MINE-based Geometric Constellation Shaping in AWGN Channel

Qian Wang, Xiuli Ji, Li Ping Qian, Zilong Liu, Xinwei Du and Pooi-Yuen Kam *Life Fellow, IEEE*

***Abstract*—The use of high-order constellation modulations is imperative to improve the spectral efficiency, for both radio frequency/laser-based satellite systems and optical wireless communications. The geometric shaping (GS) optimization as one typical constellation shaping method drives the improvement of communication capacity and system performance. This paper presents a novel mutual information neural estimation (MINE)-based GS method to optimize the high-order constellations in pure additive white Gaussian noise (AWGN) channel, which uses the deep neural network (DNN) to estimate the mutual information (MI) value and maximize the MI to approach the AWGN capacity asymptotically. The proposed system trains both the encoder and MINE networks by back propagation, and does not need to train a decoder for optimization and thus can avoid the loss caused by the decoder. Simulation results show that the MINE-based shaping design outperforms the unshaped M-ary quadrature amplitude modulation (QAM) in terms of MI values. Note that the capacity gain increases slightly as the order M increases. Furthermore, the proposed scheme is promising for constellation design in various channel models, such as the phase noise and the fading channels, once the channel model used in MINE is matched, which can be a future research topic.**

***Index Terms*—Constellation design, geometric shaping, mutual information neural estimation, mutual information maximization, high-order QAM.**

I. Introduction

With the rapid growth of communication data and the increasingly tight spectrum resources, using high-order modulations to improve the spectral efficiency is a hot trend in the field of both radio frequency/laser-based satellite systems and optical wireless communications. Constellation shaping as a typical multi-level modulation optimization technology is becoming more and more important, which is used to significantly improve the spectral efficiency. Constellation design can be divided into two types: probabilistic shaping (PS) [1], [2] and geometric shaping (GS) [3], [4]. In the former type, the constellation points are transmitted at fixed locations with unequal probabilities to maximize the mutual information (MI) in a given channel. By contrast, the constellation points in GS-based type are transmitted with equal probabilities but with changing geometric locations of the constellation points in Euclidean space. PS-based optimization has been proven to be an effective strategy to improve spectral efficiency and provide rate adaptivity [5]. However, the PS requires an external distribution matcher for efficient implementation, which limits its shaping gain. As a low-complexity alternative, GS-based optimization only requires modifying the mapper and demapper for design, which can be easily adapted to different impairments, such as fiber nonlinearity [6] and laser phase noise [7].

The basic quadrature amplitude modulation (QAM) constellations are widely recognized and extensively used for high-rate transmission. In the bandwidth-limited additive white Gaussian noise (AWGN) channel, conventional high-order QAM with square or rectangular structures can result in an asymptotic loss of 1.53 dB towards the Shannon limit [8]. With the aid of GS, various novel QAM constellations have achieved tremendous attention to further narrow the gap, such as cross QAM, star QAM, and hexagonal QAM. Regarding the GS-based constellation design, most literatures consider two available optimization criteria including average symbol error rate (SER) minimization [9], [10] and MI maximization [11], [12]. For example, [13] made a comparative analysis of average SER for several high-order QAM constellations, which justified the supremacy of hexagonal QAM for wireless communication systems. Additionally, [14] designed some advanced modulation formats based on generalized MI (GMI), and verified their better performance on GMI compared to the standard star-8QAM and 32QAM.

Note that in the MI maximization-based constellation shaping, the MI calculation is required to capture the non-linear statistical depencies between variables through a specific channel [15]. However, direct and accurate calculation of MI is difficult, since estimating MI depends on the underlying joint probability density function which should match the channel [16]. Furthermore, optimizing the constellation shape to maximize the MI in a given channel becomes more computationally complex, as the constellation size and dimensionality increase. To deal with these problems, the deep learning (DL) approach is introduced into physical layer communications [17], [18]. In this way, DL-based MI calculation and constellation design become feasible to improve the efficiency. In the conventional end-to-end DL approach that treats the transmitter and receiver as trainable neural networks, the constellation shaping communication systems adopt the autoencoder to jointly train and optimize the positions of constellation points in high-order

