ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY

TELECOM ParisTech

EDITE - ED 130

Doctorat ParisTech

T H È S E

pour obtenir le grade de docteur délivré par

Télécom ParisTech

Spécialité : « Électronique et Communications »

présentée et soutenue publiquement par

Pierre DELESQUES

Soutenance prévue en décembre 2012

Analyses of transmission and switching capacities in optical networks

Directeur de thèse : **Philippe CIBLAT**
Co-encadrement de la thèse : **Gwillerm FROC**
Co-encadrement de la thèse : **Cédric WARE**

T H È S E

Télécom ParisTech
école de l'Institut Télécom – membre de ParisTech

Contents

Acronyms

ASE	Amplified Spontaneous Emission
AWGN	Additive White Gaussian Noise
BPSK	Binary Phase-Shift Keying
BER	Bit-Error Rate
CD	Chromatic Dispersion
CDF	Cumulated Density Function
DeMux	Demultiplexer
DGD	Differential Group Delay
DPSK	Differential Phase Shift Keying
DSL	Digital Subscriber Line
DSP	Digital Signal Processing
EDFA	Erbium-Doped Fiber Amplifier
ERT	Equivalent Random Theory
FDL	Fiber Delay Line
FEC	Forward Error Code
FWM	Four Wave Mixing
FTTB	Fiber To The Business
FTTH	Fiber To The Home
GMPLS	Generalised Multi-Protocol Label Switching (MPLS)
GVD	Group Velocity Dispersion
IP	Internet Protocol
LDPC	Low-Density Parity-Check
LLR	Log-Likelihood Ratio
MIMO	Multiple Input Multiple Output
ML	Maximum-Likelihood
MPLS	Multi-Protocol Label Switching
Mux	Multiplexer
NSE	Nonlinear Schrödinger Equation
NRZ-OOK	Non Return to Zero - On/Off Keying
OADM	Optical Add/Drop Multiplexer
OBS	Optical Burst Switching

OPS	Optical Packet Switching
OWS	Optical Wavelength Switching
OFDM	Orthogonal Frequency Division Multiplexing
OSNR	Optical Signal to Noise Ratio
PMD	Polarization Mode Dispersion
PDL	Polarization Dependent Loss
PolMux	Polarization Multiplexed
PAM	Pulse-Amplitude Modulation
PDF	Probability Density Function
PDM	Polarization Division Multiplexed
PSD	Power Spectral Density
PT	Polarization-Time
QAM	Quadrature-Amplitude Modulation
QoS	Quality of Service
QPSK	Quadrature Phase-Shift Keying
RAM	Random Access Memory
ROADM	Reconfigurable Optical Add/Drop Multiplexer
SBS	Stimulated Brillouin Scattering
SDH	Synchronous Data Hierarchy
SMF	Single-Mode Fiber
SPM	Self-Phase Modulation
SNR	Signal to Noise Ratio
SRS	Stimulated Raman Scattering
WDM	Wavelength Division Multiplexing
XPM	Cross-Phase Modulation

General Introduction

The work presented in this PhD thesis has been produced thanks to the collaboration of the department "Communications et Électronique" (COMELEC) of Télécom ParisTech and MITSUBISHI ELECTRIC R&D Centre Europe (MERCE) in the framework of a "Convention Industrielle de Formation par la REcherche" (CIFRE), and has started in December 2009.

Problem statement

In the last two decades, optical communication has known an amazing expansion. It was first designed to replace the wired interconnections in the core network, and became day after day the main interconnecting solution between the network nodes. Moreover, in order to enhance the user experiment, future access networks such as Fiber To The Home (FTTH), is also based on optical communication. The success of this technique comes from its excellent trade-off between cost and performance, which cannot be achieved by any other wired or wireless system. Unfortunately, due to new Internet applications (such as video streaming), and the now huge number of users, even the optical fiber based networks become saturated, both in terms of inter-node transmission and node switching capacity.

To alleviate the transmission saturation, two solutions can be advocated: i) the first one is to improve the transmission quality of the fibers, but changing the installed base of fibers is very expensive, and would only have a marginal utility with current fiber technology. ii) the second one is to improve the spectral efficiency of existing fibers by using modern digital signal processing and digital communications tools. The second solution is a very hot topic nowadays and this thesis deals with this framework. In order to increase the spectral efficiency, multi-level modulation schemes (such as Quadrature-Amplitude Modulation (QAM)) have to be used in association with coherent detection. A lot of works have been already devoted to design new digital signal processing techniques and new coding techniques for improving significantly the performance. Nevertheless, only a few works have been focused on the fundamental limits of the fiber-based transmission using coherent detection. Therefore, the main purpose of this thesis is to **analyze the ultimate performance of a fiber based transmission using coherent detection from an**

information-theoretic point-of-view in order to obtain insights for its design.

Obviously, the ultimate performance will depend on the considered transmission impairments induced by the fiber and the devices. These impairments can be divided in two categories: linear and non-linear. The latter are mainly generated by the Kerr effect, while the former stem from fiber loss, Chromatic Dispersion (CD), Polarization Mode Dispersion (PMD), and Polarization Dependent Loss (PDL). Since the fiber loss can be easily compensated for by in-line amplification and as CD and PMD can also be perfectly compensated for assuming coherent Orthogonal Frequency Division Multiplexing (OFDM) transmission, PDL remains the unique linear impairment leading to performance loss. As a consequence, we will focus in this thesis on **the influence of PDL and non-linear impairments on the ultimate performance**.

But what do we mean by *ultimate performance* or equivalently *information-theoretic point-of-view*? Actually, we would like to find the maximum data rate for the smallest error probability. When the channel is assumed to be static, the most relevant tool is the so-called Shannon capacity (for which the error probability vanishes). When the channel is assumed to be time-varying and so random, the most relevant tool is the so-called outage probability. Therefore, our objective is twofold: **expressing in closed-form the Shannon capacity in presence of non-linear effect and the outage probability in presence of PDL**.

In addition, we also aim at the switching capacity problem. Current network nodes perform switching almost exclusively in the electronic domain. This requires optical-to-electronic conversion of the entire traffic at each node, incurring considerable expense, as high-data-rate transponders are costly and energy-consuming. Therefore, carriers would like to have transparent (but nevertheless flexible) networks. *Transparent* means that the source and the destination are optically connected, *i.e.*, without electronic regeneration at the nodes. *Flexible* means that the interconnection in Optical Add/Drop Multiplexer (OADM) can be remotely configured with respect to the network constraints (load, failures, etc). This flexibility cannot yet be achieved in the optical domain; it is currently ensured through packet switching of the data, which has to be performed in electronics. Proposed all-optical packet switches perform poorly at high network loads, due to a lack of buffering technology to temporarily store simultaneously-arriving packets competing for the same output channel.

To this end, a hybrid (electronic/optic) node architecture is envisioned, that performs switching in the optical domain and buffering using electronics. Such an architecture has been already proposed and built in the literature. Nevertheless, the performance about its overflow ability has only been analyzed roughly. Therefore, we propose to **analyze the performance-wise influence of parameters** such as the number of access ports to the buffer, the number of channels per output directions and the collision between ingress packets and re-emitted packets on the cross-layer equipment.

Outline and contributions

In this section, we give the thesis outline and we mention the most important results.

In Chapter 1, our main contribution is to derive in closed-form the Shannon capacity in presence of a guard band between transmission channels when non-linear effect occurs. The chapter is divided in three parts: firstly, we remind some definitions and we introduce the linear optical channel model. Secondly, we introduce a comparative state-of-the-art in order to choose the best framework adapted to the presence of the guard band. Thirdly, we provide new closed-form expressions for the non-linear capacity taking the guard band into account. In addition, numerous numerical evaluations are done and we notice that the guard band only has a slight influence on the Shannon capacity.

In Chapter 2, our main contribution is to derive in closed-form the outage probability due to PDL and to compare this probability to practical coding schemes. The chapter is split into three parts: firstly, the PDL phenomenon is described and a rigorous mathematical model is developed. Secondly, we provide the first (and so new) closed-form expressions for the outage probability according to different PDL statistical models. Thirdly, since the outage probability is a reachable lower-bound for Bit-Error Rate (BER) of any coding scheme, we compare it with some practical transmission schemes involving Polarization-Time (PT) codes and Forward Error Code (FEC). This comparison shows that PT codes are of interest and also the concatenation of PT codes and FEC can offer BER quite close to the outage probability.

In Chapter 3, our main contribution is to analyze the empirical performance of an optical burst switch extended by an electronic overflow mechanism. We show that, even with a small overflow mechanism, the gain is significant compared to the all-optical case and so this hybrid approach is validated. We also remark that the main drawback of this overflow mechanism is the collision between ingress packets and re-emitted packets, and that a well designed anti-collision method can significantly improve the overall performance.

Publications

The work presented in this manuscript has led to the following three publications:

- P. Delesques, E. Awwad, S. Mumtaz, G. Froc, P. Ciblat, Y. Jaouën, G. Rekaya, C. Ware, "Mitigation of PDL in Coherent Optical Communications: How Close to the Fundamental Limit?," *European Conference on Optical Communications (ECOC)*, Amsterdam (The Netherlands), September 2012.

- P. Delesques, P. Ciblat, G. Froc, Y. Jaouën, C. Ware, "Outage Probability derivations for PDL-disturbed Coherent Optical Communication," *OSA Signal Processing in Photonic Communications (SPPCom)*, Colorado Springs (USA), June 2012.
- P. Delesques, P. Ciblat, G. Froc, Y. Jaouën, C. Ware, "Influence of guard band on channel capacity for optical transmission systems," *IEEE Photonics Conference (IPC)*, Arlington (USA), October 2011.

Finally another paper has been recently submitted:

- P. Delesques, T. Bonald, G. Froc, P. Ciblat, C. Ware, "Performance analysis of an optical burst switch with shared electronic buffers," submitted for publication to *Optical Fiber Communications (OFC)*, Anaheim (USA), March 2013.

Chapter 1

Shannon Capacity in non-linear regime: influence of the guard band

1.1 Introduction

The two basic ways for increasing the data rate in a transmission system are

- increasing the spectral bandwidth, and
- increasing the spectral efficiency.

The latter can be achieved through ever more spectral efficient transmission schemes, such as Polarization Division Multiplexed (PDM) systems with coherent modulation schemes. In order to use these modulation schemes, a higher input power is required. Unfortunately, in optical transmission systems, this power growth forces system designers to take non-linearities into account.

Even if the total spectral bandwidth is limited in practice by that of Erbium-Doped Fiber Amplifiers (EDFAs), the effective used bandwidth can be increased by reducing the guard band between Wavelength Division Multiplexing (WDM) channels. Unfortunately, once again, this reduction gives rise to non-linear channel interactions.

Therefore, it is of great interest to analyze the influence of the non-linear impairments on the performance of the next system generation.

In order to be as general as possible (*i.e.* not considering a given transmission scheme), we focus on an information theoretical point-of-view. Actually, we propose to analyze the Shannon capacity which provides the maximum reachable data rate in absence of transmission error. In the literature, Shannon capacity has already been derived for some optical channels by taking into account some non-linear effects. Our main contribution is to inspect the influence of the guard band on the capacity.

This chapter is organized as follows: information-theoretic tools, especially the Shannon capacity, are introduced in Section 1.2. In Section 1.3, the optical-fiber based channel and its associated linear and non-linear impairments are given. Then, in Section 1.4, a

state-of-the-art dealing with the theoretical derivations for the Shannon capacity for the optical channel is proposed. The main contribution of this chapter lies in Section 1.5 since we derive in closed-form the Shannon capacity by taken into account the non-linearity and the guard band. Finally, in Section 1.6, numerical illustrations are done for checking our derivations and for analyzing the influence of the guard band. A conclusion is drawn in Section 1.7

1.2 Information-theoretic tools

Shannon, in his seminal work [1], has created the Information Theory for studying the fundamental limits (in terms of data rate and error probability) of any transmission system.

He has considered that any transmission system can be decomposed as on Fig. 1.1. The source is associated with one encoder which sends digital data to a destination through a propagation channel. Then the destination is associated with one decoder in order to retrieve the transmitted data.

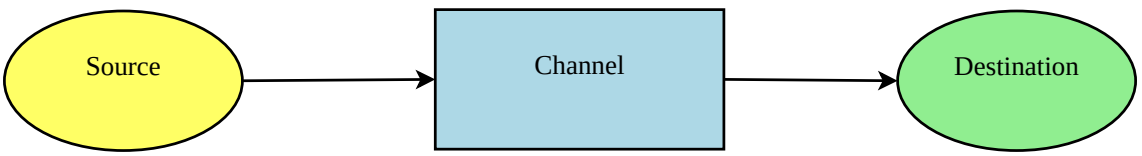

Figure 1.1: Shannon paradigm for any transmission scheme

In order to theoretically analyze such a system, he has defined three main metrics:

- the (source) entropy,
- the mutual information, and
- the (channel) capacity.

1.2.1 Definition of the entropy

Before the Shannon's work, the notion of entropy has been already introduced in physics, especially in thermodynamic and statistical physics. The entropy always represents the level of disorder in the considered system. In Thermodynamics, the system is related to the electrons. In communication, the system is associated with the message to be sent. Therefore, the entropy is a measure of the disorder in the sequence to be sent, and so a measure of the quantity of information in a message [2]. Indeed, a message without disorder (which means perfectly predictable) does not bring information. In contrast, a message without order, can not be predicted and so brings the maximum information.

In order to measure this disorder, the message/sequence will be modeled by a random variable. Let X be a discrete random variable of alphabet $\mathcal{X}$. We define the probability

mass function (for discrete alphabet) by $p_X(x) = \text{Prob}\{X = x\}$ for any $x \in \mathcal{X}$. The entropy associated with this random variable, is denoted by $H(X)$, and is defined as follows

$$H(X) = -\sum_{x \in \mathcal{X}} p_X(x) \log_2(p_X(x)). \tag{1.1}$$

The expression for continuous random variable can be found in [2] and needs the use of the so-called differential entropy.

In the following, it will be useful to measure the entropy of a random variable, let's say Y, given an other one X. Therefore, we introduce the notion of the conditional entropy (*i.e.* the entropy of a random variable given an other one) which is given as follows

$$\begin{aligned} H(Y|X) &= \sum_{x \in \mathcal{X}} p_X(x) H(Y|X = x) \\ &= -\sum_{x \in \mathcal{X}, y \in \mathcal{Y}} p_X(x) p_{Y|X}(y|x) \log_2(p_{X|Y}(y|x)). \end{aligned}$$

1.2.2 Definition of the mutual information

The mutual information provides the common information shared by two random variables. Let us consider two random variables, X and Y and their mutual information $I(X;Y)$. We have

$$\begin{aligned} I(X;Y) &= H(Y) - H(Y|X) \\ &= H(X) - H(X|Y). \end{aligned}$$

We remark that if $X = Y$, the mutual information corresponds to the maximum information available in the random variable. In contrast, if X and Y are independent, the mutual information vanishes and so there is no shared information between both variables.

1.2.3 Definition of the Shannon capacity

The channel capacity is related to the maximum data rate (useful bits transmitted per channel use or per second according to the considered definitions) that a channel can achieve without error.

Let X be the random variable representing the channel input (*i.e.*, the source), and Y be the random variable representing the channel output (*i.e.*, the destination), according to Fig. 1.1. Let p_X be the Probability Density Function (PDF) of X.

The channel Shannon capacity (per channel use), denoted by c, is the maximum of the mutual information between X and Y for any possible PDF of X, and so writes as follows

$$c = \max_{p_X} I(X;Y). \tag{1.2}$$

Notice that the optimum PDF depends on the channel.

One of the most standard channel for the communication system is the Gaussian channel, also called, Additive White Gaussian Noise (AWGN) channel. In that case, we have

$$Y_n = X_n + N_n,$$

where n is the time index and where $N = \{N_n\}_n$ is a Gaussian white random process. For this channel, the closed-form expression of the capacity has been provided in [1]. When complex-valued random variable is assumed, it has been proven that the optimum PDF is the zero-mean circularly symmetric Gaussian distribution, with a variance $\mathbb{E}[|X_n|^2] = 2P_x$ where P_x is actually the power constraint of the user at the carrier. Then we have

$$c = \log_2\left(1 + \frac{P_x}{\sigma_N^2}\right) \quad \text{bits per channel use (pcu)} \tag{1.3}$$

with $\sigma_N^2 = \mathbb{E}[|N_n|^2]/2$ the variance of the AWGN per real dimension.

If we assume that the information sequence at baud rate B is conveyed by a continuous-time transmitted signal of bandwidth B, the capacity, expressed in bits per second and denoted by C, is given by

$$\mathrm{C} = B\log_2\left(1 + \frac{P_x}{N_0 B}\right) \quad \text{bits per second.} \tag{1.4}$$

where B is the considered bandwidth and where $N_0/2$ is the noise level for the double-side noise Power Spectral Density (PSD) [1].

1.3 Optical fiber based channel model

Before going further, we need to model the optical fiber based channel since the capacity strongly depends on it. Although standard optical transmission systems are built using optical fibers, lasers, modulators, Multiplexer (Mux), Demultiplexer (DeMux), filters, amplifiers, detectors, we hereafter only consider the impairments due to the optical fiber, for the sake of simplicity.

In a optical fiber [3], the transmitted signal has to satisfy the so-called following "wave equation" strongly connect to the Maxwell's equation, and thus

$$\nabla \times \nabla \times \vec{X}(t) = -\frac{1}{c^2}\frac{\partial^2 \vec{X}(t)}{\partial t^2} - \frac{1}{\varepsilon_0 c^2}\frac{\partial^2 \vec{P}}{\partial t^2} \tag{1.5}$$

where $\vec{X}(t)$ is the electric field in time domain [2], ∇ is the (spatial) differential operator, $\times$ is the vector product, c is the light celerity, ε_0 is the vacuum permittivity, and $\vec{P}(t)$ is the polarization vector in the time domain.

[1] In this thesis, we consider that the complex envelope $x_e(t) = x_p(t) + ix_q(t)$ of the transmitted signal leads to the following transmitted signal $x(t) = x_p(t)\cos(2\pi f_0 t) - x_q(t)\sin(2\pi f_0 t)$ around the central frequency f_0. As a consequence, the power/energy at the baseband is twice that at the carrier. Unless otherwise stated, the power/energy is calculated at the carrier.

[2] Notice that the dependence of $X(t)$ with respect to the location is omitted even if this dependence exists.

It is well-known that the polarization vector in the frequency domain, denoted $\vec{\tilde{P}}$, can be decomposed as follows

$$\vec{\tilde{P}} = \varepsilon_0 \left(\chi^{(1)} \cdot \vec{\tilde{X}} + \chi^{(2)} \cdot \vec{\tilde{X}}\vec{\tilde{X}} + \chi^{(3)} \cdot \vec{\tilde{X}}\vec{\tilde{X}}\vec{\tilde{X}} + \dots \right) \tag{1.6}$$

with $\vec{\tilde{X}}$ is the Fourier transform of $\vec{X}$, and $\chi^{(j)}$ a $(j+1)$-order tensor representing the jth-order susceptibility. $\chi^{(1)}$ is called the linear susceptibility and represents generally the dominant contribution. However, when the energy increases, the other terms become significant as well. Since the silicium of the optical fiber is a symmetric material, $\chi^{(2)}$ is close to zero and can be neglected. Therefore, $\chi^{(3)}$ gives rise to the main non-linear impairments.

Let us focus on the linear regime, *i.e.*, only the term $\chi^{(1)}$ in Eq. (1.6) is considered. The wave equation can then be simplified as follows in the frequency domain

$$\nabla \times \nabla \times \vec{\tilde{X}} - \frac{(2\pi f)^2}{c^2} \left(1 + \chi^{(1)}\right) \vec{\tilde{X}} = 0 \tag{1.7}$$

where f is the frequency. Through the coefficient $\chi^{(1)}$, one can define two important parameters (see [3] for more details) :

- Energy dispersion (given by the refractive index which may depend on the frequency):

$$n = 1 + \frac{1}{2} \Re[\tilde{\chi}^{(1)}] \tag{1.8}$$

- Energy attenuation (given by the loss coefficient which depends on the frequency):

$$\alpha = \frac{2\pi f}{c} \Im[\tilde{\chi}^{(1)}] \tag{1.9}$$

Notice that, in this chapter, we neglect the birefringence aspect of the fiber, *i.e.*, we omit PMD and PDL. Consequently, only fiber attenuation and chromatic dispersion are taken into account.

1.3.1 Fiber attenuation

During the propagation the signal power is attenuated by the fiber material. This attenuation depends on the wavelength (cf. Eq. (1.9)). For standard Single-Mode Fiber (SMF) and standard used wavelengths (around 1550 nm), the attenuation is minimum. Generally, the wavelength dependency is neglected, and the relationship between the input signal power P_x and the output signal power P_y is

$$P_y = e^{-\alpha L} P_x \tag{1.10}$$

where L represents the fiber length in km.

The fiber loss/attenuation (defined as the ratio P_x/P_y) is usually expressed in dB/km and so satisfies the following relationship: $\alpha_{\text{dB}} = \frac{10}{\log(10)}\alpha$. Its standard value is around 0.2 dB/km. In long-haul optical transmission systems, the attenuation is mitigated by in-line amplifiers. Unfortunately, these in-line amplifiers lead to a noise enhancement, and so to a Signal to Noise Ratio (SNR) degradation.

1.3.2 Chromatic dispersion

The CD is due to the frequency dependency of the refractive index of the fiber $n(f)$ (cf. Eq. (1.8)). Indeed spectral components of the signal do not propagate at the same velocities. As in [3], fiber dispersion effects can be taken into account by expanding the mode propagation constant β in Taylor series about the central spectral frequency f_0:

$$\beta(f) = n(f)\frac{2\pi f}{c} = \beta_0 + 2\pi\beta_1(f - f_0) + \frac{(2\pi)^2}{2}\beta_2(f - f_0)^2 + \cdots . \tag{1.11}$$

where β_1 is the inverse of the group velocity and β_2 represents the Group Velocity Dispersion (GVD). Usually, the dispersion is measured with respect to the term $\mathrm{D}(\lambda)$ which is defined as follows

$$\mathrm{D}_\lambda = -\frac{2\pi c}{\lambda^2}\beta_2 \quad \text{ps/nm/km}. \tag{1.12}$$

If the transmitted signal is narrow band enough, high-order coefficients can be neglected in Eq. (1.11).

Notice that by doing a appropriate variable changing, the term β_1 can be removed in the wave equation.

Moreover the transfer function of CD in the frequency domain at position z (by assuming planewave) has been obtained in [4]:

$$H_{CD}(f) = e^{-i\pi\frac{D_\lambda\lambda^2 z}{c}f^2} = e^{2i\pi^2\beta_2 z f^2}. \tag{1.13}$$

Since this transfer function is unitary ($|H_{CD}(f)|^2 = 1$), the channel capacity will not be modified.

1.3.3 Nonlinear effects

We will now focus on the non-linear effects which are generated by the term $\chi^{(3)}$ in Eq.(1.6).

Actually, the first non-linear effect (which has an impact on the value of $\chi^{(3)}$), is due to the local random variation of the refractive index in fiber with respect to the signal instantaneous power. This effect is called Kerr effect [3]. As $\chi^{(3)}$ and so the Kerr effect deals with the product of three electric fields (cf. Eq. (1.6)), there is three main degradations:

- The Self-Phase Modulation (SPM) is the effect induced by its own pulse. It generates a modification of the instantaneous phase shift of the pulse.

- The Cross-Phase Modulation (XPM) occurs when several channels propagate simultaneously in the fiber as in WDM systems. Then, an other instantaneous phase shift induced by the combination with an other pulse occurs.

- The Four Wave Mixing (FWM) generates new pulses at specific frequencies and may lead to interference.

Assuming only one polarization mode and only Kerr effect for the non-linear impairments, the wave equation (Eq.(1.5)) leads to the famous Nonlinear Schrödinger Equation (NSE) [3]:

$$i\frac{\partial X}{\partial z} + i\frac{\alpha}{2}X - \frac{\beta_2}{2}\frac{\partial^2 X}{\partial t^2} + \gamma_{NL}|X|^2 X = 0 \tag{1.14}$$

where γ_{NL} is the non-linear coefficient due to Kerr Effect. This equation is a spatio-temporal differential equation which does not exhibit closed-form expressions for its solution, except in very particular cases such as, CD-free case or linear case. Notice that the absence of closed-form expressions for the solution to the NSE in the general case will be one of the most important issue for deriving the non-linear channel capacity in the remainder of this chapter.

Before going further, we have to introduce the so-called effective length defined as follows

$$L_{\text{eff}} = \frac{1 - e^{-\alpha L}}{\alpha}. \tag{1.15}$$

It represents the length of the fiber beyond which non-linear effects take place.

Notice that, other non-linearities such as Stimulated Brillouin Scattering (SBS) or Stimulated Raman Scattering (SRS) exist, but can be neglected compared to the Kerr Effect. Moreover we do not take into account the fiber polarization. In this case, the electric filed is driven by the Manakov-PMD equation [5] which proves the existence of the so-called XPolM corresponding to an extension of the the XPM into the cross-polarization. Manakov equation is strongly more complicated than the Schrödinger one and so we will not focus on it.

1.4 State of the art on the non-linear capacity derivations

The problem of the derivations of the non-linear capacity in closed-form can be split into three steps:

1. exhibiting in closed-form the input-output relationship of the non-linear optical channel,

2. finding out the PDF maximizing the mutual information of this channel,

3. deriving the mutual information.

These three steps are quite complicate for the non-linear optical channel since the NSE does not admit often a closed-form expression. Moreover, even if the solution of the NSE is known in specific cases, finding the optimal PDF and calculating the mutual information can be a very hard task.

All existing works (except [6]) consider a Gaussian PDF for the input which will lead to results similar to Eq.(1.4). The difference between these works lies in the way they take into account the non-linearities, and so they incorporate the input-output relationship. Notice that the Gaussian assumption is reasonable since high-order QAM is quite well modeled by this PDF, and it enables us to derive some expressions in closed-form.

In Section 1.4.1, the non-linear effect is viewed as an additional noise and the capacity is modified accordingly. In Section 1.4.2, the correlation between the input and output signals (since the noise is now signal-dependent) are calculated when non-linear effect occurs and the new expression for the capacity is found. In Section 1.4.3, the non-linear effect is viewed as a perturbation of the linear case and the capacity is then re-written. Finally, in Section 1.4.5, empirical evaluations of the capacity are done which enable us to discuss the validity of the different above works.

Hereafter, we consider a WDM system with N_c channels. Each channel has a bandwith B and an input power P_x. $P_N = N_0 B$ represents the noise power.

1.4.1 Derivations based on additional noise

We recall that in this subsection, the capacity will be calculated in closed-form by modifying the SNR expression in Eq. (1.4). Obviously, all the below-written expressions only correspond to capacity approximation.

1.4.1.1 When the additional noise is multiplicative [7]

In [7], the authors consider the following system

- a WDM system for which the fiber attenuation is perfectly compensated for.
- The SPM is neglected as it depends on its own channel and thus could be pre-compensated.
- The FWM is assumed to increase only the additive noise variance P_N. Nevertheless, the increase is not evaluated by the authors.
- Only the XPM is considered and thus the authors asserted that it corresponds to the main non-linear impairment.

As a consequence ($\alpha = 0$, no SPM), the input X_k of the k-th WDM channel writes as follows (cf. Eq (1.14))

$$i\frac{\partial X_k}{\partial z} - \frac{\beta_2}{2}\frac{\partial^2 X_k}{\partial t^2} + V_k X_k = 0 \tag{1.16}$$

with $V_k = 2\gamma_{NL}\sum_{\ell\neq k}|E_k|^2$. In the term V_k, the contribution of all other WDM channels interference is taken into account.

The term V_k looks like a random variable and the authors assumed (under the assumption of substantial dispersion) it is Gaussian distributed. After tedious and not straightforward derivations, they eventually obtain the following input-output relationship for the channel

$$Y(t) = \left(\int u(t,t')dt'\right)X(t) + N(t) \tag{1.17}$$

where $u(t,t')$ corresponds to the additional multiplicative noise depending on V_k.

Assuming $X(t)$ and $N(t)$ are Gaussian distributed, the authors just need to calculate all the correlations between $Y(t)$ and $X(t)$ (see more details in Section 1.4.2) also averaged on V_k. They finally obtain the following closed-form expression for the capacity:

$$\mathrm{C} = B\log_2\left(1 + \frac{e^{-\left(\frac{P_x}{P_{NL}}\right)^2}P_x}{P_N + \left(1 - e^{-\left(\frac{P_x}{P_{NL}}\right)^2}\right)P_x}\right), \tag{1.18}$$

with

$$P_{NL} = \sqrt{\frac{D_\lambda c}{2\gamma_{NL}^2\log\left(\frac{N_c}{2}\right)L_{\text{eff}}}}. \tag{1.19}$$

The maximum capacity can be roughly approximated by

$$\mathrm{C}_{\max} \simeq \frac{2B}{3}\log_2\left(\frac{2P_{NL}}{P_N}\right)$$

and is reached when $P_x = (P_{NL}P_N/2)^{1/3}$.

In [8], the authors consider the same model as in [7] (WDM, no attenuation, no SPM), but they asserted that the FWM dominates the XPM when constant-intensity modulations are used (such as, Binary Phase-Shift Keying (BPSK), Differential Phase Shift Keying (DPSK), Quadrature Phase-Shift Keying (QPSK)). Using the same framework as above, they argue that only the term P_{NL} has to be modified as follows

$$\frac{1}{P_{NL}} = N_s\gamma_{NL}^2 \sum_{\substack{-(N_c-1)/2\leq p\leq(N_c-1)/2, p\neq 0\\ -(N_c-1)/2\leq q\leq(N_c-1)/2, q\neq 0\\ |p+q|\leq(N_c-1)/2}} \frac{(2-\delta_{p,q})^2}{\alpha^2 + k_{p,q}^2} \text{ with } k_{p,q} = 2\pi\lambda^2 D_\lambda B^2 qp/c \tag{1.20}$$

where N_s is the number of spans, and $\delta_{.,.}$ stands for the Kronecker index. This expression was derived for the central channel which is the most disturbed by the FWM.

1.4.1.2 When the additional noise is additive [9, 10]

In [9, 10], the authors consider the following model

- a CD-uncompensated fiber
- potentially a PDM transmission scheme

The authors have developed a model for which the non-linearities just generate an additional additive white Gaussian noise independent of the amplification noise. This assumption is well satisfied when the channel is highly dispersive (cf. β_2 large enough). As a consequence, the capacity is obtained as follows

$$C = B\log_2\left(1 + \frac{P_x}{P_N + N_s\left(\frac{2}{3}\right)^3 \gamma_{NL}^2 L_{\text{eff}_{2D}} \frac{\log\left[\pi^2|\beta_2|L_{\text{eff}_{2D}}(N_c B)^2/2\right]}{\pi|\beta_2|}(P_x)^3}\right) \tag{1.21}$$

where $L_{\text{eff}_{2D}} = (1-\exp(-2\alpha L))/(2\alpha)$ is the non-linear effective length for the PDM case.

Actually, the closed-form expression depends on the nature of the amplification. If the amplifier is EDFAs, Eq. (1.21) holds. If the amplifier is based on Raman principle, the correct expression is provided in [10].

1.4.2 Derivations based on input-noise correlation

In [11, 12, 13, 14], all the closed-form expressions for the nonlinear optical channel are calculated thanks to the use of the so-called Pinsker's formula [15] which enables them to take into account the correlation between the input signal and the noise. We have

$$C = \int_{-B/2}^{B/2} \log_2\left(\frac{P(f)}{Q(f)}\right) df \tag{1.22}$$

where $P(f)$ and $Q(f)$ are the following terms

$$\begin{aligned} P(f) &= \det\left(\begin{bmatrix} FT(\mathbb{E}[X(t+\tau)X^*(t)]) & 0 \\ 0 & FT(\mathbb{E}[Y(t+\tau)Y^*(t)]) \end{bmatrix}\right) \\ Q(f) &= \det\left(\begin{bmatrix} FT(\mathbb{E}[X(t+\tau)X^*(t)]) & FT(\mathbb{E}[X(t+\tau)Y^*(t)]) \\ FT(\mathbb{E}[Y(t+\tau)X^*(t)]) & FT(\mathbb{E}[Y(t+\tau)Y^*(t)]) \end{bmatrix}\right) \end{aligned} \tag{1.23}$$

where $X(t)$ and $Y(t)$ are the input and the output of the fiber respectively. The operator $FT(\cdot)$ stands for the Fourier transform and the superscript $(\cdot)^*$ stands for complex conjugate.

In [11, 12, 13, 14], the derivations of the capacity are achieved only for the central channel of the WDM system. Moreover, all the WDM channels are assumed to be flat fading.

The authors consider two cases for the input-output relationship:

- the non-dispersive case ($\beta_2 = 0$) for which the NSE admits an exact closed-form solution,
- the dispersive case ($\beta_2 \neq 0$) for which only an approximate closed-form solution of the NSE can be used.

1.4.2.1 Non-dispersive case

For the non-dispersive case, the NSE admits the following closed-form solution (under a single span assumption)

$$Y(t) = e^{\gamma_{NL} L_{\text{eff}} |X(t)|^2} X(t) + N(t). \tag{1.24}$$

After tedious derivations, the author in [11] obtains that

$$C = B \log_2 \left(1 + \frac{\frac{P_x}{1+(\gamma_{NL} L_{\text{eff}} N_c P_x)^2}}{P_N + P_{\text{add.}}} \right) \tag{1.25}$$

where

$$P_{\text{add.}} = 2P_x \sum_{n=1}^{\infty} \frac{(n+1) g_{2n+1} (\gamma_{NL} L_{\text{eff}} N_c P_x)^{2n}}{[1 + (\gamma_{NL} L_{\text{eff}} N_c P_x)^2]^{n+2}}$$

with

$$g_{2n+1} = \frac{(2n+1)}{2^{2n+1}} \sum_{k=0}^{\lfloor n+1/2 \rfloor} \frac{(-1)^k (2n+1-2k)^{2n}}{k!(2n+1-k)!}.$$

As the maximum capacity is obtained when the power is small enough, the interval of interest for the power leads to small enough values and thus enables us to keep only the first term of the sum in $P_{\text{add.}}$. Then, we have

$$C = B \log_2 \left(1 + \frac{\frac{P_x}{1+(\gamma_{NL} L_{\text{eff}} N_c P_x)^2}}{P_N + \frac{3}{2} P_x \frac{(\gamma_{NL} L_{\text{eff}} N_c P_x)^2}{[1+(\gamma_{NL} L_{\text{eff}} N_c P_x)^2]^3}} \right). \tag{1.26}$$

Thanks to this equation, one can prove that the maximum capacity is

$$C_{\max} = \log_2 \left(1 + \frac{2/9^{2/3}}{(\gamma_{NL} L_{\text{eff}} N_c P_N)^{2/3}} \right)$$

and is reached when $P_x = (P_N/3)^{1/3} / (\gamma_{NL} L_{\text{eff}} N_c)^{2/3}$.

In [12], the multi-span case is studied by induction.

1.4.2.2 Dispersive case

In [13], the dispersive case is treated. The main issue lies in the absence of closed-form relationship between the input and the output. Consequently, the author proposed to overcome the problem by using an approximate relationship between the input and the output based on the Volterra series expansion [16]. Once again , for the sake of simplicity, let us consider only the single-span case.

If the non-linearities is small enough, the relationship between the input and the output can be written by considering only a third-order Volterra series expansion (the first order corresponds to the linear part; the second-order vanishes for the sake of symmetric). Then, we have, in the frequency domain,

$$Y(f) = Y_L(f) + Y_{NL}(f) \tag{1.27}$$

with

$$Y_L(f) = e^{2i\pi^2\beta_2 L f^2} X(f) \tag{1.28}$$

$$Y_{NL}(f) = e^{2i\pi^2\beta_2 L f^2}(2i\pi)\gamma_{NL} \iiint \delta(f - f_1 + f_2 - f_3)X(f_1)X^*(f_2)X(f_3) \times \frac{e^{L\left(-\alpha+2i\pi^2\beta_2(-f^2+f_1^2-f_2^2+f_3^2)\right)} - 1}{-\alpha + 2i\pi^2\beta(-f^2 + f_1^2 - f_2^2 + f_3^2)} df_1 df_2 df_3 + N(f) \tag{1.29}$$

where $Y_L(f)$ is the linear part of the received signal (associated with the Chromatic dispersion - cf. Eq. (1.13)) and $Y_{NL}(f)$ is the non-linear part and the noise. The term L stands for the fiber/span length.

Then, by using Pinsker's Formula (Eq. 1.22) and assuming white spectrum, the author asserted that the capacity takes the following form

$$C = B \log_2 \left(1 + \frac{S_X(f_0)}{K(f_0)}\right) \tag{1.30}$$

where f_0 is the carrier frequency, $S_X(f)$ is the power spectral density of the input signal, and

$$K(f) = FT(\mathbb{E}[Y_{NL}(t+\tau)Y_{NL}^*(t)]) - |FT(\mathbb{E}[Y_{NL}(t+\tau)Y_L^*(t)])|^2 / S_X(f).$$

Notice that Eq. (1.30) is actually wrong since it leads to a contradiction for the so-called opaque channel. Let us consider the opaque channel: $Y(t) = N(t)$. According to Eq. (1.30), the capacity is $C = B\log_2(1 + P_x/P_N)$ which is impossible since the output is independent of the input and so leads to a capacity equal to zero. Consequently, in Eq. (1.30), some terms are missing. In Section 1.5, devoted to our own derivations and contributions, the missing terms will be exhibited and so added. Let us go back to [13].

