Machine Learning for 100 Gb/s/λ Passive Optical Network

Lilin Yi, Tao Liao, Luyao Huang, Lei Xue, Peixuan Li, and Weisheng Hu

(Invited Paper)

Abstract—**Responding to the growing bandwidth demand by emerging applications, such as fixed-mobile convergence for fifth generation (5G) and beyond 5G, 100 Gb/s/λ access network becomes the next research focus of passive optical network (PON) roadmap. Intensity modulation and direct detection (IMDD) technology is still considered as a promising candidate for 100 Gb/s/λ PON attributed to its low cost, low power consumption, and small footprint. In this paper, we achieve 100 Gb/s/λ IMDD PON by using 20G-class optical and electrical devices due to its commercial availability. To mitigate the system linear and nonlinear distortions, neural network (NN) based equalizer is used and the performance is compared with feedforward equalizer (FFE) and Volterra nonlinear equalizer (VNE). We introduce the rules to train and test the data while using NN-based equalizer to guarantee a fair comparison with FFE and VNE. Random data have to be used for training, but for test, both random data and pseudorandom bit sequence are applicable. We found that the NN-based equalizer has the same performance with FFE and VNE in the case of linear distortion, but outperforms them in a strong nonlinearity case. In the experiment, to improve the loss budget, we increase the launch power to 18 dBm, achieving a 30-dB loss budget for 33 GBd/s PAM8 signal at the system frequency response of 16.2 GHz, attributed to the strong nonlinear equalization capability of NN.**

Index Terms—**Digital signal processing (DSP), intensity modulation and direct detection (IMDD), machine learning, neural network (NN), passive optical network (PON).**

I. Introduction

Nowadays, increasing bandwidth-consuming applications and market-demanding factors have been driving the need for higher-speed access network, such as the rapid growth of the high-definition video streaming services, the burst of smart devices of Internet of Things (IoT) and the development of wireless backhaul of 5G [1], [2]. Recently IEEE 100G Ethernet passive optical network (EPON) task force was created and focused on standardizing a solution for next-generation low-cost EPON [3]. The objectives of the task force are to standardize a 25, 50 and 100G PON. By multiplexing two or four wavelengths with the bit rate of 25 Gb/s/λ, 50 Gb/s or 100 Gb/s system capacity can be achieved [4]. However, the gain spectrum of Erbium doped fiber amplifier (EDFA) is fixed between dozens of nanometers [5]. Though many experimental laboratory techniques have been invented to increase EDFA's bandwidth [6], [7], the useable EDFA bandwidth is still limited. With more and more wavelength occupied, it becomes the most valuable resource in optical communication. Besides, the capacity improvement from already-deployed 10G PON to 25G PON is only a small step for real deployment. Currently standard groups like IEEE 802.3ca and ITU-T SG15 are working on their 50 Gb/s/λ PON industry standard, with the goal of deployment in the next few years. Several feasible 50 Gb/s/λ solutions have been proposed during the past several years. Due to the nature requirement of low cost in PON, optics with limited bandwidth, advanced modulation formats and advanced digital signal processing (DSP) are widely chosen as the main research topic of 50 Gb/s/λ PON. Algorithms like feed-forward equalization (FFE), maximum likelihood sequence estimation (MLSE), and Volterra nonlinear equalization (VNE) have been investigated to overcome the limitation of channel impairment [8]–[12]. To further increase the loss budget, some optical functions have also been introduced such as dispersion shifted fiber (DSF) and semiconductor optical amplifier (SOA) [13]–[15].

While the research and standardization of 50 Gb/s/λ are steadily progressing, we decide to move forward and pay more attentions to the next step 100 Gb/s/λ PON research demanded by emerging applications such as fixed-mobile convergence for 5G and beyond 5G. Coherent technology is regarded as a promising candidate for such high bandwidth PON due to its high receiver sensitivity. 100 Gb/s/λ coherent PON has been experimentally demonstrated with receiver sensitivity of −33.8 dBm, which is difficult to be achieved for direct detection [16]. On the other hand, coherent technology is considered with high cost, high power consumption and large physical footprint. In contrary, intensity modulation and direct detection (IMDD) transmission system is obviously simpler and more cost-effective [17], and there have already been quite a lot research on IMDD based 100 Gb/s/λ transmission especially in short reach applications [18]. Different from the point-to-point short reach applications where loss budget is not an important parameter, 100 Gb/s/λ PON applications are more

Manuscript received September 15, 2018; revised November 30, 2018; accepted December 11, 2018. Date of publication December 18, 2018; date of current version March 28, 2019. This work was supported by National Natural Science Foundation of China under Grant 61575122. *(Corresponding author: Lilin Yi.)*

The authors are with the State Key Laboratory of Advanced Optical Communication Systems and Networks, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: lilinyi@sjtu.edu.cn; liaotao1994@sjtu.edu.cn; LuyaoHuang@sjtu.edu.cn; leixue@sjtu.edu.cn; peixuanli@sjtu.edu.cn; wshu@sjtu.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JLT.2018.2888547

challenging due to the high loss budget requirement. Until now, there have been no demonstration on IMDD-based 100 Gb/s/λ PON yet. Therefore we attempt to investigate the possibility of using IMDD technologies rather than coherent technologies for 100 Gb/s/λ PON. Considering the rapid progress of 50 Gb/s PAM4 techniques for Ethernet applications, 20G-class transmitter and receiver have become quite mature and commercially available, therefore we reuse the 20G-class components in IMDD based 100 Gb/s/λ PON applications to reduce the system cost.

Due to the strict bandwidth limitation, inter-symbol-interference (ISI) becomes a severe problem to limit the system performance, therefore powerful equalizers are required as previous 50 Gb/s/λ PON applications. Apart from the traditional equalizers such as FFE, MLSE and VNE, nowadays machine learning has been considered as a powerful equalization tool to mitigate both linear and nonlinear distortions in optical communications [19]–[24]. Machine learning, especially neural network (NN) has become very popular these years and shown its strength in the domain of computer vision and machine translation. Afterwards, NN comes into view of optical communications. Several previous papers have been utilizing NN to improve system performance. Some of them use NN as an assistance to other algorithms [4], while some others directly apply it as an equalizer [19]–[21]. In many experiments, NN shows significantly better performance than the traditional equalizers, which makes the researchers believe it will be a promising candidate as an advanced DSP technique in both long-haul transmission and short reach applications. However, it was pointed out the performance of machine learning based equalization may be overestimated since in the case that psudo-random bit sequence (PRBS) is used to train the model and also used to test the model, NN may recognize the pattern of PRBS data resulting in superior performance [21]. However, if using PRBS data as training sequence, and random data as test sequence, the performance will be much worse. Unfortunately, in most of previous experiments, PRBS data is used as both the training and test sequences, therefore the previous experimental results need to be carefully re-evaluated. In the simulation of [21], using random data as both training and test sequences, NN still shows equalization capability and better performance than the no-equalization case. However, there is no performance comparison with the traditional equalizers such as FFE and VNE, therefore a natural question comes out: will NN-based equalizer under the correct training and test way still outperforms the traditional equalizers?

