Analysis and Experimental Demonstration of Orthant-Symmetric Four-dimensional 7 bit/4D-sym Modulation for Optical Fiber Communication

Analysis and Experimental Demonstration of Orthant-Symmetric Four-Dimensional 7 bit/4D-Sym Modulation for Optical Fiber Communication

Bin Chen, *Member, IEEE*, Alex Alvarado, *Senior Member, IEEE*, Sjoerd van der Heide, *Student Member, IEEE*, Menno van den Hout, *Student Member, IEEE*, Hartmut Hafermann, *Senior Member, IEEE*, and Chigo Okonkwo, *Senior Member, IEEE*

***Abstract*—We propose a new four-dimensional orthant-symmetric 128-ary modulation format (4D-OS128) with a spectral efficiency of 7 bit/4D-sym. The proposed format fills the gap between polarization-multiplexed 8- and 16-ary quadrature-amplitude modulation (PM-8QAM and PM-16QAM). Numerical simulations show that 4D-OS128 outperforms two well-studied 4D geometrically-shaped modulation formats: 128SP-16QAM and 7b4D-2A8PSK by up to 0.65 dB for bit-interleaved coded modulation at the same spectral efficiency. These gains are experimentally demonstrated in a 11×233 Gbit/s wavelength division multiplexing (WDM) transmission system operating at 5.95 bit/4D-sym over 6000 and 9000 km for both EDFA-only and hybrid amplification scenarios, respectively. A reach increase of 15% is achieved with respect to 128-ary set-partitioning 16QAM. Furthermore, the proposed 4D-OS128 is also compared to D_4 lattice-based constellation, 16QAM and probabilistically-shaped 16QAM with finite block-length via simulation.**

***Index Terms*—Coded modulation, generalized mutual information, multidimensional constellations, optical fiber communication, signal shaping.**

Manuscript received March 23, 2020; revised June 16, 2020, August 25, 2020, November 13, 2020, and January 26, 2021; accepted January 29, 2021. Date of publication February 2, 2021; date of current version May 2, 2021. This work was supported in part by Huawei France through the NLCAP Project. The work of Bin Chen was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 61701155, in part by the Fundamental Research Funds for the Central Universities under Grant JZ2020HGTB0015, and in part by Anhui Provincial Natural Science Foundation under Grant 2008085QF282. The work of Alex Alvarado was supported by the Netherlands Organisation for Scientific Research (NWO) via VIDI Grant ICONIC (Project 15685). The work of Sjoerd van der Heide, Menno van den Hout, and Chigo Okonkwo was supported in part by the Dutch NWO Gravitation Program: Research centre for Integrated Nanophotonics under Grant 024.002.033. *(Corresponding author: Bin Chen.)*

Bin Chen is with the School of Computer Science and Information Engineering, Hefei University of Technology, Hefei 230052, China, and also with the Eindhoven University of Technology, 5600 MB Eindhoven, The Netherlands (e-mail: b.c.chen@tue.nl).

Alex Alvarado is with the Information and Communication Theory Lab, Signal Processing Systems Group, Department of Electrical Engineering, Eindhoven University of Technology, 5600 MB Eindhoven, The Netherlands (e-mail: alex.alvarado@ieee.org).

Sjoerd van der Heide, Menno van den Hout, and Chigo Okonkwo are with High Capacity Optical Transmission Laboratory, Eindhoven University of Technology, 5600 MB Eindhoven, The Netherlands (e-mail: s.p.v.d.heide@tue.nl; m.v.d.hout@tue.nl; cokonkwo@tue.nl).

Hartmut Hafermann is with Optical Communication Technology Lab, Paris Research Center, Huawei Technologies France SASU, 92100 Boulogne-Billancourt, France (e-mail: hartmut.hafermann@huawei.com).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/JLT.2021.3056468.

Digital Object Identifier 10.1109/JLT.2021.3056468

I. Introduction

THE demand for higher capacity and longer transmission distances in optical fiber communications has been growing for several years. In order to further support the exponential traffic growth, various multiplexing or modulation dimensions such as time, polarization, wavelength and space (multi-mode/multi-core fibers), have been used. In particular, wavelength division multiplexing (WDM) transport systems with coherent detection have been studied to improve spectral efficiency (SE) through high-order modulation formats. These formats have been shown to be promising for high-speed long-haul transmission systems [1].

Achieving higher SE by employing polarization multiplexed M-ary quadrature amplitude modulation (PM-MQAM) formats with the same forward error correction (FEC) overhead (OH) comes at the cost of a reduced transmission reach and has been extensively studied [2], [3]. In order to maintain or extend transmission distances in high-speed fiber-optic systems, *signal shaping* has recently been the focus of considerable attention in the optical communications community.

Shaping methods can be broadly categorized into probabilistic shaping (PS) and geometric shaping (GS), both having distinct advantages and disadvantages. In PS, long coded sequences induce nonuniform probability distribution on the constellation points [4]–[10]. GS employs a uniform distribution (i.e., equiprobable symbols) on non-equidistant constellation points [11]–[14]. Despite the difference between GS and PS schemes, both techniques are employing the dimensionality to induce a nonuniform distribution via either coding or multidimensional signal space. For the additive white Gaussian noise (AWGN) channel, it has been shown that PS outperforms GS and more closely approaches Shannon's channel capacity when the number of constellation points is limited [15, Sec. 4.3], [16]. On the other hand, GS over multiple dimensions can not only reduce the gap to the Shannon capacity [17]–[21], but could also mitigate the nonlinear effects in the optical channel [22]–[27]. Multidimensional (MD) GS relies only on the selection of the

location of constellation points in a (relatively low) MD space and the design of the corresponding MD detector. MD-GS therefore offers an interesting approach to achieve shaping gains with low implementation complexity instead of using coding approach. MD-GS can also be easily coupled with FEC and only requires straightforward modifications of the mapper and demapper. However, MD-GS also increases the computational complexity of the demapper, as in this case, Euclidean distances for all multidimensional symbols need to be calculated. Nevertheless, it has been shown [28]–[30] that low-complexity MD soft demapper schemes can be designed to achieve good trade-offs between performance and complexity.

Four-dimensional (4D) modulation formats are typically optimized in the four dimensions consisting of the two quadratures (I/Q) and the two polarization states (X/Y) of the optical field. These formats are often designed to achieve large minimum Euclidean distances [17]–[20]. Conventional polarization multiplexed formats are not *true* 4D formats, because they are only optimized in I, Q, X, and Y independently. 4D modulation formats with dependency between dimensions can be obtained by using sphere packing arguments,[1] applying Ungerboeck's set-partitioning (SP) scheme [32] or warping all the points to be nonuniformly spaced coordinates. For most of good spherically bounded constellations, Gray labeling does not exist, and therefore, the combination with binary FEC and bit-interleaved coded modulation (BICM) results in a loss in achievable information rates (AIRs) [33], [34]. Set-partitioning (SP) PM-16QAM has been investigated to achieve fine granularity as 32-ary set-partitioning QAM (32SP-16QAM) [35], [36], 64-ary set-partitioning QAM (64SP-16QAM) [37] and 128-ary set-partitioning QAM (128SP-16QAM) [38]–[40]. 32SP-QAM, 64SP-QAM, and 128SP-QAM have 5, 6, 7 bit/4D-sym, resp., and can be used to achieve different transmission distances. 4D set-partitioned modulation formats based on the $\boldsymbol{D}_4$ lattice have also been investigated by using multilevel coding (MLC) and multi-stage decoding (MSD) in fiber-optical communications [41], [42]. The *warping* technique uses constellation points from a nonuniform grid. This technique has been used to geometrically transform a uniformly spaced constellation into nonuniformly spaced constellation points to cluster points near the center close to each other [43], [44]. However, this method only derives the optimal warping function to reduce the symbol error probability without considering bit labeling. Since the labeling plays an important role for designing geometrically-shaped modulation formats, existing symbol-wise based signal shaping methods are incompatible with bit-wise based coded modulation system.

An alternative to 4D SP-based modulation format is to optimize the coordinates of the 4D symbols as non-regularly spaced signal sets in two polarizations *jointly*. Previous works in this area have mainly investigated the modulation formats design in terms of minimum Euclidean distance [45]. This design criterion has been shown to be suboptimal for the medium signal-to-noise ratio (SNR) regime [26]. Sensitivity gains were achieved in [25]–[27] by optimizing the points in 4D space and also the corresponding binary labeling, by using the information theoretical performance metric called generalized mutual information (GMI) [15, Sec. 4.3], [34], [46]. GMI can be directly connected to modern binary soft-decision forward error correction (SD-FEC) based on BICM [15], [34]. Due to its simplicity and flexibility, BICM is usually considered to be a pragmatic approach to coded modulation (CM) [47] and hence, the use of GMI is preferred for optical fiber communication systems design [46]. In recent experimental demonstrations, the potential of 4D and 8D modulation formats was highlighted. 4D and 8D formats were shown to outperform conventional formats at the same SE of 6 bit/4D-sym [48], [49] and 5.5 bit/4D-sym, resp. Even though these 4D and 8D formats give interesting performance advantages, larger GMI gains (available for larger constellation sizes and higher dimensionality) are difficult to obtain due to the challenging multi-parameter constellation and labelling optimization. Previous works only solve the 4D or 8D GMI optimization problem for up to SE of 6 bits/4D-sym [26], [50].

In this paper, we propose a novel four-dimensional orthant-symmetric 128-ary modulation format (4D-OS128), which has a SE of 7 bit/4D-sym and is obtained via GS by jointly optimizing constellation coordinates and labeling in 4D to maximize GMI. For the design we use the orthant symmetry idea which can significantly reduce the dimensionality of searching space within the first orthant, and overcome the challenging multi-parameter optimization, which can be seen as a trade-off between optimization speed and GMI performance. We found that the obtained 4D format with the orthant symmetry constraint has negligible performance loss with respect to the one without orthant symmetry constraint. The obtained 4D format is compared in terms of linear performance to 128-SP-QAM and a 7 bit modulation in 4D-2A8PSK family (7b4D-2A8PSK) [25], all of them having the same SE (7 bit/4D-sym). In addition, the proposed 4D-OS128 is also compared to $\boldsymbol{D}_4$ lattice-based constellation and probabilistically-shaped 16QAM with finite blocklength via simulation [51]. The transmission performance is investigated by both numerical simulations and experiments. We experimentally demonstrate two amplifier configurations, EDFA-only and hybrid amplification, with a data rate of 233 Gb/s per channel. We target a GMI lower than the SD-FEC threshold of 5.95 bit/4D-sym [25],[2] which corresponds to a FEC rate of 0.8 (25% FEC OH). For the baseline constellations, distances around 5000 km (EDFA-only) and 8000 km (hybrid amplification) are therefore targeted. Compared at the same bit rate, the proposed 4D-OS128 format achieves a 15% longer transmission reach than 128SP-16QAM and 7b4D-2A8PSK for 11 WDM channels transmission.

This paper is organized as follows. In Sec. II, the design methodology and the proposed modulation format are introduced. In Sec. III, numerical results are shown for both AWGN and nonlinear optical fiber simulations. The experimental setup

[1] An excellent summary of MD constellations is given in the online database [31]

[2] Note that other thresholds could be used, e.g., the thresholds in [52, Table III] The use of different thresholds will however not change the general conclusions of reach increase in this paper.

of the WDM optical fiber system and the experimental results are described in Sec. IV. Conclusions are drawn in Sec. V.

II. 4D Modulation Format and Optimization

A. GS Optimization: General Aspects

The optical channel suffers from the interactions between amplified spontaneous emission (ASE) noise, dispersion and Kerr nonlinearities, which are usually classified as inter-channel and intra-channel effects. This channel can be modeled by a conditional PDF $p_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}$, where $\underline{\boldsymbol{X}}$ and $\underline{\boldsymbol{Y}}$ are the transmitted and received sequences, respectively. The transmitted symbols $\boldsymbol{X}$ in $\underline{\boldsymbol{X}}$ are assumed to be MD symbols with N real dimensions (or equivalently, with $N/2$ complex dimensions) drawn uniformly from a discrete constellation $\mathcal{X}$ with cardinality $M = 2^m = |\mathcal{X}|$. The most popular case in fiber optical communications is $N = 4$, which corresponds to coherent optical communications using two polarizations of the light (4 real dimensions). This naturally results in four-dimensional (4D) modulation formats.

In general, the channel law $p_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}$ shows memory across multiple symbols, which is introduced by the fiber optical channel even after dispersion compensation.[3] From now on, however, we consider a channel law $p_{\boldsymbol{Y}|\boldsymbol{X}}$ where the output symbols $\boldsymbol{Y}$ have also $N/2$ complex dimensions. Due to this assumption, we are actually only approximating the true optical channel. As we will explain below, this approximation is well-matched to the fact that typical optical receivers ignore potential memory across 4D symbols.

Throughout this paper, we consider BICM, which is one of the most popular CM schemes. The transmitted symbols $\boldsymbol{X}$ are jointly modulated in 4D space by a set of constellation coordinates and the corresponding labeling strategy. The ith constellation point is denoted by $\boldsymbol{s}_i = [s_{i1}, s_{i2}, s_{i3}, s_{i4}] \in \mathcal{R}^4$ with $i = 1, 2, \ldots, M$ in four real dimensions. We use the $M \times 4$ matrix $\mathbb{S} = [\boldsymbol{s}_1; \boldsymbol{s}_2; \ldots; \boldsymbol{s}_M]$ to denote the 4D constellation.[4] The ith constellation point $\boldsymbol{s}_i$ is labeled by the length-m binary bit sequence $\boldsymbol{b}_i = [b_{i1}, \ldots, b_{im}] \in \{0, 1\}^m$. The binary labeling matrix is denoted by a $M \times m$ matrix $\mathbb{B} = [\boldsymbol{b}_1; \boldsymbol{b}_2; \ldots; \boldsymbol{b}_M]$, which contains all unique length-m binary sequences. The 4D constellation and its binary labeling are fully determined by the pair of matrices $\{\mathbb{S}, \mathbb{B}\}$.

In this paper, we are interested in minimizing the SNR requirements for a target GMI, but it can be also equivalent to maximizing the GMI for a given SNR. The transmitted bits are assumed to be independent and uniformly distributed, which implies uniform symbols $\boldsymbol{X}$. The receiver assumes a memoryless channel and also uses a bit-metric decoder (i.e., a standard BICM receiver). In this case, the receiver uses a decoding metric $\mathbb{q}(\boldsymbol{y}, \boldsymbol{c})$ proportional to the product of the bit-wise metrics, i.e., the decoding metric is

$$\mathbb{q}(\boldsymbol{y}, \boldsymbol{c}) \propto \prod_{k=1}^{m} p_{\boldsymbol{Y}|C_k}(\boldsymbol{y}|c_k) \tag{1}$$

$$\propto \prod_{k=1}^{m} \sum_{b\in\{0,1\}} \sum_{j\in\mathcal{I}_k^b} p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}_j), \tag{2}$$

where $\boldsymbol{C} = [C_1, C_2, \ldots, C_m]$ is the random vector representing the transmitted bits, and $\mathcal{I}_k^b \subset \{1, 2, \ldots, M\}$ with $|\mathcal{I}_k^b| = M/2$ is the set of indices of constellation points whose binary label is b at bit position k.

