New Coding Techniques for High Bit-Rate Optical Transmission Systems

Sami Mumtaz

ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY

TELECOM ParisTech

2012-ENST

EDITE - ED 130

Doctorat ParisTech

T H È S E

pour obtenir le grade de docteur délivré par

TELECOM ParisTech

Spécialité: Communications et Électronique

présentée et soutenue publiquement par

Sami MUMTAZ

le 31 Janvier 2011

Nouvelles Techniques de Codage pour les Communications Optiques à Haut-Débit

Directeur de thèse: **Yves JAOUËN**

Co-encadrement de la thèse: **Ghaya REKAYA BEN-OTHMAN**

Jury

M. Merouane DEBBAH, Supelec	**Président**
M. Ezio BIGLIERI, Politecnico di Turino	**Rapporteurs**
M. Alberto BONONI, Università degli Studi di Parma	
M. Gabriel Charlet, Alcatel-Lucent Bell Labs France	**Examinateurs**
M. Mark SHTAIF, Tel Aviv University	
M. Yves JAOUËN, Telecom ParisTech	
Mme Ghaya REKAYA BEN-OTHMAN, Telecom ParisTech	
M. Michel Morvan, Telecom Bretagne	**Invité**

THÈSE

Á mes Parents,

Contents

Introduction

According to the CISCO VNI report of June 2010 [1], the global IP traffic has grown by 45% in 2009 to reach a rate of 176 exabytes (EB). The traffic will be quadrupled in next four years and represent 767 EB in 2014. This corresponds to a compound annual growth rate of 34%. To support this massive need of resources, the global communication network has to be continuously developed and improved. For instance, the capacity of the trans-atlantic network is 53Tb/s (in 2010) but only 13Tb/s is actually lit. However at this traffic growing rate, the limit capacity will be reached in 2014 [2].

Today, the vast majority of the communications (IP, voice...) are carried by optical transmissions. In particular, the transcontinental transmissions are quasi exclusively made on optical fibers and spatial transmissions are now negligible. Indeed, optical fiber is a perfect media for transmissions because of its very slow attenuation. Since the invention of optical amplifiers in 1987, transmissions over very large distances ($\geq$ 3000km) become possible. Hence, underseas and terrestrial optical cables have been deployed all over the world to ensure the huge data traffic between continents. This is called the long-haul core network. The optical network is also composed by smaller scale networks connected to the core network, that provide the data transport in a metropolitan or country area ($100 - 1000$km). Recently the optical network has been extended to get closer the user and fibers are progressively replacing copper networks. Fiber technologies will soon directly interconnected all the end-user applications (wireless network, storage network...) to the global network.

In the following years, the critical issue will be for the core network to support the traffic growth. Actual deployed links have a maximum capacity of $0.5 - 1$Tb/s per fibers. These rates are obtained by multiplexing 10 and 40Gb/s single carrier transmissions in a single fiber. In practice, up to 100 channels can be multiplexed. This is known as wavelength division multiplexing and has allowed to impressively expand the network capacity and thus reduce the cost of these technologies. To cope with the demand of capacity, many new underseas and terrestrial optical transmissions will be deployed in the following years. However increasing the number of links can not be enough (but at an unacceptable cost) if at the same time their capacities are not increased too.

In a short term, the capacity can be easily increased by upgrading the actual 10Gb/s and

40Gb/s systems to the recently available 100Gb/s systems. The principal evolution in these systems compared to the ones at 10Gb/s, is the major role of digital signal processing at the reception. Indeed during the last decade, intense investigations have concerned digital equalization in order to compensate optical impairments entirely in the electrical domain. 100Gb/s transmissions may also represent a transition between direct detection and coherent detection systems. Indeed coherent detection have made an outstanding come back thanks to digital signal processing. In addition to the higher receiver sensitivity, it ease the use of high spectral efficiency modulation formats and polarization multiplexing techniques. The first 100Gb/s coherent system has been proposed by Alcatel in 2010 and it is very likely that future generations of systems at 400Gb/s will all be based on this concept.

In a mid term, 400Gb/s systems are target. To reach these kind of bit-rates, a spectral efficiency increase is necessary. If the use of multi-level modulation formats such as 16-QAM is quite certain, orthogonal frequency division multiplexing may also be an interesting alternative to single carrier transmission. In the next $1-10$ years, the upgrade into $100-400$Gb/s systems can be realized by only replacing the transmitters and the receivers of actual systems. Indeed coherent detection, polarization multiplexing and orthogonal frequency division multiplexing techniques are perfectly compatible with existing fiber links.

Will these techniques be enough in 10 years to cope with the traffic demand? The answer is certainly no. A major technical breakthrough, as optical amplifier was, as wavelength division multiplexing was, is expected and required in order to keep increasing the capacity of the global network.
A promising solution that has begun to focus lot of interests, is multi-mode fiber transmissions. Indeed mode multiplexing can theoretically multiplied the capacity by the number of transmitted modes and optical fibers can support the transmission of one to hundreds modes simultaneously.

Motivations and contributions

The goal of this thesis is to propose innovative digital signal processing solutions to improve actual and future generations of optical transmission systems. Our investigation has focused particularly on forward error correction and space-time coding.

This thesis began in 2007 in the frame of the ANR (Agence Nationale pour la Recherche) project TCHATER in collaboration with Alcatel-Lucent France, E2v, ARENAIRE, and IRISA. The project goal was to develop a 40Gb/s coherent receiver on FPGAs. Our contribution in this project concerns the study of the implementation of soft decision

forward error correction.
Every optical systems use in practice a forward error correction (FEC) scheme in order to correct the transmission errors at the receiver and improve the quality of the transmission. However the FEC implemented in recent systems are far from being the state of the art. Indeed, the modern FEC (low-density parity check code or Turbo code...) major drawback is to have an excessively complex decoding. However with the development of high speed electronics and the come back of coherent detection, modern FEC implementation becomes conceivable.
The motivation of my work is to choose the most appropriate FEC for high bit-rate optical systems. The selection is made based on the decoding complexity which is here the main issue. We have proposed an original construction of low-density parity-check (LDPC) code suitable for high bit-rate implementations and that have good performance compared to the codes proposed in the literature

To improve the performances of the FEC, it is very important to consider the specificities of the channel. In particular, coherent and direct detection systems both require differential modulation and this degrades the performance. We have proposed a new structured interleaving of the FEC codewords with a corresponding decoding scheme, in order to reduce the penalties introduced by differential modulation. Moreover, the proposed scheme allows decoding complexity reduction and redundancy decrease without any performance loss.

Then, we have investigated the interest of space-time codes for optical transmission systems. They have been developed for MIMO wireless channels but can be employed in polarization multiplexed optical systems. However their implementation requires the use of optical orthogonal frequency multiplexing (OFDM). For the first time, we have shown that space-time coding can efficiently mitigate polarization dependent loss impairments. We have also shown that their performance is very different than in wireless transmission and explained the reason.

Finally, we have studied multi-mode fiber transmissions. We have investigated how to perform the equalization in such systems and unlike what is currently done in optical systems, we have proposed an equalization based on the channel estimation. This part has been realized in collaboration with R.J. Essiambre in the frame of a summer internship at Alcatel-Lucent Bell Lab's USA.

Thesis outline

Here is an brief overview of this thesis. In Chapter 1, we present an optical transmission system and describe the different technologies used in actual 100Gb/s systems. We also

discuss about the signal propagation in the fiber and the different impairments that occur.

Chapter 2 is dedicated to the forward error correction in optics. In a first part, the principles of FEC are reminded and a brief history of FEC in optical transmission systems is given. Then we investigate the decoding complexity of low-density parity-check codes and product codes. Some constructions based on algebraic objects proposed in the literature are presented before to introduce an original construction of LDPC codes.

In Chapter 3, we propose an original coding/decoding scheme to improve the performance of FEC in the case of differential modulation. We first present the problem of differential modulation and then describe our coding scheme called "structured symbol interleaving". We show that it can improve the performance and that an appropriate decoding can reduce the decoding complexity. In a last part, we show that the redundancy of the FEC can be also reduced without any performance degradation

In Chapter 4, we focus on space-time coding in OFDM systems. We first describe optical OFDM systems and then present the space-time coding principles. We discuss about the performance criteria in wireless and in optical systems. Finally we evaluate the performance of space-time code to combat polarization dependent loss.

Chapter 5 finally deals with the digital equalization in multi-mode fiber transmissions. We introduce the concept of mode multiplexing and discuss about the equalization options. We present the principle of equalization based on channel estimation and demonstrate its performance in that systems. Finally, we show that using OFDM can be profitable in order to reduce the decoding complexity

Notations

Notation	Definition	
$\|\| \cdot \|\|$	Frobenius norm	
$\lfloor \cdot \rfloor$	floor function	
$\lceil \cdot \rceil$	ceiling function	
$(\cdot)^T$	transpose conjugate	
$(\cdot)^*$	complex conjugate	
$(\cdot)^H$	Hermitian conjugate	
$\mathbb{E}\ \cdot$	Expectation	
GF	Galois field	
$\imath$	$\sqrt{-1}$	
$\mathrm{Im}(\cdot)$	Imaginary part	
lmc	least common multiple	
$\Pr\{\cdot\}$	Probability	
$\Pr\{\cdot\|\cdot\}$	Conditional probability	
$\mathrm{Re}(\cdot)$	Real part	
α	Primitive element of a Galois field	
α_a	Fiber loss coefficient	Nepper/km
α_{dB}	Fiber loss coefficient	dB/km
A_{eff}	Fiber effective area	m^2
$\beta(\omega)$	Fiber propagation constant	rad/m
β_1	Fiber group velocity inverse	ps/km
β_2	Fiber group velocity dispersion	ps^2/km
B_{ref}	Reference spectral bandwidth	nm
c	speed of the light in vacuum	m/s
C	Channel capacity	bit/s/Hz
$\mathcal{C}(\mathrm{n}, \mathrm{k})$	Forward error correction	
$\Delta\tau$	Differential group delay	$\mathrm{ps}/\sqrt{\mathrm{km}}$
d_{min}	minimal distance of a code	

Notation	Definition	
D	Dispersion coefficient	ps/nm/km
E_b	Energy per information bit	J
E_s	Energy per modulated symbol	J
E_s	Signal field complex envelope	
E_{c_m,v_k}	Edge in the Tanner graph	
$F(x,y)$	Field distribution	
$\hbar$	Planck constant	eV.s
γ_{NL}	Nonlinearity coefficient	$\mathrm{km}^{-1}\mathrm{W}^{-1}$
$\mathbf{\Gamma}_{dB}$	Polarization dependent loss	dB
$\boldsymbol{H}$	Parity-check matrix of a FEC	
I_Q,I_I	quadrature and in-phase current	A
k_c	Modulation spectral efficiency	
$\chi^{(3)}$	Silica 3^{rd} order suceptibility	
λ_c	Laser wavelength	m
L	Fiber length	m
L	CIR length	taps or T_s
L_{eff}	Effective length of the fiber	m
M	number of state in a constellation	
$\tilde{\mathbf{n}}(\omega)$	Linear fiber refractive index	
$\mathbf{n}(\omega)$	Fiber refractive index	
$\mathbf{n}_2$	Nonlinear fiber refractive index	m^2/W
n_{sp}	Spontaneous emission factor	W/Hz
$\mathrm{N}_{\mathrm{ASE}}$	ASE spectral density	W/Hz
N_0	Noise spectral density	W/Hz
ω	Pulsation (angular frequency)	rad/s
Φ	Phase	rad
$P(t)$	Average power	W
P_{out}	Outage probability	
$\mathfrak{r}$	Photodiode responsivity	A/W
r	FEC rate	
r_{ov}	Sampling rate	sample / s
r_{STBC}	STBC rank	
$\mathbf{R}$	Rotation Matrix	
R_b	Bit rate	bit/s
R_s	Symbol rate	symbol/s

Notation	Definition	
R_{STBC}	STBC rate	symb/cu
$synd$	Syndrome	
t_{max}	error correction capability of a FEC	
t	time	s
t_{max}	FEC correction capability	
T_s	Symbol duration	s
v_g	Group velocity	km/ps
V_0, V_1,V_2	Tension applied to the the MZM	V
V_π	MZ extinction tension	V

Acronyms

Notation	Definition
ASE	Amplified spontaneous emission
BCH	Bose Chaudhuri Hocquenghem
BER	Bit error rate
BPSK	Binary phase shift keying
CD	Chromatic dispersion
CMA	Constant modulus algorithm
CIDS	Cyclic invariant difference set
CIR	Channel impulse response
DGD	Differential group delay
DMD	Differential mode delay
DPSK	Differential phase shift keying
DSP	Digital signal processing
EDFA	Erbium doped fiber amplifier
FEC	Forward error correction
FFT/iFFT	(inverse) fast Fourier transform
FIR	Finite impulse response
FWM	Four-wave mixing
LDPC	Low-density parity-check
LLR	Log likelihood ratio
LMS	Least-mean square
LO	Local oscillator
LSE	Least square estimator
MGDM	Mode group diversity multiplexing
MIMO	Multiple input multiple output
MLSE	Maximum likelihood sequence estimation
MMF	Multi-mode fiber
MMSE	Minimum mean square error

Notation	Definition
MSA	Min-Sum algorithm
MZM	Mack-Zender modulator
NRZ	Non return to zero
OA	Orthogonal array
OFDM	Orthogonal frequency division multiplexing
OOK	On-off keying
OSNR	Optical signal to noise ratio
PDM	Polarization division multiplexing
PDL	Polarization dependent loss
PEG	Progressive edge growth
PMD	Polarization mode dispersion
PRBS	Pseudo random binary sequence
PT	Polarization-time
QAM	Quadrature amplitude modulation
QPSK	Quaternary phase shift keying
RS	Reed-Solomon
RZ	Return to zero
SMF	Single-mode fiber
SNR	Signal to noise ratio
SPA	Sum-Product algorithm
SPM	Self-phase modulation
SSI	Structured symbol interleaving
ST	Space-time
STBC	Space-time block code
XPM	Cross-phase modulation
XPolM	Cross polarization modulatio
WDM	Wavelength division multiplexing
ZF	Zero-forcing

Chapter 1

Principles of optical transmission systems

In this chapter, a description of optical transmission systems is given. We will present the different technologies used in the actual systems and focus more particularly on coherent detection, polarization multiplexing and digital equalization as they correspond to the state of the art of optical transmissions at 100Gb/s.