In this paper, we firstly re-visit the training and test ways of NN-based equalizer and interestingly found that the NN does not just simply recognize the pattern of PRBS data after extensively training, but characterizes the function to generate the PRBS pattern. Therefore the performance of NN is still overestimated and much better than the real case, even if using the first part sequence of a long PRBS pattern as training data and the other part of the rest sequence as test data, where the training and test data have completely different patterns. In this case, NN is considered as a powerful modeling tool rather than an equalization tool. To guarantee that NN only learns the channel characteristics, we have to use random data without pattern characteristics as the training sequence. In this case, NN is regarded as a real equalizer, and for the test data, both random data and PRBS data are applicable. Anyone who uses PRBS data as the training sequence for equalization objective will get the incorrect results. Then we compare the performance of NN-based equalizer under the correct training and test conditions with FFE and VNE, and found there is no obvious difference under linear distortions. With nonlinearity-dominated distortions, however, NN-based equalizer considerably overwhelms FFE and VNE. Therefore we fairly evaluate the performance of NN and prove NN is indeed a powerful nonlinear equalizer. In the experiment, we demonstrate 100 Gb/s/λ IMDD PON based on 20G-class optical and electrical devices using NN as both linear and nonlinear equalizer, which is designed for downstream 100G-PON. Considering the complexity, this system is not applicable for upstream at present. The equalization performance is also compared with FFE and VNE. To improve the loss budget, the launch power is increased to 18 dBm, which induces strong nonlinearity. Attributed to the powerful nonlinear equalization capability of NN-based equalizer, -12 dBm receiver sensitivity can still be achieved at the 7% FEC limit of bit-rate error (BER) of 3.8×10^{-3}, corresponding to a loss budget of 30 dB, meeting IEEE802.3av PR30 requirement for PON applications [14].

The remainder of the paper is organized as follows. In the principle part, we introduce the mathematic model and structure of FFE, VNE and NN in detail. In the simulation part, we introduce different training and test models of NN, explain which method can achieve the correct results, and provide a way to set a standard about how to correctly use NN-based equalizer. By comparing the equalization performance with FFE and VNE, we find the limitation and potential of NN-based equalizer to guide the following experiment. In the experiment part, we successfully demonstrate an 100 Gb/s/λ IMDD PON with 30-dB loss budget enabled by NN-based equalizer and discuss how to further improve the system performance. Finally, we conclude the paper in the conclusion part.

II. Principle

Since FFE and VNE can be considered as a single-layer NN, we will mainly compare the equalization performance of these three equalizers. In this part, we introduce the mathematical model and structure of the three equalizers in detail.

FFE is a common linear equalizer. This algorithm equalizes a symbol by a linear combination of the sampled sequence of the symbol's neighborhood, which can be expressed as:

$$y\left(n\right) = \sum_{i=-k}^{k} w_i x\left(n\ +\ i\right) \tag{1}$$

where $x(n)$ is the sample of the *n-th* received symbol, while $y(n)$ is the corresponding output signal after equalizer. $k = (l\text{-}1)/2$, in which l is the length of input sequence, or in other word, the number of taps of FFE. w_i is the weight for each sample in related position in the sequence. FFE is an essentially linear mapping, which means that it lacks the ability to defeat the nonlinearities originated from modulation and transmission.

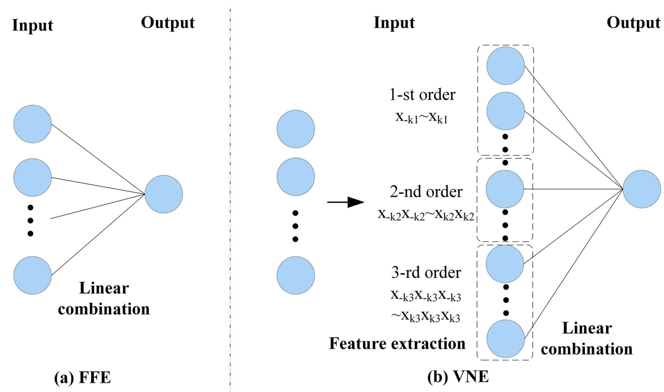

Fig. 1. (a) Structure of FFE. (b) Structure of VNE.

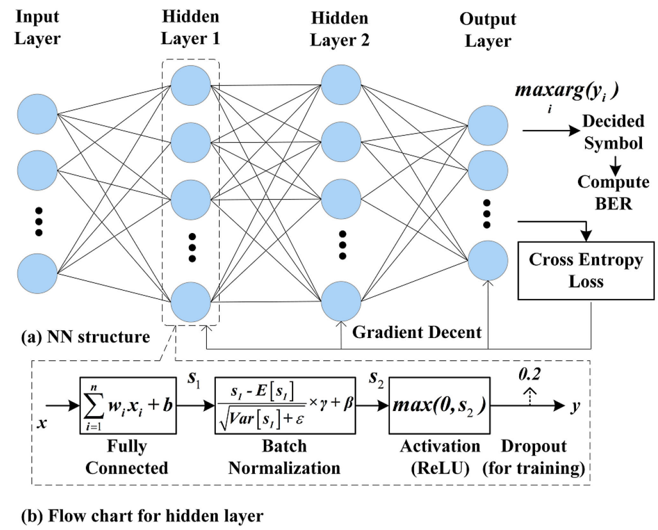

Fig. 2. (a) Structure of NN-based equalizer. (b) Detail structure of hidden layer.

VNE is inspired by Volterra series, a nonlinear model that can approximate the nonlinear system. This equalization algorithm is similar to FFE but adopt more higher-order features of the sampled signal. A 3-order VNE is shown as:

$$\begin{aligned} y\left(n\right) = &\sum_{i_1=-k_1}^{k_1} w_{i_1} x\left(n+i_1\right) \\ &+ \sum_{i_1=-k_2}^{k_2} \sum_{i_2=i_1}^{k_2} w_{i_1,i_2} x\left(n+i_1\right) x\left(n+i_2\right) \\ &+ \sum_{i_1=-k_3}^{k_3} \sum_{i_2=i_1}^{k_3} \sum_{i_3=i_2}^{k_3} w_{i_1,i_2,i_3} x\left(n+i_1\right) x\left(n+i_2\right) x\left(n+i_3\right) \end{aligned} \tag{2}$$

where$x(n)$ is the sampled signal and $y(n)$ is the equalized output. $k_j = (l_j - 1)/2$, in which l_j is the input length of the *j-th* order. $w_{i_1,\ldots,i_j}$ is the weight for each sample of the *j-th* order. As the equation expresses, VNE is still like a linear combination mapping. The difference is that it pre-processes the sampled signal to obtain the higher-order features to provide the nonlinear behavior of the system. The structures of FFE and VNE are shown in Fig. 1.

Recently, NN has been attracting more and more attentions because of its powerful capability in data processing such as classification. It was proved that an NN containing at least one hidden layer can fit and express any functions if it has enough nodes [25]. NN is powerful to characterize nonlinear model, which means that it may bring a better performance in equalization for a transmission system with strong nonlinearities. This motivates us to implement an NN-based equalizer and investigate its performance.

