On the Joint Optimization of Signal Constellations and Bit Mappings

Sandushan Ranaweera, *Student Member, IEEE*, Beeshanga Abewardana Jayawickrama, *Senior Member, IEEE*, Ying He, *Senior Member, IEEE*, Quynh Tu Ngo, *Senior Member, IEEE*, and Ren Ping Liu, *Senior Member, IEEE*

***Abstract*—This letter presents an enhanced gradient search algorithm for jointly optimizing signal constellations and bit mappings. This approach stands in contrast to the recent trend of utilizing deep learning (DL) to model communication systems as end-to-end (E2E) systems, which often involve a large number of learnable parameters and high computational complexity during training. We enhance the efficiency of the gradient search algorithm by leveraging the symmetrical properties of the I/Q plane, particularly in higher modulation orders. The resulting constellations are evaluated in terms of bit error rate (BER) performance under additive white Gaussian noise and Rayleigh flat fading channels. Our findings indicate that the optimized constellations obtained via the enhanced gradient search algorithm outperform an attention-empowered DL-based E2E system in terms of BER across both channels, notably in higher signal-to-noise ratio regimes, without encountering error floor issues.**

***Index Terms*—modulation, signal constellations, deep learning, end-to-end learning, mathematical optimization.**

I. Introduction

The notion of signal constellation, representing a set of symbols used for encoding messages by the transmitter, is fundamental in data communications. Various modulation schemes, such as phase shift keying (PSK), quadrature amplitude modulation (QAM), and pulse amplitude modulation (PAM), have been developed to improve spectral efficiency while addressing errors arising from imperfections in communication channels. The selection of binary labeling for constellations significantly influences the probability of bit errors. Studies indicate that the optimal binary labeling for PSK, QAM, and PAM can be achieved using binary reflected Gray codes [1]. Typically for other modulation schemes where constellation points are determined by maximizing the minimum distance or other metrics, there is no straightforward way of finding the optimal bit labeling scheme. For such constellations, searching for optimal labeling of constellation points usually involves exhaustive methods [2].

Recently, there has been a notable surge in employing deep learning (DL) in communication systems. In [3], kNN clustering has been used to construct an index modulation scheme for reconfigurable intelligent surfaces. In [4], [5], neural networks (NNs) have been utilized to model transmitters, receivers, and channels in an end-to-end (E2E) manner. Some works have utilized NNs to learn the channel distortions such as fading effects and noise distributions [6]–[8]. The trained NNs are then used to assist the backpropagation of gradients to the transmitter network during the training stage of E2E systems. The authors of [9] have utilized one-hot representation for both inputs and outputs of the E2E system and trained them using the softmax loss function. The concept of using two-dimensional convolution layers for DL-based E2E systems with higher spectral efficiency has also been proposed [10]. Although DL-based systems have shown comparable performance to conventional PSK/QAM systems in lower signal-to-noise ratio (SNR) regimes, they face an error floor problem in higher SNR regimes. To reduce this effect, utilization of the convolution block attention module (CBAM) [11] in the receiver side has been proposed in [12]. Although these DL-based E2E systems have been built upon NNs, which are known as universal approximators, how well they can perform compared to a carefully hand-crafted optimization problem to obtain constellations with efficient bit mappings is relatively underexplored. Such a comparison is essential since DL-based E2E systems typically entail high computational complexity while lacking an interpretation of their operation. These characteristics may not be desirable in many scenarios in communication systems that necessitate low-latency requirements and generalization abilities [13].

In this letter, we propose an algorithm to obtain optimized constellations while enforcing efficient bit mappings. First, we construct a common cost function for both additive white Gaussian noise (AWGN) and Rayleigh flat fading channels to minimize bit error rate (BER). The constructed cost function consists of scaling factors derived from the Hamming distance between bit labelings to scale their corresponding squared Euclidean distance. Natural logarithms of these scaled squared Euclidean distances are utilized during the optimization process. In contrast to [13], which focuses on analyzing a DL-based E2E system and a gradient search algorithm to obtain only the constellation points, our work employs a DL-based E2E system and a gradient search algorithm to optimize both constellations and bit mappings simultaneously. In [14], a gradient search method is proposed to jointly optimize bit mappings and constellation points for non-coherent communications. To the best of our knowledge, this is the first theoretical analysis of utilizing a gradient search method for the joint optimization of bit mappings and constellation points in coherent communication. Furthermore, we leverage the symmetrical properties of the I/Q plane to increase the efficiency of the optimization procedure in higher modulation orders. Specifically, we learn the constellation only in one

