



Nonlinearity mitigation in a fiber-wireless integrated system based on low-complexity autoencoder and BiLSTM-ANN equalizer

XIANG LIU,^{1,2} JIAO ZHANG,^{1,2,*}  MIN ZHU,^{1,2,4}  WEIDONG TONG,^{1,2} ZHIGANG XIN,^{1,2} YUNWU WANG,^{1,2} MINGZHENG LEI,²  BINGCHANG HUA,² YUANCHENG CAI,^{1,2}  YUCONG ZOU,² AND JIANJUN YU^{2,3}

¹National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China

²Purple Mountain Laboratories, Nanjing, Jiangsu 211111, China

³Fudan University, Shanghai, 200433, China

⁴minzhu@seu.edu.cn

*jiaozhang@seu.edu.cn

Abstract: We propose and experimentally demonstrate an intelligent nonlinear compensation method using a stacked autoencoder (SAE) model in conjunction with principal component analysis (PCA) technology and a bidirectional long-short-term memory coupled with ANN (BiLSTM-ANN) nonlinear equalizer for an end-to-end (E2E) fiber-wireless integrated system. The SAE-optimized nonlinear constellation is utilized to mitigate nonlinearity during the optical and electrical conversion process. Our proposed BiLSTM-ANN equalizer is primarily based on time memory and information extraction characteristics, which can compensate for the remaining nonlinear redundancy. A low-complexity 50 Gbps E2E-optimized nonlinear 32 QAM signal is successfully transmitted over a span of 20 km standard single-mode fiber (SSMF) and 6 m wireless link at 92.5 GHz. The extended experimental results indicate that the proposed E2E system can achieve a reduction of up to 78% in BER and a gain in receiver sensitivity of over 0.7 dB at BER of 3.8×10^{-3} . Moreover, computational complexity is reduced by more than 10 times compared to the classical training model.

© 2023 Optica Publishing Group under the terms of the [Optica Open Access Publishing Agreement](#)

1. Introduction

With the rapid expansion of wireless devices and bandwidth-intensive services, such as online meetings, online courses, and telemedicine, the fiber-wireless integrated system presents a promising solution to address the exponential increase in mobile broadband data [1,2]. By combining the advantages of fiber and wireless communication, it can provide both wide bandwidth and high mobility, which are essential for beyond 5 G (B5G) or even 6 G network. Millimeter wave (mmWave) (30~300 GHz) is known for large available bandwidth, making it a promising solution for high-capacity transmission [3–5]. Moreover, W-band (75~110 GHz) fiber-wireless convergence technique exhibits great potential for future long-distance and high-capacity transmission networks, owing to its wide bandwidth, directional transmission, and low transmission loss. Several reported achievements suggest that mmWave transmission systems have significant potential for all-weather communication applications [6–10].

The fiber-mmWave system faces various linear and nonlinear damage challenges when using high-order modulation formats. Currently, coherent detection for high-order quadrature amplitude modulation (QAM) signals can provide high receiver sensitivity and obtain linear impairment compensations in the digital domain [11–14]. Due to the complexity of hybrid channels and optoelectronic devices, high-speed transmission systems are facing challenges with the nonlinear distortion. Imperfect optoelectronic devices, such as multistage amplifiers, mixers, photodetectors

(PD), and optical-to-electrical (O/E) and electrical-to-optical (E/O) conversion components, as well as complex environmental factors, such as attenuation, occlusion, and perturbation [15,16], are the main causes of nonlinear distortions, as shown in Fig. 1. In particular, the nonlinear response of the fiber-mmWave integrated system causes the symbols with higher power to experience more nonlinear impairments than those with lower power. As a result, high-order symbols located in the outer region are more vulnerable to nonlinear impairments than those located in the inner region. Nonlinear impairment is inevitable due to nonlinear distortion affecting the shape of high-order M -QAM constellations, resulting in a severe deterioration of the achievable information rates (AIRs) and transmission distances. Therefore, constellation shaping can be introduced to mitigate the nonlinearity of the system. Furthermore, as the transmission rate increases, the nonlinearity becomes more pronounced, necessitating the use of more taps in the equalizer design. The computation complexity also increases with the number of taps. Therefore, a well-modeled equalizer with lower complexity can effectively compensate for the nonlinearities, significantly improving system performances, such as long-distance and high-capacity transmission.

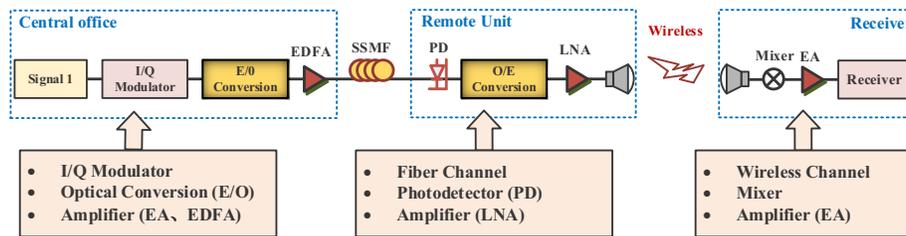


Fig. 1. The primary nonlinear distortions in a fiber-wireless integrated system.

In order to effectively mitigate nonlinearities, several nonlinear equalization approaches have been extensively studied, such as the Volterra equalizer [17], MLSE [18], and Kernel [19]. However, it is still uncertain whether individually optimized blocks can lead to the global optimization of the fiber-wireless integrated system due to the nonlinear optoelectronic devices and hybrid nonlinear channels. To guarantee optimal performance, [20] recommends optimizing the entire system as an end-to-end (E2E) solution rather than individual modules to coordinate the optimization between the transmitter and receiver. The E2E structure can be seen as autoencoder (AE) neural networks, and the gradient back-propagation (BP) algorithm helps jointly optimize hyperparameter and networks, which are extended to mitigate nonlinear impairment [21–23], residual phase noise [24,25], or optimize geometric constellation shaping (GCS) [26–31]. Combining constellation shaping and nonlinear equalizer has been shown to enhance system performance and resist nonlinearity. The AE training structure, along with an intelligent nonlinear equalizer, can further improve system performance. An intelligent end-to-end nonlinear constellation auto-optimization method has been applied to the fiber-wireless integrated network [32]. The nonlinear constellation with bit-level encoding is optimized by simulating nonlinear effects, which can compensate for the nonlinear impairments in the transmission system. However, the selected nonlinear model is suboptimal for simulating the practical nonlinearity caused by hybrid channels and optoelectronic devices. Additionally, the non-ideal matching between E2E-optimized nonlinear constellations and natural experimental environment results in the accumulation of nonlinearity. Furthermore, due to the constellation shaping and the time-varying system, traditional equalization algorithms cannot effectively compensate for the remaining nonlinearities. Moreover, the computational complexity is also a vital factor in influencing overall performance. Therefore, modeling an efficient framework and a robust

equalizer to mitigate the nonlinear impairments of the fiber-wireless convergence system remains a challenge.