Q. Wang, X. Ji and L. P. Qian are with the College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023, China (email: {wangqian18,lpqian}@zjut.edu.cn). Z. Liu is with the School of Computer Science and Electronics Engineering, University of Essex, UK (e-mail: zilong.liu@essex.ac.uk). X. Du is with the Division of Science and Technology, BNU-HKBU United International College, Zhuhai, China 519087 (email: xinweidu@uic.edu.cn). P.Y. Kam is with the School of Science and Engineering, the Chinese University of Hong Kong (Shenzhen), China 518172 (email: pykam@cuhk.edu.cn). This work was supported in part by National Natural Science Foundation of China under Grants 62201507, 62122069, 62072490, and 62071431, in part by the Intergovernmental International Cooperation in Science and Technology Innovation Program under Grants 2019YFE0111600.

modulations to maximize the MI [19], [20].

This paper will focus on proposing a DL-based geometric constellation design scheme based on MI maximization. Specifically, we introduce a novel mutual information neural estimation (MINE)-based GS method to optimize the high-order constellations in pure AWGN channel, which trains both MINE and encoder networks to calculate and maximize the MI via back propagation. Note that the principle of the MINE was originally proposed in [21], and the MI between high-dimensional continuous random variables was estimated through gradient descent of the deep neural networks (DNN). It is worth emphasizing that our MINE-based GS system replaces the decoder in the traditional autoencoder with a MINE network, and does not need to train the decoder for optimization, thus avoiding the loss caused by the decoder. For simplicity in training, the loss function of the designed system is set as the negative of MI calculation function. With this setting, the MINE-based GS system achieves back propagation by gradient descent over the neural networks, and thus can effectively train both the encoder and MINE networks to maximize the MI which asymptotically approaches the AWGN capacity.

The remainder of this paper is organized as follows. Section II describes the system model. Section III presents the specific MINE-based geometric constellation design. Section IV provides the simulation analysis. Section V summarizes our work.

II. System Model

The pure AWGN channel model is considered in this work, and thus the received signal can be written as

$$Y_k = X_k + Z_k, \tag{1}$$

where Y_k represents the received symbol in time slot $k \in \{1, 2, ...\}$, X_k represents transmitted symbol from the M possible modulated symbols $\{X_1, ..., X_M\}$ in the constellation and Z_k represents the complex AWGN variable with mean 0 and variance N_0. Here, we denote E_s as the average transmit energy per symbol, and thus the signal-to-noise ratio (SNR) can be defined as SNR = $\frac{E_s}{N_0}$. Note that the information-theoretic Shannon capacity defines the maximum information rate of the channel, and given by $C = \frac{N}{2}\log_2(1 + \text{SNR})$ where N is the number of real dimensions and we have $N = 2$ in this paper.

Here, we focus on the MI of the AWGN channel in which the input constellation symbols are equiprobable with order M. The probability distribution of transmitted symbol X is denoted as f_X, and we have $f_X(x) = \frac{1}{M}$. The MI of the AWGN channel is defined as

$$\begin{aligned} I(X;Y) &= H(X) - H(X|Y) \\ &\triangleq \mathbb{E}\left[\log \frac{f_{Y|X}(y|x)}{f_Y(y)}\right], \end{aligned} \tag{2}$$

where $I(X;Y)$ denotes the MI between X and Y, $H(X)$ denotes information entropy of X, $H(X|Y)$ denotes the conditional information entropy of X given Y, and $\mathbb{E}[\cdot]$ denotes the mathematical expectation. Besides, f_Y denotes the probability distribution of Y

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{+\infty} f_{Y|X}(y|x)\, f_X(x)\mathrm{d}(x) \\ &= \frac{1}{M}\int_{-\infty}^{+\infty} f_{Y|X}(y|x), \end{aligned} \tag{3}$$

$f_{Y|X}$ denotes the conditional distribution, that is channel law

$$f_{Y|X}(y|x) = \frac{1}{(\pi N_0)^2} exp\left(-\frac{||y-x||^2}{N_0}\right). \tag{4}$$

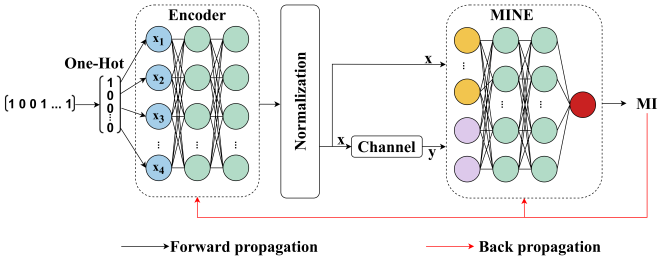


Fig. 1. Structure of the proposed MINE-based GS system.