Under the assumptions above, the GMI can be expressed as

$$G(\mathbb{S}, \mathbb{B}, p_{\boldsymbol{Y}|\boldsymbol{X}}) = \sum_{k=1}^{m} I(C_k; \boldsymbol{Y}), \tag{3}$$

where $I(C_k; \boldsymbol{Y})$ is the mutual information (MI) between the bits and the symbols, and where the notation $G(\mathbb{S}, \mathbb{B}, p_{\boldsymbol{Y}|\boldsymbol{X}})$ emphasizes the dependency of the GMI on the constellation, binary labeling, and channel law. Furthermore, for any N-dimensional channel law, (3) can be expressed as [46, eqs. (17)–(18)]

$$G(\mathbb{S}, \mathbb{B}, p_{\boldsymbol{Y}|\boldsymbol{X}}) = m + \frac{1}{M} \sum_{k=1}^{m} \sum_{b\in\{0,1\}} \sum_{i\in\mathcal{I}_k^b} \int_{\mathcal{R}^N} p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}_i) \log_2 \frac{\sum_{j\in\mathcal{I}_k^b} p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}_j)}{\sum_{j'=1}^{M} p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}_{j'})} \mathrm{d}\boldsymbol{y}. \tag{4}$$

As shown in (4), a GMI optimization requires a jointly optimization of the 4D coordinates and its binary labeling. A GMI-based optimization finds a constellation $\mathbb{S}^*$ and labeling $\mathbb{B}^*$ for a given channel conditional PDF $p_{\boldsymbol{Y}|\boldsymbol{X}}$ and energy constraints, i.e.,

$$\{\mathbb{S}^*, \mathbb{B}^*\} = \mathrm{argmax}_{\mathbb{S},\mathbb{B}:E[\|\boldsymbol{X}\|^2]\le\sigma_x^2} G(\mathbb{S}, \mathbb{B}, p_{\boldsymbol{Y}|\boldsymbol{X}}), \tag{5}$$

where σ_x^2 represents the transmitted power, $\mathbb{S}^*$ and $\mathbb{B}^*$ indicate the optimal constellation and labeling, resp.

Note that for any given channel $p_{\boldsymbol{Y}|\boldsymbol{X}}$, the optimization problem in (5) is a single objective function G with multiple parameter and constraints. From previous works [26], [50], it is known that the constellation optimization and GMI calculation for large constellations and/or for constellations with high dimensionality is computationally demanding. Therefore, an unconstrained optimization is very challenging. Unconstrained formats also impose strict requirements for the generation of the signals (i.e., high-resolution digital-to-analog converter (DAC)) as well as complex MD detectors.

B. Orthant-Symmetric (OS) Geometric Shaping Optimization

To solve the multi-parameter optimization challenges described above and to reduce the transceiver requirements, we propose to impose an "orthant symmetry" constraint to the N-dimensional modulation format to be designed. Our proposed approach makes the MD format to be generated

[3]One empirical model that properly takes this effect into account for channel capacity calculations is the so-called finite-memory GN model [53] Another example is the time-domain perturbation models in [54]–[56], where the received symbol depends on previous and future transmitted symbols (potentially across other polarizations and channels). The memory can also be considered in the demapper over multiple consecutive time slots [57].

[4]Notation convention: Random (row) vectors are denoted by $\boldsymbol{X}$ and its corresponding realization are denoted by $\boldsymbol{x}$. Matrices are denoted by $\mathbb{X}$. A semicolon is used to denote vertical concatenation of vectors, e.g., $\mathbb{X} = [\boldsymbol{x}_1; \boldsymbol{x}_2] = [\boldsymbol{x}_1^T, \boldsymbol{x}_2^T]^T$, where $[\cdot]^T$ denotes transpose.

from a first-orthant labeled constellation (see *Definition 1* below). These concepts are defined in what follows, and are based on N-dimensional orthants, defined as the intersection of N mutually orthogonal half-spaces passing through the origin. By selecting the signs of the half-spaces, the 2^N orthants available in an N-dimensional space can be obtained.

Let $\mathbb{L}_q = [\boldsymbol{l}_1; \boldsymbol{l}_2; \ldots; \boldsymbol{l}_{2^q}]$ with $\boldsymbol{l}_j \in \{0,1\}^q$ and $j = 1, 2, \ldots, 2^q$ denote a $2^q \times q$ *labeling matrix of order* q, which contains all unique length-q binary vectors. Let $\mathbb{H}_k$ be a $N \times N$ mirror matrix defined as

$$\mathbb{H}_k = \begin{bmatrix} (-1)^{l_{k1}} & 0 & \ldots & 0 \\ 0 & (-1)^{l_{k2}} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & (-1)^{l_{kN}} \end{bmatrix}, \tag{6}$$

with $k = 1, 2, \ldots, 2^N$ and $\boldsymbol{l}_k = [l_{k1}, l_{k2}, \ldots, l_{kN}]$ are the rows of the labeling matrix of order N $\mathbb{L}_N$.

Example 1 (Mirror matrices): For the 1D case ($N = 1$), there are only two orthants (an orthant is a ray), the mirror matrices are $\mathbb{H}_1 = 1$ and $\mathbb{H}_2 = -1$. For the 2D case ($N = 2$), there are four orthants (an orthant is a quadrant), the mirror matrices are

$$\mathbb{H}_1 = \begin{bmatrix} +1 & 0 \\ 0 & +1 \end{bmatrix}, \mathbb{H}_2 = \begin{bmatrix} -1 & 0 \\ 0 & +1 \end{bmatrix} \tag{7}$$

$$\mathbb{H}_3 = \begin{bmatrix} +1 & 0 \\ 0 & -1 \end{bmatrix}, \mathbb{H}_4 = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}. \tag{8}$$

For the $N = 3$ case, there are $2^3 = 8$ orthants (an orthant is a octant), the eight mirror matrices which can be obtained by (6), with $\boldsymbol{l}_j \in \{0,1\}^3$, and $j = 1, 2, \ldots, 8$.

We now define a first-orthant and orthant-symmetric (OS) labeled constellations.

Definition 1 (First-orthant labeled constellation): The pair of matrices $\{\mathbb{T}, \mathbb{L}_{m-N}\}$ is said to be a first-orthant labeled constellation if $\mathbb{T} = [\boldsymbol{t}_1; \boldsymbol{t}_2; \ldots; \boldsymbol{t}_{2^{m-N}}]$ is a constellation matrix such that $\boldsymbol{t}_j \in \mathcal{R}_+^N, \forall j \in \{1, 2, \ldots, 2^{m-N}\}$ (all entries are nonnegative), and $\mathbb{L}_{m-N}$ is a labeling matrix of order $m - N$. □

Definition 2 (Orthant-symmetric labeled constellation): The pair of matrices $\{\mathbb{S}, \mathbb{B}\}$ is said to be an OS labeled constellation if $\mathbb{S} = [\mathbb{S}_1; \mathbb{S}_2; \ldots; \mathbb{S}_{2^N}]$ is a $2^m \times N$ constellation matrix and $\mathbb{B} = [\mathbb{B}_1; \mathbb{B}_2; \ldots; \mathbb{B}_{2^N}]$ is a $2^m \times m$ binary labeling matrix, where the constellation matrix $\mathbb{S}$ is constructed via $\mathbb{S}_k = \mathbb{T}\mathbb{H}_k$ with $k = 1, 2, \ldots, 2^N$ and $\mathbb{H}_k$ is given by (6), and the labeling matrix $\mathbb{B}$ is such that $\mathbb{B}_k = [\mathbb{O}_k, \mathbb{L}_{m-N}]$, where $\mathbb{O}_k = [\boldsymbol{l}_k; \boldsymbol{l}_k; \ldots; \boldsymbol{l}_k]$ and $\boldsymbol{l}_k$ are the rows of $\mathbb{L}_N$, with $k = 1, 2, \ldots, 2^N$, and $\mathbb{L}_{m-N}$ is a labeling matrix of order $m - N$. □

For an N-dimensional OS 2^m-ary constellation in *Definition 2*, each orthant contains 2^{m-N} constellation points. The first 2^{m-N} constellation points correspond to the points in $\mathbb{T} = [\boldsymbol{t}_1; \boldsymbol{t}_2; \ldots; \boldsymbol{t}_{M/2^N}]$ (see *Definition 1*), and are found via $\mathbb{S}_1 = \mathbb{T}\mathbb{H}_1 = \mathbb{T}$ (see in *Definition 2*). The constellation points in $\mathbb{S}_k$ with $k > 1$ are generated by "mirroring" the first-orthant points in $\mathbb{S}_1$ to other orthant via $\mathbb{S}_k = \mathbb{T}\mathbb{H}_k$. The binary labeling for the proposed N-dimensional OS formats is such that the constellation points in a given orthant $\mathbb{S}_k$ use the binary labeling $\mathbb{L}_{m-N}$, i.e., $m - N$ bits are used within an orthant. The remaining N bits (which define the matrix $\mathbb{O}_k$ in *Definition 2*) are the bits used to select the orthant. To clarify this general definition, we present now two examples. We focus on traditional square 2^m-ary QAM constellations labeled by the binary-reflected Gary code [58] (in one and two polarizations), which are shown to belong to the class of OS labeled constellations.

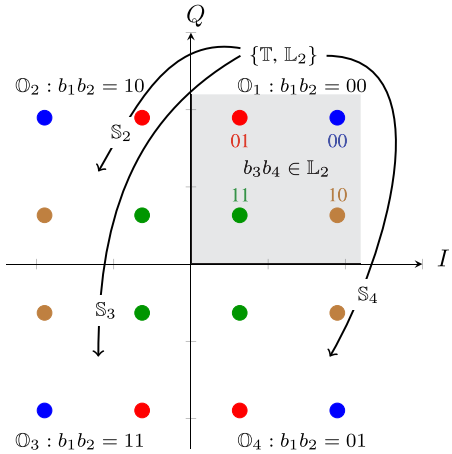


Fig. 1. OS labeled constellation example of 16QAM: $\{\mathbb{T}, \mathbb{L}_2\}$ is a first-ortant labeled constellation and $\mathbb{O}_k$ are the bits used to represent the kth orthant. The lines show the mirroring operation to obtain the points $\mathbb{S}_k$ in other orthant.

Example 2 (16QAM): Fig. 1 shows a 16QAM constellation ($N = 2$, $m = 4$) labeled by the binary-reflect Gray code independently in the first and second dimension. 16QAM is quadrant-symmetric (four quadrants) and consists of the same number of signal points (four points) in each quadrant. The corresponding *first-quadrant (orthant) labeled constellation* is

$$\mathbb{S}_1 = \mathbb{T} = \begin{bmatrix} 3 & 3 \\ 1 & 3 \\ 1 & 1 \\ 3 & 1 \end{bmatrix}, \mathbb{B}_1 = [\mathbb{O}_1, \mathbb{L}_2] = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

The constellation matrix $\mathbb{S}$ and labeling matrix $\mathbb{B}$ of 16QAM satisfy $\mathbb{S}_k = \mathbb{T}\mathbb{H}_k$ and $\mathbb{B}_k = [\mathbb{O}_k, \mathbb{L}_{m-N}]$ with $k = 1, 2, 3, 4$ in Definition 2, and thus, this format is an OS labeled constellation.

Example 3 (PM-16QAM): Consider a polarization-multiplexed QAM constellation (PM-16QAM, $N = 4$, $m = 8$) labeled by the binary-reflect Gray code, independently in each real dimension. This constellation is obtained as a Cartesian product of two 2D-16QAM, and therefore, PM-16QAM consists of 16 orthants with 16 points in each orthant. The *first-orthant*

labeled constellation $\{\mathbb{S}_1, \mathbb{B}_1\}$ of PM-16QAM is given by

$$\mathbb{S}_1 = \mathbb{T} = \begin{bmatrix} 1&1&1&1 \\ 1&3&1&1 \\ 3&3&1&1 \\ 3&1&1&1 \\ 1&1&1&3 \\ 1&3&1&3 \\ 3&3&1&3 \\ 3&1&1&3 \\ 1&1&3&3 \\ 1&3&3&3 \\ 3&3&3&3 \\ 3&1&3&3 \\ 1&1&3&1 \\ 1&3&3&1 \\ 3&3&3&1 \\ 3&1&3&1 \end{bmatrix}, \mathbb{B}_1 = [\mathbb{O}_1, \mathbb{L}_4] = \begin{bmatrix} 0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&1&0&0 \\ 0&0&0&0&1&1&0&0 \\ 0&0&0&0&1&0&0&0 \\ 0&0&0&0&0&0&0&1 \\ 0&0&0&0&0&1&0&1 \\ 0&0&0&0&1&1&0&1 \\ 0&0&0&0&1&0&0&1 \\ 0&0&0&0&0&0&1&1 \\ 0&0&0&0&0&1&1&1 \\ 0&0&0&0&1&1&1&1 \\ 0&0&0&0&1&0&1&1 \\ 0&0&0&0&0&0&1&0 \\ 0&0&0&0&0&1&1&0 \\ 0&0&0&0&1&1&1&0 \\ 0&0&0&0&1&0&1&0 \end{bmatrix}.$$

It can be shown that the symmetries of PM-16QAM and its binary labeling makes it also to be an OS modulation format.

Now that we have defined OS labeled constellations, we turn our attention back to the GMI optimization in (5). Because of the OS of the formats under consideration, the new optimization problem is

$$\{\mathbb{T}^*, \mathbb{L}^*_{m-N}\} = \underset{\mathbb{T}, \mathbb{L}_{m-N}: E[\|X\|^2] \leq \sigma_x^2}{\operatorname{argmax}} G(\mathbb{S}, \mathbb{B}, p_{\boldsymbol{Y}|\boldsymbol{X}}), \quad (9)$$

whose solution $\{\mathbb{T}^*, \mathbb{L}^*_{m-N}\}$ can then be used to obtain the complete labeled constellation $\{\mathbb{S}, \mathbb{B}\}$ using *Definition 2*. We emphasize that in (9), the GMI is still dependent on the labeled constellation $\{\mathbb{S}, \mathbb{B}\}$ (which is fully defined by 2^m N-dimensional coordinates and a labeling of order m), while the optimization is performed only over the 2^{m-N} N-dimensional coordinates in the matrix $\mathbb{T}$ and the labeling of order $m-N$ (i.e., the matrix $\mathbb{L}_{m-N}$). This reduction in the dimensionality of the search space is what will allow us to design a MD format with higher dimensionalities and higher SE for maximizing GMI.

Apart from lowering the optimization complexity, another motivation for using the OS property is that the GMI-based constellation optimization can to obtain a Gray-like mapping. It is well-known that Gray labelings (for which adjacent constellation points differ in only one bit difference) can reduce the loss between MI and GMI [59, Fig. 4], [60], [61], [62, Sec. IV]. However, since the length of the bit label is shorter than the number of nearest neighbors, it is not possible to obtain a Gray mapping and also these constellations also lack an obvious Gray-like mapping. In order to reduce the non-Gray mapping penalty with respect to Gray mapping, the orthant-symmetric structure can first guarantee the orthants are Gray-labeled and attempt to make all the neighbouring symbols in the same orthant with one bit difference to have a larger distance.

