Tetrahedral Coding and Non-Unitary Resilience in Polarization-Multiplexed Lightwave Transmissions

Arnaud Dumenil, Elie Awwad, Cyril Measson
Nokia Bell Labs Paris Saclay, F-91620 Nozay, France
Email: arnaud.dumenil@live.com, elie.awwad@telecom-paris.fr, cyril.measson@gmail.com

***Abstract*—This paper deals with polarization-multiplexed optical transmissions. It addresses the challenge of designing optimal yet practical modulation codes for the 2×2 MIMO fiber channel. This specific use case diverges from classical MIMO models encountered, e.g., in the wireless literature as, here, non-ergodicity and low-complexity are key. While rather simple from a fundamental viewpoint, the proposed modulation schemes appears to be optimal and geometrically pleasing as the *s*econd minimum distance turns out to be the object of interest and the associated vectors encode the basis of a regular tetrahedron in the affine space. From an operational viewpoint, these new schemes for single channel use form polarization codes that achieve unit dB gains while exhibiting low implementation complexity.**

I. Introduction

This work on optimal four-dimensional coding originates from the challenges arising when designing coding and modulation schemes for state-of-the-art optical systems.

a) Background: Modeling the fiber channel [9] is a tedious task and highly depends on the underlying technology. To date, modulation solutions that directly address the fundamental non-linear nature of lightwave systems [4] have yet to be developed for commercial products. Current and next-generation technologies are based on Wavelength Division Multiplexing (WDM) systems and use polarization-multiplexing and single-mode fiber. Various impairments are specific to optical transmissions, such as non-linearity [7], [8], polarization mode dispersion (PMD), or polarization dependent loss (PDL) to name but a few.[1] Nevertheless, in terms of digital system engineering, actual coding and modulation schemes are designed for the simple one-dimensional Gaussian model. The optical transceiver operating in the linear regime is often a valid assumption. Considering an advanced yet linear MIMO model permits to address the important issue of PDL.

b) State of the Art: Communication theory [1], [2] deals with the proper modulation of information into electric waveforms. This paper is an exercise dealing with the peculiar design of specific and efficient four-dimensional codes [5], [17]–[19], [21], [22] targeted to optical applications. While space-time codes have first been developed for wireless communications, they have recently become a topic of interest for polarization-multiplexed transmissions or multi-mode fiber, see [10], [11], [13], [15]. The next-generation of optical components (waveband selective switches, ultra wide-band amplifiers) could require multi-dimensional modulation for mitigating inherent PDL. Low-complexity schemes are candidates for implementation in high-speed VLSI. Polarization codes on a single channel use are investigated in, e.g., [13] and first remarkable insights on physical and non-physical four-dimensional rotations adapted to optics are found in [6], [28].

c) Notations: Let us first collect some notations found in this paper. The information source denoted by $S \in \mathbb{R}^M$ is encoded into $X \in \mathbb{R}^M$ using unitary precoding. We use $S \in N\text{-PAM}^M \stackrel{\text{def}}{=} (2\ell+1)/\sqrt{\mathcal{E}_{N,M}} : \ell \in \{-\frac{N}{2}, -\frac{N}{2}+1, \cdots, \frac{N}{2}-2, \frac{N}{2}-1\}^M$, where N even and $\sqrt{\mathcal{E}_{N,M}} = \sqrt{\frac{(N-1)N(N+1)}{3M}}$ is a normalization power constant. For $N = 2$, the modulation points form the vertices of a hypercube in $\mathbb{R}^M$. Recall that the number of edges in the hypercube of dimension M is $M \cdot 2^{M-1}$ and its number of vertices is 2^M. For $N = 2$, the modulation points are contained within the hypercube forming a regular grid on which the minimal spacing between vertices appear $M \cdot N^{M-1} \cdot (N-1)$ times. Transmissions occur via an $M/2$-dimensional complex-valued discrete channel model. In this paper, we use the special unitary and orthogonal groups

$$\mathcal{SU}_M = \{A \in \mathbb{C}^{M\times M} : \det(A) = 1\} \subset \mathcal{U}_M,$$
$$\text{and } \mathcal{SO}_M = \{A \in \mathbb{R}^{M\times M} : \det(A) = 1\} \subset \mathcal{O}_M,$$

respectively, for which the identity matrix I_M is the neutral element. Recall that elements of $\mathcal{SU}_M$ have M^2-1 parameters or degrees of freedom, and elements of $\mathcal{SO}_M$ have $\frac{M(M-1)}{2}$ parameters. A rotation in $\mathcal{SO}_2$ with angle $\theta \in [0, 2\pi)$ is represented by the matrix $R_\theta = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \in \mathcal{SO}_2$.

II. Problem Statement

Consider the complex-valued 2×2 MIMO model given by

$$Y_t = \sqrt{\rho}\mathrm{e}^{\sqrt{-1}\phi} V_t D(\sqrt{1+\gamma}, \sqrt{1-\gamma}) U_t X_t + Z_t \quad (1)$$

at the t-th channel use, where V_t, U_t are matrices from $\mathcal{SU}_2$, Z_t is a circularly-symmetric Gaussian random noise with zero mean and unit covariance matrix [3] that is independent and identically distributed over t, $D(\sqrt{1+\gamma}, \sqrt{1-\gamma})$ with $\gamma \in [0, 1)$ is a real-valued diagonal matrix with positive diagonal elements such that ρ represents the normalized signal-to-noise

[1]The general frequency flat hypothesis with small PMD and lumped noise is considered. The fiber linear regime is treated as operational model with approximate Gaussian noise.

