End-to-End Learning of Constellation Shaping for Optical Fiber Communication Systems

Wenshan Jiang, Xue Zhao, Fangfang Huang, Xiatao Huang, Taowei Jin, Hong Lin, Jing Zhang, and Kun Qiu

***Abstract*—End-to-end learning based on autoencoder can realize robust constellation shaping for optical fiber communications. The existing schemes use the symbol-wise autoencoder (SAE) or bit-wise autoencoder (BAE) to realize the constellation shaping. The SAE mainly focus on the performance of mutual information (MI), this neglects the decoding loss so that the generalized mutual information (GMI) or the post forward error correction (FEC) bit error rate (BER) has almost no performance gain in bit-wise metric systems. In this paper, we propose a probabilistic shaping (PS) based on BAE with a modified loss function, where the mean square error and source entropy are used to construct the loss function. We compare the GMI and post-FEC performance of the PS and also geometric shaping (GS) based on SAE or BAE by numerical simulations and experiments. In simulations, we transmit 64-QAM signal with GS or PS over 100-km SSFM. The simulation results show that the GS or PS based on BAE can achieve 0.13-bits/sym or beyond 0.2-bits/sym GMI gain. In experiment, the GS based on BAE obtains 0.11-bits/sym GMI gain and 0.7-dB launch optical power gain after belief propagation decoding. The PS with source entropy of 5.5-bits/sym and 5.2-bits/sym outperforms uniform 64-QAM by 0.25-bits/sym and 0.3-bits/sym, respectively.**

***Index Terms*—Optical fiber communications, end-to-end learning, constellation shaping, generalized mutual information.**

I. Introduction

RECENTLY, the explosive growth of global data traffic has put forward higher requirements for the channel capacity and transmission performance of optical fiber communications [1]. With the development of artificial intelligence, machine learning has shown advantages in nonlinear compensation, performance monitoring and modulation format recognition for optical fiber communications [2], [3], [4], [5]. However, modularized optimization in communication systems cannot achieve the overall performance optimization. The end-to-end learning based on autoencoder has been demonstrated to jointly optimize transceiver and improve transmission performance [6]. Many previous works have optimized the end-to-end learning by considering the complex impairments in optical fiber communication systems. In terms of networks adopted in autoencoder, feed-forward neural networks (FFNN) and sliding window bidirectional recurrent neural networks (SBRNN) have been investigated for intensity modulation and direct detection (IM/DD) system, which can improve the transmission performance [7]. In terms of channel model, the end-to-end optimization for IM/DD system realizes the transmission link as part of the autoencoder to yield flexible transceivers [8]. To obtain constellations robust to different distortions, the channel model of end-to-end learning considers analog-to-digital converter (ADC) quantization and limited sampling frequency, residual phase noise and optical filtering impairments in coherent optical fiber communications, respectively [9], [10], [11]. Optical fiber channel can also be realized as a differentiable model in end-to-end learning by utilizing simplified memoryless fiber channel model, nonlinear interference noise model, split-step Fourier method (SSFM) based model and first-order regular perturbation model, which can obtain gains in the systems with nonlinearity and dispersion [12], [13], [14], [15]. Where the complex channel characteristic of optical fiber can also be modeled by the data-driven based generative adversarial network (GAN) or bidirectional long short-term memory (BiLSTM) [16], [17]. The constellation with joint probability shaping (PS) and geometric shaping (GS) based on end-to-end learning obtains generalized mutual information (GMI) gains over uniform quadrature amplitude modulation (QAM) signals [18], [19].

The existing algorithms of end-to-end learning are based on symbol-wise autoencoder (SAE) or bit-wise autoencoder (BAE) to realize constellation shaping. Practical systems combine advanced modulation formats and flexible forward error correction (FEC) code to realize error free transmissions, which depends on bit-wise decoding [6], [9]. A Gray-mapping is necessary for bit-interleaved coded modulation (BICM). Therefore, end-to-end learning of optical fiber communications should optimize the binary mappings and study the labeling of learned constellations to improve the bit-wise metric performance. Gray-labeled constellation is used to initial the simple autoencoder for GS and maximize GMI performance [20].

In this paper, we conduct detailed comparisons including achievable information rate and post-FEC bit error rate (BER) of BAE or SAE for GS to verify BAE is more reliable for practical optical fiber communications. We propose a PS scheme based on BAE by using both mean square error (MSE) and source entropy

Manuscript received 9 May 2023; revised 16 July 2023; accepted 30 September 2023. Date of publication 3 October 2023; date of current version 26 October 2023. This work was supported in part by the National Key Research and Development Program of China under Grant 2022YFB2903303, in part by the National Science Foundation of China (NSFC) under Grants U22A2086 and 62111530150, and in part by the Fundamental Research Funds for the Central Universities under Grants ZYGX2020ZB043 and ZYGX2019J008. *(Corresponding author: Jing Zhang.)*

The authors are with the Key Laboratory of Optical Fiber Sensing and Communications (Education Ministry of China), School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu, Sichuan 611731, China (e-mail: wsjiang30@163.com; 202112012009@std.uestc.edu.cn; 202122010610@std.uestc.edu.cn; huangxiatao369@sina.com; jintw9705@163.com; 202121010631@std.uestc.edu.cn; zhangjing1983@uestc.edu.cn; kqiu@uestc.edu.cn).