C. Proposed 4D Format: 4D-OS128

In this paper, we focus on designing a four dimensional orthant-symmetric 128-ary (4D-OS128) constellation ($N = 4$ and $M = 128$) with SEs of 7 bit/4D-sym ($m = 7$), which consists of $M = 2^m$ points $\boldsymbol{s}_i, i \in \{1, 2, \ldots, 128\}$ labeled by 7 bits $\boldsymbol{b}_i = [b_1, b_2, \ldots, b_7]$. For the 4D-OS128 constellation, each orthant contains $2^{m-N} = 8$ constellation points, and therefore the 8 constellation points in first orthant are considered as the first-orthant labeled constellation $\mathbb{T} = [\boldsymbol{t}_1; \boldsymbol{t}_2; \ldots; \boldsymbol{t}_8]$ with labeling matrix $\mathbb{G}_3$. The jth first-orthant point is denoted by $\boldsymbol{t}_j = [t_{j1}, t_{j2}, t_{j3}, t_{j4}] \in \mathcal{R}_+^4$ with $j = \{1, 2, \ldots, 8\}$. The first orthant is labeled by four binary bits $[b_{j1}, b_{j2}, b_{j3}, b_{j4}] = \boldsymbol{l}_1$. The remaining 3 bits $[b_{j5}, b_{j6}, b_{j7}]$ determine the point $\boldsymbol{t}_j$ in the corresponding orthant.

For the 4D-OS128 constellation, only eight 4D coordinates in the matrix $\mathbb{T}$ and the corresponding binary labeling of order 3 $\mathbb{L}_3$ need to be optimized in (9). Note that there are many local optima for large constellations and/or for a constellation with high dimensionality, which has been reported by [50]. To overcome this problem, we solve the optimization problem (9) by applying the approach recently published in [63], which implement autoencoders to optimize the constellation and the binary switch algorithm [64] to find the best "swap" of two binary labels. [5]

We optimized the 4D constellation at the target GMI of $0.85m$ bit/4D-sym, where m is the number of bits transmitted per symbol. In other words, the 4D-OS128 modulation was optimized to minimize the SNR requirements for a GMI of $0.85 \times 7 = 5.95$ bit/4D-sym. The obtained 4D-OS128 modulation format has 128 nonoverlapping points in 4D space. For better visualization, these points can be projected on the two polarizations. This projection results in 20 distinct points in each 2D space, as shown in Fig. 2(a). In order to clearly show the inter-polarization dependency, we use a similar color coding strategy as in [26]: 2D projected symbols in the first and second polarization are valid 4D symbols only if they share the same color. The coordinates of the 8 vectors defining the matrix $\mathbb{T}$ are

$$\begin{aligned} \boldsymbol{t}_j \in \{&[\pm t_1, \pm t_1, \pm t_3, \pm t_3], [\pm t_2, \pm t_5, \pm t_3, \pm t_3], \\ &[\pm t_5, \pm t_2, \pm t_3, \pm t_3], [\pm t_4, \pm t_4, \pm t_3, \pm t_3], \\ &[\pm t_3, \pm t_3, \pm t_1, \pm t_1], [\pm t_3, \pm t_3, \pm t_2, \pm t_5], \\ &[\pm t_3, \pm t_3, \pm t_5, \pm t_2], [\pm t_3, \pm t_3, \pm t_4, \pm t_4]\}, \end{aligned} \quad (10)$$

where $t_5 > t_4 > t_3 > t_2 > t_1 > 0$. These 8 points are represented in Fig. 2(a) with gray, red, brown, blue, green, magenta, orange and cyan markers in the shadow area, respectively. The points with the same color in other orthants can be obtained $\mathbb{S}_k = \mathbb{T}\mathbb{H}_k$, and therefore, the proposed format is highly symmetric in 16 orthants.

The color coding scheme used in Fig. 2(a) also shows the binary labeling: 4 out of 7 bits $[b_{j1}, b_{j2}, b_{j3}, b_{j4}]$ determine the 16 orthant, and the remaining 3 bits $[b_{j5}, b_{j6}, b_{j7}]$ determine the 8 constellation points in the corresponding orthant. In other words, $[b_{j5}, b_{j6}, b_{j7}]$ determines the color of the transmitted points in Fig. 2(a), while $[b_{j1}, b_{j2}, b_{j3}, b_{j4}]$ determine the coordinate of the point in the same color. Fig. 2(b) shows an example of 4D mapping and demapping with $[b_{j5}, b_{j6}, b_{j7}] = [0, 0, 1]$, which indicates that one of the red 4D points in Fig. 2(b) is selected as the transmitted symbol $\boldsymbol{x}$.

[5]Note that initial constellation does affect the time until convergence. In this paper, we use 128SP-16QAM as initial constellation.

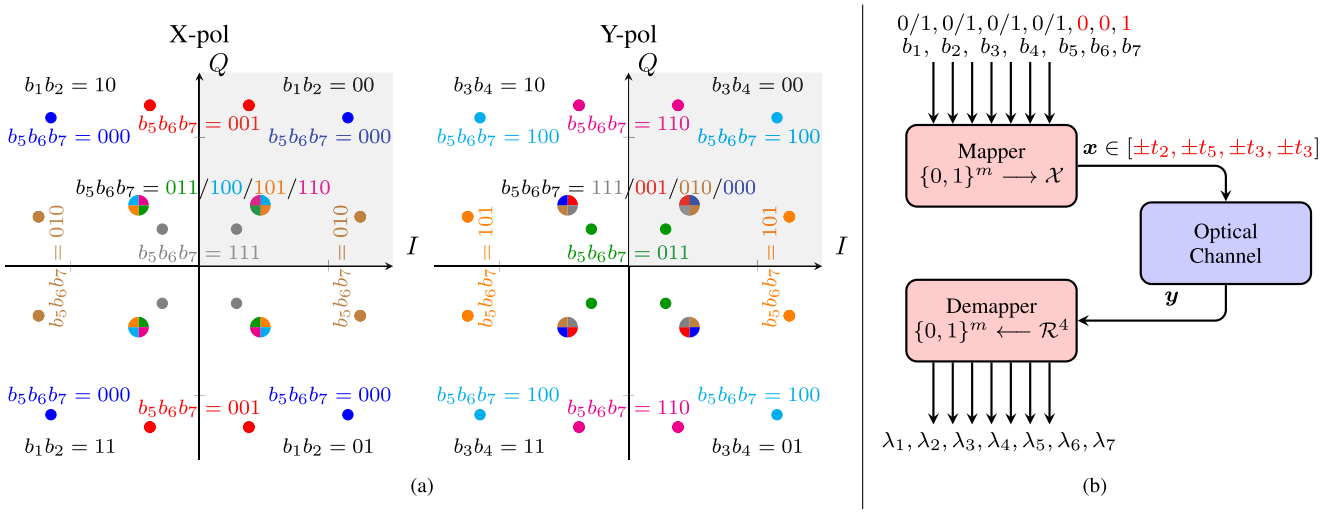


Fig. 2. (a) $2\times$ 2D-projections of the designed 4D-OS128 modulation and associated bits/polarization bits mapping. The symbols in the same color in two 2D projection indicate a 4D symbol in 4D-OS128 modulation. The shadow area indicate the first orthant in 4D space. (b) Mapping and demapping example: sixteen possible 7 bit sequences are mapped to the transmitted 4D symbol $\boldsymbol{x}$ (red points in two 2D projection) with the coordinates $[\pm t_2, \pm t_5, \pm t_3, \pm t_3]$ and the received 4D symbol $\boldsymbol{y}$ is demapped to log-likelihood ratios (LLRs) $\boldsymbol{\lambda} = [\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6, \lambda_7]$.

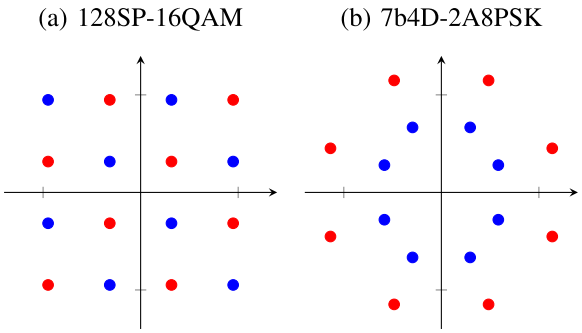


Fig. 3. 2D-projections of two 4D modulation formats: 128SP-16QAM and 7b4D-2A8PSK. The colors indicate symbol selecting strategy: the symbols indicated in blue and red are assigned to either of the polarizations.

In this paper, the 4D-OS128 modulation format indicates one of the 4D orthant-symmetric modulations with 128-ary, but optimized for the specific SNR of 9.5 dB. The coordinates of the 4D-OS128 modulation format and the corresponding binary labeling are given in Appendix (see Table II).

D. Comparison With Other 4D Modulation Formats

Here, we include two other well-known 4D 128-ary formats for comparisons at the same SE of 7 bit/4D-sym as GS modulation formats baseline. The first constellation is 128-ary set-partitioning 16QAM (128SP-16QAM) (see Fig. 3(a)), which has been demonstrated for optical communications systems in [38]–[40]. The second one is a 7 bit modulation in the 4D-2A8PSK family (7b4D-2A8PSK) [25] (see Fig. 3(b)), which is currently used in commercial programmable transponders [65]. The inner ring/outer ring ratio of 0.59 is chosen for 7b4D-2A8PSK to maximize the GMI performance as described in [25]. Note that 128SP-16QAM and 4D-2A8PSK are set-partitioned from 16QAM and 16-ary two-ring 8PSK, respectively. 128SP-16QAM is obtained by selecting points with larger Euclidean distance. However, for 4D-2A8PSK, constellation points with constant modulus are selected and the ring ratio of two-ring 8PSK is also optimized for maximizing GMI at the same time. Therefore, only 7b4D-2A8PSK are designed for an SNR region corresponding to SD-FEC overhead between 20% and 25%

To better understand the GMI performance and tolerance to fiber nonlinearities, we conduct a comparison in terms of energy per transmitted symbol, peak-to-average power ratio (PAPR), variance of the signal energy with respect to its meanaverage energy $\sigma^2_{P_s} = E[\|\boldsymbol{S}\|^2 - E[\|\boldsymbol{S}\|^2]]$, squared Euclidean distance (SED), and the numbers of pairs of constellation points at minimum squared Euclidean distance (MSED). While PAPR and energy per transmitted symbol can be seen as a rough indication of nonlinearity tolerance [66], SED and the number of pairs at MSED can be argued to be performance indicators for the AWGN channel. The analysis and discussion below only gives an intuition on the performance of the proposed format. A precise comparison of these modulation formats will be presented in both the linear and nonlinear channels by numerical simulations (Sec. III) and experimental results (Sec. IV).

We assume that all the constellations are normalized to $E_s = 2$ (i.e., unit energy per polarization). Under this assumption, the comparison of energy for each 4D symbol after sorting for all three modulation formats are shown Fig. 4. The 7b4D-2A8PSK format has a constant-modulus property, and thus, it can significantly reduce the nonlinear interference noise (NLIN) [25]. For 128SP-16QAM, 5 energy levels are visible, while the proposed 4D-OS128 shows only 3.

Table I shows four properties of four modulation formats under consideration. Three of them have SE of 7 bit/4D-sym. PM-16QAM with SE of 8 bit/4D-sym is also compared as baseline. In the first two columns, we use two performance

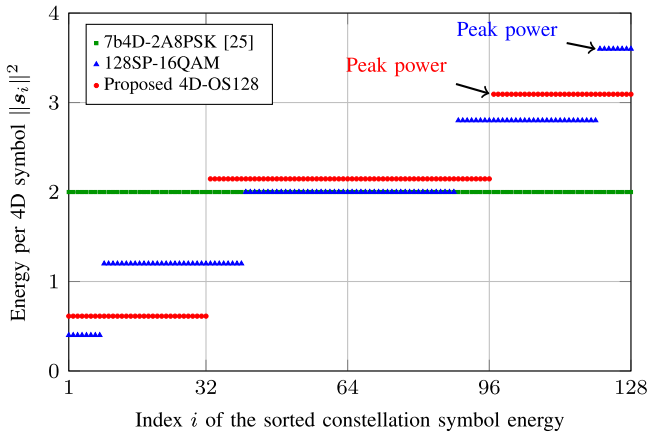


Fig. 4. Energy per 4D symbol for three SE = 7 bits/4D-sym modulations. The variation of the transmitted symbols' energy can contribute to the nonlinear interference noise (NLIN).

TABLE I
COMPARISON OF 7 BIT/4D-SYM AND 8 BIT/4D-SYM MODULATION FORMATS

Mod. Formats	PAPR [dB]	$\sigma^2_{P_s}$	d^2	n_d
4D-128SP-16QAM	2.55	0.645	0.8	864
7b4D-2A8PSK	0	0	0.23	64
4D-OS128	1.89	0.797	0.14	16
PM-16QAM	2.55	0.643	0.4	768

metrics to compare the modulation-dependent nonlinear interference: PAPR and $\sigma^2_{P_s}$ for a given modulation format. Due to the constant-modulus property, both of these two performance metrics are zero for 7b4D-2A8PSK, which is expected to be better than the other two modulation formats in terms of effective SNR. Since PAPR only depends on the few constellation points with largest energy, it cannot reflect the complete nonlinear performance. In contrast, $\sigma^2_{P_s}$ is the variation of all the possible transmitted symbols' energy, and thus, smaller $\sigma^2_{P_s}$ should in principle result in higher nonlinear noise tolerance. We expect that the small difference among 4D-128SP-16QAM, 4D-OS128 and PM-16QAM in terms of PAPR and $\sigma^2_{P_s}$ will contribute similar modulation-dependent nonlinear interference, and thus give similar effective SNR. These predictions will be confirmed in Sec. III and Sec. IV.

In addition to the nonlinear noise tolerance property, we study the structure of the formats in terms of MSED, which we denote by d^2. We also look at the number of pairs of constellation points at MSED, which we denote as n_d. These two parameters are shown in the last two columns of Table I. A large d^2 and small n_d should in principle result in high MI in the high-SNR regime, as recently proved in [61][6]. Even though the proposed 4D-OS128 has the smallest d^2, it has only 16 pairs at MSED. 4D-128SP-16QAM and 7b4D-2A8PSK have a large number of pairs of constellation points at MSED, which will degrade the GMI performance at medium SNR range. To better understand this, we also study the SED "spectrum" for the three constellations. This is shown as a histograms in Fig. 5. It has been shown in [26] that GMI does not only depend on d^2 and n_d, but also the Hamming distance (HDs) of the binary labels of the constellation points at MSED. This figure also shows a classification of the pairs at a given SED: blue bars for pairs at HD larger than one, and red bars for pairs at HD one. From the SED spectra in Fig. 5, we can see that there are less pairs at lower SED and most of the pairs are at Hamming distance one for the proposed 4D-OS128, which in principle results in a better GMI in medium SNR range.