The authors are with the School of Electrical and Data Engineering, University of Technology Sydney, Sydney, NSW 2007, Australia (e-mail: {sandushan.ranaweera, beeshanga.jayawickrama, ying.he, quynhtu.ngo, renping.liu}@uts.edu.au).

quadrant of the I/Q plane, which is then mirrored with respect to real and imaginary axes to obtain the full constellation. For comparison, we trained an attention-empowered DL-based E2E system [12] to compare with our proposed approach. Simulations have been carried out in AWGN and Rayleigh flat fading channels. The results demonstrate that the constellation obtained through our proposed approach yields superior BER performance, avoiding the error floor problem in higher SNR regimes.

II. System Model

A typical communication system consists of a transmitter, a channel, and a receiver, as shown in Fig. 1. Without loss of generality, we assume bit sequences $\boldsymbol{b} \in \{0,1\}^u$ that consist of u information bits are fed to the transmitter. The transmitter groups this information bits into message symbols consisting of $m = \log_2 M$ bits. Each of these message symbols is then mapped to a complex signal from a signal constellation $\mathcal{S} = \{s_1, \cdots, s_M\}$ where $s_i \in \mathbb{C}, \forall s_i \in \mathcal{S}$. This results in an output signal $\boldsymbol{x} \in \mathbb{C}^n$ consisting $n = \lceil \frac{u}{m} \rceil$ complex signals for each input bit sequence $\boldsymbol{b}$ where $\lceil \cdot \rceil$ denotes the ceiling operator. The receiver receives a distorted signal $\boldsymbol{y}$ through the channel which can be modeled as,

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{w}, \tag{1}$$

where $\boldsymbol{H} \in \mathbb{C}^{n \times n}$ denotes the channel coefficients matrix and $\boldsymbol{w}$ denotes additive noise introduced by the channel. In an AWGN channel, $\boldsymbol{H}$ becomes an identity matrix, while in a Rayleigh flat fading channel, it becomes a diagonal matrix with the fading coefficients along its diagonal. For both channels, we assume $\boldsymbol{w}$ to be white gaussian noise, i.e., $w_i \sim \mathcal{CN}(0, N_0)$, $\forall w_i \in \boldsymbol{w}$, where N_0 denotes the noise power, and $\mathcal{CN}(\mu, \sigma^2)$ denotes the complex normal distribution with mean μ and variance σ^2. After receiving the distorted signal $\boldsymbol{y}$, the receiver outputs an estimate $\hat{\boldsymbol{b}}$ for the transmitted bit sequence $\boldsymbol{b}$ using some decision rule. In this work, we assume perfect channel state information is available to the receiver, which enables us to use the maximum-likelihood decision rule for both AWGN and Rayleigh flat fading channels.

Fig. 1. A simple communication system.

The symbol error probability, i.e., the probability that the distortions in the communication channel shifted the transmitted symbol into the decision region of another symbol, is

$$P_s = \frac{1}{M} \sum_{i=1}^{M} Pr\left(\hat{s} \neq s_i | s_i \text{ transmitted}\right), \tag{2}$$

and the resulting BER is

$$P_b = \frac{1}{Mm} \sum_{i=1}^{M} \sum_{k=1}^{M} h(i,k) Pr\left(\hat{s} = s_k | s_i \text{ transmitted}\right), \tag{3}$$

where $h(i,k)$ is the Hamming distance between binary representations of the constellation points s_i and s_k. With the assumption of having perfect channel estimation available at the receiver, the asymptotic approximation for (3) becomes

$$P_b \approx \frac{1}{Mm} \sum_{i=1}^{M} \int h(i,\hat{k}_i) \exp\left(-\frac{\eta^2 \left|s_i - s_{\hat{k}_i}\right|^2}{8N_0}\right) p(\eta)\, d\eta, \tag{4}$$

where

$$\hat{k}_i = \underset{\substack{\forall k \in [1,M] \\ i \neq k}}{\arg\min} \left|s_i - s_k\right|^2, \tag{5}$$

$|z|$ denotes the absolute value of the complex number z, η denotes the fading coefficient, and $p(\eta)$ denotes its probability density distribution. For an AWGN channel, $p(\eta) = \delta(\eta - 1)$, where $\delta(\cdot)$ denotes Dirac-delta function, and for a Rayleigh fading channel, $p(\eta) = \frac{\eta}{\omega^2} \exp\left(-\frac{\eta^2}{2\omega^2}\right)$, where ω is the scale parameter of the corresponding Rayleigh distribution.