This work is an extension and promotion of our OFC paper [30], where we experimentally demonstrated a novel nonlinear constellation auto-optimization method using an autoencoder and a fully-connected artificial neural network (ANN) equalizer at 92.5 GHz. The key novel contributions of this work can be summarized as follows. 1) We establish a low-complexity system to optimize the stacked autoencoder (SAE) framework with the help of principal component analysis (PCA), referred to as PCA-SAE model. This model is designed to learn the nonlinear constellations and reduce the computing complexity. 2) We propose a time-memory equalizer based on a bidirectional long-short-term memory coupled with ANN (BiLSTM-ANN) to resist the time-varying interference and effectively extract nonlinear features. In particular, we evaluate the superiority of the proposed equalizer in terms of training sequence sizes, hidden layers, neuron numbers, as well as training epochs, compared with the 2D-ANN equalizer in [30]. 3) We define the reliability metrics of the training model and present more complete principle and experimental detail. In the fiber-mmWave integrated system, the low-complexity PCA-SAE is considered as a training model, which has not yet been considered by most existing works. 4) We further compare the nonlinearity mitigations and complexity under different training models and nonlinear equalization methods, including tradition algorithms and AE-ANN algorithms [30]. The results demonstrate that the proposed E2E system with the PCA-SAE training model and BiLSTM-ANN equalizer can effectively mitigate nonlinearity and reduce the computing complexity, outperforming other schemes under different nonlinear strengths.

2. Implementing an end-to-end learning for fiber-wireless system

In an actual fiber-wireless integrated system, it is possible to demonstrate the presence of nonlinear distortions by mapping the amplitudes of the transmitted (Tx) and received (Rx) symbols. Figure 2(a) displays the conventional 32QAM data transmitted through the fiber and wireless channels, symbols with higher power experience more nonlinearities. The response of the output normalized amplitude exhibits non-linear behavior when the normalized input amplitude exceeds a specific value. Due to the nonlinearity, the data deviates from the linear straight line. From the perspective of the reference constellation (yellow marks), the received symbols (blue dots) are affected by the nonlinear distortion, resulting in a squeezing in the outer part an expansion in the center area, as shown in Fig. 2(b). The majority of the incorrect symbols (represented by red dots) following whether decision symbol is equal to the original symbol are concentrated in the outer region, suggesting that the conventional algorithms have limited demodulation capabilities for outer symbols with the strong nonlinearity. By leveraging the benefits of the constellation shaping and advanced equalization algorithms, a well-designed architecture can effectively address nonlinear issues and has been successfully implemented in fiber-wireless transmission systems.

2.1. Low-complexity autoencoder architecture for high-order GCS

An autoencoder is a type of deep learning architecture that attempts to recover its input at the output in an unsupervised manner, as illustrated in Fig. 3. This architecture comprises of two sub-networks: an encoding network $c_k = NN_e(x, w_e)$, which maps an input vector $x \in X$ with v -dimensional to a k -dimensional ($k < v$) intermediate vector c consisting of $m = \log_2(M)$ measurements, and a decoding network $\hat{x} = NN_d(y, w_d)$, which attempts to recover x from the received symbols y . The neural networks are trained with the parameterized weights w_e and w_d , respectively. The MATLAB code is used to optimize the nonlinear constellation c_k , and the encoding network optimizes the nonlinear constellations c_k ($k = 2$) for GCS, while the decoding ANN learns the decision boundaries of the distorted signal y . The goal is to find optimal weight set $\{w_e, w_d\}$, that minimizes the categorical cross-entropy or mean square error (MSE), i.e.,

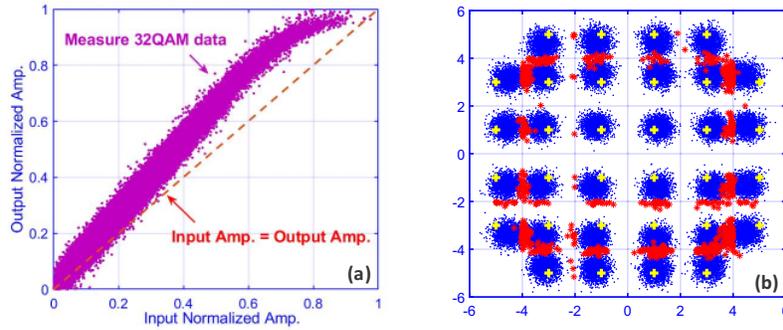


Fig. 2. Nonlinear impairments observed in an actual fiber-wireless integrated system. (a) Amplitude response, and (b) Received symbols equalized by the tradition algorithms.

bit-wise cross-entropy with bit-level optimization. For high-order M -QAM signal, N runs of the AE model are performed, as shown in Fig. 3(b). $C_{M \times N}$ is defined as the constellation set after N iterations, I and Q represent the real and imaginary parts of the complex symbol, respectively. Once the training model has converged, the encoder and decoder are fixed, and the testing is performed. The nonlinear constellations and hyper-parameters with the best accuracy are then stored and sent to the experimental fiber-wireless integrated system, and a novel nonlinear equalizer is implemented by Python to help the system bring extra resistance of nonlinearity, and is more robust.

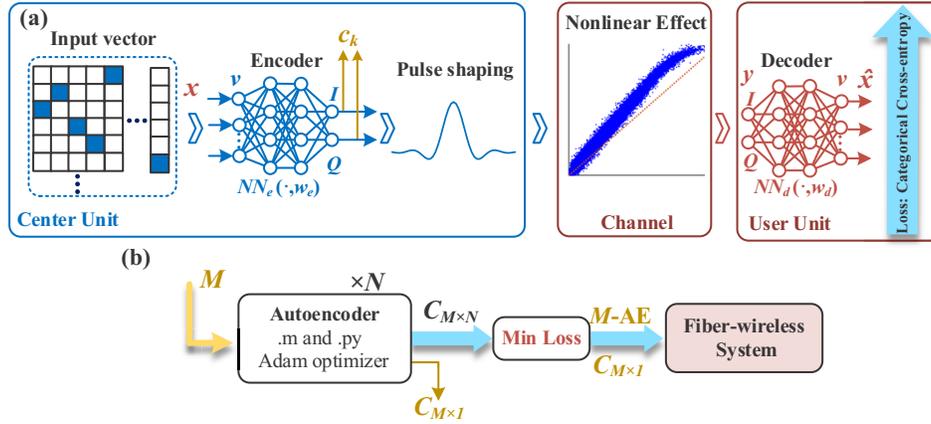


Fig. 3. A classical autoencoder architecture in a fiber-wireless system. (a) Diagram of the applied autoencoder. (b) Feeding the M -AE constellations to the system.

To achieve high input space compression for high-order modulation formats, i.e., $k \ll v$, a more complex AE called a stacked autoencoder (SAE) is designed by adding further hidden layers [33,34]. However, increasing the number of layers poses challenges in training the model \mathcal{N} , such as vanishing gradient problem, computational complexity and overfitting. To address these challenges, a sparse autoencoder framework having L layers is used. The basic cost function, including regularization and sparsity terms, is defined as

$$\mathcal{J}(X; \theta) = \frac{1}{mv} \|X - \mathcal{N}(X; \theta)\|_F^2 + \frac{\lambda}{2} \sum_{l=1}^{L-1} \|W_l\|_F^2 + \gamma \sum_{i=1}^k KL(\rho \| \hat{\rho}_i) \quad (1)$$

where $\|\cdot\|_F$ depicts the Frobenius norm, θ represents weight matrices and corresponding bias matrices, γ and λ represent the weightings that are applied to the sparsity and weight decay terms. The sparsity penalty $KL(\cdot)$ is defined as the Kullback-Leibler (KL) divergence. The mean value of the random variable $\hat{\rho}_i$ is calculated as the average of the output of the i -th neuron in the bottleneck layer, which is then averaged over the training sequences. The target value for this parameter is represented by ρ .