Though an NN with more hidden layers or more nodes of each layer is potentially more powerful to fit a complex function, the computational complexity of NN will increase a lot, while also taking more time overhead. At the same time, the increased complexity will reduce the training efficiency of NN, since the non-convexity of the network enhances. It means that the performance of such NN after training may not increase, or even worse, become poorer. In addition, a complex network cannot perform better than a simple network with proper size for a relatively simple problem, but cost more time. Based on these observations, we apply a simple structure for the NN-based equalizer, with the input size of 51, and two hidden layers of 128 nodes. We have validated that increasing the layers or the nodes of any layer will not evidently improve the performance in our evaluations.

The structure of NN-based equalizer we apply in this work is shown in Fig. 2(a). It is a 3-layer network, containing two hidden layers. The circles denote the nodes, also known as neurons. It is a symbol-spaced equalizer, in which the input layer has 51 nodes, hence the network requires 51 consecutive sampled symbols as input for a judged symbol. Each hidden layer has 128 nodes, while the output layer has 8 nodes, corresponding to 8 kind of symbols of 8-level pulse amplitude modulation (PAM-8) signal. This basic NN is also called fully connected NN, since each 2 nodes adjacent layers are connected. The output of previous layer is the input of the subsequent layer. The output of input layer is the samples themselves. Each node of hidden layer and output layer is a computing unit, as simply expressed follows:

$$y = f\left(\sum_{i=1}^{n} w_i x_i + b\right) \tag{3}$$

where y is the output of this node, x_i is the input of the node but also the output of previous layer, while w_i is the corresponding weight and parameter b is the bias. $f(\cdot)$ is a nonlinear function, called activation function, providing nonlinearity for NN. The activation function of hidden layers is ReLU, while the output activation is Softmax, which are expressed as below:

$$\mathrm{Re}LU\left(x\right) = \max\left(0, x\right), \; Soft\max\left(x_i\right) = \frac{e^{x_i}}{\sum_{j=1}^{8} e^{x_j}} \tag{4}$$

Though an NN can achieve regression as FFE or VNE, in this NN-based equalization, classification is applied to replace the equalization and decision. The output $y_i = Soft\max(x_i)$

of the output layer constructs the output vector

$$y = (y_1, y_2, \ldots, y_8) \tag{5}$$

where the max value y_i of this vector means that the detected symbol is the *i-th* symbol of PAM-8. Therefore, the target of the *i-th* symbol should be:

$$\hat{y}_i = \left(0, \ldots, \underbrace{1}_{i-th} \ldots, 0\right) \tag{6}$$

and the loss function of the network is the cross-entropy loss:

$$\text{CrossEntropy} = -\sum_{i=1}^{8} \hat{y}_i \log(y_i) \tag{7}$$

where $\hat{y}_i$ is the target, while y_i is the corresponding output of NN. As a concept of information theory, cross entropy evaluates the similarity between two probability distributions. By training, the network can minimize the cross entropy step by step, and adjust the probability distribution of output to get close to the target. At the same time, the decisions of this NN-based equalizer will gradually turn to the original symbols.

The network is trained with back-propagation and gradient descent with Adam optimization [26]. The cross entropy loss of output will be back-propagated to update the weight parameters by mini-batch gradient descent with the batch size of 100$\sim$1000, according to the training feedback. Meantime, Adam will adaptively adjust the learning rate depending on the first and second order moment estimation, to increase the training efficiency and accelerate the convergent speed. In case of overfitting, we employ the dropout strategy [27] with the dropout rate of 0.2. This strategy will be only active during training stage, which randomly removes the nodes with a probability of 0.2 over a batch training. It can improve the training efficiency in a way and guarantee the generality of the model. To further improve the training performance, we add the batch normalization [28] in our network. It can be expressed as follows:

$$y = \frac{x - E[x]}{\sqrt{Var[x] + \in}} \times \gamma + \beta \tag{8}$$

where x is the input and y is the normalized output. The parameters γ and β are the learnable vectors of the size as the same as x. The final flow chart for the hidden layer is shown in Fig. 2(b).

In our simulation and experiment, the data sets we use to train and test are all random sequences generated by the random function of MATLAB, except the evaluation about the influence of random sequences and PRBS. Through the simulation, we confirm that our NN-based equalizer is unable to characterize the random sequences we used. The length of the whole data sets are 200,000, 100,000 symbols for training, 10,000 for cross validation and 90,000 for test. The final performance is evaluated based on the BER of the test data set.

As shown in Fig. 1, FFE and VNE can be regarded as a single-layer NN without activation function. VNE obtains the nonlinearity by feature extraction for the input. Since the network is a simple linear combination, it is easy to compute the optimum weight by matrix inversion. The proposed NN-based

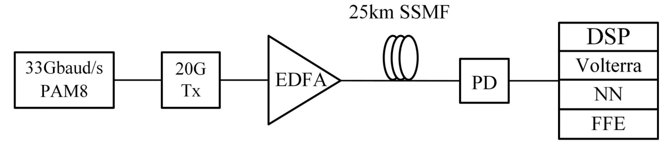

Fig. 3. Simulation setup of 33 Gbaud/s PAM8 system based on 20G-class optical devices.

equalizer is a complete 3-layer network. Its activation function guarantee that it is powerful enough to characterize the transmission system. However, the parameter optimization of NN is nonconvex, which means it is much more likely to achieve a local optimum instead of the global optimum.

To entirely evaluate the performance of the NN-based equalizer, we compare FFE, VNE and NN-based equalizer and measure the BER in our simulation and experiment. With parameters optimizing, the number of FFE taps is set to 91. The lengths of the three order l_1, l_2 and l_3 are 91, 21 and 9 respectively. As for NN, we set its configurations as we explain above. The numbers of nodes of input layer, two hidden layers and the output layer are 51, 128, 128 and 8 respectively.

We provide a simple analysis to compare the computational complexities of the three algorithms. To equalize one symbol, the FFE with 91 taps require 91 multiplications and 90 additions. As multiplication is much more complex than addition, we only consider the number of multiplications. VNE needs 1048 multiplications. Our NN model contains 3 fully connected layers, which require 2688, 16384 and 1024 multiplications respectively. If the processor supports pipeline computing, each layer can be computed in pipeline independently. We can reduce the nodes of hidden layers to significantly reduce the complexity, with a bit performance loss. For instance, let the node size of hidden layer be 32, then only 1632, 1024 and 256 multiplications are needed. With a high parallel processor, almost all of the multiplications can be operated in parallel. The complexity can be acceptable in a way. In this paper, we just show the equalization capability of NN-based equalizer. In the future study, NN with simplified structure will be evaluated to achieve the similar equalization performance.