As asserted by the author, after tedious but *straightforward* [3] derivations, he obtained that

$$\begin{aligned} K(f) &= N_c S_N(f) \\ &+ 4\pi\gamma_{NL}^2 \iiint \delta(f - f_1 + f_2 - f_3) S_X(f_1) S_X(f_2) S_X(f_3) \\ &\times \mathcal{U}(f_1, f_2, f_3)\mathcal{V}_{N_c-1}(f_1, f_2, f_3) df_1 df_2 df_3 \\ &+ 4\pi\gamma_{NL}^2 \iiint \delta(f - f_1 + f_2 - f_3) S_X(f_1) S_N(f_2) S_N(f_3) \\ &\times \mathcal{U}(f_1, f_2, f_3) \sum_{n=0}^{N_c-2} \mathcal{V}_n(f_1, f_2, f_3) df_1 df_2 df_3 \end{aligned} \tag{1.31}$$

where

- $S_N(f)$ stands for the PSD of the noise,
- $\mathcal{U}(f_1, f_2, f_3) = \left| \frac{e^{L\left(-\alpha+2i\pi^2\beta_2(-f^2+f_1^2-f_2^2+f_3^2)\right)} - 1}{-\alpha+2i\pi^2\beta(-f^2+f_1^2-f_2^2+f_3^2)} \right|^2$,

[3] We will see in Section 1.5 that these derivations are not so straightforward to our opinion.

- $\mathcal{V}_n(f_1,f_2,f_3)=\left|\sum_{m=0}^{n} e^{mL2i\pi^2\beta(-f^2+f_1^2-f_2^2+f_3^2)}\right|^2$.

Finally, in [14], the same author approximated the above three-dimensional integral in order to obtain the following approximate closed-form expression for the capacity.

$$C=B\log_2\left(1+\frac{\frac{P_x}{1+(\gamma_{NL}L_{\text{eff}}N_cP_x)^2}}{P_N+\frac{3}{2}P_x\frac{(\gamma_{NL}L_{\text{eff}}N_cP_x)^2}{\left[1+(\gamma_{NL}L_{\text{eff}}N_cP_x)^2\right]^3}\xi}\right). \tag{1.32}$$

with

$$\xi=\frac{8\pi^3}{3}\frac{\log\left(\pi^2\beta_2N_c^2B^2/\alpha\right)}{\pi^2\beta_2N_c^2B^2/\alpha}.$$

As soon as the dispersion is large enough, we have $\xi<1$ which shows that the capacity in dispersive environment is larger than in the dispersion-free case.

As a conclusion, even if the author omitted a term in Eq. (1.30), its framework is very useful and will be used later in the thesis.

1.4.3 Derivations based on perturbation theory

In [17], the authors assumed that the Kerr effect (directly related to X^3 in the wave equation) is smaller than the refractive index. Then, they developed a technique for deriving the capacity based on the perturbation theory. Let us consider the single-span case for the sake of clarity. The authors modeled the input-output relationship as follows

$$Y(f)=e^{\left(-i\left(2\pi\beta_1f-2\pi^2\beta_2f^2\right)L\right)}\left(X(f)+\sum_{\ell=1}^{\infty}\gamma_{NL}^{\ell}\mathcal{F}^{(\ell)}(X(f))\right)+N(f) \tag{1.33}$$

where $\mathcal{F}^{(\ell)}(\bullet)$ represents the perturbation due to the non-linear effect. For more details about the derivations of $\mathcal{F}^{(\ell)}(\bullet)$, the reader may refer to [17].

Assuming the input Gaussian distributed, we show that the capacity is well approximate with

$$C=B\log_2\left(1+\frac{P_x}{P_N}\right)-\Delta C \tag{1.34}$$

where

$$\begin{aligned}\Delta C &= B\left(\frac{\gamma_{NL}P_x}{\alpha}\right)^2\int_{-1/2}^{1/2}\int_{-1/2}^{1/2}\int_{-1/2}^{\max[\frac{1}{2};\frac{1}{2}+x_1+x_2]} g(x_1,x_2,x_3)dx_1dx_2dx_3\\ g(z,x_1,x_2,x_3) &= \frac{\left|1-e^{\left(-\alpha L-i\left(\frac{\beta_2^2B^4}{\alpha^2}(x_3-x_1)(x_3-x_2)\right)^2\right)}\right|}{1-\left(\frac{\beta_2^2B^4}{\alpha^2}(x_3-x_1)(x_3-x_2)\right)^2}.\end{aligned}$$

In [18], similar techniques for the capacity derivation have been used to extend this work for the WDM system. Unfortunately, only numerical illustrations are shown in [18].

1.4.4 Derivations of a lower bound

In [6], the authors proposed to derive a very simple lower-bound for the capacity. By considering that the phase of the transmitted signal is independent of the phase of the received signal (e.g., a intensity-modulated input signal), the mutual information for such a input signal leads to the following lower-bound of the capacity (since the capacity corresponds to the highest mutual information with respect to the input distribution.

$$C \geq \frac{B}{2} \log_2 \left(\frac{P_x}{P_N} \right). \tag{1.35}$$

As a consequence, the capacity is clearly unbounded when the SNR increases. Same conclusion has been drawn in [19].

The question now is: why using the Gaussian distribution and so Pinsker's formula for deriving capacity. In practice, the modulation with high efficiency mimics the Gaussian distribution, and the its phase provides information (which is not the case for intensity-modulated signal). In order to obtain high data rate, the use of modulation with high efficiency coupled with a efficient forward error coding is required. Using intensity-modulated modulation with a large number of states is irrelevant in practice since this modulation yields very poor performance that a FEC can not counter-act.

Therefore working with Gaussian distribution for deriving capacity remains of interest to obtain some insights on practical system behavior.

1.4.5 Numerical evaluations

In many papers [20, 21, 22, 18], the analysis of the capacity is only done through its numerical evaluation, which means that the mutual information is obtained through Monte-Carlo trials according to an pre-defined input distribution and an pre-defined input-output relationship (sometimes, the NSE is solved numerically in order to obtain this relationship as in [20, 21])

In this section, we only plot numerical evaluations of the closed-form expressions provided in the state-of-the-art. The simulation setup is the following one: a single-span fiber of length 80 km, the noise $P_N = 10^{-}5$ mW, $\gamma_{NL} = 1.22$ W/km, $\alpha = 0.2$ dB/km, $\beta_2 = -21.6\text{ps}^2/\text{km}$, and $N_c = 81$.

As already mentioned when [11] has been analyzed, the CD has a great influence when non-linear effect occurs. The more dispersive the fiber is, the better the capacity is. When Gaussian distribution is considered, the capacity is bounded and the maximum value is obtained for standard input power per channel ($\simeq$ 1 mW). Moreover, when the power is low enough, the capacity perfectly follows the linear regime. The true capacity (by using the optimal input distribution) is actually unbounded since the bound derived in [6] is monotonic with respect to the input power.

Nevertheless, as already said in Section 1.4.4, the derivations and analysis of the capacity relying on Gaussian input capacity remains of interest for practical coherent

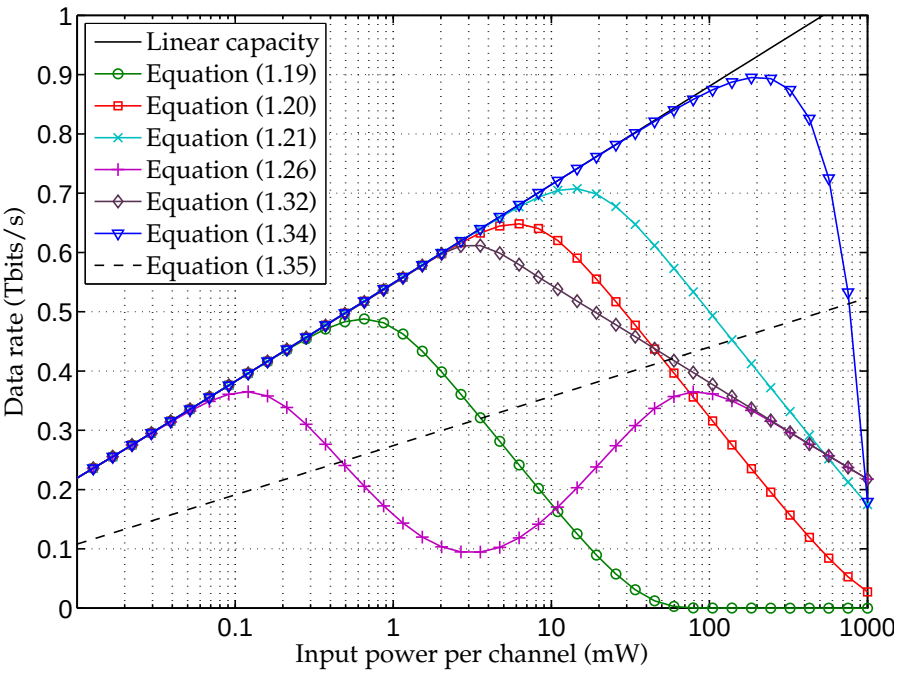

Figure 1.2: Optical nonlinear channel capacity behavior

modulation based systems.

For completing our state-of-the-art, we display in Table 1.1 the pros and cons of each existing framework and in particular which ones will be relevant for fixing our problem, *i.e.*, for analyzing the influence of the guard band on the capacity. We remark that the framework developed under the input-noise correlation assumption [11, 12, 13, 14] enables us to use different input spectrum and so enables us to insert quite easily the guard band.

1.5 New derivations of the capacity in presence of the guard band

As explained in the introduction of this chapter, the influence of the guard band raises when we would like to increase the spectral efficiency per WDM channel since we would like to have closer channels. And, due to the non-linearities, closer channels lead to performance loss. Therefore we would like to exhibit the trade-off between non-linear effect mitigation (large guard band) and high spectral efficiency (small guard band). To do that, we propose to use the information-theoretic tool, and especially, the channel capacity. As alaready explained in previous section, our derivations will use the framework, developed in [11, 12, 13, 14]. This section is organized as follows: in Section 1.5.1, we briefly introduce your system model. Then, our derivations are given and as in [11, 12, 13, 14], the non-dispersive case (see Section 1.5.2) is done separately from the dispersive case (see Section 1.5.3).

	CD	Kerr non-linearities			Multi-Span	Different input spectrum
		SPM	XPM	FWM		
Multiplicative noise (XPM) (Eq. (1.18))	×		×		×	
Multiplicative noise (FWM) (Eq. (1.20))	×			×	×	
Additive noise (Eq. (1.21))	×	×	×	×	×	
Input-noise correlation (Eqs. (1.26)-(1.32))	×	×	×	×	×	×
Perturbation theory (Eq. (1.34))	×	×	×	×	×	

Table 1.1: Comparison of framework for capacity derivation (The symbol × marks the compatibility of the framework with an impairment or an issue)

1.5.1 System model

We consider an optical coherent transmission system which contains $N_c = 2N + 1$ WDM channels. For the sake of simplicity, we assume a mono-span amplified link. Since our aim is to understand the influence on the guard band on non-linearities, each WDM channel is assumed to have a flat spectrum over the bandwidth B, and a guard band B_g between two consecutive WDM channels is introduced. Finally, P_x is the input power of each WDM channel. Such a spectrum is displayed in Fig. 1.3, and has the following mathematical form

$$S_X(f) = \frac{P_x}{B} \sum_{k=-N}^{N} \Pi\left(\frac{f + k(B + B_g)}{B}\right) \tag{1.36}$$

where $x \mapsto \Pi(x)$ is the rectangular function of width 1 and defined as

$$\Pi(f) = \begin{cases} 1 & \text{if } |f| < \frac{1}{2} \\ 0 & \text{otherwise} \end{cases} . \tag{1.37}$$

The autocorrelation of signal $X(t)$ is denoted by

$$R_X(\tau) = \mathbb{E}[X(t + \tau)X^*(t)]$$

and we have $S_X(f) = FT(R_X(\tau))$.

We remind that the capacity is calculated for the central channel which is the most disturbed by non-linearities. Moreover, the input is assumed to be Gaussian distributed, and the amplification noise is considered as AWGN.

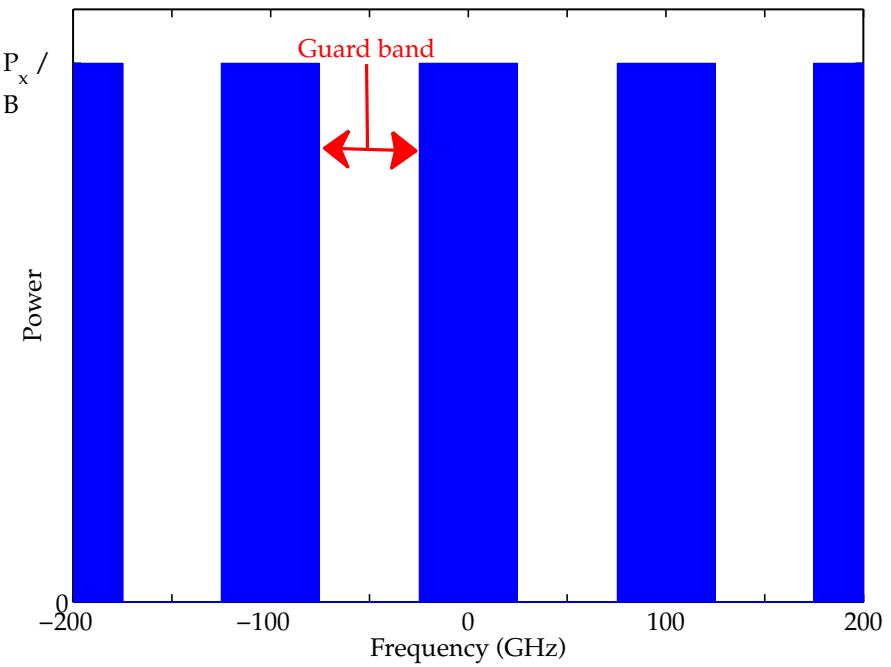

Figure 1.3: Input signal spectrum

1.5.2 Non-dispersive case

Assuming non-dispersive fiber (*i.e.* $\beta_2 = 0$), the NSE admits a closed-form solution and in [11] (after simple algebraic manipulations), the capacity is given by

$$C = \int_{-B/2}^{B/2} \log_2 \left(1 + \frac{\frac{S_X(f)}{1+\gamma_{NL}^2 L_{\text{eff}}^2 R_X^2(0)}}{S_N(f) + \sum_{n=1}^{\infty} \frac{(n+1)\gamma_{NL}^{2n} L_{\text{eff}}^{2n} FT(R_X(\tau)|R_X(\tau)|^{2n})}{(1+\gamma_{NL}^2 L_{\text{eff}}^2 R_X^2(0))^{n+2}}} \right) df. \tag{1.38}$$

Then, considering that the total input power is low (*i.e.* $N_c P_x \leq 2.7$W), one can only keep the first term in the previous series expansion at the denominator. We have

$$C = \int_{-B/2}^{B/2} \log_2 \left(1 + \frac{\frac{S_X(f)}{1+\gamma_{NL}^2 L_{\text{eff}}^2 R_X^2(0)}}{S_N(f) + \frac{2\gamma_{NL}^2 L_{\text{eff}}^2 FT(R_X(\tau)|R_X(\tau)|^2)}{(1+\gamma_{NL}^2 L_{\text{eff}}^2 R_X^2(0))^3}} \right) df. \tag{1.39}$$

Unlike [11], we decide to keep the integral in the further expression for the capacity since our spectrum is non flat due to the presence of the guard band. To complete the derivations, we just need to obtain a closed-form expression for $FT(R_X(\tau)|R_X(\tau)|^2)$.

1.5.2.1 Derivations of $FT(R_X(\tau)|R_X(\tau)|^2)$

As $S_X(f)$ in Fig. 1.3 admits a Hermitian symmetry, the autocorrelation $R_X(\tau)$ is a real-valued function. Therefore deriving $FT(R_X(\tau)|R_X(\tau)|^2)$ is equivalent to deriving $FT(R(\tau)^3)$.

Thanks to the properties of the Fourier transform, the term can be calculated as follows:

$$T_X(f) = FT(R_X(\tau)^3) = S_X \star S_X \star S_X(f). \tag{1.40}$$

where $\star$ stands for the convolution product. We will proceed into two steps: first we derive $v(f) = S_X \star S_X(f)$, and then $T_X(f) = S_X \star v(f)$.

Derivations of $v(f)$

First of all, let $U(f) = \Pi * \Pi(f)$ be the convolution between two rectangular functions. It is well known that $v_2(f)$ is the following triangle function

$$U(f) = \begin{cases} (1 - |f|) & \text{if } |f| < 1 \\ 0 & \text{elsewhere} \end{cases} \tag{1.41}$$

where the function $x \mapsto |x|$ is the absolute value function. As a consequence, we have

$$v(f) = \frac{P_x^2}{B} \sum_{j=-N}^{N} \sum_{k=-N}^{N} U\left(\frac{f + (j+k)(B + B_g)}{B}\right). \tag{1.42}$$

The double sum can be simplified by using the indices $(j, \ell = j + k)$ instead of (j, k). Then, we have $-2N \leq \ell \leq 2N$ and $\max(-N + \ell, -N) \leq j \leq \min(N + \ell, N)$. Therefore we obtain

$$v(f) = \frac{P_x^2}{B^2} \sum_{\ell=-2N}^{2N} (2N + 1 - |\ell|) B U\left(\frac{f + \ell(B + B_g)}{B}\right). \tag{1.43}$$

Derivations of $T_X(f)$

Let $W(f) = \Pi \star \Pi \star \Pi(f) = \Pi \star U(f)$ be the double convolution of the rectangular function. We have straightforwardly that

$$W(f) = \begin{cases} \frac{3}{4} - f^2 & \text{if } |f| < \frac{1}{2} \\ \frac{1}{2} f^2 - \frac{3}{2}|f| + \frac{9}{8} & \text{if } \frac{1}{2} < |f| < \frac{3}{2} \\ 0 & \text{otherwise} \end{cases}. \tag{1.44}$$

As

$$T_x(f) = \left(\frac{P_x}{B} \sum_{k=-N}^{N} \Pi\left(\frac{f + k(B + B_g)}{B}\right)\right) \star \left(\underbrace{\frac{P_x^2}{B^2} \sum_{\ell=-2N}^{2N} (2N + 1 - |\ell|) B U\left(\frac{f + \ell(B + B_g)}{B}\right)}_{v(f)}\right),$$

we have

$$T_X(f) = \frac{P_x^3}{B^3} \sum_{k=-N}^{N} \sum_{\ell=-2N}^{2N} (2N + 1 - |\ell|) B^2 W\left(\frac{f + (k + \ell)(B + B_g)}{B}\right). \tag{1.45}$$

We use again the substitution $(\ell' = k + \ell, \ell)$ instead of (k, ℓ). Then we have $-3N \le \ell' \le 3N$. As $-N \le k \le N$, we obtain the following inequality on ℓ

$$-N + \ell' \le \ell \le \ell' + N.$$

Moreover, we have $-2N \le \ell \le 2N$. So we obtain that

$$\underbrace{\max(-N + \ell', -2N)}_{a} \le \ell \le \underbrace{\min(\ell' + N, 2N)}_{b}. \tag{1.46}$$

Then $T_X(f)$ takes the following form

$$T_X(f) = \frac{P_x^3}{B} \sum_{\ell'=-3N}^{3N} \underbrace{\left(\sum_{\ell=a}^{b} (2N + 1 - |\ell|) \right)}_{\theta} W\left(\frac{f + \ell'(B + B_g)}{B} \right). \tag{1.47}$$

In order to handle easily the bounds for ℓ given by Eq. (1.46), we split our problem in two categories

- For $N < |\ell'| \le 3N$:

$$\theta = \sum_{\ell=|\ell'|-N}^{2N} (2N + 1 - |\ell|) = \sum_{\ell''=1}^{3N+1-|\ell'|} \ell'' = \frac{(3N + 1 - |\ell'|)(3N + 2 - |\ell'|)}{2},$$

- For $0 \le |\ell'| \le N$:

$$\begin{aligned} \theta = \sum_{\ell=\ell'-N}^{\ell'+N} (2N + 1 - |\ell|) &= \left(\sum_{\ell=1}^{\ell'+N \ge 0} (2N + 1 - |\ell|) \right) + \left(\sum_{\ell=0}^{N-\ell' \ge 0} (2N + 1 - |\ell|) \right) \\ &= \left(\sum_{\ell''=N+1-\ell'}^{2N} \ell'' \right) + \left(\sum_{\ell''=N+1+\ell'}^{2N+1} \ell'' \right) \\ &= (2N + 1)^2 - \frac{(N + \ell')(N + \ell' + 1)}{2} - \frac{(N - \ell')(N - \ell' + 1)}{2}. \end{aligned}$$

By replacing θ in Eq. (1.47) with the previous expressions, we have

$$T_X(f) = \frac{P_x^3}{B} \widetilde{T}_X(f)$$

with

$$\begin{aligned} \widetilde{T}_X(f) &= \sum_{k=N+1}^{3N} \frac{(3N + 1 - k)(3N + 2 - k)}{2} (W_k(f) + W_{-k}(f)) \\ &+ \sum_{k=1}^{N} \left((2N + 1)^2 - \frac{(N + k)(N + k + 1)}{2} - \frac{(N - k)(N - k + 1)}{2} \right) (W_k(f) + W_{-k}(f)) \\ &+ ((2N + 1)^2 - N(N + 1)) W_0(f) \end{aligned} \tag{1.48}$$

where the function $f \mapsto W_k(f)$ is defined as

$$W_k(f) = W\left(\frac{f + k(B + B_g)}{B} \right).$$

1.5.2.2 Final expression for non-dispersive case

According to Eq. (1.39) and as $S_X(f) = (P_x/B)$ for $f \in [-B/2, B/2]$ and $S_N(f) = P_N/B$ and $R_X(0) = P_x N_c$, we have the following closed-form expressions for the capacity

$$C = \int_{-B/2}^{B/2} \log_2 \left(1 + \frac{\frac{P_x}{1+(\gamma_{NL} L_{\text{eff}} N_c P_x)^2}}{P_N + P_x \frac{2(\gamma_{NL} L_{\text{eff}} P_x)^2 \widetilde{T}_X(f)}{(1+(\gamma_{NL} L_{\text{eff}} N_c P_x)^2)^3}} \right) df. \tag{1.49}$$

where $\widetilde{T}_x(f)$ is given in Eq. (1.48).

In this expression, the guard band is involved only in $\widetilde{T}_X(f)$. In $B_g = 0$, it is easy to check our expression is that given in [11]. If $B_g \geq B$, we have $\widetilde{T}(f) = (N_c^2 - (N_c^2 - 1)/4) W(f/B)$ since $W_k(f) = 0$ for any $k \neq 0$ as soon as $|f| < B/2$. As a consequence, we obtain that

$$C_{B_g > B} = \int_{-B/2}^{B/2} \log_2 \left(1 + \frac{\frac{P_x}{1+(\gamma_{NL} L_{\text{eff}} N_c P_x)^2}}{P_N + P_x \frac{(3N_c^2+1)(\gamma_{NL} L_{\text{eff}} P_x)^2 (3/4 - (f/B)^2)}{2(1+(\gamma_{NL} L_{\text{eff}} N_c P_x)^2)^3}} \right) df. \tag{1.50}$$

This value $C_{B_g > B}$ corresponds to the maximum value for the capacity since it is a non-decreasing function with respect to B_g. In $C_{B_g > B}$, we only viewed the non-linear effect due to intra-channel interference since the inter-channel effect has been completely removed via the guard band.

1.5.3 Dispersive case

The dispersive case is much complicate to treat since there is no closed-form solution for the NSE. To go further, once again, we use the framework developed in [13] Nevertheless, we remind that the expressions given in the dispersive case in [13] does not take into account the presence of a potential guard band, and was unfortunately flawed as already explained in Section 1.4.2. Therefore we propose here to do the whole derivations by correcting the expression of [13] and by inserting a guard band. For the sake of simplicity, a single-span link is considered.

The main issue so lies in the absence of closed-form solution for the NSE. Therefore, as in [13], we will consider that the Input/Output relationship is obtained by using a truncated Volterra series expansion (cf. Eq. (1.27)). We remind that the output signal (in time domain) can be split into two terms:

$$Y(t) \simeq Y_L(t) + Y_{NL}(t) \tag{1.51}$$

where $Y_L(t)$ represents the linear part of the received signal (its Fourier transform is given in Eq. (1.28)), and $Y_{NL}(t)$ represents the non-linear impairments and the noise (its Fourier transform is given in Eq. (1.29)).

Using Pinsker's formula (cf. Eq. (1.22)), we also remind that

$$C = \int_{-B/2}^{B/2} \log_2 \left(\frac{P(f)}{Q(f)} \right) df \tag{1.52}$$

where

$$P(f) = \det\left(\begin{bmatrix} FT(\mathbb{E}[X(t+\tau)X^*(t)]) & 0 \\ 0 & FT(\mathbb{E}[Y(t+\tau)Y^*(t)]) \end{bmatrix}\right)$$

and

$$Q(f) = \det\left(\begin{bmatrix} FT(\mathbb{E}[X(t+\tau)X^*(t)]) & FT(\mathbb{E}[X(t+\tau)Y^*(t)]) \\ FT(\mathbb{E}[Y(t+\tau)X^*(t)]) & FT(\mathbb{E}[Y(t+\tau)Y^*(t)]) \end{bmatrix}\right).$$

In the remainder of this chapter, we will use the following notation and definition for the so-called inter-spectrum between signals $a(t)$ and $b(t)$:

$$S_{a,b}(f) = FT(\mathbb{E}[b(t+\tau)a^*(t)]).$$

Then, by using Eq. (1.51), we have

$$P(f) = S_X(f)\left(S_{Y_L,Y_L}(f) + S_{Y_{NL},Y_L}(f) + S_{Y_L,Y_{NL}}(f) + S_{Y_{NL},Y_{NL}}(f)\right) \tag{1.53}$$

and

$$Q(f) = P(f) - \Theta(f) \tag{1.54}$$

with

$$\Theta(f) = S_{X,Y_L}(f)S_{Y_L,X}(f) + S_{X,Y_L}(f)S_{Y_{NL},X}(f) + S_{Y_L,X}(f)S_{X,Y_{NL}}(f) + S_{X,Y_{NL}}(f)S_{Y_{NL},X}(f). \tag{1.55}$$

As a consequence, Eq. (1.52) can be rewritten as follows

$$C = \int_{-B/2}^{B/2} \log_2\left(1 + \frac{\Theta(f)}{P(f) - \Theta(f)}\right) df. \tag{1.56}$$

Notice that in previous expression is clearly different from that of [13] since $\Theta(f)$ is not equal to $S_X(f)$, and actually vanishes when an opaque channel is considered (as $Y_L(t) = 0$ and $Y_{NL}(t)$ only depends on independent noise for such a channel).

In order to obtain closed-form expressions for $P(f)$ and $\Theta(f)$, we need three simple preliminary results given below and proven in Appendix A.

Preliminary result 1 (Spectral autocorrelation) *We have*

$$\mathbb{E}[X(f)X^*(f')] = \delta(f - f')S_X(f)$$

where $\delta(.)$ is the Dirac distribution.

Preliminary result 2 (Interference formula) *Let $a(t)$ and $b(t)$ be the outputs of two filters having the same input $x(t)$. So*

$$a(t) = h_a(t) \star x(t) \quad \text{and} \quad b(t) = h_b(t) \star x(t).$$

We have

$$S_{a,b}(f) = H_a(f)^* H_b(f) S_X(f)$$

where $H_a(f)$ and $H_b(f)$ are the Fourier transform of $a(t)$ and $b(t)$ respectively.

Preliminary result 3 (Permutation in inter-spectra) *Let $a(t)$ and $b(t)$ be the random process. We have*

$$S_{a,b}(f) = S_{b,a}(f)^*.$$

Thanks to preliminary Result 3, we have

$$P(f) = S_X(f)\left(S_{Y_L,Y_L}(f) + 2\Re[S_{Y_L,Y_{NL}}(f)] + S_{Y_{NL},Y_{NL}}(f)\right)$$

and

$$\Theta(f) = |S_{X,Y_L}(f)|^2 + 2\Re[S_{X,Y_L}(f)S_{X,Y_{NL}}(f)^*] + |S_{X,Y_{NL}}(f)|^2.$$

The inter-spectra involved in $P(f)$ and $\Theta(f)$ can be split into two categories according to the presence of the non-linear term $Y_{NL}(t)$. Therefore, we have divided our derivations in two parts: Section 1.5.3.1 (only linear terms) and Section 1.5.3.2 (with non-linear term).

1.5.3.1 Spectra derivations involving linear terms

As $Y_L(f) = H_{CD}(f)X(f)$, we have

$$S_{Y_L,Y_L}(f) = |H_{CD}(f)|^2 S_X(f), \quad \text{and} \quad S_{X,Y_L}(f) = H_{CD}(f)S_X(f).$$

Using the CD filter (cf. Eq. (1.13)), we finally obtain that

$$S_{Y_L,Y_L}(f) = S_X(f), \quad \text{and} \quad S_{X,Y_L}(f) = e^{2i\pi^2\beta_2 L f^2} S_X(f).$$

This enables us to simplify $P(f)$, $\theta(f)$, and $Q(f)$ as follows

$$P(f) = S_X(f)(S_X(f) + 2\Re[S_{Y_L,Y_{NL}}(f)] + S_{Y_{NL},Y_{NL}}(f)),$$

$$\Theta(f) = S_X(f)^2 + 2\Re[S_{X,Y_L}(f)S_{X,Y_{NL}}(f)^*] + |S_{X,Y_{NL}}(f)|^2,$$

and

$$Q(f) = S_X(f)(2\Re[S_{Y_L,Y_{NL}}(f)] + S_{Y_{NL},Y_{NL}}(f)) - 2\Re[S_{X,Y_L}(f)S_{X,Y_{NL}}(f)^*] - |S_{X,Y_{NL}}(f)|^2.$$

1.5.3.2 Spectra derivations involving non-linear terms

• **Derivations for $S_{X,Y_{NL}}(f)$**

We have

$$\begin{aligned} S_{X,Y_{NL}}(f) &= FT\left(\mathbb{E}\left[\int Y_{NL}(f)e^{2i\pi f(t+\tau)}df X^*(t)\right]\right) \\ &= FT\left(\int \mathbb{E}\left[Y_{NL}(f)X^*(t)\right]e^{2i\pi f(t+\tau)}df\right) \end{aligned} \tag{1.57}$$

We will first focus on $\mathbb{E}\left[Y_{NL}(f)X^*(t)\right]$. We have

$$\begin{aligned} \mathbb{E}\left[Y_{NL}(f)X^*(t)\right] &= \mathbb{E}\Big[\Big(e^{2i\pi^2\beta_2 L f^2}(2i\pi)\gamma_{NL}\iiint \delta(f - f_1 + f_2 - f_3)X(f_1)X^*(f_2)X(f_3) \\ &\times \nu(f_1, f_2, f_3)df_1df_2df_3 + N(f)\Big)X^*(t)\Big]. \end{aligned}$$

with

$$\nu(f_1,f_2,f_3)=\frac{e^{L\left(-\alpha+2i\pi^2\beta_2(-f^2+f_1^2-f_2^2+f_3^2)\right)}-1}{-\alpha+2i\pi^2\beta_2(-f^2+f_1^2-f_2^2+f_3^2)}.$$

As the noise is independent of the signal, we have

$$\begin{aligned}\mathbb{E}\left[Y_{NL}(f)X^*(t)\right] &= \mathbb{E}\Big[\Big(e^{2i\pi^2\beta_2Lf^2}(2i\pi)\gamma_{NL}\iiint\delta(f-f_1+f_2-f_3)X(f_1)X^*(f_2)X(f_3)\\ &\times\ \nu(f_1,f_2,f_3)df_1df_2df_3\Big)X^*(t)\Big].\end{aligned}$$

By moving the expectation only around the random terms, we obtain that

$$\begin{aligned}\mathbb{E}\left[Y_{NL}(f)X^*(t)\right] &= e^{2i\pi^2\beta_2Lf^2}(2i\pi)\gamma_{NL}\iiint\delta(f-f_1+f_2-f_3)\mathbb{E}\Big[X(f_1)X^*(f_2)X(f_3)X^*(t)\Big]\\ &\times\ \nu(f_1,f_2,f_3)df_1df_2df_3.\end{aligned}\tag{1.58}$$

To go further, we need to calculate the expectation of a product of four random variables. Assuming that the (complex-valued) random variable are zero-mean circularly-symmetric Gaussian, it is known (cf. [23] for more details) that

$$\begin{aligned}\mathbb{E}\left[X(f_1)X^*(f_2)X(f_3)X^*(t)\right] &= \mathbb{E}\left[X(f_1)X^*(f_2)\right]\mathbb{E}\left[X(f_3)X^*(t)\right]\\ &+\ \mathbb{E}\left[X(f_1)X^*(t)\right]\mathbb{E}\left[X^*(f_2)X(f_3)\right]\end{aligned}$$

Notice that the assumption of circularly symmetry is not restrictive since all the encountered modulations (PSK except BPSK, QAM) are second-order circularly symmetric. Circularly symmetry at the second order means that the real part is independent of the imaginary part and that the variance of the real part is identical to the variance of the imaginary part.

As for the preliminary Result 1, it is easy to see that

$$\begin{aligned}\mathbb{E}\left[X(f)X^*(t)\right] &= \mathbb{E}\left[\int_{-\infty}^{\infty}X(t')e^{-2i\pi ft'}dtX^*(t)\right]\\ &= \int\mathbb{E}\left[X(t')X^*(t)\right]e^{-2i\pi ft'}dt'\\ &= \int R_x(t'-t)e^{-2i\pi t'}dt'\\ &= S_X(f)e^{-2i\pi ft},\end{aligned}\tag{1.59}$$

Combining Eq. (1.59) and that dealing with the 4-th order moment decomposition along with the preliminary Result 1 , we have

$$\begin{aligned}\mathbb{E}\left[X(f_1)X^*(f_2)X(f_3)X^*(t)\right] &= e^{-2i\pi f_3t}\delta(f_1-f_2)S_X(f_1)S_X(f_3)\\ &+\ e^{-2i\pi f_1t}\delta(f_2-f_3)S_X(f_1)S_X(f_2).\end{aligned}$$

Thanks to the previous expression, Eq. (1.58) can be rewritten as follows

$$\mathbb{E}\left[Y_{NL}(f)X^*(t)\right] = e^{2i\pi^2\beta_2Lf^2}4i\pi\gamma_{NL}L_{\text{eff}}S_X(f)R_X(0)e^{-2i\pi ft}\tag{1.60}$$

with $R_X(0)=\int S_X(f)df$. Moreover as $R_X(0)=N_cP_x$, we have

$$S_{X,Y_{NL}}(f)=4i\pi\gamma_{NL}L_{\text{eff}}e^{2i\pi^2\beta_2Lf^2}N_cP_xS_X(f).\tag{1.61}$$

• **Derivations for** $S_{Y_L,Y_{NL}}(f)$

Thanks to Eq. (1.28), we remind that

$$Y_L(f) = e^{2i\pi^2\beta_2 L f^2} X(f). \tag{1.62}$$

Following the same approach than for the derivations of $S_{X,Y_{NL}}(f)$, one can easily deduce that

$$S_{Y_L,Y_{NL}}(f) = e^{-2i\pi^2\beta_2 L f^2} S_{X,Y_{NL}}(f).$$

and thus

$$S_{Y_L,Y_{NL}}(f) = 4i\pi\gamma_{NL} L_{\text{eff}} N_c P_x S_X(f).$$

which corresponds to an imaginary number.

As a consequence, we can once again simplify strongly $P(f)$ and $Q(f)$ as follows

$$P(f) = S_X(f)^2 + S_X(f) S_{Y_{NL},Y_{NL}}(f), \tag{1.63}$$

$$\Theta(f) = S_X(f)^2 + 16\pi^2\gamma_{NL}^2 L_{\text{eff}}^2 N_c^2 P_x^2 S_X(f)^2, \tag{1.64}$$

and

$$Q(f) = S_X(f) S_{Y_{NL},Y_{NL}}(f) - 16\pi^2\gamma_{NL}^2 L_{\text{eff}}^2 N_c^2 P_x^2 S_X(f)^2. \tag{1.65}$$

So it remains only to derive the term $S_{Y_{NL},Y_{NL}}(f)$ which is the most complicate one.