Introduction

In an optical transmission system the data is modulated and transmitted onto an optical carrier, usually issued from a laser. The signal propagates in a transparent media and is received and converted in the electrical domain by photodiodes. Depending on the application, transmission can be either realized in free-space (spatial communication...) or on optical fibers (submarine communications, metro, access...). The transmission distance may vary from few meters in access network to thousands of kilometers in long-haul transmissions. In this thesis, we focus particularly on long-haul optical fiber communications.

Optical fiber is a very advantageous waveguide for transmissions because of its relatively small attenuation compared to cables ($\sim$ 0.2dB/km) and its very wide spectral bandwidth ($\sim$ 100THz). In order to take advantage of the whole bandwidth of the fiber, modulated signals using different wavelengths are transmitted simultaneously. This technique is called wavelength division multiplexing (WDM) and has been the key technology during the last two decades to expand the network capacity and support the traffic growth. The other promising technology that has been intensively (re)-investigated during the last years and which has just been implemented in practical transmission system, is coherent detection. Coherent detection has a better sensitivity compared to direct detection but the main advantage is the digital signal processing that allows high spectral modulation formats and compensate more efficiently the transmission impairments. The come back of coherent

detection happened with the development of high-speed electronics making digital equalization an alternative to the optical phase-lock loop required in order to compensate the phase shift between the local oscillator and the optical carrier.

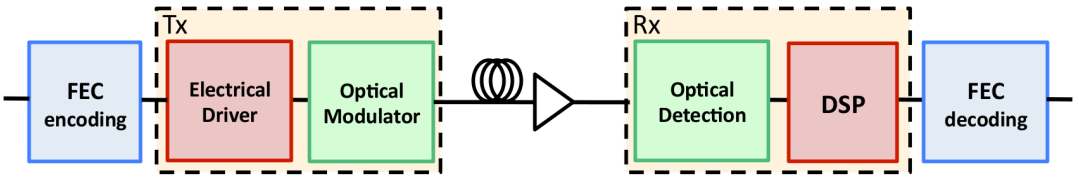


Figure 1.1: Schematic architecture of an optical transmission system

A general scheme of an optical transmission system is presented in Fig. 1.1. The transmitter is composed by an electrical driver and an optical modulator. The electrical driver input is a data bit sequence issued from a forward error correction encoder. It generates a signal that is converted in the optical domain by an electro-optical modulator. The signal is transmitted in the optical fiber link where periodically optical amplifiers are inserted to compensate the fiber attenuation. During the propagation, many transmission impairments occur that degrade the performance. The optical signal is received by an optical receiver whose structure depends on the modulation format and is converted back into the electrical domain by one or many photodiodes. Finally, digital signal processing is performed to compensate the transmission impairments and the estimated bit sequence is decoded by the forward error correction decoder.

In the following, we will describe every elements of an optical transmission system. In Sec. 1.1, we present the principle of the optical transmitter and the common modulation formats. Sec. 1.2 deals with the optical fiber propagation and give an overview of the different transmission impairments. Finally, in Sec. 1.3, we describe the structure of the optical receiver and discuss about the digital signal processing techniques employed in order to compensation the transmissions effects.

1.1 Optical Transmitter

The optical transmitter generates a modulated optical signal in function of an input bit sequence. The transmitter is composed of an electrical driver which converts the bit sequence into an electrical signal and of an optical modulator which modulated the optical carrier in function of the electrical driving signal. The transmitter output can have different states (phase, amplitude...), all associated to a certain number of input information bits.

1.1.1 Optical modulator

The optical modulator modifies the optical carrier issued from the laser in order to generate an output modulated signal:

$$E_c(t) = A_c(t)\exp(\omega_c t + \Phi_c(t)) \tag{1.1}$$

A_c is the amplitude related to the laser power, w_c is the carrier pulsation and Φ_c is the phase. It is more common to characterize the laser by its wavelength $\lambda_c = 2\pi/\omega_c$. In optical transmission systems, typical wavelengths are around 1.55μm.
We talk about amplitude modulation if the amplitude $A_c(t)$ of the optical signal is modified and phase modulation if it is the phase $\Phi_c(t)$. More advanced modulation formats such as 16-QAM, modulate at the same time amplitude and phase.
An optical modulator is characterized by:

- the electro-optical bandwidth corresponding to the limit frequency at which it keeps responding to the electrical driving signal
- the extinction ratio defined as the ratio between the maximum and minimum power it can deliver
- the insertion loss

Moreover it is expected not to be wavelength dependent and not to introduce frequency chirps.

Optical modulation can be realized by various techniques such as directly modulated laser, electro-absorption modulator and Mach-Zender modulator (MZM). MZM are the most popular device for high-speed transmissions because of their large bandwidth ($\sim$ 35GHz), their small insertion loss ($\sim$ 4dB) and their high extinction ratio ($\sim$ 20dB). Moreover they are weakly wavelength dependent and chirp free.

Mach-Zender modulator

The Mach-Zender modulator [3] is an interferometer composed by two 3dB couplers and two waveguides made of lithium niobate ($LiNbO_3$) crystal as shown in Fig. 1.2. This crystal displays Pockels effect which means that its refractive index is linearly proportional to the electric field. Hence, by applying a voltage tension on the crystal, it is possible to control the phase of the light at the output of the waveguides and thus generate constructive or destructive interferences. Fig. 1.3 represents the transmittance of the MZM in function of the tension $V_0 = V_1 - V_2$ between the arms. The transmittance is the ratio of transmitted power and is characterized by:

$$E_{out} = E_{in}\cos\left(\pi\frac{V_1 - V_2}{V_\pi}\right) \tag{1.2}$$

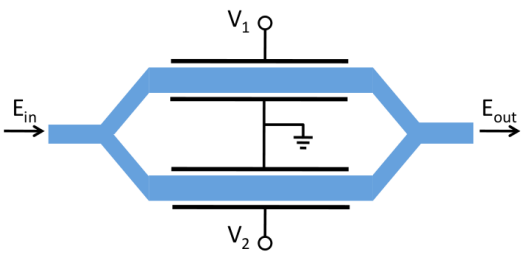


Figure 1.2: MZM

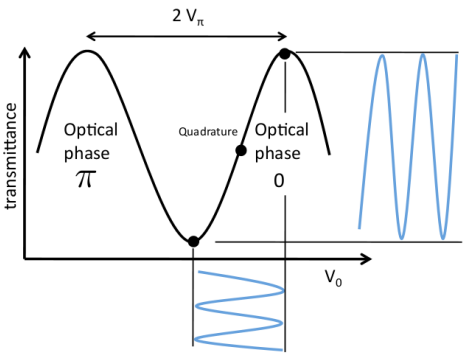


Figure 1.3: MZM transmittance

V_π represents the tension to pass from the maximum to the minimum transmittance. If V_0 is a modulated signal with an average value in the linear part of the transmittance as depicted in the figure, the electrical modulation is linearly transposed to the optical carrier. With adequate electrical driving signal V_0 it is possible to obtain amplitude or/and phase modulation.

1.1.2 Modulation formats

A modulation format is a family of signal states (amplitude, phase, frequency...). Each combination of states, for example phase and amplitude, is called a symbol. The constellation represents all the symbols that can be generated by the modulation format. Some examples of constellation are presented in Fig. 1.4. The average power of the constellation is noted P_s. In the case of a phase/amplitude modulation, the symbols can be noted:

$$E_s = A_s \exp(i\Phi_s)$$

where $A_s = \sqrt{P_s}$ is the amplitude of the signal and Φ_s is the phase of the symbol.

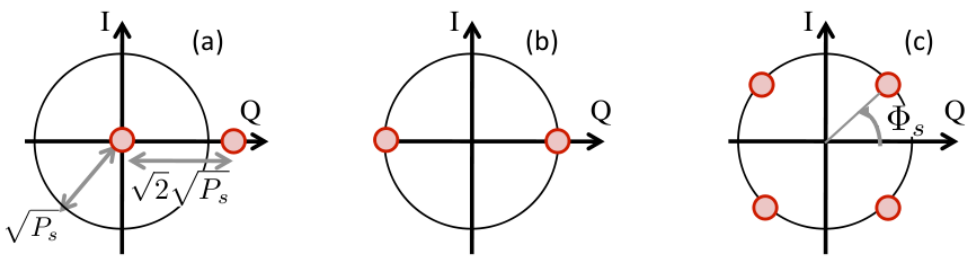


Figure 1.4: Constellations of (a) OOK, (b) BPSK, (c) QPSK, with average power P_s

A symbol is transmitted during a time T_s and the symbol rate is defined as $R_s = 1/T_s$ (symbol/s).
A distinct group of bits is associated to each symbol of the constellation. If the modulation has M states, each symbol is associated to $\log_2(M)$ bits. The spectral efficiency

$k_c = \log_2(M)$ is defined as the number of bit transmitted by a symbol. Therefore, the transmission bit-rate R_b is related to the symbol-rate R_s by:

$$R_b = k_c R_s \tag{1.3}$$

1.1.2.1 On-off keying

On-off keying (OOK) is the most basic amplitude modulation but has been employed from the first optical transmissions to the recent 40Gb/s systems. The principle is to turn the light on to transmit the bit $'1'$ and turn it off to transmit the bit $'0'$. Hence, OOK spectral efficiency is only one bit per symbol.
OOK modulation can be achieved with a single MZM by alternating constructive and destructive interference. The driving signal V_0 has an amplitude of V_π around the quadrature. Light is on (constructive interference) at the maximum transmittance and light is off (destructive interference) at the minimum as shown in Fig. 1.5(a).

OOK format is very popular because of the simplicity of its transmitter and receiver. Indeed a simple photodiode followed by a decision threshold is enough to get the bit sequence. However, this format is sensitive to the nonlinearities induced by the power fluctuations. The regular OOK format is named non-return to zero OOK (NRZ-OOK) as power does not return to zero between two symbols. Other version of OOK formats have been developed in order to improve the nonlinearity tolerance such as return to zero (RZ) OOK, chirped RZ-OOK or carrier suppressed RZ-OOK.

1.1.2.2 Binary phase shift keying

Binary phase shift keying (BPSK) is a binary phase modulation format where the bit $'0'$ is associated to the phase $\Phi_s = 0$ and the bit $'1'$ to the phase $\Phi_s = \pi$. As OOK, BPSK transmits only one bit per symbol. However BPSK has a better tolerance to nonlinearities because symbols have a constant energy $|E_s|^2$ and there are less power fluctuations. Moreover, it is less sensible to the noise and signal distortion because the distance between the symbols of the constellation is larger as depicted in Fig. 1.4.

The better performance of BPSK has the cost of a more complex implementation. The transmitter has the same architecture than that of OOK. To produce π phase shifts, V_0 has an amplitude of $2V_\pi$ as represented in Fig. 1.5(b). It is twice the amplitude required for the OOK format. The driving signal can not shift instantaneously from its maximum to its minimum value and thus passes by the minimum value of the transmittance and this causes a small fluctuation of the output power.
BPSK has also a more complex receiver than OOK scheme. Indeed a photodiode is quadratic receiver and the phase of the signal is lost by the conversion in the electrical domain. The solution is either to optically demodulate before the photodiode or to use

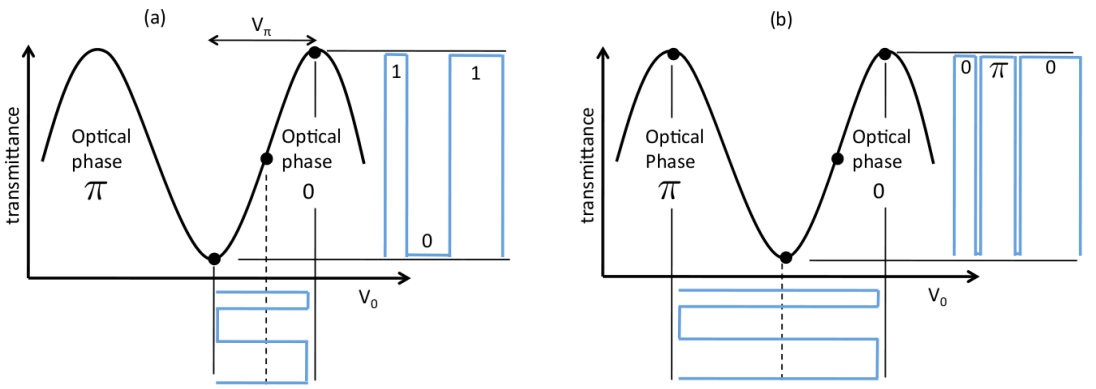


Figure 1.5: MZM driving sequence for (a) OOK (b) BPSK

coherent detection.

1.1.2.3 Quadrature phase shift keying

OOK and BPSK formats both have a spectral efficiency of one bit per symbol. This is an important limitation for the implementation of high bit-rate transmissions because the electronics devices (so the MZM driving signal) must have the same frequency as the modulation rate R_s. Therefore optical transmission bit-rates are limited by the maximum bit-rate in electronics. To overcome this problem, high spectral efficiency modulation formats have been investigated. Their constellation have more than two states in order to transmit more than one bit per symbol.