III. Simulation

Matlab and Python are used to simulate this IMDD system shown in Fig. 3. Transmitted data is 33 Gbaud/s PAM-8 with symbols length of 100,000. The reason that we adopt 33 Gb/s PAM-8 rather than 50 Gbaud/s PAM-4 to achieve 100 Gb/s bit rate is the limited bandwidth of the arbitrary waveform generator (AWG) and the optical and electrical devices used in the experiment, which will be introduced in detail later. A filter is added serving as linear distortion, whose frequency response is the same as that measured in the experiment with 3-dB bandwidth limit of 16.2 GHz. Erbium doped fiber amplifier (EDFA) is followed after the optical transmitter to amplify the signal and simultaneously introduce amplified spontaneous emission (ASE) noise. Split-step Fourier method is extensively used to solve the pulse-propagation problem in nonlinear dispersive media and we use this numerical technique to simulate dispersion

TABLE I
PARAMETERS IN FIBER

Symbol	Quantity	value
L	length	25 km
λ	wave length	1550 nm
D	dispersion parameter	17 ps/(nm*km)
n_2	constant	2.6×10^{-20} m^2/W
A_{eff}	Effective core area	80 μm^2

and nonlinear effect in optical fiber [29]. Fiber length is 25 km, and other parameters are the same with the standard single mode fiber (SSMF) presented in Table I. After fiber transmission and photo-detection (PD), signal is sent into DSP model for equalization and BER calculation.

PRBS is widely used in telecommunication to evaluate the system performance because it exhibits statistical behavior similar to a truly random sequence and can be generated easily. Naturally, when NN is used in optical communication systems as an equalizer, PRBS is also used to evaluate the equalization performance, and no one doubts about the rationality at the beginning. Different from other equalizers, NN has powerful recognizing, modeling capabilities and the potential to characterize the generation rule of PRBS. Based on this observation, the performance of NN-based equalizer using PRBS as training and test data may be overestimated. In Ref. [21], it was shown that if the random data is used for training and test, the performance of the NN-based equalizer is much worse than the PRBS-based training and test case. Unfortunately, before and even after this work was published, PRBS was used as training and test data in most of experiments so the results are not reliable. After this question was put forward, some scientists begin to doubt the effectivity of NN-based equalizer. To validate NN's effectivity and make it clear which data set should be used in training and test, we test several kinds of combination of data set and the results are shown in Fig. 4. We try to set a rule for training and test data set when using NN as an equalizer based on this investigation.

The length of training data and test data are both 100,000. In case (a), repeated PRBS15 sequences are used to train and test, resulting in amazingly low BER. In case (b), partial sequence of PRBS23 is used to train and another part sequence is used to test. BER performance is worse than case (a) but still shows good performance. In case (c) and (d), where PRBS23 is used for training, while PRBS25 and random data are used for test. BER is extremely high. In case (e) and (f), PRBS15 and PRBS17 with 50,000 symbol length are put together for training and random data is used for test. The only difference is that in case (e), the first 50,000 symbols are PRBS15 and the latter 50,000 symbols are PRBS17, while in case (f), PRBS15 and PRBS17 are symbol-interleaved. In case (g) and (h), random data is used for training while PRBS23 and random pattern are used for test. Since the two test data has different generation rules, but take the approximately equal BER curves, we validate that NN can only extract the channel feature of the transmission system rather than the inner characteristic of data. Hence, we consider (g) and (h) as standard. In case (a) and (b), the training and test data have

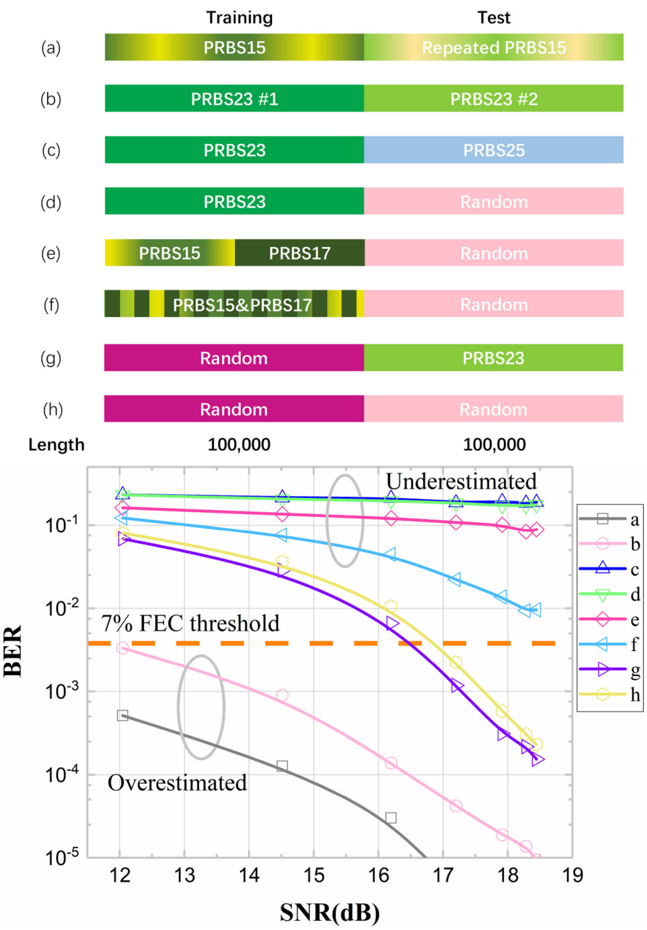

Fig. 4. BER versus SNR under the conditions of different training and test data.

completely different patterns, we attribute the good performance to the modeling capability of NN, which successfully learns the mathematical equation of generating PRBS23 the equalization performance of NN is overestimated compared with (g) and (h). The modeling capability of NN can be proved in case (c) and (d), as NN learns the PRBS23 generation equation and try to use this learned equation to predict the test data, the results are completely wrong. The equalization performance of NN is underestimated. In case (e) and (f), the BERs are slightly lower than the cases of (c) and (d) but the equalization performance of NN is still underestimated since NN can still learn the law of the special training data to an extent. The BER results correctly show the equalization capability of NN.

From the above analysis, we can know anyone who tries to use PRBS as the training data will achieve the wrong results, either overestimating or underestimating the equalization performance of NN. Once random data is used for training, no matter what kind of data is used for test, one can always fairly evaluate the equalization performance of NN. This can be regarded as a new rule of using NN as an equalizer in communication systems. In the following simulation and experiment, we use both random data for training and test to evaluate the equalization performance of NN.

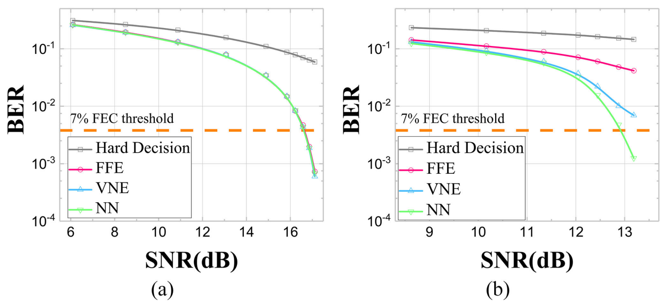

Fig. 5. The relationship between SNR and BER with launch power of (a) 1.5 mW and (b) 22 mW.