III. Gradient Search Algorithm for Constellation Optimization

In this section, we formulate a joint optimization problem to find an arrangement of signal points in a constellation and an effective way to assign bit labeling to those points. We begin with the following min-max problem to minimize P_b.

$$\text{Find } \mathcal{S} = \underset{\forall \boldsymbol{s} \in \mathbb{C}^M}{\arg\min} \left[\max_{\forall i \in [1,M]} \int \frac{h(i,\hat{k}_i) p(\eta)}{\exp\left(\frac{\eta^2}{8N_0} \left|s_i - s_{\hat{k}_i}\right|^2\right)} d\eta \right] \tag{6a}$$

$$\text{s.t. } \frac{1}{M} \sum_{i=1}^{M} |s_i|^2 = 1. \tag{6b}$$

By noting that $\exp(x)$ is a strictly increasing function, we can get an equivalent min-max problem as below.

$$\text{Find } \mathcal{S} = \underset{\forall \boldsymbol{s} \in \mathbb{C}^M}{\arg\min} \left[\max_{\forall i \in [1,M]} \frac{h(i,\hat{k}_i)}{\frac{1}{8N_0} \left|s_i - s_{\hat{k}_i}\right|^2} \int \frac{1}{\eta^2} p(\eta)\, d\eta \right] \tag{7a}$$

$$\text{s.t. } \frac{1}{M} \sum_{i=1}^{M} |s_i|^2 = 1. \tag{7b}$$

Note that we take out the terms that are independent of η out of the integral. It is straightforward to see that this is equivalent to the below max-min problem.

$$\text{Find } \mathcal{S} = \underset{\forall \boldsymbol{s} \in \mathbb{C}^M}{\arg\max} \left[\min_{\forall i \in [1,M]} \frac{\left|s_i - s_{\hat{k}_i}\right|^2}{h(i,\hat{k}_i)} \right] \tag{8a}$$

$$\text{s.t. } \frac{1}{M} \sum_{i=1}^{M} |s_i|^2 = 1. \tag{8b}$$

Interestingly, the terms inside the minimization consist of Euclidean distances between constellation points scaled by the Hamming distances between their corresponding bit labelings.

We obtain the simplified formulation below by relaxing the minimization.

$$\text{Find } \mathcal{S} = \underset{\forall \boldsymbol{s} \in \mathbb{C}^M}{\arg\max} \left[\min_{\substack{\forall i,k \in [1,M] \\ i \neq k}} \frac{|s_i - s_k|^2}{d(i,k)} \right] \tag{9a}$$

$$\text{s.t. } \frac{1}{M} \sum_{i=1}^{M} |s_i|^2 = 1, \tag{9b}$$

where $d(i,k) = \min(2, h(i,k))$. The scaling factors $d(i,k)$ compel the optimized constellation to increase the distances between points with bit labeling differing by more than 1 bit. Intuitively, the optimization process is forced to arrange constellation points such that adjacent symbols differ by only one bit in their labels.

While the Euclidean distance between two points is well known as a convex function, it's important to note that the minimum of a set of convex functions, which constitutes the objective function in this maximization problem, may not always be either convex or concave. Moreover, this optimization problem does not have a unique optimal solution since a rotation of any optimal solution results in another feasible optimal solution. On the other hand, the gradient of the squared Euclidean distance is an increasing function, which is not a desirable property for a maximization problem, particularly with gradient-based methods. So, we modify the formulation of the optimization problem (9) by taking the natural logarithm of the scaled squared Euclidean distance to yield:

$$\max_{\boldsymbol{x} \in \mathbb{R}^{2M}} \quad f(\boldsymbol{x}) \tag{10a}$$

$$\text{s.t. } \frac{1}{M} \sum_{i=1}^{2M} x_i^2 = 1, \tag{10b}$$

$$f(\boldsymbol{x}) = \min_{\substack{\forall i,k \in [1,M] \\ i \neq k}} \ln \left(\frac{|s_i - s_k|^2}{d(i,k)} \right), \tag{10c}$$