(SNR) ratio, and $\phi \in [0, 2\pi)$ is a random phase.[2] This model applies to polarization-multiplexed optical transmissions[3] experiencing non-unitary impairments ($\gamma > 0$). In standard practice, the channel input X_t is a complex-valued discrete random vector with entries taking on values for example in square N^2-QAM or, equivalently, in N-PAM^2. In this paper, we focus on orthogonal codes, hence the channel inputs take on values in $F \cdot (N\text{-PAM}^4)$, where $F \in \mathcal{SO}_4$. The random matrices V_t, U_t are picked uniformly at random from $\mathcal{SU}_2$ as well as ϕ_t from $[0, 2\pi)$. These parameters are very slowly changing when compared to the operational blocklengths encountered in optical networks. Hence the multiplicative factor may be considered as constant over t and the channel resembles a block-fading model up to rescaling. System and code performance analysis shall be specifically tailored to these features. More precisely, U_t represents the *s*tate of polarization (SOP), and $D_\gamma \stackrel{\text{def}}{=} D(\sqrt{1+\gamma}, \sqrt{1-\gamma})$ represents the imbalance of the electric field in a given direction of the polarization plane. Optical links are examples of communication links that are non-adaptive and for which no feedback is available. Therefore transmissions occur in a block-wise manner using a predefined coding and modulation scheme. As a result, because the imbalance coefficient $\gamma \in [0, 1)$ is (slowly) varying, it is at first order not the ergodic channel behavior but rather its *worst case* value that limits the transmission flow by determining streaming *outage*.

The main objective of this paper is to construct a practical linear modulation code (or precode) [17], [19] that transforms a standard N-PAM-multiplexed source[4] S_t into an *imbalance-resilient* signal. For the sake of complexity reduction, we focus on precoding over a single channel use (e.g., one time slot if the degree of freedom t represents the time) as $F : S_t \mapsto X_t = FS_t$. We then want to find the optimal orthogonal [18] code $F \in \mathcal{SO}_4$ that minimizes the worst case error probability.

III. Invariance and Fundamental Limits

a) Physical Model: Due to the noise circular symmetry, $\mathcal{SU}_2$-rotational invariance makes that the statistical channel model in Eq. 1 is equivalent to the model

$$Y_t = \sqrt{\rho} D_\gamma U_t X_t + Z_t,$$

and further equivalent to

$$Y = \sqrt{\rho} D_\gamma R_\alpha B_\beta X + Z \tag{2}$$

after skipping the indices t for simplicity and exploiting the classical $\mathcal{SU}_2$ decomposition $U_t = B_{\beta'} R_\alpha B_\beta$ using a product of 3 elementary single-parameter unitary matrices over $\mathcal{SU}_2$, where $\alpha, \beta, \beta' \in [0, 2\pi)$, see [13], [16]. In the context of optical communications, this decomposition of the SOP has an operational meaning as B_β represents a birefringence matrix and R_α the polarization rotation. Because $B_{\beta'}$ and D_γ commute, rotational invariance gives the model equivalence.

b) Capacity and Polarization-Dependent Loss: In the case of continuous alphabet with average power constraint, the singular value decomposition leads to the classical MIMO capacity formulae [3]. In short, if $VD_\gamma U$ is fixed (known by the receiver but unknown at the transmitter), then the capacity is obtained by maximizing the mutual information $\mathcal{I}(X; (Y, D_\gamma U)) = \mathcal{H}(Y|D_\gamma U) - \log(2\pi\mathrm{e})$ using X as a Gaussian with covariance $\frac{1}{2}I$. Then, it decomposes as

$$C(\gamma) \stackrel{\text{def}}{=} 2\log(1+\rho/2) + \log\left(1 - \gamma^2 \frac{\rho^2}{(2+\rho)^2}\right),$$

where the second term defines the so-called *polarization-dependent loss* (PDL) from the viewpoint of information theory. Observe that the continuous capacity only depends on the imbalance parameter γ. It is generally given in dB as $\Gamma \stackrel{\text{def}}{=} 10\log_{10}(\frac{1+\gamma}{1-\gamma})$, and does not depend on the particular orientation defined by U.

c) Operational Information Rate and Directional Loss: It is important to observe that, for practical discrete inputs such as modulation formats defined on regular square lattices, the information rate becomes highly sensitive to the orientation of the underlying grid [12]. This is illustrated in Fig.1 where the large 6dB PDL figure exemplifies an extreme *worst* case[5] that could be observed in next-gen systems. The code constructions

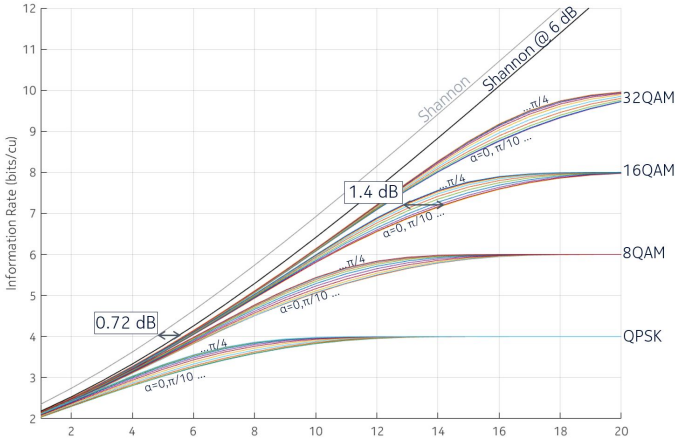

Fig. 1. Information rate as a function of the SNR for $\Gamma = 6$dB and standard dual-polarized QAM. The achievable information also depends on the SOP (in addition to the continuous PDL loss). For 16-QAM^2 = 4-PAM^4, up to 1.4dB SNR variation is reported and optical coding and modulation schemes have to to address this worst case. Polarization coding aims at minimizing the performance variation, hence effectively reducing the loss due to PDL.

presented in this paper aim at reducing the dependency on the underlying angles of $U \in \mathcal{SU}_2$. Because optical system are feed-forward communication systems characterized by a quasi-zero failure rate, the performance measure of choice is connected to the outage in capacity represented by the *worst case information rate*.