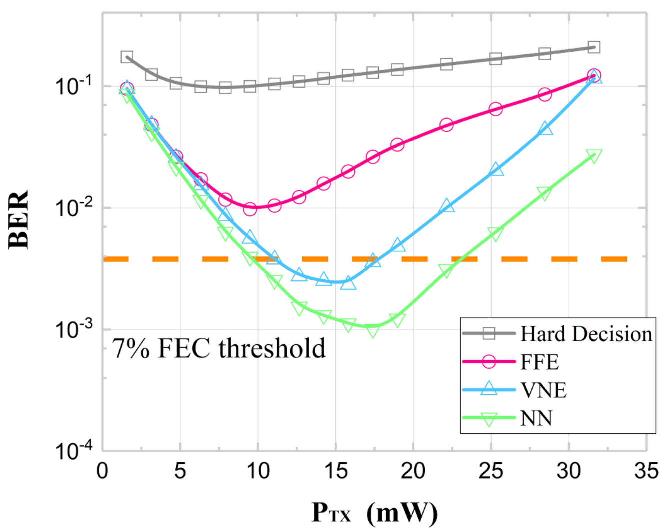

Fig. 6. The relationship between launch power and BER with fixed noise power and limited bandwidth.

To evaluate the efficiency of NN in our system, NN-based equalizer is compared with FFE and VNE under the same conditions. In Fig. 5, BER as a function of the received signal's signal-to-noise ratio (SNR) is drawn with different launch power. With low launch power, three algorithms show almost the same performance, since they all have good performance in equalizing linear distortions; but with high launch power, NN-based equalizer has the lowest BER, proving that NN has its advantage on dealing with nonlinear distortions. In Fig. 5(b), the curve cannot be extended. Because even if signal is sent without AWGN, SNR at receiver cannot be higher due to nonlinearity. In this case, VNE curve can never reach the BER target with the configuration in our simulation.

Relationship between launch power and BER is shown in Fig. 6 to illustrate that NN outperform VNE and FFE under different launch power. With the increase of the launch power, BER firstly decreases attributed to the improved SNR. Further increasing the launch power will degrade BER since the strong fiber nonlinearities exceed the equalization capability of all equalizers. Apparently, NN has the best equalization performance to deal with the nonlinearities. In the practical applications, the optimal launch power needs to be found to achieve the best system performance. Especially in PON applications, high launch power also increases the loss budget, therefore NN-based equalizer is suitable to use in high-speed PONs requiring high loss budget.

Our simulation sets a rule of how to train NN model for equalizing function, also shows the suitable working conditions for NN-based equalizer. This will guide the following experiment.

IV. Experimental Results and Discussions

In this part, we demonstrate a 100 Gb/s/λ IMDD PON using NN-based equalizer to improve the system performance. The experimental setup is shown in Fig. 7(a). A Keysight M8195A AWG with a 64 GSa/s sampling rate, a 20 GHz bandwidth and about 0.6 V output voltage generates a 33 GBd PAM-8 random sequence. The signal from AWG is directly modulated by a 20 GHz Mach–Zehnder modulator (MZM). The laser from a 1550 nm directly-modulated laser (DML) is injected to the MZM biased at its quadrature point, while a 100 Mb/s PRBS signal with a peak-to-peak voltage (Vp-p) of 1 mV from a pulse pattern generator (PPG) is modulated on the laser to broaden the optical spectrum, in order to suppress the stimulated Brillouin scattering (SBS) under the condition of high launch power. We do not directly use the DML as transmitter since its extinction ratio is too small. An EDFA is followed the MZM to control the launch optical power. After 20-km SSMF transmission, a variable optical attenuator (VOA) is used to modify the power of received signal to measure the receiver sensitivity. With another EDFA as a preamplifier in optical network unit (ONU), we can adjust the input optical power of the 20G-class photodetector (PD) to achieve the best receiver sensitivity. An optical filter (OF) is applied to suppress the out of band amplification spontaneous emission (ASE) noise induced by the preamplifier. Since the PD is not followed by trans-impedance amplifier (TIA), an electrical amplifier (EA) is required to amplify the received electrical signal, then sampled by a LeCroy digital sampling oscilloscope (DSO) with a 59 GHz bandwidth and 160 GSa/s sample rate. Note that a 20G-class APD with TIA and high receiver sensitivity may replace the EDFA, OF, PD and EA, but unfortunately, we do not have such APD at present. The sampled signal will be processed in Matlab and Python. The offline DSP first resamples the signal to one sample per symbol. The original random sequence can help extract the sample to keep synchronous. The extracted transmitted sequence is sent into the NN-based equalizer together with the original sequence. After training and test, the equalizer computes and outputs the BER.

Fig. 7(b) presents the eye diagram of electric back-to-back (BTB) 33GBd PAM-8 signals, which is tend to close due to the insufficient bandwidth of AWG. The bad quality of the electrical BTB signal also limits the final system performance. For PAM-8 signal, the power difference of different level significantly affects the BER, so extinction ratio (ER) is an important parameter to achieve good BER. PAM-8 signal with low ER will be easily affected by noise resulting in bad receiver sensitivity. Compared with externally-modulated signal by MZM, directly-modulated signal using DML with much lower ER is not suitable for PAM-8 signal generation if high receiver sensitivity is required. That is why we use MZM rather than DML

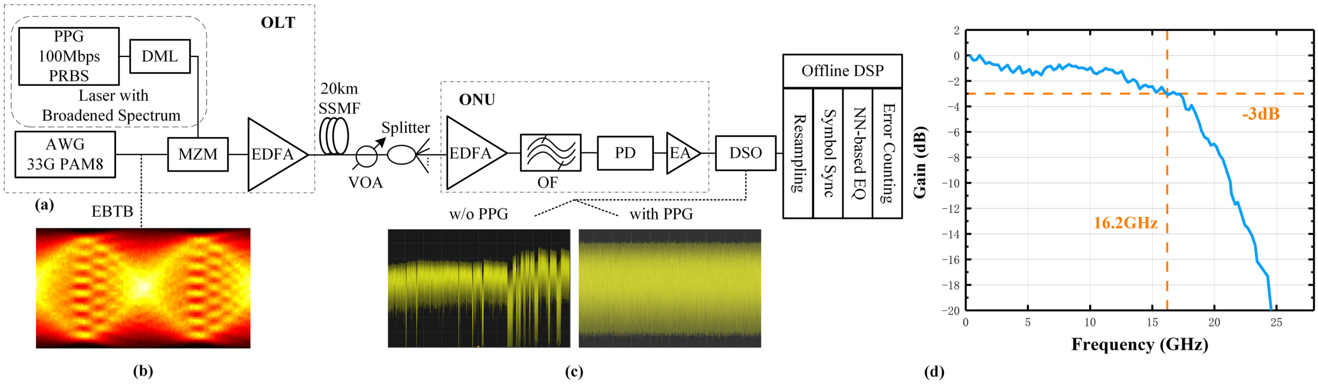

Fig. 7. (a) Experimental setup. (b) EBTB eye diagram. (c) The waveforms w/o and with broadened spectrum laser. (d) Frequency response of the system.

