

Optimization of the Non-Linearity Tolerant 4D Geometric Shaped Constellations for Optical Fiber Communication Systems Using Neural Networks

Amirhosein Soleimanzade , Masoud Ardakani , *Senior Member, IEEE*, and Hamid Ebrahimzad , *Member, IEEE*

Abstract—We propose a novel 4D geometric shaping (GS) method based on multilevel coding (MLC) for optical fiber communication systems that mitigates fiber non-linearity. To design non-linearity tolerant 4D constellations, non-linear interference noise (NLIN) and modulation dependency behavior of the fiber are considered during the optimization process. By doing so, the complexity of the 4D GS problem increases extremely, requiring novel optimization approaches to tackle this challenge. Hence, we build a novel deep neural network structure that can efficiently optimize the position of the 4D points. Moreover, we show that optimizing the geometry based on the well-known generalized mutual information rate is not efficient and results in a significant gap to the mutual information rate. To solve this issue, we take advantage of the concept of achievable information rates (AIRs) and multilevel coding. By optimizing the 4D constellations based on the proposed AIRs, our GS scheme provides higher rates than the traditional probabilistic shaping and the recent non-linearity tolerant GS methods.

Index Terms—4D geometric shaping, achievable information rate, multilevel coding, neural network-based optimization, non-linear fiber channel, optical fiber communication.

I. INTRODUCTION

THE exponential growth of Internet usage necessitates employing advanced techniques to increase the transmission rate of optical fiber communication systems. Since most of the fiber bandwidth is in use [1], [2], increasing spectral efficiency is the only way to improve the achievable rate. Among many techniques, in recent years, constellation shaping methods have attracted a lot of interest as they lead to a significant increase in the spectral efficiency [3], [4], [5], [6].

Constellation shaping aims at finding energy-efficient constellations that achieve a better performance compared to the traditional quadrature amplitude modulations (QAMs). In other

words, the goal of constellation shaping is to design constellations that mimic the input distribution that results in the channel capacity. Constellation shaping approaches can be divided into two categories: geometric shaping (GS) and probabilistic shaping (PS). In GS, the positions of the equiprobable constellation points are chosen such that the achievable information rate (AIR) is improved. PS, on the other hand, increases AIR by changing the uniform distribution of the constellation points into a non-uniform distribution.

PS transmission systems are typically based on the probabilistic amplitude shaping (PAS) [7] architecture, and in general, they are capable of providing higher AIRs compared to GS [8]. However, the PAS architecture needs a distribution matcher (DM) in order to convert uniform data into non-uniform data, which increases the complexity of the transmitter and receiver [9]. In addition, the PAS architecture suffers a rate loss due to using finite blocklengths at DMs [10].

In GS systems, the main implementation challenge is that they require high-resolution digital-to-analog and analog-to-digital converters, but they are generally less complex than PS systems [11]. In the case of non-linear fiber channels, another advantage of GS over PS is that GS is generally more non-linearity tolerant [11], [12], [13]. As a consequence, GS has attracted a lot of interest in recent years [5], [6], [11], [12], [14], [15], [16], [17], [18], [19], [20], [21], [22]. Despite all these benefits, in GS, there is a significant gap between the mutual information (MI) and generalized mutual information (GMI) rate due to the non-Gray mapping [11], [14]. Hence, improving the achievable rate of GS can be highly beneficial for fiber communications, and that is the goal of this work.

GS can be viewed as an optimization problem, and by optimizing its decision variables, the optimal position of the constellation points is obtained. To solve the GS optimization problem, an objective function, based on which the geometry is optimized, must be defined. In the literature, a wide variety of objective functions or criteria is considered to solve the GS problem. In most cases, the considered criterion falls into one of the following categories: 1) maximizing AIR and 2) minimizing the error rate. For instance, maximizing the MI or GMI rate falls into the first category while minimizing the symbol error rate (SER) or bit error rate (BER) falls into the second one. In [14] and [15], to solve the GS problem, the pairwise optimization (PO) [23] algorithm is employed to maximize the MI and GMI rate of 2D 32 and 64-point constellations, respectively. Also, in [5], the

Manuscript received 29 December 2022; revised 6 June 2023 and 4 August 2023; accepted 13 September 2023. Date of publication 19 September 2023; date of current version 16 January 2024. This work was supported by Huawei Technologies Canada. (Corresponding author: Amirhosein Soleimanzade.)

Amirhosein Soleimanzade and Masoud Ardakani are with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6G 2R3, Canada (e-mail: a.soleimanzade@ualberta.ca; ardakani@ualberta.ca).

Hamid Ebrahimzad is with the Huawei Technologies Canada Company, Ltd., Ottawa, ON N2L 0A4, Canada (e-mail: hamid.ebrahimzad@huawei.com).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/JLT.2023.3317149>.

Digital Object Identifier 10.1109/JLT.2023.3317149

particle swarm optimization (PSO) [24] is used to maximize the MI rate of 2D amplitude-phase shift keying (APSK) modulations that are non-linearity tolerant. In addition, in recent years, deep learning-based algorithms have attracted a lot of interest in solving the GS problem. For instance, [25] uses autoencoders (AE) to minimize the SER of 2D 16-point and 256-point constellations for zero-dispersion fibers. Also, in [26], deep AEs are employed to minimize the BER of 2D constellations. Moreover, recently, 4D GS has become attractive due to its potential for designing constellations that are much more capable of mitigating fiber non-linear effects compared to 2D constellations. In [12], [19], [20], [21], the constant-modulus constraint is considered to design 4D non-linearity tolerant constellations. Also, [13] uses the orthant-symmetry (OS) constraint to optimize 4D constellations with a spectral efficiency (SE) of 7 bits/4D-symbol. [27] maximizes the GMI rate of the orthant-symmetric non-linearity tolerant 4D constellations for a single-channel and single-span system using an end-to-end AE-learning scheme. [28] and [29] consider OS as well as the X-polarization-Y-polarization symmetry constraint to maximize the MI and GMI rate of the 4D non-linearity tolerant constellations using the pattern search optimizer [30].

The mentioned GS algorithms outperform the traditional QAM constellations substantially; however, due to some major drawbacks and limitations, they are not capable of providing AIRs as high as PS. The drawbacks of the previous studies can be organized as follows: 1) The 2D non-linearity tolerant GS methods such as [5] cannot design constellations as non-linearity tolerant as 4D ones due to the limitation of the search space. By switching to the 4D search space, one can assume a correlation between the X and Y-polarization and design modulations that mitigate the fiber non-linearity and non-linear interference noise (NLIN) more effectively. 2) The recent 4D non-linearity tolerant GS methods consider the constant-modulus constraint in order to combat fiber non-linearity. However, the constant-modulus constraint limits the 4D search space severely, which results in a considerable performance loss. 3) The recent 4D GS methods optimize the geometry by maximizing the GMI rate; however, the gap between the GMI and MI rate of the shaped constellations is significant, meaning that the GMI-based GS is not an efficient solution. By resolving these problems, GS could potentially provide substantially higher information rates and possibly achieve a performance comparable to PS. In this study, we propose a novel non-linearity tolerant 4D GS method based on multilevel coding (MLC) [31]. In [32], MLC is employed to considerably improve the performance of geometrically shaped constellations; however, there are substantial differences between [32] and our work: 1) Unlike our scheme, [32] proposes a 2D GS method, where constellations with much larger sizes (such as 2D 256-point GS constellations) must be considered so that the shaping gain becomes significant, and for small-to-medium size 2D constellations (such as 2D 16-point constellations), method of [32] cannot provide large gains; however, by optimizing constellations in the 4D space (which is the case in this study), small-to medium-size geometrically shaped 4D constellations can provide much higher gains. Moreover, since [32] can provide considerable gains only for large constellations, it is not proper for today's long-haul fiber optic communication systems

as these systems employ small-to-medium size constellations. On the other hand, because our method provides significant gains even for small constellations, it is much more appropriate for long-haul communication applications. 2) The second difference is in the employed objective function. [32] originally wants to use BER as the objective function of most reliable bits (MRBs); however, since BER is not a differentiable function, [32] proposes an upper bound for the BER of MRBs. This scheme may result in suboptimal constellations under certain channel conditions. On the other hand, in our study, the objective function is AIR, which is a differentiable function and perfectly represents the amounts of information that can be transmitted over a constellation. Also, in the objective function of [32], there is a parameter called α that controls the balance of the least reliable bits (LRBs) and MRBs' objective functions. [32] discusses that under different channel conditions, different α should be considered. As a result, a constellation optimized for one signal-to-noise ratio (SNR) may not be useful at other SNRs. In other words, it seems that the use of α has reduced the robustness of the performance to SNR change.