[2]The set $\{\mathrm{e}^{\sqrt{-1}\phi}, \phi \in [0, 2\pi)\}$, when used for multiplying $\mathcal{SU}_2$, permits to describe the entire group $\mathcal{U}_2 = \{A \in \mathbb{C}^{2\times 2} : |\det(A)| = 1\}$. For presentation convenience, we simply consider channel matrices in $\mathcal{SU}_2$ instead of $V_t, U_t \in \mathcal{U}_2$ which, by invariance, describe the equivalent channel.

[3]A general channel with $D(\gamma_1, \gamma_2, \cdots, \gamma_M)$ such that $\gamma_i > 0, \sum \gamma_i^2 = 1$ models forward-looking mode-multiplexed optical transmissions.

[4]The $(N^2\text{-QAM})^2$ choice for the source alphabet is also motivated by legacy and flexibility aspects of actual transceivers.

[5]The worst 6dB PDL case corresponds to a 2dB average PDL and an outage probability of 10^-5 assuming a link with i.i.d. PDL elements.

IV. Non-Physical Equivalences using Left-Isoclinic Orthonormal Codes

Let us now represent the transmission over the linear 2×2 polarization-multiplexed fiber channel model using four-dimensional real-valued vectors. The equivalent random input is $\hat{X} \in N\text{-PAM}^4 \subset \mathbb{R}^4$ and is assumed to have normalized power constraint $\mathbb{E}[X] = 1$. From the previous section, we see that the multiplicative non-unitary impairment can be expressed using the linear map $\hat{X} \mapsto \hat{D}_\gamma \hat{R}_\alpha \hat{B}_\beta \hat{X}$ and real-valued matrices of $\mathcal{SO}_4$ obtained from $\mathcal{SU}_2$ as

$$\hat{D}_\gamma \stackrel{\text{def}}{=} \begin{pmatrix} \sqrt{\gamma+1}I_2 & 0 \\ 0 & \sqrt{1-\gamma}I_2 \end{pmatrix}, \hat{B}_\beta \stackrel{\text{def}}{=} \begin{pmatrix} R_\beta & 0 \\ 0 & R_{-\beta} \end{pmatrix},$$

$$\text{and } \hat{R}_\alpha \stackrel{\text{def}}{=} \begin{pmatrix} \cos(\alpha) & 0 & -\sin(\alpha) & 0 \\ 0 & \cos(\alpha) & 0 & -\sin(\alpha) \\ \sin(\alpha) & 0 & \cos(\alpha) & 0 \\ 0 & \sin(\alpha) & 0 & \cos(\alpha) \end{pmatrix}.$$

The matrices respectively represent the polarization-imbalance, the birefringence, and the polarization rotation.

The main purpose of this paper is to construct an orthogonal code using $F \in \mathcal{SO}_4 : N\text{-PAM}^4 \to \mathbb{R}^4$ that maps $S \mapsto \hat{X} = FS$. This (pre-)code is non-physical as it cannot be obtained from physical optical components that act directly on the two-dimensional complex-valued lightwave. Instead, precoding using $\mathcal{SO}_4$ requires implementation in digital signal processing [6]. The design criteria is to minimize the probability of error (maximize the information rate) per channel use while precoding data on a single channel use basis. In the large SNR regime, Euclidean geometry arguments can be used to characterize the information rate. Then, we want to find the (not necessarily unique) code F that minimizes the worst error probability across channel realizations,

$$F^* = \text{argmax}_{F\in\mathcal{SO}_4} \Big\{ \min_{U\in\mathcal{SU}_2} \{\mathcal{I}(S; \hat{D}_\gamma \hat{U} FS + Z)\} \Big\},$$

when $\rho >> 1$. Recall that $\mathcal{SO}_4$ is generally described by 6 parameters. The 6 degrees of freedom represent the 6 possible combinations of elementary 1-parameter rotations describing elements of $\mathcal{SO}_4$. It is explicit using the Van Elfrinkhof formula or the connection to quaternions, see, e.g., [6], [16], it can also be observed using Hopf coordinates. Therefore, the raw optimization above would involve 6 parameters. Fortunately, by equivalences, we will see that the number of degrees of freedom reduces from 6 down to 2. In practice, those 2 remaining degrees of freedom are referred to as non-physical in [6]. Notice that a *numerical* optimization is performed in [20]. In this paper, we present a closed-form solution via an *analytical* optimization over $\mathcal{SO}_4$ that is an optimality proof.