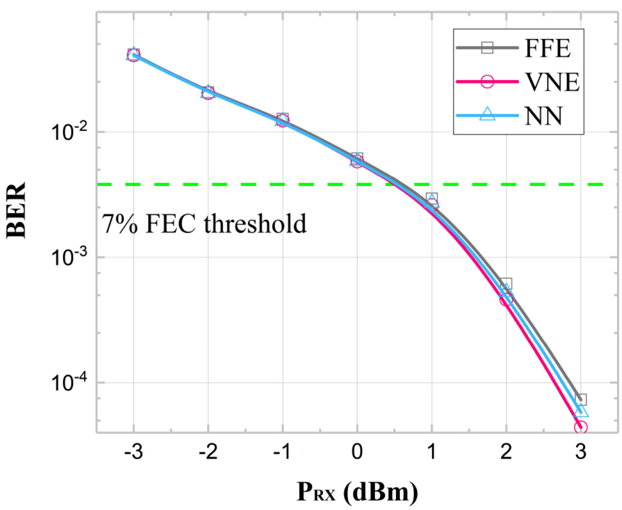

Fig. 8. BER performance comparison with different injected power to PD.

as the transmitter in this experiment even though external modulator has high cost. If PAM-4 signal is used, the case may be different. However, 50 GBd PAM-4 cannot be generated limited by the AWG. Compared with directly-modulated signal, externally-modulated signal generated by MZM also has a drawback. The strong carrier power at the high launch power case will cause severe SBS effect. To prevent this unwanted nonlinearity, we modulate the optical carrier using a 100 Mb/s PRBS data with low voltage to achieve frequency modulation. The optical spectrum is broaden therefore the SBS threshold is significantly improved. The effect of suppressing SBS is shown in Fig. 7(c). Without data modulation in the laser, strong oscillation is observed on the waveform but after data modulation the oscillation is disappeared. The DML is biased at saturation region, so residual intensity modulation can be neglected, which will not affect the PAM-8 signal detection. The frequency response of the whole system is shown in Fig. 7(d). The system 3-dB bandwidth is about 16.2 GHz.

We first investigate the sensitivity of PD in BTB case. The signal adjusted by VOA is directly injected to PD without using EDFA. Fig. 8 presents the BER for different injected power to PD. The BER rises rapidly as the injected power decreases. Since the amplitudes of adjacent symbols of PAM-8 are very close, the device noise including thermal noise and shot noise that cannot be eliminated by any algorithm tends to create huge error for the symbol equalization. To illustrate this problem, we provide a simple analysis. In this experiment, the Vp-p of the noise from electric amplifier without optical input is about 30 mV, while its output Vp-p with 3-dBm injected power to PD is just 500 mV. Consider only the influence of the noise, ignoring any other influences, for simplicity, let the amplitude difference of two adjacent symbols be 60 mV, then the Vp-p of signals is more than 420 mV. Assume the noise is additive white Gaussian noise (AWGN). It is generally known that a Gaussian distribution $X \sim N(\mu, \sigma^2)$ follows the probability:

$$P\left(|x-\mu|<3\sigma\right)=0.9974 \tag{9}$$

where μ and σ^2 are the mean and variance of the distribution respectively. Let $\mu = 0$ and $\sigma = 5$, then

$$P\left(|x|<15\right)=0.9974 \tag{10}$$

It can be considered that most of the values are located in the interval $(-15, 15)$. As the Vp-p of the noise is about 30 mV, we can simply assume the AWGN follows the distribution $X_n \sim N(0, 25)$. When the distance of adjacent symbols is 60 mV, it requires a noise with the absolute value of 30 mV. According to the probability integral table of Gaussian distribution, the probability of making an error for a symbol is

$$P\left(|x_n|>30\right)<0.0001 \tag{11}$$

An adjacent symbol error may just take 1 bit error based on Gray code, while a PAM-8 symbol has 3 bits. Hence the BER is negligible.

If the power injected to PD reduces 3 dB, the amplitude difference of symbols will be 30 mV. The probability to make an adjacent symbol error is

$$P\left(|x_n|>15\right)=0.0026 \tag{12}$$

which means the BER can reach 0.0009. If we further reduce the power by 3 dB, this value will become 0.0445, which is unacceptable. Therefore, to effectively evaluate the performance of NN-based equalizer, we fix the injected power to PD as 3 dBm in the following experiment.

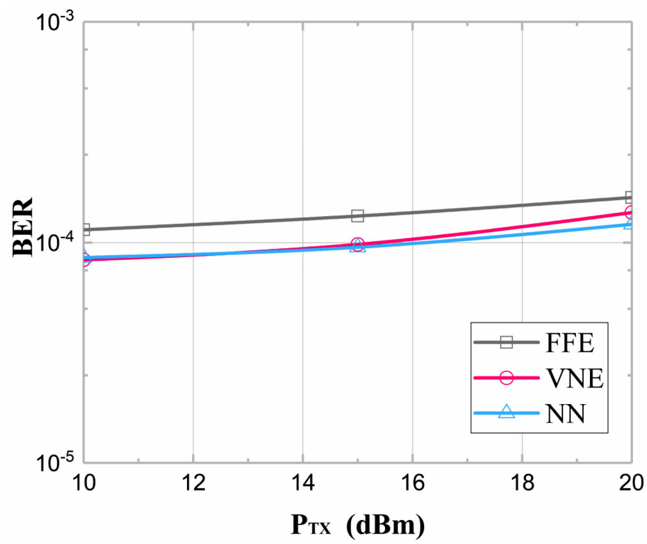

Fig. 9. BER performance comparison with different launch power for OBTB.

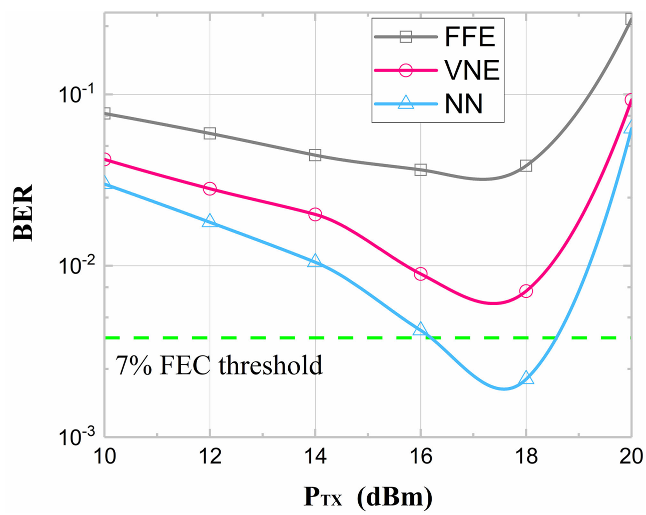

Fig. 10. BER performance comparison with different launch power for 20-km SSMF transmission.

Fig. 9 shows the optical back-to-back (OBTB) BER performance versus the launch power. The VOA keeps the received optical power before preamplifier −5 dBm, while the EDFA at receiver controls the power injected to PD to be 3 dBm. As a result, we learn that the performances of the three algorithms are approximately equal, while varying little as the launch power changes since only linear distortions exist in the system for OBTB case. Frankly speaking, the BER for OBTB is a little big. This is because the laser we use to suppress SBS is with residual intensity modulation. Though we set the amplitude of the modulation signal as small as possible, it still leads an extra error. In our experiment, we have achieved a well balance between this interference and SBS suppression.

Fig. 10 shows the BER performance versus the launch power for 20-km SSMF transmissions with −5 dBm received optical power after the VOA. Again, the EDFA at receiver adjusts the power injected to PD to 3 dBm. All the BER curves of the three algorithms first decrease and then increase, while NN get the best performance and VNE overwhelms FFE, which are consistent with the simulation results. When the launch power is small, the nonlinearity of the transmission system is weak so that the performances of three equalizers are similar. As NN is more powerful to equalize nonlinearity, it will get a significantly better performance as the nonlinearity of the system improve to a limited extent. However, as the launch power continues to grow, the nonlinearities increase beyond the capability of the algorithms, so that the BERs of all equalizers rise.