Despite the benefits that MLC provides, MLC results in a higher implementation cost compared to bit-interleaved coded modulation. Also, since MLC needs multiple stages of coding, it results in higher latency. In this work, we show that by using MLC, the gap between the GMI and the MI rate is almost completely vanished, and ultimately, our GS method outperforms the traditional PS in terms of the gap to the Shannon limit. By traditional PS, we mean a PS system that employs the well-known PAS architecture and an ideal DM to perfectly convert a uniform distribution into the corresponding optimal Maxwell-Boltzmann (MB) distribution. Please note that there exist advanced non-linearity tolerant PS schemes such as [33], [34], [35], [36] that offer better performance than the traditional PS schemes. For the MLC, we show that even a 2-level code is sufficient to surpass the traditional PS. Moreover, our design is not limited to the constant-modulus constraint. To design non-linearity tolerant 4D constellations, we use the channel model of [37] during the optimization process to consider the fiber modulation-dependency behavior for constellations that transmit correlated data over the X and Y-polarization. Considering that the GS problem is non-convex [8], solving the 4D GS problem becomes extremely complicated, especially for modulation-dependent fiber. Hence, to reduce the complexity, we consider the OS constraint on the geometry of 4D constellations. In addition, we employ deep neural networks (NNs) to optimize the position of the 4D points. Of course, neural networks do not necessarily guarantee the OS constraint. Here, we propose a novel NN-based trainable system that optimizes 4D GS constellations that are both OS-compatible and non-linearity tolerant. By employing our method, we shape 128 and 256-point 4D constellations for the fiber channel. Our 128 and 256-point GS constellations outperform the corresponding polarization-multiplexed (PM) QAM and the traditional probabilistic-shaped PM-QAM (PM-PS-QAM) constellations significantly. Also, it is observed that our 4D GS constellations introduce less NLIN to the optical fiber system than PM-QAM, PM-PS-QAM, and the recent non-linearity tolerant GS methods.

The remainder of this article is organized as follows: in Section II, the fundamentals of geometric shaping are reviewed and the 4D geometric shaping problem is defined. In Section III, the fiber channel model that is used for modeling the NLIN of 4D constellations is briefly discussed, and then, we propose our novel NN-based trainable system for optimizing the non-linearity tolerant 4D constellations. Moreover, we discuss our MLC technique that is utilized for reducing the gap between GMI and MI of the 4D constellations. In Section IV, we use our method to optimize the geometry of 4D constellations for the additive white Gaussian noise (AWGN) and the non-linear fiber channel, and the results are analyzed and discussed. Finally, Section V concludes this work.

II. FUNDAMENTALS OF GEOMETRIC SHAPING AND PROBLEM DEFINITION

In this section, first, we review the achievable information rate (AIR) analysis. We discuss mutual information (MI) and generalized mutual information (GMI) as two basic AIRs. Moreover, in the second subsection, the problem of GS is defined.

A. Achievable Information Rate Analysis

To solve the GS problem, an objective function or metric based on which the geometry of the constellation is optimized must be defined. AIR-based metrics such as MI and GMI are widely used for GS [3], [5], [13]. To compute AIR, the channel transition probability, $f_{Y|X}(y|x)$, must be known, where X and Y are the channel input and output, respectively. In the case of the non-linear fiber channel, a closed-form equation for the channel transition probability does not exist. However, by taking advantage of the mismatched decoding technique [38], [39], we can evaluate the fiber channel AIR using an auxiliary channel transition probability $q_{Y|X}(y|x)$ instead of the unknown $f_{Y|X}(y|x)$. Throughout this study, we consider an AWGN auxiliary channel, i.e., $q_{Y|X}(y|x)$ is computed as follows:

$$q_{Y|X}(y|x) = \frac{1}{(2\pi\sigma^2/n)^{\frac{n}{2}}} e^{-\frac{\|y-x\|_2^2}{2\sigma^2/n}}, \quad (1)$$

where n is the dimensionality, σ^2 is the total noise variance of the AWGN auxiliary channel over the n dimensions, x is the channel input, and $x \in \chi$, where $\chi = \{x_1, x_2, \dots, x_M\}$ is the set of n -dimensional constellation points, and y is the channel output. In this study, since the goal is to design 4D constellations, we have $n = 4$. Also, by employing the auxiliary channel transition probability of (1), a lower bound for the actual AIR of the optical fiber system is achieved [40], and using this lower bound, we measure the system AIR. Considering the auxiliary channel transition probability of (1), we proceed with our discussion on AIR analysis.

Assuming an optimal decoding scheme at the receiver side, the decoder uses the channel transition probability, which is a symbol-wise metric, to achieve the MI rate defined as:

$$\mathbb{I}(X; Y) = \mathbb{H}(X) - \mathbb{H}(X|Y), \quad (2)$$

where $\mathbb{I}(\cdot)$ and $\mathbb{H}(\cdot)$ denote mutual information and entropy, respectively. In real-world systems, we desire to achieve the rate

of (2) using binary decoders rather than symbol-wise decoders. By taking advantage of the chain rule of MI, the MI rate can be achieved using binary decoding schemes as follows: the symbol $X \in \chi$ can be represented by $\mathbf{B} = (B_1, \dots, B_m)$ which is the m bit levels of symbol X , where $M = 2^m$. Considering this representation, (2) is equivalent to $\mathbb{I}(\mathbf{B}; Y)$. Hence, using the chain rule, we have:

$$\begin{aligned} \mathbb{I}(X; Y) &= \mathbb{I}(\mathbf{B}; Y) = \mathbb{I}(B_1; Y) + \mathbb{I}(B_2; Y|B_1) \\ &+ \mathbb{I}(B_3; Y|B_1, B_2) + \dots + \mathbb{I}(B_m; Y|B_1, B_2, \dots, B_{m-1}). \end{aligned} \quad (3)$$

This can be implemented using multilevel coding (MLC) [31] techniques, where on each bit level, a different forward error correction (FEC) code is applied. In addition, one can remove the conditions in (3) in order to obtain a simpler approach to achieve the MI rate. By doing so and knowing that conditioning increases information, we have:

$$\mathbb{I}(\mathbf{B}; Y) \geq \sum_{i=1}^m \mathbb{I}(B_i; Y) = \mathbb{H}(\mathbf{B}) - \sum_{i=1}^m \mathbb{H}(B_i|Y). \quad (4)$$

Note that the derivation on the right-hand side of (4) is valid if the bit levels are independent [3], which is the case in this study. In (4), the expression $\mathbb{H}(\mathbf{B}) - \sum_{i=1}^m \mathbb{H}(B_i|Y)$ is known as the generalized mutual information (GMI), which is the proper AIR for the bit-interleaved coded modulation (BICM) systems [41]. Based on (4), the GMI rate is always lower or equal to the MI rate. For QAM constellations, the gap between MI and GMI is negligible [3], [7]. However, for geometrically shaped constellations, the gap can increase significantly due to the non-Gray mapping [11], [14], meaning that, in order to have a small gap to MI, a more efficient AIR compared to GMI must be considered for geometrically shaped constellations which will be discussed in Section III. But first, let us define the GS problem.

B. Geometric Shaping

The goal of GS is to change the position of the equiprobable constellation points in a manner that AIR increases. According to (1), AIR depends on both the position of the points and the channel signal-to-noise ratio (SNR). Considering the mentioned dependencies, the GS problem is defined as follows: for a set of constellation points χ , with m -bit binary labeling \mathcal{L} , and shaping SNR γ (the SNR for which we maximize AIR):

$$\begin{aligned} \chi^*, \mathcal{L}^* &= \underset{\chi, \mathcal{L}}{\operatorname{argmax}} \operatorname{AIR}(\chi, \mathcal{L}, \gamma(\chi)), \\ \text{subject to: } &\frac{1}{M} \sum_{x \in \chi} |x|^2 = 1. \end{aligned} \quad (5)$$

Note that in (5), the position of the constellation points has an impact on the shaping SNR γ . In the case of the AWGN channels, this statement is not valid since χ does not affect the channel SNR. However, for the fiber channel, since fiber non-linearity is a modulation-dependent impairment [42], χ does affect the channel SNR. As a consequence, a more proper notation for γ is $\gamma(\chi)$. Moreover, since (5) is non-convex [8],

the modulation-dependency behavior of fiber increases the complexity of problem (5) significantly. In the next section, we offer an efficient solution to the optimization problem (5).

III. OPTIMIZATION SCHEME OF NON-LINEARITY TOLERANT 4D CONSTELLATIONS

In this section, we propose our solution to the non-linearity tolerant 4D GS problem. We briefly review the fiber channel model that is used in the optimization process and explain our methodology for designing 4D constellations that mitigate fiber non-linearity. Moreover, we discuss different AIR-based metrics that are appropriate for selecting as the objective function of the optimization scheme in order to have a small gap to the MI rate.