Quadrature phase shift keying (QPSK) is a modulation format where the symbols can have four distinct phases Φ_s: $\pi/4$, $3\pi/4$, $5\pi/4$ or $7\pi/4$ as shown in Fig. 1.4(c). The spectral efficiency is two bits per symbol which is twice as many as in OOK or BPSK formats. Hence the symbol rate is half the bit-rate.

QPSK is less sensitive to nonlinearities than OOK because there is no power fluctuations and is less sensitive to dispersion effects (chromatic dispersion, polarization mode dispersion) than BPSK because it works at half of its rate. Note that QPSK format has actually the same noise sensitivity than BPSK [4]. Indeed although the distance between neighbor symbols is $1/\sqrt{2}$ smaller, the spectral efficiency is the double.

A QPSK signal can be created from two BPSK signals, one 'in phase' and one on 'quadrature'. The transmitter is composed of two MZM arranged in a super MZ structure as shown in Fig. 1.7. In the same way, the QPSK receiver is composed by two BPSK receivers in order to detect the signal in phase and in quadrature.

1.1.3 Polarization multiplexing

The light propagating in a fiber is an electromagnetic wave having its electric field in the transverse plane orthogonal to the direction of propagation. The field can be decomposed in an orthogonal basis of the transverse plane and this can be interpreted as the two orthogonal polarizations of the light. The optical signal can be described by a 2×1 vector as:

$$\begin{bmatrix} E_{s,1} \\ E_{s,2} \end{bmatrix} = \begin{bmatrix} \sqrt{P_{s,1}} \exp\left(i\omega_c t + \Phi_{s,1}\right) \\ \sqrt{P_{s,2}} \exp\left(i\omega_c t + \Phi_{s,2}\right) \end{bmatrix} \tag{1.4}$$

The two polarizations can be modulated independently and this is called polarization division multiplexing (PDM). Hence, twice as much information is transmitted and the spectral efficiency is doubled.

Polarization multiplexing is realized by combining two optical modulated signal with a polarization beam splitter (PBS). Therefore one optical transmitter is required for each polarization but both share the same laser.
At the receiver, the signal is split in two with a PBS and then detected by two independent optical receivers. However, due to the birefringence and other polarization mixing effects in the fiber, the states of polarization at the reception are unknown. The polarization de-multiplexing can be done optically by carefully controlling the polarization states of the received signal using a polarization controller before the PBS. However, the most efficient solution is to realize this operation in the electrical domain. PDM is often combined with coherent detection in order to ease the digital de-multiplexing.

1.2 Optical fiber propagation

The optical field propagating on one polarization in a single-mode fiber can be expressed as [5]:

$$\mathbf{E}_s(x, y, z, \omega) = F(x, y) E_s(z, \omega) \exp(\imath\beta z) \tag{1.5}$$

where z is the distance of propagation, ω is the angular frequency (or pulsation), $F(x, y)$ is the field distribution which can be approximated as frequency independent, $E_s(z, \omega)$ is the field complex envelope and $\beta = \frac{\mathbf{n}(\omega)\omega}{\mathrm{c}}$ is the propagation constant with $\mathbf{n}(\omega)$ being the refractive index and c being the light velocity. As the spectral components of the signal are all close to the optical carrier ω_c, $\beta(\omega)$ can be developed in a Taylor series:

$$\beta(\omega) \approx \beta_0 + \beta_1(\omega - \omega_c) + \frac{1}{2}\beta_2(\omega - \omega_c)^2 + \dots \tag{1.6}$$

where β_i is the i^{th} derivative of β with respect to ω. By expressing the pulse envelope in the time domain,

$$E_s(z, t) = \frac{1}{2\pi} \int E_s(z, \omega) \exp(-\imath(\omega - \omega_c)t) \mathrm{d}\omega \tag{1.7}$$

the basic propagation equation can be derived [5]:

$$\frac{\partial E_s(z,t)}{\partial z} = -\beta_1 \frac{\partial E_s(z,t)}{\partial t} + \frac{1}{2}\imath\beta_2 \frac{\partial E_s(z,t)}{\partial t^2} + \dots \tag{1.8}$$

where $\beta_1 = 1/v_g$ can be seen as the group velocity v_g inverse and β_2 as the group velocity dispersion (GVD).

1.2.1 Transmission loss

Transmission loss corresponds to the optical signal power attenuation during the propagation in the fiber. Optical fiber is a silica waveguide and the attenuation depends on the wavelength of the signal. Within the telecommunication bandwidth ($\sim 1.2\mu$m– $\sim 1.7\mu$m), the main sources of attenuation are Rayleigh scattering due to the molecular level irregularity of the silica and residual material absorption (OH^- and SIO_2). Rayleigh scattering loss decreases in $1/\lambda_c^4$ so attenuates more shorter wavelengths whereas OH^- peaks of absorption are around 1.24μm and 1.39μm. Transmission loss is minimal around 1.55μm in standard SMF (wavelengths commonly used in telecommunication) and approximatively equal to 0.2dB/km. Wavelengths around 1.3μm are also sometimes used as they correspond at the same time to a second minimum of the transmission loss and the minimum of the chromatic dispersion.

We call P_{in} the power of the optical signal at the input of a fiber of length L and P_{out} the output power. The transmission loss coefficient α_a that includes all the sources of attenuation is defined by:

$$P_{out} = P_{in} \exp\left(-\alpha_a \mathrm{L}\right) \tag{1.9}$$

Usually transmission loss is expressed in dB/km by $\alpha_{dB} = 10log_{10}(e)\alpha_a$.

1.2.2 Chromatic dispersion

Chromatic dispersion is due to the frequency dependance of the refractive index $\mathbf{n}(\omega)$ of the fiber which makes the spectral components of the signal not travel at the same velocities and leads to pulse broadening.
Eq. 1.8 can be rewritten by choosing a reference moving with the pulse ($t' = t - \beta_1 z$) as:

$$\frac{\partial E_s(z,t')}{\partial z} = -\imath\frac{1}{2}\beta_2 \frac{\partial^2 E_s(z,t')}{\partial t'^2} \tag{1.10}$$

The dispersion D expresses the variation of the group velocity in function of the wavelength $\lambda = \frac{2\pi c}{\omega}$.

$$\mathrm{D} = \frac{\mathrm{d}}{\mathrm{d}\lambda}\left(\frac{1}{v_g}\right) = -\frac{2\pi c}{\lambda^2}\frac{\mathrm{d}}{\mathrm{d}\omega}\left(\frac{1}{v_g}\right) = -\frac{2\pi c}{\lambda^2}\beta_2 \tag{1.11}$$

Therefore the propagation equation becomes:

$$\frac{\partial E_s(z,t)}{\partial z} = \imath\ \frac{\mathrm{D}\lambda^2}{4\pi c}\ \frac{\partial^2 E_s(z,t)}{\partial t^2} \tag{1.12}$$

where λ is the wavelength and D expressed in ps/nm/km is the dispersion coefficient of the fiber. The transfer function express the relation between an input and output field. By solving the previous equation, the frequency domain transfer function of chromatic dispersion is obtained [6]:

$$\mathrm{H_{CD}}(z,\omega) = \exp(-\imath\ \frac{\mathrm{D}\lambda^2 z}{4\pi \mathrm{c}}\ \omega^2) \tag{1.13}$$

The typical value of dispersion in standard single mode fiber is 17ps/nm/km.

1.2.3 Polarization mode dispersion

In an ideal fiber, both polarizations have the same group velocity v_g. However in practice, there are some material imperfections such as local stress which result in a birefringence $\mathbf{\Delta n}$. Therefore the refractive index experienced by the two components of the electrical field is different. Hence, polarizations do not travel at the same velocities and after the propagation there is a differential group delay (DGD) $\Delta\tau$. As the local birefringence is a stochastic process, the DGD has a statistical distribution. Polarization mode dispersion (PMD) is expressed in ps and defined as the average DGD $\mathbb{E}\Delta\tau$.
In the first order approximation, DGD can be considered wavelength independent. The fiber is often modeled as a concatenation of equally distributed fiber sections in which the polarizations are randomly rotated. In that case, the DGD has a Maxwellian distribution [7]. The transfer function matrix of the PMD corresponds to a concatenation of random rotation matrices and birefringent matrices [8]:

$$\mathbf{H}_{\mathrm{PMD}}(\omega) = \prod_m \mathbf{R}_m \mathbf{D}_m \tag{1.14}$$

$$\mathbf{D}_m = \begin{bmatrix} \exp\left(\imath\frac{\omega\Delta\tau}{2}\right) & 0 \\ 0 & \exp\left(-\imath\frac{\omega\Delta\tau}{2}\right) \end{bmatrix}$$

The rotation matrix $\mathbf{R}_m$ depends on two random angles θ_m and φ_m describing the local orientation mismatch between the axes of the fiber and the polarizations of the signal.

1.2.4 Polarization dependent loss

Polarization dependent loss (PDL) describes the fact that signals with distinct polarization states are not attenuated in the same way during their propagation through an optical component (isolators, couplers, amplifiers...). Although optical devices are designed to introduce only a small amount of PDL, long-haul optical transmission systems include a large number of in-line optical components that make the total PDL after transmission be significant. In polarization multiplexed systems, PDL induces a fluctuation of the polarization power levels during the transmission that change the OSNR and lead to BER degradation at the receiver.

PDL depends on the polarization states of the signal and thus is a stochastic process. We consider the model used in [9]:

$$\mathbf{H}_{\mathrm{PDL}} = \mathbf{R}_1 \begin{bmatrix} \sqrt{1-\gamma} & 0 \\ 0 & \sqrt{1+\gamma} \end{bmatrix} \mathbf{R}_2 \tag{1.15}$$

$\mathbf{R}_1$ and $\mathbf{R}_2$ are random rotation matrices representing the orientation mismatch between the polarization states and the axes of the optical component having PDL. The attenuation γ is related to the PDL noted $\mathbf{\Gamma}_{dB}$ by:

$$\mathbf{\Gamma}_{dB} = 10 \log_{10} \frac{1+\gamma}{1-\gamma} \tag{1.16}$$

This model of PDL has been chosen for its simplicity however, more detailed models of PDL exist in the literature and we suggest to refer to the work of M. Shtaif *et al* as for example in [10].

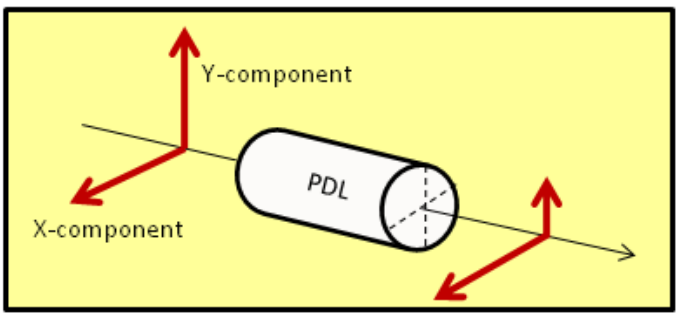


Figure 1.6: Principle of PDL when the axes of the optical element are parallel to the polarizations of the input signal

1.2.5 Erbium doped fiber amplifier

Despite the low transmission loss of the optical fiber, the signal has to be re-amplified in order to achieve long-haul transmission. Therefore, optical amplifiers are inserted periodically along the propagation, typically every 40-120km. Erbium doped fiber amplifier (EDFA) is the most popular optical amplifier as its amplification window coincides with the minimum transmission loss of the silica fiber ($\sim 1.55\mu$m) and can amplify evenly the whole WDM spectrum. EDFA was first demonstrated in 1987 by Mears [11] and Desurvire [12] and has turned to be a revolution in the domain of telecommunications.

Erbium doped fiber amplifier consists in a single mode fiber of few meters doped with Erbium ions (Er^{3+}) in which the signal co-propagates with the light issued from one or more pump lasers. The dopant ions are excited onto a high energy level by the pump lasers and relax to the lower energy level either by stimulated emission with the optical signal or by spontaneous emission.

The stimulated emission results to the creation of a photon with the same properties (direction, frequency...) than the signal and this photon is going to generate at its turn new photons and so on. This leads to the input signal amplification. To enable stimulated emission the ions have to be at an energy level corresponding to the wavelength (energy) of the input signal. Note that due to the nature of the silica, the ions are actually exited on a band of energies and this allows stimulated emission, thus amplification, of input signals with wavelengths around 1.55μm. Excitation to high energy band level is realized by pumping with a laser at 1.48μm or at 0.98μm. In the later case, the ions are transferred to the 1.48μm energy level by a non radiative relaxation. Typically, EDFA gain can be up to 40dB and the maximum output power is $\sim$ 23dBm (dBm is the power referenced to one mW).
On the other hand, the Er^{3+} ions can also decay spontaneously emitting photons in all directions. However a part of these photons are guided by the fiber in the propagation direction and added to the signal behaving as noise. Moreover they also interact with the ions by stimulated emission, generating new photons and thus, amplifying the noise. This process is referred as amplified spontaneous emission (ASE) and is the main limitation of optical transmission systems.
In [13], it is shown that ASE has the statistical property of additive gaussian noise. An EDFA that compensate the loss after a propagation into a fiber of length L, introduces some ASE with a power spectral density [14]:

$$\mathrm{N_{ASE}} = \left(e^{\alpha_a \mathrm{L}} - 1\right) \hbar \omega_s n_{sp} \tag{1.17}$$

where $\hbar$ is the Planck constant and n_{sp} is the spontaneous emission factor.
The optical signal to noise ratio (OSNR) is defined as the power ratio between the optical signal and the ASE spectral density within a reference spectral bandwidth B_{ref} of 0.1nm:

$$\mathrm{OSNR} = \frac{P_s}{2\mathrm{N_{ASE}} B_{ref}} \tag{1.18}$$

where P_s is the average optical signal power. The factor 2 at the denominator corresponds to the ASE on the two polarizations.

