Polarization dependent loss in next-generation optical networks : challenges and solutions

Arnaud Dumenil

INSTITUT POLYTECHNIQUE DE PARIS

NNT : 2020IPPAS006

Thèse de doctorat

NOKIA Bell Labs

TELECOM SudParis

IP PARIS

Polarization Dependent Loss in Next-Generation Optical Networks: Challenges and Solutions

Thèse de doctorat de l'Institut Polytechnique de Paris
préparée à Telecom SudParis

École doctorale n°626 Institut Polytechnique de Paris (ED IP Paris)
Spécialité de doctorat : Electronique et Optoélectronique

Thèse présentée et soutenue à Palaiseau, le 15/05/20, par

ARNAUD DUMENIL

Composition du Jury :

Iryna Andriyanova (présidente de jury) Professeur, ENSEA	Examinatrice
Badr-Eddine Benkelfat Professeur, Télécom SudParis	Directeur de thèse
Magnus Karlsson Professor, Chalmers University of Technology	Rapporteur
Darko Zibar Associate professor, Technical University of Denmark	Rapporteur
Ghaya Rekaya-Ben Othman Professeur, Télécom Paris	Examinatrice
Elie Awwad Maître de conférence, Télécom Paris	Encadrant de thèse

Invités :

Cyril Méasson OQ Technology	Encadrant de thèse
Jeremie Renaudier Nokia Bell Labs	

Polarization Dependent Loss in Next-Generation Optical Networks: Challenges and Solutions

(Des solutions à de nouveaux défis dans les réseaux optiques modernes avec perte dépendant de la polarisation)

Arnaud Dumenil

March 2020

Abstract

Large amounts of ever-increasing global data traffic require sound and reliable communication channels. Optical terrestrial networks and submarine links are at the very heart of the global telecom infrastructure, and carry hundreds of frequency channels modulated at very high rates. These links include not only kilometers of fiber but also optical elements such as Erbium-Doped Fiber-Amplifiers (EDFA) to amplify periodically the attenuated signals and Wavelength Selective Switches (WSS) to route the signals to their assigned destinations. In this thesis, we explore a specific rate-degrading impairment of the optical propagation that raises in those systems. Discrete optical elements often exhibit a polarization anisotropy that emerges as a gain or loss imbalance between the two polarization tributaries of the polarization-multiplexed optical signals. This non-unitary effect called Polarization Dependent Loss (PDL) impairs the quality of transmission in current and next-generation optical systems.

In the context of polarization-multiplexed signals, we assess the capacity loss induced by PDL. First, channel models are carefully studied and two approaches are described: a channel with distributed PDL elements and distributed noise or a simplified single-element equivalent channel. Making use of these models, we then analyze their fundamental limits of communications. We show that the PDL channel capacity depends in practice on the state-of-polarization orientation of the incident signal. We then review the state-of-the art of PDL-mitigating modulation schemes and propose two new multi-dimensional signaling schemes that enhance worst-case and average performance. These two modulations are unitary transforms of M-QAM symbols and do not make use of additional degrees of freedom apart from the four already-used dimensions per wavelength (in-phase and quadrature channels of two polarization states). These first results extend to space-division-multiplexed optical communications impaired by mode dependent loss (MDL) that present a similar gain imbalance. Beyond signal shaping at the transmitter side, we study the performance loss of a conventional, sequential signal processing chain at the receiver side in presence of PDL, in comparison with a joint equalization-decoding scheme. The additional capacity loss due to the mismatch sequential processing, for several modulation formats and at different operating points, is evaluated and the design of optimal architecture is explained. Finally, we report an experimental validation of the two proposed signaling schemes, both on a single PDL element and on a distributed PDL channel, demonstrating the predicted enhanced robustness to PDL.

Résumé

La demande de débit en augmentation constante requiert des canaux de communication fiables et solides. Les réseaux optiques terrestres ainsi que les liens sous-marins représentent le cœur de l'infrastructure des télécommunications au niveau mondial, et transportent des centaines de canaux fréquentiels modulés à haut débit. Ces liens se composent non seulement de kilomètres de fibre mais aussi de composants optiques comme les amplificateurs à fibre dopée à l'erbium (EDFA) pour relever régulièrement la puissance des signaux atténués, ou encore des commutateurs sélectifs en longueur d'onde (WSS) qui routent le signal vers leur destination. Dans cette thèse, nous nous intéressons à une pénalité spécifique qui survient dans ces systèmes et qui réduit le débit d'information dans une propagation optique. Les composants optiques discrets présentent typiquement une anisotropie en polarisation : un déséquilibre de perte (ou gain) apparaît entre les deux polarisations d'un signal optique multiplexé en polarisation. Cet effet non unitaire appelé pertes dépendantes de la polarisation (abrégé PDL en anglais) dégrade la qualité de transmission dans les systèmes optiques actuels et futurs.

Dans le cadre des transmissions multiplexés en polarisation, nous caractérisons la perte de capacité induite par la PDL. Pour commencer, des modèles de canal sont attentivement étudiés et deux approches sont présentées : un canal avec des éléments PDL distribués avec un bruit également distribué, ou alors un canal équivalent simplifié en un seul élément. Nous analysons les limites fondamentales de communication associées à ces modèles. Nous montrons en outre que la capacité d'un canal PDL dépend en pratique de l'orientation de l'état de polarisation du signal incident. Nous passons ensuite en revue l'état de l'art de schémas de modulation atténuant l'effet de PDL puis nous proposons deux nouveaux schémas de modulation multi-dimensionnels qui augmente la performance pire et moyenne du canal. Ces deux modulations sont des transformations unitaires de symboles M-QAM qui n'utilisent pas de degrés de liberté autres que les quatre déjà utilisés pour chaque longueur d'onde (encodage en phase et quadrature de phase sur les deux états de polarisation). Nous proposons succinctement une extension de ces premiers résultats à des communications optiques plus généralement multiplexés en dimension spatiale et impactées par des pertes dépendantes de mode, qui présentent un déséquilibre de gain identique. Au-delà de la construction de modulations au niveau du transmetteur, nous étudions la perte de performance en présence de PDL d'une chaîne de traitement de signal conventionnelle et séquentielle par comparaison avec une égalisation et décodage joint. Cette perte additionnelle de capacité résultant du traitement séquentiel non optimal est évaluée pour différents formats de modulations ainsi que différents points de fonctionnement. Enfin, nous exposons une validation expérimentale des deux schémas de modulation proposés, aussi

bien sur un élément de PDL seul ou sur un canal PDL distribué, et montrons comme attendu une robustesse augmentée vis-à-vis de la PDL.

Acknowledgment

I first want to thank my supervisor Cyril for considering my application to be an intern at Nokia Bell Labs back in 2016 and that led to this thesis. I would like to thank him and Elie for their supervision of my work at Bell Labs and the patient, devoted teaching they offered me throughout my thesis. I am also grateful for the friendly and enriching conversations I had with them and with the whole transmission group I had the chance to be part of. I am indebted to Dylan, Aymeric, Patrick, Ivan, Patrice, and Ugo for their attentive help when I was carrying out experimental measurements in the lab. Thanks also to Yann for being the academic supervisor of my thesis and to Badr-Eddine for succeeding him, and thank you to my department leaders Gabriel and then Jeremie. I would like to thank as well Iryna, Charly and Fanny for the brief collaborating work we have done during these three years. Finally, I would like to thank my family and my friends for their support of my work.

Contents

Notations

Symbol	Definition
$.^*$	Conjugate
$.^T$	Transpose
$.^\dagger$	Conjugate transpose
$E\left[\cdot\right]$	Expectation
$\left\|\cdot\right\|$	Euclidean norm
$\Re\cdot$	Real part
$\Im\cdot$	Imaginary part
i	Imaginary unit
$\mathbb{C}$	Complex field
$\mathbb{R}$	Real field
$\mathbb{K}$	Arbitrary field
$\mathbb{N}$	Set of natural integers
$\mathcal{U}\left(\cdot\right)$	Uniform distribution
$\mathcal{CN}(\cdot,\cdot)$	Circularly-symmetric Gaussian noise
$\mathrm{U}\left(\cdot\right)$	Unitary group
$\mathrm{SU}\left(\cdot\right)$	Special unitary group
$\mathrm{SO}\left(\cdot\right)$	Special orthogonal group
$\mathrm{diag}\left\{\cdot\right\}$	Diagonal matrix
$\mathcal{I}(\cdot;\cdot)$	Mutual information
$\mathcal{H}$	Entropy
$d_{\min}$ or $d_{\min}^{(1)}$	Minimum Euclidean distance
$d_{\min}^{(2)}$	Second minimum Euclidean distance
I_n	Identity matrix
$\cdot_{\sim k}$	All positions except position k
$\mathrm{span}(\cdot)$	Linear span of a set
$\lfloor\cdot\rfloor$	Floor value

List of Figures

List of Tables

Chapter 1

Introduction: Modern Optical Communications in Today's Networks

1.1 A Brief History of Optical Communications

Today, an ever increasing number of connected people and devices, driven by 5G, together with new data-consuming applications require large communications capabilities [1]. The sound and reliable communications infrastructure that are optical networks represent the backbone of this data exchange need, and are expected to follow this bitrate demand trend.

Historically, communications in optics saw a drastic increase of throughput with three major evolutions depicted in Fig. **??**. The first one in the eighties was the advent of new materials to build optical fibers, the medium of optical communications. This evolution enabled a lower propagation loss in the fiber, increasing the distance of transmission or equivalently the throughput for a given reach. Then in the late nineties came the Erbium-Doped Fiber-Amplifier (EDFA) technology [2, 3], giving access to the use of Wavelength Division Multiplexing (WDM) and that permitted to gain two orders of magnitude in the achievable bitrate exploiting the frequency dimension. The last major evolution after 2005 was the ability to retrieve the phase of the electromagnetic field enabling the coherent technology [4, 5, 6] and the use of Polarization Division Multiplexing (PDM PolMux). With the existing infrastructure, it was then immediately possible to at least quadruple the amount of sent data with the implementation of multi-amplitude and phase modulations schemes, as well as

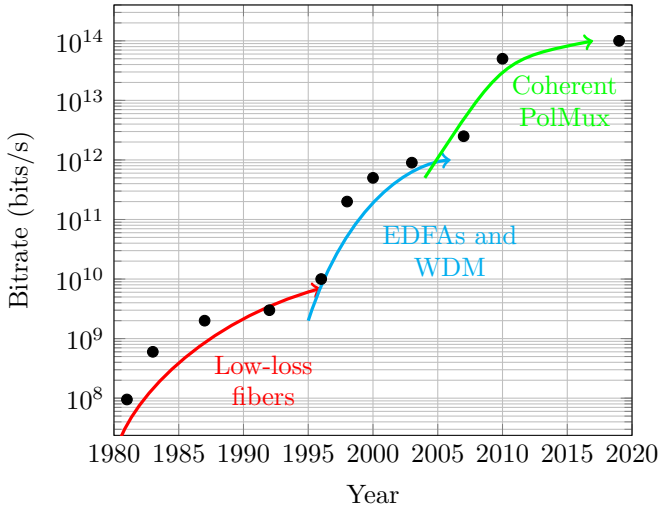


Figure 1.1 – Evolution of the throughput per fiber in long-haul optical communication systems.

the unprecedented use of a Multiple-Input Multiple-Output (MIMO) optical channel. It was then possible to apply advanced Digital Signal Processing (DSP) at the receiver to mitigate impairments and noise added by the transmission channel.

Specifically, in recent years, additional throughput gains may be hoped for by exploiting DSP and advanced modulations techniques. First, several multi-dimensional schemes have been proposed [7, 8, 9] to better encode the symbols before sending them in the transmission channel, in order to increase the link throughput. This thesis will have a strong focus on this topic and will propose new encoding techniques for a specific channel. Second, probabilistic amplitude shaping has recently been explored [10, 11] to close the gap to capacity from using equiprobable symbols. Third, a lot of efforts is made to address the nonlinearities of a transmission in an optical fiber. Modeling them make often use the Gaussian noise model [12]. Compensating them with non-linear equalizers such as the digital back propagation algorithm [13] has been proposed and a new modulation scheme called the non-linear Fourier transform has been suggested in order to provide a robustness to nonlinearities [14, 15].

With the approaching of the WDM-PDM systems fundamental limits of capacity [16], a discussed [17] candidate technique for the next major evolution is Space Division Multiplexing (SDM) and can be implemented in different technologies. Multiplexing in modes has been studied mostly for short reach applications [18, 19, 20, 21], or more recently multi-core fibers have been considered to increase the throughput in one fiber [22]. These two multiplexing techniques are sometimes compared against the simple aggregation of multiple fibers [23] in a same cable. An additional research track currently explored in order to increase the bitrate is to increase the used bandwidth by

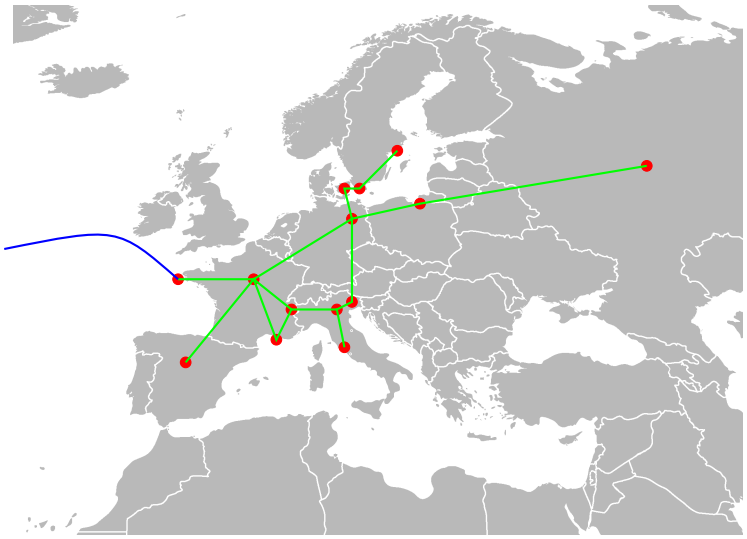


Figure 1.2 – Fictitious optical network infrastructure of Europe (green) with an illustrative submarine link (blue) for inter-continental connections.

exploiting the S-band spectrum through the use of Semiconductor Optical Amplifiers (SOAs) [24]. An orthogonal issue in optical communication systems is their consumption of energy [25] and novel works aim at reducing it.

In this thesis, we focus on particular characteristics of an optical channel called Polarization Dependent Loss (PDL). It is a non-unitary linear effect that affect PDM coherent systems. Most of the results presented after are applicable to the mentioned SDM channels, as long as they present a similar gain imbalance between two dimensions such as Mode Dependent Loss (MDL) for instance. In what follows, we furthermore always work in the linear regimes at low enough energies and leave the interactions [26, 27] between nonlinearities and PDL for future work.

1.2 Polarization Dependent Loss: a Limiting Effect in Modern Networks

In coherent optical fiber transmission networks, PDL is a linear, non-unitary impairment that is the main focus of this thesis. It is expected to have a strong impact in next-generation systems [28], and appear either in terrestrial networks or in submarine links [29, 30, 31]. For instance, to route the signals to its destination, novel Wavelength Selective Switches (WSSs) with the enabling a sharp filtering are used and can experience up to 0.3dB of PDL per port [32], a value that can also

depend on the assigned attenuation on a given port. A typical operational link can include several Reconfigurable Optical Add Drop Multiplexing (ROADM) nodes with 2 WSSs each. For example, a link constituted of 25 ROADMs introducing 0.4dB of PDL each results in an average RMS PDL exceeding $0.4 \times \sqrt{25} = 2$dB [33].

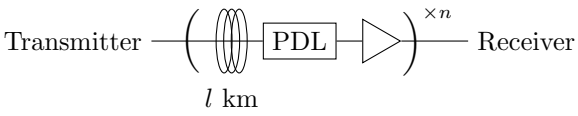


Figure 1.3 – Optical description of n concatenated spans including a long distance l of fiber, a PDL element and an EDFA represented with a triangle.

In Fig. 1.3, we represent this simplified version of a link with n concatenated spans including a portion of l kilometers of fiber, followed by a PDL element and a PDL-free EDFA plotted as a triangle. The PDL element here typically captures the WSSs effect, and the EDFAs add noise in a distributed manner along the link. EDFAs or SOAs may exhibit a small amount of PDL [33] as well, that play an impacting role in submarine links. In the modelization that follows, for simplicity we consider that the amplifiers only add noise and that the PDL they may have is in a separated PDL element. Note that optical fibers do not add PDL, but the discrete elements along the link do. This first basic illustration serves as a starting point and is later mathematically modeled and then analyzed.

Polarization Dependent Loss is a major subject of focus in optical communications because it reduces the benefit of multiplexing in polarization. Indeed, even though the throughput is theoretically doubled by the use of both polarizations of the field, the presence of PDL in the link can reduce one of the stream and reduces the benefit of having two parallel channels. More importantly, and as described in the following chapters, the performance of a channel with PDL is dependent on the input orientation into the channel. Currently, no PDL-resilient solutions is implemented in commercial products, the link designer just accounts for the PDL impairments by the addition of SNR margins to absorb the PDL-induced penalties.

In order to mitigate PDL, several methods have been proposed. One of them is the use of Space-time codes to encode the sent signal. Space-time code are algebraic codes that spread the information over consecutive timeslots in order to protect them, originally against fading environments, but also in our context against a link with polarization gain imbalance. On a PDL channel, the Golden code, the Silver code [34] and a Threaded Algebraic Space Time (TAST) code [9] have been tested. Among them, the Silver code invented originally for wireless applications by Tirkkonen and Hottinen [35] is a 2-timeslot modulation scheme. It has been proposed [34] in order to mitigate PDL and then showed to be the best performing modulation known for PDL mitigation [36]. In [36], the effective PDL-compensation performance of the Silver code is explained by the orthogonality

of most of the encoded symbols over the PDL channel. The alternative number theory 2-timeslot code [9] is a simpler modulation scheme (belonging to codes introduced in [37]) that rotates the second timeslot by a relevant angle and that shows similar performance. These 2-timeslots modulations though imply a higher decoding and equalization complexity making it unpractical in real transmission systems because of the introduced correlations between the timeslots. This thesis notably proposes to define low-complexity PDL-resilient modulation schemes that only use the existing degrees of freedom that are the in-phase and quadrature components of the two polarization states of a coherent polarization-multiplexed signal. An alternative approach uses several rotated sub-carriers [38] to achieved an average performance over a link impaired by PDL. Furthermore, by opposition to [39], we assume that Channel State Information (CSI) at the transmitter is not doable otherwise the water filling technique is directly indicated [40] for increasing the performance over a MIMO-channel with gain imbalance.

1.3 Thesis Outline

In chapter 2 we first cautiously model the effect of PDL. From a modeling on the elementary scale, we derive the complete distributed channel model and characterize its statistics. We furthermore develop the commonly used lumped model that serves later as a tool to build upon.

Then, the following chapter 3 discusses the limits of communications of the PDL channel. We study the impact on the channel capacity of its gain imbalance, and show that the information rate of discrete modulation schemes are sensitive to the incident orientation of the input. We then use information theory tools to demonstrate the influence of the receiver signal processing chain on the overall performance of the link. Namely, we show that a standard, sequential processing implies an additional loss in achievable rate in the context of transmissions over a PDL channel compared to a joint equalization-decoding process. We propose simulation results with several bit or symbol decoders and compare the performance of all the schemes.

The Dual-Polarization Quadrature Amplitude Modulation (DP-QAM) input schemes being very sensitive to the incident SOP into the channel, in chapter 4 we propose new modulation schemes in order to address this dependency and consequently increase the worst case performance. A first illustration is made on a simplified set of parameters of the PDL channel, leading to the Spatially Balanced (SB) signaling. A more general, complete solution is then provided and called the New Spatially Balanced (NSB) signaling, it represents the unitary transform that can increase the most the worst rate over a PDL channel, while keeping the same degrees of freedom as the ones of standard modulations. We analyze the SB and NSB signalings, in terms of rates and Euclidean distance. We analytically demonstrate why the NSB is the best candidate for increasing the worst

case with an analytical proof.

The closing chapter 5 is dedicated to a validation of the two built modulation schemes. We first study in simulation their performance on a distributed PDL channel for a given rate outage condition. We then perform experimental measurements in our lab of the SB signaling in a lumped PDL channel and show the obtained gains from using standard QAM. Finally, we experimentally test the NSB signaling over a distributed channel with inserted PDL mimicking a transoceanic link and assess its performance compared to DP-QAM. As a conclusion, we give some prospective studies to scale the designed schemes to other optical MIMO systems.

Chapter 2

Optical Channel Modeling in Presence of Polarization Gain Imbalance

In this opening chapter, after recalling the degrees of freedom of the optical field, we first model the optical channel that contains elements presenting a polarization gain imbalance. We start by developing the model on the elementary scale, we then concatenate several optical elements and discuss the resulting link statistics.

2.1 Exploiting the Degrees of Freedom of the Optical Field

Optical communication systems modulate lightwaves in different spectral windows. O-band window where chromatic dispersion is minimal for standard single mode fibers is preferred for short-reach applications. Direct detection and light digital signal processing are used in this case. Wavelengths in C- and L-band windows are modulated and sent over long-haul networks. Given the high capacity requirements in these networks, coherent detection and polarization multiplexing were introduced since 2000. The information that has to be exchanged, typically in the form of a bit sequence, is encoded in transmittable optical signals. This signal propagates in an electromagnetic field trapping medium that is the optical fiber. Depending on the application, different technologies exist to transmit the required bitrate on the desired distance. The optical field offers the possibility to exploit several degrees of freedom:

1. the amplitude and time, which is historically the natural degrees of freedom to be used in the first on-off keying communications systems,
2. the wavelength, made possible notably by the development of the EDFA technology after 1995,
3. the phase, mainly exploited since the coherent technology is implemented in the mid 2000s and that was the missing element to send the commonly used QAM symbols,
4. the polarization of the optical field, making possible to send two independent streams with the same signal and hence doubling the amount of information with the same legacy infrastructure,
5. and other space dimensions, such as different modes of the optical field in Few/Multi-Mode Fibers (FMF or MMF) or different cores in a single fiber, or even parallel fibers with common amplification stages.

This work principally focus on polarization-multiplexed systems (that exploits the amplitude/time and phase as well). Such a system can be viewed in communications theory as a 2×2-Multiple-Input Multiple-Output (MIMO) system here with two polarization inputs and two polarization outputs. Fig. 2.1 represents this simple schematic view of a 2×2 MIMO communication system, where the polarizations can evolve while propagating in the optical field. Typically, we model the output of this system with a transfer matrix $H \in \mathbb{C}^{2\times2}$ where the entry h_{ij} captures how the input polarization j sent by the transmitter evolves and is received in the output polarization i at the receiver side. This matrix H, that can also be called a Jones matrix, is developed in the following section. The inherent presence of noise in a transmission link will be added when properly modeling the channel transfer function.

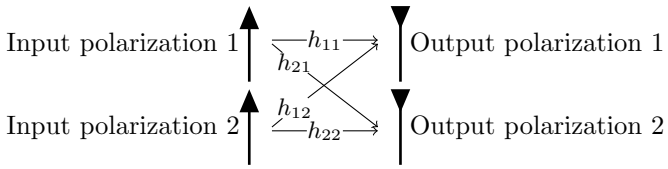


Figure 2.1 – Communication theory simplified modeling of a 2×2-MIMO channel.

In the rest of this work, we will always consider that we work in a single channel hence omitting the wavelength dimension. Indeed, it is a good assumption to consider that PDL is flat on one WDM channel, therefore we will study PDL in only one channel. PDL may vary between WDM channels but they all have the same statistics. Also, as previously mentioned, some of the results developed thereafter can be extended to space division multiplexed systems, as long as they show properties similar to the ones presented over polarizations.

2.2 Elements with Polarization Dependent Loss: a Non-Unitary Effect

Discrete components along an optical link may behave differently depending on the incident polarization. On a discrete component scale, if the gain associated with the first polarization is g_1 and the gain with the second one d_2, in general $d_1 \neq d_2$. Assuming that the two gains are along two orthogonal axes and an incident polarization-multiplexed signal X has its polarizations aligned to those axes, the output signal is $DX = \text{diag}\{d_1; d_2\} X$. This is called a Polarization Dependent Loss (PDL). PDL. As a short hand we say that the signal is aligned to the PDL element, in other words we identify the PDL elements with its two gains axis. In general though, because the signal propagates in a fiber, it meets the PDL element with a random State-of-Polarization (SOP) with respect to a given frame of reference. Similarly, the output signal may be randomly orientated with respect to this reference. When a light signal crosses an optical element, the polarization of the emerging light is computed by taking the product of the Jones matrix of the optical element and the Jones vector of the incident light. Then, a valid modeling of a PDL element inserted in an optical link is

$$H = UDV \tag{2.1}$$

where U and V are two unitary matrices taken uniformly at random that define the orientation of the input and output signal, respectively, into and from the PDL element. The distribution of these matrices are discussed in section 2.5. For the expression of the gains d_i, two different conventions may be employed to take into account the gain imbalance.

1. Applying the loss over one polarization axis, typical convention for passive components, yielding $D_\epsilon = \text{diag}\{1; \sqrt{\epsilon}\}$ with $\epsilon \in [0, 1]$ and leading to a loss of the total energy.

2. For active components, a total energy conservation can be assumed, yielding the matrix $D_\gamma = \text{diag}\left\{\sqrt{1+\gamma}; \sqrt{1-\gamma}\right\}$ with $\gamma \in [0, 1]$.

In both cases, we usually refer to the difference of gains in dB.

Definition 1 *The PDL value of an element with gains d_1 and d_2 (assuming $d_1 \geq d_2$), denoted λ or Λ in dB, is defined as the ratio between the two element squared gains*

$$\lambda = \frac{d_1^2}{d_2^2} \text{ or in dB } \Lambda = 10 \log_{10}(\lambda) \tag{2.2}$$

Note that when $\epsilon = 1$ or $\gamma = 0$, the PDL value is $\Lambda = 0$ dB. In this case, the component's two polarization gains are identical and there is no PDL. Eventually, the model 1 has a loss of energy $\text{trace}(HH^\dagger) = \text{trace}(D^2) = 1 + \epsilon$ compared to the model 2 that is energy conserving with $\text{trace}(HH^\dagger) = 2$.

2.3 Optical Link Modeling with Concatenated PDL Elements

Along an optical link, a polarization-multiplexed signal is likely to encounter several cascaded PDL elements, typically EDFAs or WSSs, which impairments accumulate. Due to the random orientation of the components, the signal experiences an tempered effect that is different from the sum of all the PDL elements. We call the pattern of: a portion of fiber of a given distance l km; a PDL element; and an EDFA adding noise a *span*. It is represented in Fig. 1.3 delimited by brackets.

We recall the following definition before modeling the distributed PDL channel.

Definition 2 *A random variable $Z \in \mathbb{C}$ is said to follow a standard complex Gaussian random variable if $\Re Z$ and $\Im Z$ are independent and both follow a Gaussian process with variance $1/2$ and mean 0. Similarly, we say that a random vector $Z \in \mathbb{C}^{n_r \times 1}$ follows a circularly-symmetric complex standard Gaussian process with covariance matrix $K = E\left[ZZ^\dagger\right]$ and zero-mean if the Z_i components Z are independent and follow themselves a standard complex Gaussian process. We denote it $Z \sim \mathcal{CN}(0, K)$.*

A block diagram of this physical model is depicted in Fig. 2.2. We consider n spans where each the fiber and the PDL element are embodied in the matrix H_i and assume the EDFAs produce Amplified Spontaneous Noise (ASE) modeled by an additive white Gaussian complex process, and simplified as $Z_i \sim \mathcal{CN}(0, I_2)$ with unit variance per polarization. The fiber loss and insertion loss of the PDL element are assumed to be perfectly compensated for by the gain of the amplifier, the matrices $H_i = U_i D_i V_i$ are hence defined according to the PDL model 2. Moreover, we intentionally discard two dispersive effects that are the Chromatic Dispersion (CD) and the Polarization Mode Dispersion (PMD). CD has the same impact on both polarization tributaries and can be perfectly compensated for in the electronic domain, whereas PMD is not modeled in this work as most novel transmission fibers exhibit very low PMD [41] (below $0.05\text{ps}/\sqrt{\text{km}}$) which can be neglected even for long transmissions and high symbol rates. The influence of moderate to high PMD values on the evolution of PDL is left for a future work.

Under these assumptions, looking from right to left in Fig. 2.2, the resulting output Y can be

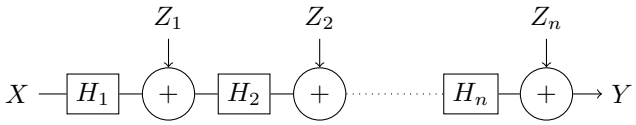


Figure 2.2 – Block representation of a concatenated chain of n PDL elements. Each span i is modeled as a matrix H_i (fiber, PDL element) and added noise Z_i (due to the amplifier).

written as

$$Y = Z_n + H_n(Z_{n-1} + H_{n-1}(\cdots(Z_2 + H_2(Z_1 + H_1 X))\cdots) \tag{2.3}$$

$$= \left[\prod_{i=n}^{1} H_i\right] X + \left[\sum_{i=1}^{n} \left(\Pi_{j=n+1}^{i+1} H_j\right) Z_i\right] \tag{2.4}$$

taking the convention $H_{n+1} = I_2$, just to be able to write the equation in a condensed form.

Note that Eq. 2.4 is of the form

$$Y = HX + Z \tag{2.5}$$

where $H = \prod_{j=n}^{1} H_j$ is the product of all the PDL element matrices and the resulting noise Z is still a zero-mean Gaussian complex noise but with covariance matrix $E\left[ZZ^\dagger\right] = \sum_{i=1}^{n} P_i P_i^\dagger$, P_i being the products $\Pi_{j=n+1}^{i+1} H_j$. Hence, a link with distributed PDL elements, called a *PDL channel*, behaves as a single-element PDL channel with an equivalent Λ aggregating the ones of the elementary PDL-impaired components, along with a correlated noise Z.

Eq. 2.4 suggests that, even though several elementary transfer matrices constitute a distributed PDL channel, it can be summarized to the study of a simple channel that is presented in the following section.

2.4 The Lumped PDL Channel: a Simplified Channel as a Tool

The previous section showed that even though several elementary transfer matrix constitute a distributed PDL channel, it can be summarized to the study of a few parameters from the point of view of linear polarization imbalance and crosstalk effects..

2.4.1 Derivation of a Channel with Three Parameters

The PDL channel in Eq. 2.4, can be recast in the product of basic matrices previously mentioned. We take the classic 2×2-MIMO modeling Eq. 2.5 where again Y is the output signal, X is the input signal, H is the channel transfer matrix, and $Z \sim \mathcal{CN}(0, Q)$ a zero-mean complex Gaussian noise with covariance matrix Q.

If the positive definite matrix $Q \neq I_2$, by performing a Cholesky decomposition, there exists $\Sigma \in \mathbb{C}^{2\times 2}$ such that $\Sigma\Sigma^\dagger = Q$ and computing $\Sigma^{-1}Y = \Sigma^{-1}HX + \Sigma^{-1}Z$ leads to a noise $Z \sim \mathcal{CN}(0, I_2)$. For simplicity, in what follows we take a unit covariance matrix $Q = I_2$ and leave the general correlation case for a future study.

In this case, note that $U \in \mathrm{U}(2)$ multiplying the noise does not change its properties: $E\left[(UZ)(UZ)^\dagger\right] = UE\left[ZZ^\dagger\right]U^\dagger = I$, and of course $E[UZ] = E[Z]$. As for the channel matrix H, we substitute it by its Singular Value Decomposition (SVD) as done for the elementary PDL matrix in Eq. 2.1, left-multiply by U^{-1} and redefine $Y := U^{-1}Y$ and $Z := U^{-1}Z$ with unchanged properties such that we can write $Y = DVX + Z$. Then according to [42, pp. 71-72] and appendix A, there exists $\delta, \alpha, \beta \in \mathbb{R}$ such that $Y = DB_\delta R_\alpha B_\beta X + Z$ where

$$R_\alpha = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \text{ and } B_\beta = \begin{pmatrix} e^{i\beta} & 0 \\ 0 & e^{-i\beta} \end{pmatrix} \tag{2.6}$$

and B_δ is defined as the latter. The real matrix R_α can physically be interpreted as a rotation of the polarizations, and the B_β matrix is called a *birefringence* or a *retardance* matrix. Commuting the diagonal matrices D and B_δ, $Y = B_\delta DR_\alpha B_\beta X + Z$, we left-multiply by B_δ^{-1} and redefine again $Y := B_\delta^{-1}Y$ and $Z := B_\delta^{-1}Z$ to end up with the equivalent form

$$Y = DR_\alpha B_\beta X + Z \tag{2.7}$$

The total gain values in the matrix $D = \mathrm{diag}\{g_1; g_2\}$ depend on the evolution of the polarizations between the individual PDL elements. It is worthy to note that the trace of $HH^\dagger$ is not constant for random drawings of concatenated fiber sections in the defined model that considers a constant-gain amplification scheme at EDFAs. We shall see notably in section 5.1.1 that this hypothesis has an impact on the measured penalties, and we will compare it to a constant-output-power amplification scheme. By factorizing D by the factor $k = \sqrt{(d_1^2 + d_2^2)/2}$, we write it in the form $D_\gamma = \mathrm{diag}\left\{\sqrt{1+\gamma}; \sqrt{1-\gamma}\right\}$ mentioned in the elementary PDL modeling and with $\gamma = (\max(d_1^2, d_2^2) - \min(d_1^2, d_2^2))/(d_1^2 + d_2^2)$. We omit this global loss factor that may be compensated by an amplifier as an additional power loss.

To conclude, we often work with the following model that is called the *lumped PDL model*.

We suppose the noise $Z \sim \mathcal{CN}(0, I_2)$ and take a transfer matrix H that has one gain parameter $\gamma \in [0, 1]$ and two angle parameters α and β in $(-\pi, \pi]$ defined as

$$H = D_\gamma R_\alpha B_\beta \tag{2.8}$$

Using the lumped PDL channel, one could possibly forget about the underlying channel that originated from concatenating PDL elements. In other words, the channel can now be seen as a single, large, PDL element. Section 2.5 discusses the distribution of the three parameters in Eq. 2.8. Next section explores an approximation of the PDL channel that is even more simplified.

2.4.2 The Simple PDL Channel Modeling with Only One Angle Parameter

A regularly used approximation [43, 44, 9] of the PDL channel is to consider $B_\beta = I_2$ to remove the dependency in β. This leaves the PDL channel with only one angle parameter

$$H = D_\gamma R_\alpha \tag{2.9}$$

that we call the *simple* PDL channel. It will serve as a basis to build PDL-resilient modulations in section 4.1. This simplification is often used in optics literature [44, 9] and can be justified as the variation in β is negligible in front of the one in α as discussed in chapter 3. Nevertheless, when applicable, it should be remembered that this is not the complete model of the lumped channel, in particular the negligible variation in β may not hold for a high value of PDL or when changing the input modulation.

2.5 Statistics of a Link with Concatenated PDL Elements

2.5.1 Angles Distribution

The unitary matrices U and V in Eq. 2.1 are mentioned to be random. Indeed, since no direction is favored in the channel, the two unitary matrices are uniformly distributed in $\mathrm{U}(2)$ with respect to the Haar measure [45].

The four angles from the unitary matrix decomposition recalled in appendix A lead to a Haar measure volume element $\mathrm{d}V = \sin^2(\alpha)\mathrm{d}\phi\mathrm{d}\delta\mathrm{d}\alpha\mathrm{d}\beta$ [46]. Therefore, ϕ, δ, β follow $\mathcal{U}((-\pi, \pi])$, the uniform distribution on the range $(-\pi, \pi]$. Furthermore, in order to respect uniformly distributed matrices in the unitary ensemble, the remaining angle $\alpha \in [0, \pi/2]$ can be taken defining $\alpha =$

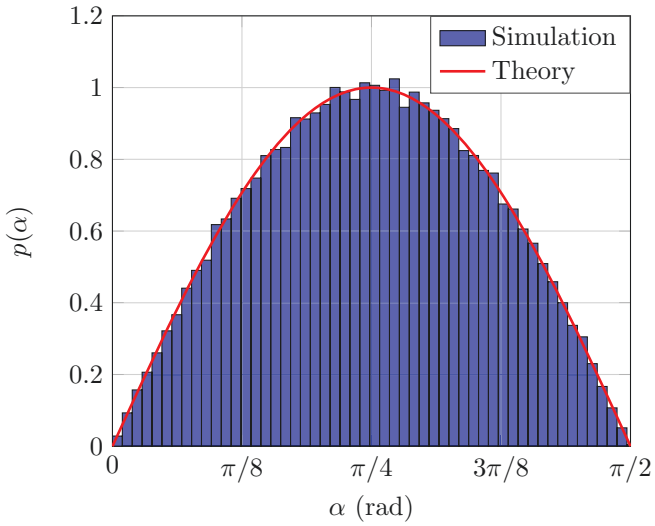


Figure 2.3 – Distribution of the equivalent channel angle α for simulated PDL channel (blue) with $n = 25$ spans including randomly oriented PDL elements of $\Lambda_i = 0.4$dB each, compared to the mentioned circle-shaped distribution (red) in Eq. 2.10.

$\arcsin\sqrt{\xi}$ where $\xi \sim \mathcal{U}([0,1])$. Equivalently, the probability density function of α, plotted in Fig. 2.3, is

$$p(\alpha) = \sin(2\alpha) \tag{2.10}$$

In Fig. 2.3 and Fig. 2.4, the distribution of respectively the angle α and β restricted in $[-\pi/2, \pi/2]$ are extracted from 10^6 simulated PDL channels with $n = 25$ spans and shows the expected probability laws. Identically, on the elementary level, the α_i corresponding to each unitary matrix element U_i taken uniformly at random should be drawn following Eq. 2.10. Nevertheless, if instead all α_i follow the uniform distribution, the impact on the resulting channel is negligible due to the concatenation.

2.5.2 Statistics of Global Polarization Imbalance

For a sufficiently large number of concatenated elements with identical PDL value, it is shown in [31] that the squared singular values ratio in decibel Λ of the equivalent channel H follows a Maxwellian distribution of which we recall the definition.

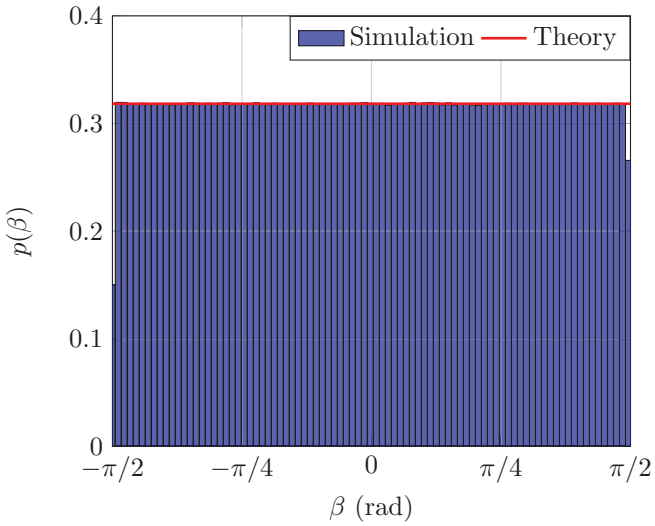


Figure 2.4 – Distribution of the equivalent channel angle β for simulated PDL channel with $n = 25$ spans including randomly oriented PDL elements of $\Lambda_i = 0.4$dB each.

Definition 3 *For $a > 0$, a random variable X is said to follow a Maxwellian distribution if*

$$X \sim \sqrt{\frac{2}{\pi}} \frac{x^2}{a^3} \exp\left(-\frac{x^2}{2a^2}\right) \tag{2.11}$$

where the parameter a is linked to the expected value of X by $a = \frac{E[X]}{2\sqrt{2/\pi}}$.

The resulting distribution of $n = 25$ spans of randomly oriented PDL elements of $\Lambda_i = 0.4$dB each is represented in Fig. 2.5. It shows a good fit between the actual singular values ratio distribution and the theoretical distribution computed with Eq. 2.11. We used an sufficiently large number n of elements so that the Maxwellian approximation is correct. Between each element, a fiber junction is emulated by a unitary matrix U_i taken uniformly at random. This scrambling of polarization enables the gain imbalance at each element to not add together: it rather follows a root mean square rule, the global PDL of the channel being roughly equal to $0.92\Lambda_i\sqrt{n}$ [31].

Now, we look into the impact of the correlated noise Z. An equivalent channel model can be derived from Eq. 2.5 where we whiten the noise with Σ^{-1} the inverse of the matrix obtained from the Cholesky decomposition of $K_{ZZ} = \Sigma Sigma^{\dagger}$. The distribution of the gain imbalance after noise whitening is represented in brown on Fig. 2.5. The gain imbalance still follows a Maxwellian distribution but with a lower mean value. Indeed, while propagating along with the signal, the noise is also impacted by the PDL elements.