[6]The results in [61] hold for 1D constellations only. However, the authors in [61] conjectured that the results holds verbatim for any number of dimensions.

III. SIMULATION RESULTS

A. Linear Performance

Considering the 128SP-16QAM and 7b4D-2A8PSK as the baselines, Fig. 6 shows the linear performance in terms of GMI for the proposed 4D-OS128 modulation format. The results in Fig. 6 indicates that 4D-OS128 can provide gains of 0.65 dB at GMI of 5.95 bit/4D-sym over 128SP-16QAM and 7b4D-2A8PSK. The shaping gain comes from the joint optimization of 4D coordinates $\mathbb{S}$ and its binary labeling $\mathbb{B}$ under the constraints of average power and orthant-symmetry. Meanwhile, the proposed 4D-OS128 can provide 0.27 bit/4D-sym gain over 128SP-16QAM at SNR=9.5 dB. Eventhough 7b4D-2A8PSK performs better than 128SP-16QAM for SNR below 10 dB, there is at least a gap of 0.5 dB between 7b4D-2A8PSK and 4D-OS128.

In order to verify whether a GMI loss is induced by using the orthant symmetry constraint, we further optimize the 4D 128-ary constellation by using (5) with 4D-OS128 as an initial constellation and removing the orthant-symmetric constraint.[7] The optimization process can reach steady state within 2000 steps as shown in the inset (b) of Fig. 6. The optimized 4D geometrically-shaped (4D-GS128) constellation without orthant-symmetric constraint is plotted in the inset (a) of Fig. 6.[8] The red and blue circles represent the symbols transmitted in X and Y polarization respectively. Despite the 0.033 bit/4D-sym improvement provided by the optimization, the result is a constellation where symbols are very close to each other in the 2D space. This rather complex constellation becomes particularly challenging to generate using a high-speed DAC with limited effective number of bits (ENOB). In this paper, we only investigate the performance of the proposed orthant-symmetric 4D-OS128 modulation format.

For verifying GMI results, we use LDPC codes from the DVB-S2 standard with code rates $R \in \{0.83, 0.8\}$ (20% and 25% OH) and blocklength $N = 64\,800$. Fig. 7 shows post-FEC BER of BER=$4.5 \cdot 10^{-3}$, between 0.55 dB and 0.65 dB, which is in excellent agreement with the prediction of the GMI. Moreover, we also evaluate the efficiency of the 4D modulations by using low complexity max-log demapper (MaxLog), which significantly reduces the computational complexity by avoiding logarithmic and exponential functions as opposed to the optimum Maximum likelihood (ML) demapper. Note that using a

[7]Note that this process is not guaranteed to find the globally optimum constellation.

[8]Note that 4D-OS128 is also a 4D geometrically-shaped modulation, but with an additional orthant symmetry constraint.

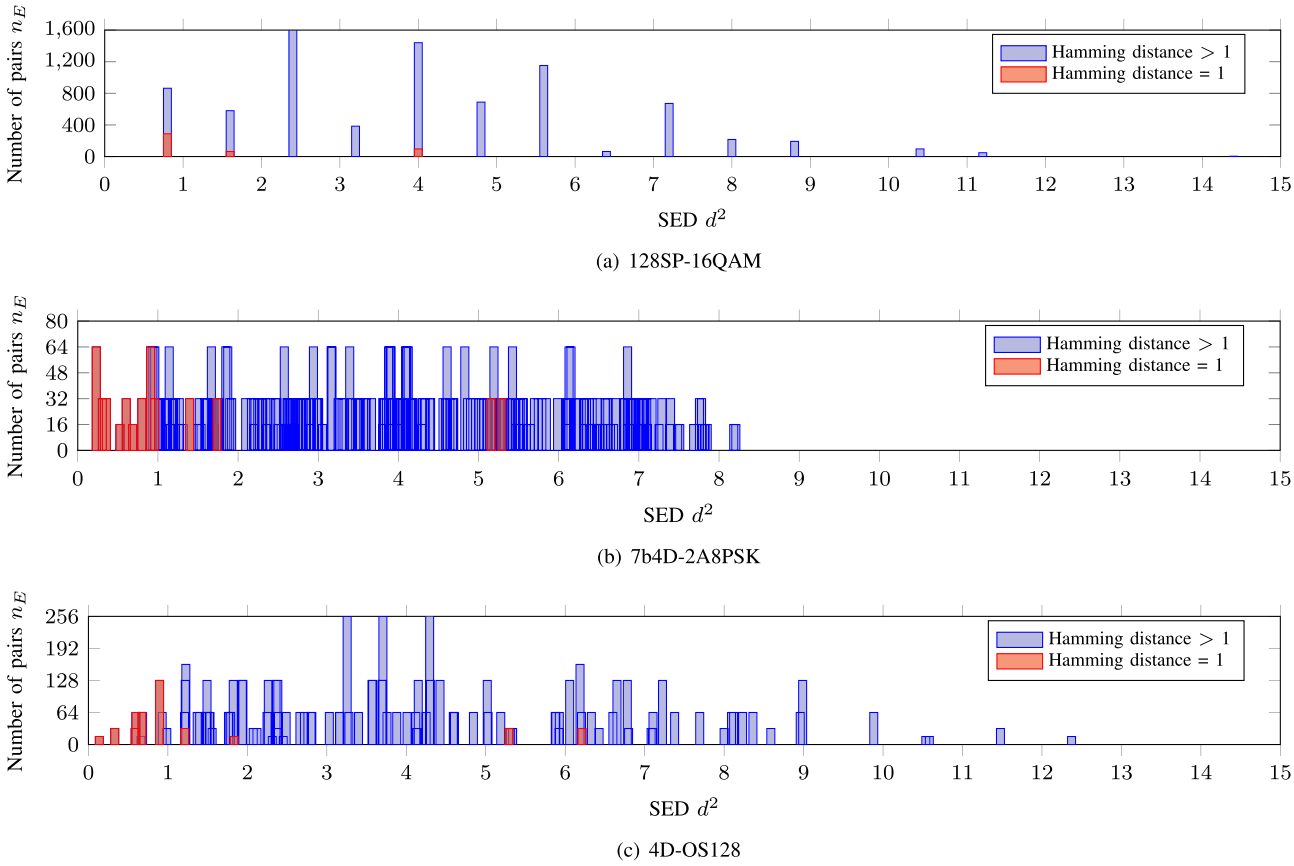


Fig. 5. Histograms of SEDs of three 4D formats: (a) 4D-128SP-16QAM, (b) 7b4D-2A8PSK, and (c) 4D-OS128. The red bars show the number of pairs with Hamming distance of 1 at the SED d^2. The MSEDs are 0.8, 0.23, and 0.14, for 4D-128SP-16QAM, 7b4D-2A8PSK, and 4D-OS128, resp.

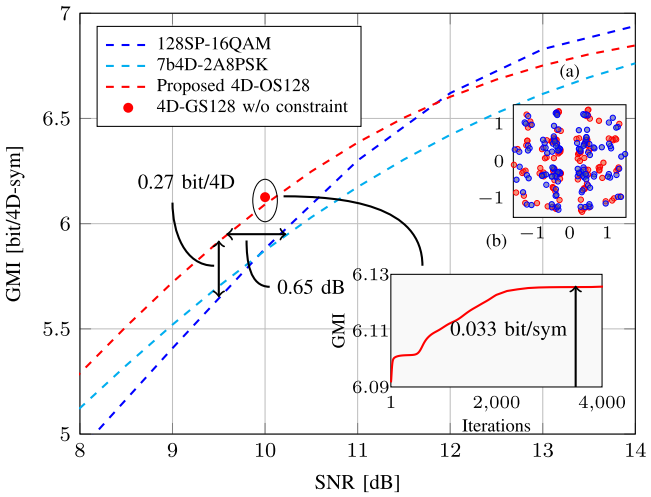


Fig. 6. GMI vs. SNR for three SE = 7 bits/4D-sym modulations. Inset: (a) Optimized constellation of 4D-GS128 w/o orthant-symmetric constraint (b) The optimization procedure after removing orthant-symmetric constraint.

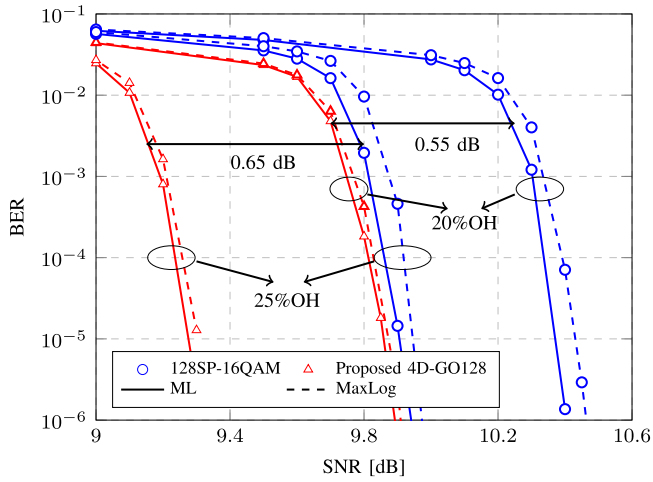
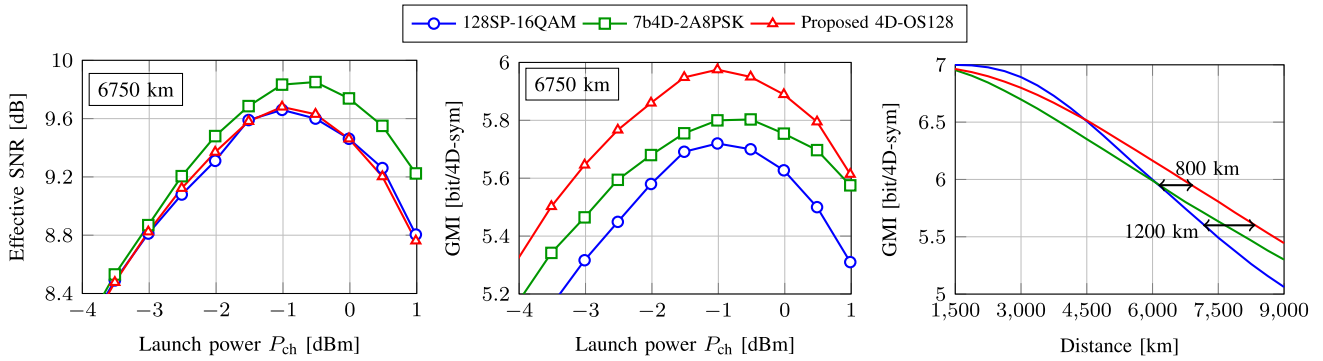


Fig. 7. Post-FEC BER performance of three SE = 7 bits/4D-sym modulations with 20% and 25% OH LDPC.

MaxLog approximation for the proposed 4D-OS128 leads to no observable degradation with respect to the ML demapper. A slightly larger penalty is observed for 128SP-16QAM, in all cases using either 20% or 25% OH.

B. Nonlinear Performance: Multi-Span WDM Transmission

We consider a dual-polarization long-haul WDM transmission system with 11 co-propagating channels generated at a symbol rate of 41.79 GBaud, a WDM spacing of 50 GHz and a root-raised-cosine (RRC) filter roll-off factor of 0.1. Each



(a) Effective SNR vs. P_{ch} at 6750 km. (b) GMI vs. P_{ch} at 6750 km. (c) GMI vs. transmission distance at optimal P_{ch}.

Fig. 8. Simulation results of multi-span optical fiber transmission with 11 WDM channels for three modulation formats with SE of 7 bit/4D-sym: 128SP-16QAM, 7b2D-4D2A8PSK and 4D-OS128.

WDM channel carries 2^{16} 4D symbols in two polarizations at the same launch power per channel P_{ch}. For the transmission link, a multi-span standard single-mode fiber (SSMF) is used with attenuation $\alpha = 0.21$ dB $\cdot$ km $^{-1}$, dispersion parameter $D = 16.9$ ps$\cdot$ nm^{-1} $\cdot$km $^{-1}$, and nonlinear coefficient $\gamma = 1.31$ W^{-1} $\cdot$ km^{-1}. Each span consists of an 75 km SSMF through a split-step Fourier solution of the nonlinear Manakov equation with step size 0.1 km and is followed by an erbium-doped fiber amplifier (EDFA) with a noise figure of 5 dB.

At the receiver side, channel selection is firstly applied and then, the signal is downsampled to 2 samples/symbol. Chromatic dispersion (CD) compensation is performed before applying an RRC matched filter and downsampling to 1 sample/symbol. An ideal phase rotation compensation is performed. Then, log-likelihood ratios (LLRs) are calculated and passed to the soft-decision LDPC decoder. We focus on evaluating the performance of the center WDM channel because it is more affected by the inter-channel crosstalk and non-linear interference.

Fig. 8(a) shows the effective SNR (after fiber propagation and receiver DSP) for different 4D modulation formats with SE=7 bits/4D-sym over a 6750 km SSMF. It is observed that the 7b4D-2A8PSK modulation format with constant modulus has less NLI penalty on effective SNR, while 128SP-16QAM and the proposed 4D-OS128 with power fluctuation between time slots have an effective SNR penalty due to fiber nonlinearities. Normally, a Gaussian-shaped modulation format will lead to a larger SNR penalty. However, the proposed 4D-OS128 modulation format experiences a similar effective SNR compared to 128SP-16QAM at the optimal launch power, which is consistent with the analysis in Sec. II-D. Therefore, the shaping gain in the linear regime can be maintained in nonlinear regime and translates into a reach increase, which will be shown in Fig. 8(b).

Fig. 8(b) shows GMI as a function of the transmitted power for different modulation formats with SE=7 bits/4D-sym over a 6750 km SSMF. As we expected, the proposed 4D-OS128 modulation format achieves the highest GMI compared to 128SP-16QAM and 7b4D-2A8PSK. The gains compared to 128SP-16QAM for different launch power are almost constant due to similar effective SNR performance. Meanwhile, the gains compared to 7b4D-2A8PSK is reduced as the launch power increases.

Fig. 8(c) shows GMI as a function of transmission distance for different modulation formats with SE=7 bits/4D-sym using the optimal launch power at each distance. The proposed 4D-OS128 modulation format leads to a 800 km (13%) and 1200 km (17%) increase in reach relative to the 128SP-16QAM modulation format at GMI of 5.95 bit/4D-sym abd 5.6 bit/4D-sym, respectively. In addition, 4D-OS128 perform better than 7b4D-2A8PSK in all the distance. For GMI above 6.5 bit/4D-sym, 128SP-16QAM provides the best performance. This can not only be attributed to the larger minimum Euclidean distance for 128SP-16QAM with respect to 4D-OS128 (see Table I). In addition, 4D-OS128 is not designed for higher SNR at shorter transmission distances.