For further reduce the computational complexity for high-order modulation, we introduce a low-complexity framework [34], as shown in Fig. 4. The PCA-SAE is a methodology that addresses the problem of dimensionality reduction by utilizing PCA and a SAE. PCA is used to reduce the dimension of the input space, and then apply SAE to learn a compressed, distributed representation of the data. The input dimension is reduced from v to k , where k represents the number of components required to achieve the desired reconstruction accuracy τ . The encoding matrix P_k , which maximizes the reliability explained, is given by the first k principal component of the PCA of X . The output of the linear neuron is denoted as $C_k = X \cdot P_k$. The nonlinear encoder block further reduces the dimension to k , and the neural network model \mathcal{N}^* is trained to obtain its parameters by minimizing

$$\mathcal{J}_{SAE}(C_{k_{lin}}; \theta) = \|C_{k_{lin}} - \mathcal{N}^*(C_{k_{lin}}; \theta)\|_F^2 \quad (2)$$

It should be noted that when performing GCS, the value of k must be set to 2. The matrix $C_{k_{lin}}$ represents the reduced dimensionality dataset, and the output of the nonlinear neurons is denoted as $\hat{C}_{k_{lin}} = \mathcal{N}^*(C_{k_{lin}}; \theta^*)$. The decoding transformation is represented by $\hat{X} = \hat{C}_{k_{lin}}\beta$, and the linear decoding matrix is $\beta = \hat{C}_{k_{lin}}^+ X$, $C_{k_{lin}}^+$ is its pseudo-inverse. The SAE is trained in two phases: a sparse AE based greedy layer-wise pre-training is performed via (1), and then the fully-connected network is fine-tuned via (2) with early stopping.

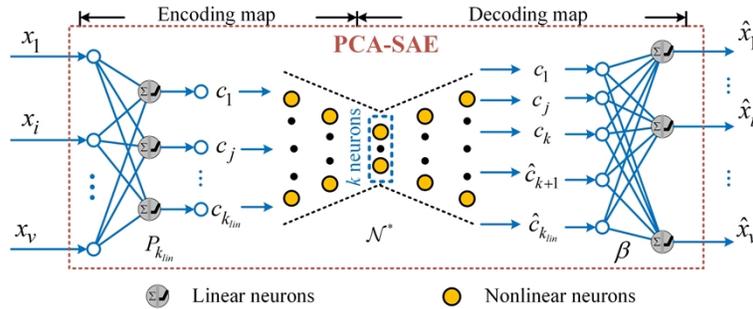


Fig. 4. The Neural network architecture interpretation of the PCA-SAE training model.

2.2. Nonlinear compensation by the BiLSTM-ANN equalizer

Although E2E-optimized nonlinear constellations can mitigate nonlinearity for fiber-wireless transmission system, the nonlinear redundancy caused by suboptimal system parameters and the device disturbance cannot be ignored. In addition, due to the time-varying channel, typical ANN equalizers are not reliable for channel information tracking. In order to further enhance nonlinear resistance, we introduce a novel BiLSTM-ANN equalizer with the complex-valued architecture for information extraction and recovery, as shown in Fig. 5. The equalizer helps the E2E-optimized system bring extra resistant of nonlinearity, and is more robust.

Figure 5 illustrates the proposed architecture for behavioral modeling, which comprises six layers: an input layer, a Bi-LSTM layer, two batch normalization layers, and three fully connected ANN layers. The received symbol $y_{i...N}$ is transformed into two-dimensional (2D) real vectors

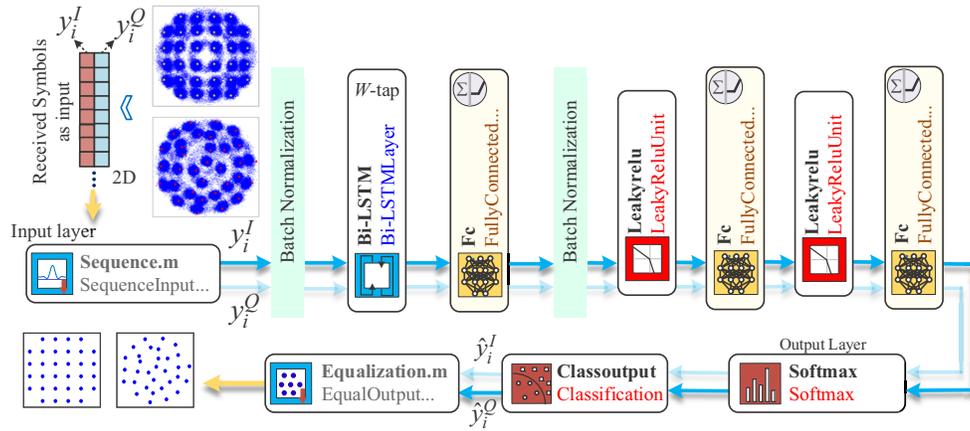


Fig. 5. The flowchart of our proposed BiLSTM-ANN equalization.

$y_{i...N}^I$ and $y_{i...N}^Q$. The normalization serves as the first layer of the Bi-LSTM neuron network, initializing and normalizing the input I/Q elements. Subsequently, the two tributary data sets pass through the Bi-LSTM layer, which includes both forward and backward layers. The LSTM network is designed and trained based on the input sequence's characteristics, and the network operates in a sequence mode, resulting in predicted output of the same size as the input sequence, and W is the weights and bias in a BiLSTM. The regression computation can be represented by

$$MSE = \frac{1}{N} \sum_{i=1}^N [(I(y(i)) - I(\hat{y}(i)))^2 + (Q(y(i)) - Q(\hat{y}(i)))^2] \quad (3)$$

where I and Q denote the real and imaginary components of a complex signal, respectively.

The Bi-LSTM network is used to make a precise decision on the BER. It is driven to fully connected layers and then to a SoftMax function that classifies symbols at the output and provides the probability of each category. The output of the hidden layers of the BiLSTM at each step can be considered as the extracted nonlinear features. We can use these features as input to another machine learning model, or use the output of the BiLSTM directly for prediction. The nonlinear impairments information in the input feature can greatly improve the accuracy of predicting these impairments. We have also investigated the impact of the number of neurons in the first fully connected layer and found that adding a new fully connected layer can be beneficial when further increasing the number of neurons in the first layer does not provide any additional benefits. Additionally, we have varied the hyperparameters to analyze their impact on the system performance. Finally, the weight values of the Bi-LSTM three gates are updated using the BP training algorithm until the target threshold or epochs is reached, which is similar to traditional ANN equalizers.

2.3. Complexity analysis of the autoencoder architecture

We conducted a complexity comparison between the proposed PCA-SAE architecture and the baseline SAE. Both PCA-SAE and SAE can be viewed as networks that have multiple sub-neural models trained independently. PCA-SAE algorithm is used to train network \mathcal{N}^* with dimensions $[k_{lin} \dots, h^i (= k), \dots, k_{lin}]$, whereas SAE train the neural network \mathcal{N} with dimensions $[v \dots, h^i, \dots, v]$. By using $k_{lin} \ll v$ components for ultra-high order signals, the reconstruction accuracy τ can be achieved, and PCA-SAE further reduces the dimension to k . However, as the value of k increases, the difference in reliability explained between algorithms decreases,

and computational complexity increases [34]. To evaluate the validity of selecting k values, especially for the lower values of GCS ($k = 2$), we calculate the reliability explained V_{EX} between the original sequence X and the reconstructed sequence \hat{X} , defined as

$$V_{EX}(X; \hat{X}) = 100(1 - \|X - \hat{X}\|_F^2 / \|X\|_F^2) \quad (4)$$

maximizing V_{EX} is equivalent to minimizing MSE.