A. Fiber Channel Model

To design constellations that mitigate fiber non-linearity, a channel model must be defined. One way to model the non-linear interference noise (NLIN) of the fiber channel is to assume that the signal disturbance caused by the non-linearity manifests itself as an additive memoryless Gaussian noise. This is one of the main pillars of the enhanced Gaussian noise (EGN) model [42]. The EGN model considers both modulation-dependent and modulation-independent parameters that impact NLIN. As a result, the EGN model is an appropriate channel model to be used for non-linearity tolerant constellation shaping since it captures the effect of modulation on the NLIN power. The issue with the EGN model is that it is valid only for modulations that transmit independent data over the X and Y-polarization. As a consequence, the EGN model cannot be used for designing non-linearity tolerant 4D constellations as they transmit correlated data over the X and Y-polarization. In [37], the authors extend the idea of the EGN model for the 4D constellations. Hence, in this study, we employ the channel model of [37], which we refer to it as the 4D EGN model throughout this article, to consider the modulation-dependency behavior of the 4D constellations. Based on the 4D EGN model, the modified SNR of the fiber can be written as follows [37]:

$$\text{SNR}_{\text{NL}} = \frac{P_{\text{ch}}}{P_{\text{ASE}} + P_{\text{NLIN}}}, \quad (6)$$

where P_{ch} is the average launch power per wavelength division multiplexing (WDM) channel, P_{ASE} is the average power of the amplified spontaneous emission (ASE) noise, and P_{NLIN} is the NLIN power. In this study, we use (6) as a metric to see how non-linearity tolerant different constellations are. Also, assuming all WDM channels are spaced equally and all with the same modulation format, equal symbol rate, and launch power, there is an optimal launch power resulting in the maximum value of the modified SNR [37], [42], [43], [44]:

$$\text{SNR}_{\text{NL}}^{\text{max}} = \frac{2}{3} \times \sqrt[3]{\frac{1}{2 \cdot P_{\text{ASE}}^2 \cdot \eta}}, \quad (7)$$

where η is a parameter that contains both modulation dependent and independent terms that affect the NLIN power and is defined in [37]. Herein, to consider the impacts of shaping on NLIN, we follow the scheme of [5], i.e., we set shaping SNR γ to

$\text{SNR}_{\text{NL}}^{\text{max}}$. Results of [5] show that this scheme is a proper method for designing 2D non-linearity tolerant constellations. Here, we follow a similar approach; but, the main difference is that we aim for 4D non-linearity tolerant constellations instead of 2D ones. By doing so, the complexity of solving (5) increases significantly since parameter γ in the optimization problem of (5) becomes dependent on χ as a result of this new consideration. In the next subsection, we propose our method for tackling this challenge.

B. Neural Network-Based Optimization of the 4D Non-Linearity Tolerant Constellations

To solve (5) for an M -point 4D constellation, $4 \times M$ unknowns must be found, which is extremely complex, especially for large values of M . To reduce the complexity, we consider the orthant-symmetry (OS) constraint [13] on the positions of the 4D points, i.e., we optimize the position of the points only for the first orthant, and then, we mirror them to obtain points of the remaining orthants. The OS constraint not only decreases the complexity of the problem but also is compatible with the assumptions of model [37]. To optimize the position of the first orthant points, we employ a deep neural network (NN). The issue with NNs is that they do not necessarily satisfy the OS constraint. To overcome this issue, we feed the NN only the bit levels that correspond to the point position in the first orthant, (B_1, \dots, B_{m-4}) , and the remaining bit levels, (B_{m-3}, \dots, B_m) , are used to map the points to the corresponding orthant. To implement this scheme, as shown in Fig. 1, $m - 4$ bits out of m bits are used to determine the first orthant points, and the remaining 4 bits are used to find the orthant index. The $m - 4$ bits are given to the geometric shaping NN, and it outputs the in-phase and quadrant components of the X and Y-polarization on the first orthant. To make sure that the points are placed on the first orthant, the output layer activation function is set to ABS(\cdot). Also, to meet the average energy constraint of (5), the generated 4D points are normalized such that the average launch energy of both X and Y-polarization becomes 0.5, which results in the total average energy of 1 in the 4D space. After the normalization, to create the full 4D constellation, points are multiplied by their corresponding mirror matrix, which is computed using the orthant index bits.

After finding the 4D points, we use them to compute $\text{SNR}_{\text{NL}}^{\text{max}}$ using (7), and then, we generate 4D modulation-dependent noise such that SNR of the system becomes equal to $\text{SNR}_{\text{NL}}^{\text{max}}$. The 4D noise is added to the 4D points, and then, the objective function of the trainable system, which is AIR, is computed. In the backpropagation process, the trainable parameters of the geometric shaping NN are updated such that AIR becomes maximum.

The remaining issue is what AIR we should use as the objective function in order to have a small gap to the MI rate. As mentioned, the gap between the GMI and MI rate of the geometrically shaped constellations is significant due to non-Gray mapping. However, by employing multilevel coding (MLC) [31] techniques, the mentioned gap can be reduced substantially, which will be discussed in the next subsection.

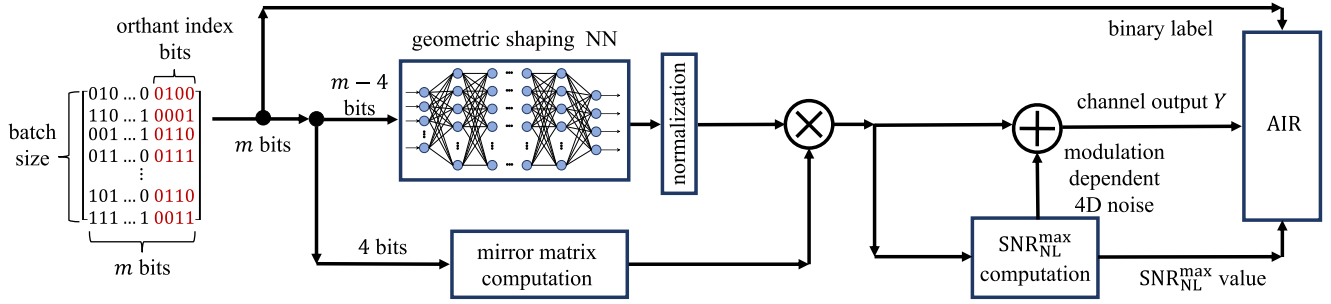


Fig. 1. Trainable system for non-linearity tolerant 4D GS. GS NN consists of 8 hidden layers whose size are 100 with the tanh activation function. The update of the trainable parameters of the geometric shaping NN is carried out using the Adam optimizer with a learning rate of 0.004 and a batch size of 2^{14} . Also, the system is trained for 300 epochs.

C. AIR Selection for the Proposed 4D GS Scheme

In Section II, we mentioned that it is desired to achieve the MI rate using the binary decoders, and (3) is the bit-wise solution to achieve it. Also, a simpler approach was introduced by removing the conditions in (3) and decoding the bit levels independently, which results in the GMI rate. For the traditional QAM constellations, the difference between the two schemes is negligible [3], [7], and as a result, one can justify employing scheme of (4) over (3). However, if the difference between (3) and (4) becomes substantial, more efficient schemes must be considered. In other words, the GMI and MI rates are the two extremes of a spectrum, and in order to approach the MI rate, the dependency between bit levels must be considered. In the following, we introduce several efficient AIRs for the 4D orthant-symmetric geometrically shaped constellations that can be implemented using the well-known MLC techniques.

According to our discussions in the previous subsection, the bit levels of the proposed 4D constellations can be separated as follows:

$$B = \left(\underbrace{B_1, B_2, \dots, B_{m-5}, B_{m-4}}_{\text{first orthant bits}}, \underbrace{B_{m-3}, B_{m-2}, B_{m-1}, B_m}_{\text{orthant index bits}} \right), \quad (8)$$

where the first orthant bits determine the position of the 4D point in the first orthant, and the orthant index bits determine at which of the 16 orthants the 4D point is located. According to the amplitude-sign factorization concept that has been discussed in [7], we can assume that the orthant index bits and the first orthant bits are statistically independent. Hence, the orthant index bits add negligible information for the decoding of the first orthant bits, i.e.,

$$\mathbb{I}(B_i; Y | B_{m-3}, B_{m-2}, B_{m-1}, B_m) \approx \mathbb{I}(B_i; Y) \quad \text{for } i = 1, 2, \dots, m-4. \quad (9)$$

Also, orthant index bits add little information for decoding of each other, i.e.,

$$\mathbb{I}(B_{m-j}; Y | B_{m-3}, \dots, B_{m-j+1}, B_{m-j-1}, \dots, B_m) \approx \mathbb{I}(B_{m-j}; Y) \quad \text{for } j = 1, 2, 3, 4. \quad (10)$$

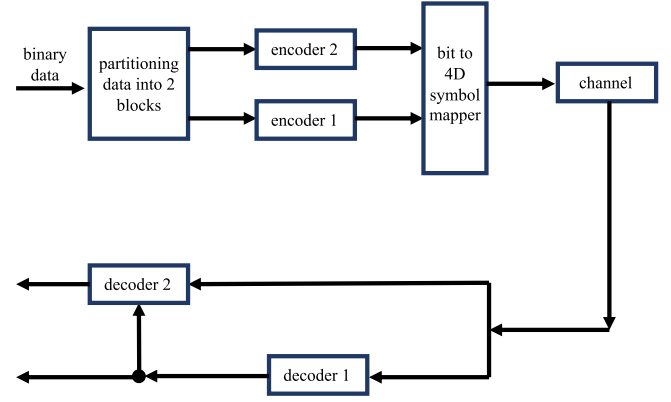


Fig. 2. Multilevel encoding and multistage decoding to achieve the rate of AIR₂.