$$\boldsymbol{x} = \langle \Re(s_1), \Im(s_1), \cdots, \Re(s_M), \Im(s_M) \rangle, \tag{10d}$$

where $\Re(z)$ and $\Im(z)$ denotes the real and imaginary parts of the complex number z respectively. Specifically, we have $x_{2i-1} = \Re(s_i)$ and $x_{2i} = \Im(s_i)$. The cost function of this optimization problem is differentiable concerning x_is, and its partial derivatives can be expressed as

$$\frac{\partial f(\boldsymbol{x})}{\partial x_{2\hat{i}}} = \frac{2d(\hat{i},\hat{k})}{\left|s_{\hat{i}} - s_{\hat{k}}\right|^2} \left(x_{2\hat{i}} - x_{2\hat{k}} \right), \tag{11a}$$

$$\frac{\partial f(\boldsymbol{x})}{\partial x_{2\hat{i}-1}} = \frac{2d(\hat{i},\hat{k})}{\left|s_{\hat{i}} - s_{\hat{k}}\right|^2} \left(x_{2\hat{i}-1} - x_{2\hat{k}-1} \right), \tag{11b}$$

$$\hat{i}, \hat{k} = \underset{\substack{\forall i,k \in [1,M] \\ i \neq k}}{\arg\min} \frac{|s_i - s_k|^2}{d(i,k)}. \tag{11c}$$

The partial derivatives respect to $x_{2\hat{k}}$ and $x_{2\hat{k}-1}$ will be same as $\frac{\partial f(\boldsymbol{x})}{\partial x_{2\hat{i}}}$ and $\frac{\partial f(\boldsymbol{x})}{\partial x_{2\hat{i}-1}}$ except an additional negative sign. The partial derivative will be zero for all other x_is. The natural logarithm function is an increasing concave function that will give smaller gradients when the scaled squared Euclidean distance is high. Intuitively, this is a desirable property for maximizing the objective function with gradient-based methods.

Algorithm 1 Constellation finding algorithm.

Input: $\boldsymbol{x}^{(0)} \in \mathbb{R}^{2M}$, $\gamma > 0$, $\boldsymbol{x}^* = \boldsymbol{x}^{(0)}$, $t = 0$, $\mathcal{T} > 0$
Output: $\boldsymbol{x}^*$

1: **repeat**
2: $\quad \boldsymbol{x}^{(t+1)} = \boldsymbol{x}^{(t)} + \gamma \nabla f\left(\boldsymbol{x}^{(t)}\right)$;
3: $\quad \boldsymbol{x}^{(t+1)} \leftarrow \frac{\sqrt{M}\boldsymbol{x}^{(t+1)}}{\left\|\boldsymbol{x}^{(t+1)}\right\|}$;
4: $\quad$ **if** $f\left(\boldsymbol{x}^{(t+1)}\right) \geq f(\boldsymbol{x}^*)$ **then**
5: $\quad\quad \boldsymbol{x}^* \leftarrow \boldsymbol{x}^{(t+1)}$;
6: $\quad\quad t \leftarrow t + 1$;
7: $\quad$ **end if**
8: **until** $t \geq \mathcal{T}$

The Algorithm 1 is used to solve the optimization problem (10). We let the binary representation of $i-1$ to be the bit labeling of s_i. We perform ℓ^2-normalization (Algorithm 1, step 3) after the gradient update (Algorithm 1, step 2) to ensure that the resulting constellation has the unit average energy as constrained in (10b). The basic mathematical optimization problem (10) can be efficiently solved for lower modulation orders, i.e., $M \leq 8$. However, the dimensionality of the problem is increasing rapidly with the modulation order. To increase the efficiency in higher modulation orders, we utilize the symmetrical properties of the I/Q plane. Specifically, we only optimize constellation points in the first quadrant of the I/Q plane and take the mirror images with respect to real and imaginary axes to obtain the full constellation for $M > 8$. We get the optimization problem

$$\max_{\boldsymbol{x} \in \mathbb{R}^{\frac{M}{2}}} \min \{ g_1(\boldsymbol{x}), g_2(\boldsymbol{x}) \} \tag{12a}$$

$$\text{s.t. } \frac{4}{M} \sum_{i=1}^{\frac{M}{2}} x_i^2 = 1, \tag{12b}$$

$$g_1(\boldsymbol{x}) = \min_{i \in \left[1, \frac{M}{2}\right]} \ln\left(4x_i^2\right), \tag{12c}$$