• **Derivations for** $S_{Y_{NL},Y_{NL}}(f)$

We obviously have that

$$S_{Y_{NL},Y_{NL}}(f) = FT\left(\mathbb{E}\left[\int Y_{NL}(f) e^{2i\pi f(t+\tau)} df \int Y_{NL}^*(f') e^{-2i\pi f'(t)} df'\right]\right).$$

By applying the expectation only on the random terms, we obtain

$$S_{Y_{NL},Y_{NL}}(f) = FT\left(\iint \mathbb{E}\left[Y_{NL}(f) Y_{NL}^*(f')\right] e^{2i\pi f(t+\tau)} e^{-2i\pi f' t} df df'\right) \tag{1.66}$$

Let us focus on the term $\mathbb{E}\left[Y_{NL}(f) Y_{NL}^*(f')\right]$. According to Eq.(1.29), we have

$$\begin{aligned}\mathbb{E}\left[Y_{NL}(f) Y_{NL}^*(f')\right] =& \mathbb{E}\Bigg[(e^{2i\pi^2\beta_2 L f^2}(2i\pi\gamma_{NL}) \iiint \delta(f - f_1 + f_2 - f_3) X(f_1) X^*(f_2) X(f_3) \\ & v(f_1, f_2, f_3) df_1 df_2 df_3] + N(f)) \\ & \times (e^{-2i\pi^2\beta_2 L f'^2}(-2i\pi\gamma_{NL}) \iiint \delta(f' - f_1' + f_2' - f_3') X^*(f_1') X(f_2') X^*(f_3') \\ & v(f_1', f_2', f_3')^* df_1' df_2' df_3' + N(f')^*)\Bigg].\end{aligned}$$

Since the amplification noise is statistically independent from the signal, the previous expression becomes (thanks to the preliminary Result 1)

$$\begin{aligned}\mathbb{E}\left[Y_{NL}(f)Y_{NL}^{*}(f')\right] =&\mathbb{E}\Bigg[(e^{2i\pi^2\beta_2 Lf^2}(2i\pi\gamma_{NL})\iiint \delta(f-f_1+f_2-f_3)X(f_1)X^{*}(f_2)X(f_3)\\ &v(f_1,f_2,f_3)df_1df_2df_3])\\ &\times(e^{-2i\pi^2\beta_2 Lf'^2}(-2i\pi\gamma_{NL})\iiint \delta(f'-f_1'+f_2'-f_3')X^{*}(f_1')X(f_2')X^{*}(f_3')\\ &v(f_1',f_2',f_3')^{*}df_1'df_2'df_3')\Bigg]+\delta(f-f')S_N(f).\end{aligned}$$

Once again, by applying the expectation only to the random terms, we obtain

$$\begin{aligned}\mathbb{E}\left[Y_{NL}(f)Y_{NL}^{*}(f')\right] =&\delta(f-f')S_N(f)+4\pi^2\gamma_{NL}^2e^{2i\pi^2\beta_2L(f^2-f'^2)}\\ &\iiint\iiint \delta(f-f_1+f_2-f_3)\delta(f'-f_1'+f_2'-f_3')\\ &\mathbb{E}\left[X(f_1)X^{*}(f_2)X(f_3)X^{*}(f_1')X(f_2')X^{*}(f_3')\right]\\ &v(f_1,f_2,f_3)v(f_1',f_2',f_3')^{*}df_1df_2df_3df_1'df_2'df_3'\end{aligned}\tag{1.67}$$

In order to go further, we need to treat the expectation involving the product of six random variables. Once again, by assuming zero-mean circularly-symmetric Gaussian process, we obtain (see [23])

$$\begin{aligned}\mathbb{E}\left[X(f_1)X^{*}(f_2)X(f_3)X^{*}(f_1')X(f_2')X^{*}(f_3')\right] &= \mathbb{E}\left[X^{*}(f_2)X(f_3)\right]\mathbb{E}\left[X^{*}(f_1')X(f_1)\right]\mathbb{E}\left[X(f_2')X^{*}(f_3')\right]\\ &+ \mathbb{E}\left[X(f_1)X^{*}(f_2)\right]\mathbb{E}\left[X^{*}(f_1')X(f_3)\right]\mathbb{E}\left[X(f_2')X^{*}(f_3')\right]\\ &+ \mathbb{E}\left[X^{*}(f_3')X(f_3)\right]\mathbb{E}\left[X^{*}(f_1')X(f_2')\right]\mathbb{E}\left[X(f_1)X^{*}(f_2)\right]\\ &+ \mathbb{E}\left[X^{*}(f_3')X(f_3)\right]\mathbb{E}\left[X^{*}(f_1')X(f_1)\right]\mathbb{E}\left[X(f_2')X^{*}(f_2)\right]\\ &+ \mathbb{E}\left[X^{*}(f_3')X(f_1)\right]\mathbb{E}\left[X^{*}(f_1')X(f_2')\right]\mathbb{E}\left[X^{*}(f_2)X(f_3)\right]\\ &+ \mathbb{E}\left[X(f_2')X^{*}(f_2)\right]\mathbb{E}\left[X^{*}(f_1')X(f_3)\right]\mathbb{E}\left[X^{*}(f_3')X(f_1)\right]\end{aligned}\tag{1.68}$$

We remark that the involved 6-th order moment is a sum of terms $\chi_1 = \mathbb{E}[X^{*}(f_3')X(f_3)]$ $\times\mathbb{E}[X^{*}(f_1')X(f_1)]\mathbb{E}[X(f_2')X^{*}(f_2)]$ or $\mathbb{E}[X(f_2')X^{*}(f_2)]\mathbb{E}[X^{*}(f_1')X(f_3)]\mathbb{E}[X^{*}(f_3')X(f_1)]$ (in which one find one frequency with the prime per second order moment) and χ_2 equal to the four other ones (in which one find at least one second-order moment with both frequencies with the prime). These two types of terms (χ_1 and χ_2) have to be calculated separately as follows.

- Let us focus on the types of term χ_1. Let us first assume that $\chi_1 = \mathbb{E}[X^{*}(f_3')X(f_3)]$ $\times\mathbb{E}[X^{*}(f_1')X(f_1)]\mathbb{E}[X(f_2')X^{*}(f_2)]$ The corresponding term in Eq.(1.67) is denoted by

$t_1(f,f')$ and writes as follows

$$\begin{aligned} t_1(f,f') &= 4\pi^2\gamma_{NL}^2 e^{2i\pi^2\beta_2 L(f^2-f'^2)} \\ &\times \iiint\!\!\iiint \delta(f-f_1+f_2-f_3)\delta(f'-f_1'+f_2'-f_3') \\ &\times \mathbb{E}[X^*(f_3')X(f_3)]\mathbb{E}[X^*(f_1')X(f_1)]\mathbb{E}[X(f_2')X^*(f_2)] \\ &\times v(f_1,f_2,f_3)v(f_1',f_2',f_3')^* df_1 df_2 df_3 df_1' df_2' df_3' \end{aligned}$$

Using the preliminary Result 1 leads to

$$\begin{aligned} t_1(f,f') &= 4\pi^2\gamma_{NL}^2 e^{2i\pi^2\beta_2 L(f^2-f'^2)} \\ &\times \iiint\!\!\iiint \delta(f-f_1+f_2-f_3)\delta(f'-f_1'+f_2'-f_3') \\ &\times \delta(f_1'-f_1)S_X(f_1)\delta(f_2'-f_2)S_X(f_2)\delta(f_3'-f_3)S_X(f_3) \\ &\times v(f_1,f_2,f_3)v(f_1',f_2',f_3')^* df_1 df_2 df_3 df_1' df_2' df_3' \end{aligned}$$

Then we have

$$\begin{aligned} t_1(f,f') &= 4\pi^2\gamma_{NL}^2\delta(f-f')\iiint \delta(f-f_1+f_2-f_3) \\ &\times S_X(f_1)S_X(f_2)S_X(f_3)\left|v(f_1,f_2,f_3)\right|^2 df_1 df_2 df_3 \end{aligned} \tag{1.69}$$

Notice that the derivations done with the other expression of χ_1 will lead to the same final closed-form expression.

- Let us focus on the types of term χ_2. Let us first assume that $\chi_2 = \mathbb{E}[X^*(f_1')X(f_2')]$ $\times\mathbb{E}[X^*(f_3')X(f_1)]\mathbb{E}[X^*(f_2)X(f_3)]$. The corresponding term in Eq.(1.67) is denoted by $t_2(f,f')$ and writes as follows

$$\begin{aligned} t_2(f,f') &= 4\pi^2\gamma_{NL}^2 e^{2i\pi^2\beta_2 L(f^2-f'^2)} \\ &\times \iiint\!\!\iiint \delta(f-f_1+f_2-f_3)\delta(f'-f_1'+f_2'-f_3') \\ &\times \mathbb{E}[X^*(f_1')X(f_2')]\mathbb{E}[X^*(f_3')X(f_1)]\mathbb{E}[X^*(f_2)X(f_3)] \\ &\times v(f_1,f_2,f_3)v(f_1',f_2',f_3')^* df_1 df_2 df_3 df_1' df_2' df_3' \end{aligned}$$

Using the preliminary Result 1 leads to

$$\begin{aligned} t_2(f,f') &= 4\pi^2\gamma_{NL}^2 e^{2i\pi^2\beta_2 L(f^2-f'^2)} \\ &\times \iiint\!\!\iiint \delta(f-f_1+f_2-f_3)\delta(f'-f_1'+f_2'-f_3') \\ &\times \delta(f_1'-f_2')S_X(f_1')\delta(f_3'-f_1)S_X(f_1)\delta(f_2-f_3)S_X(f_3) \\ &\times v(f_1,f_2,f_3)v(f_1',f_2',f_3')^* df_1 df_2 df_3 df_1' df_2' df_3' \end{aligned}$$

The sixtuple integral can be strongly simplified by using the Dirac delta distribution

$$\begin{aligned} t_2(f,f') &= 4\pi^2\gamma_{NL}^2 e^{2i\pi^2\beta_2 L(f^2-f'^2)} \\ &\times \delta(f-f')L_{\text{eff}}^2 S_X(f)\left|\int S_X(f_1)df_1\right|^2. \end{aligned}$$

Then, we have

$$t_2(f,f') = 4\pi^2\gamma_{NL}^2 L_{\text{eff}}^2 N_c^2 P_x^2 \delta(f-f') S_X(f). \tag{1.70}$$

Notice that the derivations done with the three other expression of χ_1 will lead to the same final closed-form expression.

Finally, incorporating in Eq. (1.67) the two terms associated with Eq. (1.69) and the four terms associated with Eq. (1.70) leads to

$$\begin{aligned}\mathbb{E}\left[Y_{NL}(f)Y_{NL}^*(f')\right] &= \delta(f-f')S_N(f) \\ &+ 8\pi^2\gamma_{NL}^2\delta(f-f')\iiint \delta(f-f_1+f_2-f_3) \\ &\times S_X(f_1)S_X(f_2)S_X(f_3)\left|v(f_1,f_2,f_3)\right|^2 df_1 df_2 df_3 \\ &+ 16\pi^2\gamma_{NL}^2 L_{\text{eff}}^2 N_c^2 P_x^2 \delta(f-f')S_X(f).\end{aligned} \tag{1.71}$$

As there is a Dirac distribution on f and f' for all terms in Eq. (1.71), according to Eq. (1.66), the term $S_{Y_{NL},Y_{NL}}(f)$ is just equal to $\mathbb{E}[Y_{NL}(f)Y_{NL}^*(f)]$. Therefore, we have

$$\begin{aligned}S_{Y_{NL},Y_{NL}}(f) &= S_N(f) + 8\pi^2\gamma_{NL}^2 \iiint \delta(f-f_1+f_2-f_3) \\ &\times S_X(f_1)S_X(f_2)S_X(f_3)\left|v(f_1,f_2,f_3)\right|^2 df_1 df_2 df_3 \\ &+ 16\pi^2\gamma_{NL}^2 L_{\text{eff}}^2 N_c^2 P_x^2 S_X(f).\end{aligned} \tag{1.72}$$

1.5.3.3 Final closed-form expression for the capacity

Based on the simplified expression of $\Theta(f)$ and $Q(f)$ given in Eqs. (1.64)-(1.65), the capacity (given by Eq. (1.56)) writes now as follows

$$C = \int_{-B/2}^{B/2} \log_2\left(1 + \frac{S_X(f) + 16\pi^2\gamma_{NL}^2 L_{\text{eff}}^2 N_c^2 P_x^2 S_X(f)}{S_{Y_{NL},Y_{NL}}(f) - 16\pi^2\gamma_{NL}^2 L_{\text{eff}}^2 N_c^2 P_x^2 S_X(f)}\right) df \tag{1.73}$$

Using Eq. (1.72), we have

$$C = \int_{-B/2}^{B/2} \log_2\left(1 + \frac{S_X(f)(1 + 16\pi^2\gamma_{NL}^2 L_{\text{eff}}^2 N_c^2 P_x^2)}{S_N(f) + M_X(f)}\right) df$$

with

$$M_X(f) = 8\pi^2\gamma_{NL}^2 \iiint \delta(f-f_1+f_2-f_3)S_X(f_1)S_X(f_2)S_X(f_3)\left|v(f_1,f_2,f_3)\right|^2 df_1 df_2 df_3.$$

Notice that [13] has almost the same denominator (we just have to replace the frequency with the pulsation and to replace the power defined in [13] with 2π times ours). In contrast, the numerator is different since [13] omitted the positive term equal to $16\pi^2\gamma_{NL}^2 L_{\text{eff}}^2 N_c^2 P_x^2$. This additional term will improve the capacity and so the value of its maximum. Notice

also that $M_X(f)$ depends on the CD coefficient which proves that the non-linear capacity is modified by the CD.

Once again we can simplify the capacity by forcing the input spectrum to satisfy the model provided in Eq. (1.36). Then we have

$$C = \int_{-B/2}^{B/2} \log_2\left(1 + \frac{P_x(1 + 16\pi^2\gamma_{NL}^2 L_{\text{eff}}^2 N_c^2 P_x^2)}{P_N + \widetilde{M}_X(f)}\right) df \tag{1.74}$$

with

$$\begin{aligned}\widetilde{M}_X(f) &= 8\pi^2\gamma_{NL}^2(P_x^3/B^2) \sum_{k_1,k_2,k_3=-N}^{N} \iiint \delta(f - f_1 + f_2 - f_3) \\ &\times \Pi\left(\frac{f_1 + k_1(B+B_g)}{B}\right)\Pi\left(\frac{f_2 + k_2(B+B_g)}{B}\right)\Pi\left(\frac{f_3 + k_3(B+B_g)}{B}\right)\left|v(f_1,f_2,f_3)\right|^2 df_1 df_2 df_3.\end{aligned}$$

1.6 Numerical illustrations

In the previous section, we have obtained closed-form expressions of the non-linear optical channel capacity. Hereafter, we illustrate numerically the introduced expressions in order to analyze them and to check them as well.

This section is split into three subsections: in Section 1.6.1, we introduce the simulation setup. In Section 1.6.2, we check the accuracy of the obtained closed-form expressions compared to a Monte-Carlo capacity evaluation. Finally, in Section 1.6.3 , we analyze the influence of the guard band on the capacity.

1.6.1 Simulation setup

Parameters	Value
Fiber loss	$\alpha = 0.2$ dB/km
Kerr coefficient	$\gamma_{NL} = 1.22\ \text{W}^{-1} \cdot \text{km}^{-1}$
CD coefficient	$\beta_{LEAF} = -5.1\ \text{ps}^2/\text{km}$ $\beta_{SMF} = -21.6\ \text{ps}^2/\text{km}$
Fiber length	$L = 80$ km
Number of channel	$N_c = 81$
Useful bandwidth	$B = 50$ GHz
Amplification noise power	$P_N = 10^{-5}\ \text{mW} \cdot \text{GHz}^{-1}$

Table 1.2: Simulation setup

As for the derivations, we consider a mono-span amplified optical link. The amplifier is at the receiver side and perfectly compensates for the fiber loss. Concerning the CD, we consider three fiber types: non-dispersive fiber, LEAF fiber, and standard SMF fiber.

The other parameters, such as fiber loss or Kerr coefficient, are assumed to be the same for all fiber types and are displayed in Table 1.2. Moreover the amplifier only generates AWGN.

1.6.2 Accuracy of the proposed closed-form expressions

First of all, we consider the non-dispersive case. In Fig. 1.4, we plot the capacity obtained through numerical Monte-Carlo simulation of the Pinsker's formula and through the numerical evaluation of the closed-form expression of the capacity (*i.e.*, Eq. (1.49)). The

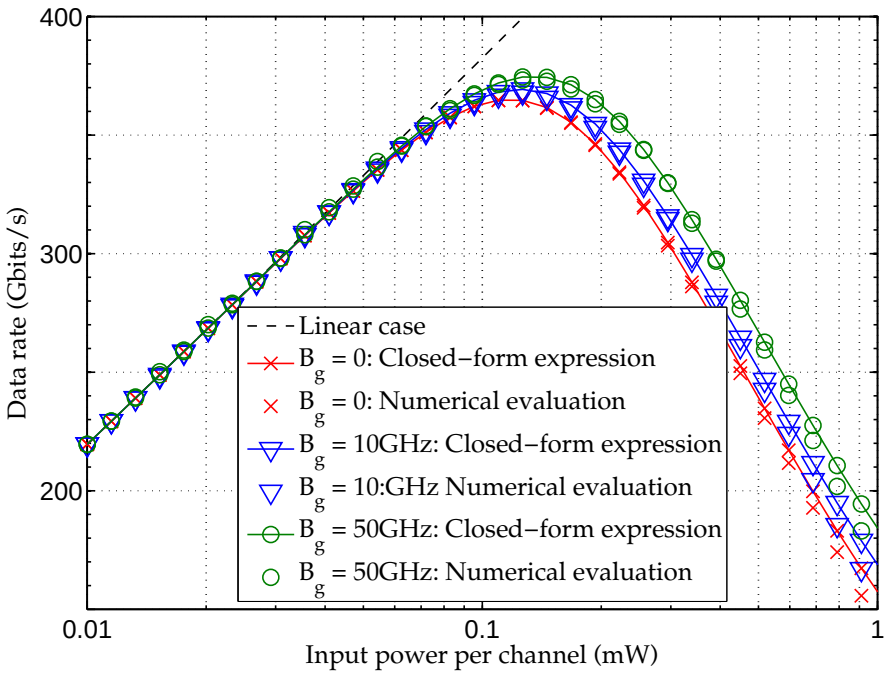

Figure 1.4: Capacity versus P_x in non-dispersive case for different values of the guard band

closed-form expressions are very accurate, especially around the maximum capacity. Indeed, when the input power is too high, a slight mismatch appears between the Monte-Carlo simulation and the closed-form expression. This can be explain by the fact we have neglected some terms in Eq. (1.38). This approximation is perfectly valid if the power is small enough.

Let us now focus on the dispersive case. Unlike the non-dispersive case, the numerical Monte-Carlo evaluation of Pinsker's Formula is much more complicate because of the lake of closed-form solution for the NSE. Therefore, rather than comparing our closed-form expressions to Monte-Carlo simulations, we propose to compare numerically our closed-form expressions with that of [13]. Therefore, in Fig. 1.5, we plot the numerical evaluation of the proposed closed-form expression of the non-dispersive capacity (*i.e.*, Eq. (1.49)) versus P_x as well as the closed-form expression proposed in [13]. We assume a guard band-free model ($B_g = 0$).

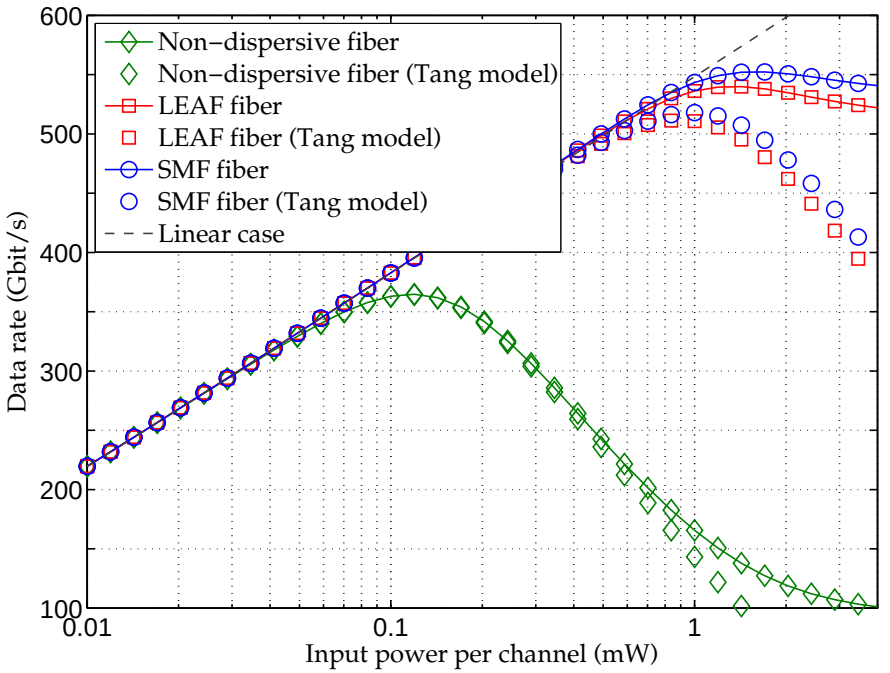

Figure 1.5: Capacity and capacity from [13] versus P_x in dispersive case

We remark that the non-dispersive capacity (evaluated through Eq. (1.74) and not Eq. (1.49)) is the same as that provided in Fig. 1.4. So the approach developed for dispersive case is also valid for the non-dispersive case although more terms have been neglected. More important, we notice that the maximum obtained with our derivations is slightly larger than that obtained in [13]. So the term omitted in [13] has a slight impact on the numerical values of the capacity.

1.6.3 Influence of the guard band

In Fig. 1.6, we plot the spectral efficiency (defined as $\eta = C/(B+B_g)$) versus the guard band in non-dispersive case. Similar figures have been obtained in dispersive case whatever the fiber type. The capacity has been evaluated through its closed-form expression.

We show that the guard band has a bad influence on the spectral efficiency. This result is quite logical since bandwidth is wasted by the guard band. In information-theoretical point-of-view, it is well known that it is more efficient to use the full bandwidth (term out of the $\log_2$) even if it is more disturbed (term within the $\log_2$). In practice, the conclusion can be different: when practical coding scheme is used, the error probability has more influence and the guard band enables us to prevent too high error probability and to satisfy the required Quality of Service (QoS).

In order to better understand the influence of the guard band on non-linearities mitigation, we plot, in Fig. 1.7, the capacity gain (instead of the spectral efficiency) versus the guard band in dispersive case. The capacity (also called "data rate") gain is defined as

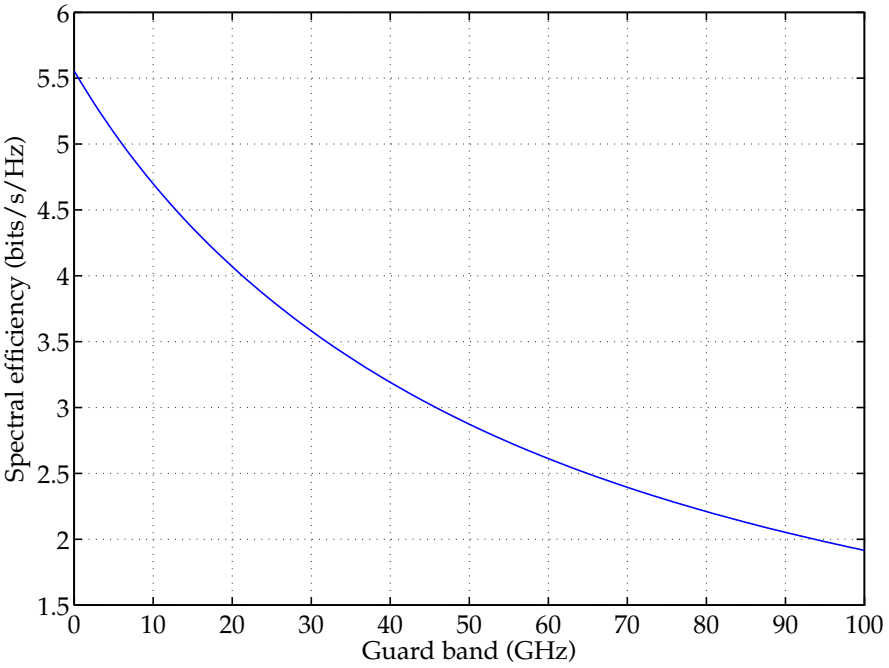

Figure 1.6: Spectral efficiency versus guard band in non-dispersive case

the ratio between the maximum capacity of the system with a given value of the guard band over the maximum capacity of the same system but without guard band. This gain is expressed in percentage.

We notice that the data rate gain is upper-bounded. This result has been predicted in the Section devoted to the non-dispersive case (for $B_g \geq B$, the capacity becomes constant and actually maximum. Here, we remind that $B = 50$ GHz). We so here remark that it remains true for the dispersive case. In addition, we remark that the more dispersive the fiber is, the less the data rate gain is. In other words, the contribution of the CD to mitigate the non-linearities is more significant than the contribution of the guard band.

1.7 Conclusion

In this chapter, after a review of the literature about the derivations of the optical non-linear capacity, we have analyzed the influence of the guard band on the non-linear capacity. To do that, we have developed closed-form expressions of this capacity taking into account to the guard band for non-dispersive case (cf. Eq. (1.49)) as well as dispersive case (cf. Eq. (1.74)). These closed-form expressions have been derived by using the framework developed in [13].

Numerical evaluations of these capacities have been made and have confirmed that CD has a great impact on the optical non-linear capacity. In contrast, the guard band has a slight influence on these capacities which means that the inter-channel interaction is not

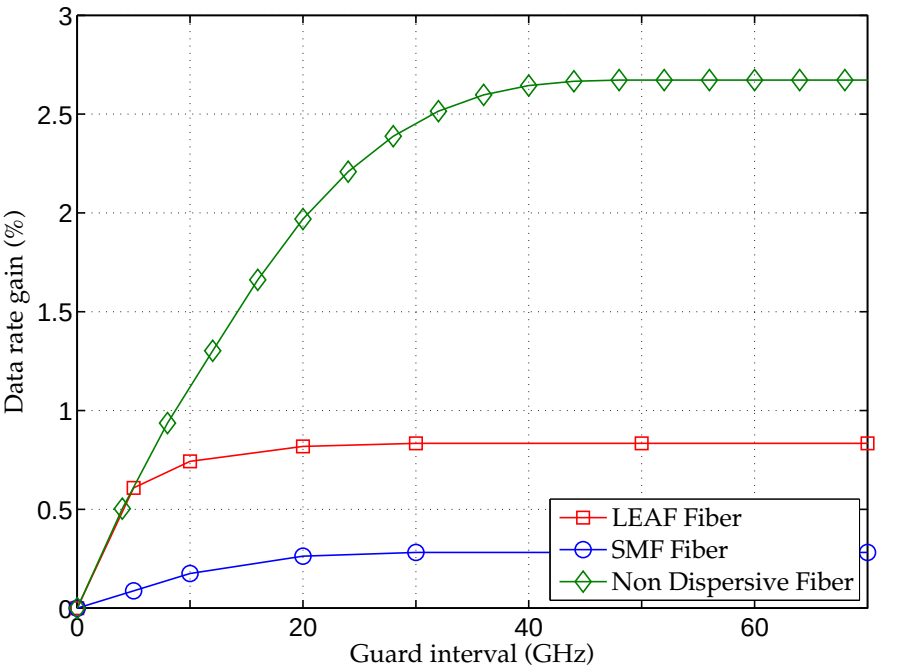

Figure 1.7: Capacity gain versus the guard band in dispersive case

the most important drawback.

Chapter 2

Outage Probability in presence of Polarization-Dependent Loss

2.1 Introduction

As seen in Chapter 1, the increase of data rate in optical transmission systems can be done by increasing the input power until a certain threshold. Beyond this threshold, the nonlinear effects degrade the performance. Therefore, other ways for increasing the data rate have to be explored. A simple idea (already implemented thanks to the coherent detection) is to use more degrees of freedom offred by the fiber, such as, the polarization (which leads to the so-called PDM transmission scheme.). Unfortunately, due to the information spread over both polarizations, some effects, such as PMD and PDL now occur and have to be compensate for.

Assuming OFDM based modulation, as PMD is an unitary perturbation (see Section 2.3.1 for more details), PMD can be completely compensated for and so does not degrade the performance. Therefore, we only focus on the performance degradation due to PDL.

As PDL is a time-varying phenomenon and so can be modeled by a random variable, the Shannon capacity can only be defined for a PDL realization and thus the Shannon capacity for the optical fiber does not exist in strict sense. To analyze such random channels, the most relevant information-theoretic tool is the so-called outage probability. Therefore, this chapter is mainly devoted to the derivations of the outage probability. In addition, we compare the fundamental limit provided by the outage probability to practical communication schemes disturbed by the PDL.

This chapter has organized as follows: in Section 2.2, we introduce and justify carefully the PDL model used in this thesis. In Section 2.3, the derivations for the considered PDL model is drawn and some numerical evaluations of the obtained closed-form expressions are also conducted. In Section 2.4, the outage probability is compared numerically to practical systems using different modulation format, FEC codes (for, instance, Low-Density Parity-Check (LDPC) ones) , PT codes. In addition, some insights related to PDL

mitigation in system design and optimization are provided. Concluding remarks are given in Section 2.5.

2.2 PDL models

As PDL is a linear impairment combining both polarizations, it can be viewed as a 2−by−2 linear Multiple Input Multiple Output (MIMO) system, that is to say that, the bivariate (on both polarizations) received signal, denoted by $\mathbf{Y}(f)$ in the frequency domain, can be written as follows

$$\mathbf{Y}(f) = \mathbf{H}_{\text{PDL}}(f)\mathbf{X}(f) + \mathbf{N}(f) \tag{2.1}$$

where

- $\mathbf{X}(f)$ is the transmitted signal,
- $\mathbf{H}_{\text{PDL}}(f)$ is the 2−by−2 matrix related to PDL.
- and $\mathbf{N}(f)$ is the accumulated noise independent of the transmitted signal.

In this section, our purpose is to introduce models for

1. the structure of the matrix $\mathbf{H}_{\text{PDL}}(f)$,
2. and the structure of the correlation matrix of $\mathbf{N}(f)$.

In this thesis, we present two models: the phenomenological one (see Section 2.2.1), and the simplified one (see Section 2.2.2) which will be used later for doing the mathematical derivations.

2.2.1 Phenomenological model

2.2.1.1 Input/Output matrix model

Assuming that the transmitted signal is narrow band compared to PDL frequency variation, we can remove in $\mathbf{H}_{\text{PDL}}(f)$ the dependency on the frequency. Generally, the link in fiber communications is composed of several spans as in Fig. 2.1 Between two consecutive

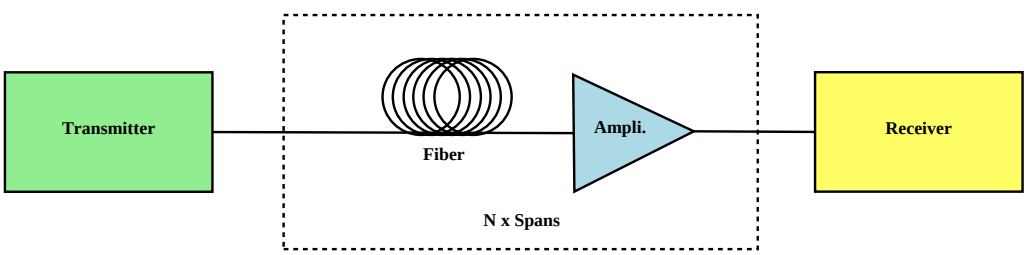

Figure 2.1: System model with amplifiers

spans, an amplifier is inserted. Therefore it is usual to consider that the transfer matrix is

a concatenation of N elementary PDL slices [24, 25, 26, 27, 28]. Each slice may correspond to a span. Thus, we have

$$\tilde{\mathbf{H}}_{\text{PDL}} = \prod_{\ell=0}^{N-1} \underbrace{(\mathbf{R}_{\alpha_\ell} \mathbf{D}_{\gamma_\ell} \mathbf{B}_{\phi_\ell})}_{\mathbf{M}_\ell} \tag{2.2}$$

where R_{α_ℓ} and R_{β_ℓ} are 2−by−2 rotation matrices of random angles α_ℓ and β_ℓ respectively. These rotations represent the polarization axis mismatch between the slice input and the slice output. The matrix $\mathbf{D}_\gamma$, defined as follows,

$$\mathbf{D}_{\gamma_\ell} = \begin{bmatrix} \sqrt{1+\gamma_\ell} & 0 \\ 0 & \sqrt{1-\gamma_\ell} \end{bmatrix} \tag{2.3}$$

represents the power mismatch between both polarizations. Finnally, $\mathbf{B}_{\phi_\ell}$ corresponds to a 2−by−2 birefringence matrix:

$$\mathbf{B}_{\phi_\ell} = \begin{bmatrix} e^{i\phi_\ell} & 0 \\ 0 & e^{-i\phi_\ell} \end{bmatrix}. \tag{2.4}$$

The matrix $\mathbf{M}_\ell$ models the $(\ell+1)$-th slice/span.

2.2.1.2 Noise matrix model

Let us now focus on the noise. We define the matrix $\mathbf{T}_j$ as the transfer matrix between the output of the j-th amplifier and the receiver. For $j \in \{0, N-1\}$, we have

$$\mathbf{T}_j = \prod_{\ell=j}^{N-1} \mathbf{M}_\ell. \tag{2.5}$$

As a consequence, $\mathbf{T}_0$ is the transfer matrix of the overall link and thus is equal to $\tilde{\mathbf{H}}_{\text{PDL}}$. In addition, the matrix $\mathbf{T}_N$ (not defined by Eq. (2.5)) is the 2−by−2 identity matrix. The noise vector for the global link can so be written as follows

$$\mathbf{N}(f) = \sum_{j=1}^{N} \mathbf{T}_j \mathbf{N}_j(f) \tag{2.6}$$

where $\mathbf{N}_j(f)$ is the i.i.d. (with respect to j) white Gaussian noise vector coming from j-th amplifier with variance σ_N^2 per complex dimension. We are interesting by the property of the noise correlation. Therefore, we introduce the correlation matrix (which will be independent of the frequency f).

$$\mathbf{R}_N = \mathbb{E}[\mathbf{N}(f)\mathbf{N}(f)^H].$$

where the superscript $(.)^H$ stands for the transpose-conjugate operator. Then we have

$$\mathbf{R}_N = \sigma_N^2 \sum_{j=1}^{N} \mathbf{T}_j \mathbf{T}_j^H. \tag{2.7}$$

Clearly, this matrix $\mathbf{R}_N$ is not diagonal, and the noise is thus spatially-correlated. Let us denote $P_N = \mathrm{tr}\left(\mathbb{E}[\mathbf{N}(f)\mathbf{N}(f)^H]\right)$ as the noise power. We obtain that

$$P_N = \sigma_N^2 \sum_{j=1}^{N} \mathrm{Tr}\left(\mathbf{T}_j \mathbf{T}_j^H\right) = \sigma_N^2 \sum_{j=1}^{N} \|\mathbf{T}_j\|_F^2$$

where Tr is the trace operator, and $\|\cdot\|_F$ is the Frobenius norm, defined as $\|\mathbf{A}\|_F^2 = \mathrm{Tr}(\mathbf{A}\mathbf{A}^H)$ for any square matrix $\mathbf{A}$.

2.2.1.3 Matrix diagonalization

As any complex-valued matrix, $\tilde{\mathbf{H}}_{\mathrm{PDL}}$ can be decomposed, according to the so-called Singular Value Decomposition (SVD), as follows

$$\tilde{\mathbf{H}}_{\mathrm{PDL}} = \mathbf{U}^H \begin{bmatrix} \sqrt{\lambda_{\max}} & 0 \\ 0 & \sqrt{\lambda_{\min}} \end{bmatrix} \mathbf{V} \tag{2.8}$$

where $\mathbf{U}$ and $\mathbf{V}$ are two unitary matrices, and where $\lambda_{\max}$ and $\lambda_{\min}$ are the maximum and the minimum singular values of the matrix $\tilde{\mathbf{H}}_{\mathrm{PDL}}$. The so-called PDL coefficient is usually expressed in dB and defined as follows

$$\Gamma = 10 \log_{10}\left(\frac{\lambda_{\max}}{\lambda_{\min}}\right). \tag{2.9}$$

Actually, this coefficient represents the power mismatch between both polarizations. Notice that this ratio is also called the condition number of the matrix.

The previous matrix decomposition can be re-written in a slightly different form which will be very useful for our analysis. For going further, let us consider the total received power (*i.e.* for the both polarization) in the frequency domain. This received power is given by

$$P_y = \mathbb{E}[\mathbf{Y}^H(f)\mathbf{Y}(f)] \tag{2.10}$$

where $\mathbf{Y}(f)$ is defined as in Eq. (2.1). In noiseless case and assuming independent symbols in each component of $\mathbf{X}(f)$, we have

$$\begin{aligned} P_y &= \mathrm{Tr}(\tilde{\mathbf{H}}_{\mathrm{PDL}} \mathbb{E}[\mathbf{X}(f)\mathbf{X}(f)^H] \tilde{\mathbf{H}}_{\mathrm{PDL}}^H) \\ &= \mathrm{Tr}(\tilde{\mathbf{H}}_{\mathrm{PDL}} \tilde{\mathbf{H}}_{\mathrm{PDL}}^H)\, P_x \\ &= \|\tilde{\mathbf{H}}_{\mathrm{PDL}}\|_F^2\, P_x \end{aligned} \tag{2.11}$$

with P_x the power/variance on each component of $\mathbf{X}(f)$. We remark that the received power is random since it depends on the realization of the PDL matrix, and is equal to

$$P_y = (\lambda_{\max} + \lambda_{\min}) P_x.$$

By forcing the norm of the matrix $\tilde{\mathbf{H}}_{\text{PDL}}$ to appear in Eq. (2.8), we obtain that

$$\begin{aligned}\tilde{\mathbf{H}}_{\text{PDL}} &= \frac{\sqrt{\lambda_{\max}+\lambda_{\min}}}{\sqrt{2}}\mathbf{U}^H\begin{bmatrix}\sqrt{\frac{2\lambda_{\max}}{\lambda_{\max}+\lambda_{\min}}} & 0\\ 0 & \sqrt{\frac{2\lambda_{\min}}{\lambda_{\max}+\lambda_{\min}}}\end{bmatrix}\mathbf{V}\\ &= \frac{\|\tilde{\mathbf{H}}_{\text{PDL}}\|_F}{\sqrt{2}}\mathbf{U}^H\begin{bmatrix}\sqrt{1+\gamma} & 0\\ 0 & \sqrt{1-\gamma}\end{bmatrix}\mathbf{V}\\ &= \frac{\|\tilde{\mathbf{H}}_{\text{PDL}}\|_F}{\sqrt{2}}\mathbf{U}^H\mathbf{D}_\gamma\mathbf{V} \qquad (2.12)\end{aligned}$$

with

$$\gamma = \frac{\lambda_{\max}-\lambda_{\min}}{\lambda_{\max}+\lambda_{\min}}. \qquad (2.13)$$

The coefficient γ lies in the interval $[0;1]$.

Finally, by considering $\mathbf{Z}(f) = \sqrt{P_N/2}\mathbf{Y}(f)$ a properly normalized output vector, we have

$$\mathbf{Z}(f) = \eta\mathbf{U}^H\mathbf{D}_\gamma\mathbf{V}\mathbf{X}(f) + \tilde{\mathbf{N}}(f) \qquad (2.14)$$

with

$$\eta = \frac{\|\tilde{\mathbf{H}}_{\text{PDL}}\|_F}{\sqrt{P_N}},$$

and $\tilde{\mathbf{N}}(f)$ the normalized but correlated noise.