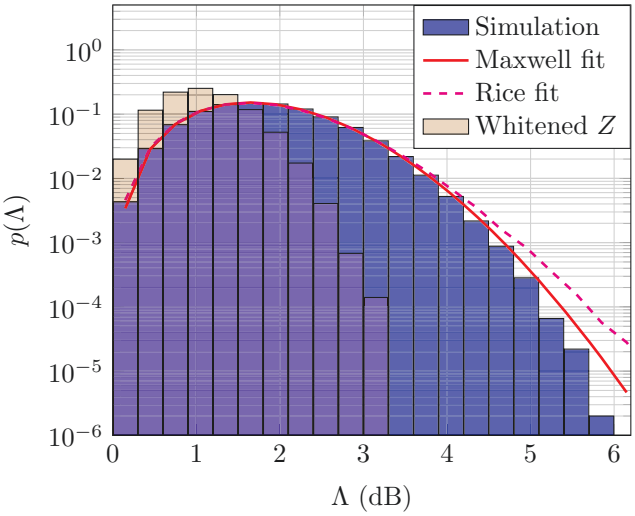


Figure 2.5 – Distribution of the equivalent channel squared singular value ratio Λ (blue) for a simulated PDL channel with $n = 25$ spans including randomly oriented PDL elements of $\Lambda_i = 0.4$dB each. The theoretical Maxwell distribution (red) and a proposed fit with singular value ratios from an emulated Rice channel (magenta dashed line) are represented. The fits are constrained with the same average as the one of the empirically obtained from the channel realizations.

σ_i^2	0	10^{-2}	10^{-1}
$E[\Lambda]$	1.852	1.853	1.907

Table 2.1 – Average PDL value $E[\Lambda]$ of the equivalent channel for different elementary PDL dispersions varying around a mean value of 0.4dB.

Also, in reality an uncertainty can impact the PDL introduced by optical components in a given link. Hence, we assume that our elementary PDL is given by a (not truncated) Gaussian distribution with mean 0.4dB and a variance σ_i^2 between 0 (fixed PDL element case) and 10^{-1} to check the validity of our model. We reported in table 2.1 the average of the link PDL Λ (with unwhitened noise) for 25-span channels in which elementary PDL are distributed as previously mentioned. We observe that for tight dispersion around the value of the fixed PDL element case, the average link PDL remains similar. But it becomes larger when PDL elements have a value that varies more. This has to be taken into account when designing a system subject to PDL effects. When discussing the limits of communications of the PDL channel, for simplicity we will keep a fixed PDL value for the concatenated elements and assume that the noise follows a complex Gaussian distribution with unit covariance.

2.5.3 An Alternative Method to Emulate Distributed PDL

In complement to the standard model distributions mentioned in section 2.5.2, we additionally present a simple model that has two purposes: it captures the link implementation impairments in an intuitive manner while presenting a reduced emulation complexity. The alternative method to model the gain imbalance distribution presented in this section is inspired from wireless communications channel models.

In wireless communications, the classical Rayleigh channel G with entries g_{ij} each following a $\mathcal{CN}(0,1)$ distribution differs from the optical channel. Indeed, one polarization totally dropping in a polarization-multiplexed propagation in a fiber is unrealistic but total fading is possible in a multi-antenna Rayleigh channel transmission. Therefore, this modeling does not suit an optical channel transmission. By adding the equivalent of a line-of-sight component to it, we then illustrate the direct path of the source in a guided medium that makes it more relevant for an optical communication. The corresponding channel transfer matrix is

$$H = aG + bI_2 \tag{2.12}$$

where G is the Rayleigh fading channel transfer matrix, a and b are in $[0,1]$. For $b = 0$ this is the Rayleigh fading channel again. For $a = 0$ the channel is deterministic. We omit these two extreme cases. The modeled channel is then equivalent to a Rice fading channel.

For $b \gg a \neq 0$, the dB squared ratio of the singular values Λ of H has a *Maxwellian-like* shape plotted in Fig. 2.5 (magenta dashed line). The figure shows a good accordance down to $p(\Lambda) = 10^{-3}$ between the proposed modeling and the Maxwellian fit curve with identical average value. Note that in section 2.5.2, the Maxwell distribution appeared for *concatenated* PDL elements in sufficiently large number $n > 15$. Here, the distribution displayed in magenta dashed line is the one of a *single* element with transfer matrix H from Eq. 2.12. Instead of precisely taking into account the model of each PDL element and concatenating them, the resulting channel may be replaced by a transfer matrix following the Rice distribution and shows a good approximation of the real resulting PDL values. This modeling offers a reduced simulation complexity and gives almost the same statistics of a PDL channel. Moreover, the unitary matrices appearing in the SVD of H are uniformly distributed in the unitary group as it is the case with the exact PDL modeling. The average value of the dB singular value squared ratio Λ of a Rice channel is given by

$$E[\Lambda] \approx 6\ln(10) \times a/b \tag{2.13}$$

This is the formula used to fit the simulated PDL distribution in Fig. 2.5 from the mean of the drawn values. Exact derivation of the singular value distribution is possible using the complex expression

of the joint distribution in [47]. The modeling presented here with a single element following a Rice distribution is valid as long as it is meant to model a PDL channel with homogeneous PDL elements.

Chapter 3

Limits of Communications of a Channel with Polarization Dependent Loss

In this chapter, we use the models derived in the previous chapter to characterize the limits of communications over a channel with polarization gain imbalance. First, the raw capacity of the PDL channel is computed as the maximal achievable information rate obtained using the mutual information between a given modulation input and the considered channel output. Second, receiver architectural mismatches are measured using the same tools associated with the relevant input-output pairs. A sequential processing chain induces rate losses compared to an optimal processing. After presenting the rate associated with this imperfect decoding, we then compare in simulation the BER after an optimal or a sequential processing at the receiver.

3.1 Fundamental System Capacity of a Link with PDL

The case of continuous inputs with power constraint permits to define the raw PDL from an information theoretical point of view. The capacity of the PDL channel modeled in Eq. 2.8 as 2×2-MIMO channel with fixed gain imbalance and AWGN noise is defined as the maximum of the mutual information [48] between its input and output over all input distributions

$$C = \max_{p(X)} \mathcal{I}(X;Y) \tag{3.1}$$

where the mutual information $\mathcal{I}(X;Y)$ depends on the entropy of the source $\mathcal{H}(X) = -\sum_x p(x)\log_2 p(x)$ and the one of the channel of transmission

$$\mathcal{I}(X;Y) = \mathcal{H}(X) - \mathcal{H}(X|Y) \tag{3.2}$$

with

$$\mathcal{H}(X|Y) = -E\left[\mathcal{H}(X|Y=y)\right] = \sum_y p(y)\mathcal{H}(X|Y=y) = -\sum_y p(y)\sum_x p(x|y)\log_2 p(x|y) \tag{3.3}$$

The capacity can then be computed by entropy maximization [49, 40] for the PDL channel modeled by Eq. 2.8. This maximum is obtained for a Gaussian input X with covariance matrix $Q = \frac{\rho}{2}I_2$ where $\rho = P_X/P_Z$ is the Signal-to-Noise Ratio (SNR). The PDL capacity is then

$$C = \log_2 \det\left(I_2 + \frac{\rho}{2}HH^\dagger\right) \tag{3.4}$$

$$= 2\log_2\left(1+\frac{\rho}{2}\right) + \log_2\left(1-\gamma^2\frac{\rho^2}{(2+\rho)^2}\right) \tag{3.5}$$

The first line is the generic capacity formula for a given MIMO channel with transfer matrix H. The second line comes from the fact that $HH^\dagger$ simplifies to $D_\gamma^2 = \mathrm{diag}\{1+\gamma, 1-\gamma\}$. The capacity can finally be written as the sum of two terms. The first term is the standard capacity of a 2×2 MIMO Gaussian channel (raw diversity) and the second, negative, term defines the PDL-induced rate loss. This quantity, which is independent of the channel angles, is depicted in Fig. 3.1 as a function of the SNR for various values of gain imbalance Λ. We see, for instance, that a PDL of $\Lambda = 6$dB translates into a loss of 0.6 bit/s/Hz in capacity at $\rho = 10$dB, which may imply a transmission outage in practice if not taken into account in the system design. This formula is valid when optimal decoding is performed at the receiver, notably by performing a joint equalization and decoding of the symbols. Otherwise, alternative metrics discussed in section 3.3.1 should be considered.

In the following sections, we often focus on a channel with 6dB of PDL, it is an arbitrary, worst case value, commonly used in systems design that corresponds to a 10^{-5} outage probability with a link average PDL of 2dB. For example, a reasonable link with 8 nodes, containing 2 WSSs each of 0.5dB of PDL, would result [31] in a link with an expected PDL $E[\Lambda] \approx 0.5\sqrt{2\times 8}$.

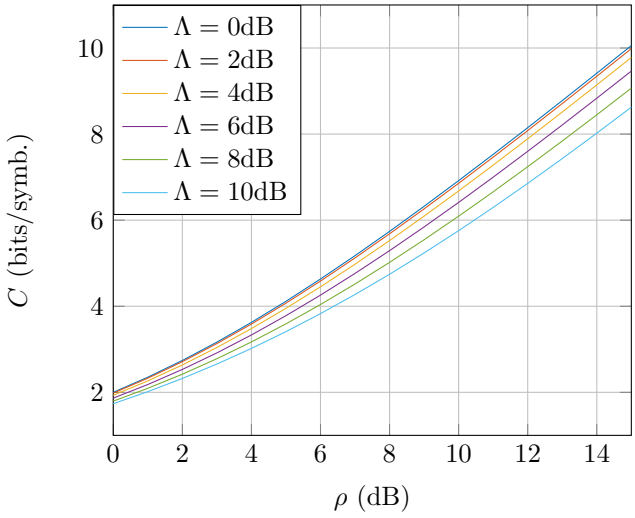


Figure 3.1 – Capacity as a function of the SNR for polarization gain imbalances Λ from 0dB (no PDL) to 10dB. A Gaussian input is used for entropy maximization.

3.2 Discrete Inputs and Polarization Rotation Dependence as Practical Limits

PDL is characterized in the previous section as a raw loss in term of information rate that even a continuous modulation with circular symmetry can not circumvent. The case of practical modulation alphabets such as QAM presents an additional directional PDL loss that is a function of the actual polarization rotation.

We now study the capacity of a PDL channel when discrete modulation schemes are employed such as DP-QAM constellations. Because the input constellation is fixed, the definition of the capacity in Eq. 3.1 reduces to the calculation of the mutual information between X and Y. Notice that the mutual information definition uses $\mathcal{H}(X|Y)$ that requires the knowledge of $p(X|Y)$ but usually the probability density function $p(Y|X)$ is more tractable. In what follows, we present how to compute information rates using Monte-Carlo simulations. Namely, for n_s samples $Y[i]$ and $X[i]$, both in $CC^{2\times 1}$ and denoted y_i and x_i in this section only for clarity, consider $h(x_i|y_i) = -\log_2 p(x_i|y_i)$; by the law of large numbers we have

$$\frac{1}{n_s}\sum_{i=1}^{n_s} h(x_i|y_i) \to -E\left[\log_2 p(X|Y)\right] = \mathcal{H}(X|Y) \text{ for } n_s \to \infty \tag{3.6}$$

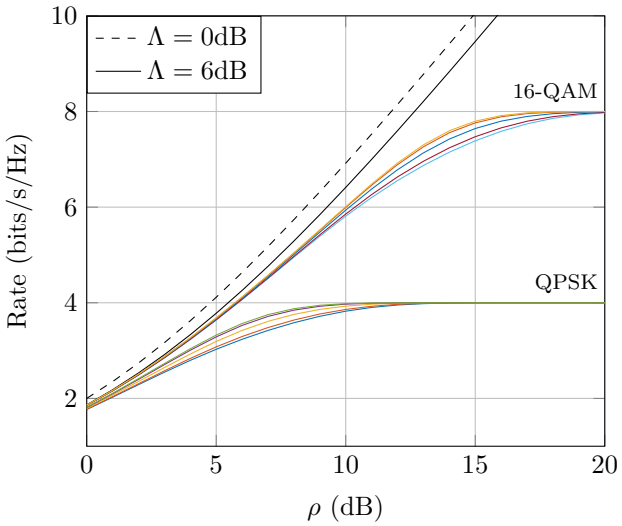


Figure 3.2 – DP-M-QAM rate function of the SNR for $\Lambda = 6$dB (full lines) and various SOPs for $M = 4$ and 16. The 2×2-MIMO capacity with no PDL (dashed line) is represented as a reference.

Denoting for clarity $x \in CC^{2\times 1}$ a symbol in the considered DP-QAM constellation, we then have

$$h(x_i|y_i) = -\log_2 p(x_i|y_i) \quad = -\log_2 \frac{p(x_i|y_i)}{\sum_x p(x|\ y_i)} = -\log_2 \frac{\frac{p(x_i)}{p(y_i)} p(y_i|x_i)}{\sum_x \frac{p(x)}{p(y_i)} p(y_i\ |x)} \quad = -\log_2 \frac{p(y_i|x_i)}{\sum_x p(y_i|x)} \tag{3.7}$$

by definition of $h(x_i|y_i)$, unit mass of a probability density function, Bayes rule and using the fact that $p(x) = p(x_i)$ is a constant when no probabilistic shaping is considered. The formula 3.7 is the one used to compute mutual information when discrete constellations are used, where the transition probability is given by

$$p(y|x,h) \propto \exp\left(-\frac{\|y - hx\|^2}{2\sigma^2}\right)$$

where σ^2 represents the total variance of the noise.

In what follows, we will talk about information rates to design the capacity associated with those constellations.

The information rates of the DP-M-QAM schemes for $M = 4$ and 16 are represented in Fig. 3.2 as a function of the SNR ρ and for various SOPs and a fixed PDL of $\Lambda = 6$dB. As a reference, we plotted the Shannon capacity previously discussed in section 3.1 for no PDL ($\Lambda = 0$dB) and

$\Lambda = 6$dB of PDL. Note the additional rate variation that strongly depends on the incident SOP. Typically, for any β and for common PDL values, the lower curve for each format is met for $\alpha = 0$ modulo $\pi/2$ and the best rate is reached for $\alpha = \pi/4$. There is in particular a great dependence in the incident angle α and a lower dependence in the second angle β. This is depicted in details in Fig. 3.3 with the information rate of DP-QPSK represented as a function of the two angle parameters, and still for $\Lambda = 6$dB, at an SNR of 8dB. A similar profile is found for any modulation M-QAM with arbitrary order of modulation M: the minimum rate is met for α multiple of $\pi/2$ and any β. In Fig. 3.3, the highest rates are achieved for $\alpha = \pi/4$ and more precisely the maximum rate over all SOPs is for $(\alpha, \beta) = (\pi/4, \pi/8)$ in $[0, \pi/4] \times [0, \pi/4]$.

Let us give here some insight into this dependency in (α, β). The SOP dependency can easily be more easily visualized when referring to the Euclidean distance associated with the modulation. Notice indeed that the information rate is dominated by the terms corresponding to the smallest Euclidean distances [50] of a given input alphabet. In our case, the alphabet of interest is the scaled modulation. In Fig. 3.4 we illustrate the minimum Euclidean distances of the DP-QPSK codebook denoted $\mathcal{X}$ after propagation in the PDL channel defined as

$$d_{\min} = \min_{\substack{(X_j, X_k) \in \mathcal{X}^2 \\ X_j \neq X_k}} \|H(X_j - X_k)\| \tag{3.8}$$

for a given SOP, and still for $\Lambda = 6$dB. While the two metrics are not comparable, observe that the variations are strongly correlated with the ones of the rate in Fig. 3.3. Maxima and minima SOPs in the rates correspond to respectively maxima and minima as well in the minimum distance profile. A more detailed analysis of the Euclidean distances in our context of gain imbalance will be discussed in chapter 4.

The average information rate of a PDL channel over all channel orientations is the mean of the presented values in Fig. 3.3 weighted by the probability laws of the parameters discussed in section 2.5. A lot of efforts is currently made in order to reduce the aforementioned angles dependency. In particular there is a need to increase the worst capacity here obtained for α angles multiple of $\pi/2$. Indeed, optical systems are designed to be robust to a worst case scenario. In other words, the transmission has to be guaranteed even in the parameters' worst configuration.

In chapter 4, we design new modulations that are built from two methods that complement one another. A first approach will be to use the information rates metric in order to provide criteria for increasing the worst rate over all SOPs. As a second approach, we will use the Euclidean distance metrics to explain some results coming from the rate study. The distance metric is more tractable because no close form of the information rates is available. This second metrics also appears to be a good approximation in high SNR regimes of the rate and enables later to understand behaviors

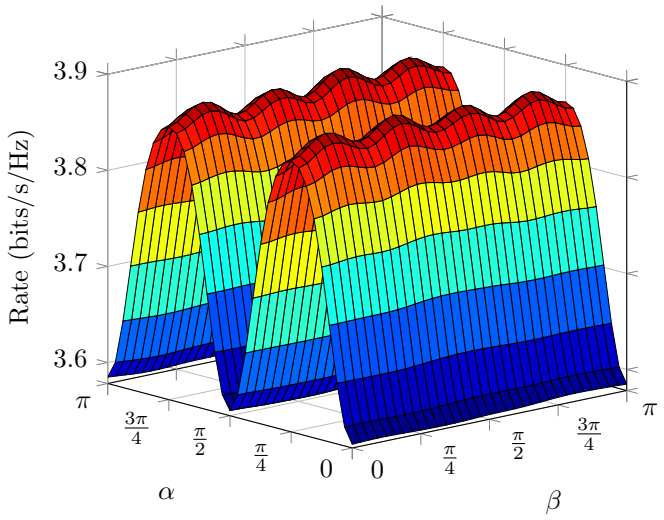


Figure 3.3 – DP-QPSK rate function of the two angles for $\Lambda = 6$dB at $\rho = 8$dB.

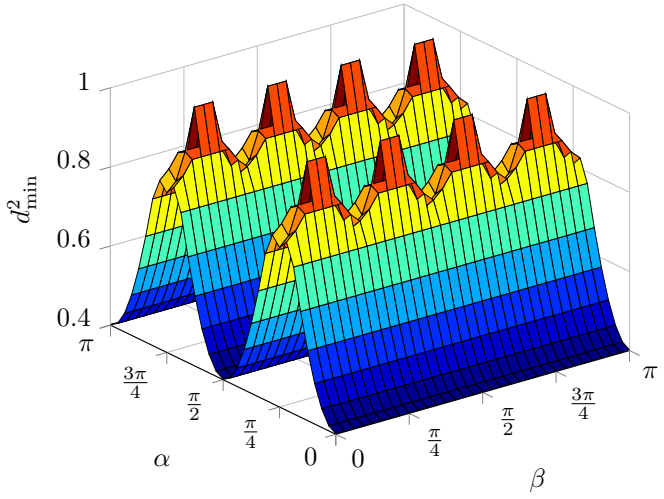


Figure 3.4 – DP-QPSK squared minimum Euclidean distance function of the two angles for $\Lambda = 6$dB.

specific to the channel of study with gain imbalance.

In the next chapter, we will see how to design PDL-resilient modulation schemes. In the next section we focus on an alternative and complementary rate metric that is suitable for conventional optical receivers.

3.3 Mismatched System Capacity and Impact of Receiver Architecture

The previous sections have shown that the system performance measured by the achievable information rate is highly dependent on the choice of the modulation alphabet. In the case of the lumped PDL channel model, the performance associated with discrete DP-QAM modulation can be severely degraded at certain SOPs whereas more rotationally invariant alphabets (such as in the ideal continuous Gaussian case) could be more *PDL-resilent*. The main objective of this thesis is therefore to derive PDL-resilient modulation schemes, also called polarization codes, and our main contribution is precisely to derive novel, optimal, and low-complexity signaling in the next chapter. Let us first discuss how, in practice, it will be possible to incorporate such advanced multi-dimensional modulation codes into high-speed receivers while keeping a low implementation complexity. To this aim, we need to discuss both system architecture using state-of-the-art receiver based on probabilistic graphical models [51, 52, 53, 54] and the need for non-mismatched receiver architecture.

In this section, we propose to discuss the implementation of a receiver processing chain to decode bits after propagation in a PDL channel. We show that a conventional receiver with a sequential processing chain induces rate losses compared to an optimal processing. We first present the rate associated with this imperfect decoding, elaborate on the method to use for an optimal decoding and then compare the BER of the two decoding schemes in simulation.

3.3.1 Sequential Processing in Conventional Receiver: the Need for an Adapted Metric

In Fig. 3.5, we represent a joint (dashed line) and sequential (full line) signal processing chain including only Maximum Likelihood (ML) symbol decoding, and bit decoding. Information theory indicates to perform both at the same time as developed in Appendix C. In practice, in conventional receivers, symbol decoding and bit decoding are performed only once in a sequential manner. It enables a significantly reduced complexity compared to the optimal joint equalization-decoding.

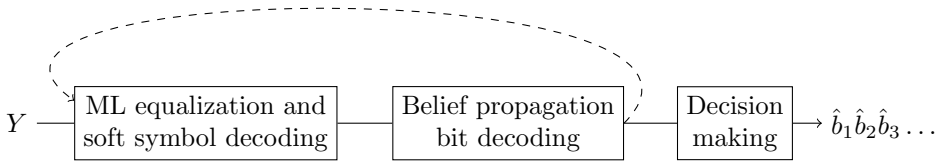


Figure 3.5 – Transmission chain: conventional, sequential processing (solid line) versus joint processing (dashed line).

Because the information rates presented in the previous sections assume optimal decoding at the receiver, there is the need to define a new metric characterizing the performance of the sequential processing chain.

Definition 4 *The* Bit-Interleaved Coded Modulation *(BICM) rate is the sum of the rates of the underlying bits as if they were sent independently, corresponding to* $\sum_i \mathcal{I}(B_i;Y)$ *where* B_i *is the bit random variable carried by the sent symbol and* Y *is the symbol output of the channel.*

The BICM metric [56] is sometimes referred to as the Generalized Mutual Information (GMI) [55], particularly in optical communications. This *mismatch* metric misses the fact that the bits are correlated by the propagation through a channel. Retrospectively, the real limits of communications are called *Coded Modulation* (CM) rates. If $\mathcal{I}(X;Y)$ is the CM rate corresponding to the properly said mutual information between the input X and output Y, then we have [56, 57]

$$\sum_i \mathcal{I}(B_i;Y) \leq \mathcal{I}(X;Y) \tag{3.9}$$

where B_i is the bit random variable of the i-th bit carried by X. Implicitly, this metric is bit-mapping dependent. In the context of gain imbalance and sensibility to the orientation of the channel, working with the BICM metrics may radically change results obtained from considerations on the CM one. Performance evaluation in terms of BICM rate will lead to different conclusions as we will explore in the rest of this chapter.

3.3.2 BICM Rates over a PDL Channel

The previous section suggested that a different performance has to be expected when implementing a conventional sequential processing at the receiver, and in this section we present this impact relative to a PDL channel. A subtle phenomenon appears when we plot the mismatch information rate in dashed lines in Fig. 3.6, corresponding to the performance of the sequential signal processing in a conventional receiver.

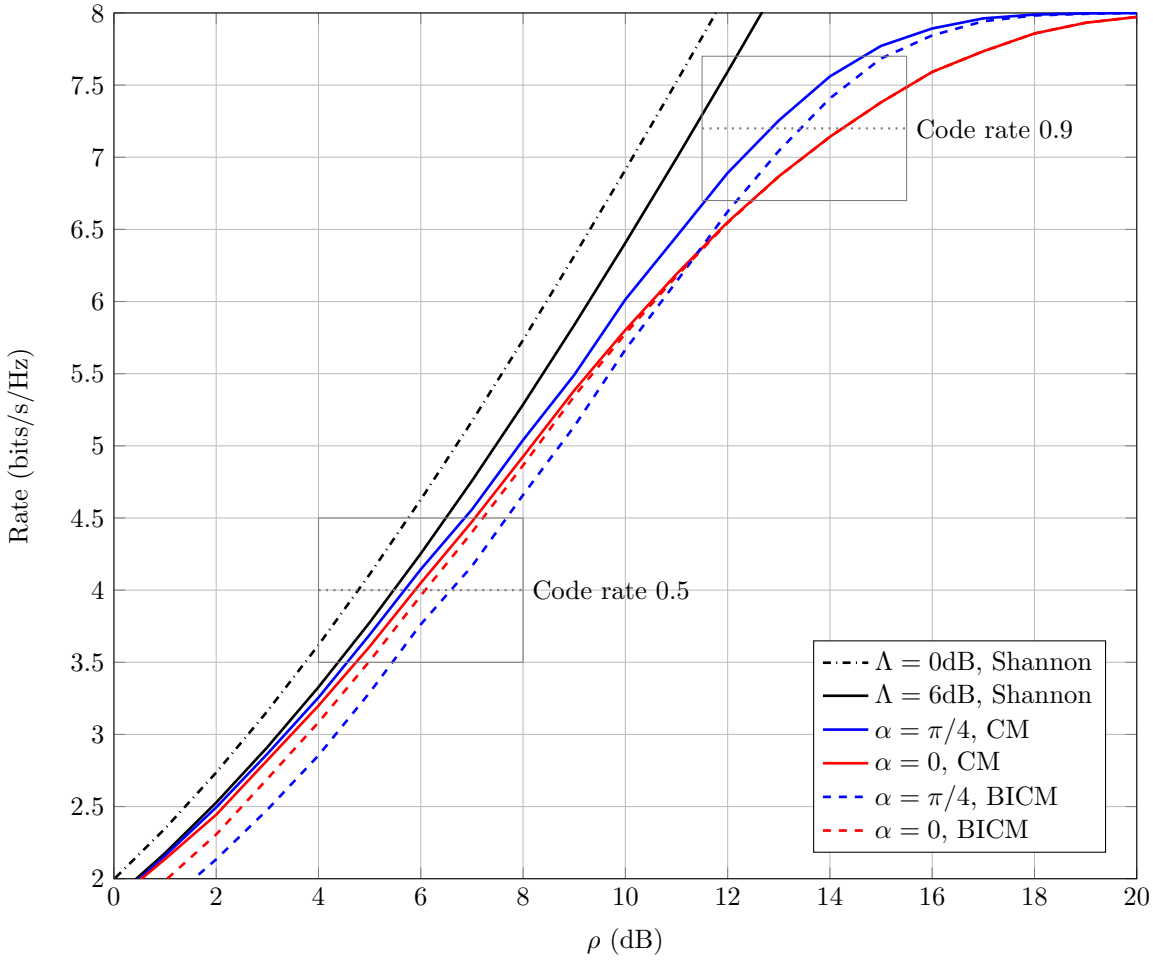


Figure 3.6 – CM and BICM rates function of the SNR for a DP-16QAM modulation over the simple PDL channnel with $\Lambda = 6$dB.

Namely, consider standard DP-16QAM modulation along with a Gray mapping, a significant loss is induced when a conventional processing is applied in presence of PDL with an angle different than 0 mod $\pi/2$. This loss can be seen on the mismatch capacity for $\alpha = \pi/4$ (dashed line) in Fig. 3.6 and appears even for (one pass of) optimal channel symbol decoding. Observe indeed that the BICM curve for $\alpha = \pi/4$ shows an impact of sequential processing greater than for $\alpha = 0$ for lower regions of SNRs. Indeed, while at $\alpha = \pi/4$ the 2×2-MIMO channel is fully used, if no joint decoding is performed the information on one polarization does not contribute to the decoding of the other. The two tributaries are then just noise to each other leading to a reduced performance compared to $\alpha = 0$ where they are well separated. This situation corresponds to the factor graph in Fig. 3.7 being cut between the two polarizations, the information on one polarization being not used for estimating the other one. Similarly, the existing loss between the CM and BICM rates for $\alpha = 0$ corresponds here with DP-16QAM to losses due to the mapping: the information on one bit is not circulated in the factor graph represented in Fig. 3.7 because no decoding loop is performed.

3.3.3 Impact of Sequential Processing on the Probability of Error

We will now exemplify the previous observations in terms of Bit Error Rate (BER). ML channel equalization and constellation demapping are performed with the optimal local sum-product rules and derived from the standard factor graph framework developed in Appendix C. Operationally, intrinsic Log-Likelihood Ratios (LLRs) l_i coming from the left of each variable bit b_i in Fig. 3.7, are fed into the decoder that implements the belief propagation algorithm in floating point.

The two architectures represented in Fig. 3.5 are then compared: a conventional processing from which a decision is made after a single pass through the equalizer-demapper-decoder chain (solid line), and a joint processing iteratively using extrinsic LLRs from the bit decoder feeding the equalizer-demapper for n times. A final decision on the bits $\hat{b}_1\hat{b}_2\hat{b}_3\ldots$ is made after the final loop. Extrinsic LLRs, arriving from the right to the bit variable in Fig. 3.7, reflect the probability of a bit calculated thanks to the probability of other received bits and the structure of the code that link them. Note that using this joint-processing scheduling is one way to approach capacity using the framework of factor graphs. It is an arbitrary choice to graphically separate decoding from equalization-detection.

For illustration, simulation results are presented in Fig. 3.8 using DP-16-QAM modulation with a half-rate spatially-coupled $(3, 6, 80000)$-LDPC with blocklength 80000 (with lifting depth 2000 and minimal coupling window [58]) for error-correction and performing $n = 9$ demapping iterations. Capacity limits drawn as vertical thresholds are reported from Fig. 3.6 at half of the maximum information rate. First, from Fig. 3.8, we see that the joint processing enables significant performance gains, at best 0.9dB when $\alpha = \pi/4$, when compared to the corresponding conventional

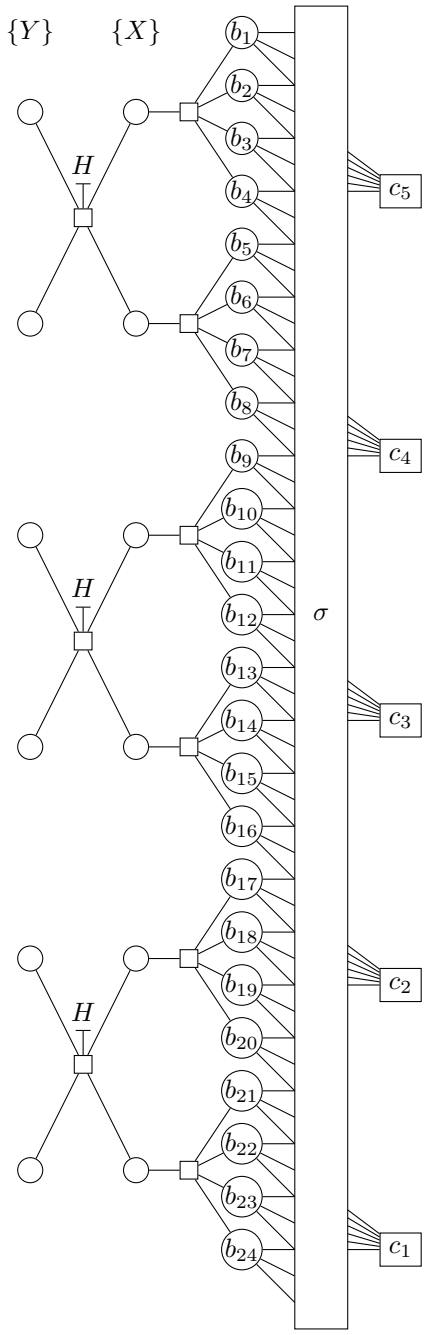


Figure 3.7 – Factor graph representation of joint receiver including a channel equalization and symbol demapper (DP-16QAM) part on the left of the bits; and the LDPC bit decoder on their right with the edge permutation matrix σ.

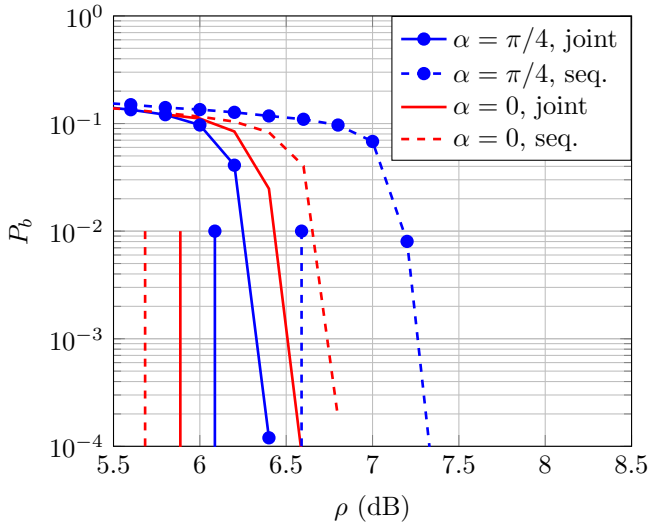


Figure 3.8 – Capacity limits (vertical lines) and BER curves using a spatially-coupled LDPC with a DP-16QAM modulation in a PDL channel with $\Lambda = 6$dB.

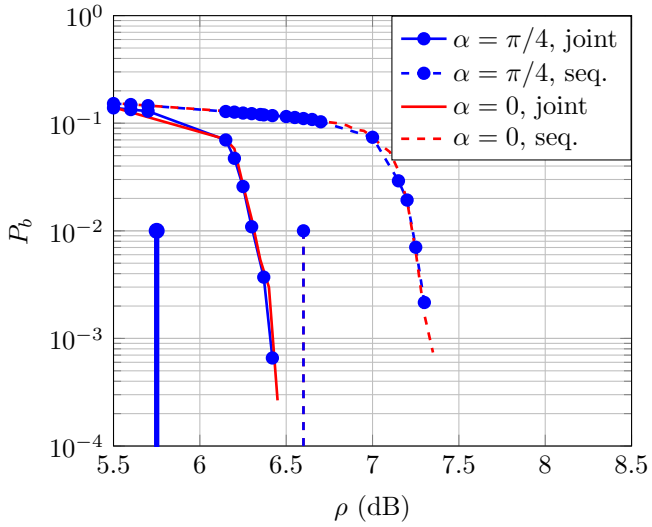


Figure 3.9 – Capacity limits (vertical lines) and BER curves using a spatially-coupled LDPC with a Silver-16QAM modulation in a PDL channel with $\Lambda = 6$dB.

architecture. Such gains can be understood as both *equalization* and *demapping* gains. Beware that the worst-case scenario can be $\alpha = \pi/4$, and not the aligned case $\alpha = 0$, when a conventional processing is used. Second, as expected from rate curves, further gains are seen when using Silver-16QAM exemplified in Fig. 3.9. Recall that the Silver code is the best-known signaling for PDL mitigation, but this property is valid only if an optimal processing is performed at the receiver. Observe indeed that using the Silver code with a sequential processing schemes actually worsen the BER compared to no coding. The mismatch capacities drawn as dashed vertical thresholds are computed from simulated BICM rate curves as in the case of standard DP-16QAM.

Finally, it is important to note that BICM and CM information rates might be ordered differently depending on the underlying code rate. For a code rate of 0.5 delimited by the lower rectangle in Fig. 3.6, we see that the mismatched system at $\alpha = \pi/4$ performs roughly 0.5dB worse than at $\alpha = 0$. Equivalently, observe that at $\alpha = \pi/4$, a rate loss of 0.4 bits/s/Hz is observed when performing sequential processing. This behavior could however be potentially reversed for a higher coding rate of 0.9 (delimited with the top rectangle) and, hence, has to be carefully studied for practical transceiver design.

3.4 Summary of the PDL Impact Over the Lumped Channel

Before proposing PDL-resilient modulations in the next chapter, we conclude that the PDL impact on the channel performance is three-fold. The first fundamental impact is the impairment due to the gain imbalance itself Λ. It corresponds to 0.6 bits/s/Hz for 6dB of PDL. The second impact is relative to the use of discrete modulation and their orientation in a channel that is unbalanced. The goal of the modulations built in the next chapter will be to mitigate the angle dependency of discrete modulation schemes by balancing their coordinate in the best possible way. Finally, the third more subtle impact is the loss induced by sequential processing usually performed in an optical receiver. Depending on the coding rate, additional rate losses are to be expected with up to 0.4 bits/s/Hz for the considered code rates. Because optimal receiver architecture can be designed, we will forget this limitation and focus on CM rates when designing PDL-resilient codes because they represent the proper limits of communications over the PDL channel.

In this chapter, we have reviewed how to integrate optimal PDL-resilient signaling into efficient receivers. This is easily done using the latest advance in information and coding theory as it is possible to approach the fundamental limits associated with a given channel input under the assumption that the receiver architecture is not mismatched. In the sequel, we focus on the critical component for PDL-resilient systems: the key enabler becomes the construction of optimal and specifically tailored modulation code.

Chapter 4

Optimal PDL-Resilent Modulation Schemes

The previous chapter presented the limits of communications associated with the standard QAM modulation scheme, and the interest in developing novel and efficient signaling schemes. In this chapter, we seek the optimal unitary precoding to apply to the QAM modulation scheme in order to offer a resilience to PDL, with a particular focus on increasing the worst rate over all channel states. Hence, we develop novel encoding techniques in order to improve this worst case, alternating between information rate optimization and Euclidean distance analysis.

4.1 Spatially-Balanced Signaling: a Modulation Resilient to Gain Imbalance by Design

The simple PDL channel introduced in 2.4.2 gives insight into the construction of PDL-resilient signaling. In this section we present the Spatially Balanced (SB) signaling that leverages the performance of square M-QAM based modulation by improving the worst-case capacity. Our first approach is based on information rates considerations, and we then elaborate on the optimal configuration based on Euclidean distances observations.

4.1.1 Decomposition of the Simple PDL Channel into Two Independent Subchannels

Recall that the simple PDL channel $H = D_\gamma R_\alpha$ presented in section 2.4.2 is real-valued. Then, consider $Y = HX + Z$ as two independent channels

$$Y = HX + Z \equiv \begin{cases} \Re Y = D_\gamma R_\alpha \Re X + \Re Z \\ \Im Y = D_\gamma R_\alpha \Im X + \Im Z \end{cases} \tag{4.1}$$

where $\Re A$ and $\Im A$ represent the respective real and imaginary part of A. As such, the two polarization streams of the standard DP-QAM modulation scheme are sent into the PDL channel with the *same* incident angle α. In order to mitigate the impact of worst-angle configurations, we propose the construction of a novel signaling called the *Spatially Balanced* (SB) signaling.

Given that the two channels are independent, we rotate by an angle η one of the complex part (real or imaginary part) of the original DP-QAM modulation. Therefore, if f_η represents the SB encoding function rotating the imaginary parts, Eq. 4.1 becomes

$$Y = Hf_\eta(X) + Z \equiv \begin{cases} \Re Y = D_\gamma R_\alpha \Re X + \Re Z \\ \Im Y = D_\gamma R_\alpha R_\eta \Im X + \Im Z \end{cases} \tag{4.2}$$

The real parts remain identical as before, but the imaginary parts now see a channel with a *different* angle $\alpha + \eta$, hence *balancing* the performance of the two subchannels.

In Fig. 4.1, we represent in blue the rate profile at an SNR of 8dB of the DP-QPSK modulation extracted from Fig. 3.3 but for the simple PDL channel taking the plane $\beta = 0$. From the symmetry of the rate profile, because the minimum and maximum are respectively in 0 and $\pi/4$, it is natural to take an offset angle of $\eta = \pi/4$. Indeed, when one complex part experiences a channel with lower achievable rate, the other part experiences a better channel, leading to an averaged performance. The rate of the novel SB-QPSK is additionally represented on the same figure in red. Observe that the SB-QPSK worst rate is increased to the middle value (50%) between the DP-QAM minimum and maximum values. By construction, as a balancing signaling, the fluctuations of the rate with respect to α are reduced. Those two observations are valid at any other SNR, and the worst case increase is the most significant in the regions of SNR where the DP-QAM has its maximum range of rate amplitudes.

The information rate of the SB-QPSK is also represented in Fig. 4.2 where we considered again the angle β, and still at an SNR $\rho = 8$dB. A curve of the DP-QPSK rate profile at the same SNR is added to have a reference of the dynamics of the SB-QPSK values. With this second angle taken into account, observe that the SB-QPSK worst rate is still increased by 50% compared to regular

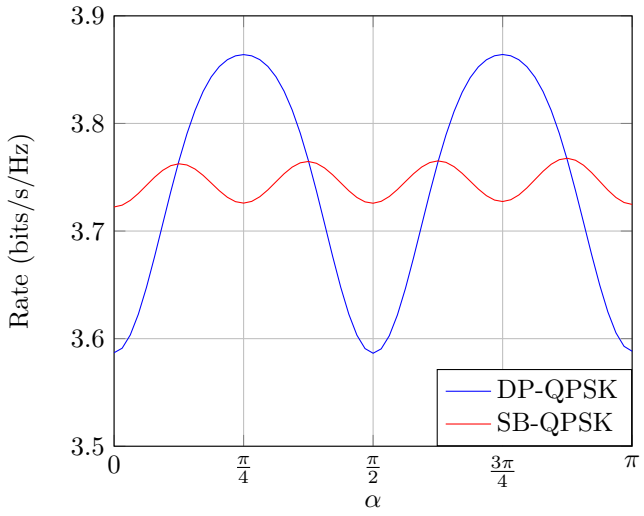


Figure 4.1 – SB-QPSK rate function of α at an SNR of 8dB and for $\Lambda = 6$dB compared to the one of DP-QPSK.

DP-QPSK. The considered simple PDL channel appears to be a worst channel modelization for the proposed signaling. Alternatively, we display in Fig. 4.3 the minimum distance of the obtained SB-QPSK codebook for all SOP pairs (α, β). Note that this time the range of $d^2_{\min}$ values is the *same* as the one of the DP-QPSK previously shown in Fig. 3.4. This may seem surprising in front of the effective increase of the information rate. Section 4.2.3 will explain that, while the minimum distance value is unchanged by the SB encoding, its occurrence is divided by 2 in the ensemble of all possible point-to-point distances after channel scaling. However this reduction of occurrence leads to rates that are enhanced since they are based on an aggregation of the Euclidean distances.