As a result, we not only do not use the orthant index bits for the decoding of the first orthant bits but also we decode the orthant index bits independently. Considering (9) and (10), we use MLC only for the first orthant bits.

For the AIR that requires 2-level codes, we propose the following AIR:

$$\begin{aligned} \text{AIR}_2 = & \sum_{i=1}^{\lfloor \frac{m-4}{2} \rfloor} \mathbb{I}(B_i; Y) + \sum_{i=\lfloor \frac{m-4}{2} \rfloor + 1}^{m-4} \mathbb{I}(B_i; Y | B_1 \cdots B_{\lfloor \frac{m-4}{2} \rfloor}) \\ & + \sum_{i=m-3}^m \mathbb{I}(B_i; Y), \end{aligned} \quad (11)$$

where $\lfloor \cdot \rfloor$ is the floor function. Here, only one term in (11) is conditioned on other bits. This means AIR₂ can be achieved using 2-level MLC. In other words, $(B_1, \dots, B_{\lfloor \frac{m-4}{2} \rfloor})$ and (B_{m-3}, \dots, B_m) are decoded independently and $(B_{\lfloor \frac{m-4}{2} \rfloor + 1}, \dots, B_{m-4})$ are decoded based on $(B_1, \dots, B_{\lfloor \frac{m-4}{2} \rfloor})$. Also, the last summation in (11) represents the total AIR of the orthant index bits. Note that there are several ways to propose an AIR with 2-level codes; however, based on our experiments, (11) results in the highest rates. As a simple explanation, Fig. 2 provides a high-level explanation of the multilevel encoding and multistage decoding in order to achieve

AIR₂, where encoder 1 and encoder 2 generate binary data corresponding to bit levels $(B_1, \dots, B_{\lfloor \frac{m-4}{2} \rfloor}, B_{m-3}, \dots, B_m)$ and $(B_{\lfloor \frac{m-4}{2} \rfloor + 1}, \dots, B_{m-4})$, respectively. Also, decoder 1 employs the channel outputs 4D symbols to decode the bit stream corresponding to $(B_1, \dots, B_{\lfloor \frac{m-4}{2} \rfloor}, B_{m-3}, \dots, B_m)$, and decoder 2 employs the channel output as well as the decoder 1 output to decode bit stream corresponding to $(B_{\lfloor \frac{m-4}{2} \rfloor + 1}, \dots, B_{m-4})$.

For the AIR that can be achieved using 3-level and 4-level codes, we propose

$$\begin{aligned} \text{AIR}_3 &= \sum_{i=1}^{\lfloor \frac{m-4}{2} \rfloor} \mathbb{I}(B_i; Y) + \mathbb{I}\left(B_{\lfloor \frac{m-4}{2} \rfloor + 1}; Y | B_1 \cdots B_{\lfloor \frac{m-4}{2} \rfloor}\right) \\ &+ \sum_{i=\lfloor \frac{m-4}{2} \rfloor + 2}^{m-4} \mathbb{I}\left(B_i; Y | B_1 \cdots B_{\lfloor \frac{m-4}{2} \rfloor + 1}\right) + \sum_{i=m-3}^m \mathbb{I}(B_i; Y), \end{aligned} \quad (12)$$

and

$$\begin{aligned} \text{AIR}_4 &= \mathbb{I}(B_1; Y) + \mathbb{I}(B_2; Y | B_1) + \mathbb{I}(B_3; Y | B_1, B_2) \\ &+ \sum_{i=4}^{m-4} \mathbb{I}(B_i; Y | B_1, B_2, B_3) + \sum_{i=m-3}^m \mathbb{I}(B_i; Y), \end{aligned} \quad (13)$$

respectively. Similar to the 2-level scenario, there are several ways to write AIRs with 3 and 4-level codes, but (12) and (13), based on our results, lead to the best performance. Also, note that AIR₄ can be used only for 4D constellations with $m > 7$. It is worth mentioning that since conditioning does not increase entropy, as the number of level of codes increases, the corresponding AIR increases as well. In other words, at a fixed SNR level, we have:

$$\text{GMI} \leq \text{AIR}_2 \leq \text{AIR}_3 \leq \text{AIR}_4 \leq \text{MI}. \quad (14)$$

Considering the defined AIRs, we use them as the objective function in the trainable system of Fig. 1 in order to see how much each scheme can fill the gap to the MI rate. The performance and comparison of the schemes will be discussed in the next section.

IV. RESULTS AND DISCUSSIONS

In the following, we employ the proposed 4D GS scheme as well as the defined AIRs to optimize 128 and 256-point 4D constellations and compare their performance with the traditional QAM constellations and other shaping schemes. Here, we consider both the AWGN and the non-linear fiber channel. By optimizing constellations for the AWGN channel, we understand how much shaping gain is achieved in the linear channels, and by optimizing non-linearity tolerant constellations for the fiber channel, we realize how much non-linear shaping gains the trainable system of Fig. 1 can provide in addition to the linear shaping gains.

A. Optimization for the AWGN Channel

In this section, the geometry of the 4D constellations is optimized for the AWGN channel. To do so, a simple modification in

the trainable system of Fig. 1 is required since the shaping SNR γ is constant during the optimization process. In other words, instead of the modulation-dependent 4D noise, we add independent circularly symmetric Gaussian noise to the 4D symbols. We use the modified scheme to optimize 128 and 256-point 4D constellations for the AWGN channel. The shaping SNRs of 9.5 and 11.5 dB are considered for 128 and 256-point constellations, respectively. We shape constellations based on the defined AIRs as well as the MI and GMI rates. Moreover, we compare our constellations with the traditional probabilistic-shaped QAM constellations and the recent 4D GS method.

The AIR performance of our 128-point 4D constellations (4D-128-GS) are provided in Fig. 3(a). Also, the performance of the polarization-multiplexed (PM) QAM, traditional probabilistic-shaped PM-QAM (PM-PS-QAM), and 4D-128OS of [13] are provided. In this study, for PM-QAM with the entropy of 7 bits/4D-symbol, we assume that star-8-QAM and 16QAM are multiplexed. Also, the same geometry is used for PM-PS-QAM. Based on Fig. 3(a), GMI-based shaping provides much lower rates compared to AIR₂ and AIR₃-based shaping. Note that in the rest of the article, we refer to optimization of the geometry such that AIR_i becomes maximum as AIR_i-based shaping. AIR₂-based shaping scheme has slightly lower AIR compared to AIR₃-based shaping, and they fill the gap to the MI rate by 85% and 86%, respectively. Moreover, our GMI, AIR₂, and AIR₃-based shaping schemes provide 0.6, 0.71, and 0.715 dB gain over PM-QAM, respectively. Also, they outperform PM-PS-QAM by 0.2, 0.34, and 0.342 dB, respectively. In addition, our GMI-based shaped constellation provides rates slightly higher than 4D-128OS of [13]; however, AIR₂ and AIR₃-based shaping methods significantly outperform 4D-128OS.

In Fig. 3(b), the performance of the 256-point constellations has been provided. The AIR₂, AIR₃, and AIR₄-based shaping schemes significantly outperform the GMI-based shaping, and they fill the gap to the MI rate by 72%, 85%, and 99%, respectively. Moreover, in the case of 256-point constellations, the GMI-based shaping is not capable of outperforming PM-PS-16QAM; however, AIR₂, AIR₃, and AIR₄-based shaping schemes significantly outperform PM-PS-16QAM, and they provide 0.11, 0.15, and 0.19 dB gain over PM-PS-16QAM, respectively. Also, our optimized constellations outperform PM-16QAM by 0.25, 0.44, 0.48, and 0.51 dB, respectively. As an example, the geometry, coordinates, and binary labels of the AIR₂-based shaped constellation are provided in Appendix A.