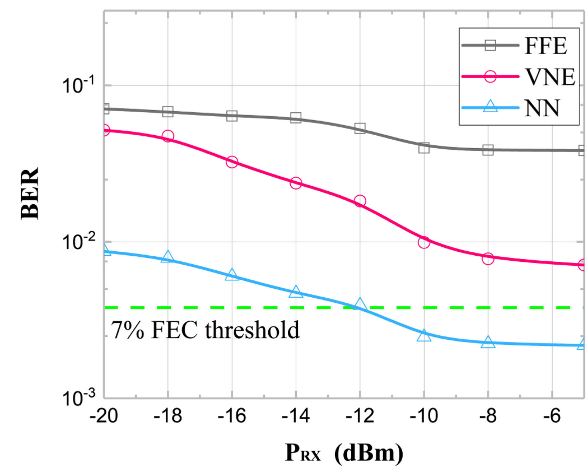

Fig. 11. BER performance comparison of different received power for 20 km-SSMF transmission.

As we have proved that NN-based equalizer can perform better than FFE and VNE in the transmission system with proper nonlinearities, it becomes possible to increase the system loss budget by increasing the launch power into an appropriate level. To measure the maximal loss budget, we set 18 dBm as the launch power to test the BER performance with different received power. Besides, we still keep the PD injected power 3 dBm. As shown in Fig. 11, NN-based equalizer can achieve a sensitivity of about −12 dBm at 7% FEC limit. With the 18-dBm launch power, the total link loss budget can reach 30 dB.

We have achieved 30-dB loss budget for 100 Gb/s/λ PON meeting IEEE802.3av PR30 requirement. Even though the value is much lower than the coherent 100 Gb/s/λ PON with 38.9 dB [13], we prove that, assisted by NN-based equalizer, IMDD technology is also feasible to achieve 100 Gb/s/λ PON with acceptable loss budget. In this experiment, EDFA is used in the ONU, which seems unacceptable for practical applications. If using a 20G-class APD+TIA with high receiver sensitivity, EDFA may be replaced by SOA. The feasibility of the system will be substantially improved. Besides, an AWG with higher bandwidth and enhanced signal quality will also improve the system performance. In this experiment, we use fully-connected NN with a typical NN structure to evaluate the system performance for a proof-of-concept. Convolutional NN (CNN) or recurrent NN (RNN) can also be used and compared for a better equalization performance. We will try to further improve the loss budget of the IMDD 100 Gb/s/λ PON, aiming at lower cost and lower complexity for practical applications in the future work.

V. Conclusion

We propose to use NN-based equalizer to mitigate both the linear and nonlinear distortions in IMDD 100 Gb/s/λ PON based on 20G-class optical and electrical devices. We revisit the training and test rules when using NN for equalization function. Using PRBS as training sequence will achieve incorrect results, either overestimating or underestimating the performance of NN-based equalizer. Once random data is used as training sequence, both random data and PRBS can be used as test sequence to fairly evaluate the equalization function of NN. Compared with the traditional equalizers such as FFE and VNE, NN shows the same equalization performance under linear distortions, but much stronger equalization performance under nonlinear distortions, verifying NN as a powerful nonlinear equalizer. By increasing the launch power to 18 dBm using a booster EDFA in OLT, assisted by the NN-based equalizer, 30-dB loss budget can be achieved, verifying the feasibility of 100 Gb/s/λ IMDD PON using 20G-class optical and electrical devices. We also analyze how to further improve the loss budget and reduce the system cost.

References

[1] P. Miguelez, "What applications are driving higher capacity in access?" in *Proc. Opt. Fiber Commun. Conf. Expo.*, 2018, Paper M2B.1.

[2] L. Yi, P. Li, T. Liao, and W. Hu, "100 Gb/s/λ IM-DD PON using 20G-clasoptical devices by machine learning based equalization," in *Proc. Eur. Conf. Opt. Commun.*, 2018, Paper Mo4B.5.

[3] G. Chen *et al.*, "Machine learning adaptive receiver for PAM-4 modulated optical interconnection based on silicon microring modulator," *J. Lightw. Technol.*, vol. 36, no. 18, pp. 4106–4113, Sep. 2018.

[4] N. Stojanovic, F. Karinou, Z. Qiang, and C. Prodaniuc, "Volterra and Wiener equalizers for short-reach 100G PAM-4 applications," *J. Lightw. Technol.*, vol. 35, no. 21, pp. 4583–4594, Nov. 2017.

[5] P. J. Winzer, D. T. Neilson, and A. R. Chraplyvy, "Fiber-optic transmission and networking: The previous 20 and the next 20 years," *Opt. Express*, vol. 26, no. 18, pp. 24190–24239, 2018.

[6] A. H. Gnauck, R. M. Derosier, A. R. Chraplyvy, and R. W. Tkach, "160 Gbit/s (8 ∗ 20 Gbit/s WBM) 300 km transmission with 50 km amplifier spacing and span-by-span dispersion reversal," *Electron. Lett.*, vol. 30, no. 15, pp. 1241–1243, Jul. 1994.

[7] N. S. Bergano, J. Aspell, C. R. Davidson, P. R. Trischitta, B. M. Nyman, and F. W. Kerfoot, "Bit error rate measurements of a 14 000 km 5 Gb/s fiber-amplifier transmission system using a circulating loop," *Electron. Lett.*, vol. 27, no. 21, pp. 1889–1890, Oct. 1991.

[8] M. Tao, L. Zhou, H. Zeng, S. Li, and X. Liu, "50-Gb/s/λ TDM-PON based on 10G DML and 10G APD supporting PR10 link loss budget after 20-km downstream transmission in the O-band," in *Proc. Opt. Fiber Commun. Conf.*, 2017, Paper Tu3G. 2.

[9] X. Li *et al.*, "4 × 28 Gb/s PAM4 long-reach PON using low complexity nonlinear compensation," in *Proc. Opt. Fiber Commun. Conf.*, 2017, Paper M3H. 4.

[10] C. Ye, D. Zhang, X. Huang, H. Feng, and K. Zhang, "Demonstration of 50 Gbps IM/DD PAM4 PON over 10 GHz class optics using neural network based nonlinear equalization," in *Proc. Eur. Conf. Opt. Commun.*, 2017, Paper W.2.B.

[11] X. Guo, D. G. Cunningham, R. V. Penty, and I. H. White, "Optical equalizers for bandwidth-limited transmitters and pam-based fiber optic links," *J. Lightw. Technol.*, vol. 36, no. 12, pp. 2484–2491, Jun. 2018.

[12] D. V. Veen and V. Houtsma, "Proposals for cost-effectively upgrading passive optical networks to a 25G line rate," *J. Lightw. Technol.*, vol. 35, no. 6, pp. 1180–1187, Mar. 2017.