C. 4D-OS128 Vs. Other Formats and Probabilistic Shaping

In the previous section, the proposed 4D-OS128 modulation comparable to 128SP-QAM and 7b4D-2A8PSK were discussed for BICM, which enables low complexity. To close the gap to Shannon limit, an alternative approach would be the use of multi-level coding (MLC) with multi-stage decoding (MSD) together with constellations based on the $\boldsymbol{D}_4$ lattice. Recently, 4D CM based on the set of Hurwitz integers has been proposed and demonstrated targeting the (IA) OIF 400ZR implementation agreement [41], [42]. The cardinality of the resulting Hurwitz constellation is $M = 2 \cdot 2^{4i} = 32, 512, 8192, \ldots$, where i is an integer. By selecting the set of Hurwitz integers from the $\boldsymbol{D}_4$ lattice, it is not possible to construct a constellation with $M = 128$. Therefore, we consider here a lattice-based constellation with SE of 7 bit/4D-sym instead, which we denote by l_{4128}. The constellation is the optimal spherical subset of $\boldsymbol{D}_4$ and thus by enumerating and testing a finite number of possible centroids inside the fundamental simplex of the lattice, it becomes optimal.

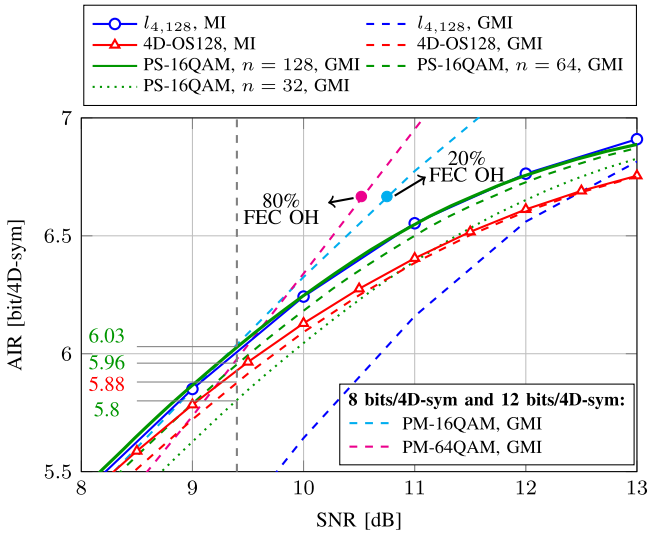


Fig. 9. AIR vs. SNR for different modulation formats with SE of 7 bit/4D-sym. PM-16QAM with SE of 8 bit/4D-sym and PM-64QAM with SE of 12 bit/4D-sym are also shown as baselines.

The l_{4128} constellation was first characterized in [67, Fig. 1(b)] and it also corresponds to the format "l4_128"in [31].

To label the constellation l_{4128}, we use a numerically optimized labeling obtained using the binary-switching algorithm [64]. A labeling was obtained and optimized for an SNR of 9.5 dB. In Fig. 9, the MI and GMI for the l_{4128} constellation are shown. For the proposed 4D-OS128 format, the GMI and the MI are almost identical due to the obtained optimized labeling. This is not the case for the constellation l_{4128}, which lacks a Gray labeling due to the large amount of nearest neighbors. The results in Fig. 9 shows that l_{4128} gives a high MI at all SNRs. However, achieving MI requires symbol-wise decoders, such as MLC-MSD, which have often been avoided in optical communications because of the potentially high complexity induced by using separate bit-level codes and the negative impact of decoding delay.[9] It is also known that MLC schemes are generally more sensitive than BICM to a mismatch between the actual channel parameters and those for which the codes are designed [15]. Furthermore, for the l_{4128} constellation, a large gap between the MI and GMI exists (more than 1 dB for low code rates). Therefore, l_{4128} will not work well with a bit-wise decoder. Similar results have been previously reported in [52] for 4D formats with $M = 16, 256, 4096$.

In Fig. 9, PM-16QAM with SE of 8 bit/4D-sym and PM-64QAM with SE of 12 bit/4D-sym are also shown as baselines. We observe that PM-64QAM has higher GMI at higher SNR region (SNR$\geq$10 dB) compared to PM-16QAM. However, the comparison need to be made between different FEC overheads. For example, PM-64QAM with 80% FEC OH can outperform PM-16QAM with 20% FEC OH at GMI of 6.67 bit/4D-sym, at the price of higher decoding complexity. The same argument applies to the comparison between PM-16QAM and all other formats with SE of 7 bit/4D-sym. More details about higher-order modulations with lower-rate FEC are discussed in Sec. III-D.

In flexible transponders, different modulation formats are used to target different transmission distances. Often in these transponders, the same FEC is used as a separate subsystem for different formats. In order to make a fair comparison against PM-16QAM (8 bits/4D-sym) for the same coding rate, we followed the methodology in [38], [39], i.e., we reduce the symbol rate by 1/8 (from 41.79 GBd to 36.57 GBd). This symbol rate reduction makes the net data rate of the considered equal for any given coding rate. Results for PM-16QAM for the same symbol rate of 41.79 GBd are also included for completeness.

The effective SNR vs. launch power and GMI vs. launch power for a transmission distance of 6750 km are shown in Fig. 10(a) and Fig 10(b), respectively. We can observe that PM-16QAM with the same symbol rate of 41.79 GBd experiences a similar effective SNR compared to the proposed 4D-OS128, and thus has a higher GMI at optimal launch power due to its better linear performance. By contrast, PM-16QAM with 36.57 GBd gives higher effective SNR due to the larger energy per symbol, and thus also achieves higher GMI at optimal launch power.

Transmission distance vs. net data rate for the optimal launch power are shown in Fig. 10(c). In addition, the required FEC OHs are also highlighted by markers. We observe that PM-16QAM with 41.79 GBd achieves higher net data rate than all other modulation formats in Fig. 10(c). However, this comes at the cost of higher FEC OH at the same transmission distance. For example, for the distances between 7400 km and 8300 km, PM-16QAM requires an FEC OH of $34\% - 41\%$ and 4D-OS128 only need the FEC OH of $20\% - 25\%$. When the same FEC OH of $20\% - 25\%$ is considered, PM-16QAM with 41.79 GBd could achieve even higher net data rate, but at the cost of shorter transmission distance (between 5400 km and 6100 km).

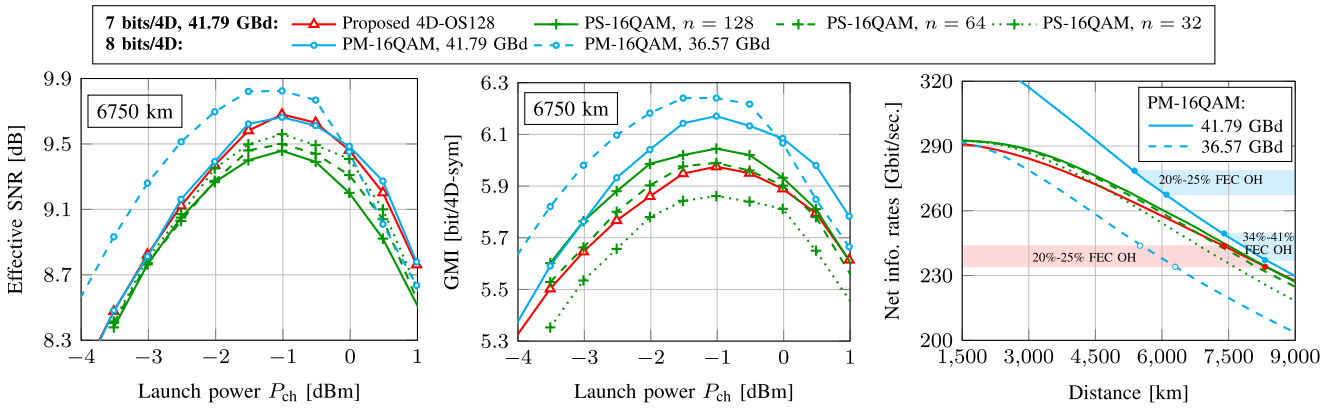
In addition to multidimensional modulation formats, we also consider PS-QAM with the same SE as baseline, which probabilistically shapes PM-16QAM to an entropy of $H(X) = 7$ bit/4D-sym. It is a well known fact that PS has superior AIRs performance for a finite number of constellation points with respect to GS, however, this is based on ideal assumptions for PS, requiring further in-depth analysis;

- **Short blocklength rate loss.** PS based on constant composition distribution matching (CCDM) with long blocklengths is very difficult to implement in high speed communications because it is based on sequential arithmetic coding. In this paper, we consider the finite-length information rate (defined in [69, eq. (15)]) AIR_n of PS with CCDM blocklength of n as $\text{AIR}_n = \text{GMI} - NR_{\text{loss}}$, where N is the number of real dimensions and R_{loss} is the rate loss of CCDM with a finite blocklength.
  The rate loss is defined in [69, eq. (4)] as

$$R_{\text{loss}} = H(P_A) - \frac{k}{n}\,[\text{bits/amp}], \tag{11}$$

  where P_A is the targeted probability distribution, $H(P_A)$ is the entropy in bits/amp, k is the number of input bits for CCDM and n is the blocklength. Note that both PS

[9]Note that the complexity and delay of MLC-MSD schemes can be reduced by wise choice of bit-level codes and efficient hardware implementation designs [68].



(a) Effective SNR vs. P_{ch} at 6750 km. (b) GMI vs. P_{ch} at 6750 km. (c) Net data rate vs. distance at optimal P_{ch}.

Fig. 10. Simulation results of multi-span optical fiber transmission with 11 WDM channels for 4D-OS128 and PS-16QAM with three difference blocklengths. The same simulation setup is implemented as Sec. III-B. Results for PM-16QAM with reduced symbol rate (36.57 Gbaud) are also shown, which give the same net data rate at a given coding rate when compared to the other formats. The net data rate results in bits per second are shown in (c).

and GS coincide, when a block code and low-dimensional constituent constellations is used. Thus, a fair comparison would be to use probabilistic shaping with a blocklength of $n = 4$ as baseline. But considering the complexity of demapping for 4D GS modulation, we use $n = 128, 64, 32$ for PS in this paper. A more comprehensive complexity comparison is left for future research. In Fig. 9, we show the GMIs of PS-16QAM with shaping blockslength $n = 128, 64, 32$. We observe that PS gives higher GMI with $n = 128$, but the resulting rate loss diminishes the efficiency of CCDM as the blocklength decreases. With $n = 32$, PS-16QAM has even worse GMI performance than 4D-OS128.

- **Lower tolerance to fiber nonlinearities**. As mentioned in [5], PS-16QAM experiences higher nonlinearity penalties, which will also lead to a decrease in effective SNR and GMI. These losses are shown in Fig. 10(a) and (b), and are particularly visible in the high nonlinear regime. In Fig. 10(a), we also observe that PS with short blocklengths can also slightly increase the nonlinear tolerance, and thus, the effective SNR. The phenomena is also reported in [9].

By considering the two aspects discussed above, we show results of net data rate with the corresponding baud rate as a function of the transmission distance in Fig. 10(c). We can see that the theoretically superior performance of PS-16QAM vanishes as the blocklength n decreases. In addition, for optical links with stronger nonlinearity, PS will have even higher penalty and lead to a reduced transmission reach and GMI.

When compared to other formats, the proposed 4D-OS128 modulation format has its own advantages in terms of complexity and performance trade-off. The 4D-OS128 modulation format can be easily coupled with FEC and only requires straightforward modifications of the mapper and demapper, and thus, it could be an alternative candidate for optical transmission systems.

D. Comparison With PM-16QAM With Identical FEC Decoding Complexity

Normally, modulation formats with the same SE are compared with the same FEC code rate to keep the identical FEC decoding complexity. For comparison of modulation formats with different SE, different FEC rates need to be considered. However, it has been shown in the literature [70]–[72] that higher-order modulations with lower-rate FEC incur a higher penalty in terms of complexity because more decoding iterations per information bit are required to converge. When the decoding complexity is constrained, the penalty will be transformed into a rate loss between the information rates of realistic hardware-implementable FEC codes and idealistic BICM limits (GMI).

For belief-propagation based decoding, the decoding complexity of LDPC codes for each symbol can be approximated to be proportional to the number of belief message updates per information bit as follows [72, Eq. (2)],[10]

$$P \propto \frac{N\bar{d}_v}{R} \cdot \log_2 M, \tag{12}$$

where N and $\bar{d}_v$ denote the number of decoding iterations and average variable-node degree, respectively. The last term of $\log_2 M$ is the total number of bits per symbol for different modulation formats. Hence, when comparing different modulation formats at the same bit rate and for a given number of decoding iterations, high order modulation formats result in higher complexity.

To keep approximately constant decoding power consumption per symbol as shown in (12), the LDPC codes with $R = 5/6$, $\bar{d}_v = 4$, and $N = 50$ iterations for $M = 4$ QAM are considered as baseline. Hence, the number of decoding iterations of complexity-constrained LDPC is adjusted

[10]Here we only compare the complexity for FEC decoding. However, 4D formats also increase the computational complexity of demapper with respect to a conventional QAM demapper. Low-complexity 4D soft demapper can be designed to reduce the number of MD Euclidean distances calculation by utilizing the property of orthant-symmetry and the schemes in [28]–[30].

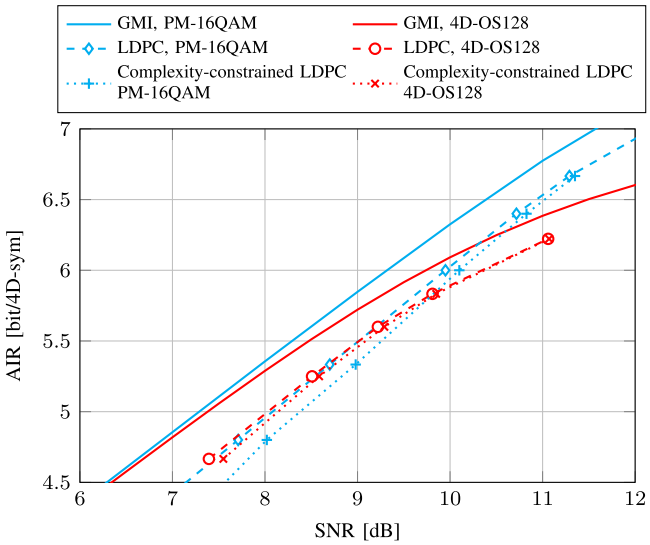


Fig. 11. AIR vs. SNR for the AWGN channel. The results obtained using LDPC codes are shown with markers, which is the required SNR to achieve a post-FEC bit error probability below 10^{-4}. Note that BERs below 10^{-15} can be achieved by for example concatenating an outer high-rate Bose-Chaudhuri-Hocquenghem (BCH) code. This code will only add a minimal additional overhead at an input BER below 10^{-4}.

for PM-16QAM and 4D-OS128 with different code rate $R \in \{1/2, 3/5, 2/3, 3/4, 4/5, 5/6, 8/9\}$. Fig. 11 shows the achievable information rates as a function of required SNR of those LDPC codes. It is observed that achievable rates with LDPC codes still have approximately 1–2 dB loss from GMI, and also that higher-order modulation and lower-rate LDPC codes have higher penalty. It is interesting to note that while 4D-OS128 cannot outperform PM-16QAM for any SNR in terms of GMI (ideal FEC), it does perform equally well (if not better) for an SNR below 9.5 dB when realistic LDPC codes are used. This suggests that idealistic GMI analysis can be too optimistic to evaluate information rates for the hardware with power consumption limitations.