Digital Object Identifier 10.1109/JPHOT.2023.3321736

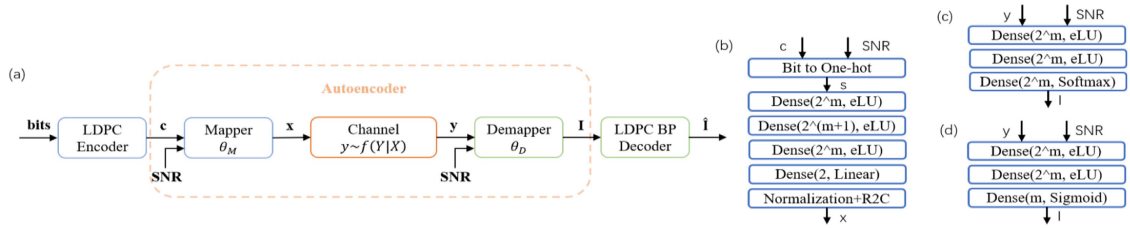

Fig. 1. Diagram of (a) end-to-end model based on autoencoder, (b) mapper DNN, (c) symbol-wise demapper DNN, (d) bit-wise demapper DNN for GS.

to construct the loss function, which can generate the continuous entropies as that with the traditional composition distribution matching (CCDM). We transmit 64-QAM signal over 100-km standard single mode fiber (SSMF) by simulations and experiments. In simulations, the GS based on BAE can learn constellation with Gray mapping, which achieves 0.13-bits/sym GMI gain and 0.1-dB optical signal-to-noise ratio (OSNR) gain after belief propagation (BP) decoding at BER of 3.8×10^{-3}. The proposed PS generating scheme outperforms uniform 64-QAM by 0.2 to 0.4-bits/sym GMI gain. In experiments, the GS based on BAE achieves about 0.11-bits/sym GMI gain and 0.7-dB launch optical power (LOP) gain after BP decoding at BER of 3.8×10^{-3}. The PS based on BAE with source entropy of 5.5-bits/sym and 5.2-bits/sym obtains 0.25-bits/sym and 0.3-bits/sym GMI gains compared to uniform 64-QAM, respectively.

II. Principles

A. End-to-End Learning for Geometric Shaping

End-to-end learning based on autoencoder regards the transmitter, channel and receiver as a single neural network. Fig. 1(a) shows the end-to-end model for GS with low-density parity-check (LDPC) channel coding, where the mapper and demapper are implemented by deep neural networks (DNN) of autoencoder. As shown in Fig. 1(b), the mapper DNN $f^{\theta_M}(s)$ with trainable parameters θ_M convert the bit vector $b = (b_1, \ldots, b_m)$ obtained from LDPC encoder to the complex symbol x with normalized average power. As shown in Fig. 1(c) and (d), the demapper DNNs with trainable parameters θ_D process the received signals y^{θ_M} to estimate the probability of inputs and generate log likelihood ratio (LLR) for LDPC decoder. Note that the SAE adopts softmax function to estimate the probability of transmitted symbol $\hat{s}^{\theta_M,\theta_D} = f^{\theta_D}(y^{\theta_M}) = g^{\theta_D}(s|y^{\theta_M})$, while BAE adopts sigmoid function to estimate the probability of transmitted bits $\hat{b}^{\theta_M,\theta_D} = f^{\theta_D}(y^{\theta_M}) = g^{\theta_D}(b|y^{\theta_M})$. The loss function based on SAE using cross entropy (CE) can be expressed by,

$$\begin{aligned} L_{CE}(\theta_M, \theta_D) &= E_{S,Y^{\theta_E}}\left\{-\log\left(g^{\theta_D}\left(s|y^{\theta_M}\right)\right)\right\} \\ &= E_Y\left\{D_{KL}\left(f\left(s|y^{\theta_M}\right)\left\|g^{\theta_D}\left(s|y^{\theta_M}\right)\right)\right\} \\ &\quad - I\left(X^{\theta_M}; Y^{\theta_M}\right) + H(S) \end{aligned} \tag{1}$$