Consequently, PDL gives rise to three main impairments when multi-span fiber are considered:

1. SNR loss and dispersion: due to the random norm $\|\tilde{\mathbf{H}}_{\text{PDL}}\|_F$ in Eq. (2.12), or alternatively, due to the term η in Eq. (2.14).

2. Polarization mismatch: due to matrix $\mathbf{D}_\gamma$ in Eq. (2.12) or in Eq. (2.14).

3. Noise correlation: due to the form of $\mathbf{R}_N$ in Eq. (2.7).

Notice that in [27], the author encompasses these three phenomena within an unique parameter, called SNR compression, which corresponds to the loss in SNR on the BER performance of an uncoded (no FEC code, no PT code) system.

For this phenomenological model, closed-form expressions for the outage probability can not be found even if a priori probability densities are given for the elementary PDL per splice/span. Consequently, only numerical evaluations of the outage probability (via Monte-Carlo trials) can be performed. In order to exhibit approximate closed-form expressions for the outage probability, simplification on Eqs.(2.7)-(2.12) have to be done. Therefore, we would like hereafter inspect the real influence of these three impairments on the outage probability value.

2.2.1.4 Influence of the noise correlation

In order to only evaluate the influence of the noise correlation on the performance, we have numerically evaluate the outage probability through Monte-Carlo runs assuming Eqs. (2.2)-(2.6) and by normalizing the noise and the PDL matrix. This is then equivalent to Eq. (2.14) by forcing $\eta = 1$. This normalization has been done in order to isolate the impact of the noise correlation.

Except the noise correlation, the simulation setup is the same as in Section 2.4. In Fig. 2.2, we plot the outage probability versus SNR (cf. Appendix Bv for more details about SNR definition) for different number of spans when a mean global PDL ($\mathbb{E}[\Gamma]$) is fixed to 3dB.

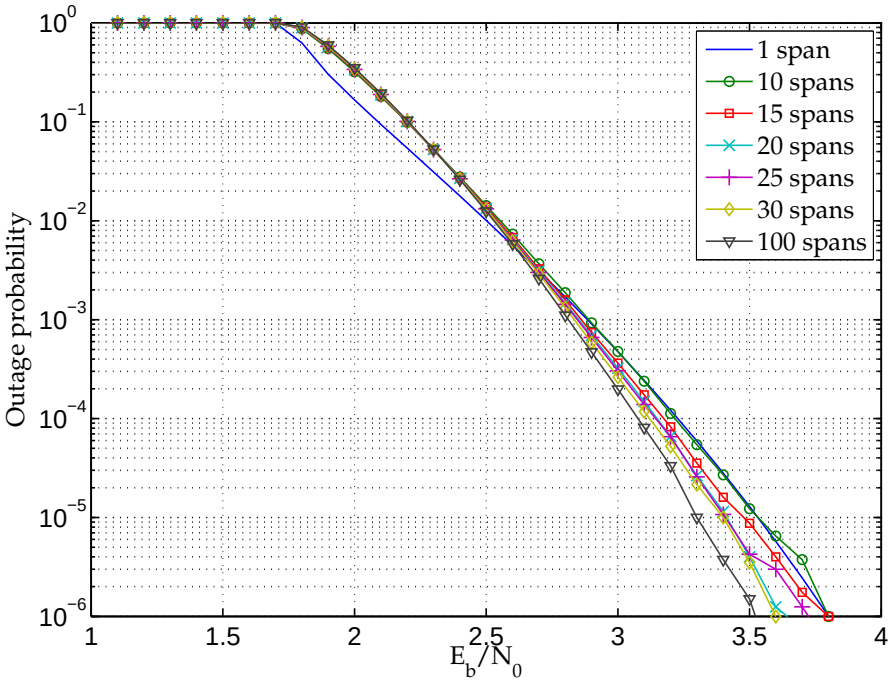

Figure 2.2: Outage probability vs E_b/N_0 for different number of spans. ($\mathbb{E}[\Gamma] = 3$dB)

We observe that the noise correlation has a very slight influence on the outage probability and tends to improve only incrementally the performance. This improvement can be explained in the following way: assuming the noise was entirely correlated between both polarizations, it could be completely removed by transmitting the signal on one polarization and using the other as a noise reference. Therefore the correlation improves the performance.

As a conclusion, without lot of generality, *the noise correlation can be neglected.*

2.2.1.5 Influence of the random SNR

First of all, in PDL-free case, it is easy to check that η is a deterministic value equal to $1/\sqrt{N}$. Assuming here $N = 20$, we thus obtain $\eta \approx 0.22$. Due to PDL, η becomes a random

variable which, for instance, has a mean $m_\eta = \mathbb{E}[\eta]$ and a variance $v_\eta = \mathbb{E}[(\eta - m_\eta)^2]$.

In Fig. 2.3, we plot the histogram (simulated PDF) of η for four values of the mean PDL.

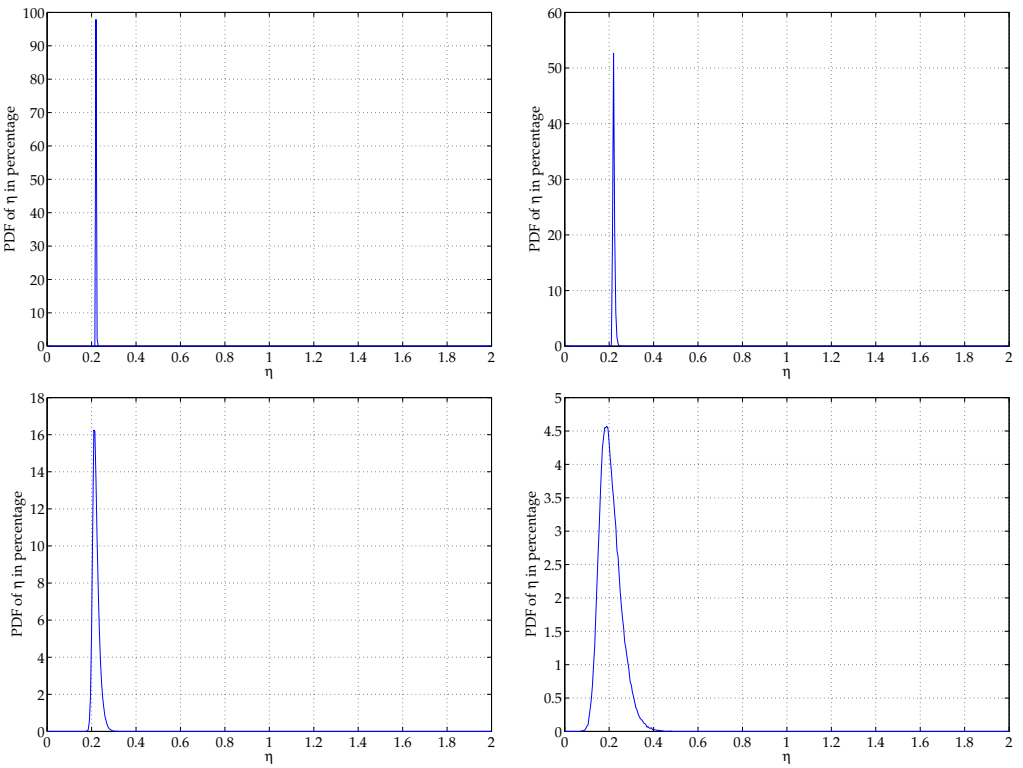

Figure 2.3: Histogram for η with $\mathbb{E}[\Gamma] = 1$dB (NW) or $\mathbb{E}[\Gamma] = 3$dB (NE), or $\mathbb{E}[\Gamma] = 6$dB (SW), or $\mathbb{E}[\Gamma] = 15$dB (SE)

We observe two phenomena :

- the average SNR (so, the mean of the histogram) is less than 0.22 and decreases very slowly with respect to the mean PDL. This assertion is also confirmed in Fig. 2.4 (left). As a consequence, without loss of generality, one can assume that *the average SNR loss can be neglected*.

- the instantaneous SNR (measured at the receiver side) is random which means that even if the average SNR is good, some channel realizations may lead to very poor performance. We remark that the fluctuation of the SNR increases with respect to the mean PDL. This can be also observed in Fig. 2.4 (right) where the variance of the SNR is displayed versus the mean PDL. If the encountered mean PDL is less than 3dB, the SNR dispersion is very small. Moreover, in many practical systems, the maximum value for the mean PDL is around 3dB. As a consequence, with loss of practical generality, *the SNR dispersion can be neglected*.

When the mean PDL is too large, we will see that we can continue to work without

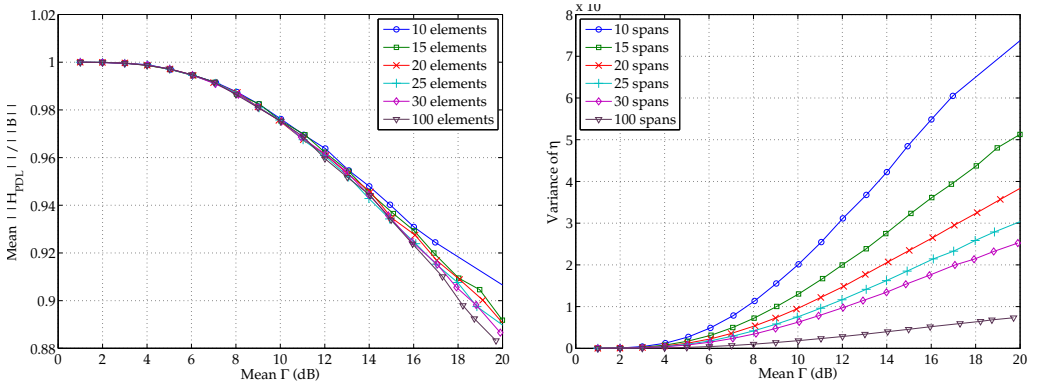

Figure 2.4: $\sqrt{N}m_\eta$ *(left) and* v_η *(right) versus* $\mathbb{E}[\Gamma]$

assuming the SNR loss and the SNR dispersion. The PDL impairment, even if random, varies very slowly compared to the symbol period. In practical system, the information symbol comes from a FEC code (typically a LDPC code), and so belongs to a codeword. Actually, the PDL matrix varies even slowly compared to the codeword [29, 30]. Consequently, it is reasonable to inspect the performance with respect to the SNR viewed by the considered codeword and not by the average SNR calculated over all the transmission duration. Therefore, in the rest of the chapter, *the SNR will be calculated within a codeword duration (and not within an entire transmission)*. This implies that *the SNR dispersion will not be considered hereafter whatever the value of the mean PDL.*

2.2.2 Simplified theoretical model

We remind that the Input/Output relationship (see Eq. (2.14)) can be written as follows

$$\mathbf{Z}(f) = \eta \mathbf{U}^H \mathbf{D}_\gamma \mathbf{V} \mathbf{X}(f) + \tilde{\mathbf{N}}(f).$$

As the noise correlation can be neglected, we force $\tilde{\mathbf{N}}(f)$ to be white spatially. Moreover, as the SNR loss is small enough and the SNR dispersion can be removed (by considering the SNR averaged over the codeword duration), one can force $\eta = 1$.

Consequently, by denoting the received signal (on which we will work) $\mathbf{Y}(f)$, we have

$$\mathbf{Y}(f) = \underbrace{\mathbf{U}^H \mathbf{D}_\gamma \mathbf{V}}_{\mathbf{H}_{\text{PDL}}} \mathbf{X}(f) + \mathbf{N}(f) \tag{2.15}$$

where $\mathbf{N}(f)$ is assumed to be white spatially with variance N_0 per real dimension [31]. This model is very close to that developed for one span [28]. Actually, when only one span is considered, the PDL matrix takes the following form

$$\mathbf{H}_{\text{PDL}} = \mathrm{R}_\alpha \mathbf{D}_\gamma \mathrm{R}_\beta \tag{2.16}$$

where the unitary matrices have been replaced with rotation matrices of angle α and β respectively. Notice that, as rotation matrices are unitary matrices as well, the outage

probability is insensitive to this modification (see derivations in Section 2.3.1 for more details). Nevertheless, in your simulations with real codes, we have used model given by Eq. (2.16).

Let us go back to the phenomenological model given in Eqs. (2.2)-(2.6). First of all, to simplify we wil consider that the noise is spatially white. Under this model, the instantaneous SNR is not constant (since $\|\tilde{\mathbf{H}}_{\text{PDL}}\|_F \neq \sqrt{2}$). In Eq. (2.16), in contrast, $\|\mathbf{H}_{\text{PDL}}\|_F = \sqrt{2}$ whatever the channel realization. To compare easily and fairly both models, we have to normalize the matrix involved in the phenomenological ones as follows:

$$\mathbf{H}_{\text{PDL}} = \tilde{\mathbf{H}}_{\text{PDL}} \cdot \frac{\sqrt{2}}{\|\tilde{\mathbf{H}}_{\text{PDL}}\|_F} \tag{2.17}$$

Finally, our new phenomenological and theoretical model enables us to only inspect the influence of the power mismatch between both polarizations. This restriction is mild since the PT codes only combat this mismatch.

2.3 Derivations of the outage probability

2.3.1 Preliminary derivations

As defined in [31], the outage probability is the probability that the instantaneous channel capacity C (*i.e.* for one channel realization) is below a given transmission rate r (expressed in bits/s)

$$P_o = \text{Prob}\{C < r\}. \tag{2.18}$$

This outage probability is in fact a lower bound on the BER whatever the considered encoder.

Assuming that the Input/output relationship is driven by Eq. (2.15) and that the input $\mathbf{X}(f)$ is a white Gaussian random variable, the instantaneous channel capacity (*i.e.*, given a channel realization) takes the following shape [32]

$$C(\mathbf{H}_{\text{PDL}}) = B \log_2 \left(\det \left[\mathbf{I}_2 + \rho \mathbf{H}_{\text{PDL}} \mathbf{H}_{\text{PDL}}^H \right] \right) \tag{2.19}$$

where $\mathbf{I}_2$ is the 2-by-2 identity matrix, B is the transmitted signal bandwidth, and $\rho = E_s/2N_0$ (cf. Appendix B for more details about SNR). Notice that only PDL has been considered here, but the result will be the same by inserting the the CD and the PMD. Recalling that CD and PMD are modeled by unitary matrices, and that the determinants of such matrices are equal to 1, they do not modify the value of the capacity. In addition, this is the reason why the unitary/rotation matrices exhibited in Eqs. (2.15)-(2.16) do not modify the value of the capacity as well.

Using Eq. (2.16) for $\mathbf{H}_{\text{PDL}}$, we easily obtain that

$$C(\mathbf{H}_{\text{PDL}}) = B \log_2((1+\rho)^2 - \rho^2\gamma^2) \tag{2.20}$$

and so the outage probability given in Eq. (2.18) becomes

$$P_o = \text{Prob}\left\{\log_2\left((1+\rho)^2 - \rho^2\gamma^2\right) < R\right\} \tag{2.21}$$

where $R = r/B$ is the spectral efficiency.

Notice that the outage probability only depends on the PDF of γ. The previous expression could be rewritten as follows:

$$P_o = \text{Prob}\{\gamma^2 > f(R,\rho)\} \quad \text{with} \quad f(R,\rho) = 1 - \frac{1}{\rho^2}(2^R - 1 - 2\rho). \tag{2.22}$$

Because of the properties of γ, one can exhibit two thresholds on the SNR denoted by ρ_l (below which $P_o = 1$) and ρ_u (beyond which $P_o = 0$).

- If $f(R,\rho) < 0$, then $P_o = 1$ since γ^2 is always positive. This constraint on $f(R,\rho)$ holds iff $\rho < \rho_l$ with
$$\rho_l = \sqrt{2^R} - 1. \tag{2.23}$$
- If $f(R,\rho) > 1$, then $P_o = 0$ since γ^2 is always less than 1. This constraint on $f(R,\rho)$ holds iff $\rho > \rho_u$ with
$$\rho_u = \frac{1}{2}(2^R - 1). \tag{2.24}$$

In the following subsections, we focus on the derivations in closed-form for the outage probability. Different statistical models for the PDL coefficient coming from an exhaustive survey of the literature have been considered. These models rely on either γ or Γ given by

$$\Gamma = 10\log_{10}\left(\frac{1+\gamma}{1-\gamma}\right) \tag{2.25}$$

thanks to Eqs. (2.9)-(2.13).

2.3.2 γ constant

This very simple model assumes γ as a constant. It has been introduced in order to deeply analyze the influence of PDL on advanced coding techniques [28, 33]. This constant value of γ is denoted by μ_c. If $\mu_c > f(R,\rho)$, then $P_o = 1$ and if $\mu_c < f(R,\rho)$, then $P_o = 0$. Therefore we are interesting to find the relationship between μ_c, R and ρ such that $\mu_c = f(R,\rho)$, *i.e.*,

$$\mu_c = 1 - \frac{1}{\rho^2}(2^R - 1 - 2\rho). \tag{2.26}$$

We thus have to solve the following second-order polynomial with respect to ρ

$$\rho^2\left(1 - \mu_c^2\right) + 2\rho - 2^R + 1 = 0. \tag{2.27}$$

As $|\mu_c| \leq 1$ and $R > 0$, the discriminant of the previous second-order polynomial is strictly positive and so this polynomial has two real-valued roots. One of them is negative and so

has no physical meaning. The second one is given by $(\sqrt{1+(1-\mu_c^2)(2^R-1)}-1)/(1-\mu_c^2)$. Therefore, we have

$$P_o = \begin{cases} 1 & \text{if } \rho < \frac{\sqrt{1+(1-\mu_c^2)(2^R-1)}-1}{(1-\mu_c^2)} \\ 0 & \text{otherwise} \end{cases} . \tag{2.28}$$

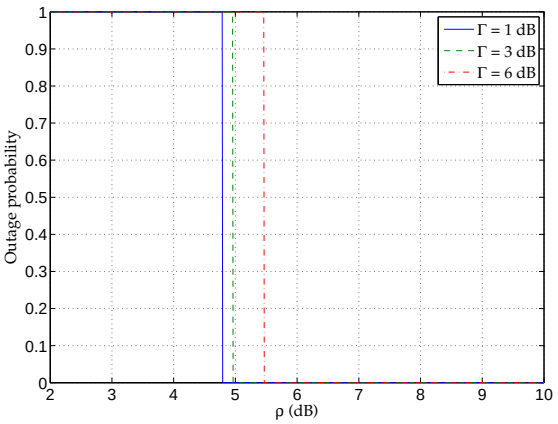

Figure 2.5: Outage probability for model γ constant and R = 4bits/s/Hz

In Fig. 2.5, we plot the outage probability for a given spectral efficiency R = 4bits/s/Hz and different values of Γ (or equivalently, of μ_c).

2.3.3 γ Gaussian

The model γ Gaussian has been roughly validated in [34, 24]. In [34, 24], the authors asserted that it is valid for low values of the PDL. It has been justified by the concatenation of some elementary PDL slices. Its probability density function is

$$p_\gamma^{(g)}(x) = \frac{1}{\sqrt{2\pi\sigma_g^2}} e^{-\frac{(x-\mu_g)^2}{2\sigma_g^2}} \tag{2.29}$$

with $\mu_g = \mathbb{E}[\gamma]$ and σ_g^2 the variance. In Fig. 2.6, such a PDF is displayed with $\mu_g = 0.3263$ and $\sigma_g^2 = 0.1240$ for ensuring $\mathbb{E}[\Gamma]$ = 3dB.

Considering $\rho > \rho_l$, the outage probability, defined in Eq. (2.22), can be written as follows

$$P_o = \int_{-\infty}^{-\sqrt{f(R,\rho)}} p_\gamma^{(g)}(x)dx + \int_{\sqrt{f(R,\rho)}}^{+\infty} p_\gamma^{(g)}(x)dx \tag{2.30}$$

This probability distribution has a known Cumulated Density Function (CDF) based on the Gaussian tail function

$$x \mapsto Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2} dt.$$

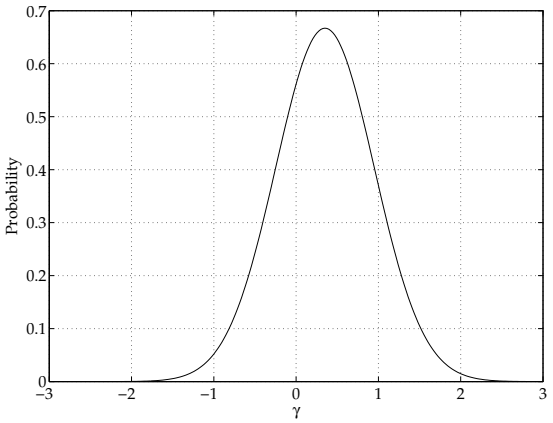

Figure 2.6: Example of γ Gaussian PDF

Consequently, Eq. (2.30) implies that

$$P_o = 1 - Q\left(\frac{-\sqrt{f(R,\rho)} - \mu_g}{\sigma_g}\right) + Q\left(\frac{\sqrt{f(R,\rho)} - \mu_g}{\sigma_g}\right). \tag{2.31}$$

Finally, the outage probability is:

$$P_o = \begin{cases} 1 & \text{if } \rho < \rho_\ell \\ Q\left(\frac{\sqrt{f(R,\rho)}+\mu_g}{\sigma_g}\right) + Q\left(\frac{\sqrt{f(R,\rho)}-\mu_g}{\sigma_g}\right) & \text{otherwise} \end{cases}. \tag{2.32}$$

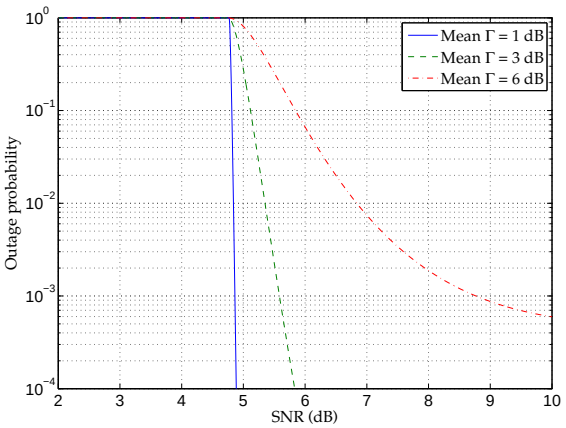

Figure 2.7: Outage probability for model γ Gaussian and R = 4bits/s/Hz

Some numerical examples of this outage probability have been plotted in Fig. 2.6 for a fixed spectral efficiency and different values of the mean PDL. In our simulations, the parameter μ_g has been fixed as $\mathbb{E}[\Gamma] = 10\log_{10}((1+\mu_g)/(1-\mu_g))$, then we have built a lookup table to find the relationship between σ_g^2 and $\mathbb{E}[\Gamma]$.

Unfortunately this model has many drawback: firstly, it is quite realistic only in few situations (low PDL); then, the evaluation of the distribution's parameters for a given

mean PDL (given by $\mathbb{E}[\Gamma]$) is intractable; finally, as Fig. 2.6 shows for high values of the mean PDL, the γ coefficient is not constrained to belong to the interval $]-1,1[$ which has no physical meaning. Therefore the outage is not zero even when $\rho > \rho_u$.

2.3.4 γ truncated Gaussian

We introduce this model in order to remove the main drawback of the γ Gaussian model: we constraint γ to belong to $[-1,1]$. Therefore the Gaussian distribution is truncated as follows

$$p_\gamma^{(tg)}(x) = \begin{cases} \frac{K}{\sqrt{2\pi\sigma_{tg}^2}} e^{-\frac{(x-\mu_{tg})^2}{2\sigma_{tg}^2}} & , \ |x| < 1 \\ 0 & \text{otherwise} \end{cases} \tag{2.33}$$

where

$$K = \frac{1}{1 - Q\left(\frac{1+\mu_{tg}}{\sigma_{tg}}\right) - Q\left(\frac{1-\mu_{tg}}{\sigma_{tg}}\right)} \tag{2.34}$$

is a normalization coefficient ensuring that $p_\gamma^{(tg)}$ is a PDF. An example of the PDF is plotted in Fig. 2.8 with $\mu_{tg} = 0.3263$ and $\sigma_{tg}^2 = 0.1240$.

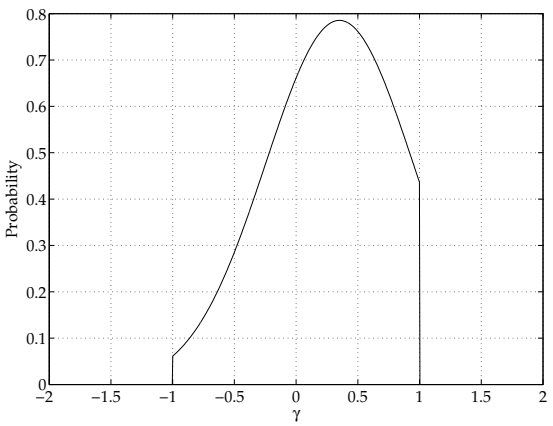

Figure 2.8: Example of γ truncated Gaussian PDF

According to Eq. (2.22), we just have to calculate the outage probability for $\rho \in [\rho_\ell, \rho_u]$ since now γ lies in the interval $]-1;1[$. Then the outage probability can be decomposed as

$$P_o = \int_{-1}^{-\sqrt{f(R,\rho)}} p_\gamma^{(tg)}(x)dx + \int_{\sqrt{f(R,\rho)}}^{1} p_\gamma^{(tg)}(x)dx. \tag{2.35}$$

Like the previous model, the CDF of the normal distribution is known, and we obtain

that

$$P_o = \begin{cases} 1 & \text{if } \rho < \rho_\ell \\ K\left[Q\left(\frac{\sqrt{f(R,\rho)}+\mu_{tg}}{\sigma_{tg}}\right) + Q\left(\frac{\sqrt{f(R,\rho)}-\mu_{tg}}{\sigma_{tg}}\right) - Q\left(\frac{1+\mu_{tg}}{\sigma_{tg}}\right) - Q\left(\frac{1-\mu_{tg}}{\sigma_{tg}}\right)\right] & \text{if } \rho_\ell < \rho < \rho_u \\ 0 & \text{otherwise} \end{cases} \tag{2.36}$$

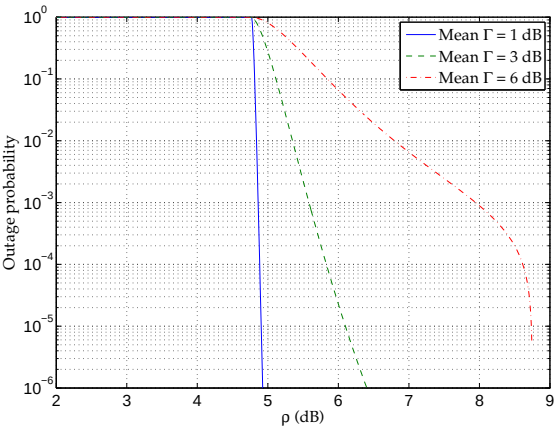

Figure 2.9: Outage probability for model γ Gaussian and R = 4bits/s/Hz

In Fig. 2.9, we plot numerical illustrations of the closed-form expressions for a fixed spectral efficiency and different values of the mean PDL. In our simulations, we have used the same relationship between μ_{tg} and $\mathbb{E}[\Gamma]$ that between μ_g and $\mathbb{E}[\Gamma]$. Similarly, the relationship between σ^2_{tg} and $\mathbb{E}[\Gamma]$ is obtained thanks to a lookup table. Although the model is now constrained, some drawbacks still occur, especially, the lack of relationship between μ_{tg}, σ_{tg} and the mean PDL.

2.3.5 Γ (positive) Gaussian

After some practical measurements, [35] has observed that the histogram of Γ fits quite well with the positive part of a Gaussian distribution. Therefore, the authors have suggested the following distribution.

$$p_\Gamma^{(pg)}(x) = \begin{cases} \frac{2}{\sqrt{2\pi\sigma^2_{pg}}} e^{-\frac{x^2}{2\sigma^2_{pg}}} & , x \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{2.37}$$

In order to obtain a relationship between the mean PDL and the distribution's parameters, we show, after some algebraic manipulations, that σ^2_{pg} has to satisfy $\sigma_{pg} = (\sqrt{2/\pi})\mathbb{E}[\Gamma]$. In Fig. 2.10, an example of the PDF is plotted for $\mathbb{E}[\Gamma]$ = 3dB.

Since the distribution is given on Γ, the inequality on γ of the Eq. (2.22) has to be replaced with an inequality on Γ. As the relationship defined in Eq. (2.9) between γ and Γ ensures that for any value of Γ the corresponding value of γ is within the interval $]-1, 1[$,

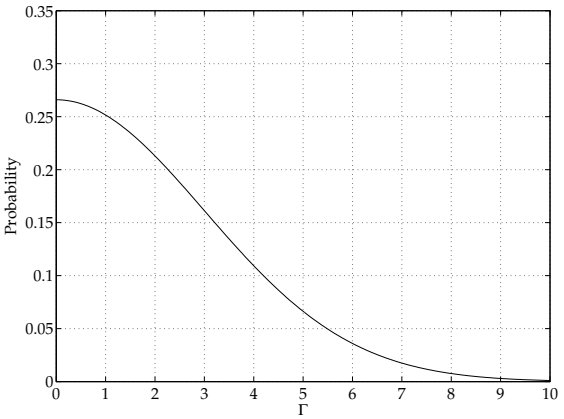

Figure 2.10: Example of Γ *positive Gaussian PDF*

the bounds given in Eqs. (2.23) and (2.24) are still valid. Therefore, we only focus on the case $\rho \in [\rho_l, \rho_u]$. Moreover, as the considered model assumes that Γ is positive, the corresponding γ is then also positive which implies that Eq. (2.22) can be rewritten as follows

$$P_o = \text{Prob}\left\{\gamma > \sqrt{f(R,\rho)}\right\}. \tag{2.38}$$

Remarking that Eq. (2.25) is equivalent to $\Gamma = \frac{20}{\log(10)}\text{artanh}(\gamma)$, and that the function $x \mapsto \text{artanh}(x)$ is strictly increasing, we obtain

$$P_o = \text{Prob}\left\{\Gamma > T(R,\rho)\right\} = \int_{T(R,\rho)}^{\infty} p_{\Gamma}^{(pg)}(x)\mathrm{d}x \tag{2.39}$$

with

$$T(R,\rho) = \frac{20}{\log(10)}\text{artanh}\left(\sqrt{f(R,\rho)}\right). \tag{2.40}$$

As the CDF of the Gaussian distribution is well known, we obtain the following closed-form expression of the outage probability:

$$P_o = \begin{cases} 1 & \text{if } \rho < \rho_\ell \\ 2Q\left(\frac{T(R,\rho)}{\sigma_{pg}}\right) & \text{if } \rho_\ell < \rho < \rho_u \\ 0 & \text{otherwise} \end{cases} \tag{2.41}$$

In Fig. 2.11, we show numerical illustrations of this closed-form expression for a given spectral efficiency and different values of the mean PDL.

Notice that in [36], the authors proposed to consider that Γ is modeled by a zero-mean Gaussian-distributed random variable. However, since Γ is defined as the ratio of the maximum transmission coefficient over the minimum (Eq. (2.9)), Γ should be positive. Therefore, such a distribution choice for Γ does not seem relevant and so has not been considered in this Chapter.

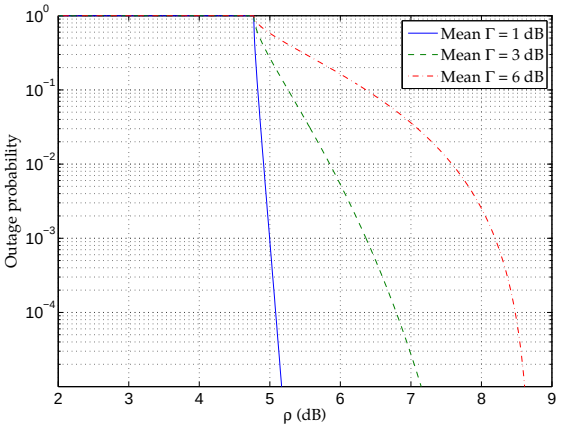

Figure 2.11: Outage probability for model Γ (positive) Gaussian and R = 4bits/s/Hz

2.3.6 Γ Rayleigh

This statistical model of Γ has been presented in [25] and has been obtained by considering the concatenation of elementary PDL slices without taking the birefringence into account. Then the PDF of Γ is given by the Rayleigh distribution as follows

$$p_{\Gamma}^{(r)}(x) = \begin{cases} \frac{x}{\sigma_r^2} e^{-\frac{x^2}{2\sigma_r^2}} & , \; x \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{2.42}$$

where the parameter $\sigma_r = \sqrt{2/\pi} \cdot \mathbb{E}[\Gamma]$. In Fig. 2.12 an example of PDF is plotted for $\mathbb{E}[\Gamma] = 3\text{dB}$.

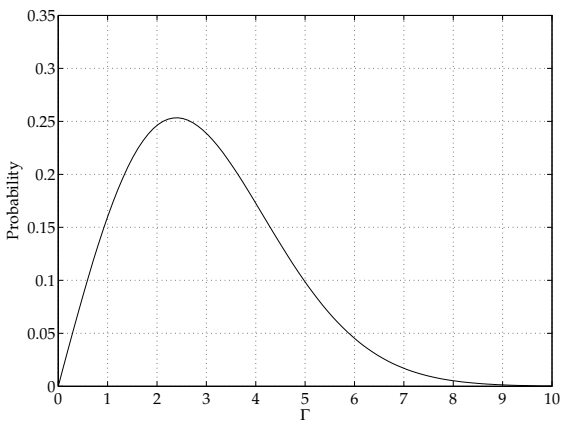

Figure 2.12: Example of Γ Rayleigh PDF

Similarly to the Γ positive Gaussian case, the definition of the PDF induces γ positive. So Eq. (2.39) can be used in order to evaluate the outage probability.

$$P_o = \int_{T(R,\rho)}^{\infty} p_{\Gamma}^{(r)}(x)\mathrm{d}x. \tag{2.43}$$

Once again using the CDF of the Rayleigh distribution, the outage probability can be written as follows

$$P_o = \begin{cases} 1 & \text{if } \rho < \rho_\ell \\ e^{-\frac{T(R,\rho)}{2\sigma_r^2}} & \text{if } \rho_l < \rho < \rho_u \\ 0 & \text{otherwise} \end{cases} \tag{2.44}$$

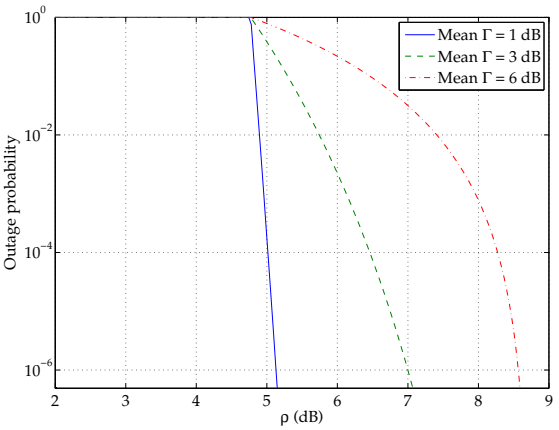

Figure 2.13: Outage probability for model Γ *Rayleigh and* $R = 4$bits/s/Hz

Numerical illustrations of the obtained closed-form expression for a given spectral efficiency and different values of the mean PDL are plotted in Fig. 2.13.

2.3.7 Γ Maxwellian

Firstly, it has been shown in [25] that the Γ Maxwellian distribution offers the best fitting with the simulated histogram of the PDL when the matrix $\mathbf{H}_{\text{PDL}}$ is the concatenation of elementary PDL slices taking the birefringence into account. Moreover this model has been well-supported by mathematical calculations presented in [26]. We thus here assume that Γ is Maxwellian distributed as follows

$$p_\Gamma^{(m)}(x) = \begin{cases} \sqrt{\frac{2}{\pi}} \frac{x^2}{\sigma_m^3} e^{-\frac{x^2}{2\sigma_m^2}} & , \ x \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{2.45}$$

where $\sigma_m = \sqrt{\pi/8} \cdot \mathbb{E}[\Gamma]$. In Fig. 2.14, an example of the PDF is plotted for $\mathbb{E}[\Gamma] = 3$dB.

Derivations to obtain the outage probability in closed-form is very similar to the two previous cases. The random variable Γ defined here is positive and so induces positive values of γ. Therefore, Eq. (2.39) can be used.

$$P_o = \int_{T(R,\rho)}^{\infty} p_\Gamma^{(m)}(x) \mathrm{d}x. \tag{2.46}$$

Using the CDF of the Maxwellian distribution, the closed-form expression of the outage

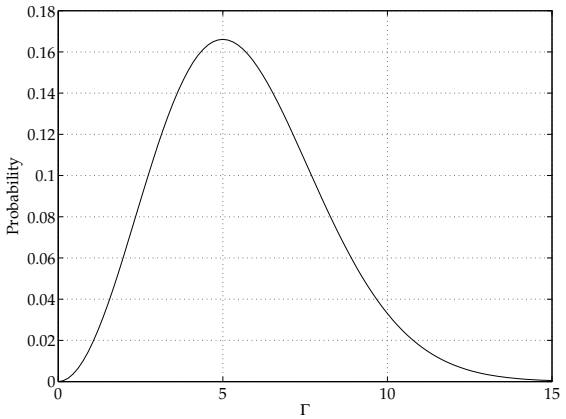

Figure 2.14: Example of Γ *Maxwellian PDF*

probability is

$$P_o = \begin{cases} 1 & \text{if } \rho < \rho_\ell \\ 2Q\left(\frac{T(R,\rho)}{\sigma_m}\right) + \sqrt{\frac{2}{\pi\sigma_m^2}}T(R,\rho)e^{\frac{-T(R,\rho)^2}{2\sigma_m^2}} & \text{if } \rho_\ell < \rho < \rho_u \\ 0 & \text{otherwise} \end{cases} \tag{2.47}$$

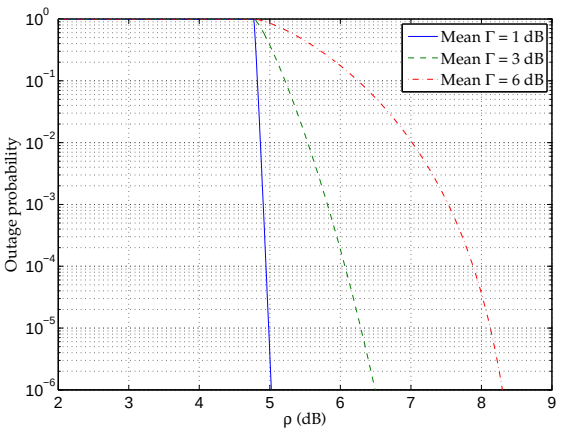

Figure 2.15: Outage probability for model Γ *Maxwellian and* R = 4bits/s/Hz

In Fig. 2.15, numerical illustrations of the obtained closed-form expression for a given spectral efficiency and different values of the mean PDL are plotted.