In digital communications, the signal to noise ratio (SNR) is derived from the capacity of the additive white gaussian noise channel (AWGN) and defined by:

$$\mathrm{SNR} = \frac{\mathrm{E}_b}{\mathrm{N}_0} \tag{1.19}$$

where N_0 is the noise spectral density and E_b is the energy per information bit which is related to the energy per modulated symbol by:

$$\mathrm{E}_b = \frac{\mathrm{E}_s}{k_c} \tag{1.20}$$

k_c being the modulation spectral efficiency. The relation between OSNR and SNR is given by [14]:

$$\mathrm{OSNR} = \frac{R_b}{2B_{ref}}\mathrm{SNR} \tag{1.21}$$

where R_b is the transmission information bit-rate.

1.2.6 Nonlinear effects

Nonlinear effects in optical fibers come from the significant value of the third-order susceptibility $\chi^{(3)}$ of the silica. This results in the intensity dependence of the refractive index which is known as optical Kerr effect [15]:

$$\tilde{\mathbf{n}}(\omega) = \mathbf{n}(\omega) + \mathbf{n}_2(\chi^{(3)})|E_s|^2/A_{eff} \tag{1.22}$$

where $\mathbf{n}_2(\chi^{(3)})$ is the nonlinear refractive index and A_{eff} is the effective area.
Eq. 1.8 can be rewritten in order to take into account fiber nonlinearities and the transmission loss:

$$\frac{\partial E_s(z,t')}{\partial z} = -\frac{\alpha}{2}E_s(z,t') + \imath\gamma_{NL}|E_s(z,t')|^2E_s(z,t') \tag{1.23}$$

$\gamma_{NL} = \omega\mathbf{n}_2/A_{eff}$ is called the nonlinearity coefficient. Note that chromatic dispersion has been omitted for a better understanding. The analytical solution is obtained by:

$$E(z,t') = E(0,t')e^{-\frac{\alpha z}{2}}e^{-\imath\Phi_{NL}(z)} \tag{1.24}$$

where Φ_{NL} is the nonlinear phase and can be expressed as:

$$\Phi_{NL}(z) = \gamma_{NL}|E(0,t')|^2\mathrm{L}_{eff}(z) \tag{1.25}$$

Because of the transmission loss, nonlinear effects only occur in the first part of the transmission when the optical power is still high. L_{eff} corresponds to the effective length of fiber where the nonlinear effects happen and is estimated by [15]:

$$\mathrm{L}_{eff}(z) = \frac{1-e^{-\alpha_a z}}{\alpha_a} \tag{1.26}$$

This nonlinear phase shift induced by the optical signal power fluctuations is known as self-phase modulation (SPM). The combined effect of SPM and chromatic dispersion causes amplitude distortion but if the value of dispersion is well chosen, the induced nonlinear dispersion can compensate partially or totally the chromatic dispersion. The later case is known as soliton propagation.
Cross-polarization modulation (XPolM) and cross-phase modulation (XPM) are two nonlinear impairments both based on the same principle as SPM. The only difference is that the refractive index fluctuations are induced not by the signal itself but respectively by the power fluctuations of a signal on another polarization and by power fluctuations of a

signal at another wavelength. XPolM occurs particularly in polarization multiplexed transmissions and results in nonlinear crosstalk between polarizations. XPM is on the other hand one of the major impairment in WDM systems, especially when intensity modulated signals co-propagate with phase modulated signals. Another kind of nonlinear impairment occurring in WDM transmissions, is four-wave mixing (FWM) which generates cross-talk between the WDM channels by creating from two co-propagating signals at ω_1 and ω_2, two new components at $\omega_3 = 2\omega_1 - \omega_2$ and $\omega_4 = 2\omega_2 - \omega_1$. SPM, XPM and FWM are the dominant nonlinear effects in 10Gb/s systems. At 40 and 100Gb/s, the main impairments are due to intra-channel XPM and intra-channel FWM where the power fluctuation is induced by neighbor pulses at the same wavelength that have been broadened by chromatic dispersion.

Raman scattering is another nonlinear effect caused by the interaction between optical signal and the silica molecules of the fiber. A photon is absorbed by the material and excites an electron to a higher energy level. This electron returns to an intermediary energy level by emitting a new photon with an energy smaller than the initial photon. Raman scattering effect is accentuated by stimulated emission in presence of co-propagating signals with the same wavelength than the scattered photon. Instead of being an impairment, this effect is used to generate continuous Raman amplification [16] which is an interesting alternative to EDFA.

1.3 Optical receiver

1.3.1 Direct detection

Direct detection is widely used in practical transmission systems at 10Gb/s and 40Gb/s in order to detect OOK or BPSK signals.
The signal is received on a pass filter characterized by its responsivity $\mathfrak{r}$ corresponding to the efficiency of the diode to convert the optical signal in current. Direct detection is particularly well adapted for amplitude modulation formats such as OOK. On the contrary, as photodiode is a quadratic receiver, the phase of the signal is discarded and direct phase modulation is impossible. However, by using a delay interferometer before the photodiodes as depicted in Fig. 1.7, the amplitude of the input signal become function of the phase shift between the modulated symbols. Hence, the modulation has to be realized differentially, by encoding the information on the phase shifts. This is known as differential modulation or differential encoding and will be discussed in Chapter 3. BPSK and QPSK with differential modulation are referred to DPSK and DQPSK.
In the receiver structure, a balanced detection between the two arms of the interferometers is realized in order to obtained a 3dB OSNR improvement compared to OOK. However the performance is very sensitive to the receiver imperfections such as time and amplitude imbalance between the photodiodes and/or between the different interferometers (in the

case of QPSK) [17].

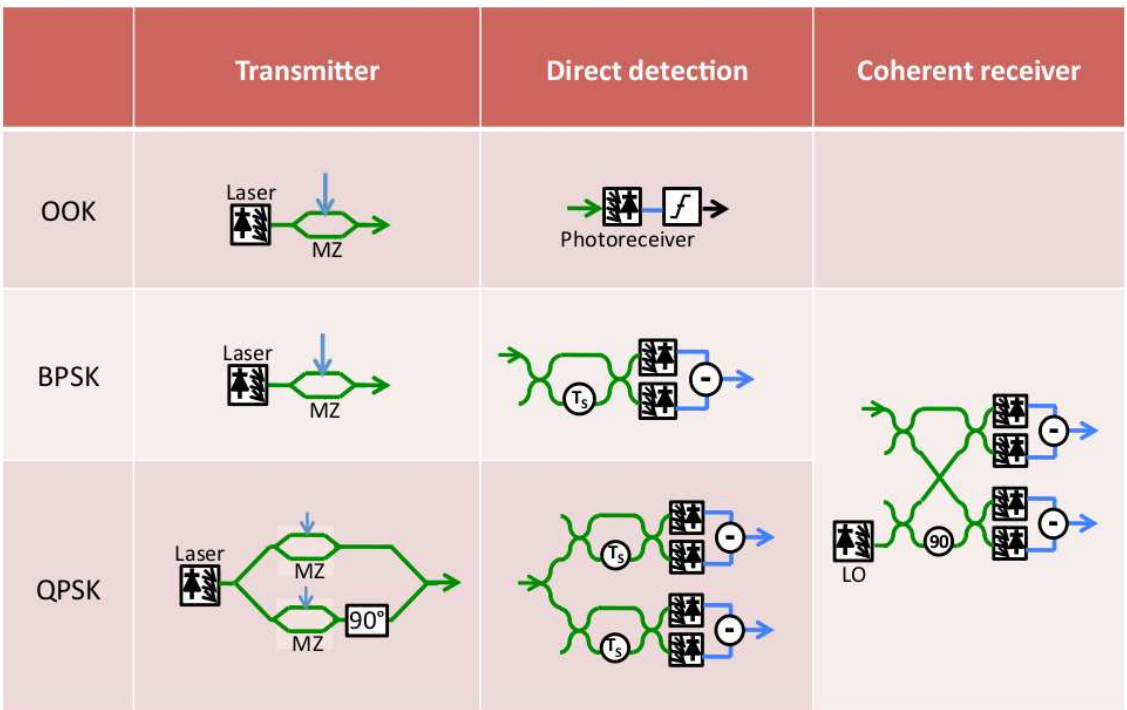


Figure 1.7: Optical transmitters and receivers

1.3.2 Coherent detection

Coherent detection corresponds to the mixing between the received signal and a local oscillator laser (LO). Coherent detection was first investigated in the 80's for its superior receiver sensitivity compared to direct detection. However, the invention of EDFA allowing pre-amplification, made this advantage less significant. Moreover, the high implementation complexity of coherent detection caused by the need of phase locking makes it uncompetitive compared to direct detection..

The revival of the coherent detection comes with the development of high speed electronic devices. Digital signal processing has become a good solution to mitigate transmission impairments. Indeed, the coherent receiver linearly transposed the received signal in the electrical domain and thus, amplitude and phase are preserved. Moreover, phase lock is no more necessary as LO phase shift can also be treated in the electrical domain by DSP. Coherent receiver are usually intradyne receivers which means that the LO frequency is very close to the carrier frequency.

In Fig. 1.7, a 90°-hybrid phase diversity coherent receiver is depicted. The receiver outputs are photocurrents corresponding to the component in phase and in quadrature of the

received signal and can be expressed as [15]:

$$I_Q(t) = \mathfrak{r}\sqrt{P_s(t)P_{LO}(t)}\cos(\Phi_s(t) - \Phi_{LO}(t)) \quad (1.27)$$

$$I_I(t) = \mathfrak{r}\sqrt{P_s(t)P_{LO}(t)}\sin(\Phi_s(t) - \Phi_{LO}(t)) \quad (1.28)$$

where $P_s(t)$, $P_{LO}(t)$, $\Phi_s(t)$ and $\Phi_{LO}(t)$ are the powers and the phases of the received signal and of the local oscillator. $\mathfrak{r}$ is the photodiode responsivity.

PDM receiver

Coherent detection is particularly adapted with polarization multiplexing because the polarization component can be transposed in the electrical domain by using a polarization diversity receiver.
The received signal is separated in two by a PBS and each polarization is mixed with the output of the LO in a 90°-hybrid phase diversity receivers. However, if the polarization states are not controlled before the PBS, the polarizations are mixed between the two received signals and digital processing is necessary to de-multiplexed the signals.

1.4 Digital equalization

Digital signal processing has been the major evolution of optical transmission systems during the last decade. By pre- or post-processing the signal in the electrical domain, it is possible to compensate signal distortion in a very efficient manner. In practical systems, optical compensation of transmission impairments is mainly restricted to chromatic dispersion compensation. However digital compensation can also deal with PMD, phase noise, frequency offset and de-multiplexing issues. Moreover it has made possible the use of high spectral efficiency modulation formats and multiplexing techniques such as polarization multiplexing or orthogonal frequency division multiplexing.

1.4.1 DSP with direct detection

Direct detection restricts the potential of DSP because only the amplitude of the signal is converted in the electrical domain. However it still demonstrates very good performance against chromatic dispersion or PMD [18].

Pre-distortion has been proposed to compensate chromatic dispersion in 10Gb/s system [19]. As it is a deterministic effect, it is possible to fully compensate it *a priori*. However, CD is also a limiting factor of nonlinear effects and thus pre-distortion of the signal reduces the nonlinearity tolerance.

Realizing the digital equalization only at the receiver has shown more potential. For instance, feed-forward and decision-feedback equalizers have been proposed [20] to compensate chromatic dispersion in a simple way. However, the discard of phase information

by the photodiodes induces a performance limitations, as high signal distortions can't be supported.

The more robust solution is the maximum likelihood sequence estimator (MLSE) which is based on the Viterbi algorithm [21]. It has very good performance against chromatic dispersion and PMD [22], but is more complex than linear equalizers. The principle is to search among all the possible emitted sequences, the one minimizing the probability to obtain the received sequence. Hence, it requires the probability density function of the transition probabilities. It corresponds to the probability that an emitted sequence of bits have been received changed into another sequence. These probability density function have to be estimated *a priori* by a training sequence. MLSE is very efficient with OOK format but it is ineffective with DPSK and DQPK [23] because of the balanced detection. Joint decision MLSE, partial DPSK and joint symbol MLSE have been proposed as solutions.

1.4.2 DSP with coherent detection

With coherent detection, the received signal can be entirely rebuilt in the electrical domain. Therefore, we have access to the amplitude, the phase, the frequency and the polarization states of the signal in order to compensate the transmission impairments.

The inputs of the coherent digital equalizer are the photocurrents issued from the optical detectors. Each of them is transmitted into an analog-to-digital converter (ADC) working at a given sampling rate. The sampling rate r_{ov} may have to be superior to 1 samples per symbol in order to satisfy Nyquist theorem. Indeed due to the various transmission effects and filtering, a spectrum broadening occurs. If the sampling rate is not an integer number of samples per symbol duration, a re-sampling can be performed.
Afterward, the received signal is recovered using the ADC ouputs corresponding to the in-phase and the quadrature signals:

$$y(k) \triangleq y(k\frac{T_s}{r_{ov}}) = I_Q + \imath I_I \tag{1.29}$$

Then this signal is digitally processed in order to compensate every transmission impairments. First chromatic dispersion is removed and polarization tributaries are demultiplexed by compensating fiber birefringence and polarization mode dispersion. Finally, frequency offset is removed and the carrier phase is recovered. The scheme of the DSP in a polarization multiplexed coherent system is depicted in Fig. 1.8.