Minimizing (1), the end-to-end model can maximize the mutual information (MI). However, the SAE focus on symbol-wise metric optimization and may not achieve bit-wise metric improvement such as GMI and BER due to decoding loss. Therefore, we consider BAE using binary cross entropy (BCE) loss function in end-to-end learning. The loss function can be expressed by,

$$\begin{aligned} L_{BCE}(\theta_M, \theta_D) &= \sum_{i=1}^{m} E_{b,Y^{\theta_M}}\left\{-\log g^{\theta_D}\left(b_i|y^{\theta_M}\right)\right\} \\ &= \sum_{i=1}^{m} E_Y\left\{D_{KL}\left(f\left(b_i|y^{\theta_M}\right)\left\|g^{\theta_D}\left(b_i|y^{\theta_M}\right)\right)\right\} \\ &\quad - \sum_{i=1}^{m} I\left(b_i; Y^{\theta_M}\right) + H(s) \end{aligned} \tag{2}$$

where b_i denotes the random variable i-th position in the vector b. The BAE can optimize the probability of transmitted bits to reduce the BER and improve the GMI for the bit-wise metric systems by minimizing (2). This is more suitable for practical bit-wise metric systems.

B. End-to-End Learning for Probabilistic Shaping

The PS has been demonstrated that it can improve the additive white Gaussian noise (AWGN) noise tolerance and realize adaptive rate transmissions in optical fiber communications [21]. Based on BAE, we propose an end-to-end learning scheme for PS. As shown in Fig. 2, the model includes a mapper with trainable probability and a demapper generating the bit-wise metric inputs. At the transmitter, the probability vector $P_M = [p_1, \ldots, p_M]^T$, $\sum_{j=1}^{M} p_j = 1$ is generated from the input vectors $b_M = [b_1, \ldots, b_M]^T$, where $b_j, j = 1, \ldots, M$ is bit vector of length m. To generate a training symbols of batch size B, the sampler takes each index j about Bp_j times by drawing from P_M and then randomly permutes all indices. Each generated index is converted to the one-hot mapping, which is sent to a single linear layer DNN of size $M \times 1$ whose weights are $\tilde{c}_m$ to realize the symbol mapping. Note that the input constellation points $C_M = [c_1, \ldots, c_M]$ are normalized by the probability vector P_M to meet the power constraint, i.e., $\tilde{c}_m = c_m / \sqrt{\sum_{j=1}^{M} p_j c_j^2}$. The resulted training batch $X = [x_1, \ldots, x_B]$ is sent to the channel. At the receiver, the signal y_n from $Y = [y_1, \ldots, y_B]$ is demapped to generate the expected bit vector $\tilde{b} = (\tilde{b}_1, \ldots, \tilde{b}_m)$.

The generation of probability vector and demapping of received signal are realized by autoencoder in the proposed model. The DNN to generate different probabilities for constellation points consisting of three dense layers, where two with eLU activations and the last one with linear activation. The output of the DNN is the logit corresponding to the bit vector

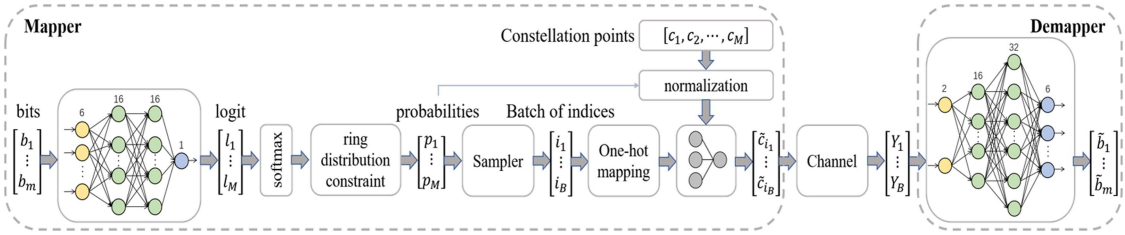

Fig. 2. Diagram of end-to-end learning model based on bit-wise autoencoder for PS.

$b = (b_1, \ldots, b_m)$, which is used to generate the probability by applying softmax function. After mapping the whole bit vectors, the mean probability of constellation points with the same Euclidean distance is calculated to make probability follows ring distributions similar with constant CCDM. Thus, the probability vector P_M can be obtained. The demapper DNN with three dense layers takes the real and imaginary parts of received signal as inputs to generate the expected bit vector. The first two layers of demapper DNN adopt eLU activations and the last layer adopts linear activation.