2.3.8 Numerical evaluations

In this section, we evaluate the accuracy of all previously-mentioned statistical models of PDL through the closed-form expressions of the outage probability with respect to the (normalized) phenomenological model introduced in Section 2.2.2. Then the most relevant statistical model (in fact the Maxwellian one) is more deeply analyzed.

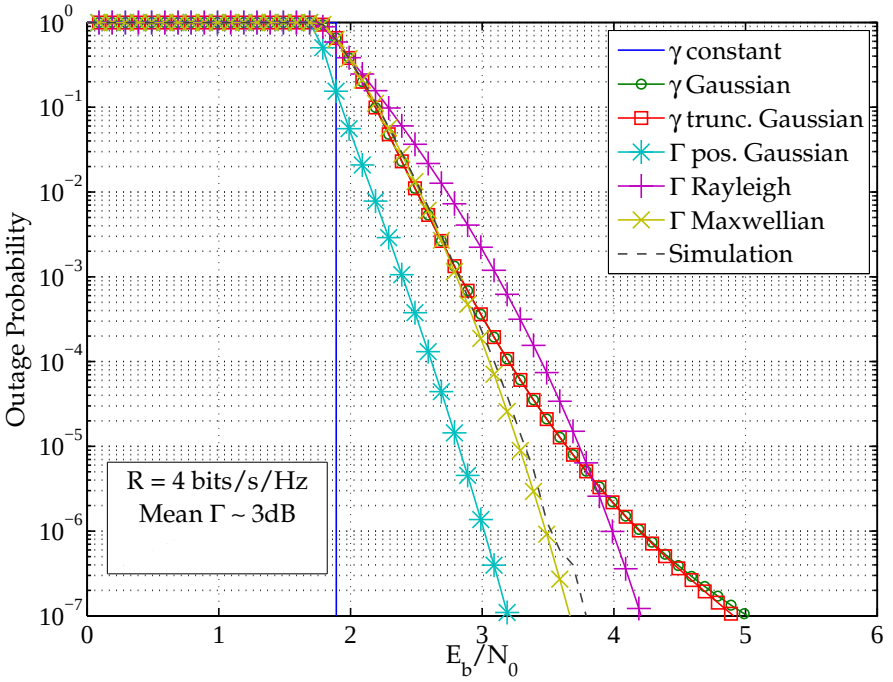

Figure 2.16: P_o versus SNR (E_b/N_0 in dB) for various PDL models, including simulation of the phenomenological model with $N = 100$, mean $\Gamma = 3$ dB, and $R = 4$ bits/s/Hz

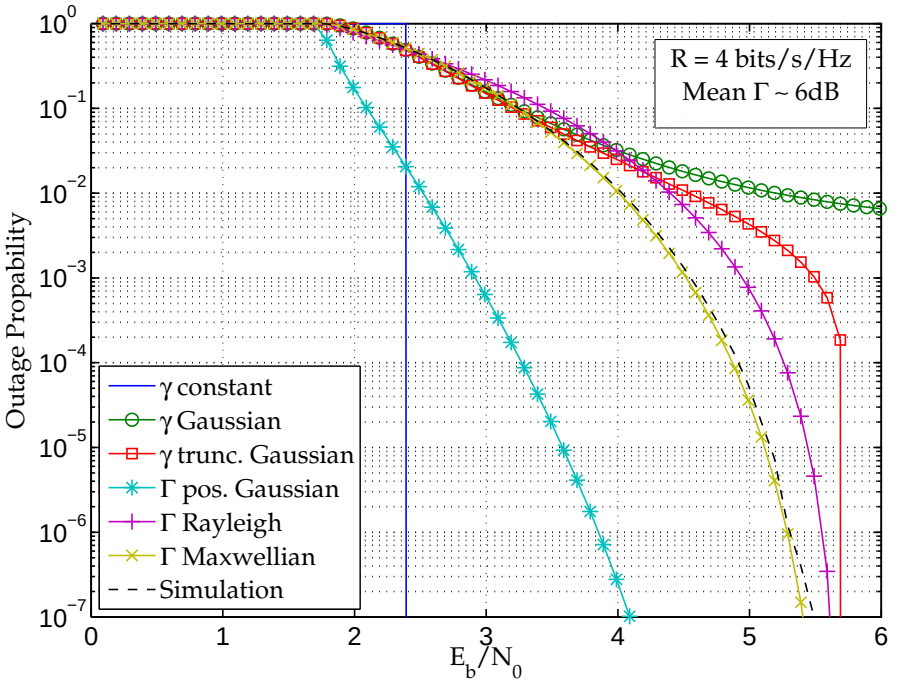

Figure 2.17: P_o versus SNR (E_b/N_0 in dB) for various PDL models, including simulation of the phenomenological model with $N = 100$, mean $\Gamma = 6$ dB, and $R = 4$ bits/s/Hz

In Figs. 2.16 and 2.17, we plot the numerical evaluations of all the obtained closed-form expressions of the outage probability as well as the outage probability of the normalized phenomenological one versus the SNR for a mean Γ equal to 3 dB and 6 dB respectively. Firstly, we remark that all curves, except the γ Gaussian and the γ constant models, vanish when the SNR is higher than the upper-bound ρ_u defined in Eq. (2.24). Moreover the γ constant model is quite far away from other models and thus seems clearly inadequate. Finally, as already asserted in [26] for the distribution, the Γ Maxwellian model is very close to the phenomenological model for the outage probability. It is also the most relevant one and will be next analyzed in depth in the rest of this Section.

Before going further, we define the so-called SNR penalty (on the outage probability) as the ratio between the value of the SNR needed to yield a given outage probability and the SNR needed for the same outage probability in a PDL-free transmission. In Figs. 2.18 and 2.19, we plot the SNR penalty for the Γ Maxwellian model for different outage probability targets and for two values of the spectral efficiency respectively

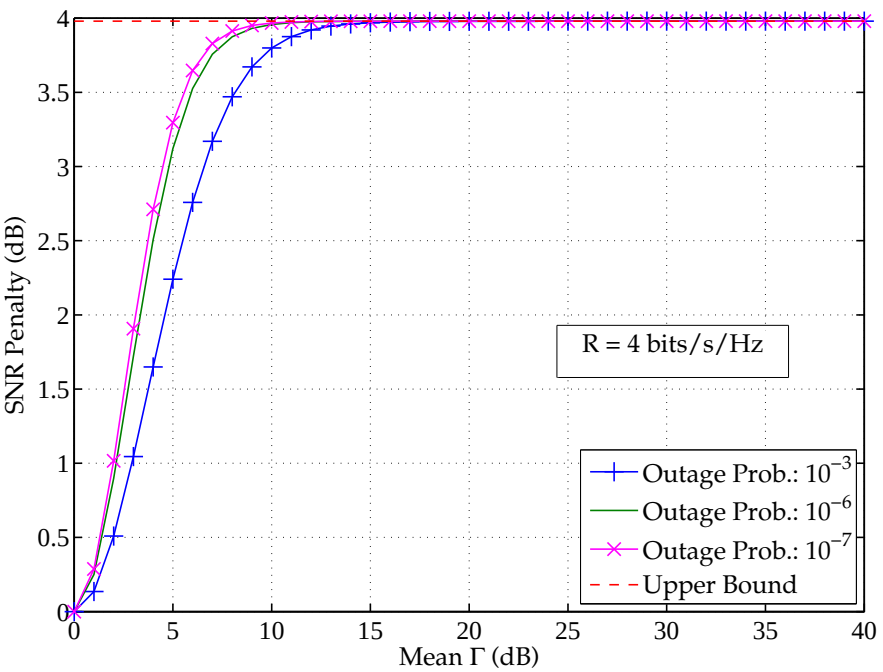

Figure 2.18: SNR Penalty (in dB*) for Maxwellian model and* $R = 4$ bits/s/Hz

We remark that the SNR penalty is upper-bounded whatever the mean PDL and the outage probability target. This bound depends only on the spectral efficiency R and has the following closed-form expression

$$\text{Penalty}_{max} = \frac{\rho_u}{\rho_l} = \frac{2^R - 1}{2(\sqrt{2^R} - 1)}. \tag{2.48}$$

This bound represents the maximal loss due to the PDL assuming only the power mismatch drawback due to the PDL. For reasonable value of the mean PDL (for instance,

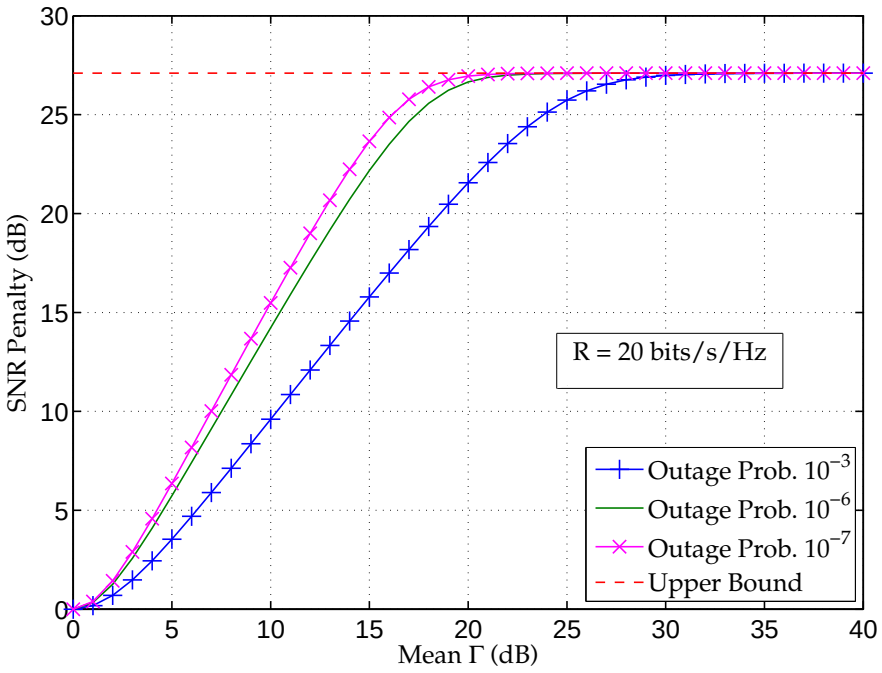

Figure 2.19: SNR Penalty (in dB*) for Maxwellian model and* $R = 20$ bits/s/Hz

3 dB), the maximal penalty is not reached and actually the SNR penalty is roughly between 1 dB and 3 dB.

In order to inspect the global effect of the PDL, we simulate in Fig. 2.20 the SNR penalty when the phenomenological model is neither normalized nor noise whitening. We remark that this new phenomenological model provides better SNR penalty at small mean PDL (this is due to the noise correlation which improves the performance and strongly worse SNR penalty at medium and high mean PDL (this is due to the SNR dispersion).

2.4 Performance of simulated systems

We would like now to compare the outage probability (which is a lower-bound of the BER of any transmission scheme) to the performance of practical systems when disturbing only by PDL.

Concerning the PDL model, we have hereafter only consider the Maxwellian theoretical model as well as the (normalized-white) phenomenological model as described in Sections 2.2.2 and 2.3.7. Consequently, only the influence of the power mismatch will be analyzed by our simulations.

This Section is organized as follows: in Section 2.4.1, we introduce in details the considered practical systems. Then, Section 2.4.3 is devoted to the performance when QPSK modulation is employed. Finally, in Section 2.4.3, we consider higher spectral

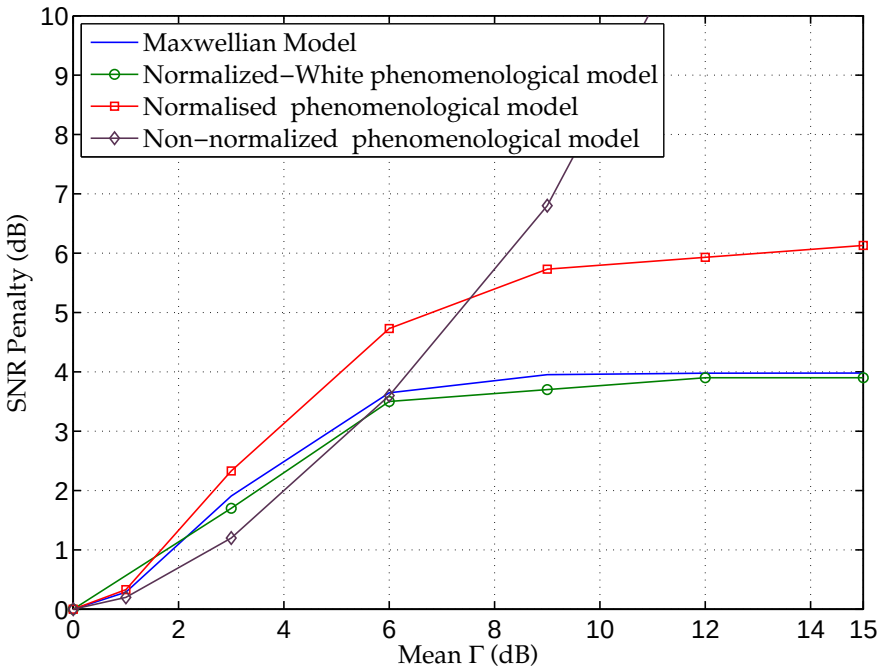

Figure 2.20: SNR Penalty (in dB*) for Maxwellian model, Phenomenological model and Phenomenological model (without normalization and noise whitening) and* $R = 4$ bits/s/Hz

efficient modulation, *i.e.*, 16-QAM.

2.4.1 Practical transmission schemes

As we would like to study only the impact of PDL, we consider that the other linear impairments (attenuation, CD, PMD) do not occur. Actually, this assumption is not restrictive at all. Indeed, assuming a OFDM transmission scheme, it is well known that the CD and the PMD can be fully compensated for since they correspond to unitary operators.

Therefore, we focus on transmission schemes based on OFDM transceiver. Then each subcarrier f_k of the considered OFDM can be modeled as follows

$$\mathbf{Y}(f_k) = \mathbf{H}_{\text{PDL}}\mathbf{X}(f_k) + \mathbf{N}(f_k) \tag{2.49}$$

where the different terms have been completely described in Section 2.2.2. The number of spans is fixed to $N = 100$.

In addition, we assume the channel transfer matrix $\mathbf{H}_{\text{PDL}}$ is independent of the channel numbering and has been perfectly estimated at the receiver side.

The practical transmission schemes simulated hereafter are summarized in Fig. 2.21. Each block in Fig. 2.21 is explained just below.

CW: Update Fig. 2.21's soft decoder.

Let us start with the transmitter side: the information bit stream is encoded using a FEC code. The objective of FEC code is to combat against the bit error generated by the channel

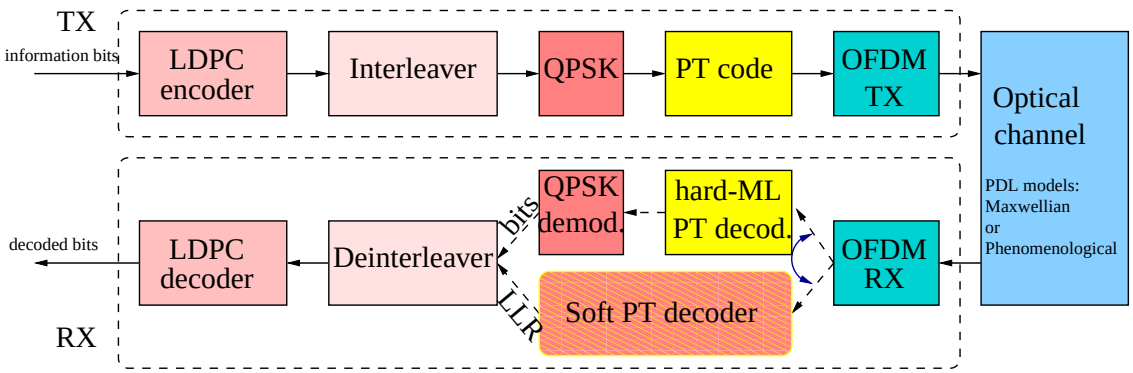

Figure 2.21: Architecture of the simulated system

by adding redundancy to the transmitted information bits. The most popular codes are the convolutive codes (used, for instance, in 2G), the Reed-Solomons (RS) codes (used, for instance, in ADSL), the Turbo codes, the LDPC codes. The most powerful codes are the Turbo codes and the LDPC codes [37, 38] which means that these codes offer performance very close to the Shannon capacity in the context of AWGN channel. As our purpose is to know if we are able to be close to the fundamental limit (given by the outage probability) or not, we propose to focus on the LDPC code. More details about the design of these codes are given in Section 2.4.1.1.

The output of the LDPC encoder associated with a packet of information bits following with its redundancy is called a codeword. The obtained codeword then passes through an interleaver in order to increase the robustness to burst errors by scrambling the bits over time and polarization. As the codeword length is huge, we can choose a simple row-column interleaver [39].

The interleaved codeword is modulated using QPSK or 16-QAM constellation. These constellations are displayed in Fig. 2.22. We remind that QPSK has a spectral efficiency of 2 bits/s/Hz and 16-QAM of 4 bits/s/Hz.

The symbols (belonging to either QPSK or 16-QAM) are then precoded thanks to a PT code. This coding technique, which significantly improves the performance of PDL-Disturbed systems [33, 40, 30, 41], is described in details in Section 2.4.1.2. Then each PT symbol is assigned to one OFDM subcarrier randomly. As PDL channel is not frequency-selective, the performance is independent of the random subcarrier assignment.

Let us move now on the receiver side: after the OFDM demodulation, two decoding schemes have been implemented:

- A "Hard" decoding approach: a hard Maximum-Likelihood (ML) decoder for the PT code is implemented, which transmits the detected symbol to the constellation demodulator. Then the detected bits is passed through the de-interleaver and a hard LDPC decoder.

- A "Soft" decoding approach: the PT decoder does not make decision on the transmit-

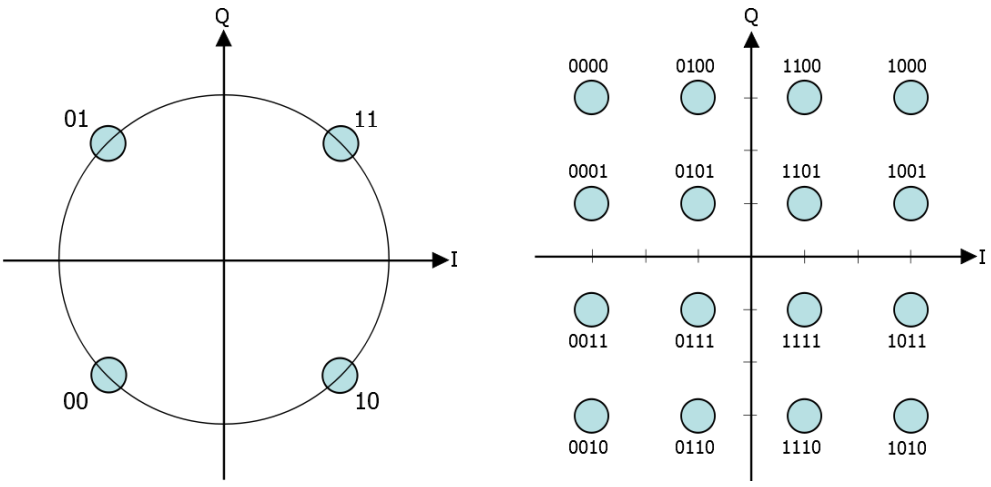

Figure 2.22: QPSK (left) and 16-QAM (right)

ted bits but calculated the so-called Log-Likelihood Ratio (LLR) defined as follows

$$\mathrm{LLR}(x_i) = \log\left(\frac{\mathrm{Prob}\{x_i = 0|y\}}{\mathrm{Prob}\{x_i = 1|y\}}\right) \tag{2.50}$$

where y_i is the received signal and x_i the i-th transmitted bit. A LLR plays the role of a detected bit and represents not only the bit value (information given by the sign of the LLR), and also a confidence index on this bit value. This extra information improves significantly the performance compared to the "Hard" approach. Nevertheless, the decoder is much more complex, which may be a crucial issue for very high data rate transmission.

Finally, we assume that the PDL remains constant within an entire LDPC codeword.

In the simulations, we have tested the following combinations of PT and LDPC codes in order to analyze the respective influence of each code and their concatenation.

- No code (neither FEC nor PT),
- PT code only,
- LDPC code only (with either hard or soft decoding),
- PT+LDPC codes with hard decoding,
- PT+LDPC codes with soft decoding.

2.4.1.1 LDPC codes

Introduced in 1962 by [42], the low-density parity-check codes (LDPC) have been forgotten for a long time, because of their implementation complexity. They have been rediscovered in 1992 by [37] and thus are well-renowned for their capacity-achieving behavior.

The LDPC codes belong to the family of linear block codes [38]. They are characterized by a sparse parity-check matrix which means that this matrix contains very few 1. The rate of the code, denoted by r_{LDPC}, is defined as the ratio between the number of information bits in a codeword and the codeword length.

In the simulations, we have chosen an LDPC code based on the quasi-cyclic progressive edge growth algorithm (QC-PEG) [36]. This kind of code is particularly well adapted to the coherent optical communications as mentioned in [36]. As the code performance depends on its length [38], we choose a long enough code by considering a codeword length of 4320 and a coding rate $r_{\text{LDPC}} = 3/4$. This rate is quite low for optical transmission systems (compared to that used currently) but enables us to improve significantly the performance.

For the "Soft" approach, the decoder is based on the Sum-Product Algorithm [38]. For the "Hard" approach, the used decoder is that proposed in [42].

2.4.1.2 PT codes

Polarization-Time (PT) codes are the adaptation to the optical channel of the space-time coding techniques used in wireless communications. This coding techniques have been developed in order to take benefit of all degrees of freedom of MIMO systems. The main principle is to combine several symbols; let's say $s_1, s_2, \cdots, s_{n_s}$, in an hyper symbol, called PT symbol, for spreading information over time and space. Thus a PT code may modify the global data rate according to this rate

$$r_{\text{PT}} = \frac{n_s}{T} \tag{2.51}$$

where n_s is the number of symbols in the PT symbol and where T is the number of channel use (symbol time) needed to transmit a PT symbol.

As in optical transmission systems, the MIMO channel is 2 by 2 (due to both polarizations), it is usual to spread the PT symbol over two symbol periods. In such a case, the PT code will be called full rate if its rate is $r_{\text{PT}} = 2$, in other words, 1 symbol / polarization / channel use. So the PT code could not add redundancy.

In wireless MIMO channel, some code design criteria [43, 31] exist in order to find the best coding scheme. In addition, many space-time codes have been proposed for this channel, such as, the Alamouti code [44], the Golden code [45] or the Silver code [46]. The Golden code is optimal according to the criterion developed in [31]. Notice that equivalent criterion does not exist in optical transmissions yet. Therefore we have to re-use codes initially developed for wireless communications. In the simulations, we consider the Golden code and the Silver code which are both full rate.

Golden Code: An hyper symbol $\mathbf{X}_{\mathcal{G}}$ associated with the Golden code has the following shape

$$\mathbf{X}_{\mathcal{G}} = \frac{1}{\sqrt{5}} \begin{bmatrix} \alpha\left(s_1 + \theta s_2\right) & \alpha\left(s_3 + \theta s_4\right) \\ i\overline{\alpha}\left(s_3 + \overline{\theta} s_4\right) & \overline{\alpha}\left(s_1 + \overline{\theta} s_2\right) \end{bmatrix} \tag{2.52}$$

where $\theta = \frac{1+\sqrt{5}}{2}$, $\overline{\theta} = \frac{1-\sqrt{5}}{2}$, $\alpha = 1 + i - i\theta$, $\overline{\alpha} = 1 + i - i\overline{\theta}$ and s_1, s_2, s_3, s_4 are four symbols. The rows of the matrix represent the polarizations and the columns the channel use.

Silver Code: this code has slightly smaller performance than the Golden code in the context of wireless MIMO channel. An hyper symbol $\mathbf{X}_S$ associated with the Silver code has the following shape

$$\mathbf{X}_S = \begin{bmatrix} s_1 & -s_2^* \\ s_2 & s_1^* \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} z_1 & -z_2^* \\ z_2 & z_1^* \end{bmatrix} \tag{2.53}$$

where z_1 and z_2 are obtained by:

$$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \frac{1}{\sqrt{7}} \begin{bmatrix} 1+i & -1+2i \\ 1+2i & 1-i \end{bmatrix} \begin{bmatrix} s_3 \\ s_4 \end{bmatrix} \tag{2.54}$$

and s_1, s_2, s_3, s_4 are four modulated symbols. As previously, the rows of the matrix represent the polarizations and the columns the channel use.

The main advantage of the Silver code compared to the Golden code is its decoding complexity. More surprisingly, as shown in [33, 40, 30, 41], performance of the Silver code in the context of optical transmission systems is slightly better than those of the Golden code.

2.4.2 Performance using QPSK

First of all, let us check that QPSK enables us to reach the standard target data rate for next generation optical communications systems based on coherent detection. This target data rate is 100 Gbit/s over a standard 50 GHz ITU channel. Because of filtering, the effective bandwidth reduces to 40 GHz. Combining QPSK and PDM leads i) to i) a spectral efficiency $R = 4$ bits/s/Hz and so a useful data rate of 160 Gbit/s in FEC-free case (with or without PT code), and ii) to a spectral efficiency $R = 3$ bits/s/Hz and so a useful data rate of 120 Gbit/s in FEC case (since the considered LDPC code has a rate $r_{\text{LDPC}} = 3/4$.

2.4.2.1 BER performance

In Fig. 2.23, outage probabilities (for the two spectral efficiency $R = 3$ and 4) and BER for the above-mentioned coding schemes have been plotted versus SNR in the context of Maxwellian model. In Fig. 2.24, same curves are considered but in the context of the phenomenological model. For both figures, we choose a mean PDL equal to 3 dB.

First of all, we notice that the concatenation of PT code and LDPC code is quite close to the fundamental limit, namely, around 1.5 dB. Consequently, even though optimizing PT codes and LDPC codes for PDL mitigation could be of some interest, the existing ones already yield remarkable performance. We also observe that performance of the "Soft" approach significantly improves the performance compared to the "Hard" one. So

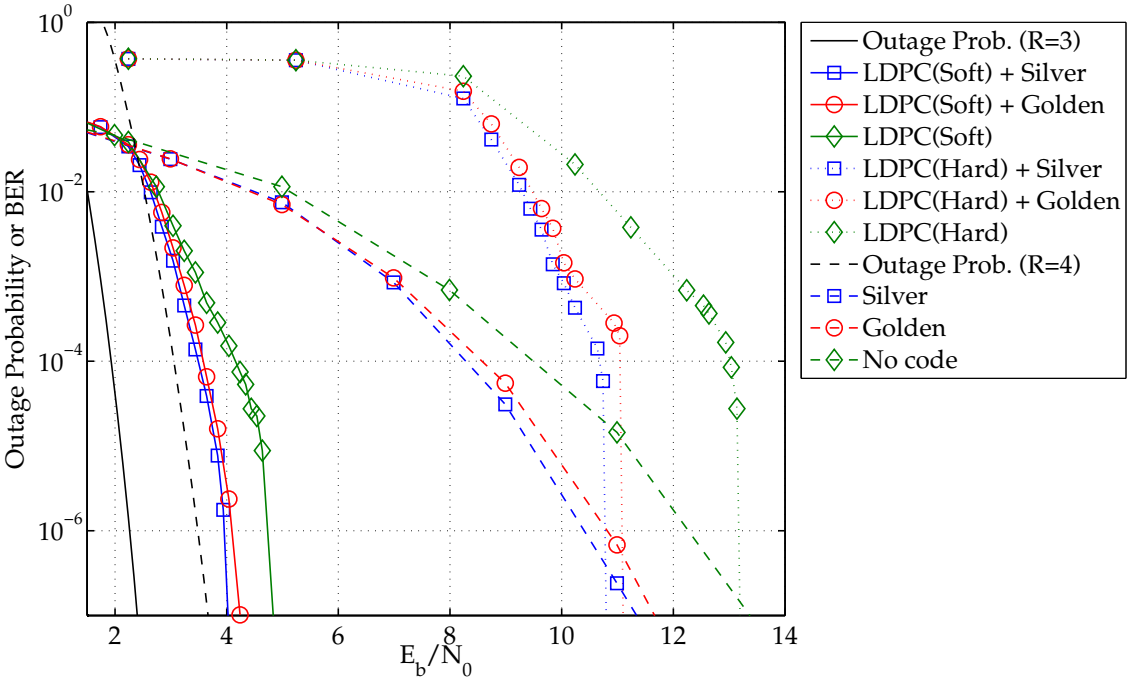

Figure 2.23: Outage probability and BER for Maxwellian model ($\mathbb{E}[\Gamma] = 3$ dB)

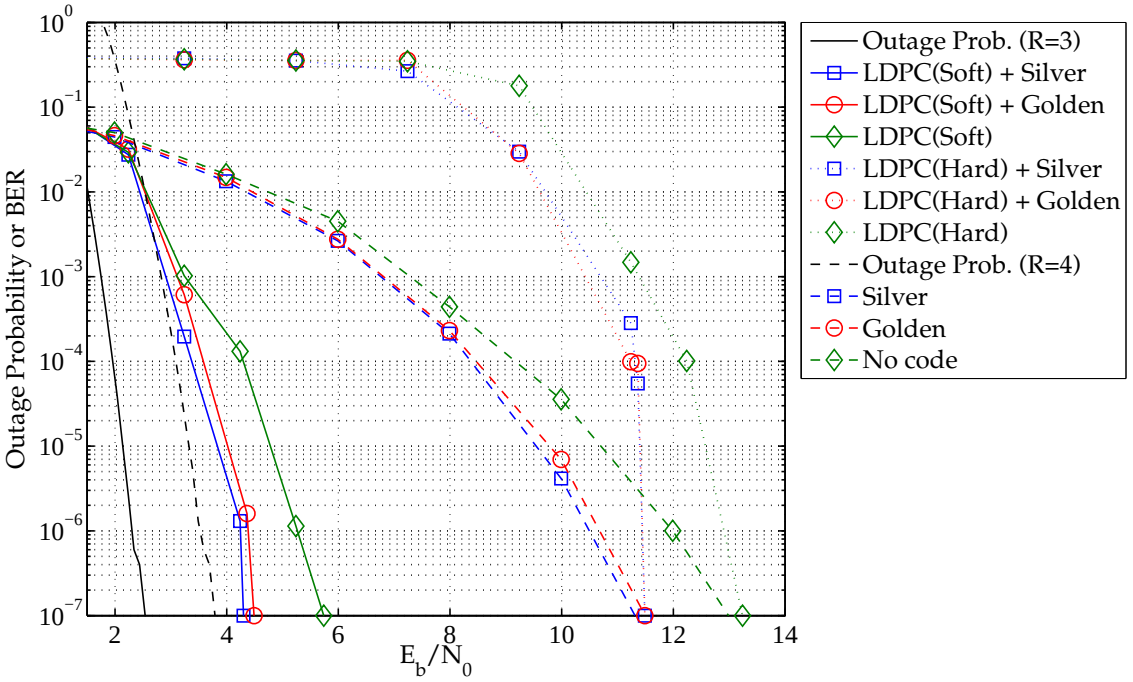

Figure 2.24: Outage probability and BER for phenomenological model ($\mathbb{E}[\Gamma] = 3$ dB)

the implementation of the "Soft" approach will be a important milestone for the PDL mitigation.

In addition we remark that, unlike the wireless MIMO channel, improvements offered by the PT code and the LDPC code are cumulative. Thus FEC coding and PT coding techniques are complementary in optical systems for combating the PDL efficiently.

Finally, as in [33, 40, 30, 41], we observe that the Silver code is slightly better than the Golden code.

2.4.2.2 SNR gap

In order to better understand the loss in performance between the fundamental limit and the practical schemes, we introduce the SNR gap which is defined as ratio between the SNR needed for a given scheme to yield a given BER and the SNR needed to ensure the outage probability equal to this given BER. Moreover, unlike Section 2.4.2.1, we would like to analyze also the influence of the mean PDL value. Therefore, in Figs. 2.25 and 2.26, we plot the SNR gap versus the mean PDL for the Maxwellian model and the phenomenological model respectively. According to Section 2.4.2.1, the "Hard" approach is definitively too far away from the fundamental limit and this is omitted here.

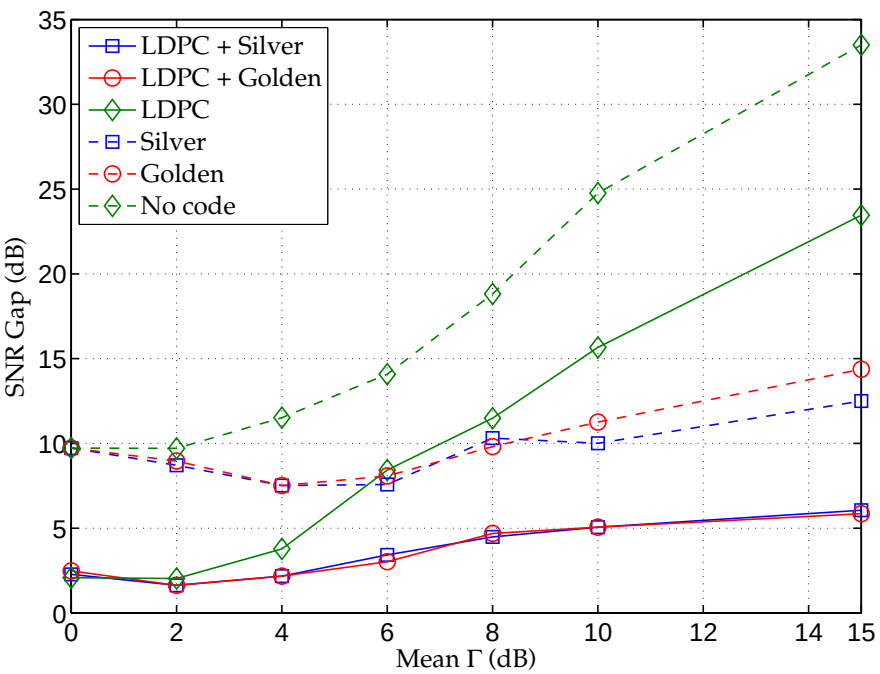

Figure 2.25: SNR gap for Maxwellian model. ($P_o = 10^{-7}$ and BER $= 10^{-7}$)

We remark that, for low mean PDL values, the LDPC alone succeeds to combat the PDL (actually the polarization power mismatch) efficiently. However, for high mean PDL values, its performance is poor. In contrast, the gain due to the PT codes is quasi-null for low mean PDL (≤ 2 dB) but is significant for high mean PDL. Moreover the best way to

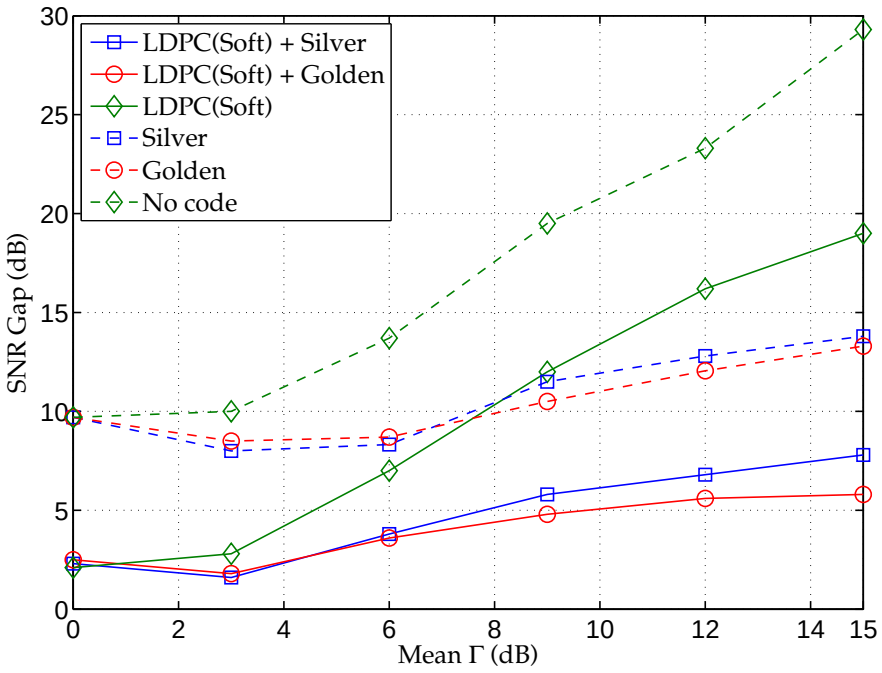

Figure 2.26: SNR gap for the phenomenological model. ($P_o = 10^{-7}$ and BER $= 10^{-7}$)

combat PDL is to concatenate these two coding techniques since the SNR gap (only due to the polarization power mismatch, since the other effects have not been considered) remains quite stable whatever the mean PDL value. Indeed, the gains provided by both coding techniques are cumulative.

2.4.2.3 SNR penalty

The SNR penalty is the metric used for comparing a system with PDL with the PDL-free case. So the SNR penalty provides insights different from those given by the SNR gap. For practical schemes, the SNR penalty is the ratio between the SNR needed to achieve a given BER and the SNR needed to achieve the same BER in a PDL-free transmission. For the fundamental limit, the SNR penalty is defined in Section 2.3.8. In Figs. 2.27 and 2.28, we plot the SNR penalty versus the mean PDL for the Maxwellian model and the phenomenological one respectively.