An experimental validation of the rate increase is exposed in section 5.3 where this first signaling is tested on an optical device with gain imbalance. More generally, and to be consistent with section 4.2, SB encoding can be defined by rotating both the real and imaginary parts of the DP-QAM constellation to get an equivalently divided offset angle η between the two rotated M-QAM-like square lattices on each complex part. In other words, instead of sending $X \in M$-QAM it can be encoded as $R_{-\eta/2}\Re X + iR_{\eta/2}\Im X$. We now build further upon this result and provide a general framework to analyze the optimal value of angle η in more general cases.

4.1.2 Optimization of the Subchannels Offset

As previously mentioned, the simple PDL channel model enables the study on either the real or the imaginary part. Instead of studying the capacity of M^2-QAM modulation, in this section we focus

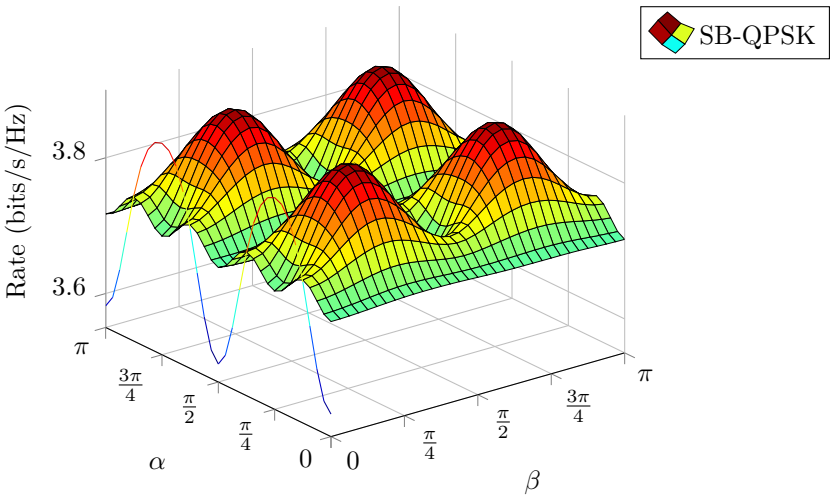


Figure 4.2 – SB-QPSK rate function of α and β for $\Lambda = 6$dB.

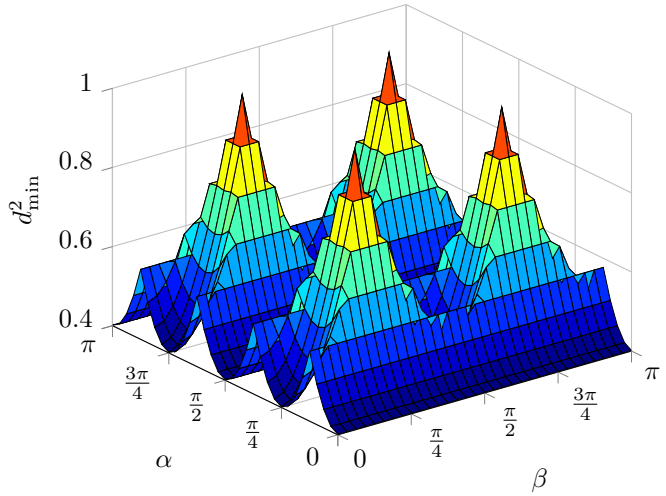


Figure 4.3 – SB-QPSK squared minimum Euclidean distance function of α and β.

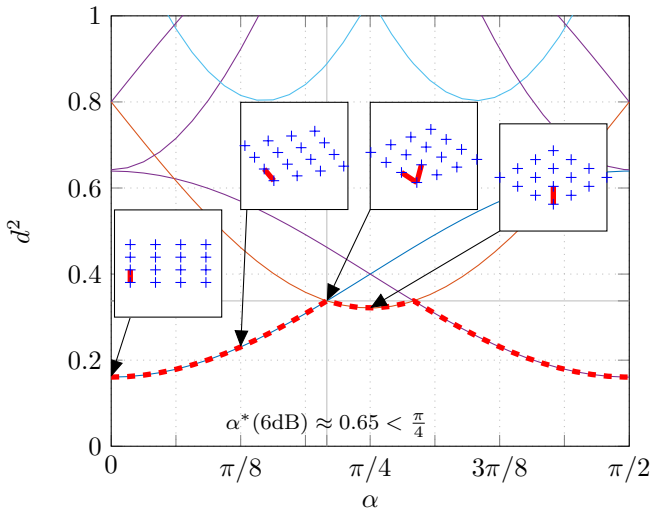


Figure 4.4 – Squared distances d^2 in $\mathcal{C}_{\lambda\alpha}$ as a function of the angle α for $\Lambda = 6$dB. The lower envelop indicates $d^2_{\min}$. It achieves its maximum in $\alpha^* \approx 0.65$.

on Pulse-Amplitude Modulation (PAM) modulation representing one complex part of the complex signal, and since there are two polarization we rather study the capacity of M-PAM^2. Specifically we propose to focus on the minimum Euclidean distance after PDL scaling as its evolution is highly correlated to the one of the information rate for high enough SNRs [50].

For a given PDL value and for any point in the lattice $A_{ij} \in M\text{-PAM}^2 = \{A_{ij} = (2i-1, 2j-1)^T | i,j \in \mathbb{Z}, -M/2 \leq i,j \leq M/2\}$, let the set of the symbols after multiplicative PDL-impairment be $\mathcal{C}_{\lambda\alpha} = \left\{H_{\gamma\alpha} A_{ij} | A_{ij} \in M\text{-PAM}^2\right\}$. Let $d(A_{ij}, A_{kl}, \gamma, \alpha)$ be the distance between any two points A_{ij} and A_{kl} of the M-PAM^2 lattice, with i,j,k and $l \in \mathbb{Z}$, for a given γ (or equivalently λ when needed) and angle α. We define the minimum distance $d_{\min}(\mathcal{C}_{\lambda\alpha}, \lambda, \alpha)$ associated with the codebook $\mathcal{C}_{\lambda\alpha}$ for a given α and λ. The search of this minimum distance should be operated over the codebook for each given α and λ. Finally, we are interested in the angle

$$\alpha^*(\lambda) = \underset{\alpha}{\operatorname{argmax}}\, d_{\min}(\mathcal{C}_{\lambda\alpha}, \lambda, \alpha) \tag{4.3}$$

that defines the maximum-diversity phase offset for SB signaling.

For symmetric square constellations, it is sufficient to consider $\alpha \in [0, \pi/4]$ by $\pi/2$-periodicity and evenness. While the presented concepts extend to any order M-QAM, Fig. 4.4 illustrates the particular case of 4-PAM^2. In Fig. 4.4, the lowest point-to-point distances of the 4-PAM^2 lattice after PDL scaling are represented for a fixed $\Lambda = 6$dB as a function of the angle α. In superimposition, the deformation of the lattice is represented at given values of α. For that particular Λ, the

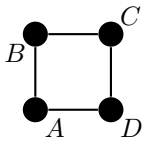
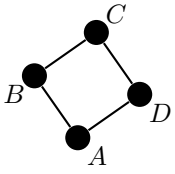
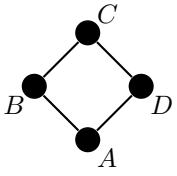
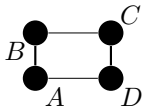
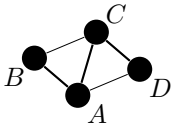
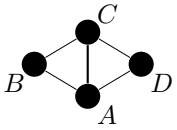
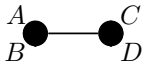
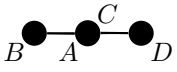
	$\alpha = 0$	$\alpha = \alpha^*(\Lambda)$	$\alpha = \pi/4$
$\Lambda = 0$dB			
$\Lambda = 6$dB			
$\Lambda \to \infty$			

Table 4.1 – Illustration of the distortion of the 2-PAM^2 fundamental polytope in presence of PDL for various angles α. For $\Lambda = 0$dB, the minimum distance is constant for any α therefore $\alpha^*(0\text{dB})$ is not unique.

minimum distance in dashed red line has several local maxima, given by the intersections of some of the represented distances, namely the ones in blue and red. In this section we give a framework to properly predict these intersections, which contain the solution of the optimal angle η to assign to the SB signaling. Notice that this study is also initiated in [59] in the context of space-time codes and wireless communications. In contrast, our work originates from optical communications with polarization and/or mode multiplexing. Hence, it extends [59] and generalizes to any square lattice scaling.

We thereafter demonstrate that different regimes are to be considered for large PDL/MDL levels. Geometric considerations show indeed that, when the imbalance increases, the distance profile evolves with the α-directional scaling of the lattice. In the general case of square two-dimensional lattices M-PAM^2 and small imbalance, the sides of the fundamental polytope give the minimum distance and $\alpha^* = \pi/4$. For large PDL values, this value of angle is not optimum as can be understood from the figures in Table 4.1. For some PDL value, at certain angles α, the diagonal of the 2-PAM^2 polytope may be lower than the one of its smaller edge. The first transition corresponds to the equality

$$\begin{aligned} & d^2(A_{ij}, A_{i+1,j+1}, \gamma, \alpha) = d^2(A_{ij}, A_{i+1,j}, \gamma, \alpha) \\ \equiv \quad & (1+\gamma)(2\cos\alpha - 2\sin\alpha)^2 + (1-\gamma)(2\cos\alpha + 2\sin\alpha)^2 \\ & = 4(1+\gamma)\sin^2\alpha + 4(1-\gamma)\cos^2\alpha \\ \equiv \quad & \tan^2\alpha + 2(1-\lambda)\tan\alpha + \lambda = 0 \end{aligned}$$

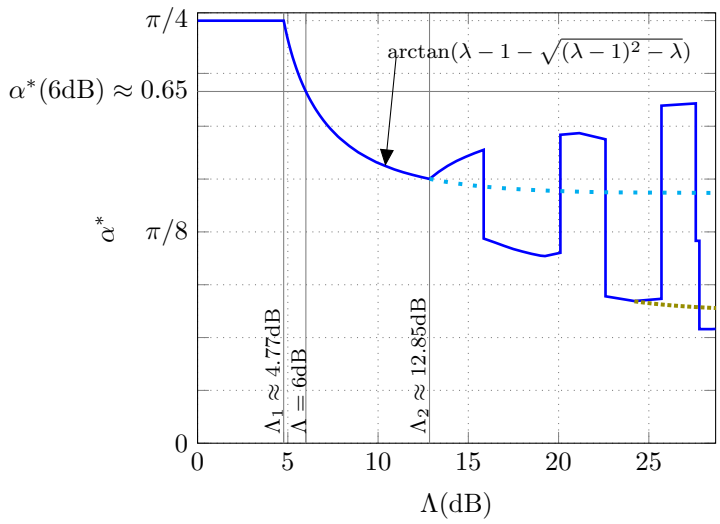


Figure 4.5 – Optimal angle α^* as a function of Λ. For 4-PAM^2 and 2-PAM^2, α^* is lower-bounded by the dashed asymptotes converging to $\arctan(1/2)$ and $\arctan(1/4)$, respectively.

where again $\lambda = (1+\gamma)/(1-\gamma)$. The above equality exists (in the real field) for $\lambda \geq (3+\sqrt{5})/2$ and makes α in $[0, \pi/4]$ for $\lambda \geq \lambda_1 = 3$ or $\Lambda_1 \approx 4.77$dB. In other words, for $\lambda < \lambda_1$, $\alpha^*(\lambda) = \pi/4$, but above λ_1, $\alpha = \pi/4$ is no longer the $d_{\min}$ maximizer. It can be understood from Fig. 4.4, the orange middle curve having intersections with the lower curve only for values of $\lambda \geq \lambda_1$. The minimum distance (dashed red line) in Fig. 4.4 was already observable in the 2-angle minimum distance profile presented in Fig. 3.4 in the plane for $\beta = 0$. At λ_1, $\alpha^*(\lambda)$ expression changes to

$$\alpha^*(\lambda) = \arctan\left(\lambda - 1 - \sqrt{(\lambda-1)^2 - \lambda}\right) \tag{4.4}$$

which is derived as the solution in α of the above mentioned distance equality. The deformations of the 2-PAM^2 fundamental polytope are summarized in table 4.1. For the singular case with no PDL, any distance of the square is minimum for all angles. For $\Lambda = 6$dB (after transition Λ_1), the angle α^* offering the maximum minimum distance is given by Eq. 4.4. Eventually, for $\Lambda \to \infty$, Eq. 4.4 gives $\alpha^* \to \arctan(1/2)$.

For $M > 2$, distances between points that were originally well-separated may shrink to become the minimum distance. While $d_{\min}(\mathcal{C}_{\lambda\alpha^*})$ is continuous and decreasing in λ, $\alpha^* = \alpha^*(\lambda)$ is a piece-wise continuous function that is fully analytically characterized and presents discontinuous phase transitions. For example, if $M > 2$, $d(A_{ij}, A_{i+1,j+2}, \gamma, \alpha)$ vanishes as λ increases. A second transition appears when a solution exists for $d(A_{ij}, A_{i+1,j+2}, \gamma, \alpha) = d(A_{ij}, A_{i+1,j+1}, \gamma, \alpha)$ at $\lambda_2 = (29 + 8\sqrt{13})/3$ or $\Lambda_2 \approx 12.85$dB.

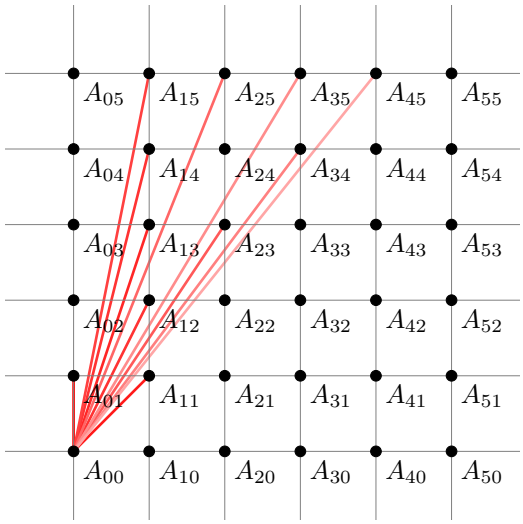


Figure 4.6 – Pair of points which distance vanishes at certain orientation when PDL increases.

The function α^* is represented up to $\Lambda = 28.7$dB in Fig. 4.5. Although these high values are unrealistic for PDL, the study is of interest for other MIMO schemes with higher imbalance such as SDM systems. The values λ_k at subsequent α^* transitions are obtained from the equality of corresponding distances in $\mathcal{C}_{\lambda\alpha}$. As the equality in α yields two roots, the one that gives the maximum distance should be taken. Note that the two roots may be reordered as λ keeps increasing. The involved distances are represented in Fig. 4.6 in red. The lighter the red, the higher the gain imbalance should be for the two points to get close. In this figure, we only represented distances in the upper left triangle corresponding to angle solutions α^* in $[0, \pi/4]$ and we draw a line linking two neighbor points in a given direction. For instance, A_{00} is not linked to A_{02} because, under PDL action, A_{00} would get closer to A_{01} and the distance to the latter point is always lower. Notice that from additional geometric observations on the lattice grid, it can be shown that $\lim_{\lambda\to\infty} \alpha^*(\lambda) = \arctan(1/M)$. This can already been seen in Table 4.1 where for $\Lambda \to \infty$, $\alpha^* \to \arctan(1/2)$.

We further refine the distance-based observations for optimizing SB signaling by studying the actual information rate for PDL-impaired systems. Recall that fundamental communication limits

Λ	2dB	4dB	6dB	15dB
$\alpha^*[d_{\min}]$	$\pi/4$	$\pi/4$	0.65	0.53
$\alpha^*[\mathcal{I}(X;Y)]$	$\pi/4$	$\pi/4$	$\pi/4$	0.52

(4.5)

Table 4.2 – Optimal SB angle η according to the information rate $\mathcal{I}(X;Y)$ or to the minimum Euclidean distance $d_{\min}$ at a 0.9-coding rate.

are measured by the information rate. For sufficiently large SNR, it becomes a function of $d_{\min}$ [50]. As shown in Fig. 4.7 the information rate is *not* always maximal for $\alpha = \pi/4$. Observe as well that the high plateau of the information rate is located where the minimum Euclidean distance is maximum. This is directly explained by our $d_{\min}$ investigations.

Another interesting example is the BICM rate depicted in Fig. 4.8 where we compare it to the standard CM rate on a PDL channel as a function of the incident angle α and with gain imbalance $\Lambda = 6$dB, using a Gray-mapped DP-16QAM at an SNR of 15dB. The CM rate aggregating the Euclidean distances value, the considered SNR has to be high enough for the found α^* to be correct. In particular, we found that for operational rates around 0.9 and at a PDL of 6dB, the approximation $\alpha^* \approx \pi/4$ is valid, but for the BICM metric the optimum value of α deviates from $\pi/4$ for practical coding rates. In Fig. 4.8, the maximum of the BICM rate is clearly not in $\alpha = \pi/4$ and the considerations on the contraction of the underlying lattice is even more critical. This amplification can be understood because, with the BICM metric, not only the distance between points count but also the affected bit difference or Hamming distance [60]. This means that in the BICM metric, the directional compression for instance at $\alpha = \pi/4$ is much more impacting than in the CM rate, because bits on a diagonal have more than one differing bit.

In Table 4.2, we reported the α^* values obtained from our Euclidean distance analysis, and compare them with the optimum angle as given from the information rate. As expected, for finite SNR, this slightly differs from Euclidean distance results. Nevertheless, the point of the study was to emphasize that the optimum offset angle is not always $\pi/4$, especially for high PDL regimes.

4.2 Derivation of the 4D Optimal Signaling: the New Spatially Balanced

After focusing on the simple PDL channel in the previous section, we now generalize our findings to the proper PDL channel We show how to derive optimal orthogonal 4-dimensional code and give a construction that we analytically prove to be exact for large SNRs and practical PDL level. We start by elaborating on the general framework regarding unitary transforms.

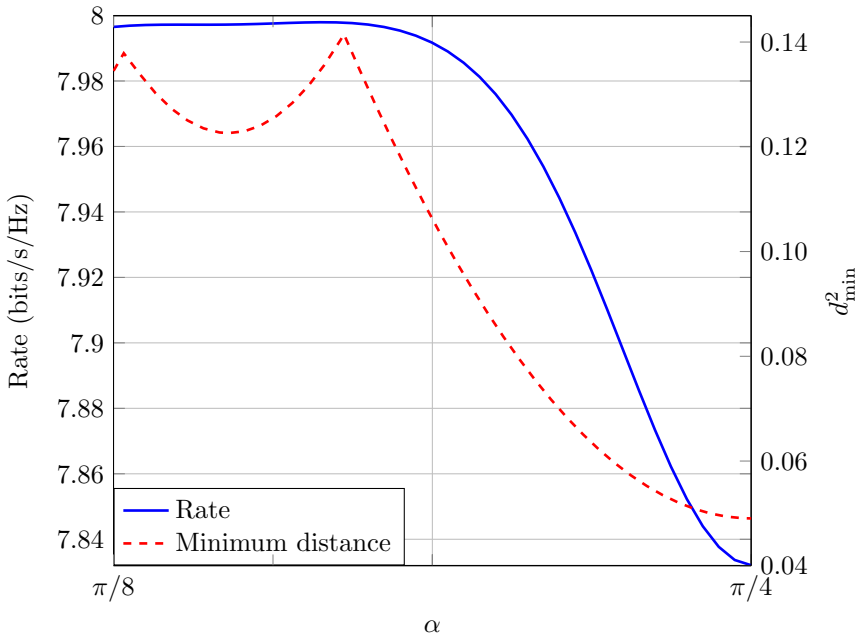


Figure 4.7 – Information rate and $d^2_{\min}$ variation for $\alpha \in [\pi/8, \pi/4]$ and 16-QAM^2 at 0.9 coding rate, $\Lambda = 15$dB PDL at an SNR of 23.1dB.

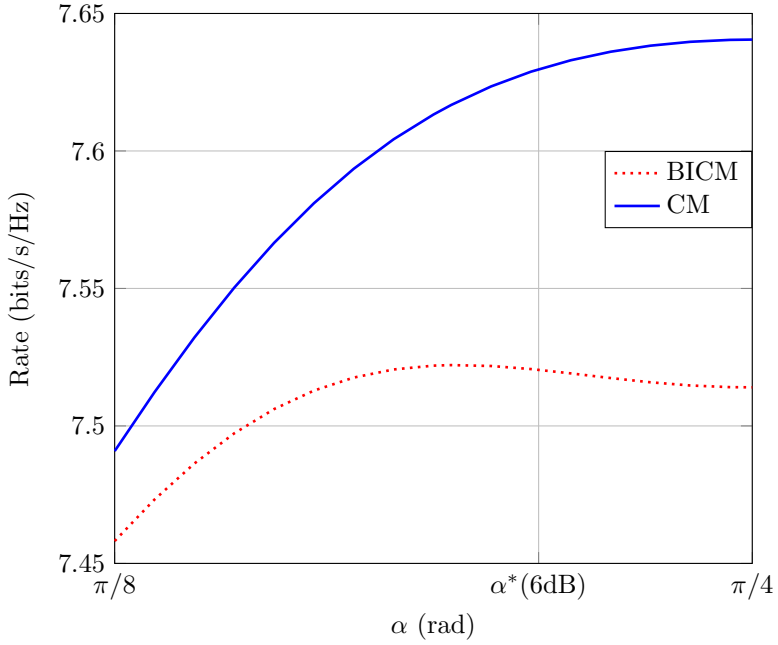


Figure 4.8 – CM and BICM rate variation for $\alpha \in [\pi/8, \pi/4]$ and DP-16QAM at 0.9 coding rate, for 6dB of PDL at SNR $\rho = 15$dB.

4.2.1 SO (4) Transforms as a Low-Complex Way to Mitigate PDL

Communication theory deals [61, 62] with the design of proper representations (coding and modulation) of information over physical electromagnetic waves. In this section, we focus on the design of efficient four-dimensional schemes targeted to optical MIMO schemes. This section deals with the design of small but efficient four-dimensional codes [63, 37, 64, 66, 65, 59] targeted to optics.

Typically, the channel input X is a complex-valued discrete random vector with entries taking on values for example in square M^2-QAM or, equivalently, in M-PAM^2, or, as in the sequel, in $F.(M\text{-PAM}^4)$, where $F \in \mathrm{SO}(4)$. The random matrices U and V are picked uniformly at random from $\mathrm{SU}(2)$ as explained in section 2.5. They are assumed to be static over the timeframe of an operational blocklength. This assumption is fair for optical fiber communication systems as fast state-of-polarization changes remain rare events. Because the multiplicative channel factor is modeled as constant, the channel resembles a block-fading model up to some rescaling. System analysis and code performance are then specific. 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 and with a pre-defined coding and modulation scheme. As a result, while 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*.

In the channel model presented in Eq. 2.8, the SOP is the matrix defined by $U = R_\alpha B_\beta \in \mathrm{SU}(2)$. Transformations in $\mathrm{SU}(2)$ that include all possible 2×2 unitary SOP rotations implemented through optical components cannot enhance the performance over the randomly oriented channel. Indeed, if the M-QAM^2 modulation worst rate is met for a SOP $R_{\alpha_0} B_{\beta_0}$, this modulation rotated by $V \in \mathrm{SU}(2)$ will encounter the same worst rate for the SOP $R_{\alpha_0} B_{\beta_0} V^{-1} \in \mathrm{SU}(2)$ leaving it unchanged.

Space-time codes introduced in wireless communications have recently become a topic of interest for polarization-multiplexed transmissions [34, 67] with the best-known performing Silver code or a simpler similar performing 2-timeslot code [9] which is no other than a 2×2 Threaded Algebraic Space-Time code (TAST) [68] for the chosen parameters. However these modulations introduce additional degrees of freedom, and imply a higher decoding complexity. Additionally, the introduced correlation between consecutive timeslots radically changes the processing paradigm at the receiver side and in particular it is not compatible with the classic blind decoding architecture.

Low-complexity four-dimensional schemes are preferred candidates for implementation in optical communications, using the 4 existing symbol dimensions that are the in-phase and quadrature phase components of the two polarizations. Remarkable insights on physical and non-physical four-dimensional rotations adapted to optics are first found in [69, 70].

Hence, we investigate orthogonal transformations in $\mathbb{R}^4$, and more precisely the Special Orthogonal group of order 4, $\mathrm{SO}(4) = \{G \in \mathbb{R}^{4\times4}, \det G = 1\}$. The canonical decomposition [71]

of a matrix $G \in \mathrm{SO}(4)$ is done through the product of a 4×4 left- and right-isoclinic matrices $G = G_L G_R$ with 3 degrees of freedom each [71]. A Hopf parametrization of these two matrices with $\eta, \theta, \nu, \alpha, \tau, \beta \in (-\pi, \pi]$ can be used to have the block-matrix form

$$G_L = \begin{pmatrix} \cos\eta R_\theta & -\sin\eta S_\nu \\ \sin\eta S_\nu & \cos\eta R_\theta \end{pmatrix} \text{ and } G_R = \begin{pmatrix} \cos\tau R_\alpha & -\sin\tau R_{-\beta} \\ \sin\tau R_\beta & \cos\tau R_{-\alpha} \end{pmatrix} \tag{4.6}$$

where $S_\nu \notin SO2$ is a *reflection* matrix with determinant -1 defined as

$$S_\nu = \begin{pmatrix} \cos\nu & \sin\nu \\ \sin\nu & -\cos\nu \end{pmatrix} \tag{4.7}$$

Each reflection in G_L is of determinant -1 but the resulting matrix G_L is of determinant 1 thus respecting the condition to be in $\mathrm{SO}(4)$. The product $G_L G_R$ commutes as shown in [69] or directly writing the two matrices parametrized with Hopf coordinates. The representation of $\mathrm{SU}(2)$ in $\mathrm{SO}(4)$ is a subset of the latter [69]; $\mathrm{SU}(2)$ matrices purely correspond to a right-isoclinic matrix G_R. These properties and additional comments on the $\mathrm{SO}(4)$ decomposition are detailed in appendix D. Therefore, it is sufficient to examine transformations with left-isoclinic only matrices that can potentially offer more robustness to PDL. These transformations applied over an M-QAM^2 symbol vector X are simplified in D.4 and can finally be written in $\mathbb{C}^2$ as $e^{i\theta} f_{\eta,\nu}(X)$ with

$$f_{\eta,\nu}(X) = \cos\eta X + e^{i\nu} \sin\eta \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} X^* \tag{4.8}$$

where X^* is the conjugate of X. The factor $e^{i\theta}$ a scalar phase rotation common to both polarization tributaries and that cannot offer any PDL resilience, we therefore omit it in what follows. This is the same phase omitted in section 2.4 when discussing the simplification of complex unitary matrices. Hence, from the 6 parameters of an $\mathrm{SO}(4)$ transform, only the two η and ν are to be optimized to construct a modulation scheme robust to PDL.

4.2.2 A First Numerical Optimization Approach as an Illustration

In this section, we want to define the optimal $\mathrm{SO}(4)$ unitary transform in terms of maximization of the worst performance over all possible channel SOPs. Hence, for the channel model

$$Y = D_\gamma R_\alpha B_\beta f_{\eta,\nu}(X) + Z \tag{4.9}$$

we want to find the pair that increases the most the worst information rate over all SOPs.

Definition 5 *For a given Λ, we define the* New Spatially Balanced *signaling as the $SO(4)$ transform of DP-M-QAM that maximizes the worst mutual information over all SOPs. The solution of this maximization reduces to finding the pair*

$$(\eta^*, \nu^*) = \underset{\eta,\nu}{\operatorname{argmax}} \min_{\alpha,\beta} \mathcal{I}(X;Y) \tag{4.10}$$

We now discuss the values of η and ν that increase the most the worst rate over all channel SOPs. Firstly, note that in the particular case where $\nu = 0$, Eq. 4.8 enables to construct the previously studied SB signaling

$$f_{\eta,\nu=0}(X) = \cos\eta \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \sin\eta \begin{pmatrix} -x_2^* \\ x_1^* \end{pmatrix} \tag{4.11}$$

$$= \begin{pmatrix} \cos\eta & -\sin\eta \\ \sin\eta & \cos\eta \end{pmatrix} \begin{pmatrix} \Re x_1 \\ \Re x_2 \end{pmatrix} + \begin{pmatrix} \cos\eta & \sin\eta \\ -\sin\eta & \cos\eta \end{pmatrix} \begin{pmatrix} i\Im x_1 \\ i\Im x_2 \end{pmatrix} \tag{4.12}$$

$$= R_\eta \Re X + i R_{-\eta} \Im X \tag{4.13}$$

This particular encoding leads to an offset angle of 2η between the two complex parts of X. In section 4.1, it is shown with information rate considerations that the optimal value is $2\eta = \pi/4$ for low values or PDL, or another value for higher gain imbalance regimes.

Now, we consider all possible (η, ν) pairs and conduct a numerical evaluation of $\mathcal{I}(X;Y)$ by means of Monte-Carlo simulations. We restrict to the channel model 4.9 for which we take a PDL value of 6dB and use a SOP grid with a step of up to $\pi/64$. In Fig. 4.9, $\min_{\alpha,\beta} \mathcal{I}(X;Y)$ is plotted when using QPSK^2 encoded with the function $f_{\eta,\nu}$ defined in 4.8, at an SNR of 8dB (corresponding to a coding rate of about 0.85). A coarse (η, ν) grid with a $\pi/32$ step is used for $(\eta, \nu) \in [0, \pi/4] \times [0, \pi/4]$ and a finer step of $\pi/64$ is considered for $(\eta, \nu) \in [\pi/8, 3\pi/16] \times [3\pi/16, \pi/4]$. We restrict the study to this range because the worst rate profile is found to be $\pi/2$-periodic, as well as even with respect to each variable η and ν.

The marked point at $(\eta, \nu) = (\pi/8, 0)$ corresponds to the SB worst rate. The point at $(0, 0)$ is the one of the uncoded QPSK^2 scheme. We notice that for η around $\pi/8$, non-zero ν values lead to an enhanced worst-case rate when compared to the SB signaling. Consequently, the worst capacity of the PDL channel after the encoding $f_{\eta,\nu}$ can be further increased. With the considered set of parameters (SNR, PDL, modulation format), we find a unique optimal pair

$$(\eta^*, \nu^*) \approx \left(\frac{5\pi}{32}, \frac{\pi}{4}\right) \approx (0.49, 0.79) \tag{4.14}$$

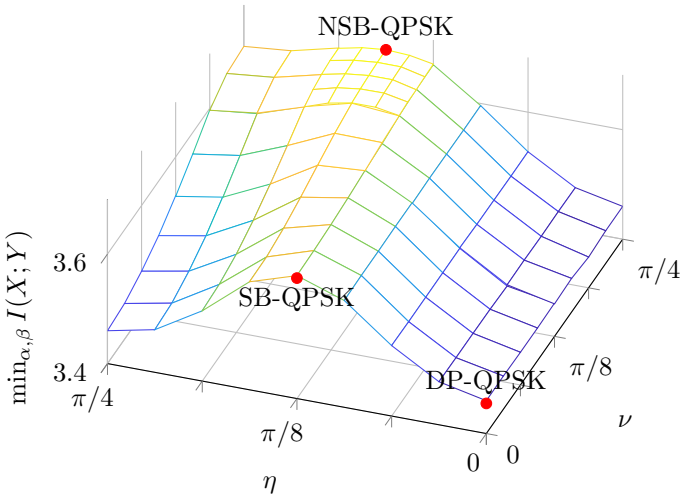


Figure 4.9 – Numerical evaluation of worst rate over all channel SOPs (α, β) for DP-QPSK encoded with the $f_{\eta,\nu}$ function for a PDL $\Lambda = 6$dB at an SNR of 8dB. By periodicity and evenness, the presented surface corresponds to all possible values that the minimum rate takes.

in the considered range $(\eta, \nu) \in [0, \pi/4] \times [0, \pi/4]$ with a $\pi/64$ precision. Other solutions follow from the periodicity outside this range. In the following section, we propose to give an analytic formula that match these approximated values.

4.2.3 Exact Polarization Code Construction Based on an Euclidean Distance Analysis

In this section, we provide a closed-form solution to our running optimization problem and we explain the result of the numerical optimization carried in the previous section using Euclidean geometry. The spatial balancing occurs now in 4 dimensions and is connected to the tetrahedron angles. Our main contribution is to give an analytic and exact solution to the initial optimization problem. This proves that the orthogonal NSB signaling is optimal in the sense of information theory as it minimizes the probability of error for large SNR and relatively small PDL.

To do this, first notice that if we first concentrate of the case of relatively small PDL, we can reduce the Euclidean distance optimization to the edges of the 4 dimensional hypercube. Let us now introduce a few notations relative to this section.

Notations

The information source denoted by $\hat{X} \in \mathbb{R}^N$ is encoded into $\hat{F}\hat{X} \in \mathbb{R}^N$ using a unitary precoding $\hat{F}$. We use

$$\hat{X} \in M\text{-PAM}^N = (2\ell+1)/\sqrt{\mathcal{E}_{M,N}} : \ell \in \{-\frac{M}{2}, -\frac{M}{2}+1, \cdots, \frac{M}{2}-2, \frac{M}{2}-1\}^N \tag{4.15}$$

where M even and

$$\sqrt{\mathcal{E}_{M,N}} = \sqrt{\frac{(M-1)M(M+1)}{3M}} \tag{4.16}$$

is a power scaling constant. The modulation points form the vertices of an hypercube in $\mathbb{R}^N$. Recall that the number of edges in an hypercube of dimension N is $N2^{N-1}$ and its number of vertices is 2^N. Transmissions occur via an $N/2$-dimensional complex-valued discrete channel model. In this section, we use the special unitary and orthogonal groups generalized to any dimension N

$$\mathrm{SU}(N) = \{A \in \mathbb{C}^{N\times N} : \det(A) = 1\} \subset \mathrm{U}(N), \tag{4.17}$$

$$\text{and } \mathrm{SO}(N) = \{A \in \mathbb{R}^{N\times N} : \det(A) = 1\} \subset \mathrm{O}(N), \tag{4.18}$$

respectively, for which the identity matrix I_N is the neutral element. Recall that elements of $\mathrm{SU}(N)$ have N^2-1 degrees of freedom and elements of $\mathrm{SO}(N)$ have $N(N-1)/2$. As an example, a rotation in $\mathrm{SO}(2)$ with angle $\theta \in (-\pi, \pi]$ is represented by the already defined rotation matrix R_θ.

The main objective of this section is to construct a practical linear modulation code (or precode) [63, 37] that transforms a standard M-PAM-multiplexed source into an *imbalance-resilient* signal modulation. We then want to find the optimal orthogonal [64] code f represented by $\hat{F} \in \mathrm{SO}(4)$ that minimizes the worst case error probability.

Optimal Left-isoclinic Polarization Codes

Let us now represent the transmission over the linear 2×2 polarization-multiplexed fiber channel model using four-dimensional real-valued vectors. The DP-QAM input $X \in M^2\text{-QAM}^2$ is equivalently represented as $\hat{X} = (\Re x_1, \Im x_1, \Re x_2, \Im x_2)^T \in M\text{-PAM}^4 \subset \mathbb{R}^4$ and is assumed to have normalized power constraint $E[X] = 1$. In what follows, the hats on the different complex variable correspond to their representation onto the real field. The correspondence between complex scalars, matrix and vectors and their real representation peers are detailed in Appendix D and briefly thereafter. From the previous sections, we see that the multiplicative non-unitary impairment can be expressed using the linear map $\hat{X} \mapsto \hat{D}_\gamma \hat{R}_\alpha \hat{B}_\beta \hat{X}$. These matrices can be expressed as real-valued

matrices of $\mathrm{SO}(4)$ as

$$\hat{D}_\gamma = \begin{pmatrix} \sqrt{\gamma+1} I_2 & 0 \\ 0 & \sqrt{1-\gamma} I_2 \end{pmatrix}, \tag{4.19}$$

$$\hat{B}_\beta = \begin{pmatrix} R_\beta & 0 \\ 0 & R_{-\beta} \end{pmatrix} \text{ and } \hat{R}_\alpha = \begin{pmatrix} \cos(\alpha) & 0 & -\sin(\alpha) & 0 \\ 0 & \cos(\alpha) & 0 & -\sin(\alpha) \\ \sin(\alpha) & 0 & \cos(\alpha) & 0 \\ 0 & \sin(\alpha) & 0 & \cos(\alpha) \end{pmatrix} \tag{4.20}$$

Let us consider 2-PAM^4 for simplicity. The main purpose of this section is to analytically construct an orthogonal encoding f represented by $\hat{F} \in \mathrm{SO}(4) : M\text{-PAM}^4 \to \mathbb{R}^4$ that maps $\hat{X} \mapsto \hat{F}\hat{X}$. This code is, in general, non-physical as it cannot be obtained from physical optical components that act directly on the two-dimensional complex-valued lightwave. Instead, precoding using $\mathrm{SO}(4)$ requires implementation in DSP [69] and cannot be directly obtained from optical components. 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, see e.g. [72, p.1073] and [73, Appendix E]. Then, we want to find the (not necessarily unique) code f defined by its encoding matrix $\hat{F}$ that minimizes the worst (across channel realizations) error probability,

$$\hat{F}^* = \underset{\hat{F}\in\mathrm{SO}(4)}{\operatorname{argmax}} \left\{ \min_{U\in\mathrm{SU}(2)} \{\mathcal{I}(\hat{X}; \hat{D}_\gamma \hat{U} \hat{F} \hat{X} + \hat{Z})\} \right\} \tag{4.21}$$

when $\rho \gg 1$. Therefore, the raw optimization above would involve 6 parameters. Nevertheless, by equivalences, we saw in the numerical optimization already performed in 4.2.2 that a $\mathrm{SO}(4)$ transform with 6 degrees of freedom reduces down to 2 for the optimization in the context of an optical channel. In practice, those 2 remaining degrees of freedom are referred to as non-physical in [69]. Compared to the numerical optimization conveyed in the previous section, we now want to present a closed-form solution via an *analytical* optimization over $\mathrm{SO}(4)$.

The 2-parameter function $f_{\eta,\nu}$ in Eq. 4.8, or equivalently its matrix representation $\hat{F}_{\eta,\nu}$, suffices to define any unitary (left-isoclinic) polarization coding of $\hat{X} \in \mathbb{R}^4$ as $\hat{X} \mapsto \hat{F}_{\eta,\nu}\hat{X}$. It is now time to reformulate our optimization objective in terms of Euclidean geometry. For large SNRs, it can be shown that the optimization of the distance profile dominates the information rate behavior (or, alternatively, notice that union bounding techniques [72, p.1073] can be used). The optimal construction nails down to an optimization using the parameter pair (η, ν) of the distance profile of codes $\{\hat{X} \mapsto \hat{D}_\gamma \hat{R}_\alpha \hat{B}_\beta \hat{F}_{\eta,\nu} \hat{X}\}_{\alpha,\beta}$.

The code construction over $\mathrm{SO}(4)$ is remarkable as the actual (worst case of the) minimum

distance $d_{\min}^{(1)}$ of $\{\hat{D}_\gamma \hat{U} \hat{F} \hat{X} : \hat{X} \in M\text{-PAM}^4\}$ cannot be improved as we will prove later. It is therefore an optimization on the distance distribution and, more critically, on the *second* minimum distance denoted $d_{\min}^{(2)}$ thereafter. We first state our main theorem, then present geometric arguments. A detailed 2-step proof with coordinate calculus is then presented.

Theorem 1 (Optimal construction over SO (4)) *Using the formalism of previous sections, and up to algebraic symmetry, an orthonormal construction $F_{\eta,\nu}$ encoding square M-QAM^2 that, asymptotically in the SNR, maximizes the information rate for $\gamma \ll 1$ 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 distribution of the 32 edges of the fundamental 4D hypercube for $\gamma \ll 1$. Up to the original constellation power scaling $\mathcal{E}_{4,M}$, the smallest minimum distance is achieved in one of the four basis directions as $\sqrt{1-\gamma}$ and the second smallest in the remaining 3 basis directions as $\sqrt{1+\gamma/3}$.*

This geometric construction is justified in two consecutive steps presented thereafter.

Preservation of the Minimum of $d_{\min}^{(1)}$ after Scaling

First, as discussed in the previous section, we focus on the case encountered in practice for which γ deviates slightly from 0. This is justified in practice as classical (extremal) PDL values Λ are typically lower than 4dB. Otherwise, scaling considerations similar to section 4.1.2 have to be done. Hence, for $\gamma \ll 1$ 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 $\mathbb{R}^4$. 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 32 smallest minimum distances (equi-distributed on the 4 orthogonal edge directions). Therefore, it suffices to study the mapping of each of the 4 elementary basis vectors of $\mathbb{R}^4$ after the action of a fixed $\hat{F}_{\eta,\nu}$ left-multiplied by the action of a random $U \in \mathrm{SU}(2)$. We now show that there is always one of the 4 directions that will experience the worst contraction, meaning that the (worst) minimum distance scales as $\sqrt{1-\gamma}$. Therefore, this distance can be encountered not less than $32/4$ dimensions $= 8$ times. 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 $\hat{F}_{\eta,\nu}$ and any $e_i \in \mathcal{B}$, there are $e_j \in \mathcal{B}$ and $U = U(e_j, e_i) \in SU(2)$ such that $\hat{U}\hat{F}_{\eta,\nu} e_j = e_i$, $\hat{U}\hat{F}_{\eta,\nu} span(\mathcal{B} \setminus \{e_j\}) = span(\mathcal{B} \setminus \{e_i\})$, and the submatrix formed by all but the i-th row and the j-th column is $(\hat{U}\hat{F}_{\eta,\nu})_{\sim i, \sim j} \in SO(3)$.*

Proof 1 *Without loss of generality, let us restrict the proof to* $e_1 = (1,0,0,0)^T$. *Then, for any* $\hat{F}_{\eta,\nu}$, $\hat{F}_{\eta,\nu}e_1 = (\cos(\eta), 0, \sin(\eta)\cos(\nu), \sin(\eta)\sin(\nu))^T$. *Setting* $\hat{U}^* = \hat{R}_{\frac{\pi}{2}-\eta}\widehat{\exp(-i\frac{\nu}{2})}\hat{B}_{\nu/2}$ *gives the result as* $\hat{U}^*\hat{F}_{\eta,\nu}e_1 = (0,0,1,0)^T = e_3$.