One of the most common distributions for PS is Maxwell-Boltzmann (MB) distribution, which is defined as,

$$MB\left(\hat{b}_i, k\right) = \frac{\exp\left(-\left(\hat{b}_i - b_i\right)^2 / k\right)}{\sum_{b_i=\{0,1\}} \exp\left(-\left(\hat{b}_i - b_i\right)^2 / k\right)} \tag{3}$$

where $k > 0$ determines the shape of the distribution function. The MB distribution including the structure of sigmoid function in the BAE can be defined as,

$$\begin{aligned} &p\left(b_i = 1|y\right) = \sigma\left(z_i\right) \\ &= \frac{1}{1 + \exp\left(-\left(\left\|\hat{b}_i - 0\right\|^2 - \left\|\hat{b}_i - 1\right\|^2\right)/k\right)} \end{aligned} \tag{4}$$

where $z_i = (\|\hat{b}_i - 0\|^2 - \|\hat{b}_i - 1\|^2)/k$ represents the similarity of the transmitted bit to '0' or '1'. Therefore, the probability of transmitted bit can be approximated by the MB distribution. The BCE loss can be expressed as,

$$\begin{aligned} L\left(\theta_M, \theta_D\right) &= -\sum_{i=1}^{m} \log\left(\frac{\exp\left(-\left(\hat{b}_i - b_i\right)^2 / k\right)}{\sum\limits_{b_i \in \{0,1\}} \exp\left(-\left(\hat{b}_i - b_i\right)^2 / k\right)}\right) \\ &= \sum_{i=1}^{m}\left[\frac{1}{k}\left\|\hat{b}_i - b_i\right\|^2 + \log\left(\exp\left(-\left(\hat{b}_i - 1\right)^2 / k\right)\right.\right. \\ &\qquad \left.\left. + \exp\left(-\left(\hat{b}_i\right)^2 / k\right)\right)\right] \end{aligned} \tag{5}$$

The second term in (5) achieves the lowest at both '0' and '1', which can be ignored in parameters optimization. The loss function can be simplified as MSE loss. Moreover, PS scheme should be consistent with the source entropy in optical fiber communications. Therefore, we propose a modified loss function of BAE for PS by combining the MSE and the source entropy to optimize the achievable information rate, which is defined as,

$$\begin{aligned} L\left(\theta_M, \theta_D\right) &= MSE\left(b, \tilde{b}\right) + \lambda \cdot \left(H\left(P_M\right) - H_x\right) \\ &\approx \sum_{k=1}^{m}\left\|\tilde{b}_k - b_k\right\|^2 + \lambda \cdot \left(\sum_{j=1}^{M} p_j \log_2 p_j - H_x\right) \end{aligned} \tag{6}$$

where H_x is the source entropy, and λ is the scaling factor.

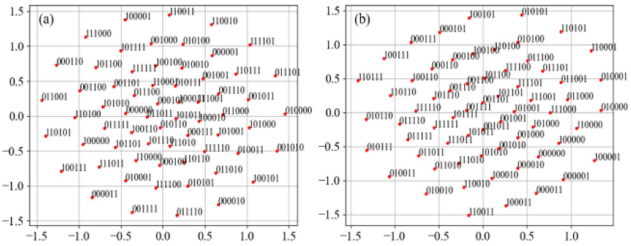

Fig. 3. Constellations learned by end-to-end model for GS based on (a) SAE, (b) BAE.

III. Numerical Simulations

A. AWGN Channel

We train the end-to-end model based on autoencoder on AWGN channel. We calculate achievable information rate and LLR to compare the performance of transmitted signal with GS or PS [22].

For GS, the signal-to-noise Ratio (SNR) is set at 15-dB with the length of bit vector $m = 6$. Glorot initializer and Adam optimizer are adopted and the training epoch numbers of both SAE with CE loss and BAE with BCE loss are set to 1000. Fig. 3 shows the constellations with bit labels learned by end-to-end model, which are used for subsequent testing. In Fig. 3, the autoencoder can realize symmetric mapping with normalized average power. The distance between constellations increases compared to conventional QAM, which can improve the noise tolerance in communication systems. Compared to SAE, the BAE can learn constellation with Gray labeling, which can reduce the decoding loss and improve the reliability of bits.

We first compare the performance by applying the learned constellations in Fig. 3 on AWGN channel. Fig. 4(a) and (b) show the MI and GMI performance versus SNRs, respectively. As shown in Fig. 4(a), the SAE and the BAE achieves 0.15-bits/sym

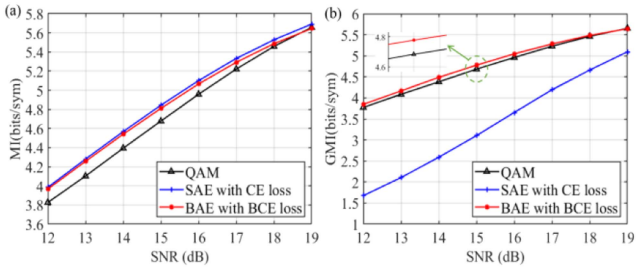

Fig. 4. (a) MI, (b) GMI performance of GS versus SNR on AWGN channel.