Let us now explain the optimization complexity reduction. From classical linear algebra, see also [6], we know that any matrix $F \in \mathcal{SO}_4$ can be written as the commutative product $F = LR$ where $L \in \mathcal{SO}_4$ is a left-isoclinic matrix and $R \in \mathcal{SO}_4$ is a right-isoclinic matrix. The matrices R and L are unique up to the multiplication by the central inversion. The subgroups of both right-isoclinic and left-isoclinic are both isomorphic to the manifold $\mathcal{S}^3 \stackrel{\text{iso}}{=} \mathcal{SU}_2$ (also isomorphic to the corresponding spinors or the unit quaternions groups). More precisely, $\mathcal{SO}_4$ is a double cover of $\mathcal{SO}_3 \times \mathcal{SO}_3$ (i.e., $\mathcal{SO}_3 \times \mathcal{SO}_3 \stackrel{\text{iso}}{=} \mathcal{SO}_4/\mathbb{Z}_2$) and $\mathcal{SU}_2$ is a double cover of $\mathcal{SO}_3 \stackrel{\text{iso}}{=} \mathcal{SU}_2/\mathbb{Z}_2$. More operationally and interestingly for us, the channel parameters from $\mathcal{SU}_2$ cover the parameters associated with the right-isoclinic rotations. Hence, if we decompose $F = RL$, then, by commutativity of R and L, the code optimization for the four-dimensional channel $\hat{Y} = \hat{D}_\gamma \hat{U} R(LS) + \hat{Z}$ is equivalently conveyed using $\hat{Y} = \hat{D}_\gamma LS + \hat{Z}$. Simple algebraic manipulations performed with Hopf coordinates, see, e.g., [20], permit to express the 3 remaining degrees of freedom of L as

$$L = \begin{pmatrix} \sin(\eta)R_\theta & -\cos(\eta)S_\nu \\ \cos(\eta)S_\nu & \sin(\eta)R_\theta \end{pmatrix}$$
$$= \begin{pmatrix} R_\theta & 0 \\ 0 & R_\theta \end{pmatrix} \cdot \begin{pmatrix} \sin(\eta)I_2 & -\cos(\eta)S_\nu \\ \cos(\eta)S_\nu & \sin(\eta)I_2 \end{pmatrix} = \hat{R}_\theta F_{\eta,\nu},$$

where $R_\theta \in \mathcal{SO}_2$ is a rotation and $S_\nu = \begin{pmatrix} \cos(\nu) & \sin(\nu) \\ \sin(\nu) & -\cos(\nu) \end{pmatrix} \notin \mathcal{SO}_2$ is a reflection. The first factor R_θ represents a phase change that is common to the two physical complex-valued polarization entries. Hence, as already mentioned in footnote, by rotational invariance, the phase θ can be absorbed into an equivalent physical channel Therefore, it is omitted in our running optimization problem that reduces to the 2 physical parameters of the (non-physical) transform $\hat{F}_{\eta,\nu}$ with

$$F_{\eta,\nu} = \sin(\eta)I_4 + \cos(\eta) \begin{pmatrix} 0 & -S_\nu \\ S_\nu & 0 \end{pmatrix}.$$

This 2-parameter formulation suffices to define any unitary (left-isoclinic) polarization coding of $S \in \mathbb{R}^4$ as $S \mapsto \hat{X} = F_{\eta,\nu}S$. It is now time to reformulate our optimization objective in terms of Euclidean geometry. Under the technical large SNR assumption, it can be shown that the optimization of the distance profile dominates the information rate behavior (alternatively, this could also be seen via union bounding techniques on the probability of error). The optimal construction nails down to an optimization using the parameter pair (η, ν) of the distance profile of codes $\{S \mapsto \hat{D}_\gamma \hat{R}_\alpha \hat{B}_\beta F_{\eta,\nu} S\}_{\alpha,\beta}$.

V. Optimal Left-Isoclinic Polarization Codes

The code construction over $\mathcal{SO}_4$ and a single t-slot is remarkable as the actual worst minimum distance value of $\{\hat{D}_\gamma \hat{U} FS : S \in N\text{-PAM}^4\}$ cannot be improved. We first state our main theorem, then present geometric arguments.

a) Main Result: It is obtained as an optimization over the distance distribution: the number of occurrences of the minimum distance is reduced and the value of the *second* minimum distance is maximized. A detailed proof with coordinate calculus is deferred to the appendix.

Theorem 1 (Optimal Construction over $\mathcal{SO}_4$): Using the formalism of previous sections, and up to algebraic symmetry, an orthonormal construction $F_{\eta,\nu}$ over $N\text{-PAM}^4$ that, asymptotically in the signal-to-noise ratio, maximizes the information rate for $\Gamma = 10\log_{10}\frac{1+\gamma}{1-\gamma} << 1$ is obtained for $\nu^* = \frac{\pi}{4}$

and $\eta^* = \frac{1}{2}\cos^{-1}(\frac{\sqrt{3}}{3})$. This construction is then optimal as it optimizes the Euclidean distance distribution of the $32 \cdot (N-1)$ smallest minimum distances for $\Gamma << 1$. Up to the original constellation power scaling $\mathcal{E}_{4,N}$, the smallest minimum distance is achieved $N^3(N-1)$ times in one of the four basis directions as $\sqrt{1-\gamma}$ and the second smallest distance is achieved $3 \cdot N^3(N-1)$ times in the complementary space as $\sqrt{1+\gamma/3}$.

This geometric construction is justified in two steps. The $N = 2$ case provides the guidelines for the general proof.