[13] L. Xue, L. Yi, P. Li, and W. Hu, "50-Gb/s TDM-PON based on 10G-class devices by optics-simplified DSP," in *Proc. Opt. Fiber Commun. Conf.*, 2018, Paper M2B. 4.

[14] J. Zhang *et al.*, "Symmetrical 50-Gb/s/λ PAM-4 TDM-PON in O-band with DSP and semiconductor optical amplifier supporting PR-30 link loss budget," in *Proc. Opt. Fiber Commun. Conf.*, 2018, Paper M1B. 4.

[15] K. Zhang *et al.*, "Demonstration of 50 Gb/s/λ symmetric PAM4 TDM-PON with 10G-class optics and DSP-free ONUs in the O-band," in *Proc. Opt. Fiber Commun. Conf.*, 2018, Paper M1B. 5.

[16] N. Suzuki, H. Miura, K. Matsuda, R. Matsumoto, and K. Motoshima, "100 Gb/s to 1 Tb/s based coherent passive optical network technology," *J. Lightw. Technol.*, vol. 36, no. 8, pp. 1485–1491, Apr. 2018.

[17] C.-Y. Chuang *et al.*, "Convolutional neural network based nonlinear classifier for 112-Gbps high speed optical link," in *Proc. Opt. Fiber Commun. Conf.*, 2018, Paper W2A. 43.

[18] F. Karinou, N. Stojanovic, and Z. Yu, "Toward cost-efficient 100G metro networks using IM/DD, 10-GHz components, and MLSE receiver," *J. Lightw. Technol.*, vol. 33, no. 19, pp. 4109–4117, Oct. 2015.

[19] E. Giacoumidis *et al.*, "Fiber nonlinearity-induced penalty reduction in CO-OFDM by ANN-based nonlinear equalization," *Opt. Lett.*, vol. 40, no. 21, pp. 5113–5116, 2015.

[20] Y. Cui, M. Zhang, D. Wang, S. Liu, Z. Li, and G.-K. Chang, "Bit-based support vector machine nonlinear detector for millimeter-wave radio-over-fiber mobile fronthaul systems," *Opt. Express*, vol. 25, no. 21, pp. 26186–26197, 2017.

[21] T. A. Eriksson, H. Bülow, and A. Leven, "Applying neural networks in optical communication systems: Possible pitfalls," *IEEE Photon. Technol. Lett.*, vol. 29, no. 23, pp. 2091–2094, Dec. 2017.

[22] C. Häger and H. D. Pfister, "Nonlinear interference mitigation via deep neural networks," in *Proc. Opt. Fiber Commun. Conf.*, 2018, Paper. W3A.4.

[23] P. Li, L. Yi, L. Xue, and W. Hu, "56 Gbps IM/DD PON based on 10G-class optical devices with 29 dB loss budget enabled by machine learning," in *Proc. Opt. Fiber Commun. Conf.*, 2018, Paper M2B.2.

[24] Z. Yang *et al.*, "Radial basis function neural network enabled c-band 4 × 50 Gb/s PAM-4 transmission over 80 km SSMF," *Opt. Lett.*, vol. 43, no. 15, pp. 3542–3545, 2018.

[25] M. A. Nielsen, *Neural Networks and Deep Learning*, Determination Press, 2015. [Online]. Available: http://neuralnetworksanddeeplearning.com/

[26] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, arXiv:1412.6980.

[27] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015, arXiv:1502.03167.

[28] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," 2012, arXiv:1207.0580.

[29] G. P. Agrawal, *Nonlinear Fiber Optics*, 4th ed. Berlin, Germany: Springer, 2015, ch. 4, sec. 4.1, pp. 79–88.

Lilin Yi received the B.S. and M.S. degrees from Shanghai Jiao Tong University (SJTU), Shanghai, China, in 2002 and 2005, respectively. He received the joint-educated Ph.D. degree from Ecole Nationale Supérieure des Télécommunications (ENST, currently named as Telecom ParisTech), Paris, France and SJTU, in March and June 2008, respectively. Since 2010, he has been with the State Key Laboratory of Advanced Optical Communication Systems and Network, SJTU where he is currently a Full Professor. He has authored or coauthored more than 150 papers in peer-reviewed journals and conferences, including more than 30 invited papers/invited talks, which have been cited more than 1600 times (Google Scholar). His main research interests include chaotic optical communications, high-speed optical access networks, optical signal processing, and machine learning based digital signal processing.

Dr. Yi is the recipient of the awards "Young scholars of the Yangtze River in China" and "National Science Fund for Excellent Young Scholars of China." He serves as the TPC member of OFC/OECC/CLEO-PR/ACP and TPC track/workshop/symposium co-chairs of OFC/OECC/CLEO-PR/ACP.

Tao Liao received the B.S. degree from the Department of Computer Science and Engineering at Shanghai Jiao Tong University (SJTU), Shanghai, China, in 2017. He is a postgraduate student with the Department of Electronic Engineering with SJTU and is currently working with the Laboratory of Optical Signal Processing and Access Networks under the supervision of Prof. Lilin Yi. His research interests include data communication and engineering, and the application of machine learning in optical access network.

Luyao Huang received the B.S. degree in optoelectronic information science and engineering from the Huazhong University of Science and Technology in 2018 and is currently a Ph.D. student with the Department of Electronic Engineering at Shanghai Jiao Tong University, Shanghai, China. She is currently working with the Laboratory of Optical Signal Processing and Access Network SJTU, Shanghai, China, under the supervision of Prof. Lilin Yi. Her research interests include optical and electrical signal processing in access work.

Lei Xue received the B.S. degree in communication engineering from the School of Information Science and Engineering, Hunan University, Hunan, China, in 2015. He is currently working toward the Ph.D. degree with the Future Ready Optical Network Technology Research Group, State Key Laboratory of Advanced Optical Communication Systems and Networks, Shanghai Jiao Tong University, Shanghai, China. His research interests include optical access network and short reach optical interconnection.

Peixuan Li received the B.S. degree in electronic engineering from the School of Electronic and Information Engineering, Beihang University, Beijing, China, in 2016. He is currently working toward the Ph.D. degree with the Future Ready Optical Network Technology Research Group, State Key Laboratory of Advanced Optical Communication Systems and Networks, Shanghai Jiao Tong University, Shanghai, China. His research interest includes machine learning based equalization technique in IM/DD transmission system.

Weisheng Hu received the B.Sc. degree from Tsinghua University, Beijing, China, in 1986, the M.Eng. degree from the University of Science and Technology Beijing, Beijing, China, in 1989, and the Ph.D. degree from Nanjing University, Nanjing, China, in 1997. In 1997, he joined Shanghai Jiao Tong University as a Postdoctorate Fellow, and was a Professor from 1999, where he was promoted to a Distinguished Professor in 2009. During 2002 and 2012, he was the Deputy Director and Director with the State Key Laboratory of Advanced Optical Communication Systems and Networks. He has authored or coauthored about 300 peer-reviewed journal/conference papers.

Dr. Hu serves in five editorial boards including *Optics Express*, JOURNAL OF LIGHTWAVE TECHNOLOGY, *Chinese Optics Letters*, and *China Communications*. He was also in program committees of a number of international conferences, including OFC, ICC, Globecom, INFOCOM, OPTICS-East, ACP, etc.