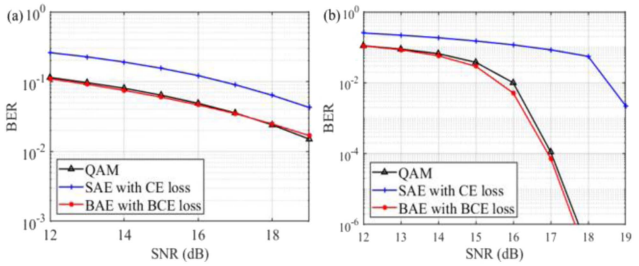

Fig. 5. (a) pre-FEC, (b) post-FEC BER performance of GS versus SNR on AWGN channel.

and 0.12-bits/sym MI gain compared to conventional 64-QAM, respectively. However, the GMI performance of SAE degrades compared with 64-QAM due to the decoding loss, where the BAE still achieves 0.11-bits/sym gain. Fig. 5(a) and (b) show the pre-FEC and post-FEC BER performance on AWGN channel, respectively. The constellation learned by SAE exists BER performance penalty due to the decoding loss. The constellation learned by BAE can achieve similar pre-FEC BER performance as conventional 64-QAM. There is about 0.1-dB SNR gain after 5-iteration BP decoding. These results show that the end-to-end learning based on SAE only improves the symbol-wise metric performance. The GS based on BAE can achieve bit-wise metric performance improvement, which is more suitable for practical communication systems.

For PS, the SNR is also set at 15-dB. The scale factor of the proposed loss function in (6) is set as $\lambda = 0.1$ and the maximum training epoch number is 50000. Early stopping is adopted during the training process to avoid overfitting. Note that the BAE is initialized with parameters obtained by pretraining the networks with uniform distribution as labels. The training process is realized with Adam optimizer.

Fig. 6(a)–(c) show the probability distributions learned by the end-to-end model with 5.2-bits/sym, 5.5-bits/sym and 5.8-bits/sym source entropies, respectively. Fig. 6(d) shows the probability distribution realized by CCDM with H_x of 5.5-bits/sym. The BAE with the proposed loss function can generate probability distributions similar to that with CCDM algorithm. Where PS with continuous H_x can realize adaptive rate transmissions in optical fiber communications.

We compare the GMI of the learned results in Fig. 6(a)–(c) on AWGN channel with different SNRs for testing. Fig. 6(e) shows the GMI performance, in low SNR region, the probability distribution learned by BAE with H_x of 5.8-bits/sym outperforms uniform 64-QAM by 0.2-bits/sym gain, which is 0.3-bits/sym with H_x of 5.5-bits/sym and 5.2-bits/sym. In higher SNR region, the proposed scheme achieves 5.8-bits/sym, 5.5-bits/sym and 5.2-bits/sym GMIs, which corresponds to the trained source entropies. Moreover, we take CCDM with H_x of 5.5-bits/sym in comparison, which shows that PS based on BAE achieves the similar performance compared with that by CCDM scheme. These results show that the end-to-end learning can realize PS and improve the transmission performance.

B. VPI System

To verify the effectiveness of the end-to-end leaning for constellation shaping, we set up a dual polarization (DP) coherent transmission simulation using VPItransmissionMakerTM 9.1. The transmitted signal is a 32-GBaud 64-QAM signal with GS or PS. A root-raised cosine filter with roll-off 0.1 is used for Nyquist pulse shaping. The linewidth of the laser is set as 100-kHz. The transmission link consists of 100-km SSMF and an Erbium-doped optical fiber amplifier (EDFA) with 5-dB noise figure to compensate for the loss. After coherent detection, the received signal is processed by digital signal processing (DSP) algorithms.

Figs. 7 and 8 show the transmission performance of GS in DP coherent system over 100-km transmission. As shown in Fig. 7, the constellations learned by SAE outperforms conventional 64-QAM by 0.2-bits/sym MI but suffers beyond 0.3-bits/sym GMI penalty. The BAE obtains 0.15-bits/sym MI gain and 0.13-bits/sym GMI gain. As shown in Fig. 8, the constellation learned by SAE has BER degradation compared with Gray mapping. Compared to conventional 64-QAM, the BAE can achieve the similar pre-FEC BER performance and obtain about 0.1-dB OSNR gain after 3-iteration BP decoding. These results show that GS based on BAE is effective in optical fiber communication system with FEC.

Fig. 9 shows the performance of the learned probability distributions in DP 100-km coherent system. The proposed scheme achieves about 0.2 to 0.4-bits/sym GMI gain in low OSNR region. The gain obtained by PS based on BAE increases with the decrease of H_x. As OSNR increases, the GMI performance becomes close to the adopted source entropy in training process. These results show that PS realized by end-to-end learning is robust to optical fiber communication system.