For a standard mean PDL of 3 dB, the SNR penalty is around a few dB if the PT codes are used. Moreover, we notice that the SNR penalty for practical transmission schemes based on the PT codes is quite the same as that for the outage probability. In contrast, the schemes only using the LDPC codes offer a SNR penalty increasing rapidly with respect to the mean PDL. Therefore, we show again that the PT codes are a crucial tool for the PDL mitigation. Once again, the concatenation of the PT codes with the LDPC ones provide a very low SNR penalty, especially for standard values of the mean PDL.

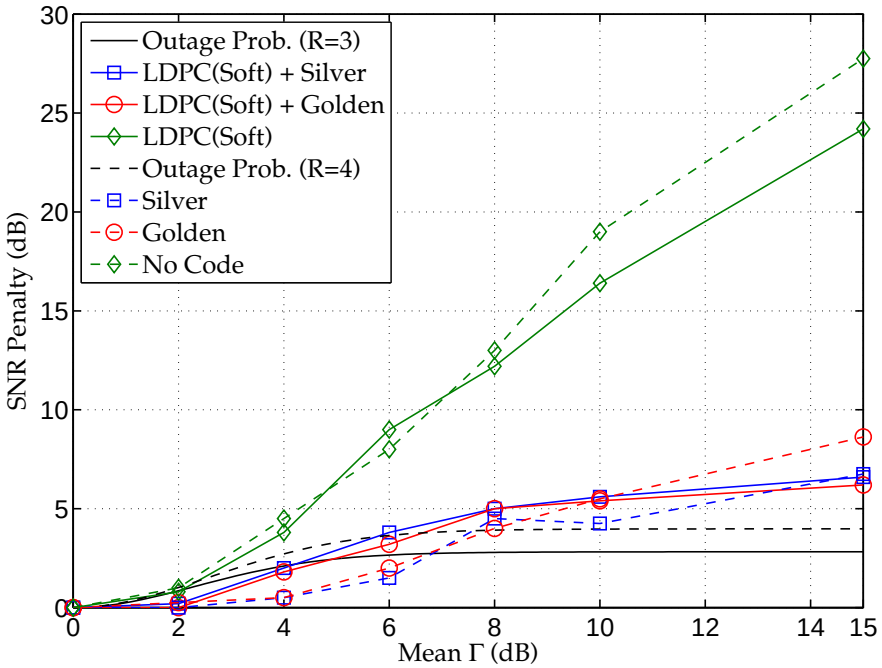

Figure 2.27: SNR penalty for Maxwellian model. ($P_o = 10^{-7}$ and BER = 10^{-7})

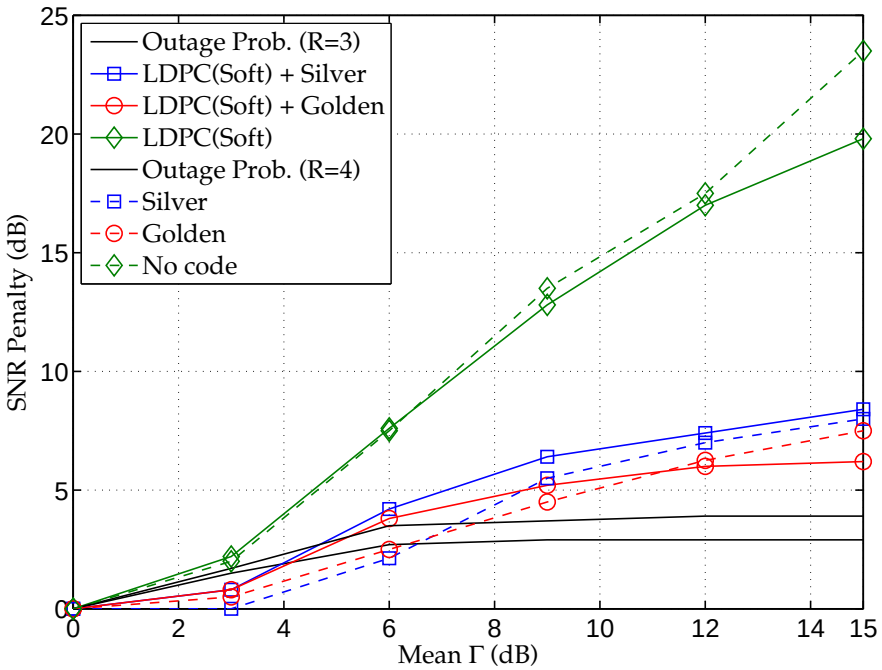

Figure 2.28: SNR penalty for the phenomenological model. ($P_o = 10^{-7}$ and BER = 10^{-7})

2.4.3 Performance using 16-QAM

In this Section, we plot figures equivalent to Figs. 2.23, 2.24, 2.25, 2.26, 2.27, and 2.28 but when 16-QAM is used. Then, using PDM, the spectral efficiency is $R = 8$ bits/s/Hz without FEC and $R = 6$ bits/s/Hz with FEC. Thus, the data rate are 320 Gbit/s without FEC and 240 Gbit/s with FEC over the standard 50 GHz ITU channel (we remind that 10 GHz are used for the guard interval). As the "Hard" approach offers too poor performance, it is omitted in next figures.

The remarks are almost the same as for QPSK. One can just see that the gains are less cumulative than for QPSK. The LDPC codes for usual mean PDL valued, offers very good performance. The improvement given by the PT codes really occurs at high mean PDL value. In addition, the SNR gap and the SNR penalty is slightly worse. This is not a surprise since it is much harder to counter-act the channel impairment when constellation with high spectral efficiency is employed.

2.5 Conclusion

We have studied the fundamental performance limit for coherent optical communications in the presence of PDL, which is one of the major impairments in such systems. This limit, represented by the outage probability, was derived in closed-form expressions for some PDL models; we then used BER simulations to evaluate how far from said limit simulated systems could be.

Comparing the outage probability expressions under the usual PDL models with the numerical evaluation of that associated with a phenomenological model, we have observed that the Maxwellian model fits quite well with the phenomenological one in this respect. Notice that our models only take into account the polarization power mismatch. In this section, we have shown why the other source of performance degradation due to the PDL can be omitted.

The outage probability has also been compared numerically to the BER offered by simulated systems using several coding schemes: LDPC and/or PT Codes, or no coding. These simulations have shown that PT codes are essential to PDL mitigation, and that the concatenation of a PT code with an LDPC is even better (especially the QPSK is used). Indeed, we have found that for a mean PDL of 3 dB, a simulated system based on PT and LDPC codes (with soft decoding) performs with an SNR gap of only 1.5 dB to the fundamental limit with QPSK and only 2 dB to the fundamental limit with 16-QAM.

While more powerful coding schemes could undoubtedly become closer to the limit, the modest 1.5 dB gain that can be hoped for must be balanced against the computational complexity, which is already the major drawback of LDPC soft decoding, especially at the very high bit rates of modern optical communication systems.

Thus, a more interesting way for further study would be whether improved PT codes

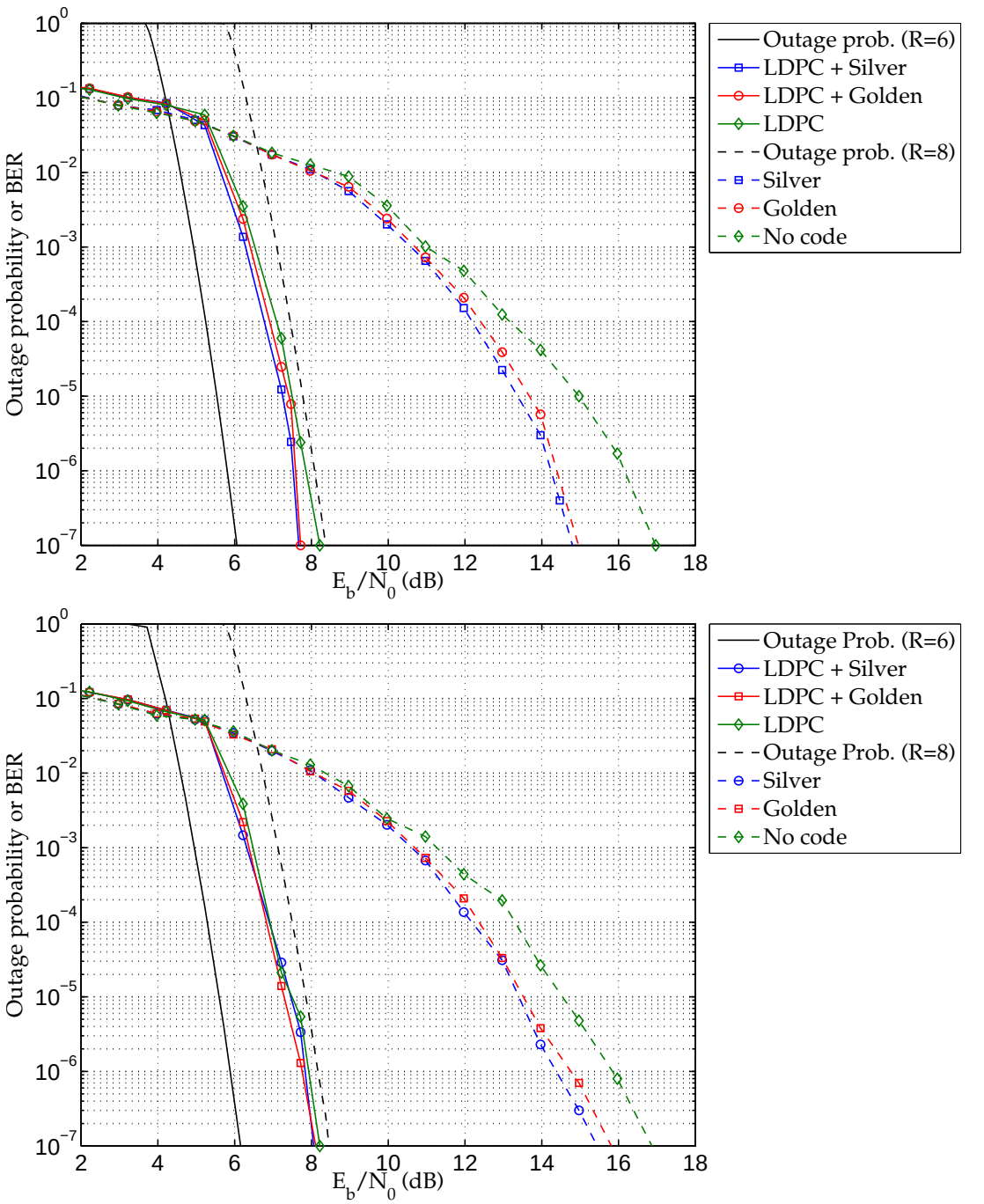

Figure 2.29: Outage probability and BER for Maxwellian model (top) and phenomenological model (bottom) (16-QAM, $\mathbb{E}[\Gamma] = 3$ dB)

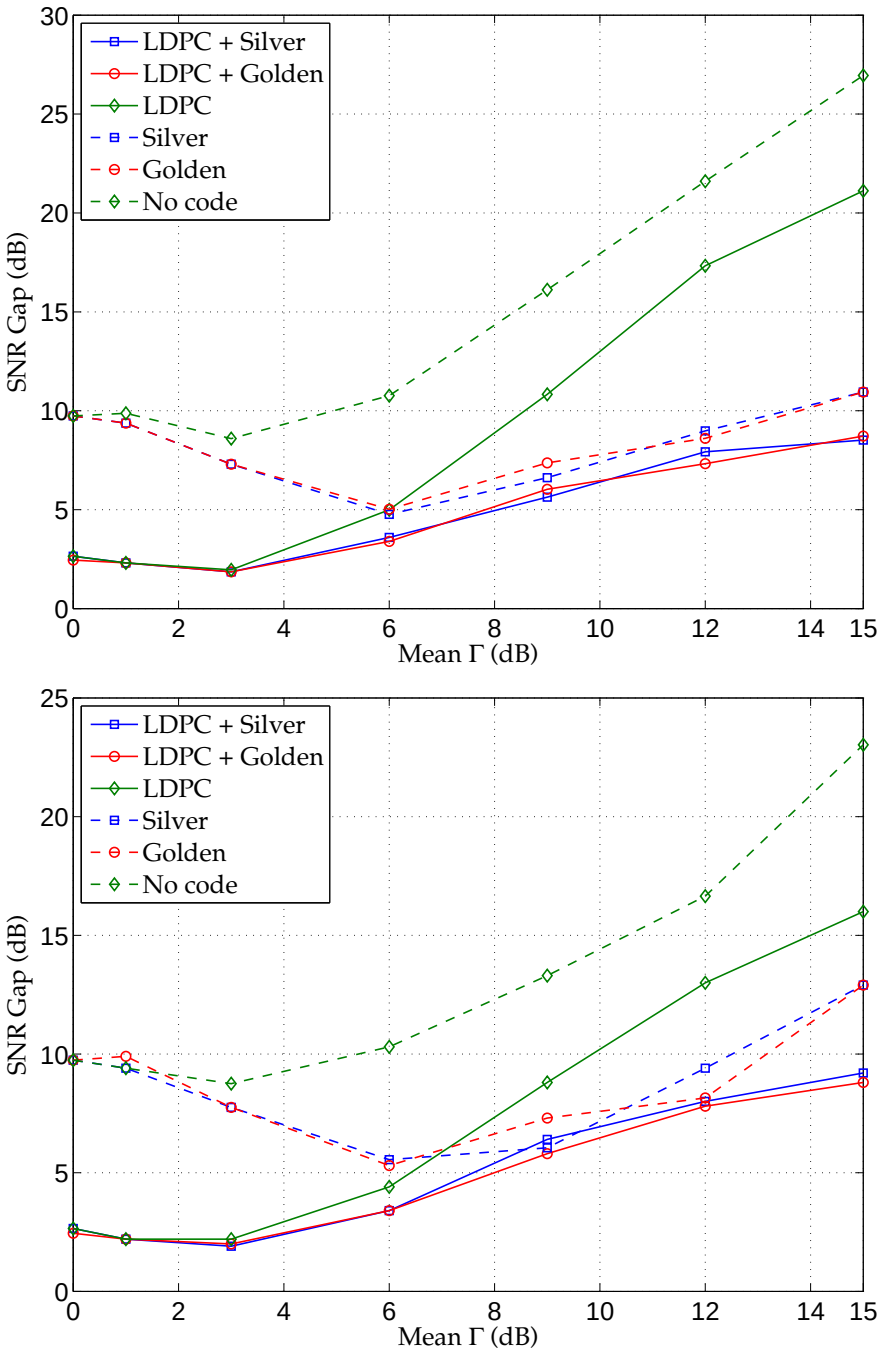

Figure 2.30: SNR gap for Maxwellian model (top) and phenomenological model (bottom) (16-QAM, $\mathbb{E}[\Gamma] = 3$ dB)

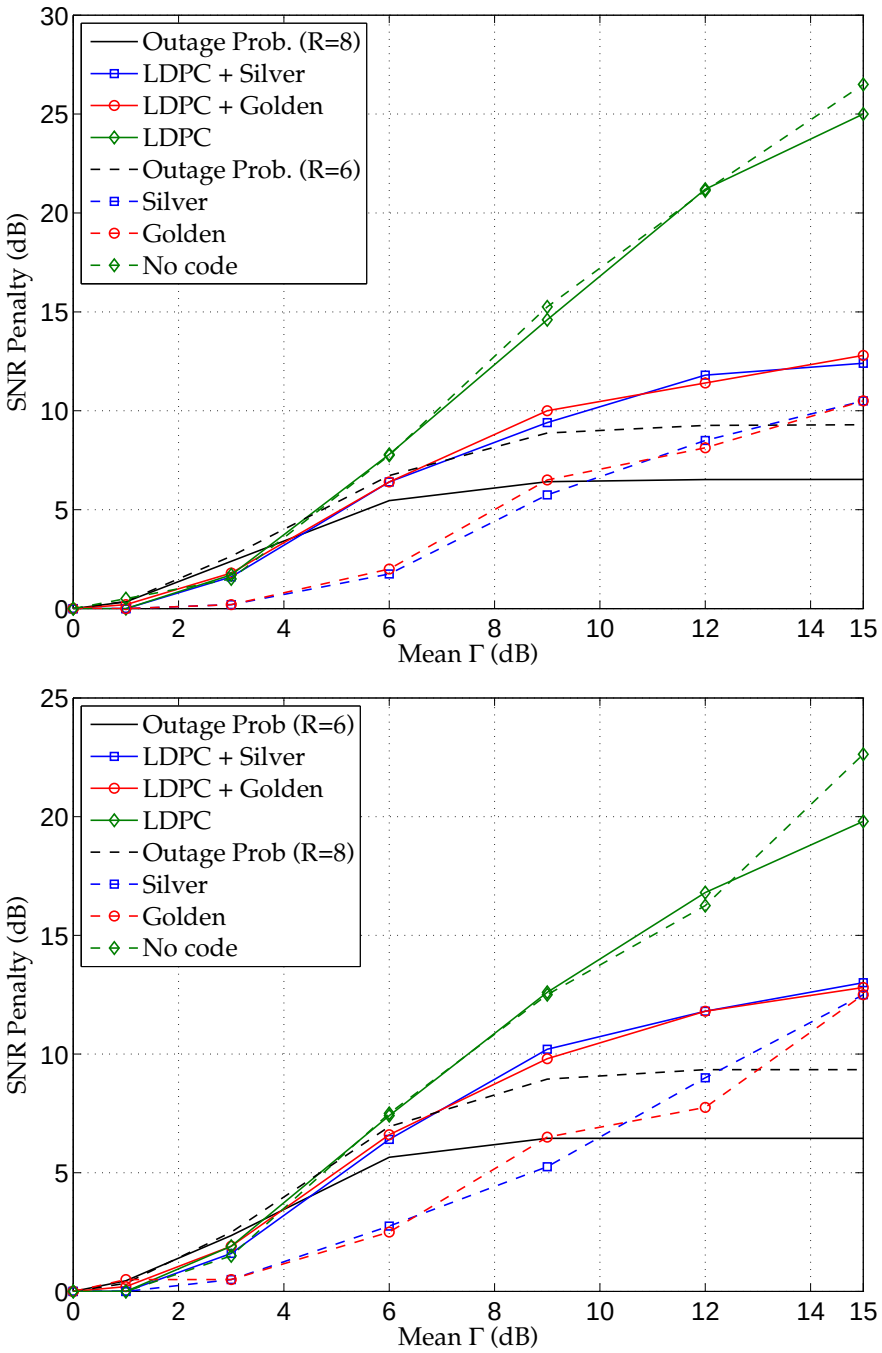

Figure 2.31: SNR penalty for Maxwellian model (top) and phenomenological model (bottom) (16-QAM, $\mathbb{E}[\Gamma] = 3$ dB)

might alleviate the need for powerful LDPC codes, allowing near-limit performance with less-expensive codes.

Chapter 3

Buffer-Assisted Optical Switching

3.1 Introduction

Historically, the optical transmission systems were deployed as transport solutions between network nodes. Indeed, first systems were installed to replace copper-based solutions which seemed to reach their limits. Then, the WDM systems have enabled a better use of the bandwidth of fibers by sharing it between independent traffics. Nowadays, optics is the preferred solution for virtually all non-wireless data links above a few meters' distance at bit rates from 1 to 100 Gbit/s per wavelength. In the most recent standard [47] about the 100 Gbits/s Ethernet, copper-based solutions are only proposed for the backplane interconnections or for the very short links (less than 7 m).However, the current optical transmission systems are almost only deployed as point-to-point links between nodes; the optical network thus can be seen as a quasi-static circuit switching.

In contrast, the packet-switched Internet Protocol (IP) traffic is the main user of optical transport networks today. Since its volume still increases, and the one from the legacy traffic gradually decreases, it will eventually become the only user of optical transport networks. Yet, all the switching and routing required by IP packets are nowadays performed electronically, requiring opto-electronic conversions of all the traffic carried optically, repeated all over the network. This separation between the optical transport and the IP switching leads to the unsustainable growth of the size and the power consumption of optical systems and electronic routers [48].

A cross-layer or hybrid equipment, in other words, an equipment which mixes the transport function of the optical layer with the switching function of the IP layer, may offer significant energy savings [49]. It may also decrease the buying cost and the maintenance cost.

Sadly all-optical routing techniques have a major loss issue; under practical load conditions, an optical packet switch has a high loss rate [50]. Indeed, several packets can go through the switch to the same destination at the same time, and thus are competitors for the output. In the electronic field, this problem has long been solved by the use of

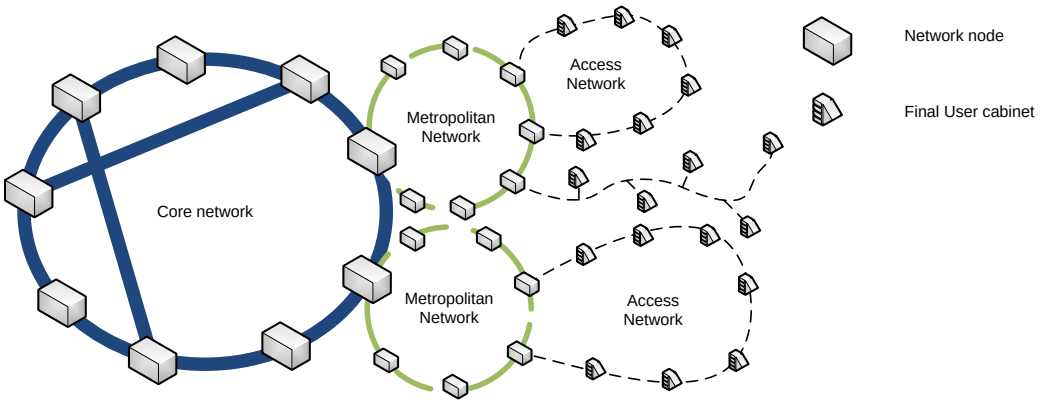

Figure 3.1: General architecture of carrier networks

buffers. Optical buffers exist [51], but have three main issues: a stored-burst can not be randomly accessed which can be a problem for the re-emitting; the memory lifetime is limited by the signal degradation over time; and they are very cumbersome, because they require amplifiers and recirculating loops of fiber [52]. Furthermore, slow-light solutions exist but are much less mature [53], especially, because of their fixed bandwidth-delay product issue which dramatically limits their interest for high data-rates.

Therefore we propose here a hybrid switch architecture. It is defined as a routing/switching equipment where packets are optically routed as long as possible, but where an electronic buffer exists in order to collect overflowed packets. We study in this chapter the gain in term of loss probability of such an equipment, and we also focus on the gauging and the impact of the electronic overflow system.

This chapter is organized as follows: Section 3.2 is dedicated to a brief overview of optical transport networks. Then, in Section 3.3, we analyze the performance of the cross-layer switch, as well as the performance of the overflow system itself. Section 3.4 deals with analytic models of the loss probability of such an equipment. Concluding remarks are given in Section 3.5.

3.2 Overview of optical transport network

3.2.1 A brief history

As illustrated in Fig. 3.1, the standard architecture of carrier's optical transport networks is divided in three areas. These different areas correspond to different aggregation levels for the network layer and to different transmission distances for the transport layer:

- **The Access Network:**
  These primary aggregation networks gather traffic from final network equipments such as legacy phone nodes, radio equipments, FTTH cabinets, Digital Subscriber

Line (DSL) nodes... They can also be the network inside a data-center. They are connected to the metropolitan network and implements optical transmission systems which cover distances between 0 and 100 km.

- **The Metropolitan Network:**
  This network gathers traffics from several access networks. It allows them to exchange data and connects them to the core network. It also permits delivery of the local traffic such as TV, or cached data. Its standard distance range is about few hundreds of kilometers.

- **The Core Network:**
  Also called backbone network, this network is the wide interconnection between several metropolitan networks and/or core networks. The aggregation at the network layer is as high as possible and the distance range for the optical transport network is over 500 km and often exceeds 1000 km.

Optical transmission systems were firstly installed in the core network, before being gradually deployed in the metropolitan network, then in the access network. As the distance ranges are very specific to a network area, suppliers often propose distinct solutions, which forces opto-electronic conversions at the border.

Since several kinds of traffic coexist in the network, choice were made to clearly separate the transport layer from the network layer, as shown in Fig. 3.2. Optical transmission systems thus are the universal transport network between nodes. The optical paths are fixed, and the smarter functions such as dynamic routing or failure recovery are made by the network layer. It leads to the deployment of more optical transponders than necessary in order to be able to re-route traffic in case of failure. In this context, fixed OADM were deployed, and physical interventions are required each time a change is needed in the optical routing plane. This method was efficient for the legacy traffic such as Sonet/Synchronous Data Hierarchy (SDH), since the changes in the optical routing plane were rare, and thus well scheduled.

From the optical transport network point-of-view, this circuit-switching technique is called Optical Wavelength Switching (OWS) [54]: A WDM channel from an ingress fiber is sent to a pre-defined egress fiber. Its wavelength may be changed thanks to wavelength converter.

Nowadays, the main part of the carried traffic comes from the IP network, and the constant growth of its bandwidth needs forces the optical transport networks to become more flexible. A first hurdle is past with the introduction of Reconfigurable Optical Add/Drop Multiplexer (ROADM) [55]. This technique enables a dynamic reconfiguration of optical paths, and may directly enable failure recoveries on the optical layer. An extension of this technique has been proposed in [56], where the granularity is the OFDM sub-band.

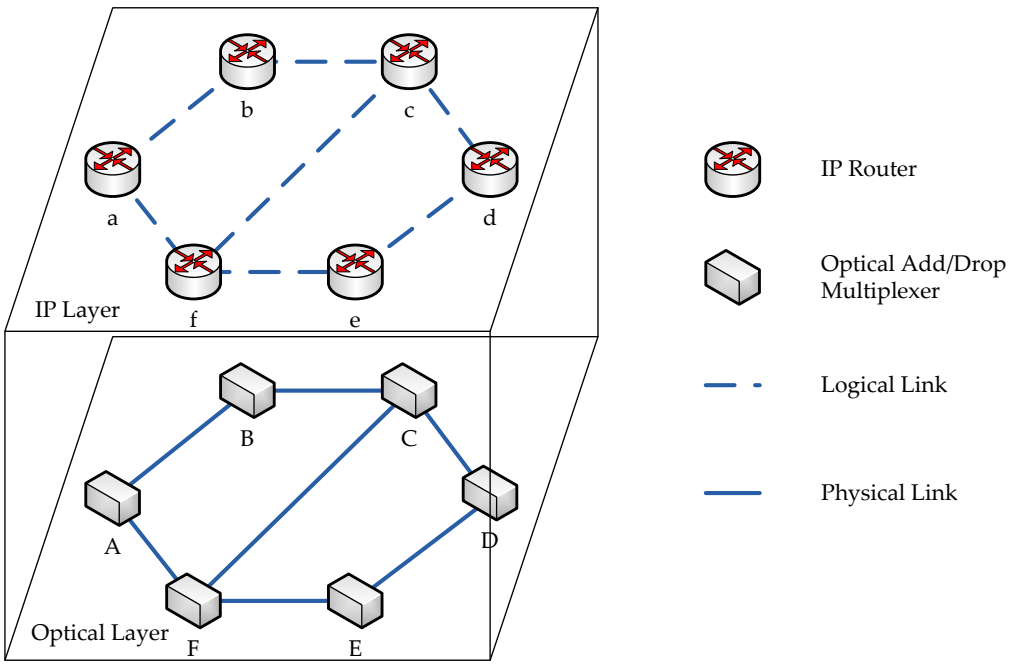

Figure 3.2: Layers separation in optical transport network

This dynamic resource allocation is only useful if the control planes of the two layers interact and collaborate. Most of time, the network layer is now based on MPLS [57]. This protocol allow a better use of the bandwidth than the IP protocol: it can transport IP packets, as well as frames from other protocols, having a strong respect to the QoS and the latency. Its control plane is decentralized, in other words, each node builds its own routing table based on its network knowledge and predefined rules. The main strength of MPLS is that it is widely used by all suppliers, and so it simplifies interconnections between the equipments.

In contrast, the optical transport network suppliers often propose closed and centralized control planes. The Generalised MPLS (GMPLS) is a first attempt to offer a standard communication framework between control planes [57]. From the optical transport network point-of-view, it permits the reservation and the release of a complete optical path, but also to find a cooperative solution between the layers and/or the equipments in case of failures. In the future, this control plane will be replaced by the implementation of the OpenFlow protocol [58]. Indeed, this protocol enables a better optimization of the forwarding plane in the network and transport layers.

3.2.2 Cross-layer techniques

Although several control planes enable a better collaboration between layers, especially for the failure recovery or the resources allocation, the bandwidth optimization is limited and the energy saving too. In fact, transponders are deployed for the prevention and/or

the further needs, and so are switched on although they transmit nothing. In this context, it becomes very attractive to mix not only the two control planes but also the two layers. This kind of equipment is called cross-layer or hybrid because the packets are as long as possible optically routed.

On the contrary of the OWS, the switching fabric may be reconfigured between each packet/burst. The working is very similar to this from an electronic switch. It allows a better use of the total bandwidth per fiber. These techniques can be classified depending on the aggregation level of data and how the destination is identified:

- **Optical Packet Switching (OPS)** [59]:
  This is the direct transposition of the electronic implementation of packet switching. Each time a packet arrives at a ingress node, a header is added to the packet, then it is sent to the next node. At each node, the header is read, and the packet is routed until its egress node. Because of the number of packets which transit on the network, this technique has a big congestion/loss issue.

- **Optical Burst Switching (OBS)** [50]:
  This technique is similar to the previous one, except that the ingress node aggregates as much as possible the ingress packets depending on their destination in order to construct bursts. Moreover, the control data are separately sent on a dedicated control channel. By reducing the load per node, this technique improves performance compared to the previous one, but it remains unusable out of the laboratory.

Even though these switching techniques are different, their nodes have the same functional architecture: a switching fabric with a control module in charge of its reconfiguration depending on the destination of ingress data. This architecture can be implemented in practice using *e.g.* semiconductor optical amplifier (SOA) gate matrices [60, 61], or fast tunable wavelength converters and wavelength switches [62, 63]. Therefore it is possible to build a global model for the switch. The different techniques are obtained by changing parameters such as the packet/burst length, the delay due to switching fabric reconfiguration, the delay at the entrance of the switch, the number of destinations and/or channels. Since the general architecture of the different techniques is similar, we only consider here an optical burst switch.

3.3 Optical Burst Switch with overflow system

As explained in [50], the OBS has a major loss issue. The simplest solution is to increase the number of channels per destination. In this way, the probability that an ingress burst find an egress channel increases. These extra-channels may be WDM channels with the sufficient number of wavelength converters, or fiber cores from multi-core fibers, or even parallel fibers [64]. The first solution is the most accurate one for current networks.

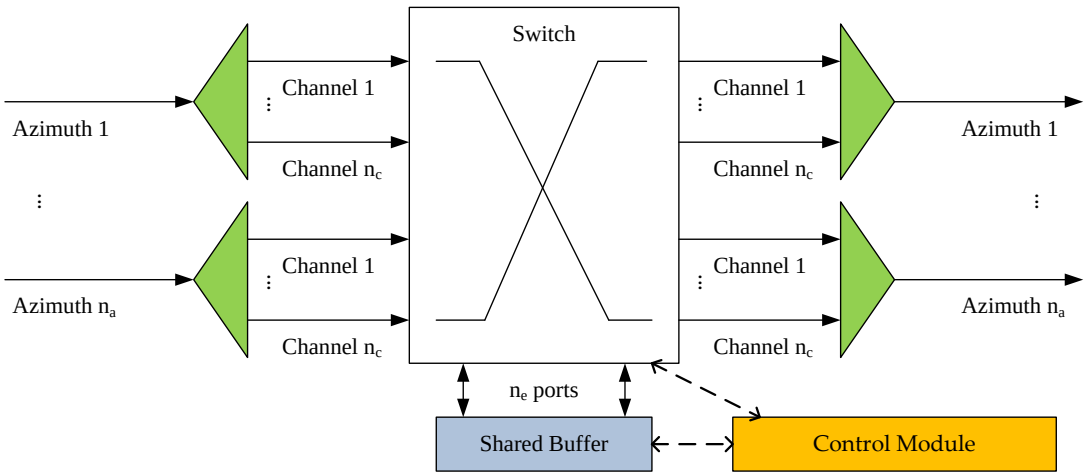

Figure 3.3: General architecture of the considered optical burst switch

However, since wavelength converters usually are electronic systems, or even back-to-back transponders, they do not enable significant energy savings or cost savings compare to standard architecture. The other ones may enable substantial power savings, but the number of extra-channels then is more limited.

Another way to improve the performance, is the use of buffers, just as in the electronic switches. In practice, a buffer can be directly implemented in the optical field [51] or using electronic systems. As explained in the introduction, an optical buffer is constructed using recirculating-loop and fiber-delay-line [52]. This solution has a memory lifetime issue, and is very cumbersome.

We choose here to consider electronic overflow systems. Although these buffers consume more energy, they have a lot of advantages: they permit a random-access to stored bursts, they avoid the life-time issue, and they occupy less space. This buffering technique has been presented in [59] and implemented in [63, 65], but performance of the architecture was not studied. We propose here to study that, using metrics and tools from queuing theory. The overall performance is evaluated using the loss probability, and we also evaluate the gain offered by this architecture compared to all-optical solutions. Moreover, we study the gauging of the overflow system (required amount of Random Access Memory (RAM)), and its influence on the performance (latency, collisions caused by re-emitting).

3.3.1 System model

Let us consider an optical burst switch, working in asynchronous mode: bursts can arrive at any instant. Its functional architecture is described in Fig. 3.3: the switch has n_a links to remote switches, called azimuths; each azimuth is supposed to be bidirectional and support n_c independent channels in each direction. Furthermore, we assume that an

egress burst can indifferently use any channel of an azimuth; although this assumption is straightforward for multiple fibers or fiber cores per azimuth, it implies a sufficient number of wavelength converters in the case of WDM channels.

In addition, our switch has a shared electronic buffer as overflow system. Thus, bursts may be stored, if no channel is available on their egress azimuth. It is accessed through n_e input ports, n_e output ports, and has an unlimited amount of RAM.

The switch operates in the following way. When a burst arrives, its egress azimuth is read from *e.g.* its header, or its control channel. If a channel is available on the egress azimuth, the burst is directly sent on its way over it. Otherwise, if an input port is available on the overflow system, the burst is sent there. Otherwise, the burst is dropped.

We also set the following two rules: first, a burst cannot get back to its incoming azimuth. Second, bursts which are stored in the electronic buffer, have priority when a channel for their destination is released.

The traffic of each ingress channel is model as an on-off process, the "on" periods corresponding to burst transmissions and the "off" periods to idle times between bursts. The idle times are assumed to have independent, exponentially distributed durations. But we consider two cases for the burst duration:

- **Pure Chance Traffic** where the burst durations are independent, exponentially distributed process. The mean burst duration is denoted t_μ. This model is considered as the worst-case scenario [66] but allows tractable derivations for the loss probability.
- **Traffic Shaping** where the burst duration is fixed to t_μ. This technique really is the burst switching. Indeed, it consists in delaying and grouping packets/bursts in order to build bursts which have almost always the same size. It permits to reduce the random aspect of the traffic (only the idle time remains random). This regulated traffic improves performance of the system and is extensively used in practice.

During the simulations, the destinations of each ingress bursts is uniformly chosen between all other azimuths than the ingress one. Furthermore, we set $t_\mu = 10$ μs. We choose this value because it represents about 100 kbit for standard 10 Gbit/s systems, and may correspond in a jumbo Ethernet frame or an aggregation of several standard Ethernet packets (12.3 kbit). The relationship between the overall system load ρ, the mean burst duration t_μ, and the mean idle time t_ν, is given by:

$$\rho = \frac{t_\mu}{t_\mu + t_\nu} \tag{3.1}$$

3.3.2 Loss probability

In order to evaluate the influence of the overflow system on the performance, we fully simulate the chosen architecture. Simulations were made under the two traffic assumptions: pure chance traffic and traffic shaping. Moreover, we have considered two broad

cases, which seems to us as representative as possible of future optical transport network: $n_a = 10$ azimuths and $n_a = 5$ azimuths.

3.3.2.1 Performance under the Pure Chance Traffic assumption

We present in this Section, results obtained under the pure chance traffic assumption. In Fig. 3.4, we plot the burst loss probability versus the switch load for the two broad cases: $n_a = 10$ azimuths, on the top, $n_a = 5$ azimuths on the bottom. In each case, we consider $n_c = 10$ channels per azimuth and different number of overflow ports: all-optical burst switch ($n_e = 0$) and $n_e = 5, 10, 15, 20$.

