. 1250–1257, Feb. 2020.

[8] C. Li, Y. Wang, L. Han, S. Chen, Q. Zhang, L. Yang, and X. Xin, "Optical fiber nonlinearity equalizer with support vector regression based on perturbation theory," Optics Communications, vol. 507, no. 3, pp. 127 627–127 634, Mar. 2022.

[9] D.-A. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and accurate deep network learning by exponential linear units (elus)," in International Conference on learning Representations (ICLR), San Juan, Puerto Rico, 2016.