IV. Experimental Setup and Results

We also conduct an experiment to transmit different constellation shaping signals over 100-km SSMF to test their performances. Fig. 10 shows the experimental setup and the DSP flow of the coherent optical fiber communication system. At the transmitter side, the bitstream is modulated by the constellations with GS or PS. Then, it is up-sampled to 2 samples per symbol for Nyquist pulse shaping. The roll-off factor of the shaping filter is 0.1. Subsequently, the sequence is loaded to an arbitrary waveform generator (AWG) operating at 92 GS/s with 32-GHz bandwidth to generate a modulated signal. The electrical signal

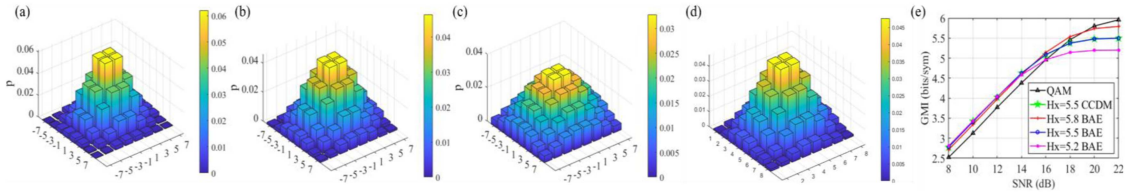

Fig. 6. Probability distribution learned by bit-wise autoencoder with H_x of (a) 5.2-bits/sym, (b) 5.5-bits/sym, (c) 5.8-bits/sym. (d) Probability distribution realized by CCDM algorithm with H_x of 5.5-bits/sym, (e) GMI performance of PS versus SNR on AWGN channel.

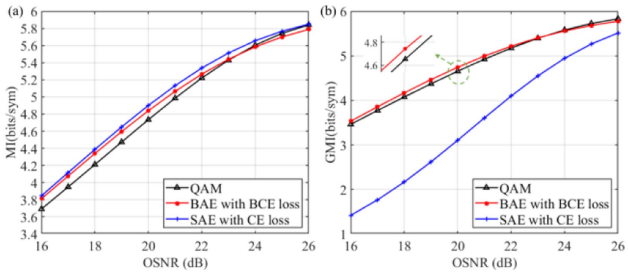

Fig. 7. (a) MI, (b) GMI performance of GS versus OSNR in DP 100-km coherent system.

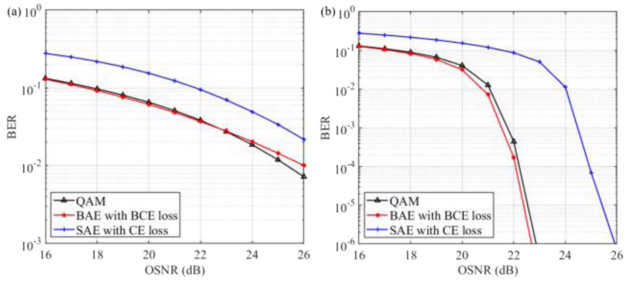

Fig. 8. (a) Pre-FEC, (b) post-FEC BER performance of GS versus OSNR in DP 100-km coherent system.

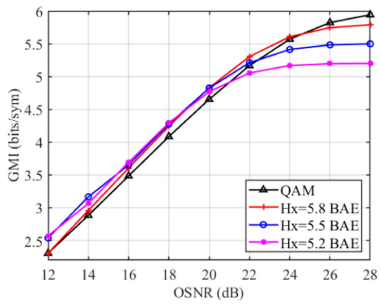

Fig. 9. GMI performance of PS versus OSNR in DP 100-km coherent system.

is fed into an optical in-phase/quadrature (I/Q) modulator to realize electrical to optical conversion. The optical carrier is working at 1550 nm (linewidth∼100-kHz). The generated signal has a symbol rate of 23-GBaud. The transmission link consists of 100-km standard SSMF, an EDFA to compensate for the fiber loss and an optical band-pass filter (OBPF) to filter out the out-of-band noise. After fiber transmission, the received optical signal is detected by the coherent receiver and digitized by a digital processing oscilloscope (DPO) operating at 50 GS/s with 23-GHz bandwidth. After the coherent detection, DSP is operated offline, e.g., I/Q imbalance, chromatic dispersion compensation, resampling, frequency offset and phase noise estimation. Then, we use the learned constellations to realize demapping and evaluate the performance.

We use 64-QAM with uniform distribution modulation as baseline, and the size of transmitted symbols is 90000. For GS, the input bitstream consists of codewords generated by LDPC (16384,12288) of rate 3/4. And the constellations learned by SAE and BAE are adopted to generate the modulated symbols. For PS, the bitstream is generated by sampling and randomly permuting all indices with the learned probability vector P_M. Moreover, the constellation is 64-QAM normalized by P_M. We analyze the achievable information rate and BER performance of signals.