First, as discussed in previous section, we consider small γ values that deviates only slightly from 0. This is justified from practice as typical extremal PDL values $\Gamma = 10\log_{10}\frac{1+\gamma}{1-\gamma}$ are observed to be lower than 5dB. See also [14] for more complete considerations. Hence, the study of the deformation of the edges of the hypercube is sufficient. The 4-dimensional hypercube fundamentally defines the 4-dimensional square lattice: it gets slightly distorted by the action of $\hat{D}_\gamma$ and contraction occurs in a two-dimensional subspace of $\mathbb{R}^4$. Focus first on 2-PAM[4]. The hypercube edges define the minimum distance of the modulation. Therefore, it suffices to study their deformation via directional scaling to estimate the evolution of the minimum distance. The code construction reduces to the optimization of the distribution of the $4 \times 8 = 32$ smallest minimum distances[6] (equi-distributed on the 4 orthogonal edge directions). Hence we are interested in the image of each of the 4 elementary basis vectors of $\mathbb{R}^4$ after the action of a fixed $F_{\eta,\nu}$ left-multiplied by the action of a random $U \in \mathcal{SU}_2$. As shown in Appendix A and Lemma 1, there is always one of the 4 directions that will experience the worst contraction. This means that the (worst) minimum distance scales as $\sqrt{1-\gamma}$. Hence, as further discussed in Appendix A, this distance can be encountered not less that 1×8 times for 2-PAM or $1 \times N^3 \cdot (N-1)$ for the general N-PAM case.

Second, while the minimum distance is defined by $\sqrt{1-\gamma}$, the second minimum distance may vary as a function of (η, ν). In fact, Appendix B shows that the sum of the squared distances observed of the 3 remaining axes scales as $1+\gamma$ and balances the minimum distance energy. Indeed, using the properties of four-dimensional rotations and Lemma 1, the (η, ν) optimization takes place in a three-dimensional subspace that is orthogonal to the contracting dimension. In this subspace, one direction is contracting while the remaining two are expanding. The maximization of the second minimum distance consists in orientating 3 orthonormal vectors in $\mathbb{R}^3$ in such a way that their orthogonal projections onto the contracting axis have minimal norm. This is illustrated in Fig. 2. An optimal placing in the three-dimensional affine space is achieved when the formed equilateral triangle becomes perpendicular to the contracting direction (z-axis in Fig. 3). Exact rotation angles are derived in Appendix B.

[6]The $M \cdot 2^{M-1}$ edges of the M-dimensional hypercube form $4 \times 8 = 32$ edges for $M = 4$ and correspond to minimum distances for square 2-PAM[4]. More generally, for square N-PAM^M, there are $M \cdot N^{M-1} \cdot (N-1)$ minimum distances, i.e., $4 \times N^3 \cdot (N-1)$ for $M = 4$.

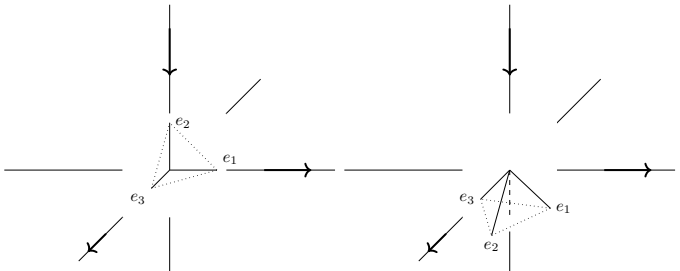

Fig. 2. Left: No precoding is performed. Right: Because there is always one basis vector that gets aligned with one of the contracting directions and get scaled (contracted) by $\sqrt{1-\gamma}$, the degrees of freedom of $F_{\eta,\nu}$ are used to orientate the remaining 3 basis vectors (in this example, e_1, e_2, and e_3) in such a way that they experience the smallest possible contraction.

b) Discussion: Notice that the angle η^* in Theorem 1 is exactly equal to 4 times the so-called *tetrahedral angle*. Appendix B gives the coordinate system

$$\tilde{M}(\eta^*, \nu^*) = \begin{pmatrix} -\sqrt{\frac{2}{3}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} \end{pmatrix} \in \mathcal{SO}_3$$

that represents 3 orthonormal basis vectors of $\mathbb{R}^3$. It can be regarded as the basis triangle of a family of tetrahedrons with summits on the z-axis. In Fig. 3, the regular tetrahedron associated with a 4-th point $(0, 0, \sqrt{3}/3)$ is represented. The construction has a physical interpretation using the normalized sphere centered at the tetrahedron center O.

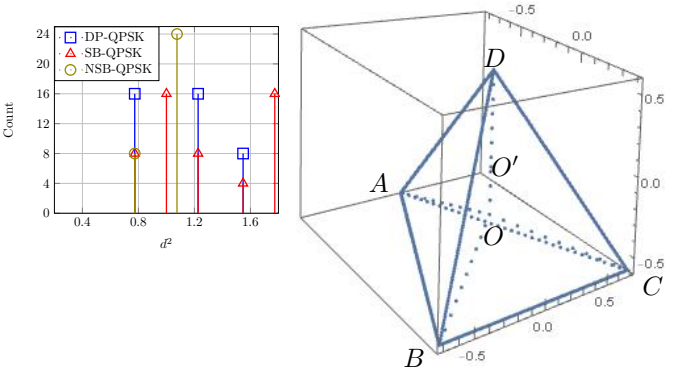

Fig. 3. Left: Squared distance profile for different polarization codes over 2-PAM[4] and $\Gamma = 2$dB. DP indicates dual-polarization (no polarization code), SB refers to the simple *spatially-balanced* code in [14] (for which the number of minimum distance occurrences reduces to the minimum of 8), and NSB refers to *new spatially-balanced* that is the *tetrahedral* code (the minimum distance occurs 8 times, but the value of the second minimum distance that occurs 24 times is now maximized). Right: Representation of the 3 orthonormal basis vectors as the basis of a regular tetrahedron ABCD with fourth vertex $(0, 0, \sqrt{3}/3)$. The coordinate origin is O'. The tetrahedron center O is located on the z-axis at $z = -\sqrt{3}/6$. The 3 vectors $\vec{O'A}, \vec{O'B}, \vec{O'C}$ remain orthonormal in the three-dimensional Euclidean space. Their joint orientation (up to reflection) maximizes the post-scaling second minimum Euclidean distance.