We compare the effectiveness of the two methodologies while ensuring that they have an equal number of parameters and layers $\theta_{PCA-SAE} = \theta_{SAE}$, as stated in the following:

$$\theta_{PCA-SAE} = k \cdot k_{lin} + k + k_{lin} \cdot k + k_{lin} = 2k \cdot k_{lin} + k + k_{lin} \quad (5)$$

$$\theta_{SAE} = h \cdot k + h + k \cdot h + k = k \cdot 2h + h + k \quad (6)$$

Equation (5) provides the definition of the parameter count for a single-hidden layer in the PCA-SAE framework, while Eq. (6) presents the corresponding expression for an SAE with h hidden neurons.

3. Experimental setup

Figure 6 illustrates the experimental configuration of our demonstrated W-band 32QAM transmission over a 20 km SSMF and 6 m wireless link, embedded with the SAE-optimized nonlinear constellations and BiLSTM-ANN equalizer. A personal computer (PC) serves as the interface between the transmitter and receiver, collecting training sequences in real-time to enable the integration of the end-to-end data transmission system.

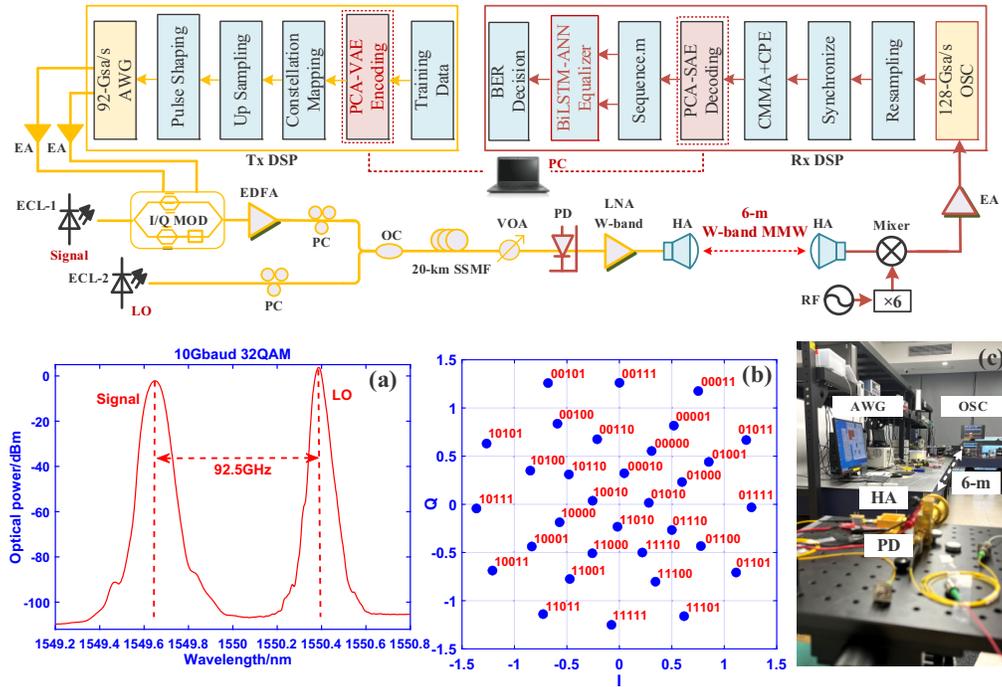


Fig. 6. Experimental setup including, (a) the optical spectrum of 10 Gbaud 32QAM signal, (b) the nonlinear constellation with bit-level mapping at R_{op} is 6 dBm and V_{pp} is 0.1 V, and (c) photo of 6 m wireless link.

The transmitter side employs two free-running tunable external cavity lasers (ECL-1 and ECL-2) with a 100 kHz linewidth to generate W-band mmWave signals in a simple, flexible, and cost-effective architecture. The data rate is 50 Gbit/s with a roll-off factor of 0.5. The training data is encoded and mapped by the PCA-SAE encoding module ($k = 2$), and the baseband 32QAM symbol is converted from digital to analog (DAC) by the arbitrary wave generator (AWG) with a sampling rate of 92 Gsa/s. The boosted 32QAM signal is amplified by a 40 GHz cascaded electrical amplifier (EA) and used to drive the 35 GHz I/Q modulator. The signal light source is ECL-1 at 1549.65 nm with a linewidth of 100 kHz and an average power of 16 dBm, which is modulated via the I/Q modulator. The MZM has a 3 dB optical bandwidth of 30 GHz, a half-wave voltage of 2.7 V at 1 GHz, and a 5 dB insertion loss. ECL-2 at the center wavelength of 1550.4 nm functions as a local oscillator (LO), which has a frequency space of 92.5 GHz with the modulated ECL-1 lightwave, as shown in Fig. 6(a). An erbium-doped fiber amplifier (EDFA) compensates for the fiber transmission loss. Two polarization controllers (PCs) are necessary to adjust the incident direction to maximize output power since the W-band photodiode (PD) is polarization sensitive. The optical signal and ECL-2 are combined by an optical coupler (OC), and the coupled light beam can be delivered over 20 km SSMF. A variable optical attenuator (VOA) adjusts the received optical power (ROP) into the PD to control the nonlinear strengths. The PD used in our experiment is implemented within the frequency range of 10~170 GHz at -2 V DC bias, and the output power range of 0~10 V.

At the wireless transmitter, as shown in Fig. 6(c), the 92.5 GHz signal is generated and emitted from the W-band antenna (HA) with a gain of 26 dBi. A paired W-band HA is used to receive the W-band signals, and the signal power is amplified by a cascaded electrical low-noise amplifier (LNA) to achieve superior input power into the HA. At the wireless receiver, the received signal at 92.5 GHz is first down-converted into an intermediate frequency (IF) signal using a commercial W-band mixer and a LO source. The IF signal is then amplified using an EA with 33 dB gain and 14 dBm saturation output power available from DC to 50 GHz frequency band. Finally, the amplified signal is captured by a 128 Gsa/s digital storage oscilloscope (OSC) with 59 GHz bandwidth and 10-bit resolution for processing. The captured signal is down-converted and resampled to two samples per symbol, followed by squaring time recovery and a 53-tap $T/2$ -spaced cascaded multi-modulus algorithm (CMMA). The frequency offset estimation (FOE) step removes the residual frequency offset of the received signal resulting from frequency drifts of the lasers. A flexible and robust algorithm for phase noise estimation is employed, consisting of a carrier phase estimation (CPE) stage and a blind phase search (BPS) stage. A 37-tap decision-directed least-mean-square (DD-LMS) equalizer is added to compensate for remaining linear damages and I/Q imbalance before the BiLSTM-ANN equalization. After the conventional optimization process, the BiLSTM-ANN equalizer is used to help the E2E-optimized system resist remaining nonlinear redundancy. Finally, the BER performance can be calculated based on the recovered signal. Figure 6(b) illustrates one E2E-optimized nonlinear constellations with bit-level encoding in the actual fiber-wireless system, which is robust to the nonlinear system.