Figs. 11 and 12 show the performance of GS based on SAE and BAE in the experiment. As shown in Fig. 11, the constellations learned by SAE and BAE both achieve up to 0.2-bits/sym MI gain. However, the SAE scheme exists GMI performance penalty, which is consistent with the simulation results. The BAE scheme outperforms 64-QAM by 0.11-bits/sym GMI gain in linear region. Fig. 12 shows the BER performance, the constellation learned by SAE cannot reach the same performance as the conventional 64-QAM with Gray mapping. Compared to 64-QAM, the constellation learned by BAE achieves similar pre-FEC BER performance and about 0.7-dB LOP gain at BER of 3.8×10^{-3} after 3-iteration BP decoding. These results show that GS based on SAE leads to decoding loss, while GS based on BAE can obtain bit-wise metric gains over optical fiber transmissions.

Fig. 13 shows the performance of PS with the proposed loss function in the experiment. Compared to uniform 64-QAM, the learned distributions with H_x of 5.5-bits/sym and 5.2-bits/sym achieve up to 0.25-bits/sym and 0.3-bits/sym GMI gains in the linear region, respectively. In the region of optimal LOP, the GMIs are 5.8-bits/sym, 5.4-bits/sym and 5.15-bits/sym, respectively, which close to the source entropies adopted in training process. These results demonstrate the effectiveness of the proposed scheme, which can realize PS robust to optical fiber communication system.

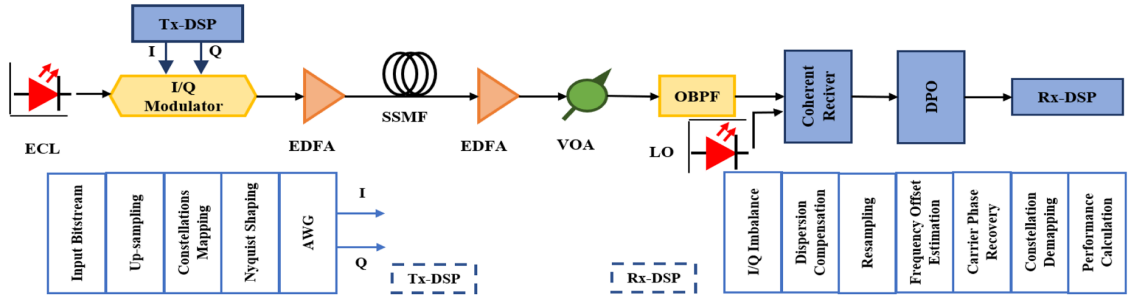

Fig. 10. System setup for 100-km coherent optical fiber communication system.

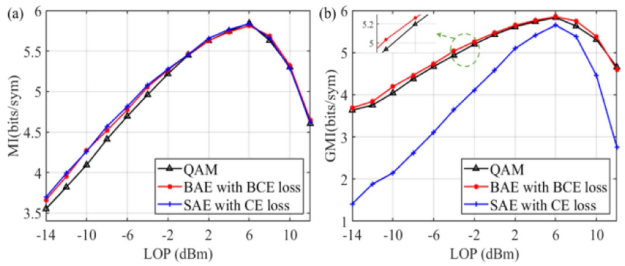

Fig. 11. (a) MI, (b) GMI performance of GS versus LOP in experimental 100-km coherent system.

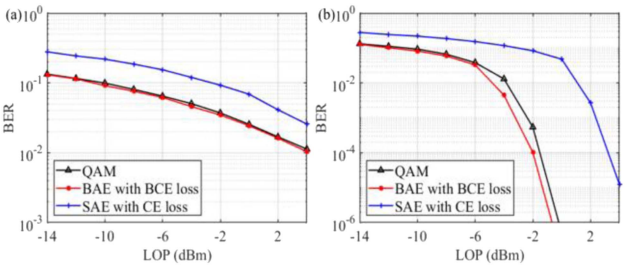

Fig. 12. (a) Pre-FEC, (b) post-FEC BER performance of GS versus LOP in experimental 100-km coherent system.

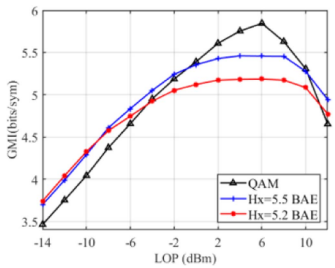

Fig. 13. GMI performance of PS versus LOP in experimental 100-km coherent system.

V. Conclusion

We have proposed and experimentally demonstrated an end-to-end learning based on BAE to generate GS and PS constellations in optical fiber communication systems. We find that the BAE is more suitable for bit-wise metric systems. The simulation and experimental results show that the GS and PS with BAE based end-to-end learning is more suitable for practical applications.

References

[1] X. Pan, X. Wang, B. Tian, C. Wang, H. Zhang, and M. Guizani, "Machine-learning-aided optical fiber communication system," *IEEE Netw.*, vol. 35, no. 4, pp. 136–142, Jul./Aug. 2021.