III. Geometric Constellation Design

This section first presents the architecture of the MINE-based GS system and explains the specific role of each part for geometric constellation design. Then, the principle of the MINE and the specific expression for calculating MI are introduced in detail. Finally, the training process and parameter settings of the proposed scheme are given for optimization.

A. Geometric design architecture

The proposed MINE-based GS system structure is illustrated in Fig.1. In this framework, there are an encoder module, a power normalization module, a channel module, and a MINE module. Both the encoder and MIEN are two fully connected neural networks, which are used for modulation and MI estimation, respectively. The constellation points are transmitted equally and the outputs from the encoder are the modulated symbols. The MINE network inputs symbols from both ends of the channel and outputs the estimated MI value, which can be fed back to the encoder for constellation optimization.

For the forward propagation process, a random string of raw bits are first generated and then mapped into constellation points according to two encoding rules which include Gray encoding and Natural encoding. These constellation symbols are converted into a one-hot code vector of length M and input into the encoder for modulation. And the elements of the one-hot code vector are defined as

$$e_{v_i}[j] = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{otherwise} \end{cases} \qquad j = 0, \ldots, M-1, \tag{5}$$

where v_i is the i-th symbol, $e_{v_i}[j]$ is the j-th vector element corresponding to the symbol v_i, and M is the modulation order, i.e. Each input symbol is encoded into the in-phase and quadrature (I/Q) components by the encoder. Subsequently, the average power of the outputs by the encoder is normalized to 1 to ensure the power efficiency of the designed constellation. After power normalization, the symbol X_k is expressed in the I/Q form and through the AWGN channel, we obtain the damaged symbol Y_k. These symbols X_k and Y_k at both ends of the channel are entered into the MINE network for estimating the MI. Finally, the system trains both the encoder and MINE networks and maximizes MI through back propagation.

B. The MINE principle

This subsection introduces the principle of the MINE that can estimate the MI between high-dimensional continuous random variables by gradient descent over neural networks. Note that the MI between X and Y is defined as

$$I(X;Y) = \int_{\mathcal{X}\times\mathcal{Y}} \log \frac{f_{XY}}{\mathrm{d}f_X \otimes f_Y} \mathrm{d}f_{XY}, \tag{6}$$

where f_{XY} is the joint distribution of X and Y, and $\otimes$ is the tensor product between two distributions.

Note that the KL divergence, also called relative entropy, is an asymmetric measure of the difference between the joint distribution and the product of the marginals [22]. Thus, the KL divergence can be used to achieve a general mutual information estimator. In information theory, the MI between X and Y is equivalent to the KL divergence between the joint distribution f_{XY} and the product of the marginals $f_X \otimes f_Y$. That is, we have

$$\begin{aligned} I(X;Y) &= D_{\mathrm{KL}}\left(f_{XY} \parallel f_X \otimes f_Y\right) \\ &= \mathbb{E}_{f_{XY}}\left[\frac{\mathrm{d}f_{XY}}{\mathrm{d}f_X \otimes f_Y}\right]. \end{aligned} \tag{7}$$

In this way, the problem of MI estimation is transformed into the estimation of KL divergence.

The focus of MINE is on the dual representation of KL divergence, also known as Donsker-Varadhan representation [23], which is specifically expressed as

$$\begin{aligned} &D_{\mathrm{KL}}\left(f_{XY} \| f_X \otimes f_Y\right) \\ &\quad = \sup_{T:\Omega\to\mathbb{R}} \mathbb{E}_{f_{XY}}[T] - \log\left(\mathbb{E}_{f_X\otimes f_Y}\left[e^T\right]\right), \end{aligned} \tag{8}$$

where e is the natural logarithmic base, T is a function in the set of all functions $\mathbb{R}$, $T:\Omega\to\mathbb{R}$ satisfying the integrability constraints of the theorem, and sup means the upper bound of the function.