The results of optimizing 4D constellations for the AWGN channel substantiate that the GMI-based shaping is inefficient for 4D GS as it results in a huge gap to the MI rate. On the other hand, AIR₂-based shaping, which is achieved by just 2-level codes, is much more efficient and has a significant potential to reduce the gap to the MI rate. Moreover, AIR₂-based shaping is capable of outperforming the traditional PS, and even in the linear regimes, the shaping gain of AIR₂-based shaping is higher than the traditional PS. In the next subsection, we employ the 4D GS scheme of Fig. 1 to optimize 4D constellations specifically for the non-linear fiber channel, and we examine how much non-linear shaping gains the proposed GS method can provide in addition to the linear gains.

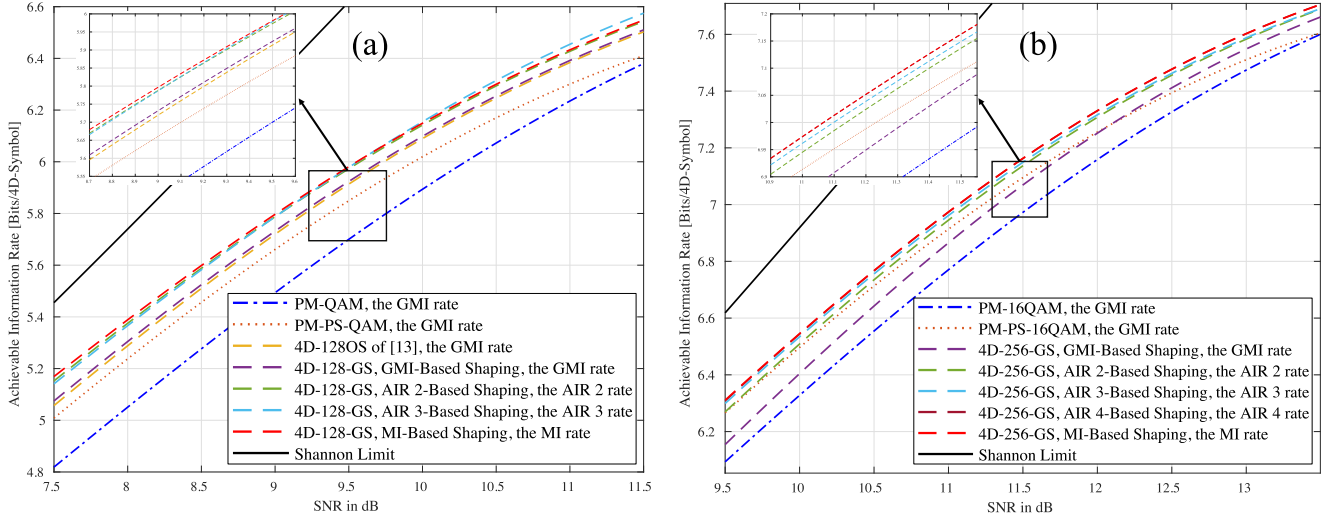


Fig. 3. AIR performance of (a) 128-point 4D constellations optimized for SNR of 9.5 dB and (b) 256-point 4D constellations optimized for SNR of 11.5 dB.

TABLE I
SYSTEM AND SIMULATION PARAMETERS

Parameter	Value
Polarization	Dual-polarization
Symbol rate	32 GBaud
WDM channels	9
WDM spacing	35 GHz
Pulse shape	Root-raised cosine (RRC)
RRC roll-off	0.01
Attenuation	0.22 dB/km
Dispersion	16.7 ps/nm/km
Non-linear coefficient	1.3 1/W/km
Span length	100 km
Amplification	EDFA
EDFA noise figure	5 dB
SSFM step size	0.1 km
Oversampling factor	16
Transmitted symbols per WDM ch.	2^{16}

B. Optimization for the Non-Linear Fiber Channel

In the following, we focus on optimizing the geometry of 128 and 256-point 4D constellations specifically for the fiber channel. We investigate the performance of our non-linearity tolerant 4D constellations in terms of reach improvement, AIR gain, and mitigation of fiber non-linearity. We also compare our 4D GS method with the traditional PS and the recent non-linearity tolerant GS methods. But first, let us discuss the optical fiber system parameters and simulation procedure.

1) *System Parameters and Simulation Procedure:* the optical fiber system parameters are provided in Table I. The generated symbols are transmitted over all WDM channels, and for 2D constellations, data of X and Y-polarization are independent and identically distributed (i.i.d). For 4D constellations, X and Y-polarization data are correlated but the overall 4D symbols are i.i.d. The central channel is the channel-under-test. Propagation of the signal over the fiber link is simulated using the split-step Fourier method (SSFM). The fiber link is made up of identical spans. At the end of each span, the loss is exactly compensated

for by an erbium-doped fiber amplifier (EDFA), then, the ASE noise is added to the amplified signal. At the receiver, the central channel is filtered, and the chromatic dispersion is compensated for. After that, the signal passes through the matched filter, then, the filtered signal is downsampled to obtain the received symbols. Also, the average non-linear phase rotation is compensated for. Finally, the AIR between the transmitted and received symbols is computed according to the Monte Carlo simulations. For 2D constellations, to compute the total AIR, we sum over the AIR of X and Y-polarization.

2) *Reach Improvement From the Proposed 4D GS Scheme:* In this part, we optimize the geometry of 4D 128 and 256-point constellations for a specific span and investigate their AIR performance in a wide range of distances. Also, the performance of our 4D constellations (4D-128-GSs and 4D-256-GSs) is compared with the corresponding PM-QAM and PM-PS-QAM. We optimize the 128 and 256-point constellations for spans 45 (4500 km) and 25 (2500 km), respectively. The results are provided in Fig. 4, where the AIR performance of the constellations at the corresponding optimal launch power is reported. According to Fig. 4(a), our 4D constellations provide much higher AIRs compared to PM-QAM and PM-PS-QAM. For AIRs between 5 and 5.8 bits/4D-symbol, our GMI, AIR₂ and AIR₃-based shaping methods give approximately 600 km (13.0%), 800 km (17.3%), and 850 km (18.4%) reach improvements compared to PM-QAM, respectively. Moreover, the GMI, AIR₂ and AIR₃-based shaped constellations provide almost 100 km (2.0%), 300 km (5.8%), and 350 km (6.8%) reach improvements over PM-PS-QAM. It is worth mentioning that based on the reported results, although the GMI-based shaping method outperforms PM-PS-QAM, it provides much lower AIRs compared to AIR₂ and AIR₃-based shaping schemes. In the case of 256-point constellations, according to Fig. 4(b), the GMI-based shaping method is not capable of outperforming PM-PS-16QAM. In other words, for AIRs between 6.4 and 7.1 bits/4D-symbol, PM-PS-16QAM provides almost 100 km (3.3%) reach improvements over the GMI-based shaping methods. However,

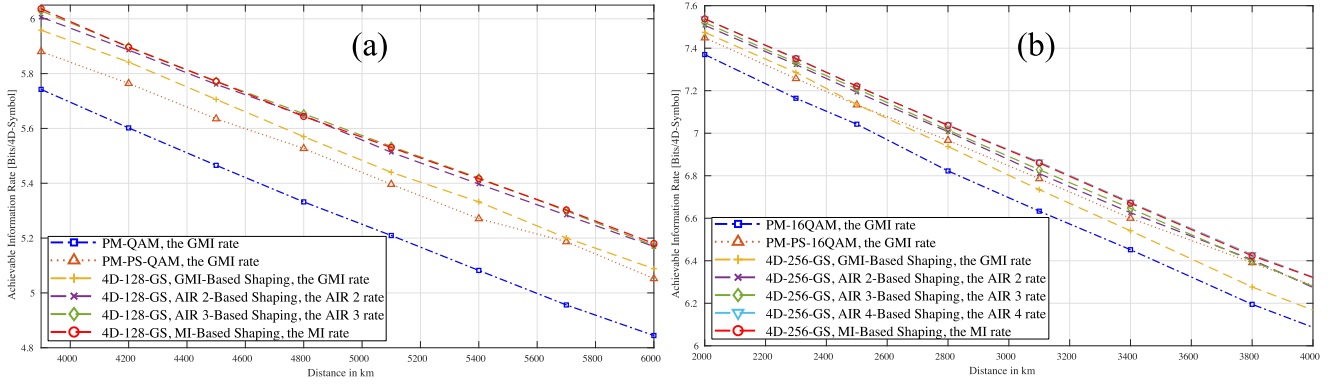


Fig. 4. AIR performance of the optimized non-linearity constellations versus the distance for (a) 128-point 4D constellations and (b) 256-point 4D constellations.

the AIR₂, AIR₃, and AIR₄-based shaping schemes outperform PM-PS-16QAM and give approximately 25 km (1.0%), 50 km (1.6%), and 100 km (3.2%) reach improvements compared to PM-PS-16QAM, respectively. Also, the AIR₂, AIR₃, and AIR₄-based shaping methods give approximately 300 km (10.7%), 325 km (11.6%), and 375 km (13.4%) reach improvements compared to PM-16QAM, respectively, for AIRs between 6.4 and 7.1 bits/4D-symbol. The reported results of our 128 and 256-point constellations substantiate that in addition to providing significant AIRs, our constellations are robust against distance and different levels of fiber non-linearity. This is an important advantage of our 4D GS method since one of the issues of GS is that the designed constellation usually performs well only at the distance for which it has been optimized [5]. However, this is not the case with our constellations, and they perform very well at high levels of non-linearity although they were specifically designed for lower levels of non-linearity.