VI. Applications and Perspectives

The presented polarization codes have been validated by simulations for various N and lab experiments for $N \in \{2, 4\}$. The results reported in [13], [20] show expected sensitivity SNR gains in the [0.5 1] range (in dB) for 4-QAM[2] or 16-QAM[2] experimental transmissions. Therefore, the performance in experiments that emulate next-generation WDM

systems validates these optimal constructions. Implementation in high-speed transceivers should be facilitated by the low-complexity of orthogonal coding over a single t-slot. Finally, for the practice of next-generation optical transceivers involving high-resolution A/D converters, notice that directional gains from orthogonal polarization codes arise when the modulation constellation is supported by an underlying cubic lattice. The presented *coding* gains are naturally and efficiently complemented by *shaping* gains obtained from probabilistically non-uniform signaling [23]–[27]. These aspects make the presented optimal four-dimensional modulation schemes very relevant for next-gen optical systems.

APPENDIX A
PRESERVATION OF THE MINIMUM OF $d^{(1)}_{\text{MIN}}$ AFTER SCALING

Let $\mathcal{B} = \{e_1, e_2, e_3, e_4\}$ be the canonical basis of $\mathbb{R}^4$, i.e., $e_i \in \mathbb{R}^4$ is a unitary column vector that has all but one entries equal to zero, its non-zero entry being 1 at the i-th position.

Lemma 1: For any $F_{\eta,\nu}$ and any $e_i \in \mathcal{B}$, there are $e_j \in \mathcal{B}$ and $U = U(e_j, e_i) \in \mathcal{SU}_2$ such that $\hat{U}F_{\eta,\nu}e_j = e_i$, $\hat{U}F_{\eta,\nu}\text{span}(\mathcal{B} \setminus \{e_j\}) = \text{span}(\mathcal{B} \setminus \{e_i\})$, and the submatrix formed by all but the i-th row and the j-th column is $(\hat{U}F_{\eta,\nu})_{\sim i,\sim j} \in \mathcal{SO}_3$.

Proof 1: Without loss of generality, let us restrict the proof to $e_2 \stackrel{\text{def}}{=} (0,1,0,0)^T$. Then, for any $F_{\eta,\nu}$, $F_{\eta,\nu}e_2 = (\sin(\eta), 0, \cos(\eta)\cos(\nu), \cos(\eta)\sin(\nu))^T$. Setting $\hat{U} \stackrel{\text{def}}{=} \hat{R}_{\pi-\eta}\exp(\hat{-}i\nu/2)\hat{B}_{\nu/2}$ gives the result as $\hat{U}F_{\eta,\nu}e_2 = e_4 \stackrel{\text{def}}{=} (0,0,0,1)^T$.

Now, consider the original discrete N-PAM^4 signal points in $\mathbb{R}^4$ and a given $F_{\eta,\nu}$. We are interested in the minimum Euclidean distance between any pair of points and, more precisely, in its minimum as a function of (η, ν). For sufficiently small γ, continuity arguments show that the minimum distance is supported by (some of) the edges of the distorted four-dimensional hypercube grid that gets $\hat{U}$-rotated in one of the two $\sqrt{1-\gamma}$-compression directions. When $\gamma = 0$, it is found $4 \times 2^{4-1} = 4 \times 8 = 32$ times in the 2-PAM case, and $4 \times N^3(N-1)$ in the general PAM case. When $\gamma > 0$, the non-zero scaling in certain directions modifies the profile of the minimum distance and the multiplicity order. For $\gamma > 0$, depending on uniquely $\hat{U}$ and $F_{\eta,\nu}$, the minimum distance appears at least $1 \times N^3(N-1)$ times.

Corollary 1: For small $\gamma > 0$, the randomly distributed $U \in \mathcal{SU}_2$ makes that the *worst* minimum Euclidean distance is $d_{\text{min}} = \frac{2}{\sqrt{\mathcal{E}_{N,4}}}\sqrt{1-\gamma}$, where $\frac{2}{\sqrt{\mathcal{E}_{\mathcal{N},\triangle}}}$ is the normalized length of the double-unit hypercube edge. I.e., in the distance profile, d_{min} scales as $\sqrt{1-\gamma}$ and is encountered at least 1×8 times for 2-PAM or, generally, $1 \times N^3(N-1)$ times.

Proof 2: Geometric alignment induces the maximal reduction in edge norm. This alignment can always be achieved as shown in 1. Among the 4 sets of parallel edges of the grid contained in the hypercube, 1 set is necessarily associated with one given (contracting) direction. Hence the minimum distance is achieved at least 8 times for 2-PAM or, generally, $1 \times N^3(N-1)$ times.