For the MI estimation based on the dual representation of the KL divergence, the idea is that assuming the set of functions $T_\theta : X\times Y \to \mathbb{R}$ parametrized by a DNN with parameters $\theta \in \Theta$. And the neural network can obtain $T_{\theta optim}$ and MI maximization by optimizing θ. The relationship between the MI of the neural network and the real MI is thus defined as

$$I_\Theta(X,Y) \le I(X;Y), \tag{9}$$

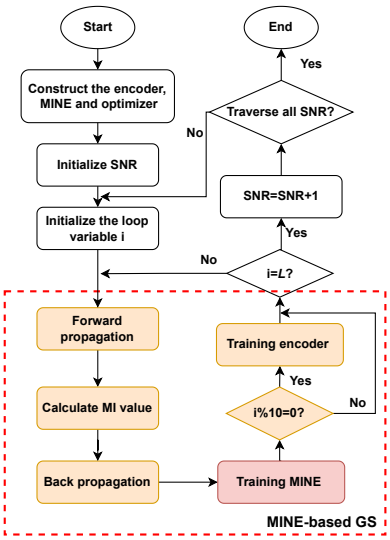


Fig. 2. Training procedure of the proposed MINE-based GS scheme.

where the neural information measure of the MINE network with parameter $\theta \in \Theta$ is defined as

$$I_\Theta(X,Y) = \sup_{\theta\in\Theta} \mathbb{E}_{f_{XY}}[T_\theta] - \log\left(\mathbb{E}_{f_X\otimes f_Y}\left[e^{T_\theta}\right]\right). \tag{10}$$

For network training, the loss function of the MINE-based system is set as the negative value of MI measurement and thus expressed as

$$\begin{aligned} Loss &= -I_\Theta(X,Y) \\ &= -\left\{\sup_{\theta\in\Theta} \mathbb{E}_{f_{XY}}[T_\theta] - \log\left(\mathbb{E}_{f_X\otimes f_Y}\left[e^{T_\theta}\right]\right)\right\}. \end{aligned} \tag{11}$$

With this specific loss function, the system achieves back propagation by gradient descent over the neural networks, and therefore, we can train both the encoder and MINE to maximize MI. The maximized MI value can be thus iteratively estimated and calculated.

C. Construction and training details

The construction and training details of the proposed network are given here. The encoder network has 3 hidden layers and each hidden layer has 256 neurons. And the Leaky ReLU is used as the activation function therein. As for the MINE network, it has 3 hidden layers and each hidden layer has 128 neurons with the Leaky ReLU as its activation function. To maximize the MI numerically, the Adam optimizer in [24] is used to optimize the network parameters of the encoder and MINE networks. The learning rate for both networks is set to 0.01. More importantly, the back propagation algorithm is used to calculate the gradients for optimizing the parameter θ.

The detailed training procedure is shown in Fig.2. Firstly, the encoder, MINE, and optimizer are constructed. A total of L iteration cycles are trained. In each training cycle, the constellation points to be optimized for each SNR are propagated forward as shown in Fig.1, and the estimated value of MI is calculated based on the trained MINE network. Specifically, in the forward propagation process, the model

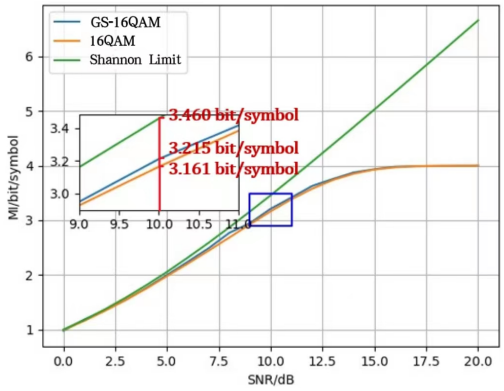


Fig. 3. MI comparison between the proposed MINE-based GS-16QAM using Gray encoding and the unshaped 16QAM in AWGN channel.