3) *AIR Gain From the Proposed 4D GS Scheme:* In this section, we investigate the shaping gains of our optimized 128 and 256-point 4D constellations (4D-128-GSs and 4D-256-GSs) at span 45 (4500 km) and 25 (2500 km), respectively. Moreover, we compare the AIR performance of our constellations with the corresponding traditional probabilistic-shaped QAMs, the non-linearity tolerant PS scheme of [45], and the recent GS methods of [13] and [5]. The AIR results of 128 and 256-point constellations are provided in Fig. 5(a) and (b) for launch powers between -1.5 dBm to 3 dBm, respectively. According to Fig. 5(a), the performance of the GMI-based shaping method is significantly lower than the MI rate. On the other hand, the AIR₂ and AIR₃-based shaping schemes provide AIRs much closer to the MI rate, and they fill the gap to the MI rate by approximately 83% and 93%, respectively. Moreover, at the corresponding optimal launch power, the GMI, AIR₂ and AIR₃-based shaping methods provide shaping gains of almost 0.25, 0.3, 0.31 bits/4D-symbol over PM-QAM, respectively, and they outperform PM-PS-QAM by approximately 0.08, 0.13, and 0.14 bits/4D-symbol, respectively. Also, the GMI-based shaping method has a similar performance compared to 4D-128OS of [13]; however, the AIR₂ and AIR₃-based shaping schemes outperform 4D-128OS substantially. In Fig. 5(b), the

AIR performance of the 256-point constellations at span 25 is provided. The GMI-based shaping method is not capable of outperforming PM-PS-16QAM and it provides comparable AIRs. However, the AIR₂, AIR₃, and AIR₄-based shaping schemes provide significantly higher AIRs and they fill the gap to the MI rate by 76%, 88%, and 92%, respectively. In addition, at the corresponding optimal launch power, the AIR₂, AIR₃, and AIR₄-based shaping schemes give shaping gains of 0.18, 0.19, and 0.21 bits/4D-symbol over PM-16QAM, respectively. Moreover, they outperform PM-PS-16QAM by almost 0.07, 0.08, and 0.09 bits/4D-symbol, respectively. In addition, the MI performance of the 2D non-linearity tolerant PM-16APSK of [5] is provided in Fig. 5(b). PM-16APSK gives shaping gains lower than the GMI-based shaping scheme and does not outperform PM-PS-16QAM. Moreover, the AIR₂, AIR₃, and AIR₄-based shaping schemes surpass PM-16APSK significantly, indicating that designing constellations in the 4D space is much more efficient compared to 2D GS.

4) *Mitigation of the Fiber Non-Linearity From the Proposed 4D GS Scheme:* As mentioned, NLIN is a modulation-dependent interference, and as a consequence, measuring NLIN for a constellation is of great importance as it provides significant insight into whether the constellation mitigates the fiber non-linearity. To investigate how non-linearity tolerant our shaped constellations are compared to other constellations, we examine the SNR_{NL} performance of the constellations. We compute SNR_{NL} of the AIR₂-based shaped 128 and 256-point constellations at span 45 and 25 whose results are provided in Fig. 5(c) and (d), respectively. According to Fig. 5(c), our optimized 128-point constellation introduces the least amount of NLIN to the optical fiber system among all shaped and non-shaped constellations, and it provides almost 0.08, 0.2, 0.12, and 0.1 dB SNR_{NL} gains over PM-QAM, PM-PS-QAM, the PS scheme of [45], and 4D-128OS of [13]. Also, based on Fig. 5(d), our optimized 256-point constellation outperforms PM-PS-16QAM and the PS scheme of [45] by almost 0.13 and 0.09 dB, respectively, and it outperforms both PM-16QAM and non-linearity tolerant PM-16APSK of [5] by approximately 0.08 dB SNR_{NL} gains, confirming that 4D GS results in much more non-linearity tolerant constellations than 2D GS.

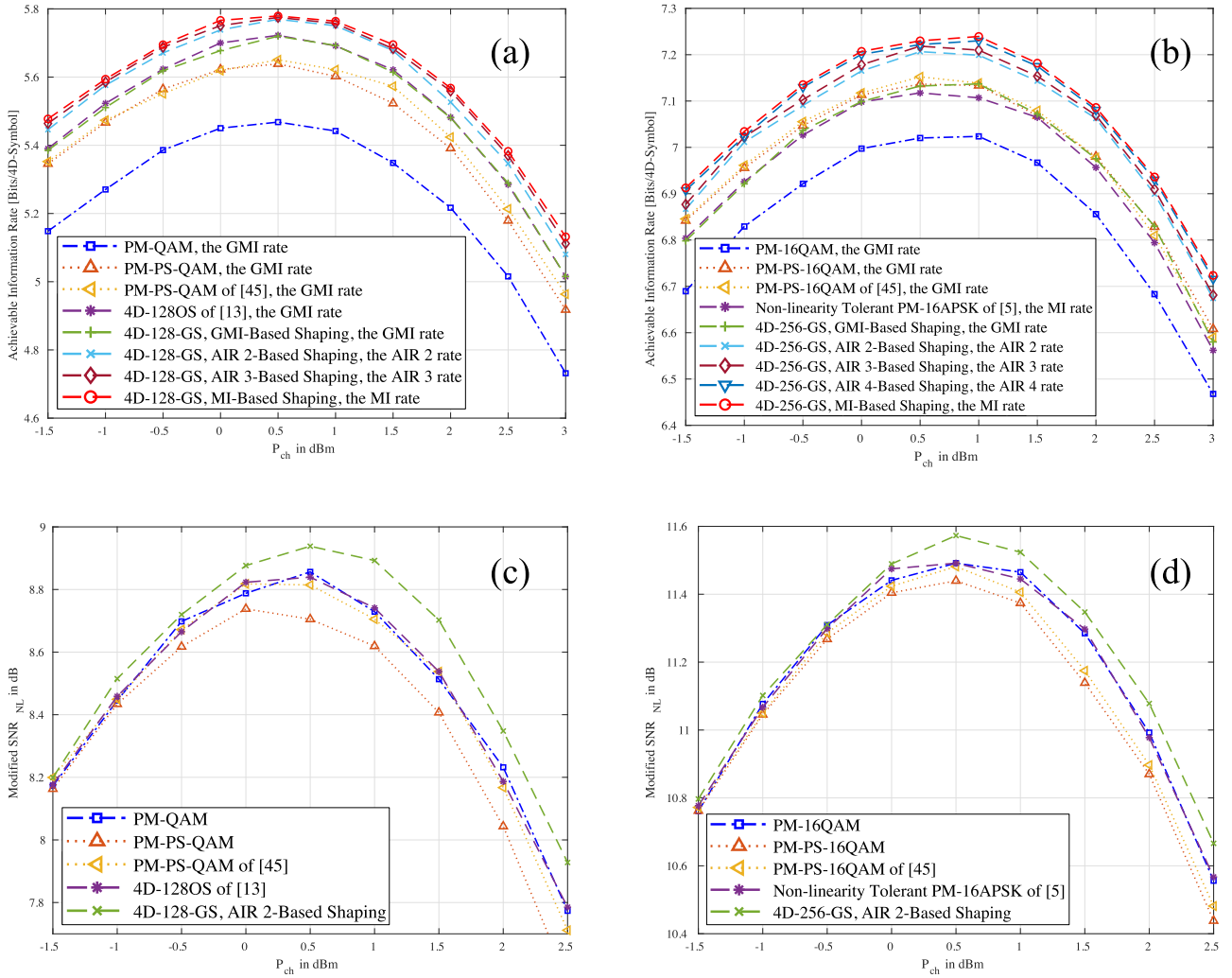


Fig. 5. AIR performance of the optimized 4D non-linearity constellations and other shaped constellations versus the launch power for (a) 128-point constellations at span 45 and (b) 256-point constellations at span 25. Also, the performance of the 128 and 256-point constellations in terms of SNR_{NL} versus the launch power at span 45 and 25 are provided in (c) and (d), respectively.