APPENDIX B
SECOND MINIMUM DISTANCE $d^{(2)}_{\text{MIN}}$ OPTIMIZATION

Lemma 1 shows that there is at least one pair (α^*, β^*) that aligns any $\mathcal{SO}_4$-transform of a hypercube edge with one of the two contracting directions e_3, e_4 by left multiplication by $U^* \stackrel{\text{def}}{=} U(\alpha^*, \beta^*)$. The associated distances, obtained after left multiplication by $\hat{D}_\gamma U^*$, give the minimum distance with maximal contraction scaling $\sqrt{1-\gamma}$. It remains to choose η, ν that maximize the second minimum distance. This can be done by simple calculus using the sub-matrices from Lemma 1. Without loss of generality, let us for instance consider

$$M(\eta,\nu) \stackrel{\text{def}}{=} \begin{pmatrix} -\sin(2\eta) & 0 & \cos(2\eta)\cos(\nu) & \cos(2\eta)\sin(\nu) \\ 0 & 0 & \sin(\nu) & -\cos(\nu) \\ -\cos(2\eta) & 0 & -\sin(2\eta)\cos(\nu) & -2\sin(2\eta)\sin(\nu) \\ 0 & 1 & 0 & 0 \end{pmatrix},$$

for which the basis vectors e_1, e_2, and e_3 represent the 3 sets of remaining edges associated with the original hypercube grid. Rotating this basis in $\mathbb{R}^3$ will permit us to maximize the second minimum distance as compression in the remaining $\mathbb{R}^3$ subspace occurs only in one direction. To do this, we use the submatrix of $M(\eta,\nu)$ that lies in $\mathcal{SO}_3$. We are interested in $\max_{\eta,\nu}\left\{\min_{i\in\{1,3,4\}}\left\{||\hat{D}_\gamma U^*(\eta,\nu)F_{\eta,\nu}e_i||\right\}\right\}$. For any $i \in \{1,3,4\}$, we have $||\hat{D}_\gamma U^*(\eta,\nu)F_{\eta,\nu}e_i||^2 = (1+\gamma)||d_i(\eta,\nu)||^2 + (1-\gamma)||a_i(\eta,\nu)||^2 = (1+\gamma)(1 - ||a_i(\eta,\nu)||^2) + (1-\gamma)||a_i(\eta,\nu)||^2$, where $d_i = d_i(\eta,\nu) \in \text{span}(e_1, e_2)$ and $a_i = a_i(\eta,\nu) \in \text{span}(e_3)$ are orthogonal projections of the unit-norm vector $U^*(\eta,\nu)F_{\eta,\nu}e_i$. A key observation follows. Maximizing over η,ν any of the 3 distances $||\hat{D}_\gamma U^*(\eta,\nu)F_{\eta,\nu}e_i||$ is equivalent to minimizing $|a_i(\eta,\nu)|^2$ over η,ν. Moreover, comparing the 3 distances $||\hat{D}_\gamma U^*(\eta,\nu)F_{\eta,\nu}e_i||$ to each other reduces to ordering the 3 values $\{|a_i(\eta,\nu)|\}_{i\in\{1,3,4\}}$. We have $|a_1|^2 = \cos(2\eta)^2 = 1 - \sin(2\eta)^2$, $|a_3|^2 = 4\cos(\eta)^2\cos(\nu)^2\sin(\eta)^2 = \cos(\nu)^2\sin(2\eta)^2$, and $|a_4|^2 = 4\cos(\eta)^2\sin(\nu)^2\sin(\eta)^2 = \sin(\nu)^2\sin(2\eta)^2$, which, because $\sum_{i\in\{1,3,4\}}|a_i|^2 = 1$, shows that $\sum_{i\in\{1,3,4\}}||\hat{D}_\gamma U^*(\eta,\nu)F_{\eta,\nu}e_i||^2 = 3+\gamma$, i.e., the sum of the squared norm of the second minimum distance always balances the squared norm of the first minimum distance.This means that $\sqrt{1+\gamma/3}$ is the largest possible scaling of the second minimum distance, which is then encountered 3 times. This maximum is achieved for $\nu^* = \frac{\pi}{4}$ and

$$\eta^* = \frac{1}{2}\cos^{-1}\left(\frac{1}{\sqrt{3}}\right) = \frac{1}{4}\cos^{-1}\left(-\frac{1}{3}\right) = \frac{1}{2}\tan^{-1}(\sqrt{2}),$$

where $4 \times \eta^*$ is called *tetrahedral angle* in chemistry. We see that the $3 \times N^3(N-1)$ second minimum distances scale as $\sqrt{1+\gamma/3}$ and the original unit quaternion matrix M evaluated in η^*, ν^* is

$$M(\eta^*,\nu^*) = \begin{pmatrix} -\sqrt{\frac{2}{3}} & 0 & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ 0 & 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{3}} & 0 & -\frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} \\ 0 & 1 & 0 & 0 \end{pmatrix}.$$

ACKNOWLEDGMENT

This work has been partially funded by ANR-MUSICO.

REFERENCES

[1] R. G. Gallager, *Information theory and reliable communication*. New York: Wiley, 1968.

[2] J.M. Wozencraft and I.M. Jacobs. *Principles of Communication Engineering*. New York: Wiley, 1965.

[3] E. Telatar, 'Capacity of Multi-Antenna Gaussian Channels,' *Eur. Trans. on Telecomm.*, vol. 10, no. 6, pp. 585—595, Nov.–Dec. 1999.

[4] M. I. Yousefi and F. R. Kschischang, "Information Transmission Using the Nonlinear Fourier Transform, Part I: Mathematical Tools," *IEEE Trans. Inf. Theory*, vol. 60, no. 7, pp. 4312–4328, July 2014.