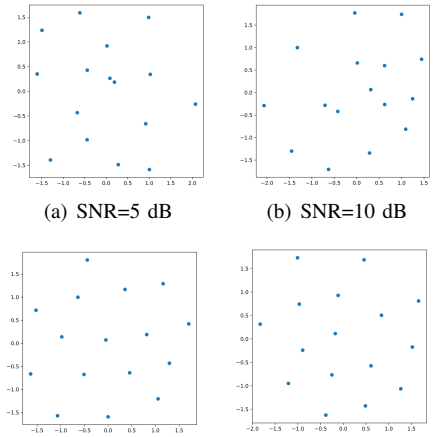
traverses the training set and performs MI calculation based on (10). In the back propagation training process, there are three key steps including gradient zeroing, gradient calculation and gradient updating. Most importantly, loss function is used to update the gradient and optimize the parameters of the neural networks through back propagation, facilitating the training of the encoder and MINE networks. Note that the number of iteration cycles is set to L=500 in this simulation to ensure loss function is convergent. The MINE network is trained at each iteration cycle, whereas the encoder neural network is trained every 10 iteration cycles.

IV. Numerical Results

To demonstrate the feasibility of our approach, we simulate the MI performance of the learned GS-based QAMs in the AWGN channel in this section. Simulations are implemented using the Pytorch DNN framework with Python programming language. Note that the MI calculation adopts the Monte Carlo estimation method, where using sufficient sample size can guarantee that the true MI can be approximated.

To verify the performance of the designed MINE-based system, the MI values of MINE-based GS-16QAM and unshaped 16QAM are compared in the SNR range from 0 dB to 20 dB as shown in Fig.3. Under the condition that the loss function converges, it can be seen that under low SNR, the MI of both modulation formats are relatively close to the Shannon limit. However, the proposed GS-16QAM outperforms the regular 16QAM under all considered SNR values. The curve at SNR = 10 dB is mainly studied. As observed, the GS-16QAM has about 0.054 bit/symbol gain compared to the unshaped 16QAM in terms of MI at 10 dB, and thus decreases the gap with the capacity limit. Fig.4 shows the MINE-based GS constellation points for 16QAM at several SNRs. Note that for higher SNR, these constellations are displayed to be more circularly symmetric compared to square QAM. With the increase of SNR, the minimum Euclidean distance between any adjacent points increases gradually, which suggests their higher AWGN tolerance.

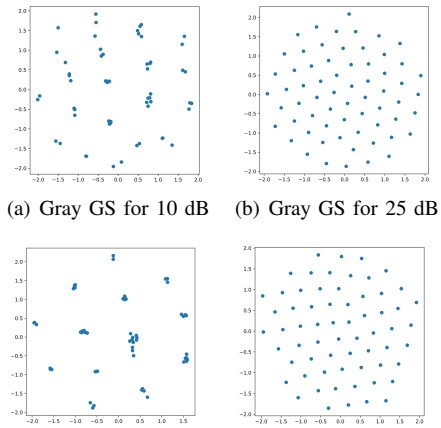
Furthermore, the proposed approach can be generalized and applied to higher-order modulations. Our simulations are



(a) SNR=5 dB (b) SNR=10 dB

(c) SNR=15 dB (d) SNR=20 dB

Fig. 4. MINE-based GS-16QAM with Gray encoding at different SNRs.



(a) Gray GS for 10 dB (b) Gray GS for 25 dB

(c) Natural GS for 10 dB (d) Natural GS for 25 dB

Fig. 5. MINE-based GS-64QAM with (a-b) Gray encoding (c-d) Nature encoding at 10 dB and 25 dB.

implemented to optimize the GS-64QAM/128QAM with two distinct encoding methods, namely Gray encoding and Natural encoding. The Gray-encoded constellations in Fig.5 (a) (b) and Natural-encoded constellations in Fig.5 (c) (d) are shown for MINE-based GS-64QAM at 10 dB and 25 dB, respectively. It is observed that the GS-64QAM constellations using the Gray encoding method is closer to circular symmetry as SNR increase, and thus Gray encoding is considered in our work for better optimization. Note that Figs.6 and 7 show the MI comparison between the MINE-based GS-64/128QAM and the unshaped 64/128QAM using Gray encoding. It can be observed that for higher-order modulations, the optimized GS-QAMs has better MI performance than the unshaped ones. Specifically, in the magnification curve at SNR = 10 dB, the MI values of GS-64QAM and GS-128QAM are 3.3365 bit/symbol and 3.3526 bit/symbol, respectively. Additionally, the proposed approach significantly reduces the gap to the Shannon limit from 0.2450 bit/symbol for GS-16QAM to 0.1074 bit/symbol for GS-128QAM, further validating the