$$g_2(\boldsymbol{x}) = \min_{\substack{\forall i,k \in \left[1, \frac{M}{4}\right] \\ i \neq k}} \ln \left(\frac{|s_i - s_k|^2}{d(i,k)} \right), \tag{12d}$$

$$\boldsymbol{x} = \left\langle \Re(s_1), \Im(s_1), \cdots, \Re\left(s_{\frac{M}{4}}\right), \Im\left(s_{\frac{M}{4}}\right) \right\rangle, \tag{12e}$$

by introducing an additional cost function for this purpose. The function $g_1(\boldsymbol{x})$ ensures that the minimum distance is maintained even after we took mirror images to obtain the final constellation while $g_2(\boldsymbol{x})$ ensures the minimum distance is maintained between the constellation points in the first quadrant. To maximize this modified cost function, Algorithm 1 is utilized with $f(\boldsymbol{x}) = \min\{g_1(\boldsymbol{x}), g_2(\boldsymbol{x})\}$ and substituting $\frac{M}{4}$ for M. Note that we take the binary representation of $i-1$ as the bit labeling of s_i. Hence, after taking mirror images, we only have to change the first two bits when obtaining the bit labeling for the full constellation.

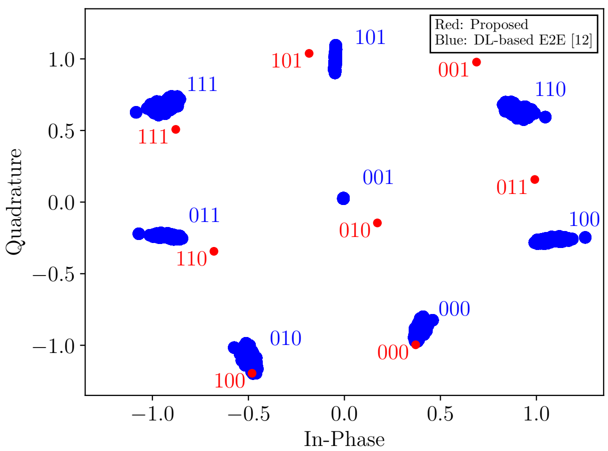


Fig. 2. Resulting constellations for $M = 8$.

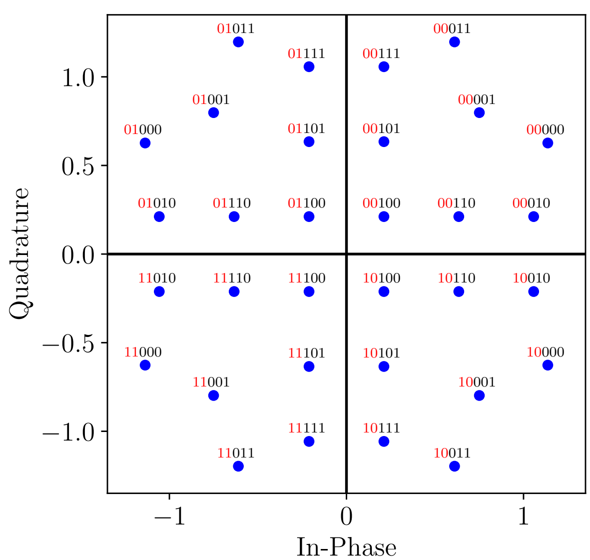


Fig. 3. Resulting constellation for $M = 32$. The first two bits of each label are in red font. We change these bits when obtaining the bit mapping for the full constellation.

IV. Simulation Results

A. Simulation Setup

We obtained constellations via the optimization procedure described in III and trained an E2E system based on [12] for modulation schemes with different M values. Specifically, we use (10) for $M \leq 8$ and (12) for $M > 8$ as the cost function when using Algorithm 1. In our simulation, we let $\gamma = 10^{-4}$, $\mathcal{T} = 10^5$ for $M \leq 8$ and $\gamma = 10^{-6}$, $\mathcal{T} = 10^6$ for $M > 8$ in Algorithm 1. Moreover, we utilize scaled and shifted versions of learned constellations for $\frac{M}{4}$ as initial points $\boldsymbol{x}^{(0)}$ when $M > 8$. A comparison of the number of learnable parameters of the two approaches is given in Table I. It can be observed that the proposed gradient search algorithm has much fewer learnable parameters than the DL-based E2E approach. This characteristic of the gradient search algorithm makes it more computationally efficient than the DL-based E2E approach. It is worth noting that the E2E system is trained separately for AWGN and Rayleigh channels, while the maximization objective for obtaining the constellation is independent of the channel model.