4. Experimental results and discussions

In this section, we present the nonlinear performances of the E2E-based system using auto-optimization GCS and the 2D BiLSTM-ANN equalizer. The training and testing models are implemented in Pytorch and Matlab. Based on the experiment system for E2E-optimized 32QAM fiber-wireless transmission at W-band, we compare the performance of the proposed 2D-ANN equalizer in [30] and the proposed BiLSTM-ANN equalizer in terms of the training size, neuron number, and the training epoch. Next, we qualitatively evaluate the nonlinearity mitigation in a fiber-wireless transmission system with the help of the nonlinear constellations and the BiLSTM-ANN equalizer. Finally, we briefly analyze the computing complexity of the PCA-SAE compared with the traditional SAE architecture.

4.1. Comparison between BiLSTM-ANN and 2D-ANN equalization

Increasing the size of the training set in a network generally improves accuracy, but it also presents a series of challenges, particularly a larger computational burden during the training process. A shorter training length results in a lower precision, making the selection of the training data scale a key influential factor in ANN networks. Figure 7(a) displays the BER performance of testing data as the training data size changes when the input power into PD is fixed at 5 dBm. The results indicate that the precision of the equalizers is dependent on the initial training sets, and BER decreases effectively with an increasing training block length. When approximately 10,000 samples are used as the training data, BER can be achieved as low as 3.8×10^{-3} . Furthermore, the BiLSTM-ANN equalizer greatly reduces the training size, demonstrating its good training accuracy and satisfactory tracking speed. Considering the accuracy, complexity, and time consuming, we train networks using 12,000 training samples in the further discussion.

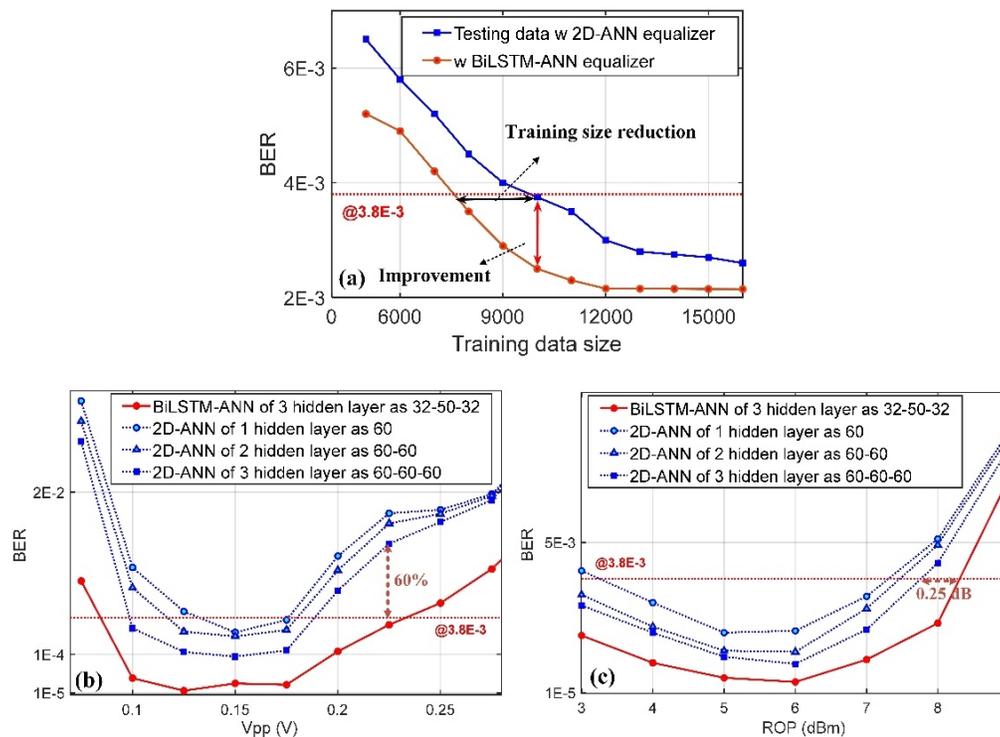


Fig. 7. (a) BER performance of the testing data versus the training data size in the 2D-ANN equalizer [30] and the BiLSTM-ANN equalizer, respectively. (b) BER versus the V_{pp} . (c) BER versus the Rop.

Furthermore, increasing the number of hidden layers in ANN also effectively improves accuracy. Figure 7(b) and Fig. 7(c) illustrate the BER of 50 Gbps 32QAM signals versus ROP and V_{pp} values, with the neuron numbers of input and output fixed at 2. The discussion is divided into four cases: (1) ANN has a hidden layer with 60 neurons in a 2-60-2 configuration. (2) ANN has two hidden layers, each with 60 neurons, in a 2-60-60-2 configuration. (3) ANN has three hidden layers, each with 60 neurons, in a 2-60-60-60-2 configuration. (4) BiLSTM-ANN structure has only one hidden layer, in a 32-50-32 configuration. During the analysis of the four cases, the training size was set to 12,000. The results indicate that the 2D-ANN with three hidden layers outperforms the other two cases in terms of BER performance. However, when the

BiLSTM-ANN equalization method is applied in case 4, there is a significant improvement in BER, with up to 60% improvement and 0.25 dB nonlinear gains at a BER of 3.8×10^{-3} , compared to the 2D-ANN equalizer.

4.2. Nonlinear constellation resist nonlinear impairments

To evaluate the nonlinear system, the peak-to-peak voltage (V_{pp}) value of the output signals from the AWG is adjusted from 0.075 V to 0.30 V, while the received power of the PD is set to 5 dBm, in order to present different strengths of nonlinear effect. Figure 8 provides a qualitative assessment of the nonlinear system in a fiber-wireless convergence configuration, showing the experimentally-received constellations at three nonlinear strengths (low, middle, and high). It also visually compares the BER performance of the nonlinear constellations to the grid constellation, with blue dots and red marks representing the correct and incorrect symbols after equalization. Figure 8(a)-(c) demonstrates that the shape of the grid constellation is affected by the nonlinearity as the nonlinear strength increases, resulting in an expansion in the center part and a squeezing in the outer area. On the other hand, Fig. 8(d)-(f) shows that the nonlinear constellations are clearly distinguishable, with significantly reduced incorrect symbols, and are resistant to the nonlinearity in the fiber-wireless integrated system.

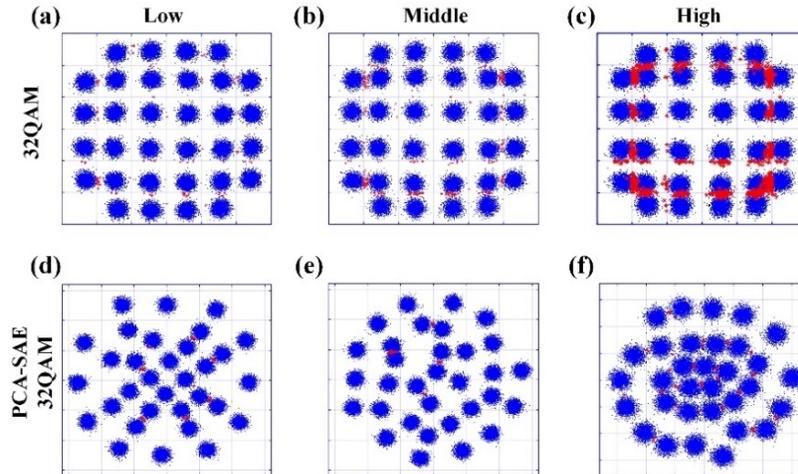


Fig. 8. Experimentally received constellations for 32QAM and AE-optimized 32QAM: (a), (d) low nonlinear strength; (b), (e) middle nonlinear strength; (c), (f) high nonlinear strength.