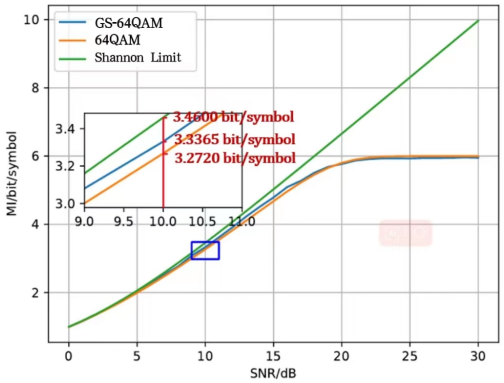


Fig. 6. MI comparison between the proposed MINE-based GS-64QAM using Gray encoding and the unshaped 64QAM in AWGN channel.

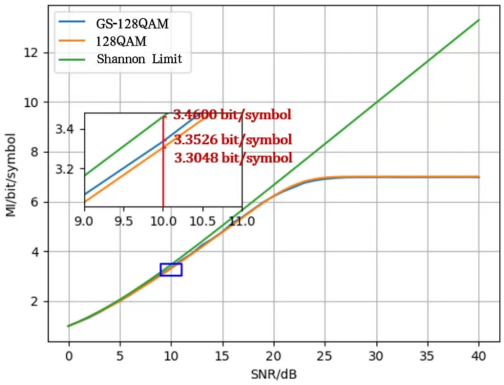


Fig. 7. MI comparison between the proposed MINE-based GS-128QAM using Gray encoding and the unshaped 128QAM in AWGN channel.

effectiveness of our design to improve the spectrum efficiency.

V. Conclusion

In this work, we have introduced a novel MINE-based GS approach to optimize the high-order constellations in the AWGN channel. To evaluate the performance of the proposed approach, we have conducted numerical comparisons of the MI between the GS-QAM and the unshaped QAM. The results demonstrate that the proposed scheme asymptotically approaches the AWGN capacity and outperforms the regular M-QAMs in terms of MI under a wide range of SNR. Note that the capacity gain exhibits a slight increase as the order M of the constellations grows. In future, our approach is expected to be applied to various channel models. And higher-order modulation formats will be investigated to improve the spectrum efficiency.

References

[1] Z. Chen, Z. He, A. Mirani, J. Schröder, P. Andrekson, M. Karlsson, M. Xiang, Y. Yu, M. Tang, Y. Qin, and S. Fu, "Transmitter optimization for PS-QAM signal in high spectral efficiency metro-transmission," *Journal of Lightwave Technology*, vol. 41, no. 9, pp. 2736–2746, Jan. 2023.

[2] W. Liu, T. Yang, X. Chen, and J. You, "Low-complexity frequency offset estimation for probabilistically shaped MQAM coherent optical systems," *IEEE Photonics Journal*, vol. 14, no. 4, pp. 1–11, Jul. 2022.

[3] Z. G. Khaki and G. Qazi, "Geometric constellation shaped non-linearity tolerant optical communication system," in *2021 IEEE 4th International Conference on Computing, Power and Communication Technologies*, Sep. 2021, pp. 1–6.

[4] O. Jovanovic, F. Da Ros, D. Zibar, and M. P. Yankov, "Geometric constellation shaping for fiber-optic channels via end-to-end learning," *Journal of Lightwave Technology*, vol. 41, no. 12, pp. 3726–3736, 2023.

[5] Z. Chen, J. Lu, S. Fu, M. Tang, D. Liu, and C. Lu, "Blind shaping rate identification for probabilistic shaping quadrature amplitude modulation formats," in *2020 Asia Communications and Photonics Conference and International Conference on Information Photonics and Optical Communications*, Oct. 2020, pp. 1–3.

[6] E. Sillekens, G. Liga, D. Lavery, P. Bayvel, and R. I. Killey, "High-cardinality geometrical constellation shaping for the nonlinear fibre channel," *Journal of Lightwave Technology*, vol. 40, no. 19, pp. 6374–6387, Oct. 2022.