1.4.2.1 CD compensation

Chromatic dispersion is a deterministic phenomenon only depending on the fiber length and dispersion coefficient D. Usually, CD is compensated in-line with the introduction of

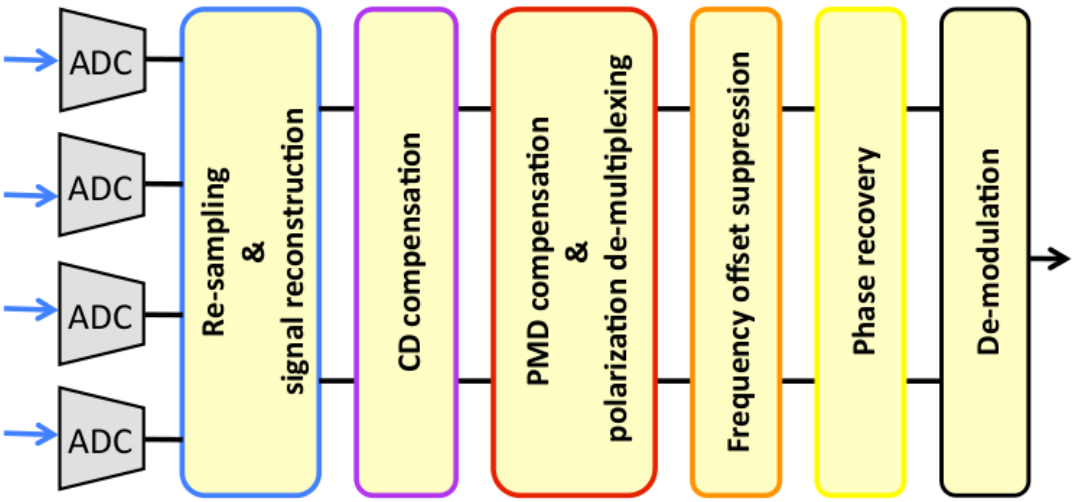


Figure 1.8: Digital signal processing of polarization diversity coherent receivers

compensating fiber with negative dispersion coefficient. In that case, CD equalization has only to compensate the residual dispersion which is relatively small. This is achieved by a finite impulse response (FIR) filter with a small number of taps. The structure of a FIR filter is represented in Fig. 1.9 and the outputs $y(k)$ can be expressed by:

$$\begin{aligned} y(k) &= \mathbf{h} * \mathbf{x} \\ &= h(L).x(k-(L-1)) + h(L-1).x(k-(L-2)) \ \ldots \ h(1).x(k) \end{aligned} \tag{1.30}$$

where L is the number of taps of the filter, $\mathbf{h} = [h(1) \ldots h(L)]$ are the coefficients of the filter and $\mathbf{x} = [x(k-(N-1)) \ldots x(k)]$ are the filter inputs.

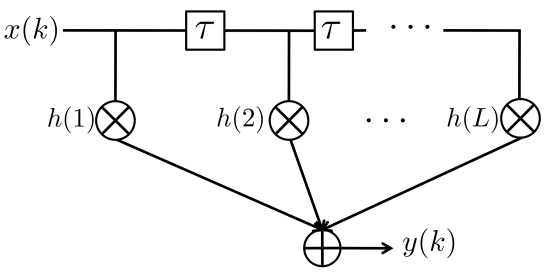


Figure 1.9: FIR filter

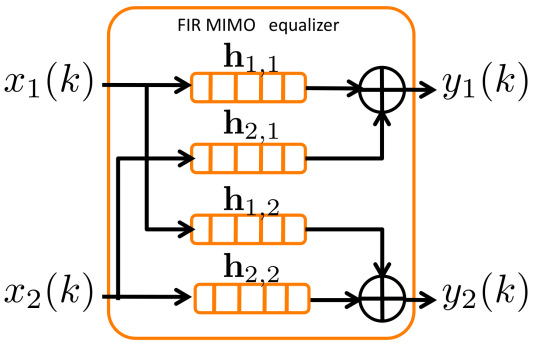


Figure 1.10: butterfly FIR structure

Chromatic dispersion is a polarization independent effect and can be compensated before polarization de-multiplexing. CD is modeled by the frequency domain transfer matrix function in Eq. 1.13. The dispersion compensating filter is the inverse of the CD transfer

function:

$$\mathrm{G_{CD}}(z,\omega) = \exp(\imath\ \frac{\mathrm{D}\lambda_c^2 z}{4\pi \mathrm{c}}\ \omega_c^2) \tag{1.31}$$

which corresponds to filter taps in the time domain given by [24]:

$$g(m) = \sqrt{\frac{\imath \mathrm{c} T_s^2}{r_{ov}\mathrm{D}\lambda^2 z}} \exp\left(-\imath \frac{\pi \mathrm{c} T_s^2}{4 r_{ov}\mathrm{D}\lambda^2 z} m^2\right) \tag{1.32}$$

with $\lfloor -N/2 \rfloor \leq m \leq \lfloor N/2 \rfloor$, $N = \lfloor \frac{r_{ov}|\mathrm{D}|\lambda^2}{2\mathrm{c}T_s^2} \rfloor$ being the total number of tap of the filter. T_s/r_{ov} corresponds to the sampling period.

Note that fibers with negative dispersion usually have a large nonlinearity coefficient and CD in-line compensation results in lower nonlinearity tolerance. Therefore it can be more interesting to compensate the CD entirely in the electrical domain [25]. The difficulty of doing this, comes from the large number of taps required to compensate the dispersion accumulated during a long-haul transmission[26]. Another solution is to realize this equalization in the frequency domain where only 1-tap FIR filters are needed. In that case the complexity comes from the FFT/iFFT operations that convert the signal from the time domain to the frequency domain.

1.4.2.2 Polarization de-multiplexing

In the case of polarization multiplexed transmissions, the transmitted signals are randomly mixed on the two polarizations because of fiber birefringence and delayed by PMD which creates inter-symbol interferences. The equalizer has the butterfly FIR structure depicted in Fig. 1.10 and the output can be expressed by:

$$y_j(k) = \sum_{i=1}^{2} \sum_{l=1}^{L} h_{ij}(l) x_i(k-l+1) \tag{1.33}$$

where the FIR filter between the i^{th} input and the j^{th} output is noted $\mathbf{h}_{i,j} = [h_{ij}(1) \ldots h_{ij}(L)]$.

Unlike CD, PMD and the birefringence are random effects depending on the polarization states of the signal and so, vary in time. Hence, the coefficients of the filters are not constant and have to be estimated and updated periodically. In PDM-QPSK transmission, the most popular equalizer is the constant modulus algorithm (CMA).

CMA is a blind adaptive equalizer based on the least-mean square (LMS) algorithm. As QPSK constellation has a constant modulus, CMA intends to find the optimum equalizer coefficients in order to minimize the distance between its output absolute values and a given radius r_{CMA}. This corresponds to minimizing the error functions ϵ_j $(j = 1, 2)$:

$$\epsilon_j = r_{CMA}^2 - |y_j|^2 \tag{1.34}$$

The optimum coefficients are searched following the stochastic gradient descent. This method is adaptive so the coefficients are updated at each new input symbol.

$$\mathbf{h}_{i,j}(k) = \mathbf{h}_{i,j}(k-1) - \mu\ \epsilon_i\ \mathbf{y}_j \mathbf{x}_i^H \tag{1.35}$$

where $\mathbf{y}_j = [y_j(k) \dots y_j(k-L+1)]$ and $\mathbf{x}_i = [x_i(k) \dots \mathbf{x}_i(k-L+1)]$ and $\mathbf{h}_{i,j}(k)$ is the FIR filter coefficients at the instant kT_s/r_{ov}.

1.4.2.3 Frequency offset compensation

Due to the frequency offset between the signal carrier and the local oscillator, the received signal is affected by a phase shift ϕ_f:

$$\arg\left(x_i(k)\right) = \phi_s(k) + \phi_0 + k\phi_f \tag{1.36}$$

where ϕ_s is the modulation phase and ϕ_0 a phase offset. To estimate the phase shift, these two phase shifts have first to be eliminated by computing $\left(x_i(k)\ x_i(k-1)^H\right)^4$ [27]. Then the phase shift is averaged on a bloc of n symbols.

$$\widehat{\phi_f}(k) = \frac{1}{4}\arg\left(\sum_{m=-n/2+1}^{n/2} \left(x_i(k+m)\ x_i^H(k+i-1)\right)^4\right) \tag{1.37}$$

1.4.2.4 Carrier phase estimation

Once the frequency offset has been removed, the carrier phase can be estimated and a common way to do it is by the Viterbi-Viterbi algorithm [28]. To estimate the phase, the contribution of the modulation has to be removed by taking the fourth power of the signal and then averaging:

$$\widehat{\phi_0}(k) = \frac{1}{4}\ \arg\left(\sum_{m=-n/2+1}^{n/2} x_i(k+m)^4\right) \tag{1.38}$$

Note that removing the contribution of the modulation by taking the fourth power of the symbols leads to a phase ambiguity of $k'\frac{\pi}{2}$. However, as this phase ambiguity concern all the symbols, the phase difference between two successive symbols remains. Therefore the encoding has to be realized on the phase shift between the symbols. This is called differential encoding.

1.4.2.5 Demodulation

We have plotted in Fig. 1.11, the QPSK constellation received, de-multiplexed and equalized. After the carrier phase recovery, the symbols are only impaired by ASE and are located in one of the four quadrants. The demodulation is realized by associating to the symbols, the bits corresponding their quadrant. This is called hard decision because a threshold decision is realized. For instance, if the equalized QPSK symbol has both positive real and imaginary values (first quadrant), it will be associated to the bits [0 0].

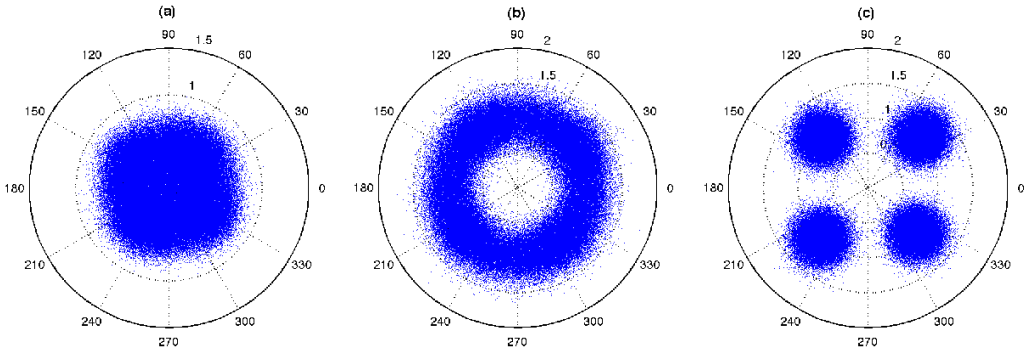


Figure 1.11: Equalization of a polarization multiplexed QPSK signal with OSNR=17.6dB: (a) Received signal (b) Demultiplexed signal (after CMA) (c) Equalized signal (after phase recovery)

On the other hand, the symbols can be directly transmitted to the FEC without taking any decisions. Hence, the FEC inputs are not bits but soft-values quantified over many levels. The soft-decision values are usually converted into log likelihood ratio (LLR) that expresses the probability of a bit to be 0 or 1 depending on the received symbols.

1.5 Forward error correction

The quality requirement of optical long-haul transmission systems is very high as an output BER of only $10^{-12}-10^{-15}$ is expected. Achieving such a BER would require a very high SNR, so a large number of optical amplifiers and a perfect compensation of transmission impairments. In such conditions transmissions over long distance are quite difficult.

Forward error correction (FEC) has been adopted very early in optical transmission systems in order to correct digitally the transmission errors and reduce the output BER. It relaxes the quality requirement and enable transmissions over larger distances, at higher bit-rates and with fewer optical amplifiers. For a given BER, we quantify the gain obtained by the FEC as the difference of SNR between the coded (with FEC) and uncoded (without FEC) cases. This gain is called "coding gain" and is expressed in dB.

An FEC introduces some redundancy to the data in order to protect the information. At the reception, the FEC decoding is able to correct a certain number of transmissions errors. The rate r of an FEC is the ratio between the number of information bits (the data) and the number of bits actually transmitted. Equivalently, the FEC overhead (or redundancy) corresponds to the percentage of redundancy bits introduced. Because of the redundancy, one modulated symbol corresponds to less information bits thus, Eq. 1.20 should be modified into:

$$\mathrm{E}_b = \frac{\mathrm{E}_s}{k_c r} \tag{1.39}$$

Summary

In this chapter , we have presented an overview of an optical fiber transmission scheme. An optical fiber transmission system is composed by a transmitter, a fiber link and a receiver. The key element of the transmitter is the electro-optical modulator that modulates the optical carrier phase and/or amplitude in function of a driving tension. We have presented the structure of the optical transmitter for OOK, BPSK and PDM-QPSK which are the common modulation formats in practical systems respectively at 10, 40 and 100Gb/s. During the propagating in the fiber, the signal is impaired by linear dispersion causing inter symbol interference and nonlinear effects occurring at high power and causing signal distortion. Moreover, the optical amplifiers introduced to compensate the fiber loss add some noise which degrade the performance.
The optical signal can be received by direct detection or coherent detection. Direct detection is preferred for its implementation simplicity in the case of OOK and BPSK. However coherent detection becomes more advantageous for high spectral efficiency formats and allows polarization multiplexing and efficient digital signal processing. Finally we have described the different steps of the DSP that are implemented in actual 100Gb/s coherent polarization multiplexed systems.
PDM-coherent transmissions and the DSP described in this chapter correspond to the state of the art of deployed optical transmission systems.