In Fig. 9(a), the average BER curves for different schemes are presented versus training epoch. The E2E-optimized nonlinear constellations with BiLSTM-ANN equalizer exhibit the best BER performance and optimal convergence speed. Although the AE-32QAM scheme with a 2D-ANN equalizer can achieve BER performance under 3.8×10^{-3} , the BiLSTM-ANN equalizer can further improve system performance. Additionally, the nonlinear performance is evaluated at different nonlinear strengths. Figure 9(b) shows the transmission performances at various V_{pp} values. As V_{pp} increases, the nonlinearity of the seamless fiber-wireless system gradually increases.

In the first region, where the V_{pp} value ranges from 0.075 V to 0.125 V, nonlinearity is weak, and the larger distance between the nonlinear constellations enable better performance under the same conditions. In the second V_{pp} region, ranging from 0.125 V to 0.175 V, the nonlinear strength is middle, and the nonlinearity effect starts to impact the shape of the 32QAM, resulting in an expansion in the center part and a squeezing in the outer area. With increasing V_{pp} , a more severe nonlinearity is introduced by the amplifiers. In the third region, ranging from

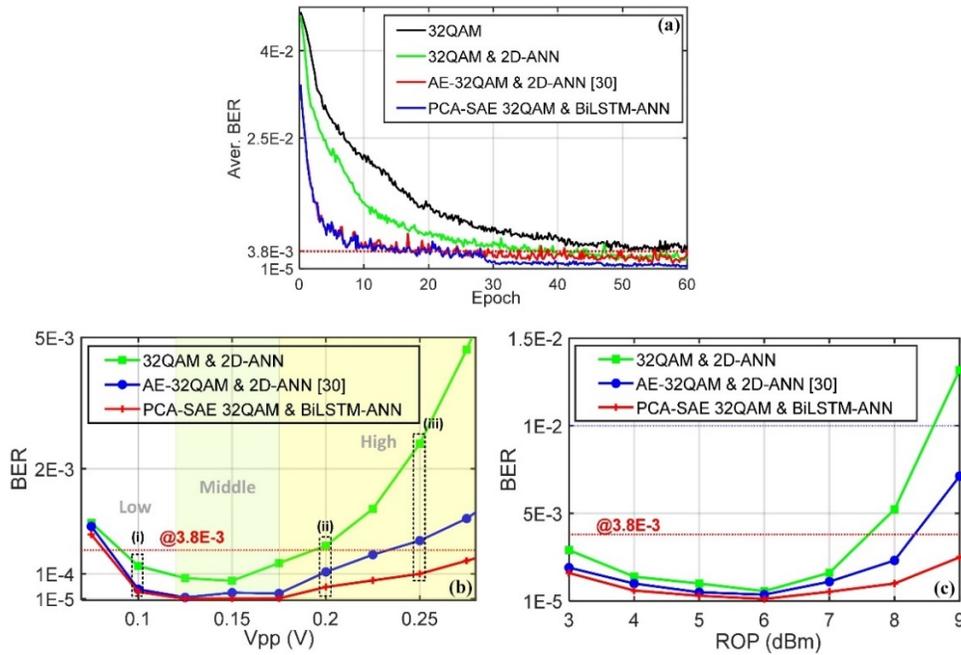


Fig. 9. (a) The iteration BER with different schemes. (b) The BER performance of the constellations with different V_{pp} when ROP is 5 dBm. (c) The BER of the constellations with different ROP when V_{pp} is 0.1 V.

0.175 V to 0.30 V, the BER performance is seriously affected by the nonlinearity. However, the point in the received constellation can be clearly distinguished. The nonlinear 32QAM with the BiLSTM-ANN equalizer achieves a 78% lower BER when V_{pp} is 0.25 V. The curves of the 2D-ANN equalizer and the 2D BiLSTM-ANN equalizer are further compared. Although the nonlinear constellations obtain similar improvement under low nonlinear strength, the resistance to nonlinearity can be further enhanced under high nonlinear strength, making it robust to the fiber-wireless system. Figure 9(c) shows the BER performance when V_{pp} is 0.1 V, varying the ROP. By using the AE-32QAM with 2D-ANN equalizer, a receiver sensitivity gain exceeding 0.7 dB is achieved over that of the method with no E2E optimization. Furthermore, it shows that the E2E system has stronger nonlinear resistance and can achieve better performance by taking advantage of the bit-level mapping and the BiLSTM-ANN equalizer.

To illustrate the mitigation of nonlinearity in the received signals, Fig. 10(i)-(iii) display the input and output normalized amplitude diagrams of the nonlinear constellations after implementing the 2D BiLSTM-ANN equalizer. The black points indicate the mean of the output versus input amplitude, while the red line represents the linear reference. The equalizer effectively mitigates nonlinear impairments, even under strong nonlinear strengths.

4.3. Complexity analysis

To evaluate the effectiveness of the proposed autoencoder layout, we compared the training models of three autoencoder architectures. As the value of k increases, the training models become more efficient, but this comes at the cost of greater complexity and longer training times. Figure 11(a) analyzes the reliability index V_{EX} between the original sequences X and the reconstruction sequences \hat{X} . The results indicate that SAE has a similar accuracy to PCA-SAE, but the AE architecture in [30] performs worse. In this work, it also shows that the auto-optimized

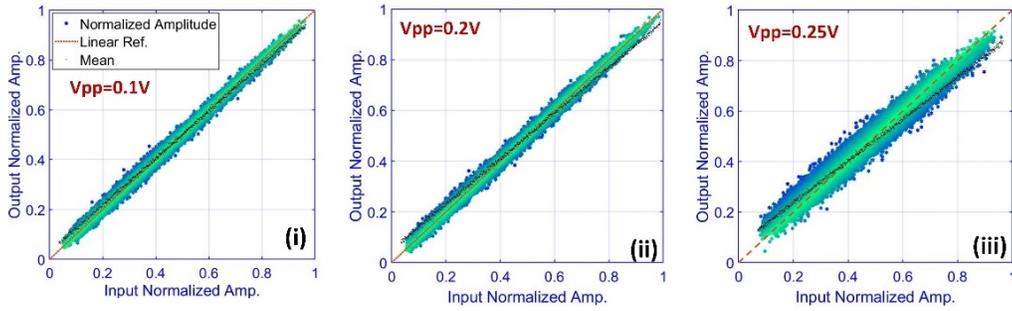


Fig. 10. The optimized nonlinearities at different nonlinear strength.

nonlinear constellations based on the PCA-SAE architecture is reliable for GCS ($k = 2$), and the reliability index V_{EX} is close to the saturation value and reaches 93. The illustration depicts the E2E-optimized nonlinear constellations.