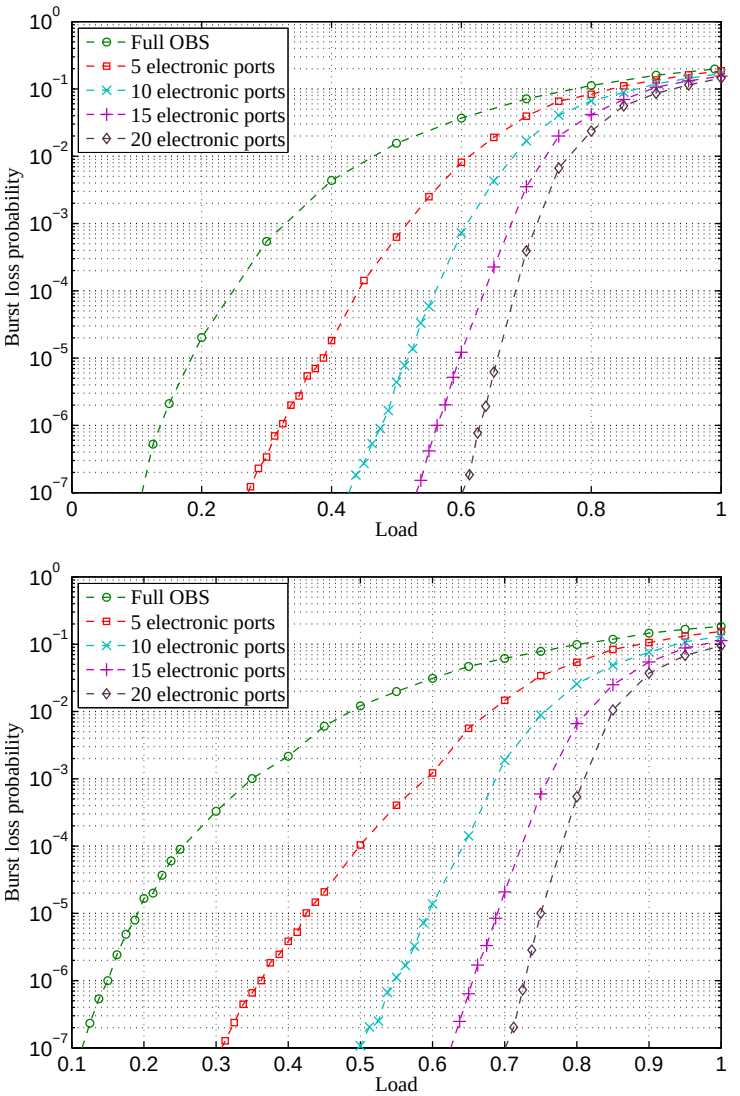

Figure 3.4: Loss probability vs system load in two cases: $n_a = 10$, $n_c = 10$ (top) and $n_a = 5$, $n_c = 10$ (down)

Simulation results confirm that the burst loss is really significant at low load for the

all-optical case: the sustainable load for a burst loss probability of 10^{-7} is around 0.1 for the two broad cases. In contrast, they show that the electronic buffer brings a significant gain, even for relatively small number of electronic ports: the sustainable load reaches 0.3 for the two broad cases with $n_e = 5$ overflow ports. The sustainable load even exceeds 0.6 for $n_e = 20$ ports for the first broad case and $n_e = 15$ for the second one. As expected, the higher the number of port on the overflow system is, the higher is this gain.

In Fig. 3.5, we display the maximum achievable system load with a loss probability of 10^{-7}, versus the number of electronic ports of the overflow system. We consider again the two broad cases: $n_a = 10$ azimuths on the top, $n_a = 5$ azimuths on the bottom; and different number of channels per azimuth in each broad case: $n_c = 1, 5, 10, 20, 50, 100$. The target loss probability of 10^{-7} corresponds to a threshold above which any switching technology would be disregarded as an industrial solution.

First of all, the results confirm that the increase of the number of channels per azimuth is another way to solve the congestion issue [64]: It is obvious when $n_e = 0$. In addition, we notice that, at low values of n_e, the sustainable load curves for $n_a = 5$ and $n_a = 10$ are rather similar; then, the sustainable load increases with n_e, as expected, and reaches 1 for $n_e = n_a \times n_c$, since having as many electronic ports as ingress channels means that all incoming bursts can always be collected. We also remark that, beyond a certain number of overflow ports, curves are close, and an equilibrium thus exists between the number of channels per azimuth and the number of electronic ports. Choosing, as a minimum acceptable operating point, a target load of 0.6, we observe that $n_e = 20$ overflow ports are sufficient in the case $n_a = 10$ azimuths and $n_e = 15$ overflow ports are sufficient in the case $n_a = 5$. These values of n_e will then be used for the performance analysis of the overflow system.

In order to evaluate the gain offered by the overflow system, we plot in Fig. 3.6 the load gain versus the number of electronic ports. The load gain is defined as the increase percentage of the sustainable load compared to the buffer-less case. This gain has been plotted for $n_c = 1, 5, 10, 20, 50, 100$ channels per azimuth, in the two broad cases $n_a = 10$ azimuths on the top, and $n_a = 5$ azimuths on the bottom.

The gain is very high, especially for a small number of channels per azimuth: over 4000% for 15 overflow ports with $n_c = 5$ and both broad cases. It is even more significant when $n_c = 1$. In contrast, although this gain grows with the number of overflow ports, it is much more limited at high numbers of channels per azimuth, where the burst loss issue is less severe. Therefore, this switch architecture is especially interesting when extra channels are expensive.

3.3.2.2 Performance under the Traffic Shaping assumption

We performed simulations under the traffic shaping assumption, in order to evaluate its discrepancy in term of performance with the pure chance traffic model.

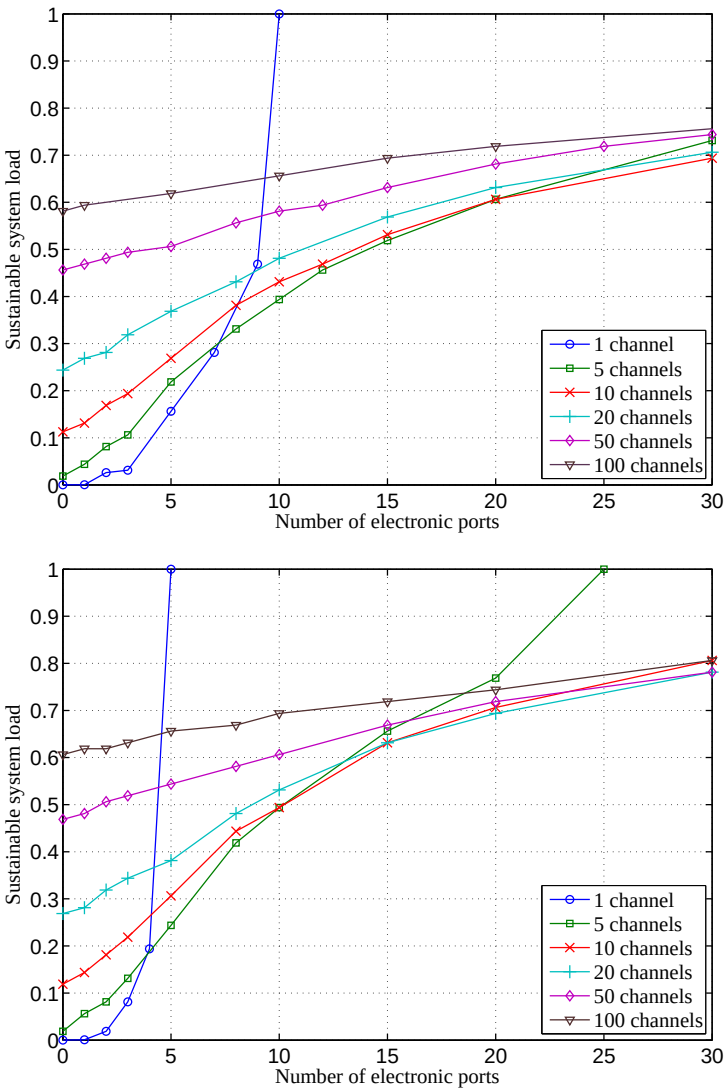

Figure 3.5: Sustainable system load vs number of electronic ports for two number of azimuths $n_a = 10$ (top) and $n_a = 5$ (down), and for different number of channels $n_c = 1, 5, 10, 20, 50, 100$

Fig. 3.7 shows the burst loss probability versus the system load for the two broad cases: $n_a = 10$ azimuths on the top, and $n_a = 5$ azimuths on the bottom. We plot, on the same figure, the simulation results under the traffic shaping assumption and under the pure chance traffic assumption, for $n_c = 10$ channels per azimuth and $n_e = 0, 5, 10, 20$ overflow ports.

We notice that our hybrid switch has the same behavior whatever the traffic model. Although we remark that for very high values of the load, the curves for traffic shaping case are slightly better, pure chance traffic can be considered as a good hypothesis for the performance evaluation.

Fig. 3.8 and Fig. 3.9 respectively are the same as Fig. 3.5 and Fig. 3.6, but under the traffic shaping assumption. The first one plots the sustainable system load for a loss

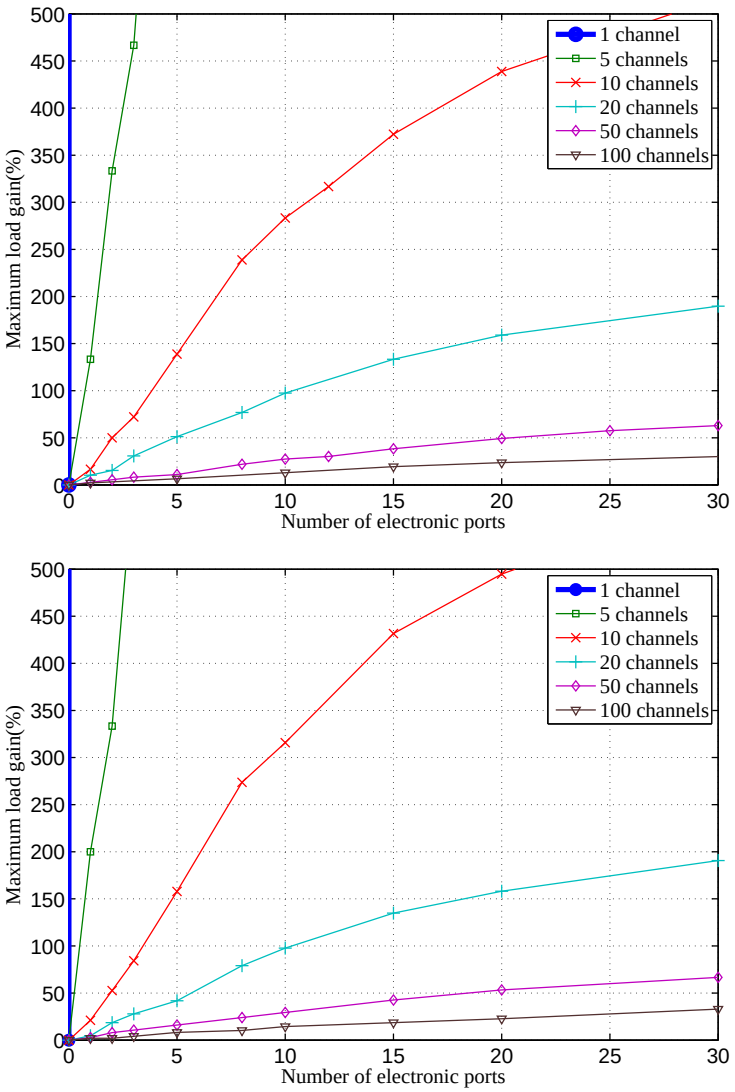

Figure 3.6: Gain in system load vs number of electronic ports for two number of azimuths $n_a = 10$ (top) and $n_a = 5$ (down), and for different number of channels $n_c = 1, 5, 10, 20, 50, 100$

probability of 10^{-7}, versus the number of electronic ports. The evaluation is done in the two broad cases $n_a = 10$ azimuths on left, and $n_a = 5$ azimuths on right; and for $n_c = 1, 5, 10, 20, 50, 100$ channels per azimuth. The second one presents the load gain, as previously defined, in the same context than the first curve.

Simulation results obtained with the traffic shaping model, are very similar to the results obtained under the pure chance traffic assumption. The electronic buffer still offers a significant gain in term of loss probability, especially when the number of channels per azimuth is low. These similarities are interesting because, as explained in the system model, frameworks exist for the derivations of the loss probability under the pure chance traffic assumption.

Since the traffic model seems to have a slight influence on the system performance,

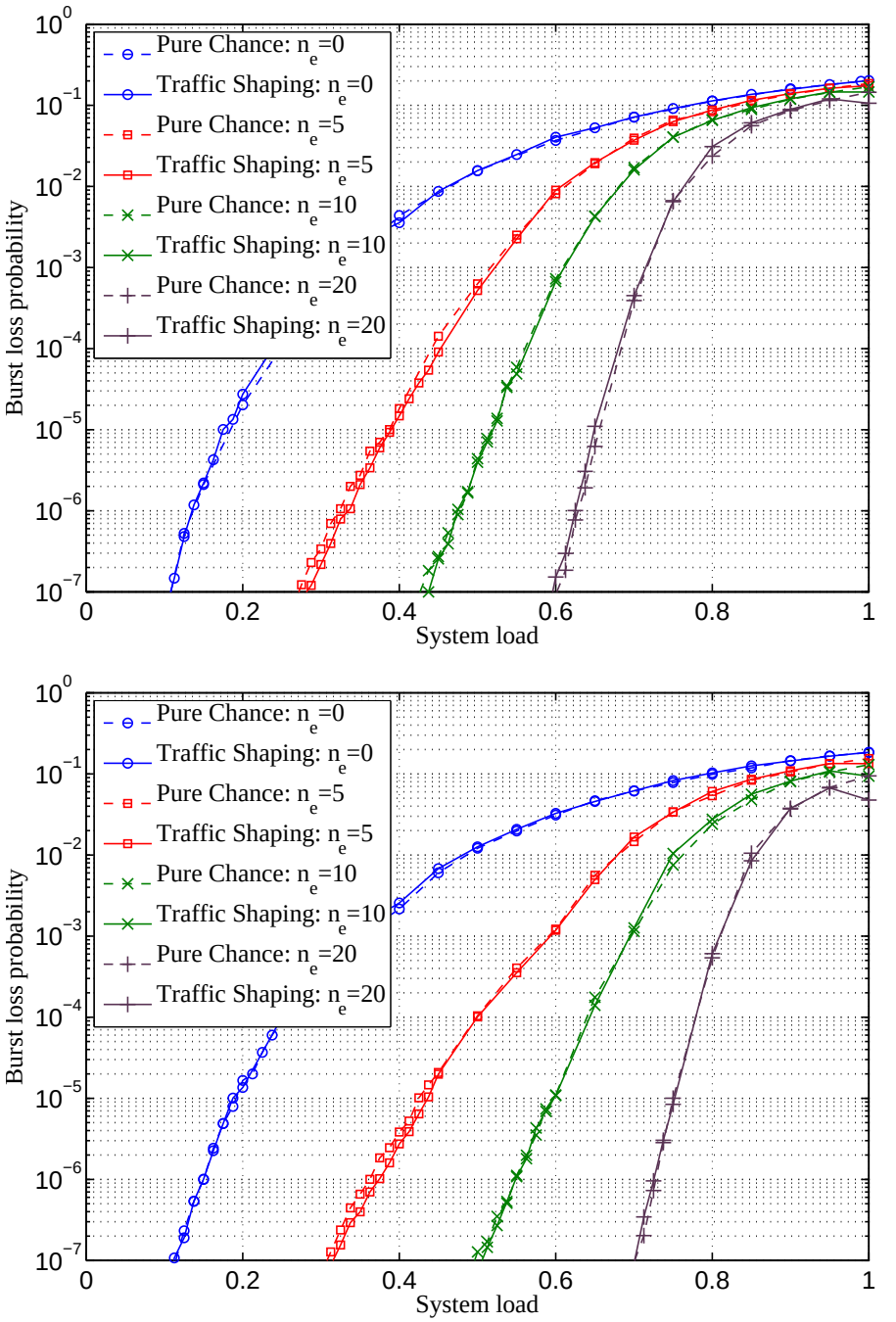

Figure 3.7: Loss probability vs system load in two cases: $n_a = 10$, $n_c = 10$ *(Top) and* $n_a = 5$, $n_c = 10$ *(Bottom); and for two models of traffic*

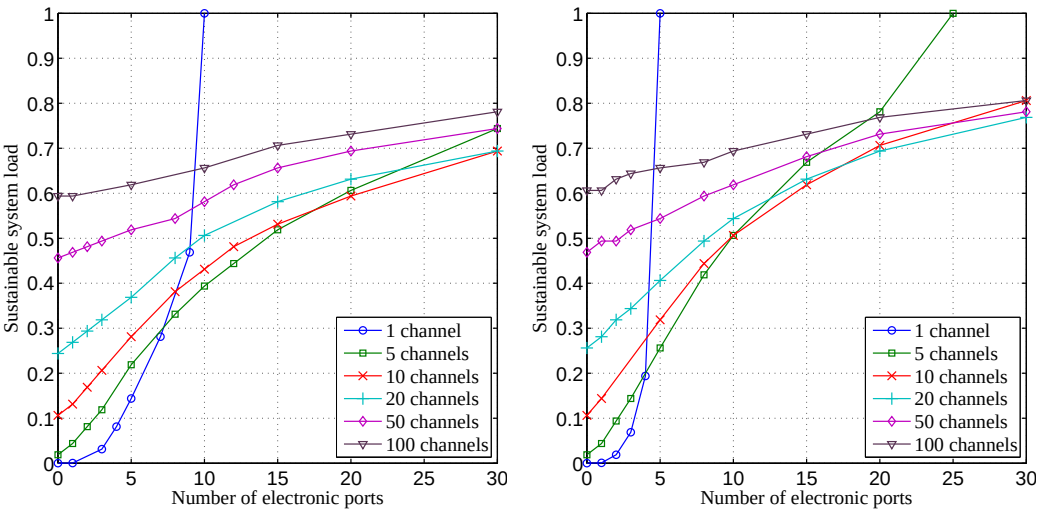

Figure 3.8: Sustainable system load vs number of electronic ports for two number of azimuths $n_a = 10$ (Left) and $n_a = 5$ (Right), and for different number of channels $n_c = 1, 5, 10, 20, 50, 100$

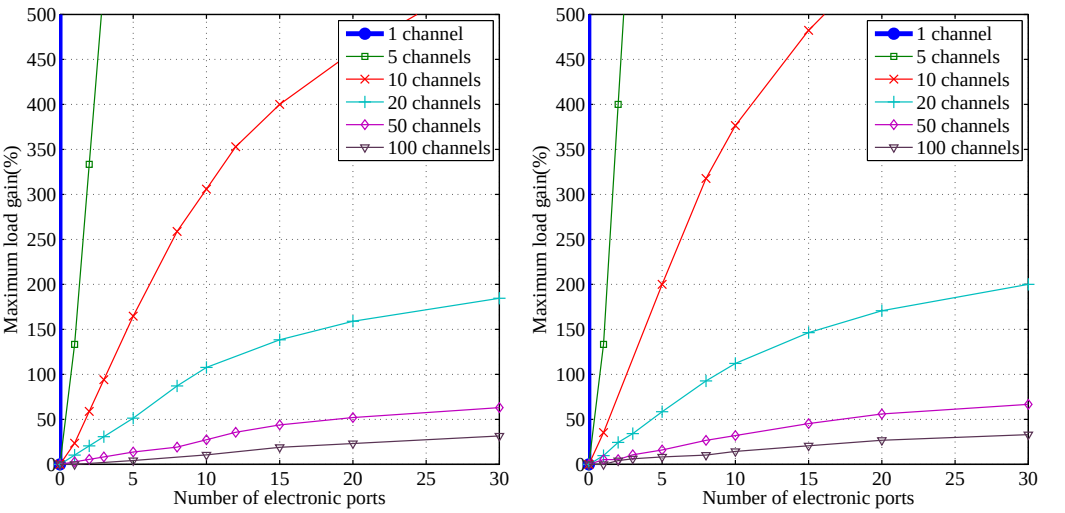

Figure 3.9: Gain in system load vs number of electronic ports for two number of azimuths $n_a = 10$ (Left) and $n_a = 5$ (Right), and for different number of channels $n_c = 1, 5, 10, 20, 50, 100$

we can draw a quick summary on energy savings of such a hybrid equipment. We can compare the number of input/output electronic ports required by the overflow system (n_e) to that of a full-up electronic IP router ($n_a \times n_c$). Assuming a node with 10 azimuths with 10 channels each, a hybrid optical switch with $n_e = 20$ overflow ports satisfies our target load and loss probability values (a sustainable load of at least 0.6 for a loss probability of 10^{-7}). And it may lead to a five-fold reduction of the energy consumption compared to the 100 ports of an all-electronic switch. The energy savings are less clear in the case of WDM channels, where the energetic cost of wavelength converters must be factored in. There may be an optimum number of wavelength converters and overflow ports to be found.

3.3.3 Performance of the overflow system

We interest now in gauging of the electronic buffer and in evaluating its influence on the overall performance. As explained before, we consider that a loss probability of 10^{-7} for a sustainable system load of 0.6 is acceptable. Therefore, we choose these two cases in next simulations: $n_a = 10$ azimuths with $n_c = 10$ channels per azimuth and $n_e = 20$ ports on the overflow system, and $n_a = 5$ azimuths with each $n_c = 10$ channels and $n_e = 15$ ports on the overflow system. The measurements are made when the system is stationary—that is, we consider that the transitory regime ends after the first burst loss occurs.

3.3.3.1 Memory consumption

In Fig. 3.10, we plot the mean and the maximum number of bursts stored in the buffer versus the system load for the two cases: $n_a = 10$ azimuths on the left, and $n_a = 5$ azimuths on the right. Since we want to evaluate the amount of memory needed by the switch, we only take into account the bursts which are waiting to be re-emitted. Notice that the results obtained for the maximum are preliminary, because they may depend on the simulation duration. It would be more significant to evaluate the probability distribution of the number of stored bursts using the percentile for a certain time; further simulations will be performed to this end. Nevertheless, we believe that this preliminary results are accurate, as the values do not vary much between different simulation durations.

Simulation results show that the mean number of stored bursts is low, even for high values of the load. For instance, in the two configurations, for a load of 0.6, there is less than one burst in memory on average. And the maximum is about tens bursts. When the system load grows, the mean is about ten bursts in the two configurations, and the maximum tends toward 50 bursts in the first case and 40 bursts in the second one. The amount of memory required is also reasonable compare to the number of bursts which can be simultaneously treated by the switch ($n_a \times n_c$). Considering a system with channels at 10Gbit/s and where the mean burst duration is 10 μs, the required quantity of RAM is about few megabytes for the two configurations.

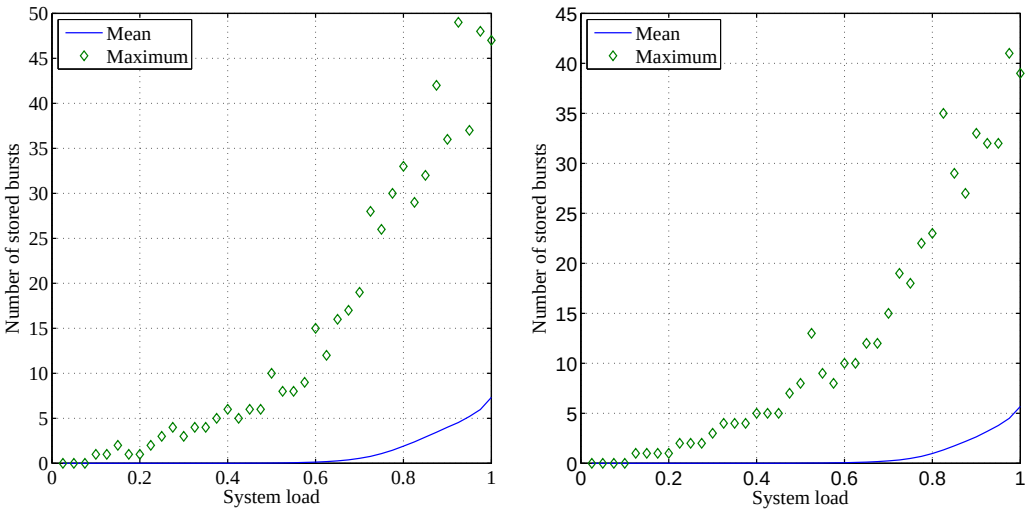

Figure 3.10: Memory consumption vs system load in two cases: $n_a = 10$, $n_c = 10$, $n_e = 20$ (Left) and $n_a = 5$, $n_c = 10$, $n_e = 15$ (Right)

3.3.3.2 Latency

Fig.3.11 proposes the mean and the maximum latency experienced by the stored bursts versus the system load, for the two study cases: $n_a = 10$ azimuths on the left, and $n_a = 5$ azimuths on the right. The latency experienced by a stored burst is defined as the time that the burst is waiting before being re-emitted.

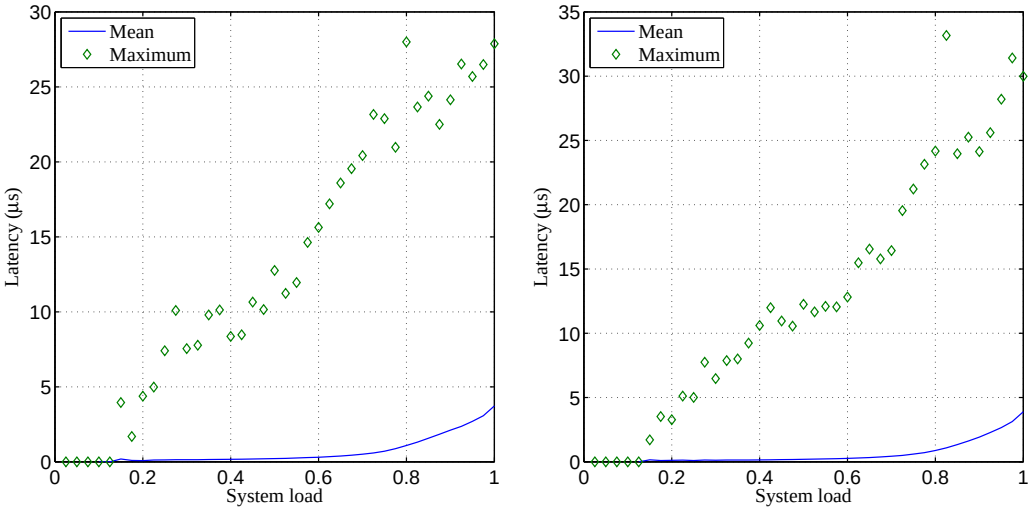

Figure 3.11: Latency vs system load in two cases: $n_a = 10$, $n_c = 10$, $n_e = 20$ (Left) and $n_a = 5$, $n_c = 10$, $n_e = 15$ (Right)

Simulation results show that even for high system loads, the mean latency is under the mean burst duration. As in the previous Section, the curves for the maximum presented here are just preliminary. However, as the values do not vary much between different

simulation durations, we believe that these results are enough accurate. Therefore, we can notice that the maximum latency is not so serious, up to 4 or 5 times the mean burst duration, in the two configurations. The latency will also not have a great influence on the global system performance.

3.3.3.3 Second order collision

On the contrary of first order collisions which occur when an incoming burst cannot be directly routed whatever the reason, a second order collision is defined as a collision between an incoming bursts and a re-emitted burst. It occurs when an incoming burst cannot be directly routed because the egress azimuth is full, and at least a stored bursts is being re-emitted over this azimuth. There are two kinds of second order collisions: second order collisions which pass through the buffer; and second order collisions which lead to burst losses.

Such phenomena could be avoided by refraining from re-emitting a stored burst, if other bursts with the same destination are incoming, and would thus cause a second-order collision. Detecting incoming bursts in advance would require an "observation window": a delay between, on the one hand, burst detection and destination processing; and, on the other hand, the bursts being actually sent into the switching matrix. Practical switches need this delay anyway, in order to give time to header-processing electronics to reconfigure the switching matrix; typically, Fiber Delay Line (FDL) are inserted on the ingress channel in order to delay bursts.

In the next simulations, we consider all the second-order collisions, and the ones which could be avoided by two observation durations: 250 ns and 500 ns which correspond respectively to 50 m and 100 m of standard SMF fiber [65]. Fig. 3.12 shows the three probabilities described above as well as the overall burst loss probability obtained previously versus the system load for the two study cases : $n_a = 10$ azimuth on the top and $n_a = 5$ azimuth on bottom.

First, simulation results show that phenomenon is non-negligible. Second order collisions exist even for relatively low values of the load (around 0.2). Most of time it leads to a useless passing through the overflow system. In addition, the second order collisions are the main issue of the overall system performance: in the two cases, the overall burst loss probability and the blocked bursts probability due to second order collisions are very close.

However, the curves obtained taking an "observation window" into account, show that an "observation window", before re-emitting stored bursts, will not only improve the system performance by reducing the burst losses, but also reduce energy consumption of the overall system by reducing the number of opto-electronic conversions. The gain in terms of performance is noticeable, even for a relatively small observation duration: about 1/10 of the burst losses can be avoided in the 250 ns case. We also notice that the longer

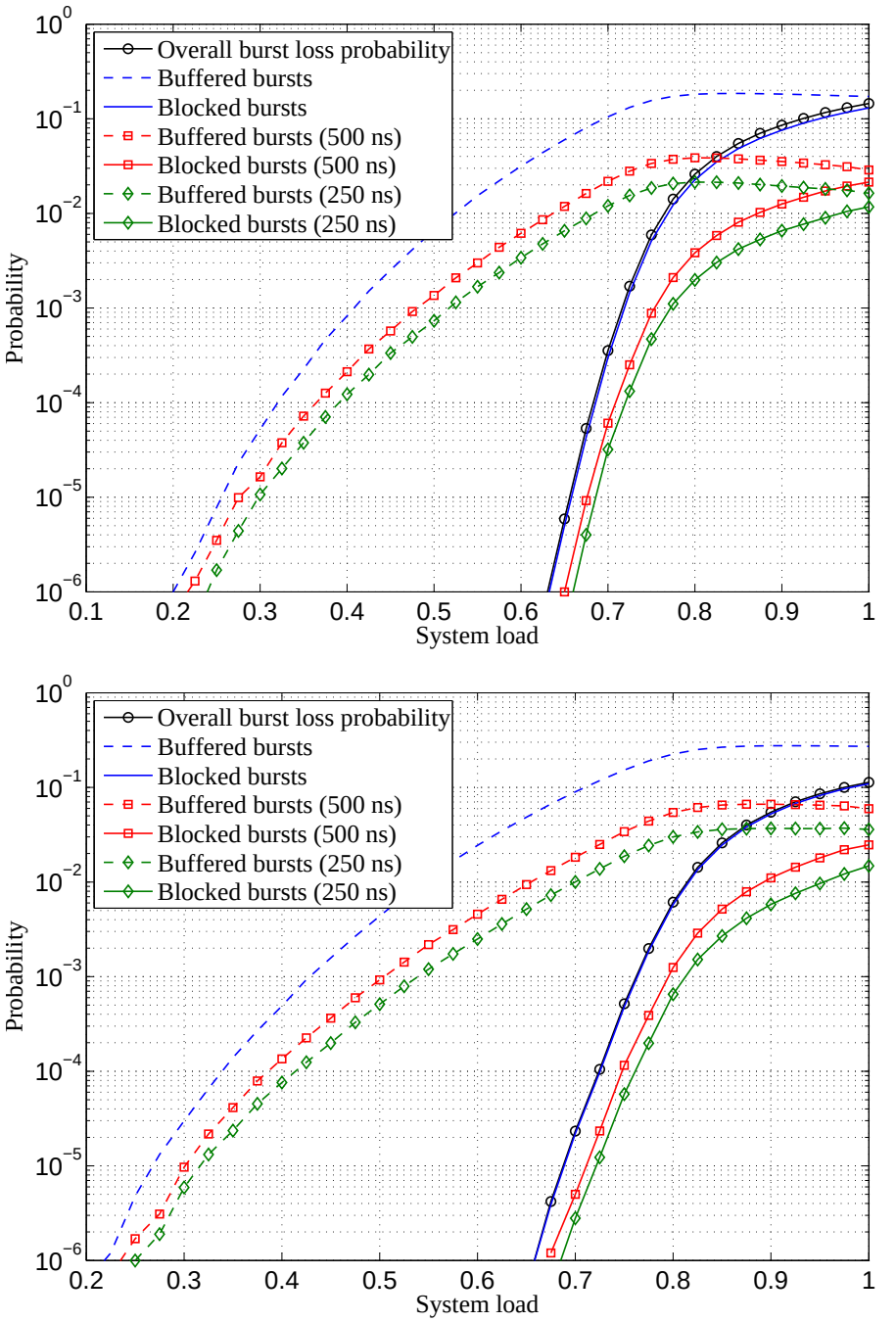

Figure 3.12: Second order collisions vs system load in two cases: $n_a = 10$, $n_c = 10$, $n_e = 20$ *(Top) and* $n_a = 5$, $n_c = 10$, $n_e = 15$ *(Down)*

the "observation window", the bigger the performance improvements, as expected. Just as the number of channels per azimuth, and the number of overflow ports, the duration of the observation window plays an important role in the hybrid switch's performance.

3.4 Analytical model for the loss probability

In the previous Section, we analyze by simulation the performance of an optical burst switch with electronic buffer. However, an analytical model allows us to better understand the overall problem, and the influence of each parameters. Moreover, it gives also predictions faster than simulations, and then enables a faster gauging of practical systems. Therefore, we propose here some closed-form expressions for the burst loss probability of such an equipment: a very tight one for the all-optical OBS System case ($n_e = 0$), and a very loose upper-bound for the complete system.

3.4.1 All-optical burst switch

3.4.1.1 Model

From a queuing theoretical point-of-view, considering pure chance traffic, two main frameworks exist for the loss probability: the Erlang one and the Engset one [66]. The main difference between them comes from how the traffic sources are taken into account. The Erlang framework is the most used, because it corresponds to a worst case in terms of traffic intensity. Indeed, the traffic is modeled by a single Poisson process, which corresponds to an infinite number of traffic sources. The Engset framework is equivalent to the Erlang one, except that the traffic is offered by a limited number of sources. When the number of sources is high, the two frameworks give similar results [66]. However, for a relatively low number of sources, the Erlang framework induces an under-evaluation of the performance; and some of the cases we simulated, *e.g.* $n_a = 5$ azimuths and $n_c = 1$ channels per azimuth, are few-source situations. Therefore, we choose the Engset framework to build our analytical model. Notice that this framework was also used in [67, 64] for a similar problematic.

In order to use the Engset formula we have to transpose our architecture in the framework. The all-optical burst switch, as shown in Fig. 3.3, is equivalent to the architecture proposed in Fig. 3.13. Indeed, each egress azimuth can be considered as an independent system which receives bursts from $(n_a - 1)n_c$ sources (because of the non-return rule). As the burst repartition between the egress azimuths is uniform, all these systems are equivalent. And the global loss probability is equal to the loss probability of one of them.

An Engset system is defined by three parameters: its number of service channels, its number of sources, and its calling rate (or traffic intensity) per source. Since we already have the number of service channels and the number of sources of the equivalent system,

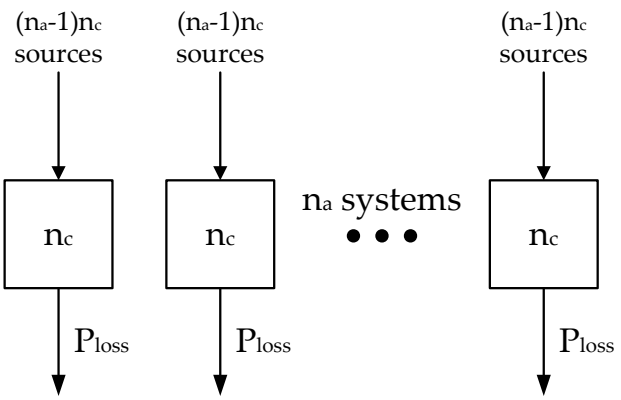

Figure 3.13: Equivalent architecture of the all-optical burst switch

we need the calling rate per source in order to obtain the loss probability. It can be easily deduced from the overall system load ρ defined in Eq. (3.1). We can reason on the offered load A (total amount of traffic), equal to the product of the number of sources for the overall switch and the calling rate per source (equal to the load in our case), of the overall system:

$$A = \rho \cdot n_a n_c \tag{3.2}$$

This offered load must also be equal to that in the equivalent architecture, that is, the calling rate per source a' multiplied by the number of sources per system, multiplied by the number of systems:

$$A = a' \cdot (n_a - 1)n_c \cdot n_a \tag{3.3}$$

Therefore:

$$a' = \frac{\rho}{n_a - 1}. \tag{3.4}$$

This expression is given as a function of the overall system load.

Thus, a closed form expression for the loss probability of an all-optical burst switch is obtained. The Engset Formula [66] can be written taking system parameters into account:

$$P_{loss}(n_a, n_c, \rho) = \frac{\begin{pmatrix} (n_a-1)n_c - 1 \\ n_c \end{pmatrix} \beta^{n_c}}{1 + \sum_{i=1}^{n_c} \begin{pmatrix} (n_a-1)n_c - 1 \\ i \end{pmatrix} \beta^i} \tag{3.5}$$

with $\beta = \left(\frac{n_a - 1}{\rho} - 1\right)^{-1}$.

In the case of a hybrid switch, this expression no longer represents the blocking probability, but will be used in Section 3.4.2.1 as the part of the ingress traffic directed into the buffer.