IV. Experimental Setup and Results

A. Experimental Transmission Setup

Fig. 12 depicts the experimental transmission setup. The transmitted signal is modulated using either 128SP-16QAM, 7b4D-2A8PSK [25], or 4D-OS128 symbols. Pseudo-random sequences of 2^{16} symbols are generated offline, pulse shaped using a RRC filter with 1% roll-off at 41.79 GBd, and uploaded to a 100-GSa/s DAC. The positive differential DAC outputs are connected to the optical multi-format transmitter (OMFT) which consists of an external cavity laser (ECL), a dual-polarization IQ-modulator (DP-IQM), an automatic bias controller (ABC) and RF-amplifiers. The channel under test (CUT), which can be defined at any of the 11 tested C-band channels, is modulated by the OMFT and subsequently amplified. The loading channels are provided by the negative outputs of the DAC and modulated onto the tones provided by 10 ECLs using a DP-IQM. These loading channels are amplified, split into even and odd channels, decorrelated by 10 200 (50 m) and 40 800 symbols (200 m), and multiplexed together with the CUT on a 50-GHz grid using an optical tunable filter (OTF). Bandwidth limitations due to transmitter electronics are initially compensated using an OTF and the residual effects are mitigated digitally as proposed in [73].

The 11-channel 50-GHz-spaced dense wavelength division multiplexing (DWDM) signal is amplified and through an acousto-optic modulator (AOM) enters the recirculating loop which consists of a loop-synchronous polarization scrambler (LSPS), a 75-km span of SSMF, an EDFA, an AOM and an OTF used for gain flattening. The inset of Fig. 12 shows the optical spectrum after 80 circulations, which corresponds to 6000 km of transmission using only EDFA-amplification. Optionally, a hybrid amplification scheme can be used by adding a 750 mW 1445 nm Raman pump in a backward pumping configuration. The output of the recirculating loop is amplified, filtered by a wavelength selective switch (WSS) and digitized by a coherent receiver consisting of a local oscillator (LO), a 90-degree hybrid, four balanced photo-diodes and an 80 GSa/s analog-to-digital converter (ADC). The offline digital signal processing (DSP) includes front-end impairment correction using blind moment estimation, chromatic dispersion compensation, frequency offset estimation and correction between transmitter and LO laser. A widely-linear multiple-input multiple-output (MIMO) equalization [74] with blind phase search (BPS) [75] inside the update loop is employed to correct for phase noise, error counting and GMI evaluation. In the following sections, we evaluate and discuss the results for two configurations as EDFA-only amplification (Sec. IV-B) and a hybrid of EDFA and Raman (Sec. IV-C).

B. Experimental Results: EDFA-Only Amplifier

Fig. 13 and Fig. 14 show the 2D projections of the received constellations after optical back-to-back and after transmission over 70 spans respectively. Note that the proposed 4D-OS128 modulation induces nonuniform probability distribution when projected onto 2D, which is similar to probabilistic amplitude shaped 16QAM with very short blocklength of 4. Instead of using four time slots as PS, the 4D-OS128 shapes the constellation using the two quadratures (I/Q) and the two polarization states (X/Y) as four dimensions.

Fig. 15(a) shows the GMI as a function of transmission distance for the optimal total launch power of 9.5 dBm. The GMI in Fig. 15(a) should be interpreted as the reach that an ideal SD-FEC would achieve. For the considered rate (6 bit/4D-sym), 7b4D-2A8PSK offers approximately the same reach as 128SP-16QAM around 5320 km. The proposed 4D-OS128 reaches 6120 km, which corresponds to a gain of 800 km (15%). This observed reach increase in percentage is in good agreement with the simulation results of Fig. 8.[11] These experimental results

[11]Note that the experimental results shows a shorter transmission distance than the simulations in Fig. 8 This is due to the loss from optical components (LSPS, AOM and OTF) in the recirculating loop.

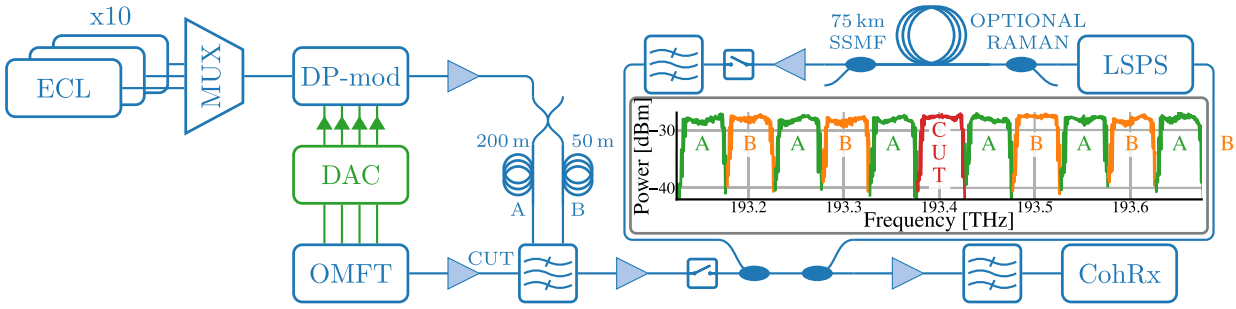
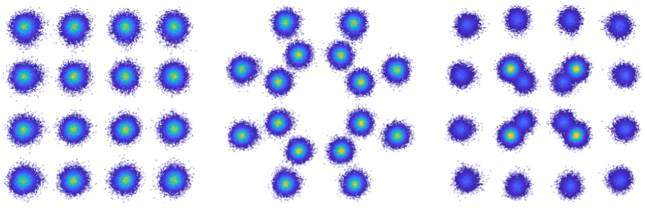
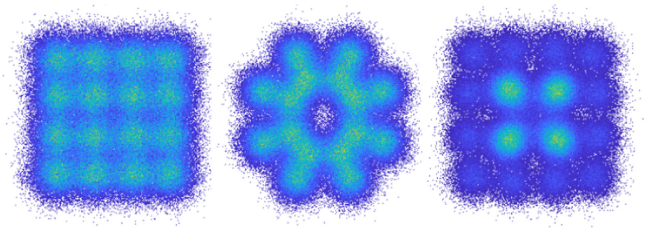


Fig. 12. Experimental optical recirculating loop setup. Inset: The received spectrum after $N = 80$ circulations (6000 km) of EDFA-only amplification. The CUT is depicted in the center position but is tested in all 11 positions in the experiment.



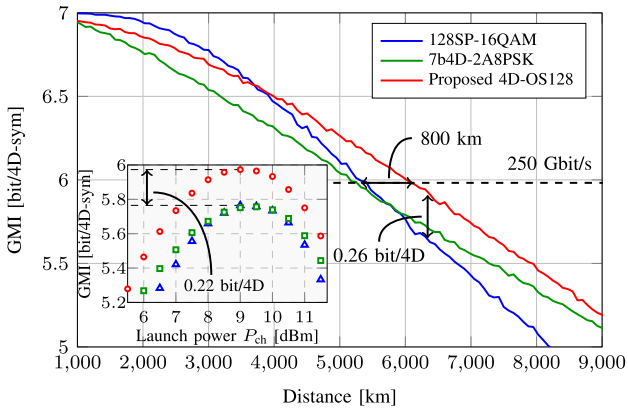
(a) 128SP-16QAM (b) 7b4D-2A8PSK (c) 4D-OS128

Fig. 13. 2D projection of the constellations after Back-to-Back measurements. The SNR of these recovered constellations is around 20 dB.

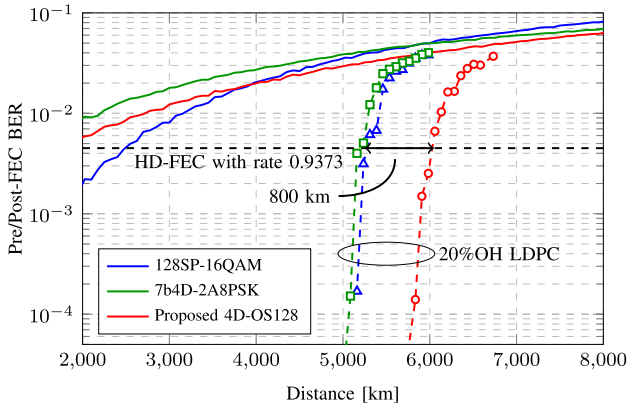


(a) 128SP-16QAM (b) 7b4D-2A8PSK (c) 4D-OS128

Fig. 14. 2D projection of the constellations after 70 spans transmission with total launch power of 9 dBm.



(a) Average GMI per channel versus transmission distance at total launch power of 9.5 dBm. Inset: Average GMI per channel versus total launch power after 6000 km.



(b) BER versus transmission distance at total launch power of 9.5 dBm.

Fig. 15. Experimental results using EDFA-only amplification.

confirm the performance of the novel 4D-OS128 modulation format obtained in simulation. Fig. 15(a) also shows that the average GMI per channel resulting in a 0.26 bit/4D increase for 4D-OS128 with respect to 128SP-16QAM after 6340 km transmission. The GMI vs. launch power for a transmission distance of 6000 km and the three modulation formats under consideration are also shown as inset in Fig. 15(a). At the optimal launch power, 4D-OS128 outperforms 128SP-16QAM and 7b4D-2A8PSK with a gain of 0.22 bits/4D-sym.

BER performance before and after FEC are shown in Fig. 15(b). For the experiment, low-density parity-check (LDPC) blocks are constructed, which are then encoded using the DVB-S2 LDPC code with 20% overhead and code length $n = 64\,800$. An outer hard decision forward error correction (HD-FEC) staircase code with rate 0.9373 [76] that corrects bit errors after LDPC decoding is assumed. The BER threshold is 4.5×10^{-3} [76, Fig. 8], which makes 4D-OS128 15% (+800 km) better in reach (6050 km) compared to 128SP-16QAM (5250 km) with the same data rate. Moreover, the 15% reach increase is preserved for the post-FEC gain using an off-the-shelf DVB-S2 LDPC. In addition to the gains shown versus 128SP-16QAM, 4D-OS128 is also shown to outperform 7b4D-2A8PSK. The combination of the rate of this staircase code together with a baudrate of 41.79 GBd and a net data rate of 5.58 bits/4D-sym results in a total data rate of just over 233 Gbit/s per channel.

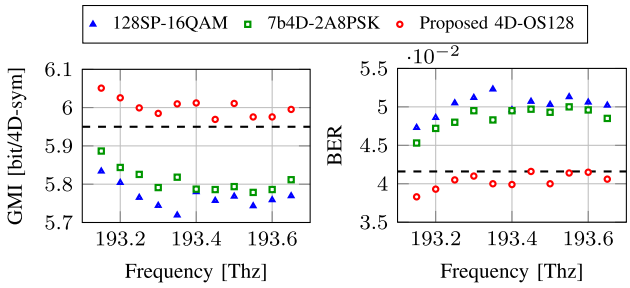


Fig. 16. Experimental results using EDFA-only amplification. Per-channel performance versus GMIs (left) and BERs (right) measured for all 11 channels individually after 6000 km showing GMIs above 5.95 bit/4D [25] and BERs below the FEC threshold $4.1 \cdot 10^{-2}$ [77].

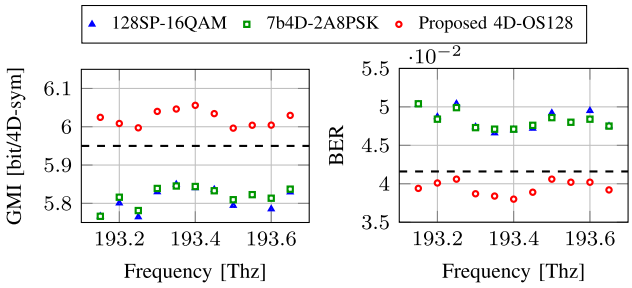
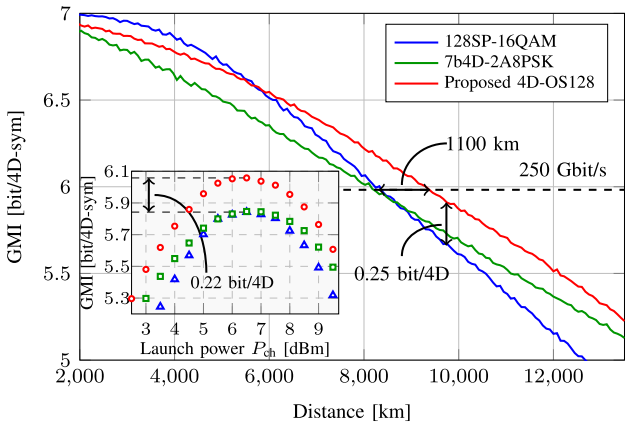
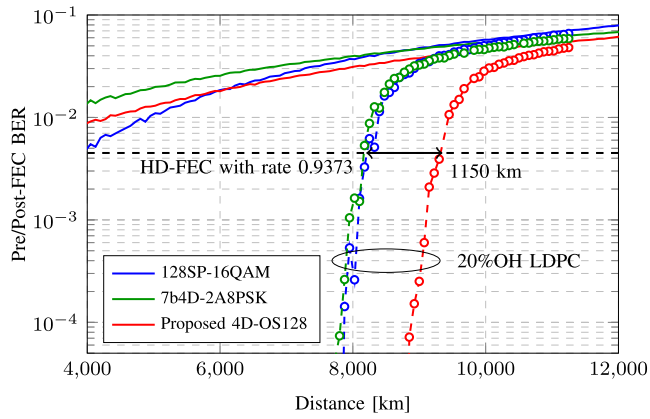


Fig. 18. Experimental results using a hybrid of EDFA and Raman amplification. Per-channel performance versus GMIs (left) and BERs (right) measured for all 11 channels individually after 9000 km showing GMIs above 5.95 bit/4D [25] and BERs below the FEC threshold $4.1 \cdot 10^{-2}$ [77].



(a) Average GMI per channel versus transmission distance at total launch power of 6.5 dBm. Inset: Average GMI per channel versus total launch power after 9000 km.



(b) BER versus transmission distance at total launch power of 6.5 dBm.

Fig. 17. Experimental results using Hybrid amplification.

At this optimal launch power, we demonstrate transmission above the GMI threshold of 5.95 bit/4D-sym (0.85 NGMI) enabling 6000 km error-free transmission of net 233 Gbit/s per channel after 25.5% overhead and for all 11 channel in Fig. 16. The GMI threshold 5.95 bit/4D-sym (0.85 NGMI) is based on a spatially-coupled type LDPC code [77] and the corresponding BER threshold of $4.1 \cdot 10^{-2}$ is derived in [25]. Note that the rate loss due to the practical FEC after 6000 km transmission is 16 Gbit/s between the net data rate of 233 Gbit/s and the maximum data rate 249 Gbit/s (41.79 GBaud$\times$5.95 bit/4D-sym). Therefore, the relative FEC loss is 6.4%.