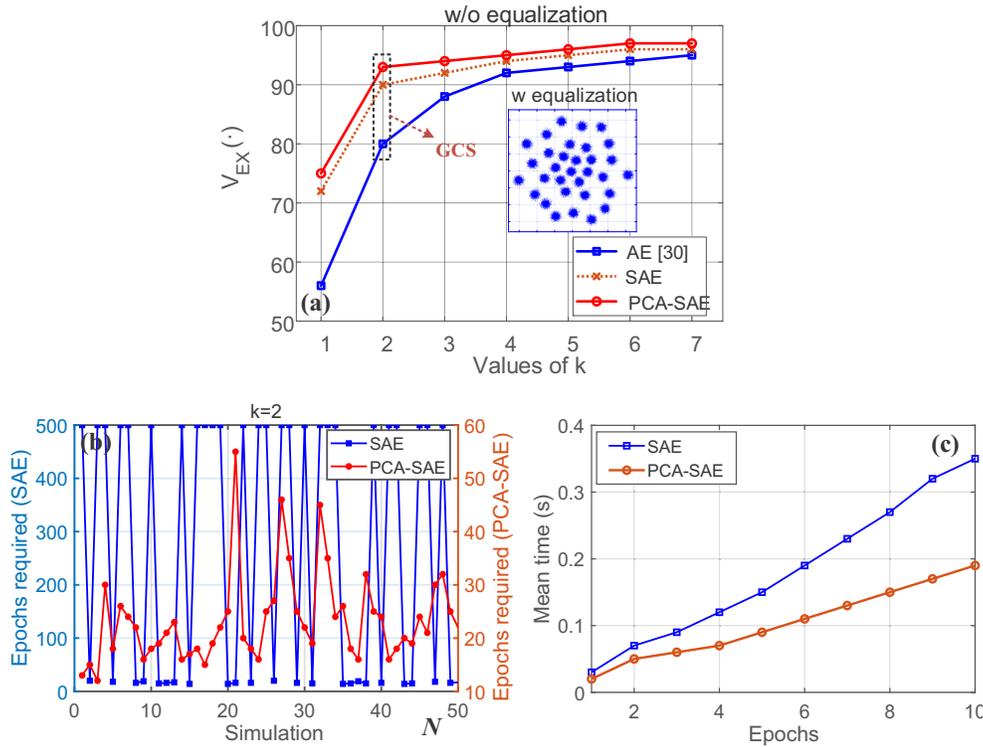


Fig. 11. (a) The V_{EX} as a function of k . (b) Number of required epochs to train in PCA-SAE and SAE. (c) The mean computation times for the first epochs.

To investigate the reason for the shorter training time of the PCA-SAE compared to SAE, despite both neural networks satisfying $\theta_{PCA-SAE} = \theta_{ASE}$ ($\tau = 99.5$, $k_{lin} = 7$ and $h = 7$), we analyzed the required number of epochs to train networks \mathcal{N} and \mathcal{N}^* in each simulation, as well as the mean computational time over the initial epoch. The analysis results are presented in Fig. 11(b) and (c), respectively. Figure 11(b) illustrates that PCA-SAE requires an average of 12 to 55 epochs to train, seldom exceeding 100, while SAE typically necessitates up to 500

epochs. The E2E transmission system can reduce the computing complexity by more than 10 times compared to the classical SAE training model, and the PCA-SAE requires less training time during each epoch. These results suggest that PCA-SAE framework is inherently better conditioned, leading to easier-to-solve training problems than the standard SAE architecture.

5. Conclusion

In this paper, we demonstrate an E2E-optimized system with a low-complexity PCA-SAE training model and a novel equalizer for a 50 Gbps 32QAM signal transmission over a 20 km SSMF and a 6 m wireless link at 92.5 GHz. The E2E-optimized system can learn nonlinear constellations with bit-level mapping and has been verified to have a strong resistance to the nonlinearity. Furthermore, we propose a 2D BiLSTM-ANN nonlinear equalizer to mitigate nonlinearities caused by the time-varying channel and device disturbance. In particular, we evaluate the superiority of the BiLSTM-ANN nonlinear equalizer compared to the 2D-ANN equalizer in terms of training sizes, training epoch, and neuron number. The fiber-wireless system benefits from the nonlinear constellations and the BiLSTM-ANN equalizer, resulting in strong nonlinear resistance and better BER performance. Compared to the classical system, the E2E system achieves a BER performance up to 70% decrement and a 0.7 dB receiver sensitivity improvement under the BER of 3.8×10^{-3} . Additionally, the computational complexity of the training system is significantly reduced. The E2E system proposed in this paper holds great potential for practical fiber-wireless integrated link in future 5 G or even 6 G mobile communication system.

Funding. National Natural Science Foundation of China (62101121, 62101126, 62201393, 62201397, 62271135); China Postdoctoral Science Foundation (2021M702501, 2022T150486); State Key Laboratory of Advanced Optical Communication Systems and Networks of Shanghai Jiao Tong University (2022GZKF003); the Major Key Project of Peng Cheng Laboratory (PCL 2021A01-2); Key Research and Development Program of Jiangsu Province (BE2020012); the Graduate Research and Practice Innovation Program of Jiangsu Province (SJCX21_0036).

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

References

1. X. H. You, Y. M. Huang, and S. H. Liu, *et al.*, "Toward 6G TK u Extreme Connectivity: Architecture, Key Technologies and Experiments," *arXiv*, arXiv:2208.01190 (2023).
2. M. Zhu, J. Zhang, B. C. Hua, M. Z. Lei, Y. C. Cai, L. Tian, D. M. Wang, W. Xu, C. Zhang, Y. M. Huang, J. J. Yu, and X. H. You, "Ultra-wideband fiber THz-fiber seamless integration communication system toward 6G: architecture, key techniques, and testbed implementation," *Sci. China Inf. Sci.* **66**(1), 113301 (2023).
3. X. Y. Li, J. J. Yu, and G. K. Chang, "Photonics-aided millimeter-wave technologies for extreme mobile broadband communications in 5 G," *J. Lightwave Technology*. **38**(2), 366–378 (2020).
4. S. Rommel, D. Dodane, and E. Grivas, "Towards a scaleable 5 G fronthaul: Analog radio-over-fiber and space division multiplexing," *J. Lightwave Technology*. **38**(19), 5412–5422 (2020).
5. X. Y. Li, J. J. Yu, Z. Dong, J. W. Zhang, N. Chi, and J. G. Yu, "Investigation of interference in multiple-input multiple-output wireless transmission at W band for an optical wireless integration system," *Opt. Lett.* **38**(5), 742–744 (2013).
6. X. Y. Li, J. N. Xiao, and J. J. Yu, "Long-distance wireless mm-wave signal delivery at W-band," *J. Lightwave Technol.* **34**(2), 661–668 (2016).
7. W. P. Li, J. J. Yu, Y. Y. Wang, F. Wang, B. W. Zhu, W. Zhou, F. Zhao, and J. J. Yu, "Photonics-based high-speed long-distance fiber-wireless-integration communication at the W-band," *Inf. Sci.* **66**, 127301 (2023).
8. N. P. Diamantopoulos, S. Inudo, Y. Yoshida, A. Maruta, A. Kanno, P. T. Dat, T. Kawanishi, R. Maruyama, N. Kuwaki, S. Matsuo, and K. Kitayama, "Mode-division multiplexed W-band RoF transmission for higher-order spatial multiplexing," in *Optical Fiber Communications Conference and Exhibition (OFC)*, (2015), pp. 1–3.
9. W. P. Li, J. J. Yu, Y. Y. Wang, F. Wang, B. W. Zhu, L. Zhao, W. Zhou, J. G. Yu, and F. Zhao, "OFDM-PS-256QAM signal delivery at 47.45Gb/s over 4.6kilometers wireless distance at the W band," *Opt. Lett.* **47**(16), 4072–4075 (2022).
10. Y. Q. Xiao, B. S. Wang, Z. Y. Wang, Z. H. Zhou, and J. He, "One-time pad scheme based on polar code and OFDM for MMW-RoF system at W-band," *Opt. Express*. **30**(3), 4412–4423 (2022).