[2] F. N. Khan, K. Zhong, W. H. Al-Arashi, C. Yu, C. Lu, and A. P. T. Lau, "Modulation format identification in coherent receivers using deep machine learning," *IEEE Photon. Technol. Lett.*, vol. 28, no. 17, pp. 1886–1889, Sep. 2016.

[3] J. Thrane, J. Wass, M. Piels, J. C. M. Diniz, R. Jones, and D. Zibar, "Machine learning techniques for optical performance monitoring from directly detected PDM-QAM signals," *J. Lightw. Technol.*, vol. 35, no. 4, pp. 868–875, Feb. 2017.

[4] J. Estaran et al., "Artificial neural networks for linear and non-linear impairment mitigation in high-Baudrate IM/DD systems," in *Proc. Eur. Conf. Opt. Commun.*, 2016, pp. 1–3.

[5] M. Zahid and Z. Meng, "Recent advances in neural network techniques for channel equalization: A comprehensive survey," in *Proc. Int. Conf. Comput. Electron. Commun. Eng.*, 2018, pp. 178–182.

[6] S. Cammerer, F. A. Aoudia, S. Dörner, M. Stark, J. Hoydis, and S. ten Brink, "Trainable communication systems: Concepts and prototype," *IEEE Trans. Commun.*, vol. 68, no. 9, pp. 5489–5503, Sep. 2020.

[7] B. Karanov, P. Bayvel, and L. Schmalen, "End-to-end learning in optical fiber communications: Concept and transceiver design," in *Proc. Eur. Conf. Opt. Commun.*, 2020, pp. 1–4.

[8] B. Karanov et al., "End-to-end deep learning of optical fiber communications," *J. Lightw. Technol.*, vol. 36, no. 20, pp. 4843–4855, Oct. 2018.

[9] R. T. Jones, M. P. Yankov, and D. Zibar, "End-to-end learning for GMI optimized geometric constellation shape," in *Proc. Eur. Conf. Opt. Commun.*, 2019, pp. 1–4.

[10] 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 *Proc. Eur. Conf. Opt. Commun.*, 2021, pp. 1–4.

[11] Z. Zhai et al., "An interpretable mapping from a communication system to a neural network for optimal transceiver-joint equalization," *J. Lightw. Technol.*, vol. 39, no. 17, pp. 5449–5458, Sep. 2021.

[12] S. Li, C. Häger, N. Garcia, and H. Wymeersch, "Achievable information rates for nonlinear fiber communication via end-to-end autoencoder learning," in *Proc. Eur. Conf. Opt. Commun.*, 2018, pp. 1–3.

[13] R. T. Jones, T. A. Eriksson, M. P. Yankov, and D. Zibar, "Deep learning of geometric constellation shaping including fiber nonlinearities," in *Proc. Eur. Conf. Opt. Commun.*, 2018, pp. 1–3.

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

[15] V. Neskorniuk et al., "End-to-end deep learning of long-haul coherent optical fiber communications via regular perturbation model," in *Proc. Eur. Conf. Opt. Commun.*, 2021, pp. 1–4.

[16] B. Karanov, M. Chagnon, V. Aref, D. Lavery, P. Bayvel, and L. Schmalen, "Concept and experimental demonstration of optical IM/DD end-to-end system optimization using a generative model," in *Proc. Opt. Fiber Commun. Conf.*, 2020, pp. 1–3.

[17] M. Li, D. Wang, Q. Cui, Z. Zhang, L. Deng, and M. Zhang, "End-to-end learning for optical fiber communication with data-driven channel model," in *Proc. Opto-Electron. Commun. Conf.*, 2020, pp. 1–3.

[18] F. A. Aoudia and J. Hoydis, "Joint learning of probabilistic and geometric shaping for coded modulation systems," in *Proc. IEEE Glob. Commun. Conf.*, 2020, pp. 1–6.

[19] V. Aref and M. Chagnon, "End-to-end learning of joint geometric and probabilistic constellation shaping," in *Proc. Opt. Fiber Cpmmun. Conf.*, 2022, pp. 1–3.

[20] K. Gümüş, A. Alvarado, B. Chen, C. Häger, and E. Agrell, "End-to-end learning of geometrical shaping maximizing generalized mutual information," in *Proc. Opt. Fiber Commun. Conf.*, 2020, Paper W3D.4.

[21] F. Buchali, F. Steiner, G. Böcherer, L. Schmalen, P. Schulte, and W. Idler, "Rate adaptation and reach increase by probabilistically shaped 64-QAM: An experimental demonstration," *J. Lightw. Technol.*, vol. 34, no. 7, pp. 1599–1609, Apr. 2016.

[22] A. Alvarado, T. Fehenberger, B. Chen, and F. M. J. Willems, "Achievable information rates for fiber optics: Applications and computations," *J. Lightw. Technol.*, vol. 36, no. 2, pp. 424–439, Jan. 2018.