TABLE I
Number of learnable parameters.

m	M	DL-based E2E [12]			Proposed
		Tx	Rx	Total	
2	4	33922	85512	119434	8
3	8	34050	85641	119691	16
4	16	34178	85770	119948	8
5	32	34306	85899	120205	16
6	64	34434	86028	120462	32
7	128	34562	86157	120719	64
8	256	34690	86286	120976	128

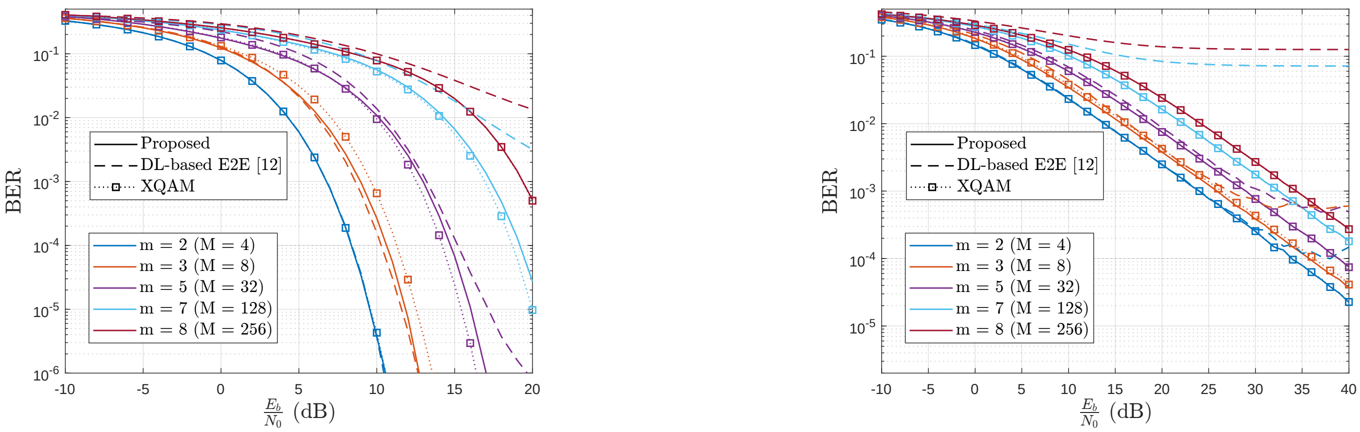
B. Signal Constellations

Fig. 2 shows the resulting constellations for both approaches for $M = 8$. The red color points and bit labelings correspond to the constellations obtained by solving the optimization problem (10). It can be observed that there is only a one-bit difference in the bit labelings of the symbols that have minimum distance separation. The blue color points and bit labeling correspond to the resulting constellations of the output of the transmitter network in the E2E system for $M = 8$. The E2E system is trained for an AWGN channel, to which we give randomly generated 128 batches of bit sequences as input. Each of these batches contains bit sequences corresponding to 128 3-bit symbols. The distortion in this constellation points is due to the operation of the normalization layer. This distortion restricts the performance of the E2E system in higher modulation orders where the separation between constellation points is low.

Fig. 3 shows the constellation obtained via the proposed gradient search algorithm for $M = 32$. In this scenario, we only learn the constellation in the first quadrant using the modified cost function in (12). The bit mapping for the full constellation can easily be obtained by changing the first two bits, i.e., 01, 11, and 10 for the second, third, and fourth quadrants. It is observed that two symbols that have the minimum distance across two quadrants are guaranteed to have bit labeling that only differs from one bit. The only difference in the bit labeling is due to the change of the first two bits, which we changed to get the bit mapping for the full constellation. All other bits in the bit labeling will be the same for any two of such constellation points.

C. BER Performance

Fig. 4a shows the BER comparison in an AWGN channel. The E2E system and the constellation obtained via the proposed approach outperform the XQAM constellation scheme for $M = 8$. For all other M, the constellation obtained via the gradient search algorithm has a closer performance to the XQAM constellation than the E2E system. Moreover, the E2E system starts to show an error floor behavior even in relatively



(a) AWGN channel.