11. Z. L. He, J. X. Song, K. Vijayan, C. Häger, G. A. Alexandre, W. Henk, P. Andrekson, M. Karlsson, and J. Schröder, "Periodicity-enabled size reduction of symbol based predistortion for high-order QAM," *J. Lightwave Technology*, **40**(18), 6168–6178 (2022).
12. M. P. Yankov, D. Zibar, K. J. Larsen, and S. Forchhammer, "Constellation shaping for fiber-optic channels with QAM and high spectral efficiency," *IEEE Photonics Technol. Lett.* **26**(23), 2407–2410 (2014).
13. X. L. Pan, D. Wu, X. Y. Li, D. Guo, Z. P. Li, H. X. Yan, C. C. Wang, J. H. Bi, C. Yu, X. Y. Liu, X. J. Xin, Q. Zhang, R. Gao, and Z. Dong, "Photonic-aided W-band dual-vector RF signal generation and detection enabled by bandpass delta-sigma modulation and heterodyne detection," *Opt. Lett.* **48**(8), 2146–2149 (2023).
14. S. Wen, G. H. Liu, C. X. Liu, H. Y. Qu, M. Tian, and Y. Chen, "Waveform design for high-order QAM faster-than-Nyquist transmission in the presence of phase noise," *IEEE Trans. Wireless Commun.* **21**(1), 2–17 (2022).
15. C. C. Wei, Z. X. Xie, P. H. Ting, Z. W. Huang, and C. T. Lin, "3×3 MIMO 60-GHz OFDM-RoF System with Long-Distance Wireless Transmission Enabled by MIMO Volterra Filtering," *J. Lightwave Technology*, **40**(20), 6860–6866 (2022).
16. S. T. Zhou, R. Gao, Q. Zhang, H. Chang, X. G. Xin, Y. Zhao, J. Liu, and Z. R. Lin, "Data-defined naïve Bayes (DNB) based decision scheme for the nonlinear mitigation for OAM mode division multiplexed optical fiber communication," *Opt. Express*, **29**(4), 5901–5914 (2021).
17. L. Zhang, X. Z. Hong, X. D. Pang, O. Ozolins, A. Udalcovs, R. Schatz, C. J. Guo, J. W. Zhang, F. Nordwall, K. M. Engenhardt, U. Westergren, S. Popov, G. Jacobsen, S. L. Xiao, W. S. Hu, and J. Chen, "Nonlinearity-aware 200 Gbit/s DMT transmission for C-band short-reach optical interconnects with a single packaged electro-absorption modulated laser," *Opt. Lett.* **43**(2), 182–185 (2018).
18. H. Yamazaki, M. Nagatani, H. Wakita, Y. Ogiso, M. Nakamura, M. Lda, T. Hashimoto, H. Nosaka, and Y. Miyamoto, "Transmission of 400-Gbps discrete multi-tone signal using > 100-GHz-bandwidth analog multiplexer and InP Mach-Zehnder modulator," in *2018 European Conference on Optical Communication (ECOC)*, (2018), pp. 1–3.
19. L. Zhang, X. D. Pang, A. Udalcovs, O. Ozolins, R. Lin, X. Yin, M. Tang, W. J. Tong, S. L. Xiao, and J. J. Chen, "Kernel mapping for mitigating nonlinear impairments in optical short-reach communications," *Opt. Express*, **27**(21), 29567–29580 (2019).
20. T. O'shea and J. Hoydis, "An introduction to deep learning for the physical layer," *IEEE Trans. Cogn. Commun. Netw.* **3**(4), 563–575 (2017).
21. J. X. Song, C. T. Häger, J. C. Schröder, A. Amat, and H. Wymeersch, "End-to-end autoencoder for superchannel transceivers with hardware impairment," in *2021 Optical Fiber Communication Conference and exhibition (OFC)*, (2021), pp. 1–3.
22. T. Koike-Akino, Y. Wang, D. S. Millar, K. Kojima, and K. Parsons, "Neural turbo equalization: Deep learning for fiber-optic nonlinearity compensation," *J. Lightwave Technol.* **38**(11), 3059–3066 (2020).
23. V. Neskorniuk, A. Carnio, V. Bajaj, D. Marsella, S. K. Turitsyn, J. E. Prilepsky, and V. Aref, "End-to-end deep learning of long-haul coherent optical fiber communications via regular perturbation model," in *2021 European Conference on Optical Communication (ECOC)*, (2021), pp. 1–4.
24. S. Li, C. Häger, N. Garcia, and H. Wymeersch, "Achievable information rates for nonlinear fiber communication via end-to-end autoencoder learning," in *2018 European Conference on Optical Communication (ECOC)*, (2018), pp. 1–3.
25. V. Talreja, T. Koike-Akino, Y. Wang, D. S. Millar, K. Kojima, and K. Parsons, "End-to-end deep learning for phase noise-robust multi-dimensional geometric shaping," in *2020 European Conference on Optical Communications (ECOC)*, (2020), pp. 1–3.
26. M. Schaedler, S. Calabrò, F. Pittalà, G. Böcherer, M. Kuschnerov, C. Bluemm, and S. Pachnicke, "Neural network assisted geometric shaping for 800Gbit/s and 1Tbit/s optical transmission," in *2020 Optical Fiber Communications Conference and Exhibition (OFC)*, (2020), pp. 1–3.
27. O. Jovanovic, M. P. Yankov, D. R. Francesco, and D. Zibar, "End-to-end learning of a constellation shape robust to channel condition uncertainties," *J. Lightwave Technology*, **40**(10), 3316–3324 (2022).
28. O. Jovanovic, M. P. Yankov, F. Da Ros, and D. Zibar, "End-to-end learning of a constellation shape robust to variations in SNR and laser linewidth," in *2021 European Conference on Optical Communication (ECOC)*, (2021), pp. 1–4.
29. R. T. Jones, T. A. Eriksson, M. P. Yankov, and D. Zibar, "Deep learning of geometric constellation shaping including fiber nonlinearities," in *2018 European Conference on Optical Communication (ECOC)*, (2018), pp. 1–3.
30. X. Liu, J. Zhang, M. Zhu, W. D. Tong, Z. G. Xin, Y. W. Wang, M. Z. Lei, B. C. Hua, Y. C. Cai, Y. C. Zou, and J. J. Yu, "Autoencoder Learning of Nonlinear Constellation Shape for Fiber-Wireless Convergence System," in *2023 Optical Fiber Communications Conference and Exhibition (OFC)*, (2023), pp. 1–3.
31. V. Aref and M. Chagnon, "End-to-end learning of joint geometric and probabilistic constellation shaping," in *2022 Optical Fiber Communications Conference and Exhibition (OFC)*, (2022), pp. 1–3.
32. J. L. Jia, J. Chen, B. Y. Dong, G. Q. Li, J. Y. Shi, C. Shen, Z. W. Li, L. Tao, J. W. Zhang, and N. Chi, "Intelligent End-to-End Nonlinear Constellation Auto-Optimization in W-band Fiber-MMW Integrated Transmission for 6 G Access," in *2022 Optical Fiber Communications Conference and Exhibition (OFC)*, (2022), pp. 1–3.
33. J. Xu, L. Xiang, Q. S. Liu, H. Gilmore, J. Z. Wu, J. H. Tang, and A. Madabhushi, "Stacked sparse autoencoder (SSAE) for nuclei detection on breast cancer histopathology images," *IEEE Trans. Med Imaging*, **35**(1), 119–130 (2015).
34. F. Zocco and S. McLoone, "Recovery of linear components: Reduced complexity autoencoder designs," *Eng. Appl. Artif. Intel.* **109**, 104663 (2022).