Now, consider the original discrete M-PAM^4 signal points in $\mathbb{R}^4$ and a given $\hat{F}_{\eta,\nu}$. 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). Notice that, for sufficiently small γ, continuity arguments show that the minimum distance is supported by (some of) the edges of the distorted four-dimensional hypercube that gets $\hat{U}$-rotated in one of the two $\sqrt{1-\gamma}$-compression directions. When $\gamma = 0$, it is encountered $4 \times 2^{4-1} = 32$ times. When $\gamma > 0$, the non-zero scaling in certain directions modifies the profile of the minimum distances and their multiplicity order. For $\gamma > 0$, depending on uniquely U and $f_{\eta,\nu}$, the minimum distance appears at least 8 times.

Corollary 1 *For small PDL value* $\gamma \ll 1$, *the randomly distributed* $U \in SU(2)$ *makes that the* worst *minimum Euclidean distance is* $d_{min} = \frac{2}{\sqrt{\mathcal{E}_{M,2}}}\sqrt{1-\gamma}$, *where* $\frac{2}{\sqrt{\mathcal{E}_{M,2}}}$ *is the normalized length of the double-unit hypercube edge. In the code distance distribution,* d_{min} *is said to scale as* $\sqrt{1-\gamma}$, *it is encountered at least 8 times.*

Proof 2 *Geometric alignment induces the maximal reduction in edge norm. This alignment can always be achieved as shown in Lemma 1. Among the 32 edges of the hypercube, 8 are then necessarily in one given (contracting) direction. Hence the distance is achieved at least 8 times.*

Second Minimum Distance $d^{(2)}_{\min}$ Analytic Optimization

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, we now want to show that the sum of the squared distances observed for the 3 remaining axes scales as $3+\gamma$ 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. 4.10. An optimal placing in the three-dimensional space is achieved when the equilateral triangle base of the tetrahedron formed by the 3 remaining basis vectors becomes perpendicular to the contracting direction (z-axis in Fig. 4.10). We now propose to provide the exact angles for one particular such rotation.

Lemma 1 shows that there is at least one pair (α^*,β^*) that aligns any SO (4)-transform of an

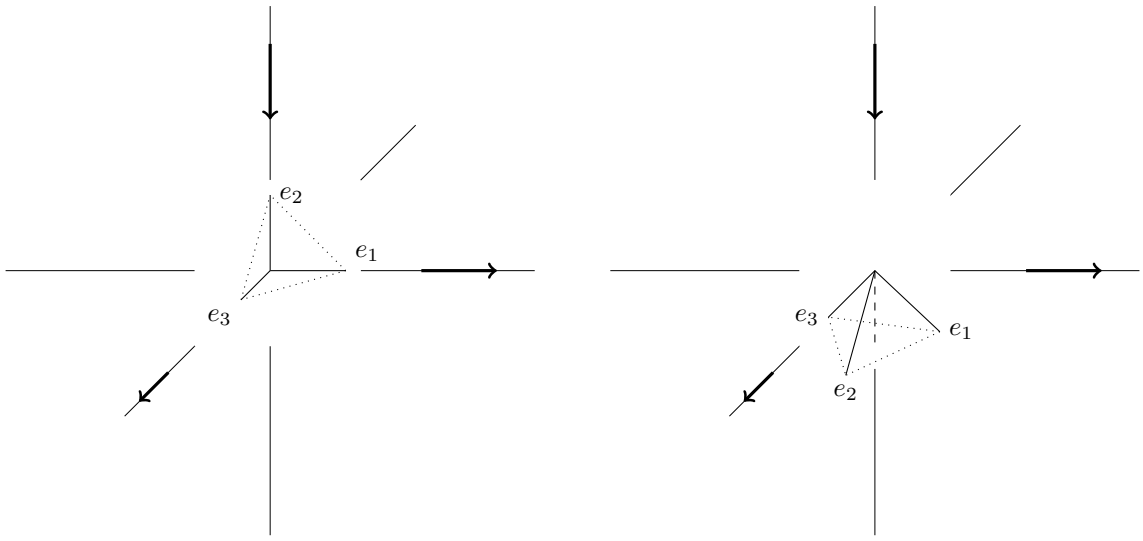


Figure 4.10 – Left: No precoding is performed. Right: Because there is always one basis vector that gets aligned with one of the contracting directions and get scaled (contracted) by $\sqrt{1-\gamma}$, the degrees of freedom of $F_{\eta,\nu}$ are used to orientate the remaining 3 basis vectors in such a way that they experience the smallest possible contraction.

hypercube edge with one of the two contracting directions e_3, e_4 by left multiplication by $U^* = U(\alpha^*, \beta^*)$. The associated distance, obtained after left multiplication by $\hat{D}_\gamma \hat{U}^*$, gives the minimum distance with maximal contraction scaling as $\sqrt{1-\gamma}$. It remains to choose η, ν that maximize the second minimum distance. This can be done by simple calculus using matrices that arise from Lemma 1. Without loss of generality, let us for instance further use

$$\hat{U}^*(\eta,\nu)\hat{F}_{\eta,\nu} = \begin{pmatrix} 0 & 0 & -\cos(\nu) & -\sin(\nu) \\ 0 & \sin(2\eta) & \cos(2\eta)\sin(\nu) & -\cos(2\eta)\cos(\nu) \\ 1 & 0 & 0 & 0 \\ 0 & \cos(2\eta) & -\sin(2\eta)\sin(\nu) & \sin(2\eta)\cos(\nu) \end{pmatrix} = M(\eta,\nu) \tag{4.22}$$

In the matrix $M(\eta, \nu)$, after multiplication by the gain imbalanced matrix $\hat{D}_\gamma$, it is clear that one axis is aligned to one of the maximum compression direction e_3. A further remarkable property is that, for this particular orientation U^*, the remaining e_i project on only one compressing direction, instead of the two. Rotating the remaining 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 $\mathrm{SO}(3)$. We want to find

$$\max_{\eta,\nu} \left\{ \min\{ \left\| \hat{D}_\gamma \hat{U}^*(\eta,\nu)\hat{F}_{\eta,\nu}e_2 \right\|, \left\| \hat{D}_\gamma \hat{U}^*(\eta,\nu)\hat{F}_{\eta,\nu}e_3 \right\|, \left\| \hat{D}_\gamma \hat{U}^*(\eta,\nu)\hat{F}_{\eta,\nu}e_4 \right\| \} \right\} \tag{4.23}$$

For any $i \in \{2, 3, 4\}$, we have

$$\left\| \hat{D}_\gamma \hat{U}^*(\eta,\nu)\hat{F}_{\eta,\nu}e_i \right\|^2 = (1+\gamma) \left\| d_i(\eta,\nu) \right\|^2 + (1-\gamma) \left\| a_i(\eta,\nu) \right\|^2 \tag{4.24}$$

$$= (1+\gamma)(1 - \left\| a_i(\eta,\nu) \right\|^2) + (1-\gamma) \left\| a_i(\eta,\nu) \right\|^2 \tag{4.25}$$

where $d_i = d_i(\eta, \nu) \in \mathrm{span}(e_1, e_2)$ and $a_i = a_i(\eta, \nu) \in \mathrm{span}(e_4)$ are orthogonal projections of the unit-norm vector $\hat{U}^*(\eta, \nu)\hat{F}_{\eta,\nu}e_i$. A key observation follows. Maximizing over η, ν any of the 3 distances $\left\| \hat{D}_\gamma \hat{U}^*(\eta,\nu)\hat{F}_{\eta,\nu}e_i \right\|$ is equivalent to minimizing each of $|a_i(\eta, \nu)|^2$ over η, ν. Moreover, comparing the 3 distances $\left\| \hat{D}_\gamma \hat{U}^*(\eta,\nu)\hat{F}_{\eta,\nu}e_i \right\|$ to each other reduces to ordering the 3 values $\{|a_i(\eta, \nu)|\}_{i\in\{2,3,4\}}$. Reading the fourth row of $M(\eta, \nu)$, we have

$$|a_2|^2 = \cos(2\eta)^2 \tag{4.26}$$

$$|a_3|^2 = \sin(2\eta)^2 \sin(\nu)^2 \tag{4.27}$$

$$|a_4|^2 = \sin(2\eta)^2 \cos(\nu)^2 \tag{4.28}$$

We want the three distances of Eq. 4.25 for all $i \in \{2,3,4\}$ to be *equal*

$$\sum_{i\in\{2,3,4\}} \left\| \hat{D}_\gamma \hat{U}^*(\eta,\nu)\hat{F}_{\eta,\nu} e_i \right\|^2 = 3d^2 \tag{4.29}$$

Otherwise, if they are not equal, then one of the distance is necessarily smaller and could have been even more optimized. On the other hand, because $\sum_{i\in\{2,3,4\}} |a_i|^2 = 1$ we have

$$\sum_{i\in\{2,3,4\}} \left\| \hat{D}_\gamma \hat{U}^*(\eta,\nu)\hat{F}_{\eta,\nu} e_i \right\|^2 = 3 + \gamma, \tag{4.30}$$

By the equality of Eq. 4.29 and 4.30, we find that $d^2 = 1 + \gamma/3$. 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 second minimum distance is optimized for

$$\nu^* = \frac{\pi}{4} \tag{4.31}$$

$$\eta^* = \frac{1}{2}\cos^{-1}\left(\frac{1}{\sqrt{3}}\right) = \frac{1}{4}\cos^{-1}\left(-\frac{1}{3}\right) = \frac{1}{2}\tan^{-1}(\sqrt{2}) \tag{4.32}$$

that gives $\|a_2\|^2 = \|a_3\|^2 = \|a_4\|^2 = \frac{1}{3}$. Different values of η or ν (other than the ones indicated by symmetry and parity) would result in one of the $\|a_i\|$ becoming larger. An alternative analysis of the variations of ν and η is reported in Appendix E.1. Hence, their projections onto the compressing axis would not have been minimized in the three $\|a_i\|$ are not equal. Those values match the ones found in the numerical optimization from section 4.2.2, it is exactly equal for ν and the found $\eta \approx 5\pi/32$ was the closest considering the $\pi/64$ rounding .

As a corollary we also saw that the 24 second minimum distances scale as $\sqrt{1+\gamma/3}$ and the original unit quaternion matrix M evaluated in η^*, ν^* is

$$M(\eta^*,\nu^*) = \begin{pmatrix} 0 & 0 & -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ 0 & \sqrt{\frac{2}{3}} & \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} \\ 1 & 0 & 0 & 0 \\ 0 & \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{pmatrix} \tag{4.33}$$

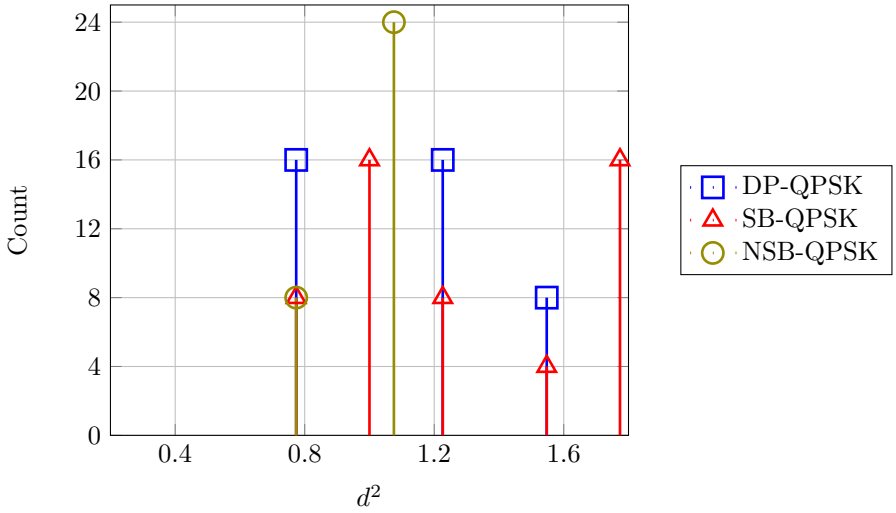


Figure 4.11 – Squared distance profile for different polarization coding schemes and $\Lambda = 2$dB. With no PDL (not represented), the 32 smallest distances, corresponding to the edges of the hypercube in four dimensions, would all be at the same value 1 for the three schemes.

Geometric View

Recall that the angle in Theorem 1 is exactly equal to 4 times the so-called *tetrahedral angle* and, more precisely, the coordinate system

$$\tilde{M}(\eta^*, \nu^*) = \begin{pmatrix} 0 & -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \sqrt{\frac{2}{3}} & \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{pmatrix} \in \mathrm{SO}(3) \tag{4.34}$$

represents the 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. It has a physical interpretation using the normalized sphere centered at the tetrahedron center O. In Appendix E.2, several tetrahedrons defined by the rotated basis are represented, for the original DP scheme (no rotation), the SB signaling (one rotation with angle η) and one with the NSB signaling (two rotations). As previously said, with the NSB the base triangle defined by the basis is exactly orthogonal to the PDL compression axis, then optimizing the second minimum distance after propagation in the PDL channel.

In Fig. 4.11, we represent the squared distances profile of the smallest distances after PDL scaling of the NSB-QPSK for a PDL of $\Lambda = 2$dB and at the worst SOP U^*, as well as the ones of the DP- and the SB-QPSK. First, observe that the NSB and SB signalings, as SO (4) encoded constellations, have a number of 8 occurrences of their minimum distance as predicted from this section. While

the minimum distance itself is not improved, its occurrences divided by 2, compared to DP-QPSK, enables to increase the information rate at the worst SOP. With a unit energy constellation, the minimum squared distance here is equal to $1-\gamma \approx 1-0.22 \approx 0.77$. Besides the 8 minimum distances, observe then that, by opposition to the SB signaling, the NSB-QPSK $32-8=24$ second minimum distances are all equal. Their value is as predicted $1+\gamma/3 \approx 1.07$. This explains why the NSB-QPSK is operationally better performing that the SB-QPSK and justifies why the worst rate point is at the top in the 3D Fig. 4.9.

4.3 Signalings on Several Timeslots: Unlocking the Minimum Distance Bound

In this section, we illustrate the performance of the Silver code [34, 36] on the lumped PDL channel. This code is a 2-timeslot, full-rate, space-time code that mixes 8 components that are the in-phase and quadrature phase of two polarizations and on two timeslots. By opposition to the Golden code [66, 65, 36], the Silver code cannot be written as a linear encoding in the complex field. Instead, its encoding matrix separates the complex parts of the 2-timeslot input vector.

According to [36], the full-rate Silver code is a transformation of two consecutive DP-QAM symbols $X[1]$ and $X[2]$ based on a modified structure of the Alamouti code [63]

$$f(X) = \frac{1}{\sqrt{2}}\begin{pmatrix}\chi_1 & -\chi_2^* \\ \chi_2 & \chi_1^*\end{pmatrix} + \frac{1}{\sqrt{2}}\begin{pmatrix}1 & 0 \\ 0 & -1\end{pmatrix}\begin{pmatrix}\xi_1 & -\xi_2^* \\ \xi_2 & \xi_1^*\end{pmatrix} \tag{4.35}$$

where $(\chi_1, \chi_2)^T$ is exactly the first timeslot input $(x_1[1], x_2[1])^T = X[1]^T$ and $\Xi = (\xi_1, \xi_2)^T$ is a unitary transform of the second timeslot input

$$\Xi = \frac{1}{\sqrt{7}}\begin{pmatrix}1+i & -1+2i \\ 1+2i & 1-i\end{pmatrix} X[2] \tag{4.36}$$

Defining $\hat{x}_j[k] = (\Re x_j[k], \Im x_j[k])^T$, Eq. 4.35 can be recast in the real field to the block encoding matrix F as

$$\hat{F}\hat{X} = \frac{1}{\sqrt{2}}\left(\begin{array}{cc|c} \begin{pmatrix}I_2 & 0 \\ 0 & I_2\end{pmatrix} & & G_1 \\ \hline \begin{pmatrix}0 & -S_0 \\ S_0 & 0\end{pmatrix} & & G_2 \end{array}\right)\begin{pmatrix}\hat{x}_1[1] \\ \hat{x}_2[1] \\ \hat{x}_1[2] \\ \hat{x}_2[2]\end{pmatrix} \tag{4.37}$$

where

$$G_1 = \begin{pmatrix} R_{-\pi/2} & 0 \\ 0 & R_{-\pi/2} \end{pmatrix} \begin{pmatrix} \cos(\frac{\pi}{4})R_{-\frac{\pi}{2}} & -\sin(\frac{\pi}{4})R_{-(\arctan(2)+\frac{\pi}{2})} \\ \sin(\frac{\pi}{4})R_{\arctan(2)+\frac{\pi}{2}} & \cos(\frac{\pi}{4})R_{\frac{\pi}{2}} \end{pmatrix} \tag{4.38}$$

$$G_2 = \begin{pmatrix} 0 & -S_{-\frac{\pi}{2}} \\ S_{\frac{\pi}{2}} & 0 \end{pmatrix} \begin{pmatrix} \cos(\frac{\pi}{4})R_{-\frac{\pi}{2}} & -\sin(\frac{\pi}{4})R_{-(\arctan(2)+\frac{\pi}{2})} \\ \sin(\frac{\pi}{4})R_{\arctan(2)+\frac{\pi}{2}} & \cos(\frac{\pi}{4})R_{\frac{\pi}{2}} \end{pmatrix} \tag{4.39}$$

Observe that the four matrices appearing in Eq. 4.37 are all in SO (4). The matrices G_1 and G_2 have been factorized to reveal their left- and right-isoclinic part, specifically appearing in that order. The Silver code appears to be a particular solution of a more general SO (8) encoding framework that would involve $8 \times 7/2 = 28$ optimization parameters. This optimization requiring the analysis of SO (8) group structure is out of the scope of this thesis and is left for a future study. We focus in this section of an analysis of the Silver code rate and explain its good performance with Euclidean distances considerations.

The information rate of the Silver-encoded QPSK modulation scheme is represented in Fig. 4.12 as a function of the two channel angle parameters. We display also in the same figure the information rate of the previously derived NSB signaling in order to compare the two. The profile of the DP-QPSK is plotted as a reference amplitude scale. While the NSB signaling shows an increase of the rate for a significant number of angle pairs, it is locked to around 70% of the DP-QPSK rate range. By construction the minimum rate of the NSB signaling over the grid is the maximum worst rate of all possible unitary 4D transforms as demonstrated in section 4.2.3. The Silver code on the other hand represents the best-known upper limit of achievable information rate in a PDL channel and outperforms the NSB signaling. Not only it significantly elevates the worst case rate, it appears to be a lot less sensitive to the incident polarization orientation, the rate profile being almost a plateau.

A key observation explains this. In Fig 4.13, we represent $d_{\min}(U) = \min_{A \neq B} d(A, B, \gamma, U)$ the minimum Euclidean distance between 2 points A and B in the codebook after PDL scaling for a given $\Lambda = 6$dB. Observe that the minimum distance itself is increased for most of the channel SOPs $U \in \mathrm{SU}(2)$. In particular, and by opposition to SO (4) encodings on 1-timeslot, the minimum over the different U of the minimum Euclidean distances $\min_{U \in \mathrm{SU}(2)} d_{\min}(U)$ *is* increased. This increase of minimum distance is a radical difference compared to SO (4) transforms, and is only possible when encoding on additional degrees of freedom, with here two timeslots. Likewise, in Fig. 4.14 the same figure is represented this time for a PDL of $\Lambda = 2$dB. Observe that for lower PDL values, the profile of the DP-QPSK minimum distance can be summarized to a 2D figure only as there is no more variation in the angle β. Similar conclusions hold for the information rates, not represented here. This is what justifies the use of the simple PDL channel with only one angle parameter.

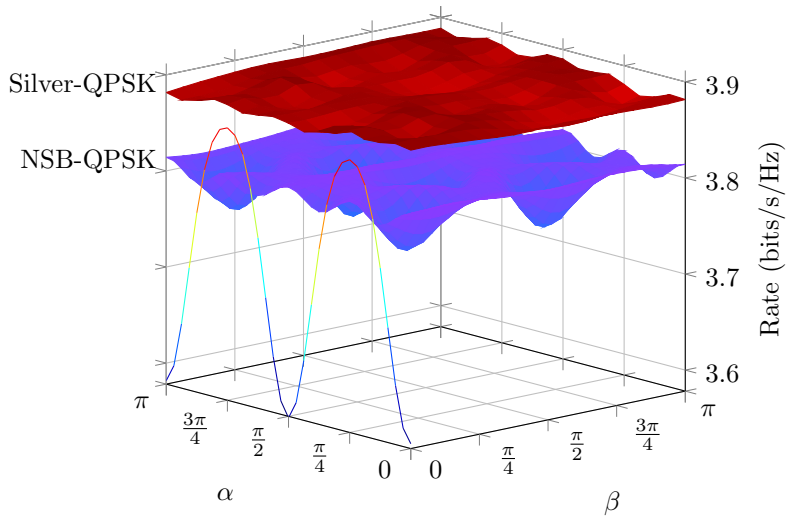


Figure 4.12 – NSB- and Silver-QPSK information rates at $\rho = 8$dB on the lumped PDL channel with $\Lambda = 6$dB for various angles α and β. The line represents the amplitude range of the DP-QPSK rates.

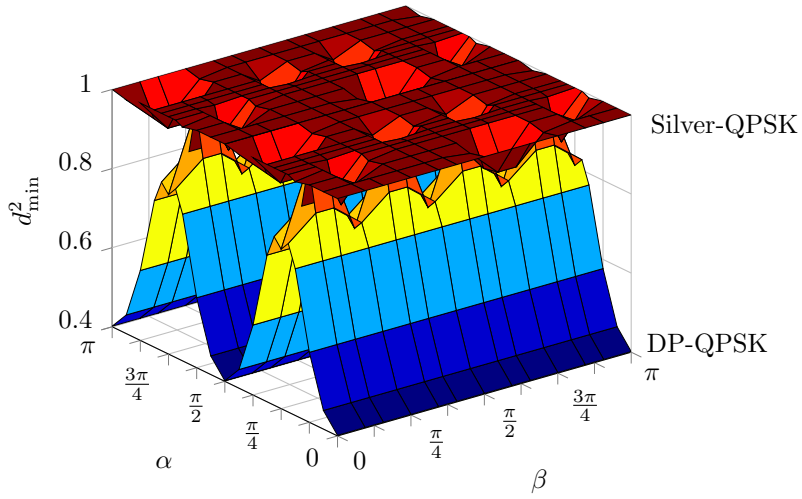


Figure 4.13 – Silver-QPSK minimum distance after PDL scaling with $\Lambda = 6$dB as a function of the angles α and β restricted in the range $[0, \pi]$ compared to the one of DP-QPSK (on 2 or 1 timeslot, the two being equivalent). The complete profile on $[0, 2\pi] \times [0, 2\pi]$ is replicas from this grid.

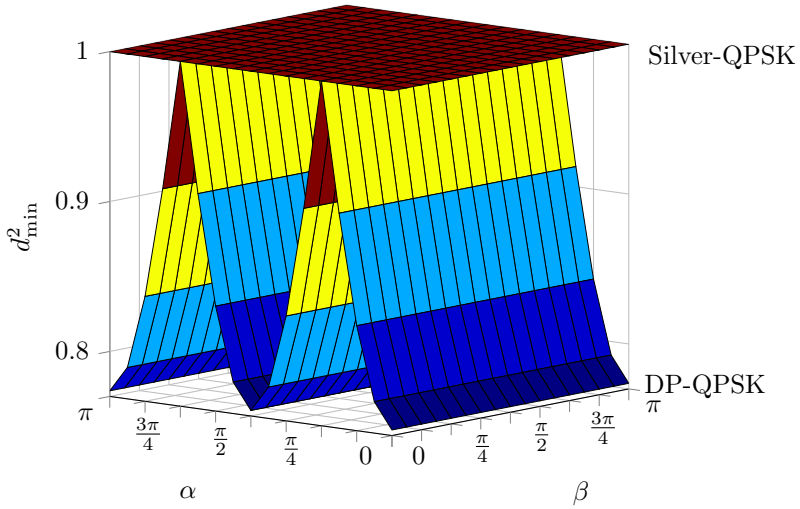


Figure 4.14 – Silver-QPSK minimum distance after PDL scaling with $\Lambda = 2$dB as a function of the angles α and β restricted in the range $[0, \pi]$ compared to the one of DP-QPSK (on 2 or 1 timeslot, the two being equivalent). The complete profile on $[0, 2\pi] \times [0, 2\pi]$ is replicas from this grid.

Finally, for low PDL values, the minimum distance of the Silver encoded QPSK becomes a variable independent of the input SOP resulting in a flat plane in Fig. 4.14. This is not true for $\Lambda = 6$dB where periodic depressions break the flat plane for certain regions of SOPs.

4.4 Summary on PDL-Resilient Modulation Schemes

In this chapter, we focused on the lumped PDL channel model and introduced the principles of PDL-resilience as a four-dimensional modulation with spatial balancing. After having proposed the simple Spatially Balanced for the simple PDL channel, we derived the optimal orthogonal PDL-resilient code over one time slot, which we called New Spatially-Balanced signaling. Finally, we discussed how the new modulation schemes compare with schemes spreading over several time slots.

Notice that directional gains from orthogonal polarization codes occur when the modulation constellation is supported by an underlying cubic lattice. Such *coding* gains naturally and efficiently come in addition to *shaping* gains obtained, e.g., from non-uniform signaling [74, 75, 76, 11, 77].

While performance can be improved by using several time, space or frequency slots, the increased computational complexity per slot use becomes almost prohibitive when compared to the optimal SB or NSB signalings. Because of their reduced complexity, in the following chapter we study the performance of the two signalings that we developed on an actual optical link, both in simulation

and in a lab experiment.

Chapter 5

Practical Validation of PDL-Resilient Signaling

We now propose to validate the performance of the SB and NSB signalings, first by a simulation on an optical link including distributed PDL elements with a rate outage analysis. We then report experimental measurements on both a single (lumped) PDL element and on a link with distributed elements after developing the required signal processing specifically needed to decode our SO (4) encoded schemes.

5.1 Outage Comparison Between Signalings on a Distributed PDL Link

We study the performance over the PDL channel modeled as a distributed optical link with inserted PDL elements and distributed noise. We first present an impairing effect due to EDFAs in a distributed link and analyze the performance of several modulation schemes in terms of verification of an outage condition.

5.1.1 SNR Degradation with EDFAs in Constant Output Power Mode

In optical links EDFAs are used inline in order to boost the signal that has lost energy along its propagation in a 40 to 100 kilometers span. To model it, we consider two modes, either EDFAs operating with a fixed *gain* (abbreviated G), or with a fixed *Constant Output Power* (COP). In

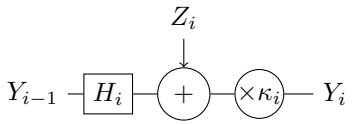


Figure 5.1 – Span representation when EDFAs are in constant output power mode with κ_i a energy rescaling factor to meet identical energy at each span output.

this section we study the difference of behavior between the two.

When designing a system, the number of spans the signal will go through is to take into account. Namely, in a submarine cable for instance, the link between the transmitter and the receiver is constituted of a number of spans n, it means the signal will propagate on a long fiber distance and then re-amplified n times. Assuming each amplifier produces noise according to a Gaussian process $\mathcal{CN}(0, I_2)$, in order to target a given SNR ρ^* at the receiver, the energy at the transmitter should be of

$$P_X = 2n\rho^* \tag{5.1}$$

The targeted SNR cannot be guaranteed as the PDL elements may either amplify or attenuate the signal depending on the PDL elements random orientation at the input of the element. Indeed, the energy modifying matrix H_i makes the effective output energy dynamic and impossible to exactly predict when designing a system. Then, in what follows we distinguish when applicable between the notations of the target SNR ρ^* and the effective SNR ρ.

Amplifiers in COP mode on the other hand make sure that the energy at the output of the amplification at each span is set to a fixed value P_0. In practice, $P_0 = P_X$, the energy of the signal at the transmitter and $P_{Z_i} = 2$ from the (uncorrelated) contribution of unit variance noise from the two polarizations. In Fig. 5.1 the modeling of an amplifier with constant output power is represented. Again, we omitted here the (static) amplification gain the amplifier is designed for: we assume it amplifies the input signal Y_{i-1} to compensate for the total energy loss it has been subject to in span $i-1$. The transfer matrix H_i captures the effects of the current span i including its orientation but excluding the total energy loss and compensation just described. The scaling factor κ_i is meant to rescale the output power of the amplifier to a set value. This scaling factor is expressed as

$$\kappa_i = \sqrt{\frac{P_0}{P_{H_i Y_{i-1}} + P_{Z_i}}} \tag{5.2}$$

the term 2 coming from the noise energy added by the amplifier and $P_{H_i Y_{i-1}}$ the energy of the signal Y_{i-1} coming from the previous span and modified by the current span matrix H_i.

Assume now that we have a link with n spans that does not contain any PDL. Then the scaling factor is a constant $\kappa = \sqrt{P_0/(P_0+2)}$ in the case of a PolMux transmission. At the end of the link

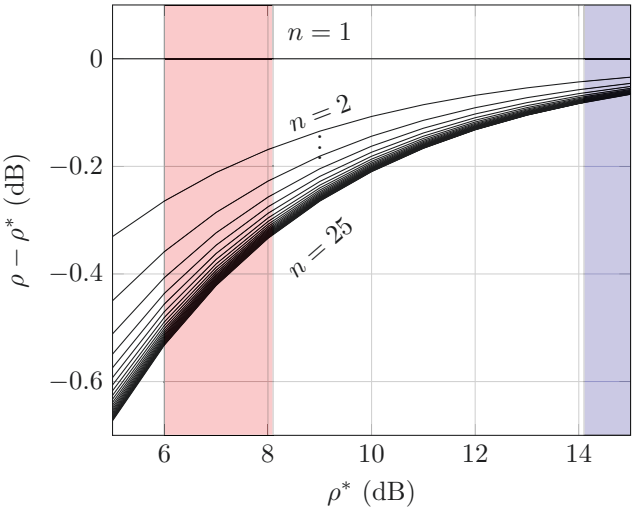


Figure 5.2 – SNR degradation with EDFAs in COP mode with QPSK (red) and 16QAM (blue) operating regions

the energy of the signal HX propagated through the equivalent channel H and the total noise Z are

$$P_{HX} = (\kappa^2)^n P_0 \text{ and } P_Z = 2\sum_{i=1}^{n}(\kappa^2)^i \tag{5.3}$$

In Fig. 5.2 the SNR degradation $\rho^* - \rho$ (also known as the droop effect [78, 79]) is depicted for various numbers n of spans with no PDL as a function of the targeted SNR ρ^* in the sense of formula 5.1. The value of ρ is the actual SNR defined as the ratio between expression 5.3. In superposition the areas of operation of the QPSK and 16-QAM are represented with an example targeted pre-bit decoding BER of 10^{-2}. Observe that the QPSK scheme is more impacted by the COP mode of the amplifier and requires additional energy at the transmitter to take into account the SNR degradation shown in the figure. Even with no PDL, the constant output power of the EDFA degrades the effective SNR at the receiver.

5.1.2 An Outage Condition Study on the Information Rate

We now propose to compare the performance of the channel in terms of verified outage condition. From an optical system operator's perspective, because a distributed channel is random, it may be impossible to guarantee information transmission in 100% of the time. This is depicted in Fig. 5.3 with the rate distribution of 5000 distributed optical links with $n = 15$ randomly oriented PDL elements modeled with a gain imbalance of $\Lambda = 1.1$dB of PDL and EDFAs in gain mode using

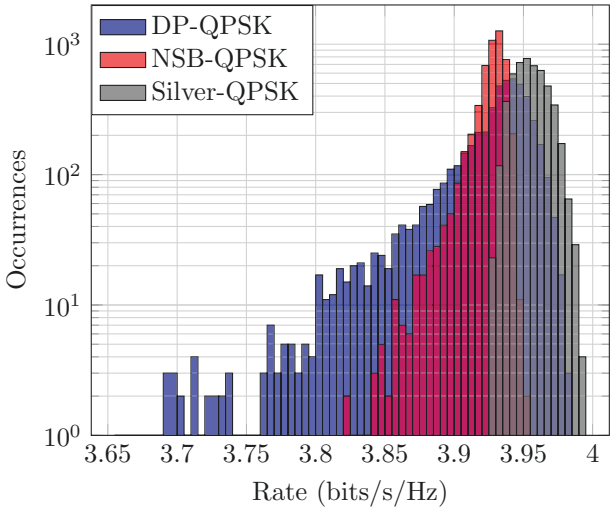


Figure 5.3 – Distributed channel impact on the information rates for 5000 randomly distributed channels with 15 spans of elementary PDL 1.1dB at SNR 9.2dB with EDFAs in COP mode.

DP-, NSB-, and Silver-encoded QPSK at the same SNR $\rho^* = 9.2$dB. As can be expected, the rate is a random variable. A rate realization depends on the equivalent resulting channel gain imbalance as well as the signal incident orientation to it. Rates depend also on the total noise construction in the channel propagation. The distribution of the rates with the PDL model 1 with a parameter ϵ presented in section 2.2 or with EDFAs in COP mode are similar but with a shifted mean value. Observe that the NSB-QPSK (red) is as desired less impacted by the PDL channel with the minimum rate realization 0.13 bits/s/Hz greater than the one of the DP-QPSK (blue). The Silver-QPSK (gray) minimum rate is even larger, and its rate distribution is significantly more narrow.

Based on this unpredictable characteristic of the information rate over a distributed PDL channel, it is useful to define a system outage probability condition and seek criteria that enable to verify them. In this work, we present the following outage condition based on information rates.

Definition 6 *A system is said to verify the outage condition if the rates I below a given threshold value $I_{\text{th.}}$ occurs with a probability of at most a*

$$P(I < I_{th.}) \leq a \tag{5.4}$$

For n given PDL elements i with PDL value Λ_i, the maximum value of PDL of the resulting channel is $\Lambda = n \sum_i \Lambda_i$. Considering this worst PDL value and the minimum rate over all SOPs (typically in $(\alpha, \beta) = (0, 0)$ for DP-QAM), one could design a system at an SNR that guarantees

rates to always be above a given $I_{\text{th.}}$. This is however too strict, the distributed channel balancing the occurrence of the worst capacities. Indeed, a perfect alignment of all the PDL elements leading to the summation of their imbalance is very unlikely. Furthermore, and as explained in section 2.5.1, the resulting SOP of the channel makes the equivalent channel angle aligned with a very low probability.

We compute the rate over 10^5 link realizations with elementary PDL 0.6dB at n spans between of 0 and 25 by increment of 5. We then explore RMS PDL values up to 3dB. We test three different modulations: the regular DP-QPSK, the proposed NSB-QPSK, and the Silver-QPSK on 2 timeslots. We seek the lowest SNR that verifies the outage condition 5.4 with $a = 10^{-3}$ and $I_{\text{th.}} = 3.6$ for the three modulations. This particular rate threshold value corresponds to the commonly-used coding rate of 0.9 in optical communications. Eventually, we use the PDL model 2 with the γ parameter, EDFAs are in COP mode and an ideal, maximum-likelihood, equalization and demapping is assumed at the receiver when computing the mutual information.

Fig. 5.4 displays the required target SNR at the receiver in order to respect the outage condition as the RMS PDL increases. For a given n, this is equivalent to increasing the launch energy at the transmitter. The COP mode here guarantees the output signal Y to be at a set value. With no PDL, the channel is just an SOP rotation and the three modulations check the outage condition for the same SNR value. This can be seen in Fig. 5.4 as the three modulation curves originate from the same point at $n = 0$. When PDL is added, the outage condition validation is met for a higher SNR and differs depending on the modulation. Besides the addition of 5 PDL elements of 0.6dB, the steep increase from $n = 0$ to 5 may be understood from a sudden drop in the effective SNR according to Fig. 5.1 as DP-QPSK operating regions are strongly impacted by the EDFAs COP mode. For higher operating SNR regions, this impact would be reduced. We also note that this additional penalty is not seen when amplifiers operate in a constant gain mode. This loss of SNR is translated in a global decrease of the information rates hence the requirement for even additional energy at the transmitter. Observe that the Silver-QPSK as expected is the best modulation of the three tested and gives the highest SNR relaxation of 0.4dB compared to DP-QPSK for 3dB of RMS PDL ($n = 25$). At this same PDL value, the NSB-QPSK enables a relaxation of 0.2dB of required SNR but for the same decoding complexity as the DP-QPSK. The gap between the DP-QPSK and the other two modulations tends to increase as the number of span increases which shows their interest to be used in practice for an enhanced PDL-resilience in optical links.

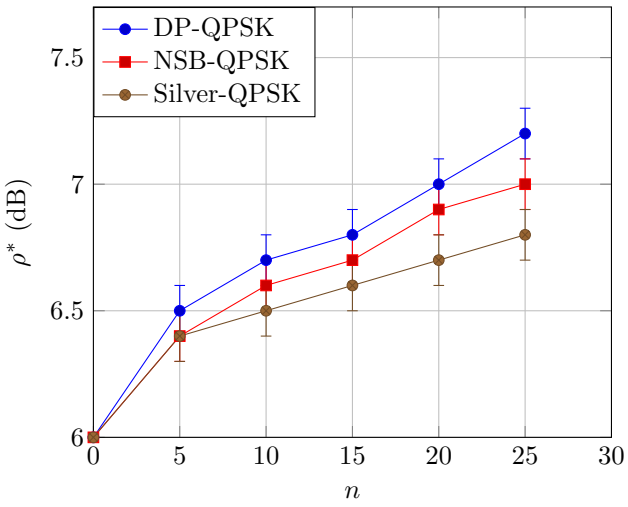


Figure 5.4 – Outage verification for n spans for PDL elements of 0.6dB (PDL model 2 and in COP mode), for DP-, NSB- and Silver-encoded QPSK.

5.2 Digital Signal Processing for Polarization-Encoded Modulations

The presented modulations in previous chapters are sophisticated compared to regular QAM and present a challenge to be implemented in practice. By construction, the SB and NSB signalings are modulations where information is jointly encoded on the four dimensions of the signal. In other words, the information on one polarization is dependent on the other. In this section we develop the specific signal processing required to recover our symbols after transmission in the channel. The remaining chromatic dispersion correction is not detailed as it just corresponds to a deterministic unitary effect that can be properly removed.

5.2.1 Channel Estimation with Dedicated Pilot Sequences

The major change required to implement the jointly encoded modulations is the channel equalization, ordinarily performed by a Constant Modulus Algorithm (CMA) equalization module [80]. This blind algorithm converging to a Linear-Minimum Mean Square Error (L-MMSE or MMSE) equalizer does not require prior knowledge of the channel as long as the input modulation verifies some conditions. One of them assumes symbols to be of unit energy, or possibly have several energy radii when using the upgraded Multi Modulus Algorithm [81]. More importantly, the spatially-multiplexed input is supposed to carry independent streams on each spatial dimension. This is

not the case of our modulations. Furthermore, even with the original DP-QAM constellation, the presence of PDL may cause the CMA equalizer block converge to a singularity because of the weak polarization being hidden by the strong one, and resulting in a filter outputting only one of the tributaries. When designing a transmission link in the presence of PDL, this issue has to be taken into account and methods to escape from this singularity should be implemented. We consequently propose an alternative equalization scheme. In order to estimate the channel to be able to equalize it, we decide to place regular pilots in each symbol frame. In this section, we first elaborate in a pedagogical manner on the used channel estimation method in the general case of $n_r \times n_t$-MIMO channel with memory l_h.

Modeling

We first precise the notations and concepts for a memory-less channel that has been used in this document until this section. We then extrapolate to channels with memory.

Memoryless Channel A memory-less $n_r \times n_t$ possibly MIMO channel is modeled as

$$Y = HX + Z \tag{5.5}$$

with $n_t > 0$, $n_r > 0$, $X \in \mathbb{C}^{n_t \times 1}$, $Y, Z \in \mathbb{C}^{n_r \times 1}$ and $H \in \mathbb{C}^{n_r \times n_t}$.

When sending n symbols over the same channel H, Eq. 5.5 remains true but with $X \in \mathbb{C}^{n_t \times n}$, Y and Z in $\mathbb{C}^{n_r \times n}$ (and still $H \in \mathbb{C}^{n_r \times n_t}$).

$$\begin{pmatrix} y_1[1] & \dots & y_1[n] \\ \vdots & \ddots & \\ y_{n_r}[1] & & y_{n_r}[n] \end{pmatrix} = \begin{pmatrix} h_{11} & \dots & h_{1n_t} \\ \vdots & \ddots & \\ h_{n_r 1} & & h_{n_r n_t} \end{pmatrix} \begin{pmatrix} x_1[1] & \dots & x_1[n] \\ \vdots & \ddots & \\ x_{n_t}[1] & & x_{n_t}[n] \end{pmatrix} + \begin{pmatrix} z_1[1] & \dots & z_1[n] \\ \vdots & \ddots & \\ z_{n_r}[1] & & z_{n_r}[n] \end{pmatrix} \tag{5.6}$$

Notation $X[k] = (x_1[k], \dots, x_{n_t}[k])^T$ indicates the k-th element of a sent sequence $X \in \mathbb{C}^{n_t \times n}$.