C. Experimental Results: Hybrid Amplification

In the hybrid amplification scheme, a 750 mW 1445 nm Raman pump is also used in a backward configuration as shown in Fig. 12.

The GMI vs. launch power for a transmission distance of 9000 km and the three modulation formats under consideration are also shown as inset in Fig. 17(a). At the optimal launch power of 6.5 dBm, 4D-OS128 maximizes the average GMI per channel resulting in a 0.22 bits/4D-sym increase for with respect to 128SP-16QAM and 7b4D-2A8PSK. Therefore, we use 6.5 dBm as launch power to evaluate the transmission performance. Fig. 17(a). shows the GMI as a function of transmission distance. We can observe that the relative GMI gains of 0.25 bit/4D-sym and 1100 km (13.5%) are similar to the EDFA-only case.

BER performance before and after FEC are shown in Fig. 17(b). Under the assumption of concatenated LDPC and staircase code with rate 0.9373, 4D-OS128 shows a post-FEC reach increase of 1150 km (14%). By measuring all the 11 WDM channels' performance, Fig. 18 shows all channels above the GMI threshold of 5.95 bit/4D-sym (0.85 NGMI) enabling 9000 km error-free transmission of net 233 Gbit/s per channel after 25.5% overhead. Comparing to the results of the EDFA-only amplification in Sec. IV-B, utilizing Raman amplifier causes an apparent improvement (50% reach increase) on the transmission performance compared to using the EDFA for the 4D-OS128 modulation.

V. Conclusion

A new modulation format (4D-OS128) with a spectral efficiency of 7 bits/4D-sym was introduced using the concept of orthant symmetry. The format was designed based on the generalized mutual information, and thus, it finds applications to systems with soft-decision forward error correcting codes and bit-wise decoding. The 4D-OS128 format provides sensitivity gains of up to 0.65 dB after LDPC decoding versus two

TABLE II
COORDINATES AND BINARY LABELING OF THE PROPOSED 4D-OS128 FORMAT FOR THE TARGET SNR OF 9.5 dB. THE COORDINATES ARE ROUNDED TO FOUR DECIMAL POINTS:
$(t_1, t_2, t_3, t_4, t_5) = (0.2875, 0.3834, 0.4730, 1.1501, 1.2460)$

Coordinates	Labeling	Coordinates	Labeling
$(+t_3, +t_3, +t_1, +t_1)$	0000011	$(+t_2, +t_5, +t_3, +t_3)$	0000001
$(-t_3, +t_3, +t_1, +t_1)$	1000011	$(-t_2, +t_5, +t_3, +t_3)$	1000001
$(-t_3, -t_3, +t_1, +t_1)$	1100011	$(-t_2, -t_5, +t_3, +t_3)$	1100001
$(+t_3, -t_3, +t_1, +t_1)$	0100011	$(+t_2, -t_5, +t_3, +t_3)$	0100001
$(+t_1, +t_1, +t_3, +t_3)$	0000111	$(+t_3, +t_3, +t_5, +t_2)$	0000101
$(-t_1, +t_1, +t_3, +t_3)$	1000111	$(-t_3, +t_3, +t_5, +t_2)$	1000101
$(-t_1, -t_1, +t_3, +t_3)$	1100111	$(-t_3, -t_3, +t_5, +t_2)$	1100101
$(+t_1, -t_1, +t_3, +t_3)$	0100111	$(+t_3, -t_3, +t_5, +t_2)$	0100101
$(+t_3, +t_3, +t_2, +t_5)$	0000110	$(+t_3, +t_3, +t_4, +t_4)$	0000100
$(-t_3, +t_3, +t_2, +t_5)$	1000110	$(-t_3, +t_3, +t_4, +t_4)$	1000100
$(-t_3, -t_3, +t_2, +t_5)$	1100110	$(-t_3, -t_3, +t_4, +t_4)$	1100100
$(+t_3, -t_3, +t_2, +t_5)$	0100110	$(+t_3, -t_3, +t_4, +t_4)$	0100100
$(+t_5, +t_2, +t_3, +t_3)$	0000010	$(+t_4, +t_4, +t_3, +t_3)$	0000000
$(-t_5, +t_2, +t_3, +t_3)$	1000010	$(-t_4, +t_4, +t_3, +t_3)$	1000000
$(-t_5, -t_2, +t_3, +t_3)$	1100010	$(-t_4, -t_4, +t_3, +t_3)$	1100000
$(+t_5, -t_2, +t_3, +t_3)$	0100010	$(+t_4, -t_4, +t_3, +t_3)$	0100000
$(+t_3, +t_3, -t_1, +t_1)$	0010011	$(+t_2, +t_5, -t_3, +t_3)$	0010001
$(-t_3, +t_3, -t_1, +t_1)$	1010011	$(-t_2, +t_5, -t_3, +t_3)$	1010001
$(-t_3, -t_3, -t_1, +t_1)$	1110011	$(-t_2, -t_5, -t_3, +t_3)$	1110001
$(+t_3, -t_3, -t_1, +t_1)$	0110011	$(+t_2, -t_5, -t_3, +t_3)$	0110001
$(+t_1, +t_1, -t_3, +t_3)$	0010111	$(+t_3, +t_3, -t_5, +t_2)$	0010101
$(-t_1, +t_1, -t_3, +t_3)$	1010111	$(-t_3, +t_3, -t_5, +t_2)$	1010101
$(-t_1, -t_1, -t_3, +t_3)$	1110111	$(-t_3, -t_3, -t_5, +t_2)$	1110101
$(+t_1, -t_1, -t_3, +t_3)$	0110111	$(+t_3, -t_3, -t_5, +t_2)$	0110101
$(+t_3, +t_3, -t_2, +t_5)$	0010110	$(+t_3, +t_3, -t_4, +t_4)$	0010100
$(-t_3, +t_3, -t_2, +t_5)$	1010110	$(-t_3, +t_3, -t_4, +t_4)$	1010100
$(-t_3, -t_3, -t_2, +t_5)$	1110110	$(-t_3, -t_3, -t_4, +t_4)$	1110100
$(+t_3, -t_3, -t_2, +t_5)$	0110110	$(+t_3, -t_3, -t_4, +t_4)$	0110100
$(+t_5, +t_2, -t_3, +t_3)$	0010010	$(+t_4, +t_4, -t_3, +t_3)$	0010000
$(-t_5, +t_2, -t_3, +t_3)$	1010010	$(-t_4, +t_4, -t_3, +t_3)$	1010000
$(-t_5, -t_2, -t_3, +t_3)$	1110010	$(-t_4, -t_4, -t_3, +t_3)$	1110000
$(+t_5, -t_2, -t_3, +t_3)$	0110010	$(+t_4, -t_4, -t_3, +t_3)$	0110000
$(+t_3, +t_3, -t_1, -t_1)$	0011011	$(+t_2, +t_5, -t_3, -t_3)$	0011001
$(-t_3, +t_3, -t_1, -t_1)$	1011011	$(-t_2, +t_5, -t_3, -t_3)$	1011001
$(-t_3, -t_3, -t_1, -t_1)$	1111011	$(-t_2, -t_5, -t_3, -t_3)$	1111001
$(+t_3, -t_3, -t_1, -t_1)$	0111011	$(+t_2, -t_5, -t_3, -t_3)$	0111001
$(+t_1, +t_1, -t_3, -t_3)$	0011111	$(+t_3, +t_3, -t_5, -t_2)$	0011101
$(-t_1, +t_1, -t_3, -t_3)$	1011111	$(-t_3, +t_3, -t_5, -t_2)$	1011101
$(-t_1, -t_1, -t_3, -t_3)$	1111111	$(-t_3, -t_3, -t_5, -t_2)$	1111101
$(+t_1, -t_1, -t_3, -t_3)$	0111111	$(+t_3, -t_3, -t_5, -t_2)$	0111101
$(+t_3, +t_3, -t_2, -t_5)$	0011110	$(+t_3, +t_3, -t_4, -t_4)$	0011100
$(-t_3, +t_3, -t_2, -t_5)$	1011110	$(-t_3, +t_3, -t_4, -t_4)$	1011100
$(-t_3, -t_3, -t_2, -t_5)$	1111110	$(-t_3, -t_3, -t_4, -t_4)$	1111100
$(+t_3, -t_3, -t_2, -t_5)$	0111110	$(+t_3, -t_3, -t_4, -t_4)$	0111100
$(+t_5, +t_2, -t_3, -t_3)$	0011010	$(+t_4, +t_4, -t_3, -t_3)$	0011000
$(-t_5, +t_2, -t_3, -t_3)$	1011010	$(-t_4, +t_4, -t_3, -t_3)$	1011000
$(-t_5, -t_2, -t_3, -t_3)$	1111010	$(-t_4, -t_4, -t_3, -t_3)$	1111000
$(+t_5, -t_2, -t_3, -t_3)$	0111010	$(+t_4, -t_4, -t_3, -t_3)$	0111000
$(+t_3, +t_3, +t_1, -t_1)$	0001011	$(+t_2, +t_5, +t_3, -t_3)$	0001001
$(-t_3, +t_3, +t_1, -t_1)$	1001011	$(-t_2, +t_5, +t_3, -t_3)$	1001001
$(-t_3, -t_3, +t_1, -t_1)$	1101011	$(-t_2, -t_5, +t_3, -t_3)$	1101001
$(+t_3, -t_3, +t_1, -t_1)$	0101011	$(+t_2, -t_5, +t_3, -t_3)$	0101001
$(+t_1, +t_1, +t_3, -t_3)$	0001111	$(+t_3, +t_3, +t_5, -t_2)$	0001101
$(-t_1, +t_1, +t_3, -t_3)$	1001111	$(-t_3, +t_3, +t_5, -t_2)$	1001101
$(-t_1, -t_1, +t_3, -t_3)$	1101111	$(-t_3, -t_3, +t_5, -t_2)$	1101101
$(+t_1, -t_1, +t_3, -t_3)$	0101111	$(+t_3, -t_3, +t_5, -t_2)$	0101101
$(+t_3, +t_3, +t_2, -t_5)$	0001110	$(+t_3, +t_3, +t_4, -t_4)$	0001100
$(-t_3, +t_3, +t_2, -t_5)$	1001110	$(-t_3, +t_3, +t_4, -t_4)$	1001100
$(-t_3, -t_3, +t_2, -t_5)$	1101110	$(-t_3, -t_3, +t_4, -t_4)$	1101100
$(+t_3, -t_3, +t_2, -t_5)$	0101110	$(+t_3, -t_3, +t_4, -t_4)$	0101100
$(+t_5, +t_2, +t_3, -t_3)$	0001010	$(+t_4, +t_4, +t_3, -t_3)$	0001000
$(-t_5, +t_2, +t_3, -t_3)$	1001010	$(-t_4, +t_4, +t_3, -t_3)$	1001000
$(-t_5, -t_2, +t_3, -t_3)$	1101010	$(-t_4, -t_4, +t_3, -t_3)$	1101000
$(+t_5, -t_2, +t_3, -t_3)$	0101010	$(+t_4, -t_4, +t_3, -t_3)$	0101000

well-studied 4D modulation formats: 128SP-16QAM and 7b4D-2A8PSK. The numerical results of optical multi-span transmission results shows significant improvement over GS formats and also a comparable performance versus PS-QAM with short blocklength for distribution matching. The experimental results confirm the overall superior receiver sensitivity of 4D-OS128 versus previously published GS formats. Transmission reach extensions of more than 15% is demonstrated. We believe that the proposed format is a good alternative for future high capacity long haul transmission systems, which provides an intermediate solution between PM-8QAM and PM-16QAM. The design of orthant-symmetric constellations for higher dimensions (e.g., 16 dimensions) and larger constellation sizes (e.g., 256-ary, 512-ary and 1024-ary formats) are left for further investigation.

APPENDIX
COORDINATES AND BINARY LABELING FOR 4D-OS128

Table II lists the coordinates of the constellation points and the bit-to-symbol mapping of 4D-OS128 for the target SNR of 9.5 dB. The constellation is assumed to be normalized to $E_s = 2$, i.e., to unit energy per polarization.

ACKNOWLEDGMENT

Fraunhofer HHI and ID Photonics are gratefully acknowledged for providing their Optical-Multi-Format Transmitter.

REFERENCES

[1] R. Essiambre *et al.*, "Capacity limits of optical fiber networks," *J. Lightw. Technol.*, vol. 28, no. 4, pp. 662–701, Feb. 2010.
[2] H. Bülow, "Polarization QAM modulation (POL-QAM) for coherent detection schemes," in *Proc. Opt. Fiber Commun. Conf.*, Mar. 2009, pp. 1–3.
[3] P. J. Winzer, A. H. Gnauck, C. R. Doerr, M. Magarini, and L. L. Buhl, "Spectrally efficient long-haul optical networking using 112-Gb/s polarization-multiplexed 16-QAM," *J. Lightw. Technol.*, vol. 28, no. 4, pp. 547–556, Feb. 2010.
[4] F. Buchali, F. Steiner, G. Böcherer, L. Schmalen, P. Schulte, and W. Idler, "Rate adaptation and reach increase by probabilistically shaped 64-QAM: An experimental demonstration," *J. Lightw. Technol.*, vol. 34, no. 7, pp. 1599–1609, Apr. 2016.
[5] 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.
[6] G. Böcherer, F. Steiner, and P. Schulte, "Fast probabilistic shaping implementation for long-haul fiber-optic communication systems," in *Proc. Eur. Conf. Opt. Commun.*, Sep. 2017, pp. 1–3.
[7] F. Buchali, W. Idler, R. Dischler, T. Eriksson, and L. Schmalen, "Spectrally efficient probabilistically shaped square 64QAM to 256QAM," in *Proc. Eur. Conf. Opt. Commun.*, Sep. 2017, pp. 1–3.
[8] R. Maher, K. Croussore, M. Lauermann, R. Going, X. Xu, and J. Rahn, "Constellation shaped 66 GBd DP-1024QAM transceiver with 400 km transmission over standard SMF," in *Proc. Eur. Conf. Opt. Commun.*, Sep. 2017, pp. 1–3.
[9] 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.
[10] S. Goossens *et al.*, "First experimental demonstration of probabilistic enumerative sphere shaping in optical fiber communications," in *Proc. Opto-Electron. Commun. Conf. Int. Conf. Photon. Switch. Comput.*, Fukuoka, Japan, Jul. 2019, pp. 1–3..
[11] Z. Qu and I. B. Djordjevic, "Geometrically shaped 16QAM outperforming probabilistically shaped 16QAM," in *Proc. Eur. Conf. Opt. Commun.*, Sep. 2017, pp. 1–3.

[12] S. Zhang, F. Yaman, E. Mateo, T. Inoue, K. Nakamura, and Y. Inada, "Design and performance evaluation of a GMI-optimized 32QAM," in *Proc. Eur. Conf. Opt. Commun.*, Sep. 2017, pp. 1–3.