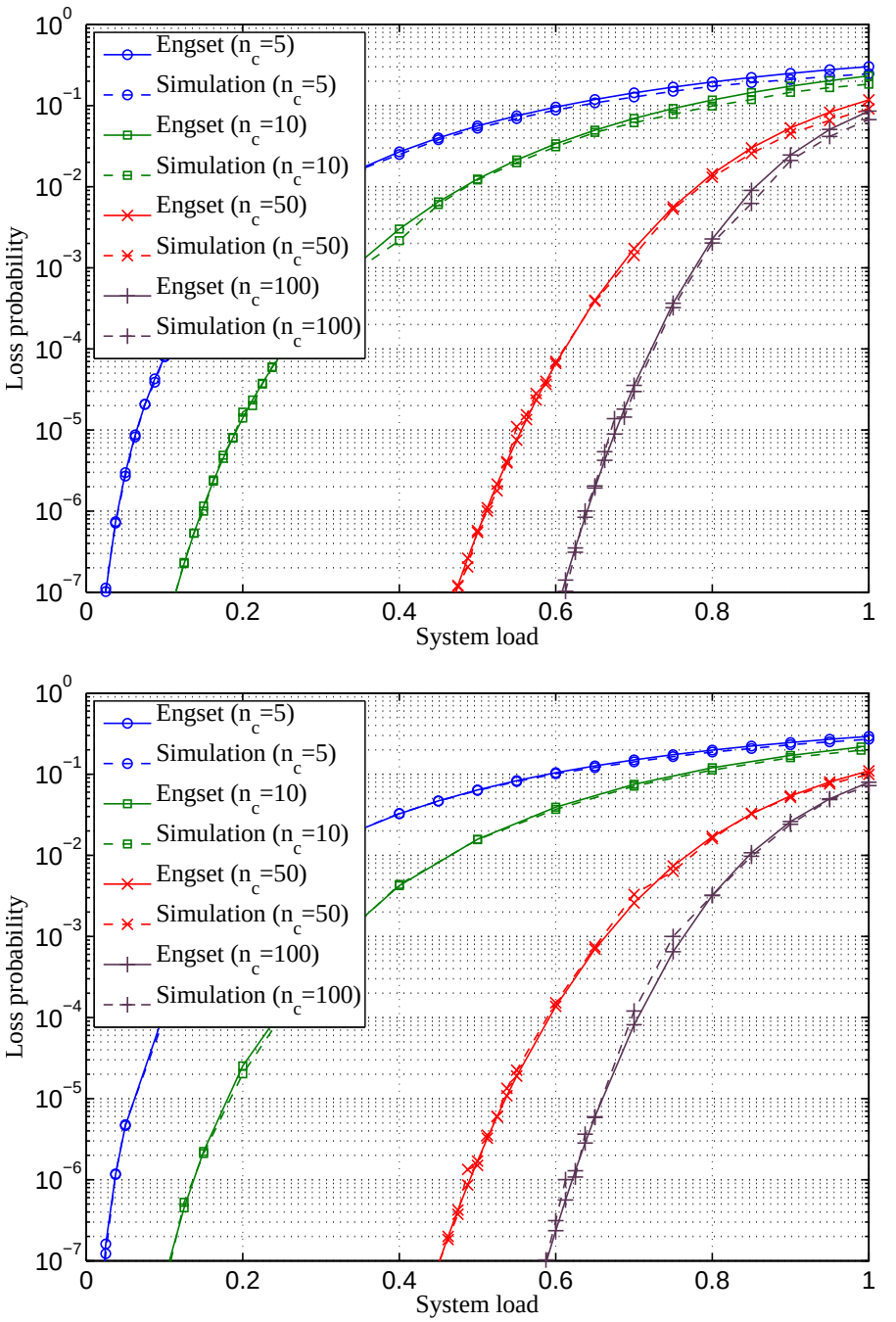

Figure 3.14: Theoretical and simulated blocking probability of the all-optical burst switch versus system load for $n_a = 5$ azimuths (Top) and $n_a = 10$ azimuths (Down) and different number of channels.

3.4.1.2 Numerical illustrations

In aim to evaluate the accuracy of the closed-form expression, we compare it with simulations of the all-optical burst switch (without electronic ports) as the ones presented in previous Section. Fig. 3.14 shows the theoretical and the simulated burst loss probability versus the system load for different configurations: $n_a = 5$ azimuths on the top and $n_a = 10$ azimuths on the bottom; and $n_c = 5, 10, 50, 100$ channels per azimuth.

Apart from a slight deviation between theory and simulation for high values of the load, we notice that theoretical curves fit well with simulated one. This deviation comes from simulation, where the respect of pure chance traffic assumption becomes questionable at high load value. Nevertheless, the closed-form expression presented here offers a good evaluation of the loss probability of the all-optical burst switch, or of the probability that an incoming burst tries to go in the buffer.

3.4.2 OBS with overflow system

3.4.2.1 Model

Although the Engset framework offers an expression of the loss probability of a system if the ingress traffic is pure chance, it does not give any further information about statistics of the blocked or overflowed traffic. Sadly, these inputs are required to evaluate the blocking probability of the complete system. [51, 68] work on the blocking probability of optical switch with optical buffer, but they do not take the number of channels per azimuth into account.

In [69], a framework was developed in order to compute the loss probability of several Engset systems which have a common overflow system. It corresponds to the general architecture of our optical switch, except that re-sending of bursts is not taken into account.

A primary group is defined as an Engset system equivalent to an egress azimuth. Considering a single primary group, the complete system is equivalent to an Engset system, where the number of sources and their calling rate are the ones from primary group; and where the number of channels is equal to the sum of channels from the primary group with the number of channels of the overflow system.

If there are several primary groups, a method gives an equivalent primary group. This method is similar to the standard Equivalent Random Theory (ERT) method from the Erlang framework [70].

Fig. 3.15 sums up the problem for our architecture. We thus have to evaluate the number of sources q^*, their traffic intensity a^*, and the number of channels n_c^* of the equivalent primary group. However, we notice that this equivalent primary group has been evaluated in the previous Section. The parameters thus are: $q^* = (n_a - 1)n_c$ sources, with a traffic intensity $a^* = \frac{\rho}{n_a - 1}$, and a number of channels $n_c^* = n_c$. Thanks to these equivalent parameters, we obtain the following closed-form expression for the blocking

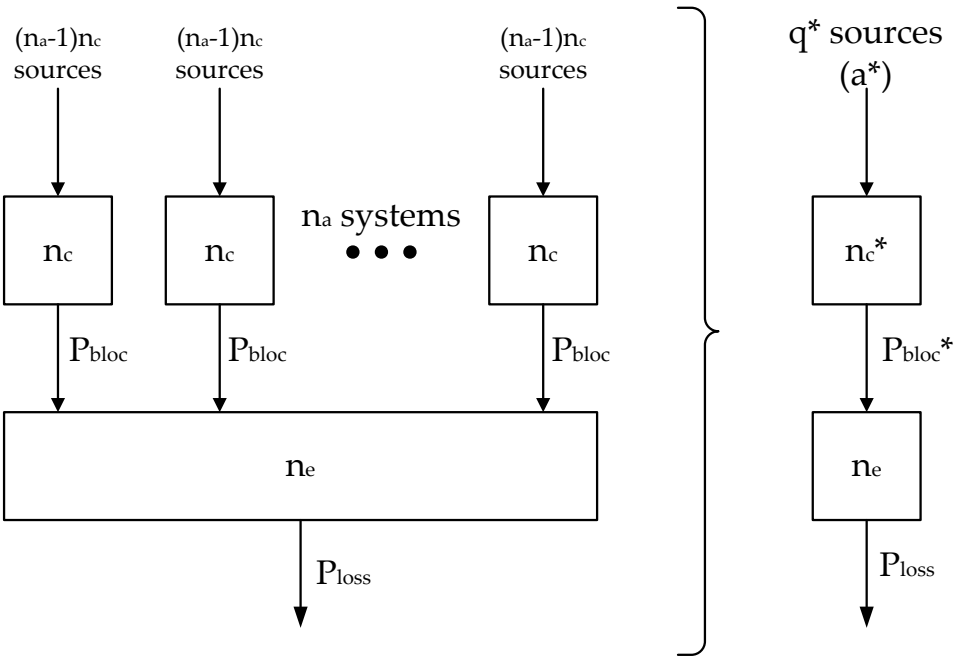

Figure 3.15: Equivalent architecture of the optical burst switch

probability of the complete system:

$$P_{loss}(n_a, n_c, n_e, \rho) = \frac{\begin{pmatrix} (n_a-1)n_c - 1 \\ (n_c + n_e) \end{pmatrix} \beta^{(n_c+n_e)}}{1 + \sum_{i=1}^{n_c+n_e} \begin{pmatrix} (n_a-1)n_c - 1 \\ i \end{pmatrix} \beta^i}, \tag{3.6}$$

again with $\beta = \left(\frac{n_a-1}{\rho} - 1\right)^{-1}$ as in equation (3.5), to which equation (3.6) is manifestly equivalent when $n_e = 0$.

3.4.2.2 Numerical illustrations

Once again, we evaluate the accuracy of our closed-form expression by comparing it to simulations proposed in Section 3.3. In Fig. 3.16, we plot the theoretical and the simulated burst loss probability versus the system load for different configurations: $n_a = 5$ azimuths on the bottom or $n_a = 10$ azimuths on the top. We also consider $n_c = 10$ channels per azimuth, and $n_e = 5, 10, 15, 20$ ports on the overflow system.

We remark that a large gap exists between the analytical results and the simulated one. Moreover, it grows with the system load and with the number of electronic ports. It can be explained by the second order collisions in the simulation which are not taken into account in the theoretical model. In other words, the analysis considers that our hybrid switch is able to predict them and so to avoid them. Therefore, it is of interest to add an

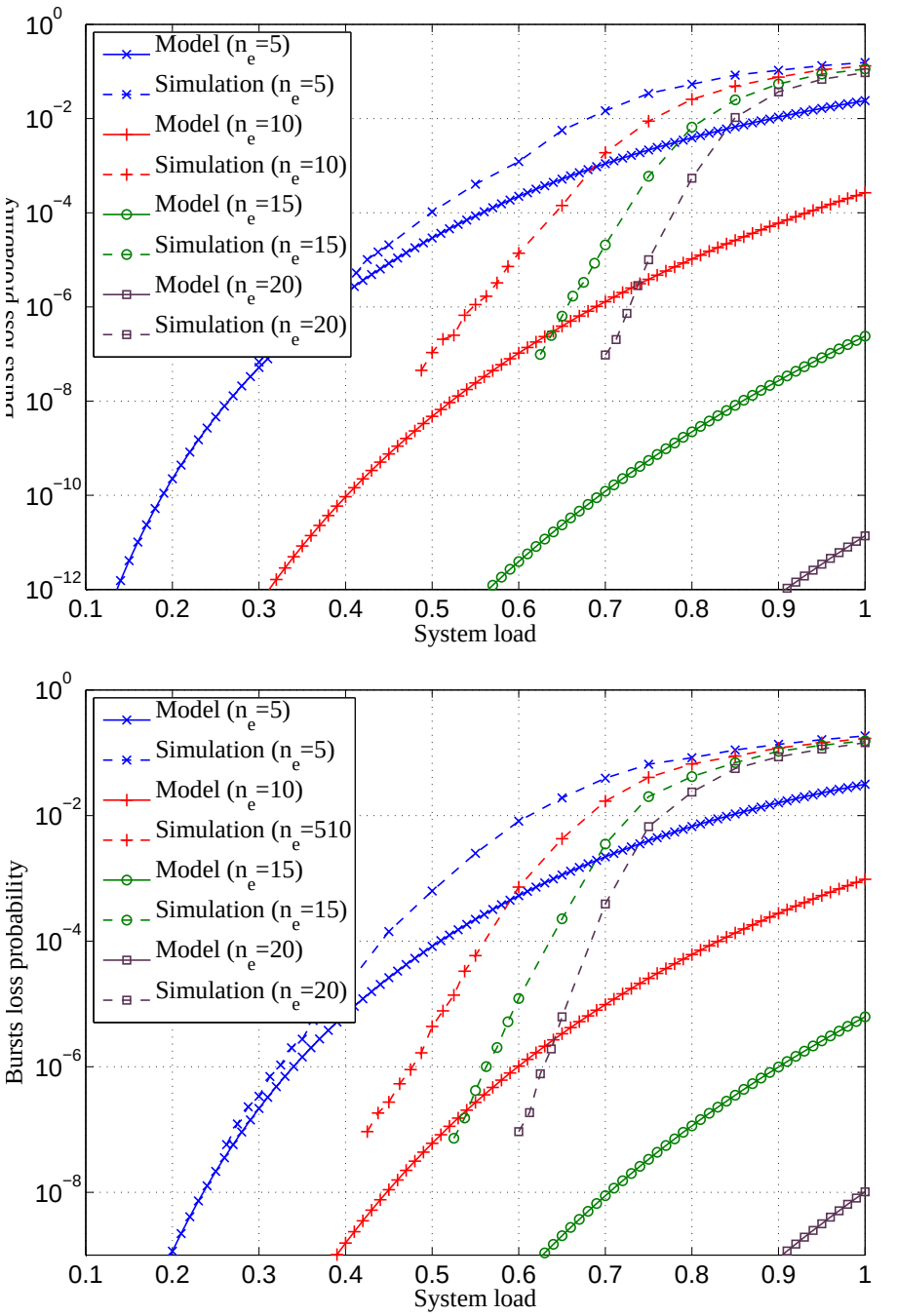

Figure 3.16: Analytical and simulated burst loss probability of the hybrid switch versus system load for $n_a = 5$ azimuths (Top) and $n_a = 10$ azimuths (Down), $n_c = 10$ channels and different number of electronic ports.

anti-collision system in the re-sending process, in order to further improve performance of the hybrid switch.

3.5 Conclusion

Since it decreases global cost and improves the flexibility, optical packet/burst switching is a key enabling technique for the future transport network. Sadly, the lack of all-optical buffers results in these techniques having a big packet/burst loss issue, especially when few channels are available per azimuth. As multiplying these channels may not be economically feasible, it seems better to look into buffering solutions.

In this chapter, we consider an optical burst switch assisted by an electronic buffer. We analyze, by simulation, the performance in term of burst loss probability of such an equipment. Simulation results lead to the identification of three key parameters for the performance improvement: the number of channels per azimuth, the number of electronics ports of the buffer, and the observation duration of the system against second order collisions.

The electronic buffer enables a great increase of the sustainable load even with a small number of overflow ports. Its combination with extra-channels on egress azimuth improves even more the performance.

In addition, we shown that the second order collisions are responsible of the main part of burst losses. Therefore, a well designed observation window before burst re-emitting, will significantly improves the overall system performance, and the closed-form expressions show us the maximum gain we can expect.

Finally, such a hybrid equipment may enable significant energy savings. Indeed, an network of hybrid switches is less energy consuming than a standard two-layered network. And a system which limits the second order collisions, may even more decrease this consumption. Main issue of hybrid techniques also remains the financial and/or energetic cost of extra-channels per azimuth.

Conclusions and Perspectives

The work carried out in this thesis deals with the analysis of coherent optical transmission systems, which will be one of the key techniques for future high data-rate networks. Since commercial systems able to transmit 100 Gbit/s over thousands of kilometers now exist, the main objective of the work was to find theoretical insights for the design of future systems. Indeed, the use of more powerful modulation schemes, and digital signal processing techniques, forces system suppliers to improve their theoretical knowledge of the optical channel. But, the need of energy-efficient, transparent and flexible systems also obliges them to combine these techniques with cross-layer solutions, in order to improve the efficiency of networks. That is the reason why we have proposed in the thesis, not only theoretical analysis of the propagation impairments in fiber, but also detailed studies of hybrid switching techniques.

Chapter 1 has been dedicated to the study using the Shannon capacity of the main non-linear impairment in optical fiber transmissions: the Kerr Effect. The data-rate increase often induces an increase of the input power per channel. Therefore non-linearities, which could be neglected until now, have to be considered. Moreover, the guard band between channels is reduced in order to improve the overall spectral efficiency, and it seems to us that studying its impact on system performance is relevant. The first part of the chapter was dedicated to an extended state-of-the-art about non-linear Shannon capacity. It led to the choice of a framework used for derivations of closed-form expressions of the optical channel capacity taking guard-band into account. The second part of this Chapter has been devoted to express the capacity in closed-form. Numerical illustrations of these expressions have shown that guard-band has a slight influence on the capacity.

Since another way to improve the achievable data-rate is to use all the freedom degrees of fibers, the impairments related to the polarization such as PMD and PDL, have now to be taken into account in the design. In the same spirit as previously, after a general overview of the PDL impaired channel model, Chapter 2 focused on the derivations of the outage probability due to PDL for different PDL models. Notice that the model chosen in the derivations only takes the polarization mismatch into account, since the other issues due to PDL can firstly be neglected, as shown in the chapter. Comparisons of the closed-form expressions with phenomenological simulations shows that a model (the Maxwellian one) is especially relevant with this respect. In addition, comparisons

between the theoretical results and simulated systems have been also achieved. Indeed, the outage probability is a bound of the achievable BER of practical transmission systems. Simulations of practical systems have been done with a powerful FEC, several modulations schemes and Polarization-Time codes. Simulation results confirm that PT codes are essential for the PDL mitigation. However, their combination with a well-designed FEC improves significantly the performance. At standard values of the mean PDL, the combination of a PT code with a powerful soft-decoded FEC is only 1.5 or 2 dB of the outage probability. Unfortunately, this powerful combination induces a high computational load which still prevents its use to practical systems.

Another way to improve the performance of optical transmission systems has been proposed in Chapter 3 with cross-layer techniques: the separation between the transport layer, and the network layer was a good idea when the network layer was heterogeneous. However, IP traffic nowadays is the main user of the transport network and, will eventually become the only one. Therefore, real synergies exist between the IP network layer and the optical transport layer. However all-optical techniques have many drawbacks. Therefore, the general architecture of a cross-layer equipment, also called optical burst switch with electronic buffer, has been proposed and studied. Numerical evaluations show the interest of such an equipment, because it significantly improves the performance in term of bursts loss probability, even for a small number of access ports on the buffer. Comparisons between simulations and the analytic model show that the performance can still be improved by the implementation of a system which reduces the number of collisions between re-sent bursts and ingress bursts (also called second-order collisions). Finally, this kind of solutions not only improves the network transparency and flexibility but also may allow significant energy savings by reducing the number of opto-electronic conversions.

This thesis has raised several issues which deserve to be treated in the future. These problems are listed below:

- The closed-form expressions in Chapter 1 have been established under the single polarization and the single span assumptions. Further works may focus on extending the expressions in a more general setting.
- The general PDL model used for the derivations in Chapter 2 takes only the polarization mismatch into account. Although this model is accurate for standard values of the mean PDL, a more advanced model, which considers the SNR compression and/or the noise correlation, needs to be develop, and then to be applied for the outage probability derivations in order to have a global view of the PDL phenomenon.
- Moreover, numerical illustrations in Chapter 2 show a gap of about 1.5 even 2 dB between the simulated BER and the outage probability. Further works should be to

find better coding techniques in order to reduce this gap.

- The analytic model of Chapter 3 can be improved in order to take the second-order collisions into account. It will enable a faster gauging of the whole equipment.
- The study about the hybrid switch and the related energy savings can be extended by *e.g.* taking interest in the λ converters issue in the case of WDM channels.

Appendix A

Proofs of Results 1, 2, and 3

A.1 Proof for Result 1

We consider $X(t)$ a Gaussian distributed random variable which has a PSD denoted $S_X(f)$ and an auto-correlation function denoted $R_X(t)$. $X(f)$ denotes the Fourier transform of $X(t)$. Let us derive $\mathbb{E}[X(f)X^*(f')]$.

$$\mathbb{E}[X(f)X^*(f')] = \mathbb{E}\left[\int_{-\infty}^{\infty} X(t)e^{-2i\pi ft}dt \int_{-\infty}^{\infty} X^*(t')e^{2i\pi f't'}dt'\right].$$

By permuting the expectatino and the integrals and then by applying the expectatino only on the random variables, we have

$$\mathbb{E}[X(f)X^*(f')] = \iint \mathbb{E}[X(t)X^*(t')]\, e^{-2i\pi ft}e^{2i\pi f't'}dtdt'$$

and so

$$\mathbb{E}[X(f)X^*(f')] = \iint R_X(t-t')e^{-2i\pi ft}e^{2i\pi f't'}dtdt'.$$

We will first integrate with respect to t. We thus have

$$\mathbb{E}[X(f)X^*(f')] = \int \left(\int R_X(t-t')e^{-2i\pi t}dt\right)e^{2i\pi f't'}dt'$$

which simplifies as follows

$$\mathbb{E}[X(f)X^*(f')] = \int S_X(f)e^{-2i\pi ft'}e^{2i\pi f't'}dt'.$$

According to the property of the Dirac distribution, we obtain

$$\mathbb{E}[X(f)X^*(f')] = \delta(f-f')S_X(f),$$

where $x \mapsto \delta(x)$ is the Dirac delta distribution.

A.2 Proof for Result 2

Let us denote

$$R_{a,b}(\tau) = \mathbb{E}[b(t+\tau)a^*(t)].$$

As

$$a(t) = \int A(f)e^{2i\pi ft}df$$

and

$$b(t) = \int B(f)e^{2i\pi ft}df,$$

we have

$$R_{a,b}(\tau) = \int\int \mathbb{E}[A(f)^*B(f')]e^{-2i\pi ft}e^{2i\pi f'(t+\tau)}dfdf'.$$

Moreover, since $A(f) = H_a(f)X(f)$ and $B(f) = H_b(f)$, we obtain

$$R_{a,b}(\tau) = \int\int H_a(f)^*H_b(f')\mathbb{E}[X(f)^*X(f')]e^{-2i\pi ft}e^{2i\pi f'(t+\tau)}dfdf'.$$

Thanks to Result 1, we simplify as follows

$$R_{a,b}(\tau) = \int H_a(f)^*H_b(f)S_X(f)e^{2i\pi f\tau}df$$

which implies that

$$S_{a,b}(f) = H_a(f)^*H_b(f)S_X(f).$$

A.3 Proof for Result 3

It is easy to check that

$$R_{a,b}(\tau) = R_{b,a}(-\tau)^*.$$

As a consequence, we obtain that

$$S_{a,b}(f) = S_{b,a}(f)^*.$$

Appendix B

Definition of Signal-to-Noise Ratio

In digital communication [31], the signal to noise ratio (SNR) is defined as

$$\text{SNR} = \frac{E_b}{N_0} \tag{B.1}$$

where

- E_b represents the consumed energy (at the carrier level) per information bit,
- N_0 is the noise variance per real dimension.

Notice that we consider that if $x(t)$ is the signal at the carrier and $x_e(t)$ is its corresponding complex envelope, then $x(t) = \Re[x_e(t)e^{2i\pi f_0 t}]$ with f_0 the carrier frequency.

According to the used practical transmission scheme, the term E_b can be expressed in different way.

Let E_s be the energy needed for transmitting one symbol (in baseband). We obviously have that

$$E_b = \frac{E_s}{2m}$$

where m is the average number of bits per symbol.

Now, it is easy to check that

$$m = k_c r_{\text{FEC}}(r_{\text{PT}}/n_t)$$

where

- $k_c = \log_2(M)$ is the number of bits per symbol when a constellation of size M is employed
- r_{FEC} the FEC coding rate.
- r_{PT}/n_t corresponds to the number of symbols per polarization and per channel use where r_{PT} is the PT code rate and n_t the number of inputs in the considered MIMO system.

Finally, we have

$$\mathrm{SNR} = \frac{n_t E_s}{2 k_c r_{\mathrm{FEC}} r_{\mathrm{PT}} N_0}. \tag{B.2}$$

and in our case (since $n_t = 2$),

$$\mathrm{SNR} = \frac{E_s}{k_c r_{\mathrm{FEC}} r_{\mathrm{PT}} N_0}. \tag{B.3}$$

Notice that in the expression of the outage probability, the term ρ is actually equal to

$$\rho = \frac{E_s}{2 N_0}$$

Bibliography

[1] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–423, 1948. Cited pages 6 and 8

[2] M. Cover and A. Thomas, *Elements of information theory (Second Edition)*. Wiley, 2006. Cited pages 6 and 7

[3] G. Agrawal, *Nonlinear Fiber Optics (Third Edition)*. Academic Press, 2001. Cited pages 8, 9, 10, and 11

[4] S. Savory, "Digital filters for coherent optical receivers," *Optics Express*, vol. 16, pp. 804–817, 2008. Cited page 10

[5] M. Karlsson, "Probability density functions of the differential group delay in optical fiber communication systems," *IEEE/OSA Journal of Lightwave Technology*, vol. 19, no. 3, pp. 324–331, 2001. Cited page 11

[6] K. Turitsyn, S. Derevyanko, I. Yurkevich, and S. Turitsyn, "Information capacity of optical fiber channels with zero average dispersion," *Physical review letters*, vol. 91, no. 20, p. 203901, 2003. Cited pages 12 and 18

[7] P. Mitra and J. Stark, "Nonlinear limits to the information capacity of optical fiber communications," *Nature*, vol. 411, pp. 1027–1030, 2001. Cited pages 12 and 13

[8] J. Kahn and K. Ho, "Spectral efficiency limits and modulation/detection techniques for dwdm systems," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 10, no. 2, pp. 259–272, 2004. Cited page 13

[9] P. Poggiolini, A. Carena, V. Curri, G. Bosco, and F. Forghieri, "Analytical modeling of nonlinear propagation in uncompensated optical transmission links," *IEEE Photonics Technology Letters*, vol. 23, June 2011. Cited page 13

[10] G. Bosco, P. Poggiolini, V. Curri, and F. Forghieri, "Analytical results on channel capacity in uncompensated optical links with coherent detection," *Optics Express*, vol. 19, November 2011. Cited pages 13 and 14

[11] J. Tang, "The shannon channel capacity of dispersion-free nonlinear optical fiber transmission," *IEEE/OSA Journal of Lightwave Technology*, vol. 19, August 2001. Cited pages 14, 15, 18, 19, 21, and 24

[12] J. Tang, "The multispan effects of kerr nonlinearity and amplifier noises on shannon channel capacity of a dispersion-free nonlinear optical fiber," *IEEE/OSA Journal of Lightwave Technology*, vol. 19, no. 8, 2001. Cited pages 14, 15, and 19

[13] J. Tang, "The channel capacity of a multispan dwdm system employing dispersive nonlinear optical fibers and an ideal coherent optical receiver," *IEEE/OSA Journal of Lightwave Technology*, vol. 20, no. 7, 2002. Cited pages 14, 15, 16, 19, 24, 25, 31, 33, 34, and 35

[14] J. Tang, "A comparison study of the shannon channel capacity of various nonlinear optical fibers," *IEEE/OSA Journal of Lightwave Technology*, vol. 24, no. 5, 2006. Cited pages 14, 17, and 19

[15] M. Pinsker, "Information and information stability of random variables and processes," *Holden Bay (San Francisco, CA)*, pp. 160–201, 1964. Cited page 14

[16] K. Peddanarappagari and M. Brandt-Pearce, "Volterra series transfer function of single-mode fibers," *IEEE/OSA Journal of Lightwave Technology*, vol. 15, no. 12, pp. 2232–2241, 1997. Cited page 15

[17] E. Narimanov and P. Mitra, "The channel capacity of a fiber optics communication system: Perturbation theory," *IEEE/OSA Journal of Lightwave Technology*, vol. 20, no. 3, p. 530, 2002. Cited page 17

[18] L. Xiang and X. Zhang, "The study of information capacity in multispan nonlinear optical fiber communication systems using a developed perturbation technique," *IEEE/OSA Journal of Lightwave Technology*, vol. 29, no. 3, pp. 260–264, 2011. Cited pages 17 and 18

[19] E. Agrell, "The channel capacity increases with power," *arXiv:1108.0391v2*, 2011. Cited page 18

[20] R. Essiambre, G. Foschini, G. Kramer, and P. Winzer, "Capacity limits of information transport in fiber-optic networks," *Physical Review Letters*, vol. 101, no. 16, p. 163901, 2008. Cited page 18

[21] R. Essiambre, G. Kramer, P. Winzer, G. Foschini, and B. Goebel, "Capacity limits of optical fiber networks," *IEEE/OSA Journal of Lightwave Technology*, vol. 28, no. 4, pp. 662–701, 2010. Cited page 18

[22] I. Djordjevic, "Ultimate information capacity of fiber optic network," in *SPIE Photonic West*, no. 4, 2010. Cited page 18

[23] B. Porat, *Digital Processing of Random Signals*. PRENTICE-HALL, INC., 1993. Cited pages 27 and 29

[24] A. ElAmari, 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," *IEEE/OSA Journal of Lightwave Technology*, vol. 16, no. 3, pp. 332–339, 1998. Cited pages 39 and 47

[25] P. Lu, L. Chen, and X. Bao, "Statistical distribution of polarization dependent loss in the presence of polarization mode dispersion in single mode fibers," *IEEE Photonics Technology Letters*, vol. 13, pp. 451–453, 2001. Cited pages 39, 52, and 53

[26] A. Mecozzi and M. Shtaif, "The Statistics of Polarization-Dependent Loss in Optical Communication Systems," *IEEE Photonics Technology Letters*, vol. 14, pp. 313–315, March 2002. Cited pages 39, 53, and 56

[27] M. Shtaif, "Performance degradation in coherent polarization multiplexed systems as a result of polarization dependent loss," *Optics Express*, vol. 16, pp. 13918–13932, 2008. Cited pages 39 and 41

[28] F. A. C. Garcia, D. A. A. Mello, and H. Waldman, "Feedforward carrier recovery for polarization demultiplexed signals with unequal signal to noise ratios," *Optics Express*, vol. 17, pp. 7958–7969, 2009. Cited pages 39, 44, and 46

[29] D. S. Waddy, L. Chen, and X. Bao, "Polarization effects in aerial fibers," *Optical Fiber Technology*, vol. 11, pp. 1–19, 2005. Cited page 44

[30] E. Meron, A. Andrusier, M. Feder, and M. Shtaif, "Use of space-time coding in coherent polarization-multiplexed systems suffering from polarization-dependent loss," *Optics Letters*, vol. 35, no. 21, pp. 3547–3549, 2010. Cited pages 44, 59, 62, and 64

[31] D. Tse and P. Viswanath, *Fundamentals of Wireless Communications*. Cambridge Press, 2005. Cited pages 44, 45, 61, and 103

[32] I. E. Telatar, "Capacity of multi-antenna gaussian channels," *AT&T Technical Memorandum*, 1995. Cited page 45

[33] S. Mumtaz, G. Rekaya, and Y. Jaouën, "PDL mitigation in PolMux OFDM systems using Golden and Silver Polarization-Time codes," in *Optical Fiber Communication (OFC)*, 2010. Cited pages 46, 59, 62, and 64

[34] N. Gisin, "Statistics of polarization dependent losses," *Optics Communications*, vol. 114, pp. 399–405, 1995. Cited page 47

[35] J. Jiang, D. Richards, S. Oliva, P. Green, and R. Hui, "In-situ monitoring of PMD and PDL in a traffic-carrying transatlantic fiber-optic System," in *Optical Fiber Communication (OFC)*, 2009. Cited page 50

[36] S. Mumtaz, *Modern Coding Techniques for Optical Fiber Communications*. PhD thesis, Telecom ParisTech, 2011. Cited pages 51 and 61

[37] D. MacKay and R. Neal, "Near shannon limit performance of low density parity check codes," *Electronics letters*, vol. 33, no. 6, pp. 457–458, 1997. Cited pages 59 and 60

[38] W. Ryan and S. Lin, *Channel Codes: Classical and Modern*. Cambridge Press, 2009. Cited pages 59 and 61

[39] S. Johnson and S. Weller, "Practical Interleavers for Systematic Repeat-Accumulate Codes," in *Vehicular Technology Conference (VTC)*, vol. 3, pp. 1358–1362, IEEE, 2006. Cited page 59

[40] S. Mumtaz, G. Rekaya, and Y. Jaouën, "Space-Time Codes for Optical Fiber Communication with Polarization Multiplexing," in *International Conference on Communications (ICC)*, IEEE, 2010. Cited pages 59, 62, and 64

[41] E. Meron, A. Andrusier, M. Feder, and M. Shtaif, "Increasing the PDL tolerance of systems by use of the Golden-Code," in *European Conference on Optical Communication (ECOC)*, 2010. Cited pages 59, 62, and 64

[42] R. Gallager, "Low-density parity-check codes," *IEEE Transactions on Information Theory*, vol. 8, no. 1, pp. 21–28, 1962. Cited pages 60 and 61

[43] V. Tarokh, N. Seshadri, and A. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Transactions on Information Theory*, vol. 44, no. 2, pp. 744–765, 1998. Cited page 61

[44] S. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, pp. 1451–1458, 1998. Cited page 61

[45] J. C. Belfiore, G. Rekaya, and E. Viterbo, "The golden code: a 2× 2 full-rate space-time code with nonvanishing determinants," *IEEE Transactions on Information Theory*, vol. 51, no. 4, pp. 1432–1436, 2005. Cited page 61

[46] A. Tirkkonen and A. Hottinen, "Improved mimo performance with non-orthogonal space-time block codes," in *Global Telecommunications Conference (GLOBECOM)*, vol. 2, pp. 1122–1126, IEEE, 2001. Cited page 61

[47] IEEE Std 802.3ba™-2010, *Carrier Sense Multiple Access With Collision Detection (CSMA/CD) Access Method and Physical Layer Specifications*. IEEE, 2010. Cited page 73

[48] M. Gupta and S. Singh, "Greening of the internet," in *Conference on applications, technologies, architectures, and protocols for computer communications (SIGCOMM)*, vol. 33, pp. 19–26, ACM, 2003. Cited page 73

[49] S. Yoo, "Energy efficiency in the future internet: the role of optical packet switching and optical-label switching," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 17, no. 2, pp. 406–418, 2011. Cited page 73

[50] Z. Rosberg, H. Le Vu, M. Zukerman, and J. White, "Performance analyses of optical burst-switching networks," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 7, pp. 1187–1197, 2003. Cited pages 73 and 77

[51] T. Zhang, K. Lu, and J. Jue, "Shared fiber delay line buffers in asynchronous optical packet switches," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 4, pp. 118–127, 2006. Cited pages 74, 78, and 93

[52] Y. Liu, M. Hill, R. Geldenhuys, N. Calabretta, H. de Waardt, G. Khoe, and H. Dorren, "Demonstration of a variable optical delay for a recirculating buffer by using all-optical signal processing," *IEEE Photonics Technology Letters*, vol. 16, no. 7, pp. 1748–1750, 2004. Cited pages 74 and 78

[53] R. S. Tucker, P. Ku, and C. J. Chang-Hasnain, "Slow-Light Optical Buffers: Capabilities and Fundamental Limitations," *IEEE/OSA Journal of Lightwave Technology*, vol. 23, pp. 4046–4066, 2005. Cited page 74

[54] D. Awduche and Y. Rekhter, "Multiprotocol lambda switching: combining MPLS traffic engineering control with optical crossconnects," *IEEE Communications Magazine*, vol. 39, no. 3, pp. 111–116, 2001. Cited page 75

[55] P. Roorda and B. Collings, "Evolution to colorless and directionless ROADM architectures," in *Optical Fiber communication/National Fiber Optic Engineers Conference (OFC/NFOEC)*, IEEE, 2008. Cited page 75

[56] E. Le Rouzic, N. Brochier, S. Blouza, G. Froc, C. Mangin, M. Morvan, P. Gravey, J. Jauffrit, J. M., and M. van der Keur, "On the Interest of Multi-Band OFDM in backhaul networks," in *European Conference on Optical Communications (ECOC)*, no. We.2.D.4, 2012. Cited page 75

[57] A. Farrel and I. Bryskin, *GMPLS: architecture and applications*. Morgan Kaufmann Pub, 2006. Cited page 76

[58] S. Azodolmolky, R. Nejabati, E. Escalona, R. Jayakumar, N. Efstathiou, and D. Simeonidou, "Integrated OpenFlow–GMPLS control plane: an overlay model for software defined packet over optical networks," *Optics Express*, vol. 19, no. 26, pp. B421–B428, 2011. Cited page 76

[59] R. Takahashi, T. Nakahara, K. Takahata, H. Takenouchi, T. Yasui, N. Kondo, and H. Suzuki, "Ultrafast optoelectronic packet processing for asynchronous, optical-packet-switched networks [invited]," *Journal of Optical Networking*, vol. 3, no. 12, pp. 914–930, 2004. Cited pages 77 and 78

[60] E. Almstrom, C. Larsen, L. Gillner, W. van Berlo, M. Gustavsson, and E. Berglind, "Experimental and analytical evaluation of packaged 4×4 InGaAsP/InP semiconductor optical amplifier gate switch matrices for optical networks," *IEEE/OSA Journal of Lightwave Technology*, vol. 14, pp. 996–1004, June 1996. Cited page 77

[61] C. Lai, D. Brunina, C. Ware, B. Bathula, and K. Bergman, "Demonstration of failure reconfiguration via cross-layer enabled optical switching fabrics," *IEEE Photonics Technology Letters*, vol. 23, no. 22, pp. 1679–1681, 2011. Cited page 77

[62] S. Danielsen, B. Mikkelsen, C. Joergensen, T. Durhuus, and K. Stubkjaer, "WDM packet switch architectures and analysis of the influence of tunable wavelength converters on the performance," *IEEE/OSA Journal of Lightwave Technology*, vol. 15, pp. 219–227, Feb. 1997. Cited page 77

[63] D. Chiaroni, R. Urata, J. Gripp, J. E. Simsarian, G. Austin, S. Etienne, T. Segawa, Y. Pointurier, C. Simonneau, Y. Suzaki, *et al.*, "Demonstration of the interconnection of two optical packet rings with a hybrid optoelectronic packet router," in *European Conference on Optical Communications (ECOC)*, no. PD3.5, 2010. Cited pages 77 and 78

[64] J. Zhang, E. Wong, and M. Zukerman, "Modeling an OBS node under critical load and high utilization conditions," *IEEE Communications Letters*, vol. 16, no. 4, pp. 544–546, 2012. Cited pages 77, 81, and 90

[65] Y. Yin, R. Proietti, X. Ye, R. Yu, V. Akella, and S. Yoo, "Experimental Demonstration of LIONS: A Low Latency Optical Switch for High Performance Computing," in *International Conference on Transparent Optical Networks (ICTON)*, no. We-S31-O02, 2012. Cited pages 78 and 88

[66] T. Bonald and M. Feuillet, *Network performance analysis*. Wiley ed., 2011. Cited pages 79, 90, and 91

[67] H. Overby, "Performance modelling of optical packet switched networks with the Engset traffic model," *Optics express*, vol. 13, no. 5, pp. 1685–1695, 2005. Cited page 90

[68] M. Yasuji, "An Approximation for Blocking Probabilities and Delays of Optical Buffer With General Packet-Length Distributions," *IEEE/OSA Journal of Lightwave Technology*, vol. 30, no. 1, pp. 54–60, 2012. Cited page 93

[69] R. Schehrer, "On the calculation of overflow systems with a finite number of sources and full available groups," *IEEE Transactions on Communications*, vol. 26, no. 1, pp. 75–82, 1978. Cited page 93

[70] M. Glabowski, "Modeling Systems with Multi-service Overflow Erlang and Engset Traffic Streams," *International Journal On Advances in Telecommunications*, vol. 1, no. 1, pp. 14–26, 2008. Cited page 93