SISO Channel with Memory For any integer symbol-time $k > 0$, the output of a SISO channel with memory $l_h \geq 0$ can be modeled by

$$y[k] = (h * x)[k] + z[k] = \sum_{i=0}^{l_h} h_i x[k-i] + z[k] \tag{5.7}$$

The parameter l_h is the memory of the channel, in other word $\forall i > l_h$ we have $h_i = 0$. The

number of channel taps is then $l_h + 1$. A memory-less channel is modeled by the same equation taking $l_h = 0$.

MIMO Channel with Memory Similarly, the output of MIMO channel with n_t input dimensions and n_r output dimensions is

$$Y[k] = (H * X)[k] + Z[k] = \sum_{i=0}^{l_h} H_i X[k-i] + Z[k] \tag{5.8}$$

where for each $0 \geq i \geq l_h$ and integer k, $X[k-i] \in \mathbb{C}^{n_t \times 1}$, $Y[k], Z[k] \in \mathbb{C}^{n_r \times 1}$ and $H_i \in \mathbb{C}^{n_r \times n_t}$. Memory-less channels correspond to Eq. 5.5. In order to write the output in the sequential form 5.6, define $Y = (Y[1], \dots Y[n])$ and $Z = (Z[1], \dots Z[n])$, consider $H = (H_0, \dots, H_{l_h})$ and define $\mathbb{X}$ the Toeplitz form of the sequence $X = (X[1], \dots, X[n])$:

$$\mathbb{X} = \begin{pmatrix} X[l_h+1] & X[l_h+2] & \dots & X[n] \\ X[l_h] & X[l_h+1] & \dots & X[n-1] \\ \vdots & & & \\ X[1] & X[2] & \dots & X[n-l_h] \end{pmatrix} \tag{5.9}$$

such that we have

$$\begin{aligned} Y &= \begin{pmatrix} Y[1], & Y[2], & \dots, & Y[n] \end{pmatrix} \\ &= H\mathbb{X} + Z \\ &= \underset{(H_0, H_1, \dots, H_{l_h})}{} \begin{pmatrix} X[l_h+1] & X[l_h+2] & \dots & X[n] \\ X[l_h] & X[l_h+1] & \dots & X[n-1] \\ \vdots & & & \\ X[1] & X[2] & \dots & X[n-l_h] \end{pmatrix} + Z \end{aligned}$$

Least-square Channel Estimation Using a Training Sequence

Assuming knowledge of n_p sent symbols at the receiver (pilots), a least square estimate $\hat{H}$ of the channel is [82] $\hat{H} = Y(X^\dagger X)^{-1} X^\dagger / n_p$. It corresponds to the right-multiplication of Y by the pseudo-inverse of X in $Y = HX$. Constant Amplitude Zero Auto Correlation (CAZAC) sequences [83] $s \in \mathbb{C}^{1 \times l_c}$ of length l_c are specifically designed to check $s^\dagger s = I_{l_c}$ in a SISO configuration. Concatenating n_t circularly-shifted versions of a CAZAC sequence in the spatial dimension of X [82], and then

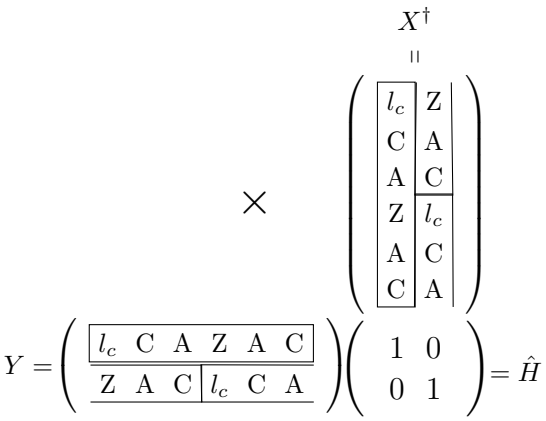


Figure 5.5 – Channel estimation product $Y\mathbb{X}^\dagger$ illustration with no crosstalk and no memory for two input and output spatial dimensions $n_t = n_r = 2$ interpreted as polarizations. The input first polarization carries a CAZAC sequence of length l_c and the second one has the same sequence shifted by $l_c/2$. The product of Y first row with $\mathbb{X}$ first column gives 1 by the CAZAC properties. The one of Y second row with the first column of $\mathbb{X}$ gives 0 because the two sequences are orthogonal. For real-life channel memory estimation, $\mathbb{X}$ has to be larger.

repeating them n_{rep} times, an estimate $\hat{H}$ of the MIMO channel H with memory as

$$\hat{H} = \frac{1}{n_p} Y\mathbb{X}^\dagger \tag{5.10}$$

with $\mathbb{X}$ defined in Eq. 5.9 taking $n = n_p = n_{\text{rep}} l_c$. The dimensions are $Y \in \mathbb{C}^{n_r \times n_p}$, $\mathbb{X}^\dagger \in \mathbb{C}^{n_p \times n_t(l_h+1)}$

For $n_t = n_r = 2$, the ideal shift of the second spatial dimension is $l_c/2$ [84]. Note that l_c should be [83] a power of 4. A CAZAC sequence of length l_c can fully recover a SISO channel with memory $l_h + 1 \leq l_c$. Fig. 5.5 exemplifies the channel estimation product in a simple 2×2 case with no memory. In general for a MIMO channel with memory, the sequence length of spatially concatenated CAZAC sequences should verify

$$l_c \geq n_t(l_h + 1) \tag{5.11}$$

That being said, the pilots sequence length should also be

- long enough to be robust to added noise
- short enough so that the channel does not vary during the etimation window (typically because of laser phase noise)

In what follows we assume that we use CAZAC sequences that are long enough for the considered number of channel taps. For notation simplification, we identify the two lengths: $l_c = n_t(l_h + 1)$. Consequently, the variable dimensions in Eq. 5.10 are

$$Y \in \mathbb{C}^{n_r \times n_{\text{rep}} l_c} \text{ and } \mathbb{X}^\dagger \in \mathbb{C}^{n_{\text{rep}} l_c \times l_c} \tag{5.12}$$

Extension to Multiple Samples per Symbol Estimation

Performing sub-symbol time channel estimation may be useful as for instance matched filtering may have not been performed at this stage of signal processing in the receiver. To do so, we fill the original sequence X with zeros between each symbols, in order to upsample the original sequence to the number of sample per symbol n_s that match the one of Y. Eq. 5.10 becomes

$$\hat{H} = \frac{1}{n_p n_s} Y \mathbb{X}^\dagger \text{ with } Y \in \mathbb{C}^{n_r \times n_s n_{\text{rep}} l_c} \text{ and } \mathbb{X}^\dagger \in \mathbb{C}^{n_s n_{\text{rep}} l_c \times n_s l_c} \tag{5.13}$$

Illustration for $n_s = 2$ Below is represented the channel estimation product for better understanding.

$$\begin{pmatrix} Y[1] & Y[1+\frac{1}{2}] & Y[2] & \dots & Y[n+\frac{1}{2}] \end{pmatrix} \begin{pmatrix} X^\dagger[l_h+1] & 0_{1\times n_r} & X^\dagger[l_h] & \dots & 0_{1\times n_r} \\ 0_{1\times n_r} & X^\dagger[l_h+1] & 0_{1\times n_r} & \dots & X^\dagger[1] \\ X^\dagger[l_h+2] & 0_{1\times n_r} & X^\dagger[l_h+1] & \dots & 0_{1\times n_r} \\ \vdots & & & & \\ 0_{1\times n_r} & X^\dagger[n] & 0_{1\times n_r} & \dots & X^\dagger[n-l_h] \end{pmatrix} \begin{pmatrix} H_0 & H_{0+\frac{1}{2}} & H_1 & \dots & H_{l_h+\frac{1}{2}} \end{pmatrix}$$

In the matrix multiplication, odd column indices of Y are multiplied by the original sequence X for odd indices of X columns capturing the 'physical' channel. Even indices of Y capture sub-baudrate transformations of the generalized channel $H_0, H_{0+\frac{1}{2}}, \dots, H_{l_h+\frac{1}{2}}$.

5.2.2 Channel MMSE Equalization

In this section we now derive the MMSE equalizer for a channel with memory in order to use it when equalizing our experimental measurements in the following section.

We need the channel transfer matrix to be written in a Toeplitz form. This matrix $\mathbb{H} \in$

$\mathbb{C}^{n_r l_w \times n_t(l_h+l_w)}$ is defined as

$$\mathbb{H} = \begin{pmatrix} H_0 & H_1 & H_2 & \dots & H_{l_h} & 0 & 0 & \dots & 0 \\ 0 & H_0 & H_1 & \dots & & H_{l_h} & 0 & \dots & 0 \\ & & \ddots & \ddots & & & & & \\ 0 & \dots & 0 & H_0 & H_1 & \dots & & & H_{l_h} \end{pmatrix} \tag{5.14}$$

In $\mathbb{H}$, each of the l_w repetitions of the $H = (H_0, \dots, H_{l_h})$ sequence is translated by n_r in the horizontal dimension and by n_t in the vertical dimension. From the lower-left triangle of zeros, the length of $\mathbb{H}$ last row is of size $n_r \times n_t(l_w - 1) + n_t(l_h + 1)$ hence the size of the Toeplitz matrix $\mathbb{H}$. From this matrix, the MMSE equalizer matrix for the SNR ρ can be derived

$$\mathbb{W} = \left(\mathbb{H}^\dagger \mathbb{H} + \frac{1}{\rho} I_{n_t(l_h+l_w)} \right)^{-1} \mathbb{H}^\dagger \tag{5.15}$$

As such the equalizer matrix $\mathbb{W} \in \mathbb{C}^{n_t(l_h+l_w) \times n_r l_w}$ is a large matrix containing copies of the desired equalizer but with unwanted edge effects. The correct equalizer has to be carefully selected as n_t rows among the $n_t(l_h + l_w)$ rows of $\mathbb{W}$, the farther possible from the edges. In other words, a suitable equalizer matrix W is a submatrix of $\mathbb{W}$ selecting the rows $\delta n_t + 1, \delta n_t + 2, \dots, \delta n_t + n_t$ with $\delta = \lfloor \frac{l_w + l_h}{2} \rfloor$. A particular attention should be paid for the inversion appearing in Eq. 5.15. Indeed, the matrix $\mathbb{H}$ may contain a lot of zeros making the matrix to be inverted bad conditioned. This is particularly true if the number of taps of the channel is identified to l_c (because typically $l_h + 1 < l_c$ meaning that $H_j \approx 0$ for $j > l_h + 1$) and also with multiple samples per symbol estimation.

Illustration with $l_h = 0$ Taking $H = H_0$ we have

$$\mathbb{H} = \begin{pmatrix} H_0 & 0 & 0 & \dots & 0 \\ 0 & H_0 & 0 & \dots & 0 \\ & & & \ddots & \\ 0 & \dots & & & H_0 \end{pmatrix} \tag{5.16}$$

and

$$\mathbb{W} = \begin{pmatrix} (H_0^\dagger H_0 + I_{n_t}/\rho)^{-1} H_0^\dagger & 0 & 0 & \dots & 0 \\ 0 & (H_0^\dagger H_0 + I_{n_t}/\rho)^{-1} H_0^\dagger & 0 & \dots & 0 \\ & & & \ddots & \\ 0 & \dots & & & (H_0^\dagger H_0 + I_{n_t}/\rho)^{-1} H_0^\dagger \end{pmatrix} \tag{5.17}$$

In that case any of the n_r block rows in $\mathbb{W}$ (including the ones with the suggested shift $\delta = \lfloor \frac{l_w+l_h}{2} \rfloor$) yields the equalizer

$$W = \begin{pmatrix} 0_{n_t \times n_r \delta} & (H_0^\dagger H_0 + I_{n_t}/\rho)^{-1} H_0^\dagger & 0_{n_t \times n_r (l_w - \delta - 1)} \end{pmatrix} \tag{5.18}$$

to be able to calculate $\hat{X} = W\mathbb{Y} \in \mathbb{C}^{n_t \times n_s}$ where $\mathbb{Y}$ is a Toeplitz version of Y defined as in Eq. 5.9.

$$\begin{pmatrix} 0_{n_t \times n_r \delta} & (H_0^\dagger H_0 + I_{n_t}/\rho)^{-1} H_0^\dagger & 0_{n_t \times n_r (l_w - \delta - 1)} \end{pmatrix} \begin{pmatrix} Y[l_h+1] & Y[l_h+2] & \dots & Y[n_s] \\ Y[l_h] & Y[l_h+1] & \dots & Y[n_s-1] \\ \vdots & & & \\ Y[1] & Y[2] & \dots & Y[n_s - l_h] \end{pmatrix} = \begin{pmatrix} \hat{X}[\delta] & \hat{X}[\delta+1] & \dots & \end{pmatrix}$$

In high SNR regimes $\rho \to \infty$, $I_{n_t}/\rho \to 0_{n_t}$ and by continuity we have the zero-forcing approximation

$$(H_0^\dagger H_0 + I_{n_t}/\rho)^{-1} H_0^\dagger \approx H_0^{-1} (H_0^\dagger)^{-1} H_0^\dagger = H_0^{-1} \tag{5.19}$$

In general though, at low to moderate SNRs, the zero-forcing equalization amplifies the noise and MMSE equalization should be preferred.

5.2.3 Optical Carrier Phase Correction

Inherent to the coherent technology [4, 5], a laser serving as a local oscillator beating with the signal has to be used. This laser may not have the exact frequency as the one used at the emission. Furthermore, both lasers typically have a phase that varies with time. To address these issues, optical carrier and phase recovery are performed as a two-step process.

The first pass is to remove the gross frequency offset between the lasers on both sides. It can be done in signal processing on standard DP-QPSK [85, 88] and happen to work on higher order modulation formats or even on the designed 4D modulation schemes tested in section 5.4.

A more challenging issue is the fast varying carrier phase that can be modeled as a multiplicative noise with Gaussian jumps

$$Y[k] = e^{i\phi[k]} H X[k] + Z[k] \tag{5.20}$$

Phase jumps $\Delta = \phi[k] - \phi[k-1]$ are modeled as a Wiener process [86], two consecutive samples at $k-1$ and k follow a zero-mean Gaussian distribution with variance $E\left[\Delta^2\right] = 2\pi\nu/R_S$ where ν is the linewidth of the laser around its central wavelength λ_0 and R_S is the baud rate. Phase noise

can be simply removed by window averaging M-PSK symbols raised to the power M. For more advanced modulation schemes $\mathcal{X}$, two main solutions exist:

1. a blind [87, 89] solution by resolving $\text{argmin}_{\phi\in(-\pi,\pi],x\in\mathcal{X}} \left\|Y[k] - e^{i\phi}HX[k]\right\|^2 := \hat{\phi}[k]$ locally for each k using the transition law.
2. a pilot-aided solution where part of the sequence sent by the transmitter is known at the receiver: by polarization-wise computing the ratio between Y and HX and then averaging in order to remove the additive circularly-symmetric noise Z, the phase noise is estimated and corrected.

In the two experiments that follow, we will use the second option.

5.3 A First Experimental Proof of Resilience to PDL Using the Spatially Balanced Signaling

In this section, we we experimentally test the Spatially Balanced signaling. We first present the adapted signal processing algorithms. We then report experimental measurements and discuss the interest of this modulation.

5.3.1 Proof-of-Concept Frame Design and Proposed Signal Processing

Figure 5.6 – Symbol frame proposed structure to process the SB and NSB signaling on experimental measurements. The frame includes synchronization symbols, CAZAC sequence symbols for channel estimation and equalization, and payload constituted of DP- and encoded QAM in which carrier phase correcting pilots are periodically inserted.

In order to experimentally validate the SB modulation format, we first want to simulate the full communication link. This section also permits us to estimate implementation penalties. Compared to usual DP-QAM-based DSP architecture, a few modifications are required in particular regarding equalization. The commonly used CMA [80, 82] for adaptative MIMO equalization again assumes that independent input streams are sent on the two polarization components. In the case of SB-QAM-based architecture, these two components are no longer independent as the imaginary parts are linked by the applied $\pi/4$-rotation.

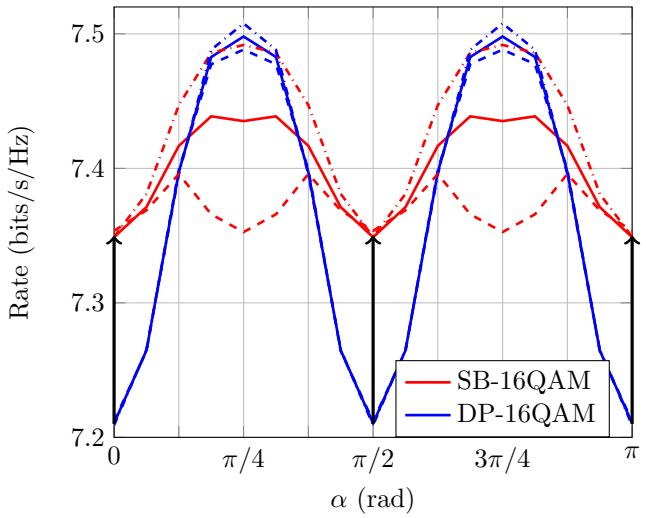


Figure 5.7 – Rate β dependence at SNR $\rho = 15$dB: minimum (at $\beta = 0$), maximum (at $\beta = \pi/8$ for DP, $\beta = \pi/4$ for SB) values in dashed lines, and for an intermediate value in full line. Simulated 6dB-PDL channel with $\alpha \in [0, \pi]$. The arrows depict relative rate gains $\Delta R = R_{\text{SB-16QAM}} - R_{\text{DP-16QAM}}$ between the two modulation formats that reach 0.14 bits/cu in the worst case (black arrows).

Recall that in section 4.1, the SB signaling was built using considerations on the simple PDL channel. For mimicking the complete fiber channel, birefringence is reincluded in the channel through a phase retarder . Carrier phase noise is added as well at the receiver and transmitter side by emulating 40kHz-linewidth lasers.

As [115] suggests, we use a pilot-based chain to deal with these impairments. As a proof-of-concept, we partition the signaling frame using as many standard DP-16QAM symbols as PDL-resilient SB-16QAM symbols. The first half of the symbol frame contains the DP-16QAM symbols. The frame structure is represented in Fig. 5.6. For simplicity, at the receiver we choose to first learn the optical channel using the CMA on the DP-16QAM symbols. Second, the whole frame is equalized with the previously learnt MIMO filter. Every 500 symbols, 12 phase recovering pilots are inserted (2.4% pilots) for carrier phase correction. Time-synchronization symbols are placed at the frame front for pilot identification.

In Fig. 5.7, the lower dashed rate curves for both signaling schemes correspond to $\beta = 0$ (worst case discussed in previous section) at a (total) SNR of 15dB. Unlike the DP-16QAM modulation, the SB-16QAM exhibits a high dependency on β as observed in section 4.1. Indeed, observe that the three DP-16QAM curve are very close, but this is not the case for SB-16QAM where the minimum and maximum largely differ. The presence of birefringence enhances the performance of our signaling leading to a higher average performance. The SB signaling still improves the link

performance as its lowest rate is raised at least to the middle value between the uncoded best and worst cases (black arrows at $\alpha = 0 \bmod \pi/2$).

5.3.2 Experimental Validation on a Single PDL Element

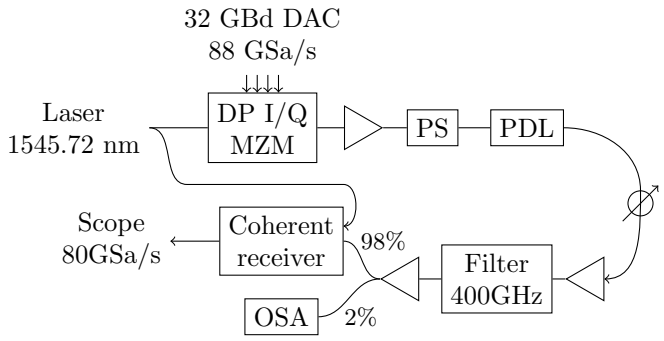


Figure 5.8 – Lumped PDL experimental setup using notably a single PDL element with $\Lambda = 6$dB preceded by a Polarization Scrambler (PS) and with an Optical Spectrum Analyzer (OSA) to retrieve the OSNR.

The experimental setup is depicted in Fig. 5.8. We voluntarily choose to not focus on carrier frequency offsets: the experimental validation is made using a homodyne configuration, where the emitting laser at $\lambda_0 = 1545.72$nm is split via a 3dB-coupler. One branch modulates the symbols and the other one plays the role of a local oscillator beating with the incoming signal. The 32 Gbd polarization-multiplexed signal is generated by 88 GSa/s Digital-to-Analog Converters (DAC), modulating a PDM-I/Q Mach-Zender Modulator (DP I/Q MZM) and then sent through a PDL channel. The PDL channel is emulated with a Polarization Scrambler (PS) that changes the polarization orientation every 10 seconds, followed by a 6dB polarization gain imbalance device denoted PDL in Fig. 5.8. The polarization scrambler is tuned in order to make sure the channel is fixed at each signal propagation between the transmitter and the receiver. At the receiver, the signal is retrieved with a coherent mixer and a 33 GHz scope at 80GSa/s.

Offline DSP processing follows. In back-to-back experiments with no PDL, rate measurements present a 0.9dB-SNR implementation penalty at SNR 15dB for the two formats. This is an implementation penalty of our lab setup and is therefore not correlated to the new signaling. Hence, this penalty has been removed in the results for better understanding. Fig. 5.9 shows the experimental fluctuations of the (total) rate associated with conventional DP-16QAM modulation. This is in accordance with the simple PDL channel model. The variations along the horizontal axis can be understood as the incident angle modifies the total SNR for each acquisition.

Having observed that the theoretical model captures the essence of the PDL dependence, let us

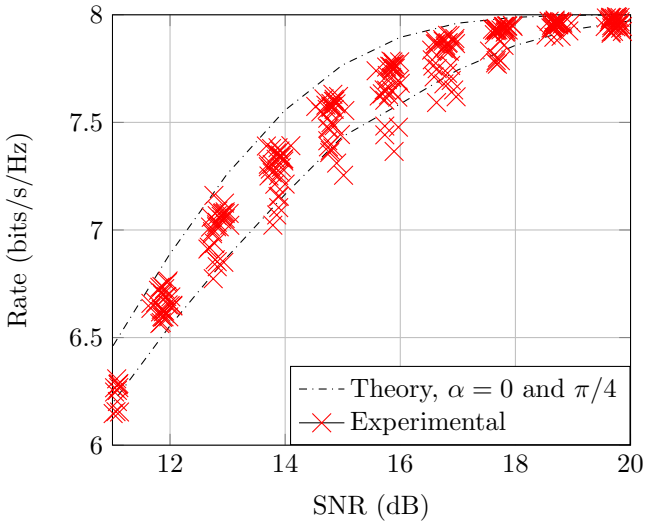


Figure 5.9 – PDL channel rate against SNR (deduced from the OSNR measured by the OSA) curve for DP-16QAM in a conventional DSP chain with a CMA and then polarization-wise processing.

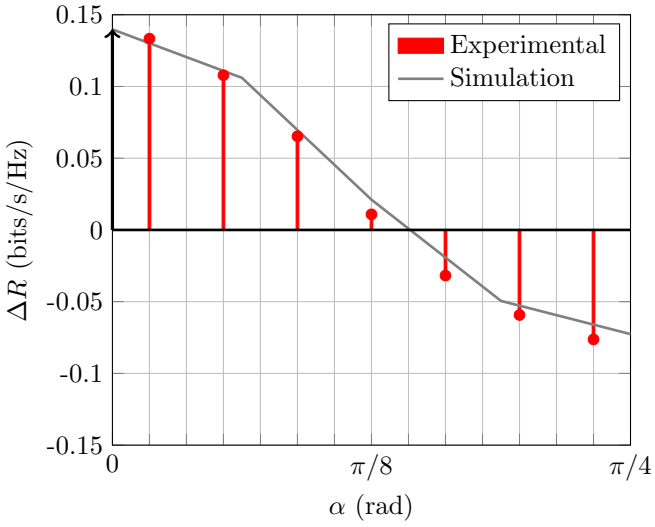


Figure 5.10 – Experimental gains ΔR of rate (quantization over 7 bins) using SB-16QAM over DP-16QAM with our pilot-based solution at SNR $\rho = 15$dB compared to simulated gains (gray line). The black arrow indicates the expected worst case enhancement averaged over all β reported from Fig. 5.7.

now focus on Fig. 5.10. Out of 100 acquisition points at an SNR of 15dB, we evaluated the gain of the new signaling measured as the difference ΔR between the rate of SB-16QAM and the one of DP-16QAM. The values are then sorted according to the channel incident angle α. We then classified the values into 7 bins that uniformly partition $[0, \pi/4]$ and represented the average gain value for each bin with red bars. The grey curve represent the predicted simulated gains and can be compared to the experimental values. The incident α angle is retrieved from a Singular Value Decomposition (SVD) of the CMA equalizer matrix at the null frequency component in the frequency domain. We observe a very satisfying matching between experiments and simulations, which fully validates our approach from theory to real-life scenarios. The novel SB signaling experimentally achieves the original expected gains. This is importantly observed at angles near $\alpha = 0$ where conventional PDL worst cases (and design margins) are located while it only slightly reduces the rates for the best SOP orientation (near $\pi/4$ as expected in theory). In this region, experiments report gains of 0.14 bits/s/Hz as predicted by simulations.

5.4 The New Spatially-Balanced Performance on a Trans-oceanic Link

After the previous experiment on a single PDL element mimicking a lumped PDL channel transmission, in this section we explore the NSB signaling this time on a distributed link containing in-line PDL elements.

5.4.1 Adapted Data-Aided Equalization Scheme

Similarly to the SB signaling, NSB signaling is a modulation where information is jointly encoded on the two polarization. Therefore, the classical 2×2 CMA equalizer is again not suitable.Consequently, we choose a pilot-based DSP approach for channel estimation and equalization and compare the performance of standard DP and NSB-coded signals using identical digital signal processing.

The sent sequences consists in two complex vectors of 2^{14} symbols each including a real-valued 64-symbol synchronization sequence, l_c = 64-symbol CAZAC sequences repeated $n_{\text{rep}} = 10$ times, and the payload symbols where the first half is standard DP-QPSK modulated and the second half is NSB-QPSK coded. This frame structure, represented in Fig. 5.6 and identical in principle to the one presented in the last section, enables a performance comparison between the two schemes for the same observed channel at each acquisition. For carrier phase recovery, we insert 8 QPSK pilots every 400 data symbols. The added pilots represent 6.25% of the total frame.

The processing chain consists in normalization and resampling of the two acquired complex

signals, compensation of chromatic dispersion when needed, time and frequency synchronization, data-aided channel estimation and multi-tap MMSE equalization, followed by carrier phase noise compensation.

5.4.2 Properties of the PDL Channel from Experimental Measurements

The channel under test at 1545.72nm is modulated with polarization-multiplexed 32GBaud signals. The experimental setup is represented in Fig. 5.11 where the tested channel is inserted in a recirculating loop along with 87 loading channels modulated at 49GBaud and covering the C-band. The loop consists of 11 spans of 55km low-PMD fibers as well as a $\Lambda = 1.1$dB PDL emulator preceded by a controllable polarization scrambler. The scrambler randomly changes the SOP for each loop. Without this scrambler, all successive propagation in the recirculating loop would be with the same incident SOP at each new introduction in the loop, which is unrealistic and does not mimic the random fiber junctions orientation in a commercial links. After 15 loops yielding a total distance of 9075km, the signal is acquired through a coherent receiver and a 80GSa/s scope.

We perform a channel estimation following the method presented in section 5.2.1. Note that this method is optimal for estimating a channel that has uncorrelated noise, but should be adapted in our correlated noise case. Though, because we do not have the noise properties at the transmitter, we still use the presented CAZAC sequences even if they may not be optimal. Then, from the estimated channel, the global PDL Λ and the incident SOP are extracted through an SVD of the estimated channel matrix H at the central tap. A total of 5000 acquisitions are recorded to cover different channel states. In Fig. 5.12, we show the statistics of the estimated overall gain imbalance Λ that fit well with the theoretical Maxwell distribution of mean $0.92 \times 1.1 \times \sqrt{15} = 3.9$dB [31]. The polarization scramblers are fixed during one acquisition to emulate the slow-varying statistics of the PDL channel (constant on one blocklength) but still changing the channel between each acquisition. Interestingly, the gain imbalance obtained from an emulated Rice channel with identical mean (magenta dashed curve) has a distribution that match even more closely the experimental values. Fig. 5.13 shows the statistics of the rotation angle α defined in section 2.5.1. The obtained semi-circular-shaped distribution depicts a uniform sampling of the Poincaré sphere. The extracted retardance angle β in $[-\pi/2, \pi/2]$ represented in Fig 5.14 is uniformally distributed as expected.

5.4.3 Demonstration of the New Spatially-Balanced Interest compared to Conventional QAM

After measuring the observed channel statistics, we are interested in comparing the information rates of the two modulation schemes. All of the following presented results are considered for a

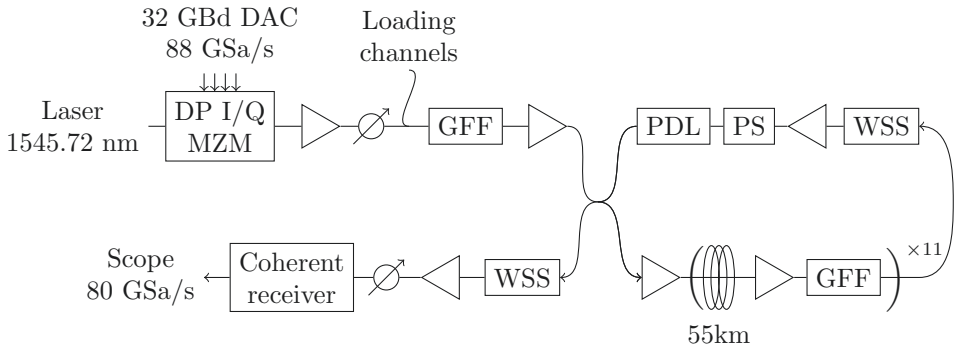


Figure 5.11 – Experimental setup for distributed PDL study constituted of a recirculating loop with in-line PDL preceded by a Polarization Scrambler (PS), 11x55km fiber spools and EDFAs (triangle). Gain Flattening Filters (GFF) and WSSs are inserted to ensure a flat C-band spectrum.

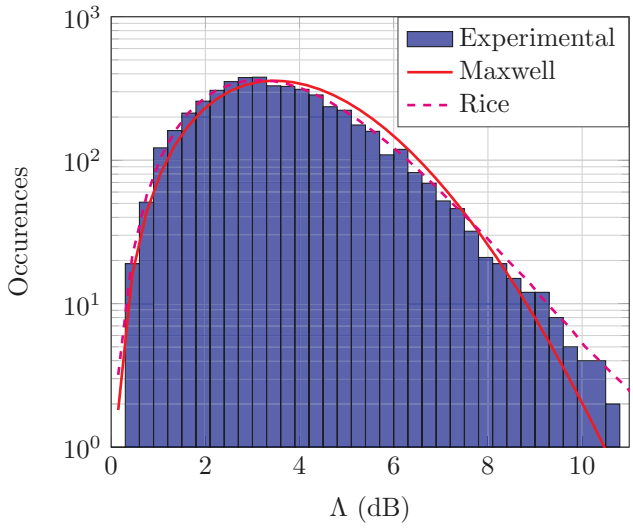


Figure 5.12 – Distribution of the parameter Λ of 5000 channels constituted of $n = 15$ concatenated, randomly oriented, PDL elements of 1.1dB following the experimental setup in Fig. 5.11.

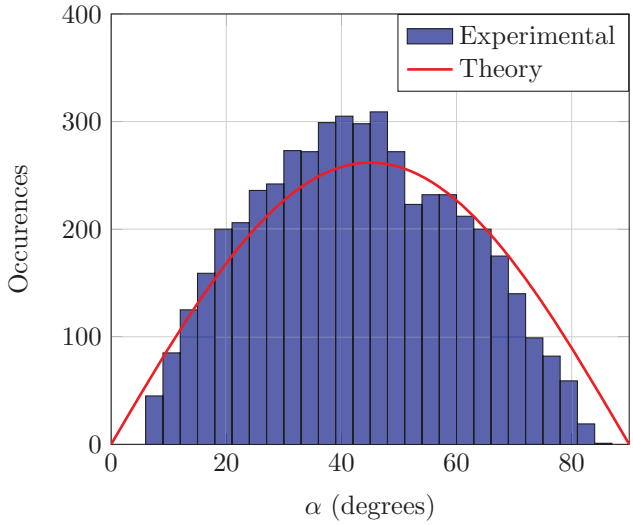


Figure 5.13 – Parameter distribution α of 5000 channels constituted of $n = 15$ concatenated, randomly oriented, PDL elements of 1.1dB following the experimental setup in Fig. 5.11.

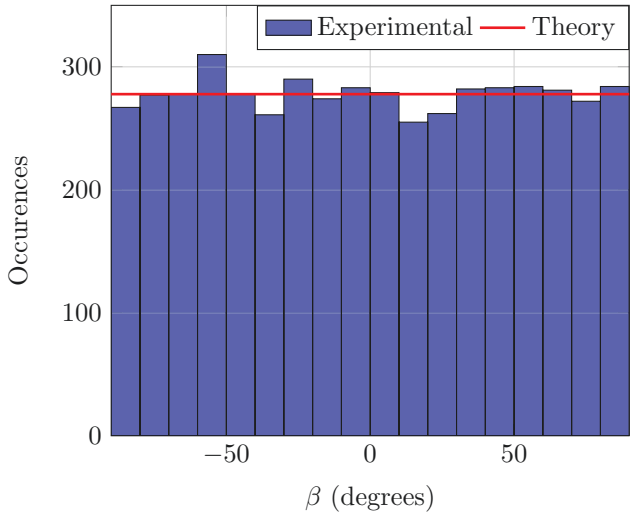


Figure 5.14 – Parameter distribution β of 5000 channels constituted of $n = 15$ concatenated, randomly oriented, PDL elements of 1.1dB following the experimental setup in Fig. 5.11.

channel configuration in the linear regime. First, we evaluated the total rate over both polarization tributaries in a back-to-back experiment without PDL (no markers). We see in Fig. 5.15 that the NSB-coded part suffers from a small penalty compared to standard DP-QPSK mainly due to imperfect phase estimation. Indeed, given that the NSB consists in mixing the in-phase and quadrature dimensions of the two polarizations, it has an increased sensitivity to phase recovery imperfections. At an OSNR of 13dB in 0.1nm corresponding to the measured value after 15 loops, the penalty is 0.0125bit/s/Hz.

While remaining in back-to-back, we add a 3.9dB loss (corresponding to the average overall PDL after 15 loops) over one polarization at the transmitter to emulate the aligned PDL case where the coding gains are expected to be the highest and measure the total rate shown in squared markers. The rate enhancement is clearly illustrated over all the measured OSNR range. Next, we analyze the information rates per polarization after transmission over 9055km for all encountered Λ values and SOP states. For the NSB-coded part, the measured penalty in back-to-back at an OSNR of 13dB was added to assess the PDL-mitigation gain independently of imperfect phase recovery. Fig. 5.16 displays the statistics of all the measured rates per polarization for the two schemes. We notice a net worst case enhancement and a reduction of rate fluctuations for the NSB part. Finally, in Fig. 5.17, we represent the measured rates of the most-impaired polarization for the NSB-QPSK and standard DP-QPSK as a function of the corresponding estimated PDL. The worst polarization rate is statistically higher for the NSB coded scheme (gray triangles) than for the standard DP one (black points). Again, the alleviation of the worst case is well observed when the joint 4D signaling is employed. For each plotted point, the realization of the channel (notably α and Λ) is different but for the worst channel states, the NSB-QPSK scheme shows a rate gain compared to the DP-QPSK one.

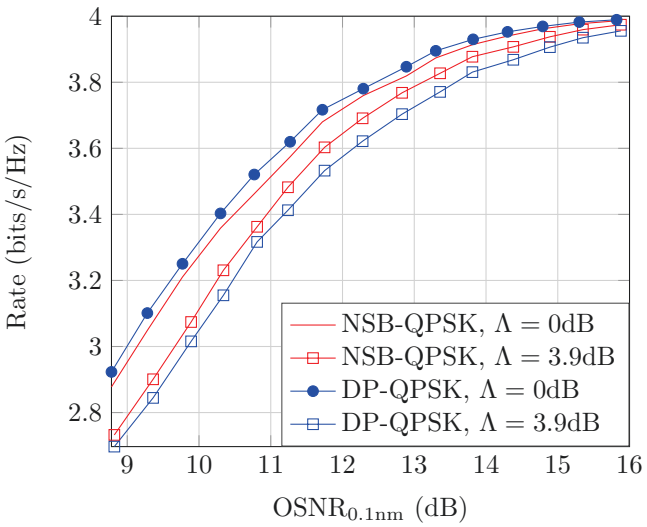


Figure 5.15 – Back-to-back total rate loss in the presence of $\Lambda = 3.1$dB of PDL in an aligned ($\alpha = 0$) configuration. The rate loss is plotted as a function of the OSNR measured in a bandwidth of 0.1nm at 32Gbauds.

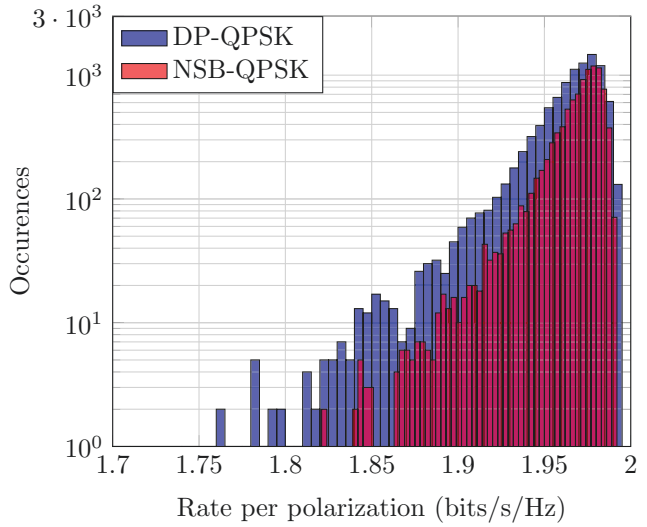


Figure 5.16 – Distribution of the rate per polarization of 5000 channels with $n = 15$ randomly oriented PDL elements of 1.1dB of PDL.

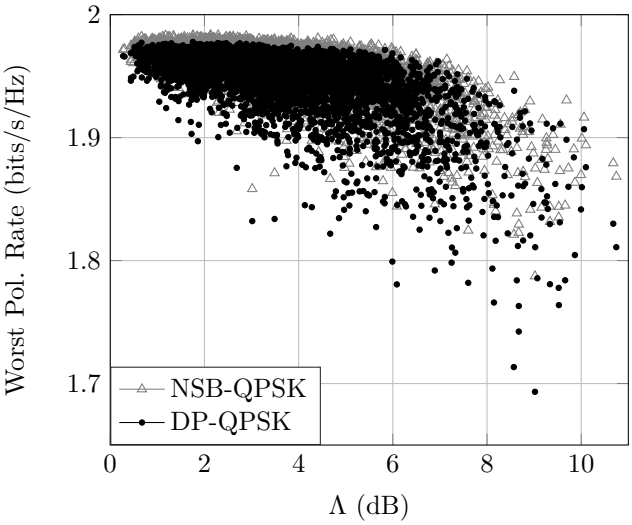


Figure 5.17 – Worst of the two rates per polarization as a function of the PDL for $n = 15$ randomly oriented PDL elements.

Chapter 6

Conclusions and Prospective Work

In this thesis, we studied in detail the impact of polarization gain imbalance on an optical link by constructing relevant channel models. In order to fully capture the discrete nature of this gain imbalance mainly introduced in optical networks by wavelength selective switches, amplifiers and possibly other optical elements, we decomposed a link in elementary spans and modeled the aggregate behavior. Besides the analysis of these channel models, we derived the fundamental limits of communications for the modulations typically used in conventional systems, and for different receiver architectures.

While it is known how to build efficient and low-complexity receivers using probabilistic coding and learning, the question of the optimal modulation targeted towards PDL-resilience along with a complexity constraint dictated by the use a single timeslot (or time-like slot) was an open question. Compared with early works on PDL-resilient signaling [34, 36], we indeed propose to overcome the issue of implementation complexity with the use of polarization code over a single time (or time-like) slot.

Using the framework of information theory, we have developed such optimal four-dimensional modulation, not only by a numerical optimization aiming at a maximization of the information rate, but also through a rigorous proof of this optimality using Euclidean geometry. The novel signal has an elegant and pleasing geometric interpretation. It is also remarkable from the viewpoint of communication theory as it consists in analytically optimizing the second minimum distance. Notice that the directional gain that the PDL-resilient modulation brings are roughly additive to other, more classical, gains such as shaping gains.

Further, we conducted simulations as well as experiments to validate the performance of the proposed schemes in real conditions. First, the Spatially Balanced signaling was tested along with

a single PDL element showing an increase of the worst encountered rate similarly to what was expected in theory. Regarding the New Spatially Balanced signaling, while exhibiting an increased sensitivity to phase recovery imperfections, it still offered rate increases not only over a channel with a single PDL element, but also over an emulation of distributed PDL in a multi-span channel with distributed noise.