[5] S. K. Mohammed, E. Viterbo, Y. Hong, and A. Chockalingam, "MIMO Precoding With X- and Y-Codes," *IEEE Trans. on Inf. Theory*, vol. 57, no. 6, June 2011, pp. 3542-3566.

[6] M. Karlsonn, ''Four-Dimensional Rotations in Coherent Optical Communications," *Journal of Lightwave Technology*, vol. 32, no. 6, pp. 1246–1257, 2014.

[7] G.P.Agrawal, "Nonlinear Fiber Optics," 5-th Edition, *Academic Press*, Oct. 2012.

[8] R. Dar, M. Feder, A. Mecozzi, and M. Shtaif, "Properties of nonlinear noise in long dispersion-uncompensated fiber links," *Optics Express*, vol. 21, no. 22, pp. 25685–25699, Oct. 2013.

[9] R.J. Essiambre, G. Kramer, P.J. Winzer, G.J. Foschini, B. Goebel, "Capacity limits of optical fiber networks," *IEEE J. Lightwave Technol.*, vol. 28, no. 4, pp. 662–701, 2010.

[10] S. Mumtaz, G. Rekaya-Ben Othman, and Y. Jaouën, ''Space-Time Code for Optical Fiber Communication with Polarization Multiplexing," *IEEE Int. Conf. on Comm. (ICC)*, Cape Town, pp. 1-5, May 2010.

[11] E Awwad, Y. Jaouën and G. Rekaya-Ben Othman "Polarization-time coding for PDL mitigation in long-haul PolMux OFDM systems," *Optics Express, OSA,* vol. 21, no. 19, pp. 22773–22790, 2013.

[12] A. Dumenil, E. Awwad, C. Measson, "Polarization Dependent Loss: Fundamental Limits and How to Approach Them," *Signal Processing in Photonic Commun. Conf.,* New Orleans, Louisiana, USA, Jul. 2017.

[13] A. Dumenil, E. Awwad, and C. Measson, ''Low-Complexity Polarization Coding for PDL-Resilience," *Eur. Conf. on Opt. Comm. (ECOC)*, Rome, Sep. 2018.

[14] A. Dumenil, E. Awwad, and C. Measson, "Low-Complexity PDL-Resilient Signaling Design," *Signal Processing in Photonic Commun. Conf.,* OSA Advanced Photonics Congress, SpTh2E.3, Jul. 2019.

[15] T. Oyama, G. Huang, H. Nakashima, Y. Nomura, T. Takahara, and T. Hoshida, ''Low-Complexity Low-PAPR Polarization-Time Code for PDL Mitigation," *Opt. Fiber Conf. (OFC)*, San Diego, 2019.

[16] A. Perez-Garcia and F. Thomas, ''On Cayley's Factorization of 4D Rotations and Applications," *Advances in Applied Clifford Algebras*, vol. 27, no. 1, pp. 523–538, 2017.

[17] V. Tarokh and N. Seshadri and A. R. Calderbank, ''Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Trans. Inf. Theory*, vo. 44, no. 2, pp. 744–765, 1998.

[18] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, ''Space–time block codes from orthogonal designs." *IEEE Trans. on Inf. Theory* vol. 45, no. 5, Jul. 1999.

[19] S.M. Alamouti, ''A simple transmit diversity technique for wireless communications," *IEEE Journal on Sel. Areas in Comm.*, vol. 16, no. 8, pp. 1451-–1458, Oct. 1998.

[20] A. Dumenil, E. Awwad, and C. Measson, ''Low-Complexity PDL-Resilient Signaling Design," *OSA Advanced Photonics Congress, Optical Society of America*, paper SpTh2E.3, 2019.

[21] F. Oggier, "On the Optimality of the Golden Code," *Inf. Theory Workshop (ITW)*, Punta del Este, pp 468–472, Nov. 2006.

[22] J. C Belfiore, G. Rekaya, and E. Viterbo "The golden code: a 2 times 2 full-rate space-time code with nonvanishing determinants," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1432—1436, April 2005.

[23] F. R. Kschischang and S. Pasupathy, "Optimal Nonuniform Signaling for Gaussian Channels," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 913–929, May 1993.

[24] A. R. Calderbank and L. H. Ozarow, "Non-equiprobable signaling on the Gaussian channel," *IEEE Trans. Inf. Theory*, vol. 36, no. 4, pp. 726–740, Jul. 1990.

[25] F. Buchali, G. Böcherer, W. Idler, L. Schmalen, P. Schulte, F. Steiner, "Experimental Demonstration of Capacity Increase and Rate-Adaptation by Probabilistically Shaped 64-QAM," *ECOC*, Aug. 2015.

[26] A. Ghazisaeidi, I. Fernandez de Jauregui, R. Rios-Mueller, L. Schmalen, P. Tran, P. Brindel, A. Carbo Meseguer, Q. Hu, F. Buchali, G. Charlet, and J. Renaudier, "65Tb/s Transoceanic Transmission using Probabilistic Shaping," *ECOC*, Sep. 2016.

[27] J. Boutros, F. Jardel, amd C. Measson, "Probabilistic Shaping and Non-Binary Codes," *Int. Symp. Inf. Theory (ISIT)*, Jul. 2017.

[28] T. Eriksson, T. Fehenberger, P. Andrekson, M. Karlsson, N. Hanik, and E. Agrell, "Impact of 4D Channel Distribution on the Achievable Rates in Coherent Optical Communication Experiments," *IEEE J. Lightwave Technol.*, vol. 34, pp. 2256–2266 , May 2016.