Chapter 2

Forward error correction in optical transmission systems

Introduction

Forward error correction (FEC) has been employed in optical transmission systems for more than 20 years. In the same way as optical amplifier, it has been a major contributor to the increase of transmission distances and bit-rates. There are been two generations of FEC implemented in practical systems, the linear block codes and the concatenated FEC. Today lot of investigation is done to implement a third generation, the soft decoding FEC.

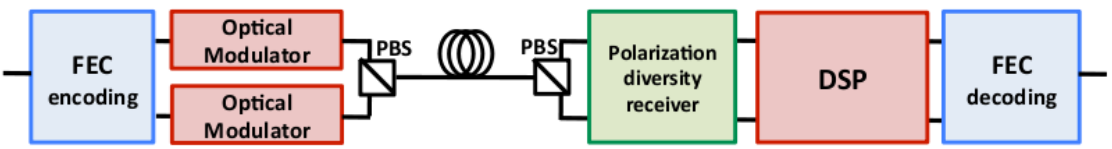


Figure 2.1: System model

In this chapter, we give an overview of the FEC in optical transmission systems. It will be separated in two parts. The first part is dedicated to the state of the art of FEC. In Sec. 2.1, we remind the FEC principles and present the different families of FEC that have been implemented optics. In Sec. 2.2, a brief review of the history of FEC in optical transmission systems is proposed.
The second part focuses on the soft-decoding FEC implementation in high bit-rate optical systems. In Sec. 2.3, a comparative study between LDPC and Product codes is realized in order to evaluate the most appropriate code. Finally Sec. 2.4 focuses on the construction of LDPC codes. In this last section an original construction is presented, performing better than most LDPC codes proposed in the literature and being very suitable for optical transmissions.

Part I : State of the art of FEC in optical transmission systems

2.1 Forward error correction

The principle of forward error correction is to introduce redundancy to the transmitted information (encoding) following a coding rule known at the transmitter and at the receiver. The redundancy bits protect the information and let the receiver correct some of the transmission errors. Linear block codes are a family of FEC where the information bits are encoded by blocks. It associates to k information bits, n coded bits. Hence the rate of the code is:

$$r = \frac{k}{n} \tag{2.1}$$

The code $\mathcal{C}(n,k)$ is said systematic if the coded sequence is composed of k information bits and $n-k$ redundancy bits as depicted in Fig. 2.1.

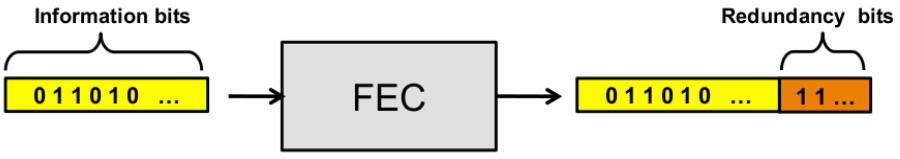


Figure 2.2: FEC encoding

The information bits can been seen as a vector of the finite field $GF(2)^k$ and the encoding operation corresponds to a linear matrix multiplication:

$$\boldsymbol{c} = \boldsymbol{m}\mathbf{G}$$

where $\boldsymbol{c} \in GF(2)^n$ is the coded bit sequence called codeword, $\boldsymbol{m} \in GF(2)^k$ is the information bits vector and $\boldsymbol{G}$ is the FEC generator matrix. The set of codewords forms a vector subspace of $GF(2)^n$ and the lines of the matrix $\boldsymbol{G}$ represent a basis of this subspace. Let $\boldsymbol{H}$ denote the basis of the dual subspace, it satisfies:

$$\boldsymbol{c}\boldsymbol{H}^t = 0 \tag{2.2}$$

$\boldsymbol{H}$ is called the parity-check matrix of the code.

Let us assume that a codeword $\boldsymbol{c} \in \mathcal{C}$ is transmitted and the vector $\boldsymbol{r}$ is received. Some errors may have occurred during the transmission so $\boldsymbol{r} = \boldsymbol{c} + \boldsymbol{e}$ where $\boldsymbol{e}$ is the error vector having a 1 at the positions of each error. The syndrome is defined as:

$$\begin{aligned} synd &= \boldsymbol{r}\boldsymbol{H}^t \\ synd &= \boldsymbol{c}\boldsymbol{H}^t + \boldsymbol{e}\boldsymbol{H}^t \\ synd &= \boldsymbol{e}\boldsymbol{H}^t \end{aligned} \tag{2.3}$$

If the syndrome is null, the received vector is a codeword. Therefore the syndrome computation tells us if the received message is a valid codeword.

The minimal distance d_{min} of a code is defined as the minimal number of different bits between two codewords. A linear block code is usually noted $\mathcal{C}(n, k, d_{min})$. The minimal distance is related to the code performance. When FEC are employed, we have two strategies:
– Error detection: we can detect errors only if the received vector is not a codeword (Eq. 2.2 is not satisfied). This is always possible when the number of error t is less than d_{min}.
– Error correction: in order to correct the errors, one has to find the most likelihood codeword. If the number of error is superior to $d_{min}/2$, the most likelihood codeword is not the emitted codeword. Hence the maximum number of errors t_{max} that a code $\mathcal{C}(n, k, d_{min})$ is able to correct is:

$$t_{max} = \left\lfloor \frac{d_{min} - 1}{2} \right\rfloor \tag{2.4}$$

2.1.1 Hamming code

Hamming code is a family of linear block codes defined by $\mathcal{C}(2^m - 1, 2^m - m - 1, 3)$ with $m > 2$. It can detect up to 2 errors or correct 1 error. The parity-check matrix corresponds to the bit representation of the integers between 1 and $2^m - 1$. In general, Hamming codes are decoded by syndrome decoding.

Syndrome decoding

Syndrome decoding is a decoding algorithm adapted for FEC with a low correction capability, typically $t_{max} = \{1, 2\}$.
The syndrome $\boldsymbol{eH}^t$ is equal to the sum of the columns of $\boldsymbol{H}$ corresponding to the error positions. Hence, it gives an indication on the location of the errors. It can be shown that if $t \leq t_{max}$ the syndrome is unique and thus, determines precisely the error positions. Therefore a look-up table including the values of all the syndromes and the corresponding error configurations can be created.
When a vector is received, its syndrome is computed and we search in the table which is the corresponding error configuration.

In the case of the Hamming code, if an unique error occurred, the syndrome is equal to one of the column and gives directly the error position. If there are more than one error, the syndrome is the sum of many columns which is always equal to another column of $\boldsymbol{H}$. So the decoding fails and creates a new error.

2.1.2 Bose-Chaudhuri-Hocquenghem codes

BCH codes have been discovered by Bose, Chaudhuri and Hocquenghem in 1959 [29][30]. They are a sub-family of linear block codes constructed using algebraic tools on finite fields in order to correct up to t_{max} errors.
BCH codes are usually described by using a polynomial representation of the bit vectors. To $\boldsymbol{b}=[b_1\ b_2 \dots b_q]$, we associate the polynomial:

$$\boldsymbol{b}(x) = \sum_{i=1}^{q} b_i x^{i-1} \tag{2.5}$$

Let $\boldsymbol{c}(x)$ and $\boldsymbol{m}(x)$ be respectively the polynomial representation of a codeword and an information vector. The encoding is performed by:

$$\boldsymbol{c}(x) = \boldsymbol{m}(x)\boldsymbol{g}(x) \tag{2.6}$$

where $\boldsymbol{g}(x)$ is called the generator polynomial and is defined in $GF(2)[X]$. $\boldsymbol{c}(x)$ and $\boldsymbol{m}(x)$ have respectively a degree $n=2^m-1$ ($\boldsymbol{c}$ has a length n) and $k=n-\deg(\boldsymbol{g}(x))$.

The generator polynomial $\boldsymbol{g}(x) \in GF(2)[X]$ is chosen in order to correct t_{max} errors. Let α be the primitive element of $GF(2^m)$ so $\alpha, \alpha^2, \alpha^3 \dots \alpha^{n-1}$ are the roots of X^n-1, where $n=2^m-1$. $g(x)$ is constructed in order to have $\alpha^k, \alpha^{k+1} \dots \alpha^{k+2t_{max}-1}$, with $1 \leq k \leq n-1$ being its roots. If we note $f_{\alpha^k}(x) \in GF(2)[X]$ is the minimal polynomial of α^k, the generator polynomial of a BCH code is defined as:

$$\boldsymbol{g}(x) = \text{lcm}\left[f_{\alpha^k}(x), f_{\alpha^{k+1}}(x) \dots f_{\alpha^{k+2t_{max}-1}}(x)\right] \tag{2.7}$$

The decoding of BCH codes is realized in three steps:

1. Syndrome calculation
2. Computation of the error locator polynomial using Peterson-Gorenstein-Zierler algorithm [31] or Berlekamp-Massey algorithm [32].
3. Factorization of the error locator polynomial by Chien search [33] in order to determine the errors positions

This decoding will be referred in Chapter 3 as the "algebraic" decoding of BCH codes.

2.1.3 Reed-Solomon codes

Reed-Solomon (RS) codes are a sub-family of BCH codes but have the particularity of coding blocks of bytes instead of blocks of bits. We call byte a group of bits.

The generator polynomial $\boldsymbol{g}(x)$ belongs to $GF(2^m)[X]$ and is defined as:

$$\boldsymbol{g}(x) = (x-\alpha)(x-\alpha^2)\dots(x-\alpha^{2t_{max}})$$

where α is the primitive element of $GF(2^m)$. Note that as $\boldsymbol{g}(x) \in GF(2^m)[X]$, $\boldsymbol{c}(x)$ and $\boldsymbol{m}(x)$ are also in $GF(2^m)[X]$. It means that the entries are no more binary input $\{0,1\}$ but belong to a larger alphabet $\{\alpha_1 \dots \alpha_{2^m-1}\}$ where each α_i represents a byte.
A RS code can correct t_{max} erroneous bytes. Therefore they are particularly well adapted to correct burst of errors. For instance RS$(255, 239)$ is constructed over $GF(2^8)$ and is able to correct 8 bytes of 8 bits each. Note that RS codes are optimal in the sense that they have the largest possible d_{min} and thus introduce the fewest redundancy $n-k=2t_{max}$

The decoding of RS codes follows the same steps as the BCH decoding. However, as errors occur on bytes, a final step is necessary to find the values of the errors. We add for that a fourth step to the algebraic decoding:

4. Computation of the error values using Forney algorithm [34].

2.1.4 Low-Density Parity-Check codes

Low-density parity-check codes (LPDC) have been invented in 1962 by Gallager [35]. For long time forgotten, they have been re-discovered by Mackay and Neal in 1996 [36].
LDPC are linear block codes characterized by a sparse parity-check matrix. It means that there are very few entries at 1 in the parity-check matrix. A regular LDPC $\mathcal{C}(n, l, c)$ has l one on each line of $\boldsymbol{H}$ and c one on each column. The codeword size is n and the rate of the code is:

$$r = \frac{n - \text{rank}(\boldsymbol{H})}{n} \tag{2.8}$$

The parity-check matrix can be represented by a bipartite graph called Tanner graph having two kind of nodes: the check nodes corresponding to the parity-check equations and the variable nodes corresponding to the bits. There is an edge E_{c_m,v_k} between the check node c_m and variable node v_k if the k^{th} bit is involved in the m^{th} parity-check equation.

$$E_{c_m,v_k} \longleftrightarrow \boldsymbol{H}(m,k) = 1 \tag{2.9}$$

Each node is connected to a set of neighbor nodes noted $\mathcal{S}_c$ if it is a check node and $\mathcal{S}_v$ if it is a variable node. The size of this set is called the node degree.
Fig. 2.3 represents the parity-check matrix of the Hamming code $\mathcal{C}(7, 4)$ and the corresponding Tanner graph representation. The neighbor set of c_1 are the variable nodes $\mathcal{S}_{c_1} = \{v_1, v_2, v_3, v_5\}$.

In the Tanner graph we define a cycle as a path starting and ending at the same node and passing by each edge only once. The length of the cycle is the number of edges in the path. As a Tanner graph is a bipartite graph, the cycle size are even and superior or equal to 4. The size of the shortest cycle is called the girth. For instance, the colored edges $[E_{c_1,v_1}\ E_{v_1,c_3}\ E_{c_2,v_3}\ E_{v_3,c_1}]$ form as cycle of length 4.

$$\mathbf{H} = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix}$$

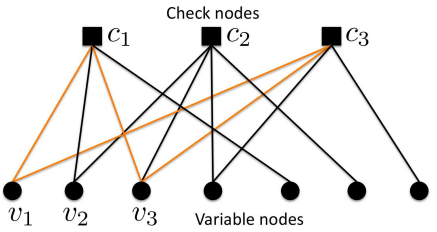


Figure 2.3: Tanner graph representation of the Hamming code $\mathcal{C}(7,4)$

Short cycles degrades the performance of the FEC. Hence LDPC are constructed in order to have large girths. In Sec. 2.4, some constructions proposed in the literature will be presented. Now let us present the classical decoding algorithms of LDPC codes.

2.1.4.1 Belief Propagation Algorithm

The belief propagation algorithm is an iterative soft-input decoding which corresponds to the maximum *a posteriori* decoding if Tanner graph is cycle-free. When cycles are present, especially short ones, the performance is degraded and it results to an error floor. The error floor level depends on the girth and on the number of short cycles.
The algorithm consists on a two-step exchange of information between the two sides of the graph and the exchanged messages are usually log-likelihood ratio (LLR).