The presented work brings new research avenues. First, a key practical challenge is to translate a joint receiver design involving multi-dimensional modulation schemes and complex graphical message-passing structure into an efficient high-speed VLSI implementation. As discussed in chapter 3, the joint probabilistic framework that enables efficient architectures is well-understood. Its optimization, whether it implies non-binary information processing or bit-interleaved processing, relying on efficient message quantization, remains to be correctly conducted in practice.

Nevertheless, PDL-resilient solutions are a promising direction for next-generation optical systems. Indeed they permit unit-dB gains in SNR by exploiting the complex nature of the light and capturing it in an advanced MIMO channel model. Such gains are in the same order as shaping gains while building on more complex models. The actual implementation of the New Spatially Balanced Signaling for commercial transceivers is still an open challenge and has to be considered as a prospective work. Technical issues such as the optimization of the proposed pilot frame structure, or the addition of a probabilistic shaping layer have to be addressed. Note again that these optimizations are really dependent on the targeted operating region of the designed optical link which has to be carefully characterized.

In the light of the New Spatially Balanced analysis, we initiated an interpretation of the performance of the 8D Silver code. We believe that this study paves the way to the design of optimal N dimensional ($N > 4$) PDL-resilient signalings. It is fundamental to understand which design parameters are the most relevant. Subsequently, simplified schemes can be defined taking into account implementation constraints.

Last but not least, note that many results produced in this work can largely be extended to optical or wireless channels presenting a similar gain imbalance. For instance, multi-mode channels with MDL typically have gain imbalance values larger that the ones observed in PDL channels. Those channels could benefit even more from the gain balancing signaling that were proposed in this thesis. In general, the approach we adopted could be extended to any $N \times N$-MIMO channel in order to find the adapted $\mathrm{SO}(N)$ encoding that increases the most the link capacity.

Appendix A

Unitary Jones Matrices Decomposition

The matrices U and V mentioned in Eq. 2.1 are in the unitary matrices group $\mathrm{U}(2) = \{U \in \mathbb{C}^{2\times 2} | UU^\dagger = I\}$ and can be decomposed [42] as $U = e^{i\phi} B_\delta R_\alpha B_\beta$ where $\phi, \alpha, \beta, \nu \in \mathbb{R}$. Observe that $\det U = e^{i\phi}$. The two B matrices have the same expression which is function of an angle parameter, for instance

$$B_\beta = \mathrm{diag}\left\{e^{i\beta}; e^{-i\beta}\right\} \tag{A.1}$$

and R_α is a real-valued rotation matrix defined by

$$R_\alpha = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \tag{A.2}$$

The scalar phase ϕ is common to both polarizations and does not play any role in the imbalance of the two polarization tributaries. We often omit it and only keep the three other parameters. In this case, the matrices are in the Special Unitary group of order 2 denoted $\mathrm{SU}(2) = \{U \in \mathrm{U}(2) \,|\, \det U = 1\}$ are expressed as a function of the three angle parameters α, β and δ only

$$U = B_\delta R_\alpha B_\beta \tag{A.3}$$

The distribution of the three angles are discussed in section 2.5.1.

Appendix B

Results on Singular Values of Rayleigh Channels

For better understanding, we recall the following well-known definitions in the context of a 2×2-MIMO communication link.

Definition 7 *A 2×2-MIMO* Rayleigh *fading channel is a channel with transfer matrix $G \in \mathbb{C}^{2\times 2}$ which entries are distributed following a complex Gaussian process $G_{ij} \sim \mathcal{CN}(0, 1/2)$ for $i, j \in \{1; 2\}$. We will call G a* Gaussian matrix.

Definition 8 *If G is a Gaussian matrix, then $W = GG^\dagger$ is said to follow a* Wishart *distribution.*

Definition 9 *The singular values of a matrix H are the square roots of the eigenvalues of $HH^\dagger$.*

We now focus on the distribution law $(S_1, S_2) = S$ of the singular values s_1 and s_2 of the channel matrix $G \in \mathbb{C}^{2\times 2}$ defined above by means of change of variable from a known distribution. The Gaussian matrix represents the model with the least knowledge about the channel and is insightful to get properties of the transmission channel. First, the joint density distribution of the *eigenvalues* of $W = GG^\dagger$ is [45]

$$p_S(s_1, s_2) \propto e^{-(s_1+s_2)}(s_1 - s_2)^2 \tag{B.1}$$

where $s_1 \geq s_2$.

The distribution of the ratio $R = S_1/S_2$ of the eigenvalues of W can be computed by a change of variable as $p_R(r) = \int |s_2| p_S(rs_2, s_2)\mathrm{d}s_2 = \int s_2 p_S(rs_2, s_2)\mathrm{d}s_2$ because s_i are positive as eigenvalues

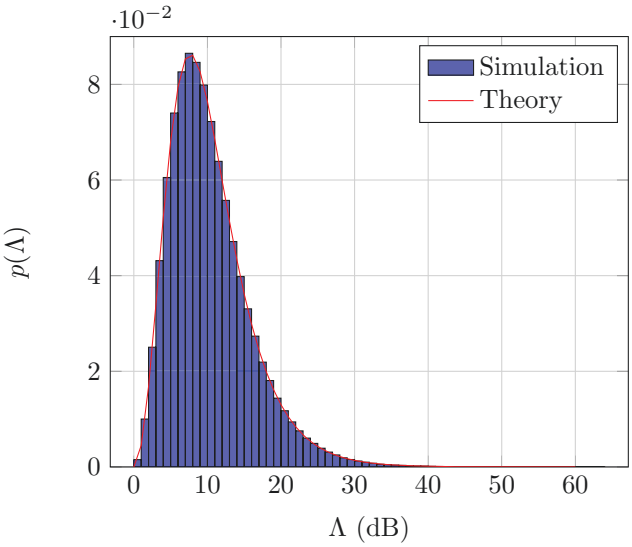


Figure B.1 – Singular value ratio distribution for 10^6 Rayleigh fading channels.

of $GG^\dagger$ positive definite. Note that r is in $[1, +\infty[$ by the ratio definition and the order relation of s_i. Its expression is finally $p(r) \propto (1-r)^2/(1+r)^4$. To relate to G singular values distribution, notice that $\sqrt{r} = \sqrt{s_1}/\sqrt{s_2} = \rho$ is defined on $[1, +\infty[$ and if $r = \rho^2$, then $\mathrm{d}r = 2\rho\mathrm{d}\rho$ and

$$p(r)\mathrm{d}r \propto \frac{(1-r)^2}{(1+r)^4}\mathrm{d}r = \frac{(1-\rho^2)^2}{(1+\rho^2)^4}2\rho\mathrm{d}\rho = p(\rho)\mathrm{d}\rho \tag{B.2}$$

with $\int p = 1/6$ so the expression should be scaled by 6.

Similarly, with $\Lambda = 10\log_{10}(\rho^2)$, we have $\rho = 10^{\Lambda/20}$, $\mathrm{d}\rho = \frac{\ln(10)}{20}10^{\Lambda/20}\mathrm{d}\Lambda$ and finally

$$p(\rho)\mathrm{d}\rho = \frac{3}{5}\ln(10)\frac{(1-10^{\Lambda/10})^2}{(1+10^{\Lambda/10})^4}10^{\Lambda/10}\mathrm{d}\Lambda = p(\Lambda)\mathrm{d}\Lambda \tag{B.3}$$

with now $\Lambda \in [0, +\infty[$.

In Fig. B.1 we represent in blue the squared singular value ratio distribution of 10^6 simulated channels with $H \in \mathbb{C}^{2\times 2}$ with $h_{ij} \sim \mathcal{CN}(0,1)$. We added in red the theoretical fit from Eq. B.3. Note that this distribution cannot match well the one of distributed PDL channels. Indeed, by opposition to the Rice fading channel, no parameter is accessible to tune the mean value of the plotted distribution. The average value is about 10dB and is incompatible with common PDL channels. This is why we investigate the Rice distribution in section 2.5.3 instead of the Rayleigh one.

Appendix C

A Glimpse into Information and Coding Theory: Derivation of Capacity-Approaching Receivers for Joint Decoding-Demodulation

In this appendix, we briefly recall some results regarding information theory, and on Forward Error Correcting (FEC) code structures. We forst present two linear block codes: the basic Hamming code and the more advanced Low-Density Parity-Check (LDPC) codes. We then elaborate on the nearly optimal message-passing techniques in order to reach the information rates.

C.1 Hamming Code Introduction to Bit Coding

The benefit of FEC or channel coding can be understood by the simple Hamming code scheme. Assume that an information bit sequence $b = b_1 b_2 \dots b_k \in \mathbb{F}_2^k = \{0,1\}^k$ with $p \in \mathbb{N}$ is encoded in a

$+$	0	1
0	0	1
1	1	0

Table C.1 – Bit sum operation in $\mathbb{F}^2$.

BPSK symbol sequence $x \in \{1, -1\}^k$ with the mapping rule

$$\mathbb{F}_2 \to \{1, -1\}\,; \begin{cases} 0 \to +1 \\ 1 \to -1 \end{cases} \tag{C.1}$$

In a classical transmission, the estimate $\hat{x}$ at the receiver of the sent sequence x will likely be altered by the additive Gaussian noise coming from the channel, resulting in incorrect bit transmission. The Hamming coding scheme [60] in its simplest form proposes to append a redundant parity bit b_{k+1} to the sequence b following the rule

$$b_{k+1} = b_1 + b_2 + \cdots + b_k \tag{C.2}$$

called the parity check operation.

Equivalently, because $b_1 + b_2 = b_1 - b_2$ in $\mathbb{F}^2$ (see table C.1), the previous equation can be written

$$\sum_{i=1}^{k+1} b_i = 0 \tag{C.3}$$

Then, at the reception side, if $\sum_{i=1}^{k+1} \hat{b}_i \neq 0$, a transmission outage occurred. Though, two incorrect bit estimations cancel out each other and this simple scheme is inefficient when the noise power increases. Then, the general Hamming scheme contains several parity checks verifying the bits at different position that even enables to detect the incorrect bits location.

Definition 10 *A (n, k)-Hamming code is a linear code with k information bits for $n > k$ transmitted bits. The parity check matrix $H \in \mathbb{F}^{k \times n}$ is such that for b in the codebook $\mathcal{C}$ of authorized bit sequences, $Hb = 0$ with H of the form*

$$\left(\begin{array}{c|c} C & I_{n-k} \end{array} \right) \tag{C.4}$$

Each row C_i of $C \in \mathbb{F}_2^{n-k \times k}$ represents the parity check done by the redundant bit b_i, therefore C_i contains zeros except at the positions j that appear in the parity check sum $b_i = \sum_j b_j$ where it is 1.

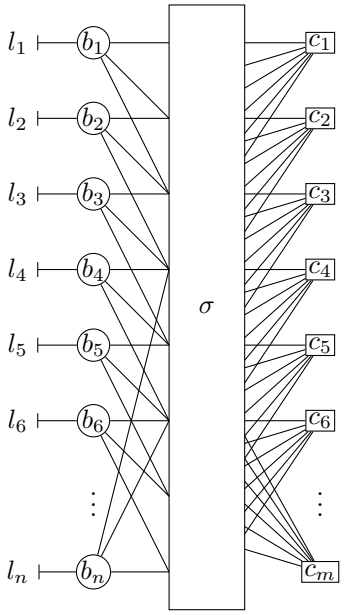


Figure C.1 – Factor graph representation of a $(3, 6)$-LDPC with blocklength n. The variables l_i are log-likelihood ratio calculated from the channel, variable nodes (bits) are represented with circles and check nodes are represented by squares.

Example

$$H = \begin{pmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 \end{pmatrix} \tag{C.5}$$

is the parity check matrix of a $(7, 4)$-Hamming code that captures the rules

$$\begin{cases} b_1 + b_2 + b_3 + b_5 = 0 \\ b_1 + b_2 + b_4 + b_6 = 0 \\ b_1 + b_3 + b_4 + b_7 = 0 \end{cases} \tag{C.6}$$

The last bit in the three equations are *parity check* bits and the other three ones are *information* bits.

C.2 Low-Density Parity-Check Codes

The commonly used Low-Density Parity-Check (LDPC) codes introduced by Gallager [101] are similar to Hamming codes: their parity check matrix H verifies the same type of bit parity equations. The innovation is that H is *sparse*, meaning that it contains significantly more zeros than ones.

Definition 11 *A regular (l, r, n)-LDPC code is a linear code with encoding matrix $H \in \mathbb{F}_2^{m \times n}$ sparse with $l, r \ll n$ and $m = nl/r$ that has l ones in each column and r ones in each row. The integers l and r are called left and right degrees, respectively, and n is the length of the codeword.*

Example LDPC parity check matrices are generally large by definition. By space constraint we only show the first values of an example H of a $(3, 6)$-LDPC for illustration:

$$H = \begin{array}{c} \begin{matrix} b_1 & b_2 \end{matrix} \\ \left(\begin{array}{cccccccc} 1 & 1 & 1 & 0 & 1 & 0 & 0 & \cdots \\ 0 & 1 & 0 & 1 & \cdots & & & \\ 0 & 1 & 1 & 1 & & & & \\ 0 & 0 & 0 & 0 & & & & \\ \vdots & \vdots & & & & & & \end{array} \right) \\ \downarrow \\ \Sigma_i h_{ij} = 3 \end{array} \longrightarrow \Sigma_j h_{ij} = 6$$

The location of the ones in a row i are the indices of the bits used by the check node i. And the ones location in a column j correspond to the indices of the check nodes the bit b_j is sent to. For large enough bit blocklength, it is shown [51] that LDPC codes enable to be very close to the capacity, motivating their use in a transmission chain. An illustration of a $(3, 6)$-LDPC is represented in Fig. C.1 as a factor graph. Bit variable nodes b_i are represented in circles on the left, and the check nodes c_j represented by squares on the right perform the parity check operations. Bit decoding using an LDPC is discussed in [51] with the belief propagation algorithm used on a factor graph formalism, and is briefly presented thereafter.

C.3 Bit Decoding Using the Belief Propagation Algorithm

The decoding algorithm consists in recursively estimating the value of the bits and computing the parity check information. The bit nodes b_i (circle) in the context of factor graphs are called *variable* nodes, and the parity check function nodes c_j (rectangles) are called *check nodes*. Either type of

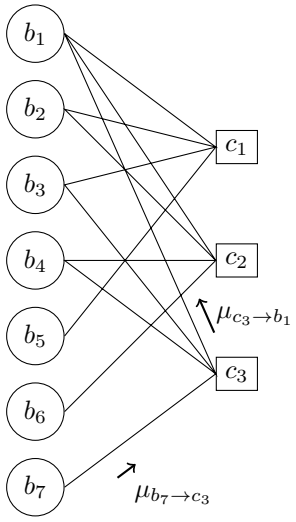


Figure C.2 – Message passing in a $(7, 4)$-Hamming code bit decoder: each variable node state b_i state can be inferred from the parity check sums with other variables.

nodes use the connections of the factor graph called *edges*, to send information or *messages* for the next iteration. In bit soft decoding, instead of the value of the variable node, the sent message is the probability of the variable to be in a given state. In the particular case of bit variables, the message μ of the state of a bit b_i is $\mu = (p(b_i = 0), p(b_i = 1))$ or equivalently, because the sum of the probabilities is unitary, $\mu = l_i$ where $l_i = \log_{10}(p(b_i = 0)/p(b_i = 1))$ is the (Log) Likelihood Ratio (LLR) associated to the bit b_i. Fig. C.2 illustrates a simple Hamming code message passing. The information $\mu_{c_3 \to b_1}$ on the state of the bit b_1 is inferred from the parity bit b_7 and the other bits b_3 and b_4 using the parity check sum at c_3. The same happens for LDPCs but with a larger scale and with more connections. The rectangle σ in Fig. C.1 captures the permutations between the bits and the check nodes defined by the matrix H. Finally, the initial LLR values are calculated from the symbol-to-bit demapping, derived in the following section, and serve as input to the LDPC decoder. Working with LLRs, the standard operations in the (l, r)-LDPC decoder produces left-to-right and right-to-left messages that are respectively [51]

$$\mu_{b_i \to c_j} = l_i + \sum_{k=1,\dots,l, k \neq j} \mu_{c_k \to b_i} \tag{C.7}$$

$$\mu_{c_j \to b_i} = 2 \operatorname{arctanh} \prod_{\substack{k=1,\dots,r \\ k \neq i}} \tanh \frac{\mu_{b_i \to c_k}}{2} \tag{C.8}$$

In section 3.4 we use an LDPC as a bit decoder, more specifically we use a spatially-coupled LDPC which is a subclass of LDPCs. Such LDPCs, typically designed for large blocklengths, are formed with ℓ randomly interconnected stacked copies of a protograph that is solely composed of local interactions of a pattern [51, 109, 58]. The parameter ℓ is called the *lifting depth* of the code, and the span of the coupling between each copy is the *coupling window*.

C.4 Optimized Joint Receiver Architecture with Symbol-to-Bit Demapping

Following [51], we develop the Maximum-Likelihood (ML) expression of the LLR from its definition by marginalization of the bit probability

$$l_i = \frac{p(b_i = 0|Y, H)}{p(b_i = 1|Y, H)} \tag{C.9}$$

$$= \frac{\sum_{b_1,\dots,b_{i-1},b_{i+1}\dots,b_n} p(b_1, \dots, b_i = 0, \dots, b_n|Y, H)}{\sum_{b_1,\dots,b_{i-1},b_{i+1}\dots,b_n} p(b_1, \dots, b_i = 1, \dots, b_n|Y, H)} \tag{C.10}$$

$$= \frac{\sum_{X\in\mathcal{X}|b_i=0} p(X|Y, H)}{\sum_{X\in\mathcal{X}|b_i=1} p(X|Y, H)} \tag{C.11}$$

$$= \frac{\sum_{X\in\mathcal{X}|b_i=0} p(Y|X, H)}{\sum_{X\in\mathcal{X}|b_i=1} p(Y|X, H)} \tag{C.12}$$

noting the Bayes rule $p(X|Y, H) = p(Y|X, H)p(X)/p(Y)$ with $p(X)$ (constant) and $p(Y)$ disappearing in the ratio l_i. The LLR can then been calculated with the last formula, assuming that the channel transfer function $p(Y|X, H)$ is known. In a lumped PDL model with circularly-symmetric additive noise with unit covariance, because $p(Z) \propto \exp(-\|Z\|^2/2)$, we have $p(Y|X, H) \propto \exp\left(-\|Y - HX\|^2/2\right)$. From a Maximum-A-Posteriori (MAP) definition of the LLR, we derived a ML definition that leads to the low-complexity message-passing receiver.

In this appendix, we have skimmed through the construction method for designing efficient low-complexity receiver that can benefit from our optimal 4-dimensional PDL-resilient modulation newly introduced in this thesis. Notice that more construction information can be found in [51]. Let us eventually represent the type of optimal receiver architecture we envision for taking full benefits of our modulation. It is based on a binary representation of the computational information while non-binary receiver architectures could also be envisioned, see [52, 54]. Hence, in Fig. 3.7 we represent a factor graph capturing the previously presented two main operations involved in a receiver that are

1. Channel equalization and symbol-to-bit demapping

2. Bit decoding implemented as a LDPC decoder

We suppose a polarization-multiplexed input $X = (X_1, X_2)^T$ and a channel transfer matrix $H \in \mathbb{C}^{2\times 2}$ assumed to be known. Information theory suggests to perform both operations jointly for optimal decoding. The CM information rates presented before represented the limits of communications are valid if optimal signal processing is performed at the receiver. Symbol decoding and channel equalization should be performed at the same time [98] in order to meet these limits. In practice though, this implies a high complexity at the receiver that is not wanted as explored in section 3.3.

Appendix D

Properties on SO (4) Matrices Decomposition

In this appendix we focus on unitary transforms in $\mathbb{R}^4$. A unitary transform can be represented by a matrix $U \in \mathbb{R}^{4\times4}$ such that $UU^T = U^TU = I$ and $(\det U)^2 = 1$. Among the unitary matrix, the special orthogonal group of order 4, denoted $\mathrm{SO}(4)$, are the matrices with determinant 1.

D.1 Left- and Right-Isoclinic

$\mathrm{SO}(4)$ is the special orthogonal group for the matrix multiplication operation and is of dimension 6. Any matrix $G \in \mathrm{SO}(4)$ can be decomposed [71] as the product of two matrices

$$G = \begin{pmatrix} a & -b & -c & -d \\ b & a & -d & c \\ c & d & a & -b \\ d & -c & b & a \end{pmatrix} \begin{pmatrix} p & -q & -r & -s \\ q & p & s & -r \\ r & -s & p & q \\ s & r & -q & p \end{pmatrix} = G_L G_R \tag{D.1}$$

where $a, b, c, d, p, q, r, s \in \mathbb{R}$ and such that $a^2 + b^2 + c^2 + d^2 = 1$ and $p^2 + q^2 + r^2 + s^2 = 1$.

The first factor matrix G_L is called a *left-isoclinic* rotation matrix, and the second one G_R is a *right-isoclinic* rotation. The ensemble of left- and right-isoclinic tranforms can be denoted S^3_L and S^3_R, respectively. They represent sub-groups of $\mathrm{SO}(4)$. We have $G_LG_R = G_RG_L$ but S^3_L and S^3_L are not commutative groups. The decomposition is unique up to the G_L and G_R permutation.

Using a Hopf parametrization $\theta, \eta, \nu, \alpha, \beta, \tau \in (-\pi, \pi]$

$$\begin{cases} a = \cos\theta\cos\eta \\ b = \sin\theta\cos\eta \\ c = \cos\nu\sin\eta \\ d = \sin\nu\sin\eta \end{cases} \quad \text{and} \quad \begin{cases} p = \cos\alpha\cos\tau \\ q = \sin\alpha\cos\tau \\ r = \cos\beta\sin\tau \\ s = \sin\beta\sin\tau \end{cases} \tag{D.2}$$

we can describe SO (4) and we can write the left- and right- isoclinic rotation as the following block matrices

$$G_L = \left(\begin{array}{c|c} \cos\eta R_\theta & -\sin\eta S_\nu \\ \hline \sin\eta S_\nu & \cos\eta R_\theta \end{array}\right) \text{ and } G_R = \left(\begin{array}{c|c} \cos\tau R_\alpha & -\sin\tau R_{-\beta} \\ \hline \sin\tau R_\beta & \cos\eta R_{-\alpha} \end{array}\right) \tag{D.3}$$

where the R and S matrices correspond to the rotation and reflection matrices respectively, defined before. Note that we have again that SO (4) is of dimension 6. The left- and right- isoclinic rotations can be found [71] using the relations

$$\begin{cases} G^R = \mathcal{L}_i(G) / \left[\det \mathcal{L}_i(G)\right]^{1/4} \\ G^L = G\mathcal{L}_i(G)^T / \left[\det \mathcal{L}_i(G)\right]^{1/4} \end{cases} \tag{D.4}$$

for $i = 1, 2, 3, 4$ where $\mathcal{L}_i(G)$ is a linear operator on G, for instance

$$\mathcal{L}_1(G) = -\frac{1}{4}\left(-G + A_1 G A_1 + A_2 G A_2 + A_3 G A_3\right) \tag{D.5}$$

where A_i are matrices forming a base of S^3_L

$$A_1 = \begin{pmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}, \; A_2 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \text{ and } A_3 = \begin{pmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \tag{D.6}$$

D.2 SU (2) and Scalars Representation on SO (4)

First, observe that a unit scalar multiplication $e^{i\theta}X$ is equivalent to left-multiply by $\operatorname{diag}\left\{e^{i\theta}, e^{i\theta}\right\}$ hence if $\hat{X} = (\Re x_1, \Im x_1, \Re x_2, \Im x_2)^T$, in $\mathbb{R}^4$ we have

$$e^{i\theta}X \leftrightarrow \begin{pmatrix} R_\theta & 0 \\ 0 & R_\theta \end{pmatrix} \hat{X} \tag{D.7}$$

We do not strictly have $\mathrm{SU}(2) \subset \mathrm{SO}(4)$ as the two ensembles are not defined on the same field, but the projection of $\mathrm{SU}(2)$ of dimension 3 in $\mathrm{SO}(4)$ is a subset of $\mathrm{SO}(4)$. In this section we describe how $\mathrm{SU}(2)$ and $\mathrm{SO}(4)$ are related.

Matrix Exponential Based Following [8], recall that in Jones (complex) vector formalism, a transfer matrix $U \in \mathbb{C}^{2\times 2}$ is such that $Y = UX$ with $X, Y \in \mathbb{C}^{2\times 1}$. In particular with $U \in \mathrm{SU}(2)$, we define the matrix

$$S = \begin{pmatrix} s_1 & s_2 - is_3 \\ s_2 + is_3 & -s_1 \end{pmatrix} \tag{D.8}$$

with $s_i \in \mathbb{R}$ such that we write U as the exponential of the matrix S

$$U = \exp(iS) = I\cos s + \frac{iS}{s}\sin s \tag{D.9}$$

where $s = \sqrt{s_1^2 + s_2^2 + s_3^2}$.

Now if we define the vector $\hat{X} = (\Re x_1, \Im x_1, \Re x_2, \Im x_2)^T \in \mathbb{R}^4$ the projection of X on the real dimensions, we can define $G \in \mathrm{SO}(4)$ and call $\hat{Y} = G\hat{X}$. Define the *skew-symmetric* matrix

$$M = \begin{pmatrix} 0 & -l_1 - r_1 & l_3 + r_3 & l_2 - r_2 \\ l_1 + r_1 & 0 & l_2 + r_2 & -l_3 + r_3 \\ -l_3 - r_3 & -l_2 - r_2 & 0 & -l_1 + r_1 \\ -l_2 + r_2 & l_3 - r_3 & l_1 - r_1 & 0 \end{pmatrix} \tag{D.10}$$

where $l_i, r_i \in \mathbb{R}$. Note that $M^T = -M$. We can write $M = M_L + M_R$ where M_L and M_R depends only in l_1, l_2, l_3 and r_1, r_2, r_3, respectively. We finally have

$$G = \exp M = \exp M_L \exp M_R = G_L G_R = G_R G_L \tag{D.11}$$

where G_L and G_R are left- and right- isoclinic rotations defined before.

To relate U to G, we just have to take $l_1 = l_2 = l_3 = 0$ and identify $r_i = s_i$ for $i = 1, 2, 3$. Indeed, as in Eq. D.9 we have also $G_R = I\cos r + M_R \sin(r)/r$ with $r = \sqrt{r_1^2 + r_2^2 + r_3^2} = s$ in our case such that

$$G_R X = \cos s \begin{pmatrix} \Re x_1 \\ \Im x_1 \\ \Re x_2 \\ \Im x_2 \end{pmatrix} + \frac{\sin s}{s} \begin{pmatrix} 0 & -s_1 & s_3 & -s_2 \\ s_1 & 0 & s_2 & s_3 \\ -s_3 & -s_2 & 0 & s_1 \\ s_2 & -s_3 & -s_1 & 0 \end{pmatrix} \begin{pmatrix} \Re x_1 \\ \Im x_1 \\ \Re x_2 \\ \Im x_2 \end{pmatrix} \tag{D.12}$$

Thus using Eq. D.8, D.9 and Eq. D.12, remembering $x_i = \Re x_i + i\Im x_i$ and $Y = UX$ we conclude

$$\begin{pmatrix} \hat{Y}_1 + i\hat{Y}_2 \\ \hat{Y}_3 + i\hat{Y}_4 \end{pmatrix} - Y = \begin{pmatrix} \hat{Y}_1 + i\hat{Y}_2 \\ \hat{Y}_3 + i\hat{Y}_4 \end{pmatrix} - \left\{ \cos s \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + i\frac{\sin s}{s} \begin{pmatrix} s_1 & s_2 - is_3 \\ s_2 + is_3 & -s_1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \right\} = 0 \quad \text{(D.13)}$$

We then have $G = G_R$ meaning that SU (2) projects in SO (4) as right-isoclinic rotations. As the dimensions of the two ensembles are identical, we conclude that S^3_R is exactly the unitary transforms in $\mathbb{R}^4$ that correspond to SU (2) transforms in $\mathbb{C}^2$.

Hopf Coordinates Based $U \in \mathrm{SU}(2)$ parameterized as $U = B_\nu R_\alpha B_\beta$ where $\nu, \alpha, \beta \in [0, 2\pi[$ is represented in SO (4) as $G_R(\alpha, \beta + \nu, \beta - \nu)$.

The Van Elfrinkhof formula [71] can be written using Hopf coordinates as

$$G\hat{X} \leftrightarrow \begin{pmatrix} e^{i\xi_1} \sin\gamma & e^{i\xi_2} \cos\gamma \\ -e^{-i\xi_2} \cos\gamma & e^{-i\xi_1} \sin\gamma \end{pmatrix} \begin{pmatrix} \Re x_1 + i\Im x_1 \\ \Re x_2 + i\Im x_2 \end{pmatrix} \quad \text{(D.14)}$$

where the G matrix represented as above is the same as the right-isoclinic rotation in Eq. D.3 transcripted in the complex field with $\theta = \xi_1, \beta = \xi_2, \tau = \pi - \gamma$.

D.3 The Spatially Balanced Signaling

We know that the Spatially Balanced signaling by construction is not a SU (2) transform. Its decomposition in left- and right-isoclinic rotations raises that $G_L \neq I$ otherwise it would be a SU (2) transform according to the previous section. Also we know from section 4.2 that SU (2) transform do not change the worst case performance value, which is not the case for the Spatially Balanced signaling. Using Eq. D.4 we calculate the projections in S^3_L and S^3_R of the encoding matrix of the Spatially Balanced signaling originally defined in section 4.1 with the imaginary parts rotated by an angle θ:

$$G_L = \begin{pmatrix} \cos\theta/2 & 0 & \sin\theta/2 & 0 \\ 0 & \cos\theta/2 & 0 & -\sin\theta/2 \\ -\sin\theta/2 & 0 & \cos\theta/2 & 0 \\ 0 & \sin\theta/2 & 0 & \cos\theta/2 \end{pmatrix} = G_L(\theta/2, 0, 0) \tag{D.15}$$

$$G_R = \begin{pmatrix} \cos\theta/2 & 0 & -\sin\theta/2 & 0 \\ 0 & \cos\theta/2 & 0 & -\sin\theta/2 \\ \sin\theta/2 & 0 & \cos\theta/2 & 0 \\ 0 & \sin\theta/2 & 0 & \cos\theta/2 \end{pmatrix} = G_R(-\theta/2, 0, 0) \tag{D.16}$$

Recall that we proposed a second version of the SB transform as an equally divided offset χ between the real and imaginary parts, i.e., $f(X) = R_{-\chi/2}\Re X + iR_{\chi/2}\Im X$. Its encoding matrix is:

$$G = \begin{pmatrix} \cos\chi & 0 & \sin\chi & 0 \\ 0 & \cos\chi & 0 & -\sin\chi \\ -\sin\chi & 0 & \cos\chi & 0 \\ 0 & \sin\chi & 0 & \cos\chi \end{pmatrix} \tag{D.17}$$

which is already a pure left-isoclinic matrix, meaning that $G_R = I_4$.

We discussed that the optimum angle rotation for the original definition is $\theta = \pi/4$ and $\chi = \pi/8$ for the form D.17 for low PDL values. With those angle values, the two G_L expressions match exactly, and the original SB encoding matrix carries an extra useless right-isoclinic rotation component. The uselessness is again that it just performs a rotation of the entire signal (both complex parts) by the same angle $\theta/2$. The left-isoclinc is the part that effectively raises the worst rate of the signaling as it balances the two complex parts.

D.4 General Modulations Based on SO (4) Transforms

In this section we manipulate the G_L expression to have a simple decomposition of the left-isoclinic rotation in the complex field.

$$G_L = \left(\begin{array}{c|c} \cos\eta R_\theta & -\sin\eta S_\nu \\ \hline \sin\eta S_\nu & \cos\eta R_\theta \end{array}\right) = \left(\begin{array}{c|c} \cos\eta R_\theta & 0 \\ \hline 0 & \cos\eta R_\theta \end{array}\right) + \left(\begin{array}{c|c} 0 & -\sin\eta S_\nu \\ \hline \sin\eta S_\nu & 0 \end{array}\right) \tag{D.18}$$

Then from the reflection matrix definition in Eq. 4.7, we have

$$G_L X \equiv \begin{pmatrix} e^{i\theta}\cos\eta & 0 \\ 0 & e^{i\theta}\cos\eta \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} 0 & -e^{i\nu}\sin\eta \\ e^{i\nu}\sin\eta & 0 \end{pmatrix} \begin{pmatrix} x_1^* \\ x_2^* \end{pmatrix} \tag{D.19}$$

$$= e^{i\theta} \left\{ \cos\eta X + \sin\eta e^{i(\nu-\theta)} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} X^* \right\} \tag{D.20}$$

As in appendix A, we drop the common phase $e^{i\theta}$ — it is the missing phase in $\mathrm{SU}(2)$ matrices to be a unitary matrix — and redefine $\nu := \nu - \theta$. Only the study of the two parameters η and ν are necessary for deriving a $\mathrm{SO}(4)$ transform for PDL resilience, which justifies the study of the function $f_{\eta,\nu}$ defined in Eq. 4.8.

Appendix E

Optimal Basis Orientation for PDL-Resilience

E.1 Optimum Angles in $F_{\eta,\nu}$

We want to minimize the $\|a_i\|$ defined as

$$|a_2|^2 = \cos(2\eta)^2 \tag{E.1}$$

$$|a_3|^2 = \sin(2\eta)^2 \sin(\nu)^2 \tag{E.2}$$

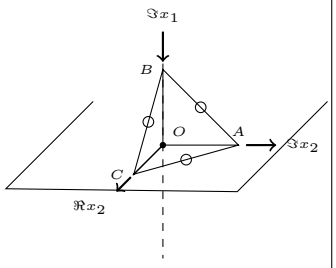
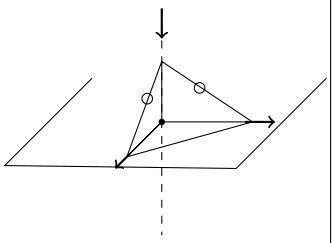
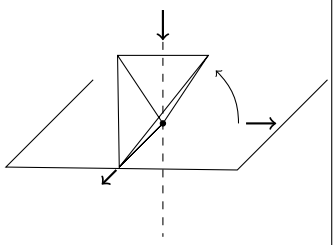
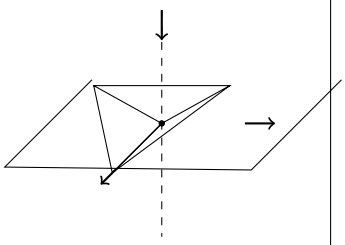
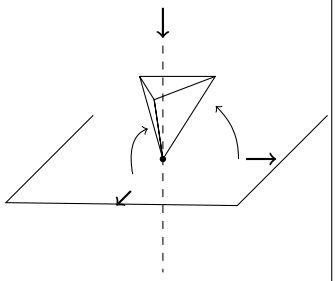
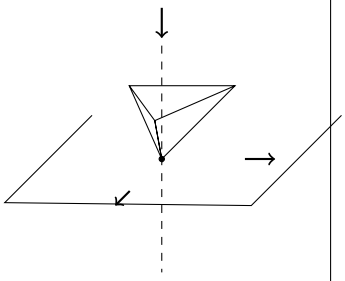
$$|a_4|^2 = \sin(2\eta)^2 \cos(\nu)^2 \tag{E.3}$$

Consider the case $\nu \in [-\pi/4, \pi/4]$. Then $\sin(\nu)^2 \geq \cos(\nu)^2$ therefore $\min(|a_2|^2, |a_3|^2, |a_4|^2) = \min(|a_2|^2, |a_4|^2)$. Studying the map $\eta \mapsto |a_2(\eta,\nu)|^2 - |a_4(\eta,\nu)|^2 = \cos(2\eta)^2 - \sin(2\eta)^2 \sin(\nu)^2$ with derivative $\eta \mapsto \sin(4\eta)(\cos(2\nu) - 3) \leq 0$ over $[0, \pi/4]$ reveals that $\min(|a_2|^2, |a_3|^2, |a_4|^2) = |a_4|^2 = \sin(\nu)^2 \sin(2\eta)^2$ for $\eta \in [0, \frac{1}{2}\cos^{-1}\left(\sqrt{\frac{\sin^2(\nu)}{\sin^2(\nu)+1}}\right)$ and $\min(|a_2|^2, |a_3|^2, |a_4|^2) = |a_2|^2 = \cos(2\eta)^2$ above. From the opposite monotonicities of $\cos(2\eta)$ and $\sin(2\eta)$ over $\eta \in [0, \pi/4]$, we see that the maximum over all η is achieved for $\eta^* \stackrel{\text{def}}{=} \frac{1}{2}\cos^{-1}\left(\sqrt{\frac{\sin^2(\nu)}{\sin^2(\nu)+1}}\right)$. In other words, we obtain $\max_\eta \min(|a_2|^2, |a_3|^2, |a_4|^2) = \frac{\sin^2(\nu)}{\sin^2(\nu)+1}$. As this function is increasing for $\nu \in (0, \pi/4]$, it has a

maximum in $\nu^* = \pi/4$. In this case, we get

$$\begin{aligned}
\nu^* &= \frac{\pi}{4} \\
\eta^* &= \frac{1}{2}\cos^{-1}\left(\frac{1}{\sqrt{3}}\right) \\
&= \frac{1}{4}\cos^{-1}(-\frac{1}{3}) \\
&= \frac{1}{2}\tan^{-1}(\sqrt{2})
\end{aligned}$$

E.2 Illustration of How the Optimize the Basis Orientation Before PDL Compression

	Before PDL	After PDL	Notes
DP-QPSK			The polytope $OABC$ is the tetrahedron of reference. For DP-QPSK, no PDL, it corresponds to the object delimited by the canonical orthonormal basis. Weak axis is the vertical dashed line and the strong axes are in the represented plane. While $AB = AC = BC$ with no PDL, only the distances AB and BC remains equal after PDL scaling. Assuming that the minimum distance $d\sqrt{1-\gamma}$ is met for the (not represented) $\Re x_1$ dimension, OA, OB and OC correspond to the $\Im x_2, \Im x_1$ and $\Re x_2$ dimensions, resp.
SB-QPSK			SB consists in a rotation with axis OC by $\theta = \pi/4$ (rotation of the imaginary parts or equivalently rotating the real/imag. parts by resp. $\pm\pi/4/2$. After PDL scaling, the rotation enables to have points A and B farther apart than with no rotation. The object is represented on purpose with slightly modified coordinates for a better visual understanding of the 3D figure.
NSB-QPSK			NSB is more subtle. It is equivalent to mixing $\Re x_2$ and $\Im x_1$; and $\Im x_1$ and $\Im x_2$. Before PDL, the tetrahedron is placed such that one face is parallel to the expanding plane. This way, the three distances AB, AC and BC remains equal after PDL hence optimizing the distribution of distances and A, B and C go apart from each other. The reference tetrahedron has also been stretched for better visualization.

Appendix F

Condensed French Version

F.1 Introduction

La demande de débit toujours plus importante année après année nécessite une infrastructure internet fiable qui est en pratique basée sur les réseaux optiques. Ceux-ci permettent de véhiculer de larges quantités de données. Dans ces réseaux qui sont aujourd'hui multiplexés en polarisation, certains éléments comme les routeurs ou les amplificateurs peuvent ne pas avoir la même atténuation ou gain selon l'état de polarisation incident. Cet effet, appelé perte dépendant de la polarisation ou PDL en anglais réduit la capacité du réseau. Dans ce manuscrit nous nous proposons de l'étudier en dérivant un modèle de canal précis, puis de donner des solutions, tant au transmetteur qu'au receveur, pour diminuer cette perte de capacité. Après avoir dérivé ces formats de modulation robustes à la PDL, nous présentons une validation de ceux-ci en simulation puis en laboratoire.

F.2 Modélisation d'un canal optique avec disparité de gain entre polarisations

Dans cette thèse nous nous focalisons sur un effet dit de perte dépendant de la polarisation, abrégé PDL pour Polarization Dependent Loss en anglais. Cet effet apparaît lorsqu'un signal traverse un élément optique discret, présentant une disparité de gain entre les deux polarisations du champ électromagnétique. Dans cette section, à partir d'une modélisation au niveau élémentaire de l'effet PDL, nous présentons le modèle d'un canal optique contenant des éléments PDL concaténés.

F.2.1 Éléments avec perte dépendant de la polarisation : un effet non unitaire

Lorsqu'un symbole $X = (x_1, x_2)^T \in \mathbb{C}^2$ multiplexé en polarisation traverse un élément PDL, ce dernier se comporte différemment selon l'état de polarisation incident. Si l'on nomme g_1 et g_2 les gains associés respectivement à la première et deuxième polarisation, en général ces deux gains diffèrent. On appelle cette différence de gain la valeur de PDL et on l'exprime usuellement en décibel (dB).

Definition 12 *La PDL d'un élément est définie comme le ratio λ entre les deux gains au carré associés à cet élément, ou en dB*

$$\Lambda = 10 \log_{10} \lambda = 10 \log_{10} \frac{g_1^2}{g_2^2} \tag{F.1}$$

Cette disparité de gain est usuellement prise en compte dans le formalisme de Jones par une matrice $D = \mathrm{diag}\{g_1, g_2\}$. En toute généralité, le signal X entre avec une orientation quelconque sur cette élément, et en ressort avec une autre orientation quelconque par rapport aux axes principaux de référence. Ainsi, on modélise un élément PDL par la matrice de transfert

$$H = UDV \tag{F.2}$$

où D est la matrice de déséquilibre de gain et U et V sont deux matrices de $\mathrm{U}(2)$, l'ensemble des matrices unitaires complexes d'ordre 2. Les deux matrices sont prises aléatoirement et de manière uniforme dans $\mathrm{U}(2)$ en respectant la mesure de Haar [45].