[7] H. Dzieciol, G. Liga, E. Sillekens, P. Bayvel, and D. Lavery, "Geometric shaping of 2-D constellations in the presence of laser phase noise," *Journal of Lightwave Technology*, vol. 39, no. 2, pp. 481–490, Jan. 2021.

[8] G. Forney, R. Gallager, G. Lang, F. Longstaff, and S. Qureshi, "Efficient modulation for band-limited channels," *IEEE Journal on Selected Areas in Communications*, vol. 2, no. 5, pp. 632–647, Sep. 1984.

[9] B. Chen, W. Ling, Y. C. Gültekin, Y. Lei, C. Okonkwo, and A. Alvarado, "Low-complexity geometrical shaping for 4D modulation formats via amplitude coding," *IEEE Photonics Technology Letters*, vol. 33, no. 24, pp. 1419–1422, Dec. 2021.

[10] J. Ding, J. Zhang, Y. Wei, F. Zhao, C. Li, and J. Yu, "Comparison of geometrically shaped 32-QAM and probabilistically shaped 32-QAM in a bandwidth-limited IM-DD system," *Journal of Lightwave Technology*, vol. 38, no. 16, pp. 4352–4358, Aug. 2020.

[11] K. Gümüş, A. Alvarado, B. Chen, C. Häger, and E. Agrell, "End-to-end learning of geometrical shaping maximizing generalized mutual information," in *2020 Optical Fiber Communications Conference and Exhibition*, Mar. 2020, pp. 1–3.

[12] B. Wiens and D. C. Lee, "Constellation design with equal-probability partition of a cropped gaussian distribution," in *2020 IEEE 92nd Vehicular Technology Conference*, Dec. 2020, pp. 1–5.

[13] P. K. Singya, P. Shaik, N. Kumar, V. Bhatia, and M.-S. Alouini, "A survey on higher-order QAM constellations: Technical challenges, recent advances, and future trends," *IEEE Open Journal of the Communications Society*, vol. 2, pp. 617–655, Mar. 2021.

[14] S. Zhang and F. Yaman, "Design and comparison of advanced modulation formats based on generalized mutual information," *Journal of Lightwave Technology*, vol. 36, no. 2, pp. 416–423, Jan. 2018.

[15] J. B. Kinney and G. S. Atwal, "Equitability, mutual information, and the maximal information coefficient," *Proceedings of the National Academy of Sciences*, vol. 111, no. 9, pp. 3354–3359, Feb. 2014.

[16] L. Paninski, "Estimation of entropy and mutual information," *Neural Computation*, vol. 15, no. 6, pp. 1191–1253, 2003.

[17] T. O'Shea and J. Hoydis, "An introduction to deep learning for the physical layer," *IEEE Transactions on Cognitive Communications and Networking*, vol. 3, no. 4, pp. 563–575, Dec. 2017.

[18] T. Han, Q. Yang, Z. Shi, S. He, and Z. Zhang, "Semantic-preserved communication system for highly efficient speech transmission," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 1, pp. 245–259, Jan. 2023.

[19] O. Jovanovic, M. P. Yankov, F. Da Ros, and D. Zibar, "End-to-end learning of a constellation shape robust to channel condition uncertainties," *Journal of Lightwave Technology*, vol. 40, no. 10, pp. 3316–3324, May. 2022.

[20] V. Aref and M. Chagnon, "End-to-end learning of joint geometric and probabilistic constellation shaping," in *2022 Optical Fiber Communications Conference and Exhibition*, Mar. 2022, pp. 1–3.

[21] M. I. Belghazi, A. Baratin, S. Rajeshwar, S. Ozair, Y. Bengio, A. Courville, and D. Hjelm, "Mutual information neural estimation," in *International Conference on Machine Learning*, Jul. 2018, pp. 531–540.

[22] S. Kullback, *Information theory and statistics*. Courier Corporation, 1997.

[23] M. D. Donsker and S. S. Varadhan, "Asymptotic evaluation of certain markov process expectations for large time. iv," *Communications on Pure and Applied Mathematics*, vol. 36, no. 2, pp. 183–212, Mar. 1983.

[24] D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, Dec. 2014.