Let us consider that a codeword $X = [x_1, x_2 \dots x_n]$ has been transmitted over a memoryless channel and that $Y = [y_1, y_2 \dots y_n]$ is the received vector. We associate to each received symbol y_i, a soft-input value called log-likelihood ratio and defined as:

$$L_{ch}(y_i) = \log \frac{\Pr\{x_i = 0 | y_i\}}{\Pr\{x_i = 1 | y_i\}} \tag{2.10}$$

The LLR expressed the reliability of the received bits to be 1 or 0. The larger the LLR is, the more reliable the bit is. If the LLR is positive, the probability to be a 0 is larger and if the LLR is negative the probability to be a 1 is larger.

We note μ_{c_m,v_k} the message passing from the check node c_m to the variable node v_k and μ_{v_k,c_m} the other way.

- *Initialization*: The messages from variable nodes to check nodes are initialized by:

$$\mu_{v_k,c_m} = L_{ch}(y_k) \tag{2.11}$$

- *Check node step*: Each check node receives messages from its neighbor variable nodes and computes the output messages:

$$\mu_{c_m,v_k} = 2 \tanh^{-1} \left(\prod_{i \in \mathcal{S}_{c_m} \setminus v_l} \tanh \left(\frac{\mu_{i,c_m}}{2} \right) \right) \tag{2.12}$$

• *Variable node step*: Each variable node receives messages from its neighbor check nodes and computes the output messages:

$$\mu_{v_k,c_m} = L_{ch}(y_k) + \sum_{i \in \mathcal{S}_{v_k} \backslash c_m} \mu_{i,c_m} \tag{2.13}$$

Finally the LLR of the estimated bits are obtained by:

$$L(\hat{x}_k) = L_{ch}(y_k) + \sum_{i \in \mathcal{S}_{v_k}} \mu_{i,c_m} \tag{2.14}$$

If $\hat{X} = [\hat{x}_1 \hat{x}_2 \dots \hat{x}_n]$ is a codeword (the syndrome is null) the procedure stops, else a new iteration can be performed.

The output LLR are obtained by successive sum and product operations therefore, this algorithm is usually referred as the Sum-Product Algorithm (SPA).

2.1.4.2 Min-Sum Algorithm (MSA)

In Eq. 2.12, the operations tanh and $\tanh^{-1}$ bring a significant complexity to the SPA algorithm. Therefore one can simplify the check node step doing [37]:

• *Check node step*:

$$\mu_{c_m,v_k} = \prod_{i \in \mathcal{S}_{c_m} \backslash v_k} \operatorname{sgn}(\mu_{i,c_m}) \times \min_{i \in \mathcal{S}_{c_m} \backslash v_k} |\mu_{i,c_m}| \tag{2.15}$$

This approximation leads to performance degradation compared to the SPA. An offset factor β_{off} and/or a scaling factor α_{sca} can be introduced to get closer to the SPA performance [38]:

$$\mu'_{c_m,v_k} = \max\left(\alpha_{sca}\mu_{c_m,v_k} - \beta_{off}, 0\right) \tag{2.16}$$

Note that α_{sca} and β_{off} are usually found empirically and may change at each iteration. Offset-MSA performs close to SPA but has a reduced decoding complexity.

2.1.5 Concatenated codes

Concatenated codes consist of serially concatenating FEC as illustrated in Fig. 2.4. They have been first proposed by D. Forney in 1966 [39]. Between the two encoders, the bits are interleaved in order to spread the bursts of errors over many codewords. The first FEC is called outer code and the second one is called inner code. The idea is to use two FEC with different properties, for example a BCH for the inner code and a RS code for the outer code.

Let us consider the concatenation of $\mathcal{C}_1(n_1, k_1)$ and $\mathcal{C}_2(n_2, k_2)$, the concatenated code obtained is $\mathcal{C}(n_1+n_2-k_2, k_1)$. For cxample, the code $\text{BCH}(239, 233) + \text{BCH}(255, 239)$ has a rate 223/255 and a codeword length of 225 bits.

The decoder is a mirror scheme of the encoder. The inner FEC decodes first, then the bits are de-interleaved and finally the outer FEC finishes the decoding.

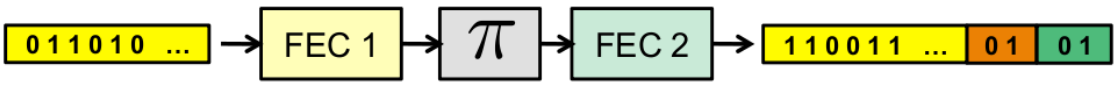


Figure 2.4: Concatenated codes

2.1.6 Product codes

Product codes are a particular case of concatenated codes where the interleaving is performed over k_2 codewords of $\mathcal{C}_1$ in a way that the i^{th} input vector of the second encoder is formed by the i^{th} bits of each codewords from $\mathcal{C}_1$. Fig. 2.5 shows that information bits can be seen $k_1 \times k_2$ matrix being a vertical concatenation of information vectors. The lines are first encoded by FEC 1 and then the columns are encoded by FEC 2. The product of $\mathcal{C}_1(n_1, k_1)$ and $\mathcal{C}_2(n_2, k_2)$ results in the code $\mathcal{C}(n_2 \times n_1,\ k_2 \times k_1)$.

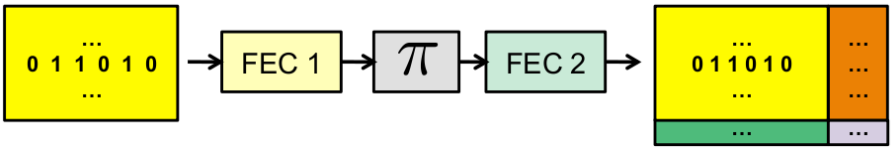


Figure 2.5: Product codes

The decoding of Product codes is realized iteratively. At each iteration, FEC 1 decoding is realized (horizontal decoding) and then FEC 2 decoding is realized (vertical decoding). This can be performed many times. The decoding of the codewords can be done in a soft or a hard way. The soft decoding is usually performed by the ChaseII algorithm [40].

2.1.6.1 ChaseII algorithm

Let $Y = [y_1, y_2 \ldots]$ denote the received symbols. We first compute the LLR $L_{ch}(y_i)$ as in Eq. 2.10. The hard decision vector is noted $\bar{Y} = [\bar{y}_1, \bar{y}_2 \ldots \bar{y}_n]$ and obtained from the the LLR signs. ChaseII algorithm follows:

1. Find the positions of the p least reliable bits.

   We note $i_1 \ldots i_p$ the positions of the p smallest LLR values in $L_{ch}(Y)$

2. Create a pattern of 2^p possible vectors $\bar{Y}^{(j)} = [\bar{y}_1^{(j)}, \bar{y}_2^{(j)} \ldots \bar{y}_n^{(j)}]$ with $j \in [0 \ldots 2^p{-}1]$.

   A pattern of binary vectors is created by flipping the bits of $\bar{Y}$ at the least reliable positions in all possible ways. Let us call $bin_p(j)$ the binary representation on p bits of the integer $j \in [0 \ldots 2^p{-}1]$. For instance: $bin_4(3)$=[0 0 1 1]. Then $\bar{Y}^{(j)}$ are created

such as:

$$\bar{Y}_i^{(j)} = \begin{cases} [\bar{y}_{i_1}^{(j)} \dots \bar{y}_{i_p}^{(j)}] = [y_{i_1} \dots y_{i_p}] + bin_p(j) \\ \\ y_i^{(j)} = y_i \qquad\qquad \text{For} \quad i \notin [i_1 \dots i_p] \end{cases} \tag{2.17}$$

3. Decode each vector of the pattern.

   The decoding is usually performed by a syndrome decoding and the decoded pattern $C^{(j)}$ is obtained

4. Compute the reliability of each valid codeword.

$$\Pr\{C^{(j)}|Y\} = \prod_{i=1}^{n} \Pr\{x_i = c_i^{(j)}|y_i\} \tag{2.18}$$

   Note that:

$$\Pr\{x_i = 0|y_i\} = \frac{\exp(L_{ch}(y_i))}{1 - \exp(L_{ch}(y_i))}$$
$$\Pr\{x_i = 1|y_i\} = \frac{1}{1 - \exp(L_{ch}(y_i))}$$

5. Compute the reliability of each bit.

   We call S_i^1 the set of codewords in the pattern having the bit $c_i^{(j)} = 1$ and S_i^0 the set of codewords having the bit $c_i^{(j)} = 0$. Finally the LLR of the estimated bits are obtained by:

$$L(\hat{x}_i) = \log \frac{\sum\limits_{C^{(j)} \in \mathrm{S}_i^0} \Pr\{C^{(j)}|Y\}}{\sum\limits_{C^{(j)} \in \mathrm{S}_i^1} \Pr\{C^{(j)}|Y\}} \tag{2.19}$$

2.1.6.2 Min-ChaseII algorithm

At high SNR, Eq. 2.19 can be simplified into [41]:

$$L_i \approx \max_{C^{(j)} \in \mathrm{S}_i^0} (\log \Pr\{C^{(j)}|Y\}) - \max_{C^{(j)} \in \mathrm{S}_i^1} (\log \Pr\{C^{(j)}|Y\}) \tag{2.20}$$

On the AWGN channel $\log \Pr\{C^{(j)}|Y\})$ depends on the Hamming distance between the received codeword $\bar{Y}$ and $C^{(j)}$. So the last two steps of ChaseII algorithm can be modified into:

4. Compute the weight of each valid codeword.

$$m^{(j)} = -\sum_{k=1}^{n} \bar{y}_k \oplus c_k^{(j)} \; |y_k| \tag{2.21}$$

5. Compute the reliability of each bit.

$$L_i \approx \max_{C^{(j)} \in \mathrm{S}_i^0} (m^{(j)}) - \max_{C^{(j)} \in \mathrm{S}_i^1} (m^{(j)}) \tag{2.22}$$

In the case of product code decoding, a scaling factor α_{sca} can be applied between each horizontal/vertical decoding to compensate the approximation made in Eq. 2.20.

2.2 FEC in optical fiber transmission systems

In Fig. 2.6, we summarize the implementation of the different FEC generations in optical transmission system.

2.2.1 1^{st} generation: Hamming, BCH and Reed-Solomon

One of the first implementation of a FEC in an optical fiber transmission system was realized by Grover in 1988 [42]. Hamming code $\mathcal{C}(224, 216)$ has been employed in a 565 Mb/s transmission system providing a 2.5 dB coding gain at BER= 10^{-13}. With the improvement of electronic equipments, more powerful codes such as BCH and Reed-Solomon codes started to be considered. In 1991, the BCH$(167, 151)$ was used in the Italian submarine system FESTONI offering a 2.5 dB gain at 10^{-10} [43]. The following year, Gabla realized a 622 Mb/s transmission over 401 km using the RS$(142, 126)$ and a 5 dB coding gain was obtained [44]. The same code is used by Pamart for a 5 Gb/s (7×622 Mb/s) transoceanic transmission over 6400km [45]. In the following years, Reed-Solomon codes become commonly used in optical transmissions and the RS$(255, 239)$ is specified in the norm ITU-T G975 [46].

2.2.2 2^{nd} generation: concatenated codes

With the development of 10 Gb/s WDM systems, RS codes are still used [47] but more efficient codes are needed to face the new types of fiber impairments such as XPM and PMD which decrease the performance and reduce the achievable transmission distances. To outperform the RS$(255, 239)$, concatenated schemes were investigated. These new codes could be easily implemented as they are based on the same linear block codes used in the first generation of FEC.

In 1999, O. Ait Sab proposed for the first time a concatenated code with 22% overhead performing 1.9 dB better than RS$(255, 239)$ [48]. Concatenated codes are then implemented in real WDM systems and experimental demonstration are made at 2.5 Gb/s [49],

10 Gb/s [50], 25 Gb/s [51] and 40 Gb/s [52]. Many combination of RS codes have been proposed with performance approximatively 2 dB better than that of RS(255, 239). In general overheads were chosen between 7% [50][52] and 23%[51][53].

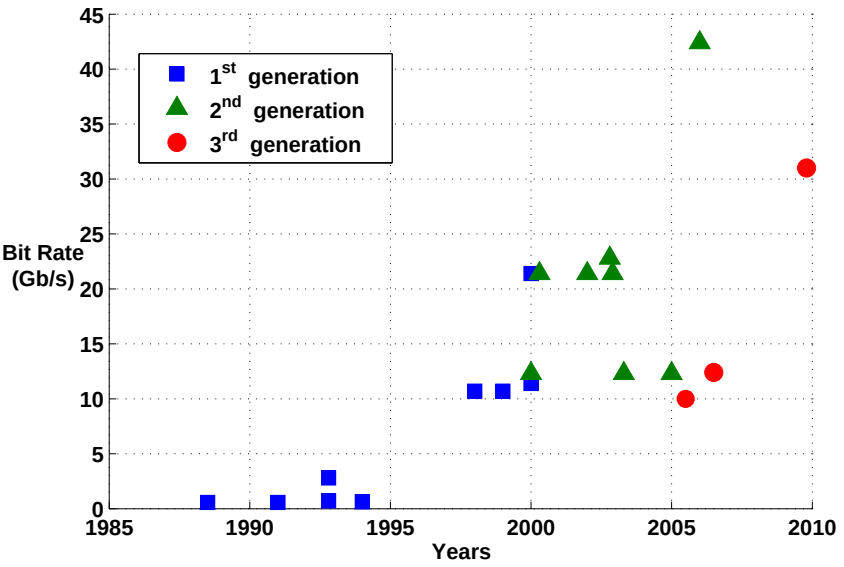


Figure 2.6: FEC implementations in optical fiber transmission systems

2.2.3 Modern coding

Since 2004, first 40 Gb/s systems are investigated and face severe impairments due to PMD. More powerful FECs were soon required and research focused on FECs with soft-input decoding algorithm. Two families of FEC are investigated: the Product codes an the LDPC codes.