F.2.2 Modélisation d'un lien optique avec des éléments PDL concaténés

Un lien optique où l'effet PDL est prépondérant (où les effets de dispersion et de non-linéarité sont négligés) peut se modéliser par une succession d'éléments PDL de matrice de transfert H_i et de sources ajoutant chacune un bruit $Z_i \sim \mathcal{CN}(0, I_2)$. Les sources de bruit sont dues aux amplificateurs optiques, typiquement de la fibre dopée à l'Erbium (EDFA), qui permettent de relever la puissance du signal après transmission dans des kilomètres de fibre.

Le signal sortant de ce canal succession d'éléments PDL et d'amplificateurs s'écrit

$$Y = Z_n + H_n(Z_{n-1} + H_{n-1}(\cdots(Z_2 + H_2(Z_1 + H_1 X))\cdots) \tag{F.3}$$

$$= \left[\prod_{i=n}^{1} H_i\right] X + \left[\sum_{i=1}^{n} \left(\Pi_{j=n+1}^{i+1} H_j\right) Z_i\right] \tag{F.4}$$

$$= HX + Z \tag{F.5}$$

avec la convention $H_{n+1} = I_2$.

Cette équation suggère qu'un canal PDL se comporte comme un canal classique 2×2 avec un canal résultant $H = \prod_i H_i$ et un bruit $Z \sim \mathcal{CN}(0, K)$ avec $K = \sum_{i=1}^{n} P_i P_i^\dagger$ où P_i est le facteur devant Z_i. La matrice du canal en Eq. F.5, comme toute matrice de $\mathbb{C}^{2\times 2}$, peut se factoriser en faisant une décomposition en valeurs singulières

$$H = UDV \tag{F.6}$$

où cette fois D contient le déséquilibre de gain *total* du lien optique, et U et V sont encore dans $\mathrm{U}(2)$ et aléatoires. Cette forme est très similaire à celle modélisant l'élément PDL, mais ici elle modélise la matrice du canal résultant de toute la concaténation.

Dans la section F.3, nous étudierons la capacité associée à ce canal PDL, à partir de la précédente équation que nous pouvons simplifier. Nous considérons désormais que $K = I_2$. Pour commencer, la matrice U à gauche n'influe pas sur les performances du canal. En effet, en tant que matrice unitaire, l'inverser en multipliant à gauche par U^{-1} ne change pas les propriétés du bruit. Ainsi, le signal en sortie de canal peut être redéfinie en $Y = DVX + Z$. La matrice V peut se décomposer canoniquement en $e^{i\phi} B_\delta R_\alpha B_\beta$ où les matrices B et R représentent des rotations de signal respectivement complexes et réelle et sont définies par

$$B_\beta = \begin{pmatrix} e^{i\beta} & 0 \\ 0 & e^{-i\beta} \end{pmatrix} \text{ et } R_\alpha = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \tag{F.7}$$

avec ϕ, δ, α et β trois angles dans $[-\pi, \pi[$.

Le scalaire $e^{i\phi}$ peut être omis puisque il s'agit d'une phase pouvant être inversée sans changer les propriétés du bruit. Les deux matrices diagonales B_δ et D peuvent être commutées et B_δ peut finalement être inversée comme U sans influence sur les performances du lien.

Finalement, en modélisant la matrice de gain en $D_\gamma = \mathrm{diag}\left\{\sqrt{1+\gamma}, \sqrt{1-\gamma}\right\}$ (conservant ainsi l'énergie totale), on modélise un canal PDL en $Y = HX + Z$ avec H retenant uniquement trois paramètres

$$H = D_\gamma R_\alpha B_\beta \tag{F.8}$$

et Z suit un processus $\mathcal{CN}(0, I_2)$. Ce modèle est appelé *canal PDL lumped*. Parfois, la matrice B_β est supprimée car elle ne modifie pas beaucoup les performances du canal, on parle alors de *canal PDL simple*.

F.2.3 Statistiques d'un lien avec éléments PDL concaténés

Dans cette section nous étudions les lois de distribution des différents paramètres introduits précédemment. Les matrices U et V sont tirées uniformément dans $\mathrm{U}(2)$. Pour respecter [46] la mesure de Haar et le volume élémentaire $dV = \sin^2(\alpha)\mathrm{d}\phi\mathrm{d}\delta\mathrm{d}\alpha\mathrm{d}\beta$ associé aux quatre angles de leur décomposition canonique, les trois angles ϕ, δ et β suivent des lois uniformes $\mathcal{U}([-\pi, \pi[)$. L'angle α suit une loi de forme circulaire définie par $p(\alpha) = \sin(2\alpha)$.

La loi de distribution du paramètre de gain γ est présentée dans [31]. Le ratio défini en Eq. F.1 suit un processus de Maxwell représenté dans la Fig. F.1 Une bonne approximation de la distribution du ratio est obtenue lorsque l'on simule un canal avec matrice de transfert

$$H = aG + bI_2 \tag{F.9}$$

où G est une matrice avec des coefficients suivant une distribution $\mathcal{CN}(0, 1)$, et a et b sont deux réels. Elle est représentée dans la Fig. F.1. Cette approximation alternative a l'avantage de ne pas avoir à simuler plusieurs éléments PDL concaténés. Seulement un tirage de H défini dans Eq. F.9 permet d'avoir un canal émulant un canal PDL, dont la moyenne peut être arbitrairement contrôlée. En effet, le ratio en dB des valeurs singulières au carré Λ d'un canal tiré de Eq. F.9 vérifie $E[\Lambda] = 6\ln(10) \times a/b$.

F.3 Limites de communications d'un canal avec PDL

À partir de la modélisation de la section F.2, nous étudions maintenant les limites de communication associées à un canal avec PDL.

F.3.1 Capacités associés un canal PDL

La capacité d'un système MIMO est classiquement calculable par maximisation de l'entropie [40] en utilisant une entrée continue suivant une distribution Gaussienne. Pour un canal connu et déterminé

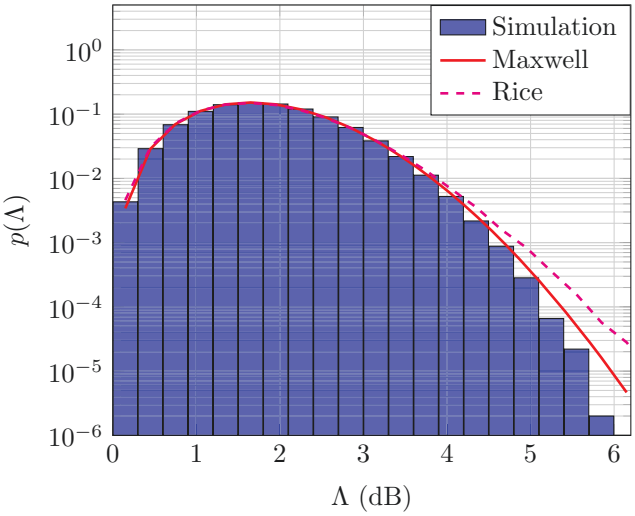


FIGURE F.1 – Distribution des ratios de gain d'un canal PDL de 25 étages avec éléments de 0.4dB de PDL ainsi qu'un ajustement de courbe de type Maxwell (analytique) et Rice (émulé à partir de la même valeur de PDL moyenne).

de matrice H, cette capacité peut s'exprimer comme

$$C = \log_2 \det\left(I_2 + \frac{\rho}{2} H H^\dagger\right) \tag{F.10}$$

$$= 2\log_2\left(1 + \frac{\rho}{2}\right) + \log_2\left(1 - \gamma^2 \frac{\rho^2}{(2+\rho)^2}\right) \tag{F.11}$$

où ρ est le rapport de la puissance du signal sur celle du bruit (SNR).

Dans Eq. F.11, le premier terme correspond à la capacité usuelle d'un système MIMO 2×2 sans PDL. Le deuxième terme, négatif, capture la perte de capacité associée à la présence de PDL dans le canal. Les capacités d'un canal PDL avec différentes valeurs de Λ sont représentées dans la Fig. F.2. Il est à noter que cette formule ne dépend pas des paramètres angulaires introduits précédemment.

Puisque les entrées continues sont irréalisables en pratique, nous dérivons maintenant les capacités associées à l'utilisation de modulations usuelles discrètes de type QAM multiplexé en polarisation. Dans ce cas, il n'existe pas de formule exacte et des simulations Monte-Carlo sont nécessaires pour les calculer. Typiquement, pour une modulation DP-M-QAM (c'est-à-dire M-QAM sur les deux polarisations), outre la saturation à l'entropie de la source $m = \log_2 M$, une dépendance en l'état de polarisation apparaît, et ce pour n'importe quelle valeur de M. Elle est représentée dans

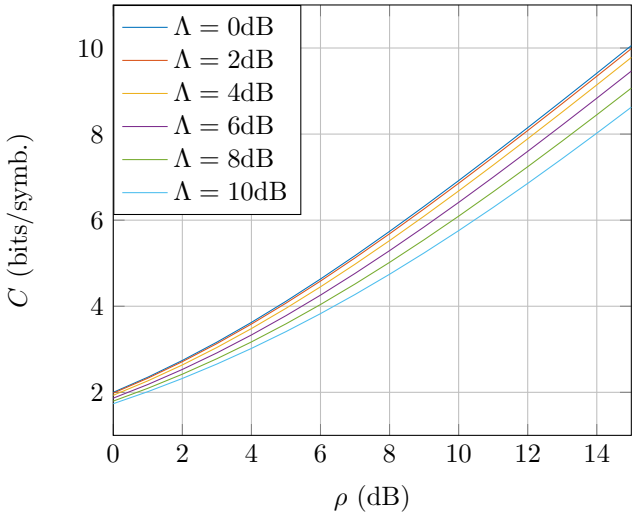


FIGURE F.2 – Capacité du canal PDL en fonction du SNR pour Λ allant de 0 à 10dB de PDL.

la Fig. F.3. Typiquement, pour une PDL $\Lambda = 6$dB, les états de polarisation offrant les meilleures capacités sont atteintes pour l'angle $\alpha = \pi/4$ et le pire cas est rencontré en $\alpha = 0$, modulo $\pi/2$. La dépendance en β est moins importante et les pire et meilleure capacités sont atteintes respectivement pour $\beta = 0$ et $\pi/8$, modulo $\pi/4$.

Cette dépendance en angle est problématique en terme de communication sur un canal PDL. En effet, la recherche actuelle vise à réduire cette dépendance, et en particulier à rehausser la capacité du pire cas. Ce profil de capacité peut aussi être analysé dans un formalisme de distance Euclidienne et nous alternerons régulièrement entre une étude de cette métrique avec celle de la capacité du lien. Par exemple, le profil de la distance minimum en fonction des angles du canal observe les mêmes variations que celle de la capacité.

F.3.2 Impact de l'architecture de receveur sur la capacité du système

En plus de la perte en capacité dû à la PDL, et de la perte additionnelle due à l'utilisation de modulations discrètes, s'ajoute potentiellement une troisième perte de capacité associée cette fois à l'architecture du receveur optique choisie. En effet, les capacités dérivées plus tôt sont valables tant que le décodage de symbole et de bit après canal sont faites de manière optimale au sens de la théorie de l'information. Cela implique de les décoder de manière récursive, en faisant un retour par exemple après le décodage de bit vers un nouveau décodage des symboles QAM bruités. En pratique cependant, ces deux décodages sont réalisés séquentiellement, en une seule passe.

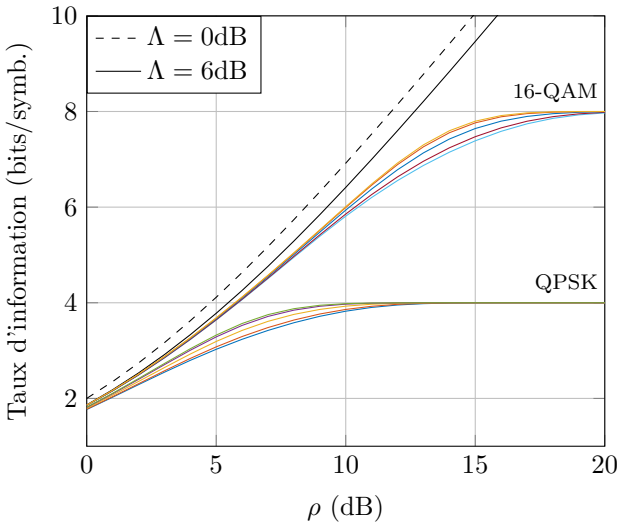


FIGURE F.3 – Taux d'information pour les modulations QPSK et 16-QAM pour une PDL de $\Lambda = 6$dB.

Ce choix d'architecture séquentielle se traduit par une perte de capacité représentée en Fig. F.4, ces valeurs réduites de capacités sont appelées capacités BICM pour Bit-Interleaved Coded Modulation en anglais. Les capacités BICM sont toujours inférieures à celles du canal calculées dans la section précédente (appelées a posteriori capacité CM pour Coded Modulation). En plus de cela, alors que l'ordre des courbes CM associé à chaque état de polarisation était le même pour tout SNR, ce n'est plus le cas pour les capacités BICM. Les états de polarisation donnant la meilleure capacité BICM dans une plage de SNR deviennent le pire cas dans un autre régime de SNR. Cela impose une attention particulière sur le paramétrage du lien optique car selon le taux de codage recherché, les conclusions sur les capacités de canal peuvent différer selon la plage de SNR visée. Ces pertes de capacités se traduisent évidemment en des pertes similaires de seuils FEC à partir desquels les bits sont décodés sans erreur, et s'illustrent de façon équivalente en une perte de BER à SNR identique comme représenté en Fig. F.5.

F.4 Formats de modulation optimaux contre la PDL

Dans cette section, nous nous reconcentrons sur les capacités CM d'un canal PDL et proposons des modulations pour élever la capacité du pire cas.

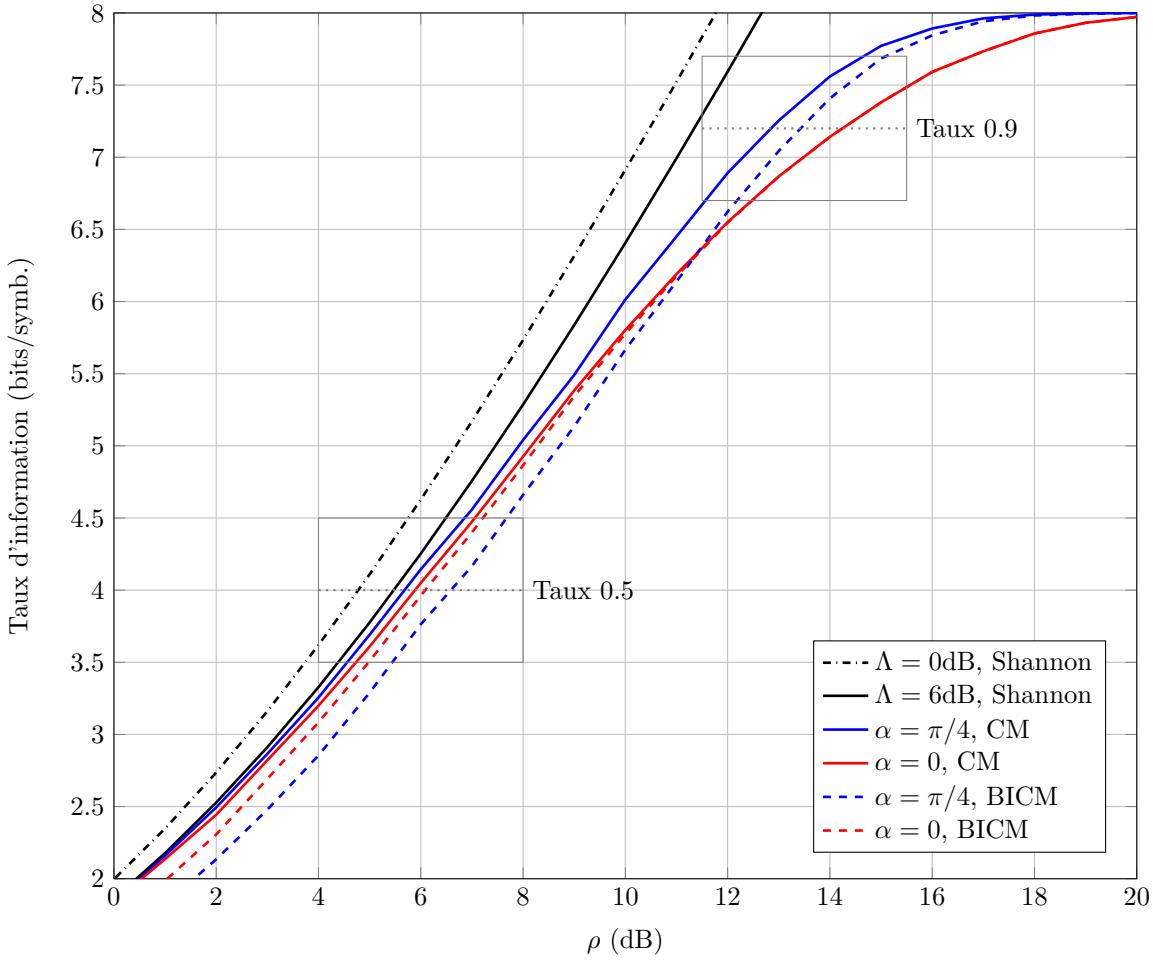


FIGURE F.4 – Taux d'information BICM et CM pour la modulation 16-QAM pour une PDL de $\Lambda = 6$dB.

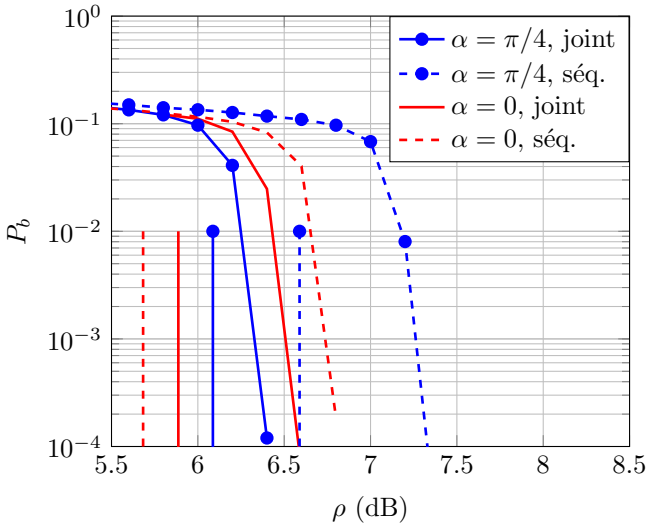


FIGURE F.5 – Taux d'erreur en fonction du SNR pour plusieurs angle α et avec un codage joint ou séquentiel.

F.4.1 Le format "Spatially Balanced" adapté à la disparité de gain

Pour commencer, nous dérivons un premier format de modulation basé sur l'étude du canal PDL *simple* avec seulement un paramètre angulaire $H = D_\gamma R_\alpha$. Nous pouvons observer que ce canal est à valeurs réelles, ainsi nous avons

$$Y = D_\gamma R_\alpha X + Z \equiv \begin{cases} \Re Y = D_\gamma R_\alpha \Re X + \Re Z \\ \Im Y = D_\gamma R_\alpha \Im X + \Im Z \end{cases} \tag{F.12}$$

Nous notons que les deux sous-canaux à droite sont identiques et indépendants, en particulier les parties réelle et complexe du signal DP-QAM voient le canal avec un même angle d'incidence. Ainsi, nous proposons d'encoder le signal en tournant artificiellement une des deux parties, par exemple la partie imaginaire, d'un angle η. Désormais, les parties réelles voient un angle de canal α, mais les parties imaginaires entrent alors dans le canal avec un angle d'incidence $\alpha + \eta$. D'après les profils de capacité présentés en bleu en Fig. F.6 qui sont périodiques et maximums en $\pi/4$, il apparaît judicieux d'affecter $\eta = \pi/4$. En effet, lorsque qu'une partie complexe du signal verra un mauvais canal (proche de $\alpha = 0$), alors l'autre partie verra un bon canal (proche de $\pi/4$), résultant in fine en un moyennage des amplitudes de capacité du canal PDL représenté en rouge sur la Fig. F.6. Nous appelons ce format de modulation le "Spatially Balanced" (SB).

Cependant cette valeur de $\eta = \pi/4$ n'est pas universelle. On peut déterminer la valeur optimale

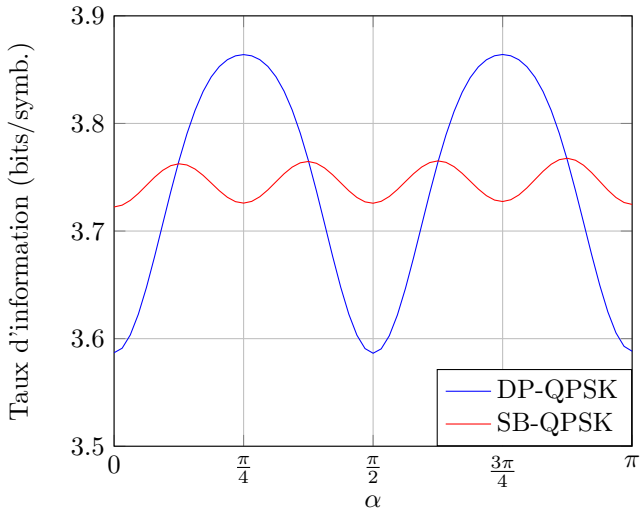


FIGURE F.6 – Taux d'information à un SNR de 8dB du DP-QPSK et du SB-QPSK pour $\Lambda = 6$dB en fonction de α.

de η comme étant celle de α^*, la valeur d'angle de canal pour lequel la capacité est maximale. Puisque le calcul de capacité est lourd et difficilement maniable, nous nous proposons plutôt d'étudier les distances Euclidiennes associées à la constellation de points déformée après canal. En effet, pour de forts SNR, la capacité est entièrement déterminée par la distance minimum [50].

La distance minimum de la constellation en fonction de l'angle α est l'enveloppe basse de toutes les distances du dictionnaire des symboles DP-QAM après transmission, et dépend de la valeur de PDL. Nous nous concentrons pour commencer sur la constellation DP-QPSK et puisque les deux parties complexes subissent la même déformation, nous regardons uniquement celle d'une partie complexe, qui correspond à un carré $ABCD$ formé à partir des quatre points possibles. Pour toute valeur de PDL, l'angle $\alpha = 0$ donne universellement la plus petite distance AB possible parallèle à l'axe de compression de l'effet PDL. Par opposition, pour de faibles valeurs de PDL, l'angle $\alpha = \pi/4$ donne toujours la meilleure distance minimum, mais à partir de $\Lambda_1 \approx 4.77$dB la diagonale du carré BD peut s'effondrer et devenir plus petite que AB. Ces différentes déformations sont représentées dans le tableau F.1.

L'angle optimal α^* qui donne la plus grand distance minimum est celui où les deux BD et AB sont égales. Il peut être trouvé par des considérations géométriques et est finalement défini par

$$\alpha^*(\lambda) = \arctan(\lambda - 1 - \sqrt{(\lambda-1)^2 - \lambda}) \tag{F.13}$$

Si une modulation d'un ordre M supérieur à 2 est utilisée, alors d'autres compressions de

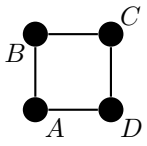
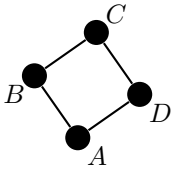
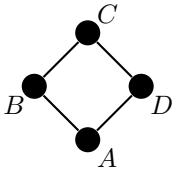
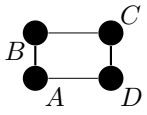
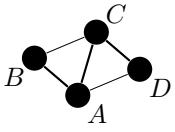
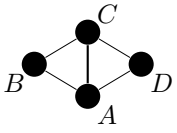
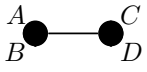
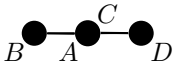
	$\alpha = 0$	$\alpha = \alpha^*(\Lambda)$	$\alpha = \pi/4$
$\Lambda = 0$dB			
$\Lambda = 6$dB			
$\Lambda \to \infty$			

TABLE F.1 – Distorsion du carré formé par les points 2-PAM^2 en présence de différentes valeurs de PDL et différents angles. Excepté en $\Lambda = 0$dB, α^* est l'angle pour lequel la distance minimum est maximale.

distances arrivent et l'angle peut être donné par une expression similaire, découlant d'une autre égalité de distance. Cette angle, donnant la plus grande distance minimum, et représenté dans Fig. F.7 jusqu'à $M = 8$, est à affecter à η pour que la modulation SB soit optimisée.

F.4.2 Dérivation de la modulation optimale pour la PDL : le "New Spatially Balanced"

Nous dérivons maintenant l'encodage unitaire en 4 dimensions optimal pour le canal PDL. Par optimalité est entendu l'encodage qui permet d'augmenter le plus la pire capacité. Une première classe d'encodage imaginable est l'emploi de transformation U de SU (2). Cependant, puisque l'état de canal est une matrice de ce même ensemble, alors le pire cas n'est pas augmenté. En effet, si avant codage le pire cas est en U_0, alors il sera encore rencontré après codage en $U_0 U^{-1}$.

Nous nous concentrons donc plus généralement sur les modulations unitaires en 4 dimensions, autrement dit les encodages $G \in$ SO (4). Chaque matrice de ce groupe de dimension 6 peut se factoriser canoniquement en deux matrices dites isoclinique gauche G_L et droite G_R [71]. La partie isoclinique droite, représentée dans $\mathbb{C}$, équivaut à une matrice de l'ensemble SU (2) (de dimension 3, inclus dans U (2)) et ne peut pas augmenter la pire capacité comme évoqué précédemment.

Nous nous focalisons donc sur les transformations de SO (4) uniquement isocliniques gauches qui peuvent se définir à l'aide de trois angles d'Euler η, ν et $\phi \in [-\pi, \pi[$ dont le dernier correspond à une phase commune aux deux polarisations et est omise. Après simplification, la matrice d'encodage

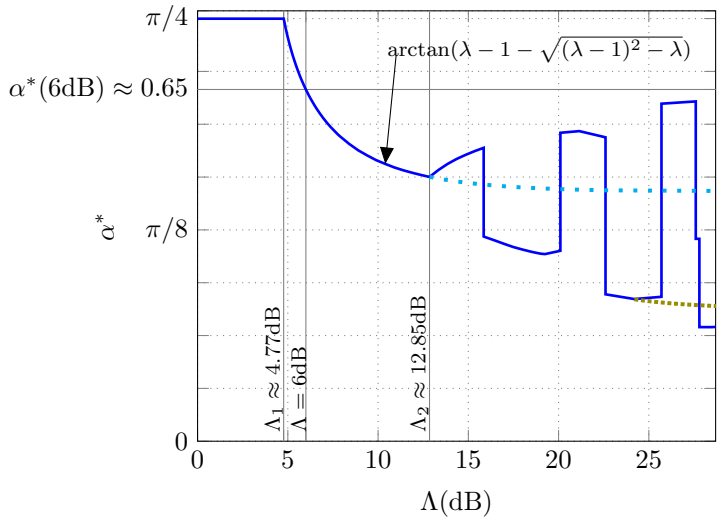


FIGURE F.7 – Angle optimal α^* en fonction de la PDL. Les courbes en pointillés sont dans les limites respectives de $M = 2$ et 4.

peut être représentée dans $\mathbb{C}$ en la transformation

$$f_{\eta,\nu}(X) = \cos\eta X + e^{i\nu} \sin\eta \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} X^* \tag{F.14}$$

ce qui permet d'introduire le théorème principal de cette thèse.

Théorème L'encodage unitaire en de QAM multiplexé en polarisation qui optimise la pire capacité sur tous les états de polarisation, pour de faibles valeurs de PDL et à un SNR suffisamment élevé est celui défini par $f_{\eta,\nu}$ avec les valeurs $(\eta^*, \nu^*) = (\frac{1}{2}\arctan\sqrt{2}, \pi/4)$. On appelle cet encodage le format *New Spatially Balanced* (NSB).

Cela peut s'observer si l'on trace le profil $w(\eta, \nu)$ de la pire capacité sur l'ensemble des états de polarisation, en fonction des angles η et ν de l'encodage. Avec une précision de $\pi/64$ près représenté en Fig. F.8, la plus grande valeur de $w(\eta, \nu)$ se trouve pour les valeurs $(\eta, \nu) = (5\pi/32, \pi/4)$ qui se trouve être la meilleure approximation, avec la finesse évoquée, de $(\frac{1}{2}\arctan\sqrt{2}, \pi/4)$.

Une autre approche, géométrique, permet d'observer que le couple proposé est optimal. Il correspond en effet aux angles de deux rotations qui place le réseau de points de la constellation QAM orientée de telle sorte que, pour la capacité la pire, la compression PDL écrase identiquement les distances au lieu de certaines d'entre elles.

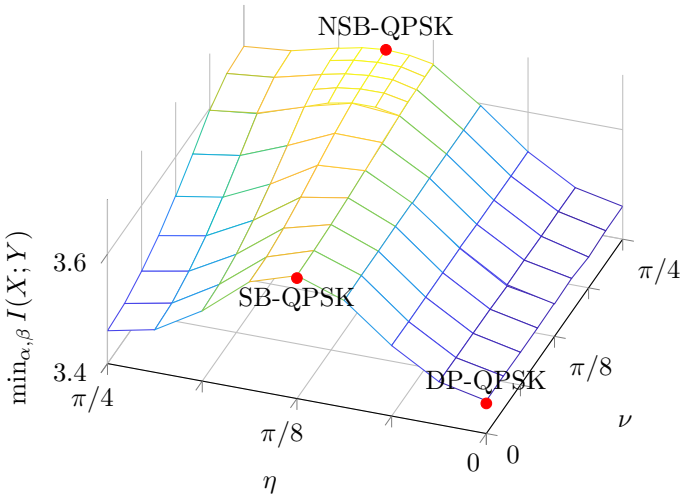


FIGURE F.8 – Evaluation numérique de la pire capacité sur tous les états de canal pour un encodage $f_{\eta,\,\nu}$ de DP-QPSK à une PDL de 6dB et un SNR de 8dB.

	DP-QPSK	SB-QPSK	NSB-QPSK
$d_{\min}^{(1)}$	16	8	8
$d_{\min}^{(2)}$	16	16	24
$d_{\min}^{(3)}$	(pas un côté)	8	(pas un côté)

TABLE F.2 – Comptage des 32 plus petites distances des modulations DP-, SB- et NSB-QPSK.

A proprement parler, la distance minimum du QAM multiplexé en polarisation ne peut pas augmenter, mais son nombre d'occurrences peut. La remarque précédente d'écrasement identique des distances porte sur la *deuxième* distance minimum noté $d_{\min}^{(2)}$ qui est maximisée pour le format NSB.

Comme illustré dans le tableau F.2, pour le cas du QPSK, le format SB n'a que 16 de ses côtés qui sont deuxièmes distances minimums là où le NSB en a un nombre maximal de 24. En se rappelant que le nombre de côtés d'un hypercube de dimension N est $N2^{N-1}$, les deuxièmes distances minimums ainsi que les 8 distances minimums se trouvent être les 32 côtés de l'hypercube en 4 dimensions. Pour le DP-QPSK, deux des directions de l'hypercube sont exactement alignées avec deux axes de compression de la PDL portant un nombre de distance minimum à 16. Pour les formats SB et NSB, seulement 8 de ces distances sont alignées avec un axe de compression. Pour le NSB, la deuxième distance est en plus optimisée par rapport au second axe de compression, leur projections sur le plan perpendiculaire à cet axe étant égales et maximales.

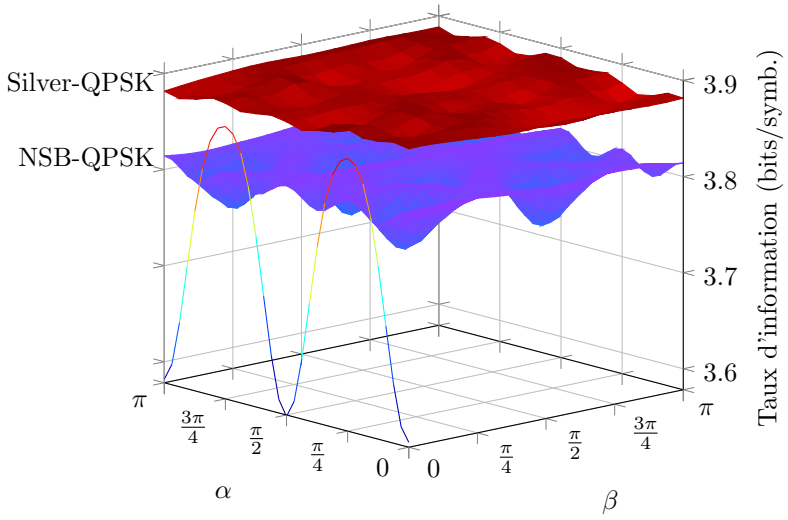


FIGURE F.9 – Taux d'information du NSB- et Silver-QPSK pour un SNR de 8dB et 6dB de PDL. La ligne en plusieurs couleurs représente l'amplitude du DP-QPSK pour avoir une référence.

F.4.3 Les modulations sur plusieurs temps-symboles pour augmenter la distance minimum

Les formats de modulation sur plusieurs temps-symboles permettent de s'affranchir de l'impossibilité d'augmenter la distance minimum de la constellation, toujours après canal. En effet, dans [36], il est déjà montré que l'encodage Silver permet de garantir la meilleure capacité quasiment pour tous les angles de canal, et ce par des considérations de distances Euclidiennes, pour des PDL usuelles inférieures à 6dB. La capacité du Silver est affichée dans la Fig. F.9 qui se trouve supérieure à celle du NSB-QPSK.

F.5 Validation pratique de modulations robustes à la PDL

Dans cette section nous explorons la performance des formats proposés précédemment, qui ont été élaborés sur un canal PDL "lumped".

F.5.1 Probabilités d'échec de communication sur un canal PDL

Dans cette section nous proposons l'étude de trois formats de modulation sur un lien optique avec PDL distribuée. On peut définir une mesure de confiance sur l'aptitude du canal à transmettre les

symboles encodés sans erreur. Plus précisément, on veut garantir que le taux d'information I du canal soit supérieur à une valeur déterminée I_{th} avec une probabilité $1 - a$

$$P(I < I_{\text{th}}) \leq a \tag{F.15}$$

Pour étudier la capacité avec PDL distribuée, il faut déjà notifier un effet relatif aux EDFA présent dans la ligne. En effet, ces amplificateurs peuvent fonctionner en mode *puissance constante*, induisant des pertes de SNR en bout de ligne. Ce mode compresse la proportion de signal par rapport au bruit pour garantir une puissance identique en sortie de chaque amplificateur. Même sans PDL, ce facteur de compression après l'amplificateur i s'élève à

$$\kappa_i = \sqrt{\frac{P_0}{P_{H_i Y_{i-1}} + 2}} \tag{F.16}$$

où P_0 est la puissance du signal au transmetteur, 2 est la puissance du bruit ajouté à l'amplificateur courant, et $P_{H_i Y_{i-1}}$ est la puissance reçue depuis l'étage précédent d'amplification. Cette formule permet de calculer la perte en SNR effectif $\rho - \rho^*$ représenté en Fig. F.10 entre le signal X et le bruit accumulé Z

$$P_{HX} = (\kappa^2)^n P_0 \text{ et } P_Z = 2 \sum_{i=1}^{n} (\kappa^2)^i \tag{F.17}$$

Pour un nombre d'étages de 25, cela induit des pertes de SNR de l'ordre de 0.5dB dans les régions d'opérations du QPSK (SNR faibles). Cela est moins critique pour le 16QAM qui vise des régions de puissance de signal plus importantes. Ces pertes sont à ajouter à celles intrinsèques à la PDL et se manifeste lorsque l'on recherche le SNR pour lequel une même condition $P(I < I_{\text{th}}) \leq a$ doit être respectée pour un nombre différent d'étages d'amplification. Cela est réprésenté en Fig. F.11. Comme attendu, le format NSB offre une relaxe de SNR d'environ 0.2dB par rapport au DP-QPSK pour 25 étages, quand le format Silver (sur deux temps-symboles) offre 1dB de relaxe de SNR.

F.5.2 Traitement de signal adapté aux modulations 4D

Dans l'optique de tester nos formats de modulation avec des mesures expérimentales, nous abordons dans cette partie un traitement de signal adapté. En effet, les formats proposés SB et NSB sont tous deux des modulations en 4 dimensions où l'information est encodées conjointement sur les deux polarisations. Cela est incompatible avec l'égaliseur communément utilisé dans les chaînes de

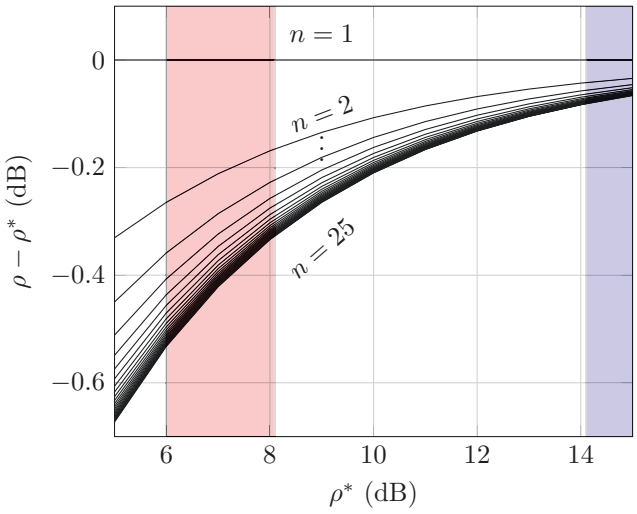


FIGURE F.10 – Compression $\rho - \rho^*$ du SNR ρ^* avec des EDFA en mode puissance constante et les régions d'opérations du QPSK (rouge) et du 16QAM (bleu).

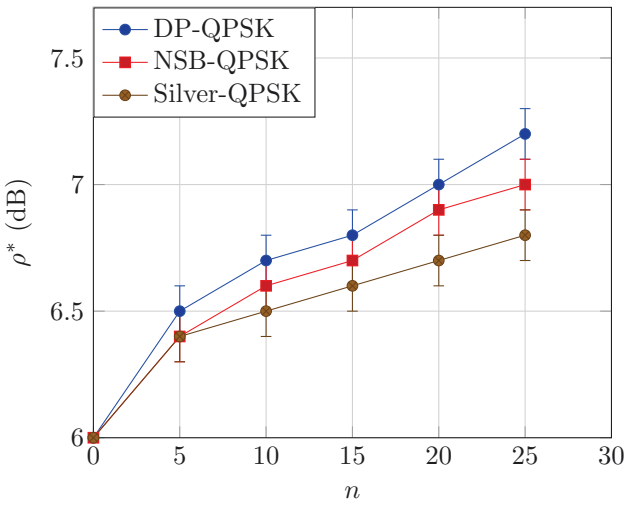


FIGURE F.11 – SNR ρ^* à partir duquel une condition sur le nombre de taux d'erreur inférieurs à un seuil est vérifié, pour trois formats de modulations et pour différents nombre d'étages.

traitement optiques, dérivé de l'algorithme à module constant [80] ou CMA en anglais.

Ainsi notre chaîne de traitement se base sur l'utilisation de symboles pilotes introduits dans les symboles d'information. Ils sont de trois sortes

1. des symboles (réels) qui servent à la synchronisation temporelle afin d'identifier les pilotes qui suivent
2. des symboles qui serviront à estimer le canal, dits CAZAC pour "Constant Amplitude Zero AutoCorrelation" en anglais, répétés un certain nombre n_{rep} de fois
3. des symboles de type QPSK, introduit périodiquement tout au long des symboles d'intérêt, et connu du receveur pour défaire le bruit de phase

Les symboles CAZAC permettent d'avoir une estimée $\hat{H}$ du canal [82] depuis les symboles reçus Y en utilisant

$$\hat{H} = \frac{1}{n_p} Y \mathbb{X}^\dagger \tag{F.18}$$

où $\mathbb{X}$ est l'écriture Toeplitz du vecteur envoyé X. A partir de cette estimée écrite en forme Toeplitz $\mathbb{H}$, éventuellement adaptable sur plusieurs temps-symboles, on obtient l'égaliseur d'erreur quadratique minimum, ou MMSE en anglais, contenu dans la matrice

$$\mathbb{W} = \left(\mathbb{H}^\dagger \mathbb{H} + \frac{1}{\rho} I_{n_t(l_h+l_w)} \right)^{-1} \mathbb{H}^\dagger \tag{F.19}$$

où n_t est la dimension du canal MIMO à l'entrée, et l_h et l_w sont des longueurs respectivement du canal et de l'égaliseur souhaité. La construction de l'égaliseur final consiste à sélectionner les indices des colonnes de $\mathbb{W}$ pour que les effets de bord soient négligeables.