(b) Rayleigh Channel.

Fig. 4. BER comparison in AWGN and Rayleigh flat fading channels. *Solid lines* correspond to constellations obtained via the proposed optimization approach, *dashed lines* corresponds to attention empowered DL-based E2E system [12] and *dotted lines* corresponds to equivalent XQAM constellation.

lower SNRs when we increase the modulation order. Fig. 4b shows the BER comparison in a Rayleigh channel. A similar behavior can be observed again as in the AWGN channel. However, in the Rayleigh channel, the error floor behavior of the E2E system can be observed more clearly than in the AWGN channel.

V. Conclusion

We presented an enhanced gradient search algorithm for obtaining signal constellations with efficient bit mappings. Furthermore, we leveraged the symmetrical properties of the I/Q plane to increase the efficiency of the optimization procedure for higher modulation orders. The BER performance of the signal constellation obtained by the proposed algorithm is compared with that of an XQAM constellation and an attention-empowered DL-based E2E communication system. Comparisons have been carried out under AWGN and Rayleigh flat fading channels. The constellation obtained via the proposed algorithm gave better BER performance in both channels without showing error floor behavior, as observed with the DL-based E2E system.

References

[1] E. Agrell, J. Lassing, E. G. Ström, and T. Ottosson, "On the optimality of the binary reflected gray code," *IEEE Trans. Inf. Theory*, vol. 50, pp. 3170–3182, 2004.

[2] 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 J. Commun. Soc.*, vol. 2, pp. 617–655, 2021.

[3] H. Liu, J. An, W. Xu, X. Jia, L. Gan, and C. Yuen, "K-means based constellation optimization for index modulated reconfigurable intelligent surfaces," *IEEE Commun. Lett.*, vol. 27, no. 8, pp. 2152–2156, 2023.

[4] T. O'Shea and J. Hoydis, "An introduction to deep learning for the physical layer," *IEEE Trans. Cogn. Commun. Netw.*, vol. 3, pp. 563–575, 2017.

[5] S. Dorner, S. Cammerer, J. Hoydis, and S. T. Brink, "Deep learning based communication over the air," *IEEE J. Sel. Top. Signal Process.*, vol. 12, pp. 132–143, 2018.

[6] H. Ye, L. Liang, G. Y. Li, and B. H. Juang, "Deep learning-based end-to-end wireless communication systems with conditional GANs as unknown channels," *IEEE Trans. Wirel. Commun.*, vol. 19, pp. 3133–3143, 2020.

[7] Y. An, S. Wang, L. Zhao, Z. Ji, and I. Ganchev, "A learning-based end-to-end wireless communication system utilizing a deep neural network channel module," *IEEE Access*, vol. 11, pp. 17 441–17 453, 2023.

[8] D. Yan, M. Jia, Q. Guo, and X. Gu, "Unknown channel end-to-end learning of communication system with residual DCGAN," *Proc. IEEE International Symposium on BMSB*, IEEE Computer Society, Jun. 2023.

[9] N. Wu, X. Wang, B. Lin, and K. Zhang, "A CNN-based end-to-end learning framework toward intelligent communication systems," *IEEE Access*, vol. 7, pp. 110 197–110 204, 2019.

[10] Q. Liu, J. Liu, G. Yu, and X. Zhou, "Research on end-to-end wireless communication system based on two dimensional convolution," *Proc. ICCT*, 2022.

[11] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," *Proc. ECCV*, Cham: Springer International Publishing, 2018.

[12] M. Lu, B. Zhou, and Z. Bu, "Attention-empowered residual autoencoder for end-to-end communication systems," *IEEE Commun. Lett.*, vol. 27, pp. 1140–1144, 2023.

[13] J. Liu, H. Zhao, D. Ma, K. Mei, and J. Wei, "Theoretical analysis of deep neural networks in physical layer communication," *IEEE Trans. Commun.*, vol. 70, no. 10, pp. 6589–6603, 2022.

[14] D. Cuevas, J. Álvarez-Vizoso, C. Beltrán, I. Santamaria, V. Tuček, and G. Peters, "Union bound minimization approach for designing grassmannian constellations," *IEEE Trans. Commun.*, vol. 71, no. 4, pp. 1940–1952, 2023.