[13] B. Chen, C. Okonkwo, H. Hafermann, and A. Alvarado, "Increasing achievable information rates via geometric shaping," in *Proc. Eur. Conf. Opt. Commun.*, Sep. 2018, pp. 1–3.

[14] B. Chen, C. Okonkwo, D. Lavery, and A. Alvarado, "Geometrically-shaped 64-point constellations via achievable information rates," in *Proc. 20th Int. Conf. Transparent Opt. Netw.*, Jul. 2018, pp. 1–4.

[15] L. Szczecinski and A. Alvarado, *Bit-Interleaved Coded Modulation: Fundamentals, Analysis and Design*. Wiley, 2015.

[16] F. Steiner and G. Böcherer, "Comparison of geometric and probabilistic shaping with application to ATSC 3.0," in *Proc. 11th Int. ITG Conf. Syst., Commun. Coding*, Feb. 2017, pp. 1–6.

[17] E. Agrell and M. Karlsson, "Power-efficient modulation formats in coherent transmission systems," *J. Lightw. Technol.*, vol. 27, no. 22, pp. 5115–5126, Nov. 2009.

[18] M. Karlsson and E. Agrell, "Which is the most power-efficient modulation format in optical links?" *Opt. Exp.*, vol. 17, no. 13, pp. 10 814–10 819, Jun. 2009.

[19] T. Koike-Akino, D. S. Millar, K. Kojima, and K. Parsons, "Eight-dimensional modulation for coherent optical communications," in *Proc. Eur. Conf. Opt. Commun.*, Sep. 2013, pp. 1–3.

[20] D. S. Millar *et al.*, "High-dimensional modulation for coherent optical communications systems," *Opt. Exp.*, vol. 22, no. 7, pp. 8798–8812, Apr. 2014.

[21] D. S. Millar, T. Fehenberger, T. Koike-Akino, K. Kojima, and K. Parsons, "Coded modulation for next-generation optical communications," in *Proc. Opt. Fiber Commun. Conf.*, Mar. 2018, pp. 1–3.

[22] M. Chagnon *et al.*, "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. 30 204-30 220, Dec. 2013.

[23] A. D. Shiner *et al.*, "Demonstration of an 8-dimensional modulation format with reduced inter-channel nonlinearities in a polarization multiplexed coherent system," *Opt. Exp.*, vol. 22, no. 17, pp. 20 366–20 374, Aug. 2014.

[24] M. Reimer, S. O. Gharan, A. D. Shiner, and M. O'Sullivan, "Optimized 4 and 8 dimensional modulation formats for variable capacity in optical networks," in *Proc. Opt. Fiber Commun. Conf.*, Mar. 2016, Paper M3A–4.

[25] 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.

[26] B. Chen, C. Okonkwo, H. Hafermann, and A. Alvarado, "Polarization-ring-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.

[27] B. Chen, C. Okonkwo, H. Hafermann, and A. Alvarado, "Eight-dimensional polarization-ring-switching modulation formats," *IEEE Photon. Technol. Lett.*, vol. 31, no. 21, pp. 1717–1720, Nov. 2019.

[28] T. Yoshida *et al.*, "Hardware-efficient precise and flexible soft-demapping for multi-dimensional complementary APSK signals," in *Proc. Eur. Conf. Opt. Commun.*, Sep. 2016, pp. 1–3.

[29] D. F. Bendimerad, H. Zhang, and I. Land, "Ultra-low complexity and high performance soft-demapper for 8D set-partitioned PDM-QPSK modulation formats," in *Proc. Eur. Conf. Opt. Commun.*, Sep. 2018, pp. 1–3.

[30] M. Nakamura *et al.*, "Low-complexity iterative soft-demapper for multidimensional modulation based on bitwise log likelihood ratio and its demonstration in high baud-rate transmission," *J. Lightw. Technol.*, vol. 36, no. 2, pp. 476–484, Jan. 2018.

[31] "Database of Sphere Packings," [Online]. Available: https://www.codes.se/packings/

[32] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Trans. Inf. Theory*, vol. 28, no. 1, pp. 55–67, Jan. 1982.

[33] F. Waeckerle and R. F. H. Fischer, "Multistage bit-wise receivers for 4D modulation formats in optical communications," in *Proc. 10th Int. ITG Conf. Syst., Commun. Coding*, 2015, pp. 1–6.

[34] A. Alvarado and E. Agrell, "Four-dimensional coded modulation with bit-wise decoders for future optical communications," *J. Lightw. Technol.*, vol. 33, no. 10, pp. 1993–2003, May 2015.

[35] R. Rios-Muller *et al.*, "Experimental comparison between hybrid-QPSK/8QAM and 4D-32SP-16QAM formats at 31.2 GBaud using Nyquist pulse shaping," in *Proc. 39th Eur. Conf. Opt. Commun.*, Sep. 2013, pp. 1–3.

[36] H. Sun, R. Egorov, B. E. Basch, J. McNicol, and K. Wu, "Comparison of two modulation formats at spectral efficiency of 5 bits/dual-pol symbol," in *Proc. 39th Eur. Conf. Opt. Commun.*, Sep. 2013, pp. 1–3.

[37] T. Nakamura, E. L. T. de Gabory, H. Noguchi, W. Maeda, J. Abe, and K. Fukuchi, "Long haul transmission of four-dimensional 64SP-12QAM signal based on 16QAM constellation for longer distance at same spectral efficiency as PM-8QAM," in *Proc. Eur. Conf. Opt. Commun.*, Sep. 2015, pp. 1–3.

[38] T. A. Eriksson, M. Sjödin, P. Johannisson, P. A. Andrekson, and M. Karlsson, "Comparison of 128-SP-QAM and PM-16QAM in long-haul WDM transmission," *Opt. Exp.*, vol. 21, no. 16, pp. 19 269–19 279, Aug. 2013.

[39] A. S. Kashi *et al.*, "Information rates for the SP 128-QAM and DP 16-QAM modulation formats," in *Proc. Eur. Conf. Opt. Commun.*, Sep. 2015, pp. 1–3.

[40] K. Wang, J. Yu, P. Gou, M. Zhao, K. Lyu, and X. Xin, "Transmission performance comparison of 128-SP-QAM and PM-16QAM in a WDM system," *Opt. Fiber Technol.*, vol. 43, pp. 158–162, 2018.

[41] F. Frey, S. Stern, R. Emmerich, C. Schubert, J. K. Fischer, and R. Fischer, "Coded modulation using a 512-ary Hurwitz-integer constellation," in *Proc. Eur. Conf. Opt. Commun.*, Sep. 2019, pp. 1-4.

[42] F. Frey, S. Stern, J. K. Fischer, and R. Fischer, "Two-stage coded modulation for Hurwitz constellations in fiber-optical communications," *J. Lightw. Technol.*, vol. 38, no. 12, pp. 3135–3146, Jun. 2020.

[43] W. Betts, A. R. Calderbank, and R. Laroia, "Performance of nonuniform constellations on the Gaussian channel," *IEEE Trans. Inf. Theory*, vol. 40, no. 5, pp. 1633–1638, Sep. 1994.

[44] I. Kalet and B. R. Saltzberg, "Qam transmission through a companding channel - signal constellations and detection," *IEEE Trans. Commun.*, vol. 42, no. 234, pp. 417–429, Feb.-Apr. 1994.

[45] J. Renaudier *et al.*, "Comparison of set-partitioned two-polarization 16QAM formats with PDM-QPSK and PDM-8QAM for optical transmission systems with error-correction coding," in *Proc. Eur. Conf. Opt. Commun.*, Sep. 2012, pp. 1–3.

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

[47] B. P. Smith and F. R. Kschischang, "A pragmatic coded modulation scheme for high-spectral-efficiency fiber-optic communications," *J. Lightw. Technol.*, vol. 30, no. 13, pp. 2047–2053, Jul. 2012.

[48] S. van der Heide *et al.*, "11,700 km transmission at 4.8 bit/4D-sym via four-dimensional geometrically-shaped polarization-ring-switching modulation," in *Proc. Opto-Electron. Commun. Conf. Int. Conf. Photon. Switch. Comput.*, Jul. 2019, pp. 1–3.

[49] G. Liga *et al.*, "30% reach increase via low-complexity hybrid HD/SD FEC and improved 4D modulation," *IEEE Photon. Technol. Lett.*, vol. 32, no. 13, pp. 827–830, Jul. 2020.

[50] S. Zhang, F. Yaman, E. Mateo, T. Inoue, K. Nakamura, and Y. Inada, "A generalized pairwise optimization for designing multi-dimensional modulation formats," in *Proc. Opt. Fiber Commun. Conf.*, Mar. 2017, pp. 1-3.

[51] B. Chen, C. Okonkwo, and A. Alvarado, "Nonlinear interference analysis of probabilistic shaping vs. 4D geometrically shaped formats," in *Proc. Eur. Conf. Opt. Commun.*, Dec. 2020, pp. 1–4.

[52] A. Alvarado, E. Agrell, D. Lavery, R. Maher, and P. Bayvel, "Replacing the soft-decision FEC limit paradigm in the design of optical communication systems," *J. Lightw. Technol.*, vol. 33, no. 20, pp. 4338–4352, Oct. 2015.

[53] E. Agrell, A. Alvarado, G. Durisi, and M. Karlsson, "Capacity of a nonlinear optical channel with finite memory," *J. Lightw. Technol.*, vol. 32, no. 16, pp. 2862–2876, Aug. 2014.

[54] R. Dar, M. Feder, A. Mecozzi, and M. Shtaif, "Inter-channel nonlinear interference noise in WDM systems: Modeling and mitigation," *J. Lightw. Technol.*, vol. 33, no. 5, pp. 1044–1053, Mar. 2015.

[55] R. Dar, M. Feder, A. Mecozzi, and M. Shtaif, "Pulse collision picture of inter-channel nonlinear interference in fiber-optic communications," *J. Lightw. Technol.*, vol. 34, no. 2, pp. 593–607, Jan. 2016.

[56] A. Ghazisaeidi, "A theory of nonlinear interactions between signal and amplified spontaneous emission noise in coherent wavelength division multiplexed systems," *J. Lightw. Technol.*, vol. 35, no. 23, pp. 5150–5175, Dec. 2017.

[57] T. A. Eriksson, T. Fehenberger, and W. Idler, "Characterization of nonlinear fiber interactions using multidimensional mutual information over time and polarization," *J. Lightw. Technol.*, vol. 35, no. 6, pp. 1204–1210, Mar. 2017.

[58] E. Agrell, J. Lassing, E. G. Strom, and T. Ottosson, "On the optimality of the binary reflected gray code," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3170–3182, Dec. 2004.
[59] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 927–946, May 1998.
[60] E. Agrell and A. Alvarado, "Optimal alphabets and binary labelings for BICM at low SNR," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6650–6672, Oct. 2011.
[61] A. Alvarado, F. Brännström, E. Agrell, and T. Koch, "High-SNR asymptotics of mutual information for discrete constellations with applications to BICM," *IEEE Trans. Inf. Theory*, vol. 60, no. 2, pp. 1061–1076, Feb. 2014.
[62] A. Alvarado, F. Brännström, and E. Agrell, "High SNR bounds for the BICM capacity," in *IEEE Inf. Theory Workshop*, Paraty, Brazil, Oct. 2011, pp. 360–364.
[63] K. Gümüs, A. Alvarado, B. Chen, C. Häger, and E. Agrell, "End-to-end learning of geometrical shaping maximizing generalized mutual information," in *Proc. Opt. Fiber Commun. Conf.*, Mar. 2020, pp. 1–3.
[64] F. Schreckenbach, N. Gortz, J. Hagenauer, and G. Bauch, "Optimization of symbol mappings for bit-interleaved coded modulation with iterative decoding," *IEEE Commun. Lett.*, vol. 7, no. 12, pp. 593–595, Dec. 2003.
[65] "1FINITY T600 Transport Blade," [Online]. Available: https://www.fujitsu.com/us/Images/1FINITY-T600-Data-Sheet.pdf
[66] O. Geller, R. Dar, M. Feder, and M. Shtaif, "A shaping algorithm for mitigating inter-channel nonlinear phase-noise in nonlinear fiber systems," *J. Lightw. Technol.*, vol. 34, no. 16, pp. 3884–3889, Aug. 2016.
[67] M. Karlsson and E. Agrell, "Spectrally efficient four-dimensional modulation," in *Proc. Opt. Fiber Commun. Conf.*, Mar. 2012, pp. 1–3.
[68] M. Barakatain, D. Lentner, G. Böecherer, and F. R. Kschischang, "Performance-complexity tradeoffs of concatenated fec for higher-order modulation," *J. Lightw. Technol.*, vol. 38, no. 11, pp. 2944–2953, Jun. 2020.
[69] T. Fehenberger, D. S. Millar, T. Koike-Akino, K. Kojima, and K. Parsons, "Multiset-partition distribution matching," *IEEE Trans. Commun.*, vol. 67, no. 3, pp. 1885–1893, Mar. 2019.
[70] K. Sugihara, S. Kametani, K. Kubo, T. Sugihara, and W. Matsumoto, "A practicable rate-adaptive FEC scheme flexible about capacity and distance in optical transport networks," in *Proc. Opt. Fiber Commun. Conf.*, 2016, pp. 1–3.
[71] T. Koike-Akino *et al.*, "Iteration-aware LDPC code design for low-power optical communications," *J. Lightw. Technol.*, vol. 34, no. 2, pp. 573–581, 2016.
[72] T. Koike-Akino, K. Kojima, D. S. Millar, K. Parsons, T. Yoshida, and T. Sugihara, "Pareto optimization of adaptive modulation and coding set in nonlinear fiber-optic systems," *J. Lightw. Technol.*, vol. 35, no. 4, pp. 1041–1049, Feb. 2017.
[73] J. Lin, H. Sepehrian, L. A. Rusch, and W. Shi, "Joint digital and optical pre-compensation for 720 Gb/s all-silicon IQ modulator single carrier transmission," in *Proc. Eur. Conf. Opt. Commun.*, Sep. 2018, pp. 1–3.
[74] E. P. da Silva and D. Zibar, "Widely linear equalization for IQ imbalance and skew compensation in optical coherent receivers," *J. Lightw. Technol.*, vol. 34, no. 15, pp. 3577–3586, Aug. 2016.
[75] T. Pfau, S. Hoffmann, and R. Noé, "Hardware-efficient coherent digital receiver concept with feedforward carrier recovery for m-QAM constellations," *J. Lightw. Technol.*, vol. 27, no. 8, pp. 989–999, Apr. 2009.
[76] B. P. Smith, A. Farhood, A. Hunt, F. R. Kschischang, and J. Lodge, "Staircase codes: FEC for 100 Gb/s OTN," *J. Lightw. Technol.*, vol. 30, no. 1, pp. 110–117, Jan. 2012.
[77] K. Sugihara *et al.*, "A spatially-coupled type LDPC code with an NCG of 12 dB for optical transmission beyond 100 Gb/s," in *Proc. Opt. Fiber Commun. Conf.*, Mar. 2013, pp. 1–3, Paper OM2B-4.