V. CONCLUSION

In this study, we introduced a novel non-linearity tolerant 4D geometric shaping method based on multilevel coding for the non-linear fiber channel by considering the effects of shaping on NLIN. We employed the proposed method to optimize 128 and 256-point 4D constellations for both the AWGN and fiber channels. We discussed that shaping the geometry of 4D constellations based on the GMI rate is not efficient as it leads to a huge gap to the MI rate. Instead, we proposed AIRs that are achieved using the MLC technique. We showed that using an MLC scheme that requires only 2-level codes, the performance of GS can significantly be improved. In some reported cases, we reported 18.4% and 6.8% reach improvements over the regular and PM-PS-QAM constellations, respectively. Moreover, we showed that our shaped constellations mitigate the fiber non-linearity and introduce the least NLIN to the system compared to PM-QAMs, PM-PS-QAMs, and the recent non-linearity tolerant geometric shaping methods.

APPENDIX A

THE GEOMETRY OF THE OPTIMIZED 4D CONSTELLATION

In this appendix, we provide the geometry, coordinates, and binary labels of the AIR₂-based shaped 256-point constellation that is optimized for the AWGN channel with a shaping SNR of 11.5 dB. The geometry of the 4D constellation in the first orthant is provided in Fig. 6. Moreover, the coordinates and the binary labels of the first orthant are provided in Table II. Also, the mirror matrices can be computed based on the orthant index bits as follows:

$$M_{(B_{m-3}, B_{m-2}, B_{m-1}, B_m)} = \text{diag} \left(\begin{array}{c} -2 \cdot B_{m-3} + 1 \\ -2 \cdot B_{m-2} + 1 \\ -2 \cdot B_{m-1} + 1 \\ -2 \cdot B_m + 1 \end{array} \right), \quad (15)$$

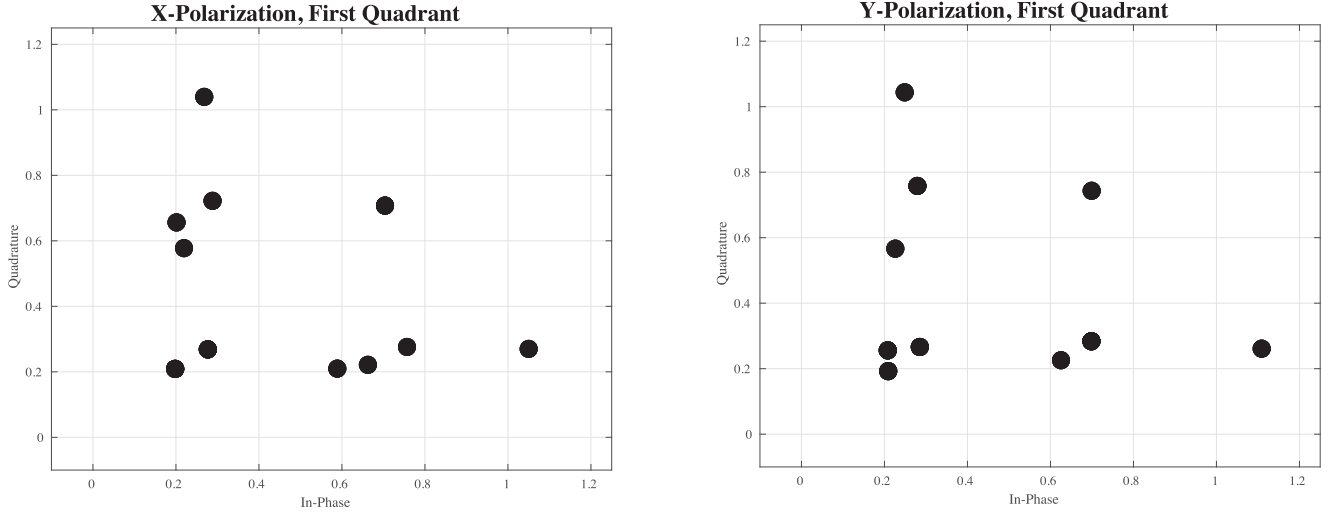


Fig. 6. Geometry of the AIR₂-based shaped 256-point constellation with shaping SNR of 11.5 dB in the first quadrant of the X-polarization (left) and Y-polarization (right).

TABLE II
COORDINATES AND THE CORRESPONDING FIRST ORTHANT BINARY LABELS OF
THE AIR₂-BASED SHAPED 256-POINT CONSTELLATION

Coordinates	(B_1, B_2, B_3, B_4)
$(0.75678, 0.27595, 0.27931, 0.75798)^T$	$(0, 0, 0, 0)$
$(0.27688, 0.26860, 0.24879, 1.04383)^T$	$(0, 0, 0, 1)$
$(0.19811, 0.20897, 0.20898, 0.19251)^T$	$(0, 0, 1, 0)$
$(0.19811, 0.20897, 0.22612, 0.56645)^T$	$(0, 0, 1, 1)$
$(0.27688, 0.26860, 1.10902, 0.26108)^T$	$(0, 1, 0, 0)$
$(0.28821, 0.72211, 0.27931, 0.75798)^T$	$(0, 1, 0, 1)$
$(0.19811, 0.20897, 0.62545, 0.22583)^T$	$(0, 1, 1, 0)$
$(0.27688, 0.26860, 0.69936, 0.74321)^T$	$(0, 1, 1, 1)$
$(0.21944, 0.57763, 0.20898, 0.19251)^T$	$(1, 0, 0, 0)$
$(0.70379, 0.70745, 0.69860, 0.28389)^T$	$(1, 0, 0, 1)$
$(0.58879, 0.20982, 0.20801, 0.25599)^T$	$(1, 0, 1, 0)$
$(1.05004, 0.26999, 0.28525, 0.26633)^T$	$(1, 0, 1, 1)$
$(0.20137, 0.65626, 0.69860, 0.28389)^T$	$(1, 1, 0, 0)$
$(0.26815, 1.03952, 0.28525, 0.26633)^T$	$(1, 1, 0, 1)$
$(0.66256, 0.22136, 0.69860, 0.28389)^T$	$(1, 1, 1, 0)$
$(0.70379, 0.70745, 0.20801, 0.25599)^T$	$(1, 1, 1, 1)$

where $\text{diag}(\cdot)$ returns a square diagonal matrix with the elements of its argument on the main diagonal. Using Table II and (15), the whole 4D constellation can be constructed.

REFERENCES

- [1] R. Essiambre and R. W. Tkach, "Capacity trends and limits of optical communication networks," *Proc. IEEE*, vol. 100, no. 5, pp. 1035–1055, May 2012.
- [2] P. Bayvel et al., "Maximizing the optical network capacity," *Philos. Trans. Roy. Soc.*, vol. 374, no. 2062, Mar. 2016, Art. no. 20140440.
- [3] T. Fehenberger, A. Alvarado, G. Böcherer, and N. Hanik, "On probabilistic shaping of quadrature amplitude modulation for the nonlinear fiber channel," *J. Lightw. Technol.*, vol. 34, no. 21, pp. 5063–5073, Nov. 2016.
- [4] J. Cho and P. J. Winzer, "Probabilistic constellation shaping for optical fiber communications," *J. Lightw. Technol.*, vol. 37, no. 6, pp. 1590–1607, Mar. 2019.
- [5] A. Soleimanzade and M. Ardakani, "EGN-based optimization of the APSK constellations for the non-linear fiber channel based on the symbol-wise mutual information," *J. Lightw. Technol.*, vol. 40, no. 7, pp. 1937–1952, Apr. 2022.
- [6] R. T. Jones, M. P. Yankov, and D. Zibar, "End-to-end learning for GMI optimized geometric constellation shape," in *Proc. IEEE 45th Eur. Conf. Opt. Commun.*, 2019, pp. 1–4.
- [7] G. Böcherer, F. Steiner, and P. Schulte, "Bandwidth efficient and rate-matched low-density parity-check coded modulation," *IEEE Trans. Commun.*, vol. 63, no. 12, pp. 4651–4665, Dec. 2015.
- [8] F. Steiner and G. Boecherer, "Comparison of geometric and probabilistic shaping with application to ATSC 3.0," in *Proc. IEEE 11th Int. ITG Conf. Syst. Commun. Coding*, 2017, pp. 1–6.
- [9] G. Böcherer, F. Steiner, and P. Schulte, "Fast probabilistic shaping implementation for long-haul fiber-optic communication systems," in *Proc. IEEE Eur. Conf. Opt. Commun.*, 2017, pp. 1–3.
- [10] A. Amari et al., "Introducing enumerative sphere shaping for optical communication systems with short blocklengths," *J. Lightw. Technol.*, vol. 37, no. 23, pp. 5926–5936, Dec. 2019.
- [11] Z. Qu and I. B. Djordjevic, "On the probabilistic shaping and geometric shaping in optical communication systems," *IEEE Access*, vol. 7, pp. 21454–21464, 2019.
- [12] K. Kojima et al., "Nonlinearity-tolerant four-dimensional 2A8PSK family for 5–7 bits/symbol spectral efficiency," *J. Lightw. Technol.*, vol. 35, no. 8, pp. 1383–1391, Apr. 2017.
- [13] B. Chen, A. Alvarado, S. van der Heide, M. van den Hout, H. Hafermann, and C. Okonkwo, "Analysis and experimental demonstration of orthant-symmetric four-dimensional 7 bit/4D-sym modulation for optical fiber communication," *J. Lightw. Technol.*, vol. 39, no. 9, pp. 2737–2753, May 2021.
- [14] S. Zhang and F. Yaman, "Design and comparison of advanced modulation formats based on generalized mutual information," *J. Lightw. Technol.*, vol. 36, no. 2, pp. 416–423, Jan. 2018.