F.5.3 Etude expérimentale des formats proposés sur un canal PDL

Le format SB

Avant d'effectuer des mesures expérimentales avec le format proposé SB, nous proposons d'étudier son comportement en simulation en ajoutant des effets de canal non considérés lors de la construction du format. Notamment, nous ajoutons un laser avec un bruit de phase de plusieurs centaines de kilohertz, tant à l'émission qu'à la réception. Pour cette première expérience, nous choisissons d'utiliser la séquence CAZAC pour avoir une première estimée du canal afin de le passer à un bloc d'égalisation CMA non-adaptatif pour égaliser le signal. Les pilotes de phase sont ensuite utilisés

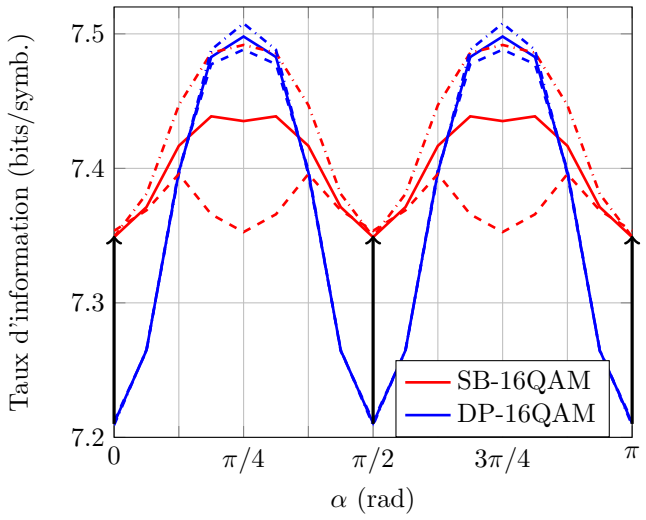


FIGURE F.12 – Taux d'information après implémentation du solution de traitement de signal pour utiliser le format SB-16QAM proposé et considérer ses gains par rapport au DP-16QAM, pour 6dB de PDL à un SNR de 15dB.

pour défaire le bruit de phase des lasers. Ainsi on trouve que la capacité pire est élevée d'un gain de 0.14 bits par utilisation de canal en utilisant le format SB, ce qui correspond à la flèche noire dans la Fig. F.12.

Dans le laboratoire, on monte un banc d'expérience représenté en Fig. F.13 en configuration homodyne où le laser, à 1545.72 nm est utilisé à la fois en émission et en réception. Cela permet de s'affranchir d'une correction de biais de fréquence. Nous utilisons un lien court avec un élément PDL de 6dB devant lequel est placé un brouilleur de polarisation pour pouvoir explorer différents états de canal, et succédé par une source de bruit. Le signal à 32 Gbauds est envoyé via un DAC à 88 GSa/s injecté dans un modulateur Mach-Zender I/Q et double polarisation. Du côté du récepteur, un filtre à 400 GHz est appliqué après le bruit pour ne retenir que la bande d'intérêt et mesurer le SNR optique avec un OSA. Le signal est ensuite injecté dans un receveur cohérent où il bat avec le laser d'émission, puis est récupéré sur un oscilloscope à 80 GSa/s. Nous effectuons donc 100 mesures que nous traitons, et classons en fonction de l'angle α d'incidence obtenu à partir de l'égaliseur. Nous observons alors des gains reportés en Fig. F.14 qui suivent ceux observés en simulation, et en particulier la pire capacité est augmentée de la même valeur de 0.14 bits par utilisation de canal.

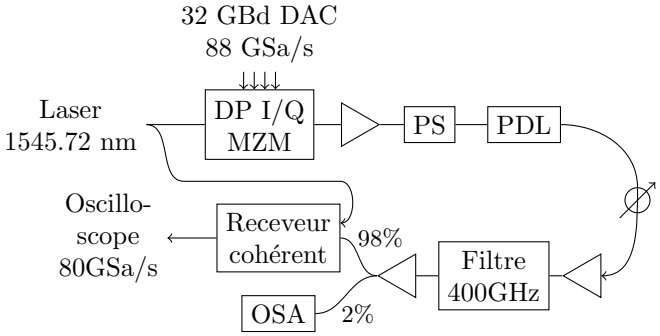


FIGURE F.13 – Banc d'expérience avec notamment un élément de PDL à 6dB précédé par un brouilleur de polarisation (PS) et d'un analyseur de spectre optique (OSA).

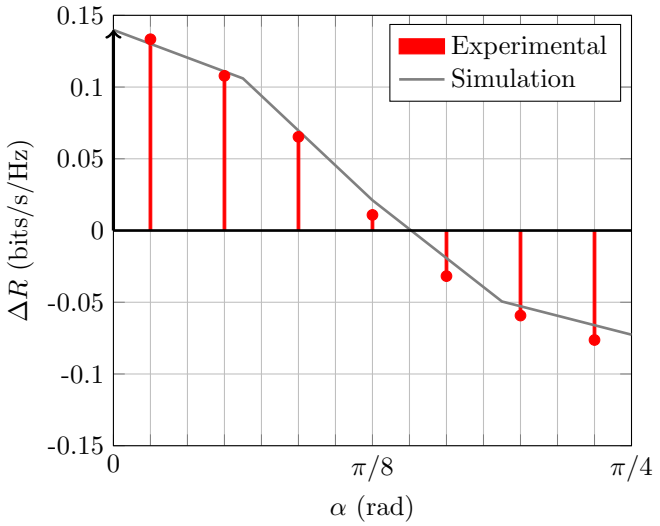


FIGURE F.14 – Gain de taux d'informations depuis les mesures expérimentales entre le SB-16QAM et le DP-16QAM à un SNR de 15dB.

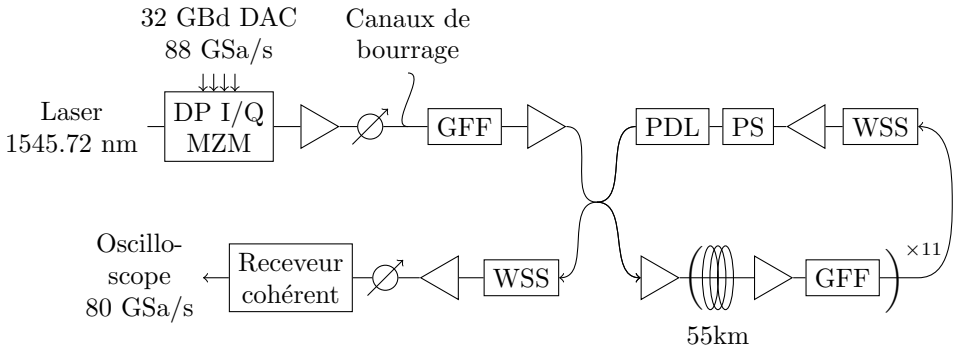


FIGURE F.15 – Banc d'expérience émulant un lien transpacifique où le signal se propage plusieurs fois dans une boucle à recirculation. Des filtres (GFF) et des WSS sont insérés dans la boucle afin d'assurer une bande C plate.

Le format NSB

De la même manière, nous voulons valider le format NSB-QPSK sur un canal expérimental avec PDL. Cette fois-ci, nous poussons l'étude pour le tester sur un lien transocéanique émulé par une boucle à recirculation réprésenté en Fig. F.15. A chaque tour dans cette boucle, le signal traverse un brouilleur de polarisation et un élément PDL pour simuler l'envoi à travers différents nœuds d'un réseau ou différents amplificateurs dans une ligne sous-marine. Le signal est enfin extrait après 15 tours à travers un WSS qui filtre, puis un mixeur cohérent et enfin un oscilloscope. Afin d'étalonner notre boucle, nous faisons parcourir le signal sans passer par la boucle pour émuler un canal "lumped" et valider notre format dans des conditions simples. Celui-ci montre des pénalités d'implémentation, dues à l'imperfection de l'étape de suppression de phase de laser. Une fois ces pénalités retirées dans un canal avec PDL, on observe que le format NSB augmente effectivement la capacité du pire cas par rapport à celle du DP-QPSK comme indiqué par la Fig. F.16. Sur le lien transocéanique, à 15 tours on observe également pour chaque réalisation que la polarisation la pire offre toujours une meilleure capacité pour le NSB-QPSK que pour le DP-QPSK ce qui montre que le NSB augmente bien la pire capacité du lien. De façon similaire, la distribution des valeurs des deux capacités par polarisation du NSB-QPSK est en moyenne et au minimum meilleure que celles du DP-QPSK.

F.6 Conclusion

Dans cette thèse, nous avons étudié en détail l'impact d'un déséquilibre de gain entre états de polarisation dans les communications optiques. Après avoir dérivé un modèle de canal et étudié

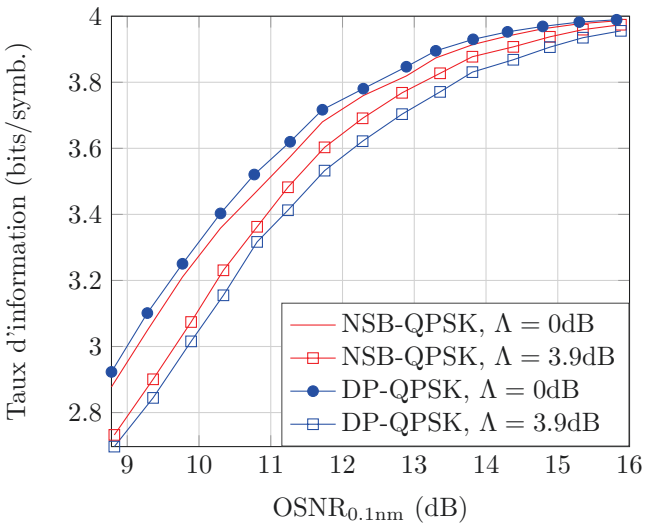


FIGURE F.16 – Taux d'information en fonction du SNR optique dans 0.1nm pour les deux modulations DB- et NSB-QPSK à des PDL nulle et de 3.1dB.

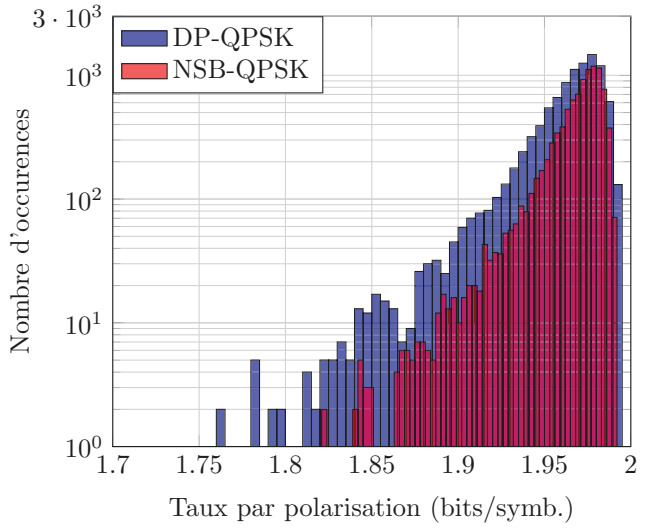


FIGURE F.17 – Distribution du taux d'information par polarisation pour les deux formats DP- et NSB-QPSK pour 5000 réalisations de canal PDL avec 15 éléments de PDL de 1.1dB orientés aléatoirement.

les capacités associées, nous avons proposé principalement deux formats de modulation qui sont robustes à la PDL, basés sur des considérations algébriques, et qui permettent de diminuer la perte en capacité. Nous avons en outre validé que la pire capacité augmente en simulation puis dans deux expériences lorsque ces formats sont utilisés plutôt que le DP-QAM communément utilisé.

Au delà de ces résultats, il reste à adapter une chaîne de transmission pour la rendre compatible avec notre solution, ainsi qu'étudier son effet sur les non linéarités du canal optique. Ces modulations sont compatibles avec l'utilisation de distribution de symboles non uniforme permettant de s'approcher de la capacité de Shannon. Elles peuvent aussi être utilisées dans n'importe quel canal ayant une disparité de gain similaire (perte dépendant du mode par exemple), et peuvent servir comme une première étape pour la dérivation de modulations avec davantage de dimensions spatiales ou temporelles.

Publications

Conference proceedings

1. A. Dumenil, E. Awwad, and C. Measson, "Tetrahedral Coding and Non-Unitary Resilience in Polarization-Multiplexed Lightwave Transmissions," accepted at the 2020 IEEE International Symposium on Information Theory (ISIT), Los Angeles, CA, USA, June 2020.
2. A. Dumenil, E. Awwad, and C. Measson, and Dylan Le Gac, "Experimental Demonstration of a 4D PDL-resilient Signaling for Long-haul Networks," 2020 Conference on Lasers and Electro-Optics (CLEO), San Jose, CA, USA, May 2020, paper SW3L.3.
3. A. Dumenil, E. Awwad, and C. Measson, "Rate Optimization Using SO(4) Transforms for PDL Mitigation," 2019 ECOC, Dublin, September 2019, W.2.D.4.
4. 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. Awarded Best Student Paper.
5. A. Dumenil, E. Awwad, and C. Measson, "Low-Complexity Polarization Coding for PDL-Resilience," *Eur. Conf. on Opt. Comm. (ECOC)*, Rome, Sep. 2018, paper Th1F.5.
6. A. Dumenil, E. Awwad, and C. Measson, "Polarization Dependent Loss: Fundamental Limits and How to Approach Them," *Signal Processing in Photonic Commun. Conf.,* New Orleans, Louisiana, USA, Jul. 2017.

Journal paper

1. A. Dumenil, E. Awwad, and C. Measson, "Tetrahedral Coding and Non-Unitary Resilience in Polarization-Multiplexed Lightwave Transmissions," IEEE Transactions on Communications, in redaction.

2. A. Dumenil, E. Awwad, and C. Measson, "PDL in Optical Links: a Model Analysis and a Demonstration of a PDL-Resilient Modulation," submitted to IEEE Journal of Lightwave Technology, received on January 15, 2020.

Granted patent

1. A. Dumenil, E. Awwad, and C. Méasson, "Low-Complexity Methods Using Dynamical Signaling For Mitigating Polarization Dependent Loss And Other Non-Unitary Losses In Optical Communications," European patent EP3404853, granted in September 2019.

Bibliography

[1] Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2017–2022 White Paper, accessible at https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white-paper-c11-738429.html

[2] R. J. Mears, L. Reekie, I. M. Jauncy, and D. N. Payne, "High-gain rare-earth-doped fiber amplifier at 1.54 μm," in Optical Fiber Communication, 1987 OSA Technical Digest Series (Optical Society of America, 1987), paper WI2.

[3] E. Desurvire, C. R. Giles, and J. R. Simpson, "Saturation-induced crosstalk in high-speed erbium-doped fiber amplifiers at l = 1.53 μm," in Optical Fiber Communication Conference, Vol. 5 of 1989 OSA Technical Digest Series (Optical Society of America, 1989), paper TUG7.

[4] R. A. Linke and A. H. Gnauck, "High-capacity coherent lightwave systems," in Journal of Lightwave Technology, vol. 6, no. 11, pp. 1750-1769, Nov. 1988.

[5] K. Kikuchi, "Coherent optical communication systems," *Optical Fiber Telecommunications VB: systems and networks*, 2008.

[6] G. Charlet, "Coherent detection associated with digital signal processing for fiber optics communication," *Comptes Rendus Physique*, 2008, vol. 9, no 9-10, p. 1012-1030.

[7] T. A. Eriksson et al., "Experimental investigation of a four-dimensional 256-ary lattice-based modulation format," 2015 Optical Fiber Communications Conference and Exhibition (OFC), Los Angeles, CA, 2015, pp. 1-3.

[8] Magnus Karlsson, Erik Agrell, "Four-Dimensional Optimized Constellations for Coherent Optical Transmission Systems," Proceedings in ECOC, Torino, Italy, September 2010

[9] T. Oyama, G. Huang, H. Nakashima, Y. Nomura, T. Takahara and T. Hoshida, "Low-Complexity, Low-PAPR Polarization-Time Code for PDL Mitigation," in Optical Fiber Conference (OFC), San Diego, 2019.

[10] G. Böcherer, F. Steiner and P. Schulte, "Bandwidth Efficient and Rate-Matched Low-Density Parity-Check Coded Modulation," in IEEE Transactions on Communications, vol. 63, no. 12, pp. 4651-4665, Dec. 2015.

[11] 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.

[12] P. Poggiolini, "The GN Model of Non-Linear Propagation in Uncompensated Coherent Optical Systems," in Journal of Lightwave Technology, vol. 30, no. 24, pp. 3857-3879, Dec.15, 2012.

[13] A. Ghazisaeidi et al., "Submarine Transmission Systems Using Digital Nonlinear Compensation and Adaptive Rate Forward Error Correction," in Journal of Lightwave Technology, vol. 34, no. 8, pp. 1886-1895, April, 2016.

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

[15] H. Bülow, "Experimental Demonstration of Optical Signal Detection Using Nonlinear Fourier Transform," in Journal of Lightwave Technology, vol. 33, no. 7, pp. 1433-1439, April, 2015.

[16] René-Jean Essiambre, Gerhard Kramer, Peter J. Winzer, Gerard J. Foschini, and Bernhard Goebel, "Capacity Limits of Optical Fiber Networks," J. Lightwave Technol. 28, 662-701 (2010)

[17] Peter J. Winzer and Gerard J. Foschini, "MIMO capacities and outage probabilities in spatially multiplexed optical transport systems," Opt. Express 19, 16680-16696 (2011)

[18] V. A. J. M. Sleiffer, P. Leoni, Y. Jung, H. Chen, M. Kuschnerov, S. U. Alam, M. Petrovich, F. Poletti, N. V. Wheeler, N. Baddela, J. Hayes, E. N. Fokoua, D. J. Richardson, L. Grüner-Nielsen, Y. Sun, and H. de Waardt, "Ultra-high Capacity Transmission with Few-mode Silica and Hollow-core Photonic Bandgap Fibers," in Optical Fiber Communication Conference, OSA Technical Digest (online) (Optical Society of America, 2014), paper Tu2J.3.

[19] E.M. Amhoud, G. Rekaya-Ben Othman and Y. Jaouën, "Capacity Enhancement of Few-Mode Fiber Transmission Systems impaired by MDL", applied sciences journal, vol. 8, pp. 326, April 2018.

[20] K. Benyahya et al., "Multiterabit Transmission Over OM2 Multimode Fiber With Wavelength and Mode Group Multiplexing and Direct Detection," in Journal of Lightwave Technology, vol. 36, no. 2, pp. 355-360, 15 Jan.15, 2018.

[21] D. A. A. Mello, H. Srinivas, K. Choutagunta and J. M. Kahn, "Impact of Polarization- and Mode-Dependent Gain on the Capacity of Ultra-Long-Haul Systems," in Journal of Lightwave Technology.

[22] A. Abouseif, G. Rekaya-Ben Othman and Y. Jaouën, "Channel Model and Optimal Core Scrambling for Multi-Core Fiber Transmission System", Optics Communications, August 2019

[23] R. Dar et al., "Cost-Optimized Submarine Cables Using Massive Spatial Parallelism," in Journal of Lightwave Technology, vol. 36, no. 18, pp. 3855-3865, Sept., 2018.

[24] J. Renaudier and A. Ghazisaeidi, "Scaling Capacity Growth of Fiber-Optic Transmission Systems Using 100+nm Ultra-Wideband Semiconductor Optical Amplifiers," in Journal of Lightwave Technology, vol. 37, no. 8, pp. 1831-1838, April, 2019.

[25] D. G. Foursa, "Performance and Impairments of Submarine Systems," 2019 Optical Fiber Communications Conference and Exhibition (OFC), San Diego, CA, USA, 2019, pp. 1-3.

[26] N. Rossi, P. Serena and A. Bononi, "Polarization-Dependent Loss Impact on Coherent Optical Systems in Presence of Fiber Nonlinearity," in IEEE Photonics Technology Letters, vol. 26, no. 4, pp. 334-337, Feb.15, 2014.

[27] P. Serena et al., "The Gaussian Noise Model Extended to Polarization Dependent Loss and its Application to Outage Probability Estimation," 2018 European Conference on Optical Communication (ECOC), Rome, 2018, pp. 1-3.

[28] H.-M. Chin, D. Charlton, A. Borowiec et al., "Probabilistic Design of Optical Transmission Systems," Journal of Lightwave Technology, vol. 35, no. 4, pp. 931-940, 2017.

[29] A. El Amari, N. Gisin, B. Perny, H. Zbinden and C. W. Zimmer, "Statistical prediction and experimental verification of concatenations of fiber optic components with polarization dependent loss," in Journal of Lightwave Technology, vol. 16, no. 3, pp. 332-339, March 1998.

[30] B. Huttner, C. Geiser and N. Gisin, "Polarization-induced distortions in optical fiber networks with polarization-mode dispersion and polarization-dependent losses," in IEEE Journal of Selected Topics in Quantum Electronics, vol. 6, no. 2, pp. 317-329, March-April 2000.

[31] A. Mecozzi, M. Shtaif, "The statistics of polarization-dependent loss in optical communication systems," IEEE Photon. Technol. Lett., vol. 14, pp. 313-315, Mar. 2002

[32] T. A. Strasser and J. L. Wagener, "Wavelength-Selective Switches for ROADM Applications," in IEEE Journal of Selected Topics in Quantum Electronics, vol. 16, no. 5, pp. 1150-1157, Sept.-Oct. 2010

[33] L. E. Nelson, C. Antonelli, A. Mecozzi, M. Birk, P. Magill, A. Schex, and L. Rapp, "Statistics of Polarization Dependent Loss in an Installed Long-Haul WDM System," Opt. Express 19, 6790-6796, 2011

[34] S. Mumtaz, G. Rekaya-Ben, and Y. Jaouën, "Space-time codes for optical fiber communication with polarization multiplexing," 2010 IEEE International Conference on Communications (ICC), 2010, pp. 1-5.

[35] O. Tirkkonen and A. Hottinen, "Improved MIMO performance with non-orthogonal space-time block codes," GLOBECOM'01. IEEE Global Telecommunications Conference (Cat. No.01CH37270), San Antonio, TX, 2001, pp. 1122-1126 vol.2.

[36] Elie Awwad, "Emerging Space-Time Coding Techniques for Optical Fiber Transmission Systems", Thesis, Telecom ParisTech, January 2015

[37] V. Tarokh, N. Seshadri and A. R. Calderbank, "Space-time codes for high data rate wireless communication: performance criterion and code construction," in IEEE Transactions on Information Theory, vol. 44, no. 2, pp. 744-765, March 1998.

[38] T. Tanimura, T. Oyama, H. Nakashima and J. C. Rasmussen, "PDL-tolerant signal generation by digital spectrum slicing and polarization control," 2015 Optical Fiber Communications Conference and Exhibition (OFC), Los Angeles, CA, 2015, pp. 1-3.

[39] G. Huang et al., "Efficient Polarization Dependent Loss Monitoring and Compensation in Coherent Transmission System," 2018 European Conference on Optical Communication (ECOC), Rome, 2018, pp. 1-3.

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

[41] L. E. Nelson and R. M. Jopson, "Introduction to polarization mode dispersion in optical systems," *Journal of Optical and Fiber Communications Reports*, Springer, New York, NY, 2004. pp. 1-33.

[42] J. N. Damask, Polarization Optics in Telecommunications. New York, NY, USA: Springer-Verlag, 2005

[43] T. Duthel, C. R. Fludger, J. Geyer, and C. Schulien, "Impact of Polarisation Dependent Loss on Coherent POLMUX-NRZ-DQPSK," in Optical Fiber Communication Conference/National Fiber Optic Engineers Conference, OSA Technical Digest (CD) (Optical Society of America, 2008), paper OThU5.

[44] C. Zhu, B. Song, B. Corcoran, L. Zhuang, and A. Lowery, "Improved polarization dependent loss tolerance for polarization multiplexed coherent optical systems by polarization pairwise coding," Opt. Express , vol. 23, no. 21, pp. 27434-27447, 2015.

[45] A. Edelman, "Eigenvalues and Condition Numbers of Random Matrices," Ph.D. dissertation, Dept. Mathematics, MIT, Cambridge, MA, USA, 1989

[46] M. Ozols, "How to generate a random unitary matrix," lecture notes from March 16, 2009, accessible online.

[47] A. Maaref and S. Aissa, "Joint and Marginal Eigenvalue Distributions of (Non)Central Complex Wishart Matrices and PDF-Based Approach for Characterizing the Capacity Statistics of MIMO Ricean and Rayleigh Fading Channels," in IEEE Transactions on Wireless Communications, vol. 6, no. 10, pp. 3607-3619, October 2007

[48] T. Cover and J. Thomas, *Elements of Information Theory*, New York; Wiley, 2006.

[49] C. E. Shannon, "A mathematical theory of communication," Bell System Tech. J., vol. 27, pp. 379–423, 623–656, July and Oct. 1948.

[50] A. Alvarado, F. Brännstroöm, E. Agrell and T. Koch, "High-SNR Asymptotics of Mutual Information for Discrete Constellations With Applications to BICM," in IEEE Transactions on Information Theory, vol. 60, no. 2, pp. 1061-1076, Feb. 2014.

[51] R. Urbanke and T. Richardson, *Modern Coding Theory*, New York; Wiley, 2008.

[52] I. Andriyanova and A. Graell i Amat, "Threshold Saturation for Nonbinary SC-LDPC Codes on the Binary Erasure Channel," in IEEE Transactions on Information Theory, vol. 62, no. 5, pp. 2622-2638, May 2016.

[53] D. Zibar, M. Piels, R. Jones and C. G. Schäeffer, "Machine Learning Techniques in Optical Communication," in Journal of Lightwave Technology, vol. 34, no. 6, pp. 1442-1452, 15 March, 2016.

[54] T. Benaddi, A. D. Yardi, C. Poulliat and I. Andriyanova, "Estimating the Maximum a Posteriori Threshold for Serially Concatenated Turbo Codes," 2019 IEEE International Symposium on Information Theory (ISIT), Paris, France, 2019, pp. 1347-1351.

[55] L. Schmalen, "Performance Metrics for Communication Systems with Forward Error Correction," 2018 European Conference on Optical Communication (ECOC), Rome, 2018, pp. 1-3.

[56] G. Caire, G. Taricco and E. Biglieri, "Bit-interleaved coded modulation," in IEEE Transactions on Information Theory, vol. 44, no. 3, pp. 927-946, May 1998.

[57] A. Martinez, A. Guillen i Fabregas, G. Caire and F. M. J. Willems, "Bit-Interleaved Coded Modulation Revisited: A Mismatched Decoding Perspective," in IEEE Transactions on Information Theory, vol. 55, no. 6, pp. 2756-2765, June 2009.

[58] S. Kudekar, C. Measson, T. Richardson and R. Urbanke, 'Threshold Saturation on BMS Channels via Spatial Coupling,' *IEEE Symp. on Turbo Codes* (2010)

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

[60] R. W. Hamming, "Error detecting and error correcting codes," in The Bell System Technical Journal, vol. 29, no. 2, pp. 147-160, April 1950.

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

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

[63] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," in IEEE Journal on Selected Areas in Communications, vol. 16, no. 8, pp. 1451-1458, Oct. 1998.

[64] 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.

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

[66] 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.

[67] E. Awwad, Y. Jaouen, and G. Rekaya, "Polarization-Time Coding for PDL Mitigation in Long-Haul PolMux OFDM Systems," Opt. Express 21, 22773-22790 (2013)

[68] H. El Gamal and M. O. Damen, "Universal space-time coding," in IEEE Transactions on Information Theory, vol. 49, no. 5, pp. 1097-1119, May 2003.

[69] M. Karlsson, "Four-dimensional Rotations in Coherent Optical Communications," Journal of Lightwave Technology, vol. 32, no. 6, pp. 1246-1257, 2014.

[70] 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.

[71] Alba Perez-Gracia, Federico Thomas, "On Cayley's Factorization of 4D Rotations and Applications", Advances in Applied Clifford Algebras, 2017, vol. 27, no 1, p. 523-538.

[72] F. Perez-Cruz, M. R. D. Rodrigues and S. Verdu, "MIMO Gaussian Channels With Arbitrary Inputs: Optimal Precoding and Power Allocation," in IEEE Transactions on Information Theory, vol. 56, no. 3, pp. 1070-1084, March 2010.

[73] D. Duyck, J. J. Boutros and M. Moeneclaey, "Precoding for Outage Probability Minimization on Block Fading Channels," in IEEE Transactions on Information Theory, vol. 59, no. 12, pp. 8250-8266, Dec. 2013.

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

[75] F. R. Kschischang and S. Pasupathy, "Optimal Nonuniform Signaling for Gaussian Channels," vol. 39, no. 3, pp. 913–929, May 1993.

[76] 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.

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

[78] M. Shtaif, "Performance degradation in coherent polarization multiplexed systems as a result of polarization dependent loss," *Optics Express 16*, 2008, pp. 13918-13932.

[79] J. Antona, A. C. Meseguer, and V. Letellier, "Transmission Systems with Constant Output Power Amplifiers at Low SNR Values: A Generalized Droop Model," 2019 Optical Fiber Communications Conference and Exhibition (OFC), San Diego, CA, USA, 2019, pp. 1-3

[80] Seb J. Savory, "Digital filters for coherent optical receivers," Opt. Express 16, 804-817 (2008)

[81] Jian Yang, J. -. Werner and G. A. Dumont, "The multimodulus blind equalization and its generalized algorithms," in IEEE Journal on Selected Areas in Communications, vol. 20, no. 5, pp. 997-1015, June 2002.

[82] M. Kuschnerov, M. Chouayakh, K. Piyawanno, B. Spinnler, E. de Man, P. Kainzmaier, M. S. Alfiad, A. Napoli, and B. Lankl, "Data-Aided Versus Blind Single-Carrier Coherent Receivers," in IEEE Photonics Journal, vol. 2, no. 3, pp. 387-403, June 2010.

[83] D. Chu, "Polyphase codes with good periodic correlation properties (Corresp.)," in IEEE Transactions on Information Theory, vol. 18, no. 4, pp. 531-532, July 1972.

[84] F. Pittalà, A. Mezghani, F. N. Hauske, Y. Ye, I. T. Monroy, and J. A. Nossek, "Efficient Training-Based Channel Estimation for Coherent Optical Communication Systems," in Advanced Photonics Congress, OSA Technical Digest (online) (Optical Society of America, 2012), paper SpTu3A.4.

[85] Hlaing Minn, V. K. Bhargava and K. B. Letaief, "A robust timing and frequency synchronization for OFDM systems," in IEEE Transactions on Wireless Communications, vol. 2, no. 4, pp. 822-839, July 2003.

[86] Maurizio Magarini, Arnaldo Spalvieri, Francesco Vacondio, Marco Bertolini, Marianna Pepe, and Giancarlo Gavioli, "Empirical modeling and simulation of phase noise in long-haul coherent optical transmission systems," Opt. Express 19, 22455-22461 (2011)

[87] A. Viterbi, "Nonlinear estimation of PSK-modulated carrier phase with application to burst digital transmission," in IEEE Transactions on Information Theory, vol. 29, no. 4, pp. 543-551, July 1983.

[88] A. Leven, N. Kaneda, U. Koc and Y. Chen, "Frequency Estimation in Intradyne Reception," in IEEE Photonics Technology Letters, vol. 19, no. 6, pp. 366-368, March15, 2007.

[89] I. Fatadin, D. Ives and S. J. Savory, "Blind Equalization and Carrier Phase Recovery in a 16-QAM Optical Coherent System," in Journal of Lightwave Technology, vol. 27, no. 15, pp. 3042-3049, Aug.1, 2009.

[90] C. Xie and L. F. Mollenauer, "Performance Degradation Induced by Polarization-Dependent Loss in Optical Fiber Transmission Systems With and Without Polarization-Mode Dispersion," Journal of Lightwave Technology, vol. 21, no. 9, 2003, pp. 1953-1957.

[91] C. Xie, "Polarization-dependent loss induced penalties in PDM-QPSK coherent optical communication systems," 2010 Conference on Optical Fiber Communication (OFC/NFOEC), collocated National Fiber Optic Engineers Conference, San Diego, CA, 2010, pp. 1-3.

[92] A. Iwaki et al., "Combined effect of DGD and PDL evaluated in field testbed for 128-Gb/s PDM-QPSK signal," 2012 17th Opto-Electronics and Communications Conference, Busan, 2012, pp. 35-36.

[93] T. Tanimura et al., "FPGA-based 112Gb/s coherent DP-QPSK receiver and multi-stage PMD-PDL emulator for fast evaluation of digital signal processing algorithms," 36th European Conference and Exhibition on Optical Communication, Torino, 2010, pp. 1-3.

[94] K. Ho, "Exact Model for Mode-Dependent Gains and Losses in Multimode Fiber," in Journal of Lightwave Technology, vol. 30, no. 23, pp. 3603-3609, Dec.1, 2012.

[95] A. Abouseif, G. R. Ben-Othman, and Y. Jaouën, "Multi-Core Fiber Channel Model and Core Dependent Loss Estimation," in Advanced Photonics 2018 (BGPP, IPR, NP, NOMA, Sensors, Networks, SPPCom, SOF), OSA Technical Digest (online) (Optical Society of America, 2018), paper SpW1G.3.

[96] Erik Agrell and Magnus Karlsson, "Power-Efficient Modulation Formats in Coherent Transmission Systems," J. Lightwave Technol. 27, 5115-5126 (2009)

[97] Amirhossein Ghazisaeidi, "Theory of Coherent WDM Systems Using In-Line Semiconductor Optical Amplifiers," J. Lightwave Technol. 37, 4188-4200 (2019)

[98] A. Ganti, A. Lapidoth, and E. , "Mismatched Decoding Revisited: General Alphabets, Channels with Memory, and the Wide-Band Limit," *IEEE Trans. on Inf. Theory*, vol. 46, pp. 2315–2328, Nov. 2000.

[99] D. Forney and G. Ungerboeck, 'Modulation and Coding for Linear Gaussian Channels,' *IEEE Trans. on Inf. Theory*, pp. vol. 44, 2384–2415, Oct. 1998.

[100] G. D. Forney and L. -. Wei, "Multidimensional constellations. I. Introduction, figures of merit, and generalized cross constellations," in IEEE Journal on Selected Areas in Communications, vol. 7, no. 6, pp. 877-892, Aug. 1989.

[101] R. G. Gallager, *Low Density Parity Check Codes, Monograph*, M.I.T. Press, Cambridge, MA, USA, 1963.

[102] E. Arikan, 'Channel polarization: A Method for Constructing Capacity-Achieving Codes for Symmetric Binary-Input Memoryless Channels', *IEEE Trans. on Inf. Theory* (2008)

[103] F. Kschischang, B. Frey and H.-A. Loeliger, 'Factor Graphs and the Sum-Product Algorithm,' *IEEE Trans. on Inf. Theory* (2001)

[104] G. Kramer, M. Yousefi, and F. Kschischang, "Upper Bound on the Capacity of a Cascade of Nonlinear and Noisy Channels," *arXiv:1503.07652v3*, Apr. 2015.

[105] L. H. Ozarow and A. D. Wyner, "On the Capacity of the Gaussian Channel with a Finite Number of Input Levels," in IEEE Transactions on Information Theory, vol. 36, no. 6, pp. 1426-1428, Nov. 1990.

[106] Joseph Boutros, Emanuele Viterbo, "Signal Space Diversity: A Power- and Bandwidth-Efficient Diversity Technique for the Rayleigh Fading Channel," IEEE Transactions on Information Theory, Vol. 44, No. 4, July 1998

[107] R. Elschner, F. Frey, C. Meuer et al., "Experimental demonstration of a format-flexible single-carrier coherent receiver using data-aided digital signal processing," Opt. Express 20, 28786-28791 (2012)

[108] Anton Andrusier, Eado Meron, Meir Feder, and Mark Shtaif, "Optical implementation of a space–time-trellis code for enhancing the tolerance of systems to polarization-dependent loss," Opt. Lett. 38, 118-120 (2013)

[109] D. Costello, etc *Error Control Coding: Fundamentals and Applications*, Academic Press, 2nd Edition, 2014

[110] Andreas Leven and Laurent Schmalen, "Status and Recent Advances on Forward Error Correction Technologies for Lightwave Systems," J. Lightwave Technol. 32, 2735-2750 (2014)

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

[112] 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.

[113] G. Caire, G. Taricco and E. Biglieri, "Bit-interleaved coded modulation," Proceedings of IEEE International Symposium on Information Theory, Ulm, Germany, 1997, pp. 96-.

[114] C.R. Johnson et al., "The core of FSE-CMA behavior theory," in *Unsupervised adaptive filtering: Blind deconvolution*, Vol. II, edition S. Haykin, pp. 13-122

[115] M. Mazur, J. Schröder, A. Lorences-Riesgo, T. Yoshida, M. Karlsson, and P. A. Andrekson, "Overhead-optimization of pilot-based digital signal processing for flexible high spectral efficiency transmission," Opt. Express 27, 24654-24669 (2019)

Titre: Des solutions à de nouveaux défis dans les réseaux optiques modernes avec perte dépendant de la polarisation

Mots clés : Codage, Modulation, Traitement de signal, PDL

Résumé : La demande de débit en augmentation constante requiert des canaux de communication fiables et solides. Les réseaux optiques terrestres ainsi que les liens sous-marins représentent le cœur de l'infrastructure des télécommunications au niveau mondial, et transportent des centaines de canaux fréquentiels modulés à haut débit. Ces liens se composent non seulement de kilomètres de fibre mais aussi de composants optiques comme les amplificateurs à fibre dopée à l'erbium (EDFA) pour relever régulièrement la puissance des signaux atténués, ou encore des commutateurs sélectifs en longueur d'onde (WSS) qui routent le signal vers leur destination. Dans cette thèse, nous nous intéressons à une pénalité spécifique qui survient dans ces systèmes et qui réduit le débit d'information dans une propagation optique. Les composants optiques discrets présentent typiquement une anisotropie en polarisation : un déséquilibre de perte (ou gain) apparaît entre les deux polarisations d'un signal optique multiplexé en polarisation. Cet effet non unitaire appelé pertes dépendantes de la polarisation (abrégé PDL en anglais) dégrade la qualité de transmission dans les systèmes optiques actuels et futurs.

Dans le cadre des transmissions multiplexés en polarisation, nous caractérisons la perte de capacité induite par la PDL. Pour commencer, des modèles de canal sont attentivement étudiés et deux approches sont présentées : un canal avec des éléments PDL distribués avec un bruit également distribué, ou alors un canal équivalent simplifié en un seul élément. Nous analysons les limites fondamentales de communication associées à ces modèles. Nous montrons en outre que la capacité d'un canal PDL dépend en pratique de l'orientation de l'état de polarisation du signal incident. Nous passons ensuite en revue l'état de l'art de schémas de modulation atténuant l'effet de PDL puis nous proposons deux nouveaux schémas de modulation multi-dimensionnels qui augmente la performance pire et moyenne du canal. Ces deux modulations sont des transformations unitaires de symboles M-QAM qui n'utilisent pas de degrés de liberté autres que les quatre déjà utilisés pour chaque longueur d'onde (encodage en phase et quadrature de phase sur les deux états de polarisation). Nous proposons succinctement une extension de ces premiers résultats à des communications optiques plus généralement multiplexés en dimension spatiale et impactées par des pertes dépendantes de mode, qui présentent un déséquilibre de gain identique. Au-delà de la construction de modulations au niveau du transmetteur, nous étudions la perte de performance en présence de PDL d'une chaîne de traitement de signal conventionnelle et séquentielle par comparaison avec une égalisation et décodage joint. Cette perte additionnelle de capacité résultant du traitement séquentiel non optimal est évaluée pour différents formats de modulations ainsi que différents points de fonctionnement. Enfin, nous exposons une validation expérimentale des deux schémas de modulation proposés, aussi bien sur un élément de PDL seul ou sur un canal PDL distribué, et montrons comme attendu une robustesse augmentée vis-à-vis de la PDL.

Title: Polarization Dependent Loss in Next-Generation Optical Networks: Challenges and Solutions

Keywords: Coding, Modulations, Signal processing, PDL

Abstract: Large amounts of ever-increasing global data traffic require sound and reliable communication channels. Optical terrestrial networks and submarine links are at the very heart of the global telecom infrastructure, and carry hundreds of frequency channels modulated at very high rates. These links include not only kilometers of fiber but also optical elements such as Erbium-Doped Fiber-Amplifiers (EDFA) to amplify periodically the attenuated signals and Wavelength Selective Switches (WSS) to route the signals to their assigned destinations. In this thesis, we explore a specific rate-degrading impairment of the optical propagation that raises in those systems. Discrete optical elements often exhibit a polarization anisotropy that emerges as a gain or loss imbalance between the two polarization tributaries of the polarization-multiplexed optical signals. This non-unitary effect called Polarization Dependent Loss (PDL) impairs the quality of transmission in current and next-generation optical systems.

In the context of polarization-multiplexed signals, we assess the capacity loss induced by PDL. First, channel models are carefully studied and two approaches are described: a channel with distributed PDL elements and distributed noise or a simplified single-element equivalent channel. Making use of these models, we then analyze their fundamental limits of communications. We show that the PDL channel capacity depends in practice on the state-of-polarization orientation of the incident signal. We then review the state-of-the art of PDL-mitigating modulation schemes and propose two new multi-dimensional signaling schemes that enhance worst-case and average performance. These two modulations are unitary transforms of M-QAM symbols and do not make use of additional degrees of freedom apart from the four already-used dimensions per wavelength (in-phase and quadrature channels of two polarization states). We briefly extend these first results to space-division-multiplexed optical communications impaired by mode dependent loss (MDL) that present a similar gain imbalance. Beyond signal shaping at the transmitter side, we study the performance loss of a conventional, sequential signal processing chain at the receiver side in presence of PDL, in comparison with a joint equalization-decoding scheme. The additional capacity loss due to the mismatch sequential processing is evaluated for several modulation formats and at different operating points. Finally, we report an experimental validation of the two proposed signaling schemes, both on a single PDL element and on a distributed PDL channel, demonstrating the predicted enhanced robustness to PDL.

Institut Polytechnique de Paris

91120 Palaiseau, France