- [15] B. Chen, C. Okonkwo, D. Lavery, and A. Alvarado, "Geometrically-shaped 64-point constellations via achievable information rates," in *Proc. IEEE 20th Int. Conf. Transparent Opt. Netw.*, 2018, pp. 1–4.
- [16] S. Zhang and F. Yaman, "Constellation design with geometric and probabilistic shaping," *Opt. Commun.*, vol. 409, pp. 7–12, 2018. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0030401817307538>
- [17] R. T. Jones, T. A. Eriksson, M. P. Yankov, and D. Zibar, "Deep learning of geometric constellation shaping including fiber nonlinearities," in *Proc. IEEE Eur. Conf. Opt. Commun.*, 2018, pp. 1–3.
- [18] B. Chen et al., "Shaped four-dimensional modulation formats for optical fiber communication systems," in *Proc. Opt. Fiber Commun. Conf.*, 2022, Paper Th3F.4. [Online]. Available: <https://opg.optica.org/abstract.cfm?URI=OFC-2022-Th3F.4>
- [19] M. Chagnon, M. Osman, Q. Zhuge, X. Xu, and D. V. Plant, "Analysis and experimental demonstration of novel 8PolSK-QPSK modulation at 5 Bits/symbol for passive mitigation of nonlinear impairments," *Opt. Exp.*, vol. 21, no. 25, pp. 30204–30220, Dec. 2013. [Online]. Available: <https://opg.optica.org/oe/abstract.cfm?URI=oe-21-25-30204>
- [20] B. Chen, C. Okonkwo, H. Hafermann, and A. Alvarado, "Polarization-switching for nonlinearity-tolerant geometrically shaped four-dimensional formats maximizing generalized mutual information," *J. Lightw. Technol.*, vol. 37, no. 14, pp. 3579–3591, Jul. 2019.
- [21] R.-J. Essiambre et al., "Increased reach of long-haul transmission using a constant-power 4D format designed using neural networks," in *Proc. Eur. Conf. Opt. Commun.*, 2020, pp. 1–4.
- [22] E. Sillekens, D. Semrau, D. Lavery, P. Bayvel, and R. I. Killey, "Experimental demonstration of geometrically-shaped constellations tailored to the nonlinear fibre channel," in *Proc. IEEE Eur. Conf. Opt. Commun.*, 2018, pp. 1–3.
- [23] B. Moore, G. Takahara, and F. Alajaji, "Pairwise optimization of modulation constellations for non-uniform sources," *Can. J. Elect. Comput. Eng.*, vol. 34, no. 4, pp. 167–177, 2009.
- [24] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proc. IEEE Int. Conf. Neural Netw.*, 1995, pp. 1942–1948.
- [25] S. Li, C. Häger, N. Garcia, and H. Wymeersch, "Achievable information rates for nonlinear fiber communication via end-to-end autoencoder learning," in *Proc. IEEE Eur. Conf. Opt. Commun.*, 2018, pp. 1–3.
- [26] M. Schaedler et al., "Neural network assisted geometric shaping for 800Gbit/s and 1Tbit/s optical transmission," in *Proc. IEEE Opt. Fiber Commun. Conf. Exhib.*, 2020, pp. 1–3.
- [27] B. Chen et al., "Geometrically-shaped multi-dimensional modulation formats in coherent optical transmission systems," *J. Lightw. Technol.*, vol. 41, no. 3, pp. 897–910, Feb. 2023. [Online]. Available: <https://opg.optica.org/jlt/abstract.cfm?URI=jlt-41-3-897>
- [28] S. Goossens et al., "4D geometric shell shaping with applications to 400ZR," in *Proc. Optica Adv. Photon. Congr.*, 2022, Paper SpM2I.1. [Online]. Available: <https://opg.optica.org/abstract.cfm?URI=SPPCom-2022-SpM2I.1>
- [29] S. Goossens et al., "Introducing 4D geometric shell shaping," 2022, *arXiv:2206.03341*.
- [30] C. Audet and J. E. Dennis Jr, "Analysis of generalized pattern searches," *SIAM J. Optim.*, vol. 13, no. 3, pp. 889–903, 2002.
- [31] U. Wachsmann, R. Fischer, and J. Huber, "Multilevel codes: Theoretical concepts and practical design rules," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1361–1391, Jul. 1999.
- [32] T. Matsumine, M. P. Yankov, and S. Forchhammer, "Geometric constellation shaping for concatenated two-level multi-level codes," *J. Lightw. Technol.*, vol. 40, no. 16, pp. 5557–5566, Aug. 2022.
- [33] M. Fu et al., "Parallel bisection-based distribution matching for nonlinearity-tolerant probabilistic shaping in coherent optical communication systems," *J. Lightw. Technol.*, vol. 39, no. 20, pp. 6459–6469, Oct. 2021.
- [34] Y. C. Gültekin et al., "Mitigating nonlinear interference by limiting energy variations in sphere shaping," in *Proc. Opt. Fiber Commun. Conf.*, 2022, Paper Th3F.2. [Online]. Available: <https://opg.optica.org/abstract.cfm?URI=OFC-2022-Th3F.2>
- [35] Z. Gan, X. Li, and S. J. Savory, "Nonlinearity aware bisection-based sphere shaping for optical digital subcarrier multiplexing systems," *Opt. Exp.*, vol. 30, no. 24, pp. 44118–44131, Nov. 2022. [Online]. Available: <https://opg.optica.org/oe/abstract.cfm?URI=oe-30-24-44118>
- [36] K. Wu, G. Liga, A. Sheikh, Y. C. Gültekin, F. M. J. Willems, and A. Alvarado, "List-encoding CCDF: A nonlinearity-tolerant shaper aided by energy dispersion index," *J. Lightw. Technol.*, vol. 40, no. 4, pp. 1064–1071, Feb. 2022. [Online]. Available: <https://opg.optica.org/jlt/abstract.cfm?URI=jlt-40-4-1064>
- [37] H. Rabbani et al., "Analytical modeling of nonlinear fiber propagation for four dimensional symmetric constellations," *J. Lightw. Technol.*, vol. 39, no. 9, pp. 2704–2713, May 2021.
- [38] A. Ganti, A. Lapidoth, and I. E. Telatar, "Mismatched decoding revisited: General alphabets, channels with memory, and the wide-band limit," *IEEE Trans. Inf. Theory*, vol. 46, no. 7, pp. 2315–2328, Nov. 2000.
- [39] G. Böcherer, P. Schulte, and F. Steiner, "Probabilistic shaping and forward error correction for fiber-optic communication systems," *J. Lightw. Technol.*, vol. 37, no. 2, pp. 230–244, Jan. 2019.
- [40] D. M. Arnold, H. Loeliger, P. O. Vontobel, A. Kavcic, and W. Zeng, "Simulation-based computation of information rates for channels with memory," *IEEE Trans. Inf. Theory*, vol. 52, no. 8, pp. 3498–3508, Aug. 2006.
- [41] A. Martinez, A. Guillen i Fabregas, G. Caire, and F. M. J. Willems, "Bit-interleaved coded modulation revisited: A mismatched decoding perspective," *IEEE Trans. Inf. Theory*, vol. 55, no. 6, pp. 2756–2765, Jun. 2009.
- [42] A. Carena, G. Bosco, V. Curri, Y. Jiang, P. Poggiolini, and F. Forghieri, "EGN model of non-linear fiber propagation," *Opt. Exp.*, vol. 22, no. 13, pp. 16335–16362, Jun. 2014. [Online]. Available: <http://www.opticsexpress.org/abstract.cfm?URI=oe-22-13-16335>
- [43] P. Poggiolini, G. Bosco, A. Carena, V. Curri, Y. Jiang, and F. Forghieri, "The GN-model of fiber non-linear propagation and its applications," *J. Lightw. Technol.*, vol. 32, no. 4, pp. 694–721, Feb. 2014.
- [44] A. Carena, V. Curri, G. Bosco, P. Poggiolini, and F. Forghieri, "Modeling of the impact of nonlinear propagation effects in uncompensated optical coherent transmission links," *J. Lightw. Technol.*, vol. 30, no. 10, pp. 1524–1539, May 2012.
- [45] E. Sillekens et al., "A simple nonlinearity-tailored probabilistic shaping distribution for square QAM," in *Proc. IEEE Opt. Fiber Commun. Conf. Expo.*, 2018, pp. 1–3.