In 2006, K. Ouchi et. al present the first demonstration of a Product code [54]. BCH(144, 128)× BCH(256, 239) has been implemented in a 10 Gb/s system bringing a 10.1 dB coding gain. This is 2 dB better than concatenated schemes. However this code as most of product codes, suffers from large overhead (23%) and large codeword length (37376 bits) which are not compatible with high bit-rate optical transmissions.

Djordjevic et.al have done a significant work on LDPC codes for optical transmissions. They have proposed many combinatorial constructions of LDPC codes [55][56][57] and validate them by simulation in various kind of optical systems. For instance, LDPC have shown their efficiency on a 40 Gb/s OOK transmission [58], a 100 Gb/s coherent transmission [59] and a 100 Gb/s OFDM transmission [60]. The proposed LDPC codes have a reduced overhead ($< 20\%$) and small codeword length (< 10000) and outperforms Product codes.

One of the major problem of LDPC code is their error floors. Indeed they appear around BER$= 10^{-10}$ for LDPC code having a girth 8 [61] whereas target BER are below 10^{-12}. To increase the girth and thus, reduce the error floor level, long codeword length and large overhead have to be considered. However this is not acceptable in high bit-rate transmissions. The proposed solution for practical implementation is to concatenate the LDPC code with a linear block code in order to remove the error floor. Mizuochi *et.al* have experimentally demonstrated for the first time a concatenation LDPC$(9216, 7936)$ + RS$(992, 956)$ at 31.3 Gb/s [62]. The observed coding gain is 9dB with 4 decoding iterations. This is 3.2dB better than RS$(255, 239)$.

Part II : Soft decoding FEC in optical transmission systems

Recently major efforts have been made to implement the new generation of FEC in 40G/100G transmission systems. The main difficulty comes from the very high complexity of their soft-input decoding. These FEC have been developed for wireless or wireline transmissions where bit-rates are lower and the transmission quality requirement is very different. For example, overhead of 50% can be considered in wireless transmissions whereas in optical systems, 7% has been for a long time the maximum admitted value. Moreover, target BER in optics are far smaller (i.e BER$= 10^{-12}$) and at such low values, soft decoding FEC may present some error floors.

The goal of this part is to propose a soft decoding FEC for optical transmissions systems. Product codes and LDPC codes are the two candidates for the future generation of FEC. Practical realizations have demonstrated a coding of about 2dB compared to concatenated schemes. However performance is not the unique criterion to base a choice between these two families. Indeed as high bit-rates optical transmission have limited resources, complexity is a main issue. Therefore we will first compare the decoding complexity of Product codes and LDPC codes.

In the following of this chapter, the simulations are realized considering a coherent polarization multiplexed transmission. We neglect the transmission nonlinearities and assume that digital signal processing perfectly compensate the signal distortion. Hence, the dominant source of noise at the FEC input is the ASE which can be modeled as an additive

gaussian noise. The noise is actually colored by bandpass filtering in our simulations, but in a matter of simplicity, the channel is modeled as an AWGN channel. This assumption is often realized in the literature [63] [64]. Indeed, with coherent detection, the optical noise is clearly transposed in the electrical domain and with direct detection, ASE noise results in a χ^2 distribution but can be approximated by a Gaussian distribution for typical FEC input BER (10^{-1}-10^{-3}).

2.3 LDPC Vs Product code

In this section, we evaluate the complexity of the soft input decoding algorithms of LDPC and Product codes presented in Sec. 2.1.4.2 and 2.1.6.2. The complexity will be evaluated by the number of logical operation required to decode a codeword. We note $\oplus$ and $\otimes$ respectively the binary addition and multiplication. In Tab. 2.1 and Tab. 2.2, we have summarized the XOR and AND logical operations. When soft inputs are considered, the received symbol is quantified on q bits . We note $<$ the comparison operation and $+$ the additive operation between quantified values.

Table 2.1: XOR

0	0	0
1	0	1
0	1	1
1	1	0

Table 2.2: AND

0	0	0
1	0	0
0	1	0
1	1	1

2.3.1 LDPC decoding complexity

Let us consider a regular LDPC code $\mathcal{C}(n, l, c)$. We note n_v the number of columns (the number of variable nodes) and n_c the number of lines (the number of check nodes) of the parity-check matrix.

The MSA is a soft input iterative decoding and each iteration occurs in two steps. The check node step is given by equation Eq. 2.15. The output messages μ_{c_m,v_k} are function of the sign product $\prod_{i\in\mathcal{S}_{c_m}\setminus v_k} \text{sgn}(\mu_{i,c_m})$ and function of the minimum LLR. Instead of looking for the minimum input LLR at each time an output is computed (this would mean searching l times the minimum), one would rather compute only once the two minimal inputs $|\mu_{min1}|$ and $|\mu_{min2}|$ (with $|\mu_{min1}| \leq |\mu_{min2}|$). This requires $2l-3$ comparisons. We note v_{min} the variable node corresponding to the minimum input LLR. On the other hand, the sign products can be easily obtained from π_{v_k} the product of all input signs. Indeed

$$\prod_{i\in\mathcal{S}_{c_m}\setminus v_k} \text{sgn}(\mu_{i,c_m}) = \prod_{i\in\mathcal{S}_{c_m}} \text{sgn}(\mu_{i,c_m}) \times \text{sgn}(\mu_{v_k,c_m}) \tag{2.23}$$

Therefore the output message can be efficiently evaluated by:

$$\pi_{v_k} = \prod_{i \in \mathcal{S}_{c_m}} \text{sgn}(\mu_{i,c_m}) \tag{2.24}$$

$$\mu_{c_m,v_k} = \begin{cases} \pi_{v_k} \times \text{sgn}(\mu_{v_k,c_m}) \times \mu_{min1} & \text{If } v_k \neq v_{min} \\ \pi_{v_k} \times \text{sgn}(\mu_{v_k,c_m}) \times \mu_{min2} & \text{If } v_k = v_{min} \end{cases} \tag{2.25}$$

Sign products are done by $\oplus$ operations on the sign bits. Therefore we need $l-1$ $\oplus$ to have π_{v_l} and l $\oplus$ to obtain every outputs.

The variable node step is expressed in Eq. 2.13 and Eq. 2.14. The LLR of the estimated bits $L(\hat{x}_i)$ are obtained by the addition of the c input messages and the channel LLR $L_{ch}(\hat{x}_i)$. To compute the output messages we can just subtract from $L(\hat{x}_i)$ the corresponding input message:

$$\mu_{v_l,c_k} = L(\hat{x}_l) - \mu_{c_k,v_l} \tag{2.26}$$

We need $c-1$ $+$ operations to get $L(\hat{x}_l)$ and then c $+$ operations to obtain the output.

We have evaluated the number of operations at each check and variable node thus in order to obtain the total complexity of an iteration, we have to multiply it by the number check and variable nodes. The total number of operations required for one iteration of the Min-Sum algorithm is summarized on Tab. 2.3.

Table 2.3: Decoding complexity of one iteration of the MSA

operators	MSA
$<$	$n_c.(2l-3)$
$\oplus$	$n_c.(2l-1)$
$+$	$n_v.(2c-1)$

2.3.2 Product code decoding complexity

Let us evaluate the decoding complexity of the min-ChaseII algorithm for a single codeword decoding. We consider a block linear code $\mathcal{C}(n,k)$ which is able to correct t_{max} errors on a codeword. The inputs of the decoder are LLR derived from the received vector Y. First we obtain the hard decision vector $\bar{Y}$ by checking the sign of $L_{ch}(y_i)$. It requires n comparisons. Then the p least reliable bits are chosen with $p.n$ comparisons and the pattern of 2^p vector is created with $2^p.\frac{p}{2}$ $\oplus$ operations. We consider a decoding of the pattern by syndrome decoding (see Section. 2.1.1). This is possible for small values of t_{max}, typically $t_{max} \leq 2$. $2^p.n.(n-k)$ $\otimes$ operations and $2^p.t_{max}$ $\oplus$ are necessary to compute the syndromes and correct the pattern vectors. Then the weights of the valid codewords are computed following Eq. 2.21 which requires $2^p.n$ $+$ operations. Finally the reliabilities are obtained by Eq. 2.22. We need $2^p.n$ comparisons for the two max functions and n $+$

operations for the subtractions. Note that in the normalized version of the algorithm, a scaling factor is applied on the extrinsic information thus n $\times$ operations and n $+$ operations have to be added. The total number of operations in a ChaseII algorithm is presented on Tab. 2.4.

Table 2.4: Decoding complexity of min-ChaseII algorithm

operators	min-ChaseII
$<$	$n.(2^p + p + 1)$
$\oplus$	$2^p.(\frac{p}{2} + \mathrm{t}_{max})$
$\otimes$	$n.(n-k).2^p$
$+$	$n.(2^p + 1)$

The number of operations required for the soft decoding of a BCH$(255, 239)$ is presented on Tab. 2.5 for two values of p (2 different pattern sizes).

Table 2.5: min-ChaseII on BCH$(255, 239)$

operators	$p = 3$	$p = 5$
$<$	3 060	9 690
$\oplus$	28	144
$\otimes$	32 640	130 560
$+$	2 295	8 415

It has been seen in Sec. 2.1.6, that one iteration of the decoding of Product codes occurs in two steps. First the decoding of n_2 codewords by the inner FEC and then the decoding of n_1 codewords by the outer FEC. For a soft version of the decoding, each codeword is decoded using the min-ChaseII algorithm. The operations involved for one iteration of the soft decoding of Product codes $\mathcal{C}_1(n_1, k_1, t_{max,1}) \times \mathcal{C}_2(n_2, k_2, t_{max,2})$ are summarized on Tab. 2.6.

Table 2.6: Decoding complexity of one iteration of $C \in \mathcal{C}_1(n_1, k_1, t_{max,1}) \times \mathcal{C}_2(n_2, k_2, t_{max,2})$ with min-ChaseII algorithm

operators	Product code soft decoding
$<$	$n_2.(n_1.(2^p + p + 1)) + n_1.(n_2.(2^p + p + 1))$
$\oplus$	$n_2.(2^p.(\frac{p}{2} + t_{max,1})) + n_1.(2^p.(\frac{p}{2} + t_{max,2}))$
$\otimes$	$n_2.(n_1.(n_1 - k_1).2^p) + n_1.(n_2.(n_2 - k_2).2^p)$
$+$	$n_2.(n_1.(2^p + 1)) + n_1.(n_2.(2^p + 1))$

2.3.3 LDPC and Product codes decoding complexity comparison

In order to compare Product and LDPC codes decoding complexity, we chose codes with equivalent codeword length n and rate r. The complexity is evaluated for different values

of p in the case of Product codes and for different couples (l, c) in the case of LDPC. We denote by $\mathcal{C}_1(n_1, k_1, t_{max,1}) \times \mathcal{C}_2(n_2, k_2, t_{max,2})$ the product codes and by (n_v, n_c, l, c), the LDPC codes.

Table 2.7: $n = 36720, r = 0.83$

	BCH$(255, 239) \times$ BCH$(144, 128)$		LDPC $(36720, 6120, l, c)$	
	$p = 3$	$p = 5$	l=18, c=3	l=60, c=10
$<$	881 280	2 790 720	201 960	716 040
$\oplus$	11 172	57 456	214 200	728 280
$\otimes$	9 400 320	37 601 280	0	0
$+$	660 960	2 423 520	183 600	697 680

Table 2.8: $n \approx 11400, r = 0.87$

	BCH$(107, 100)^2$		LDPC $(11472, 1434, l, c)$	
	$p = 3$	$p = 5$	l=24, c=3	l=48, c=6
$<$	274 776	870 124	64 530	133 362
$\oplus$	4 280	23 968	67 398	136 230
$\otimes$	1 282 288	5 129 152	0	0
$+$	206 082	755 634	57 360	126 192

Table 2.9: $n \approx 4800, r = 0.79$

	BCH$(64, 57)^2$		LDPC $(4788, 1024, l, c)$	
	$p = 3$	$p = 5$	l=14, c=3	l=82, c=18
$<$	98 304	331 296	25 650	165 186
$\oplus$	2 560	14 336	27 702	167 238
$\otimes$	458 752	1 835 008	0	0
$+$	73 728	270 336	23 940	167 580

We present in Tab. 2.7, Tab. 2.8 and Tab. 2.9, the number of operation $\oplus$, $\otimes$, $+$, $<$ for the decoding of one iteration of both kind of codes. In Tab. 2.7, we consider codes with size n=36720 and rate 0.83. In Tab. 2.8, we consider codes with size n=11400 and rate 0.87. In Tab. 2.7, we consider codes with size n=4800 and rate 0.79.

We can see that for Product code the choice of the parameter p has a strong influence on the complexity. For instance, there is about four times more $+$ operations with p=5 than with $p = 3$. Indeed the size of the pattern increases exponentially with p and syndrome computation and reliabilities have to be computed for each valid codeword of the pattern. Therefore there is a tradeoff to make between performance and complexity.

For LDPC, we have compared two codes having the same size and the same rate but with different sparseness. The sparseness is defined as the number of one in the parity-check matrix:

$$spar = \frac{lc}{n_c n_v} \tag{2.27}$$

From Tab. 2.3 we know that the number of operation increases linearly with c and l. For example, in Tab. 2.8, we have compared two LDPC codes with $c = 3$ and $c = 6$ and there is a ratio of 2 between the number of operations $+$ of each code. Note that sparseness is related to code performance as it can result to less cycles in the Tanner graph. So choosing sparse code can improve the performance and reduce the decoding complexity at the same time.

We can see through this comparison that MSA has a lower complexity than min-ChaseII algorithm. We notice that min-ChaseII requires a large number of $\otimes$ operations for the syndrome decoding of the pattern whereas MSA does not require any. Tab. 2.8 shows that even with a LDPC with low sparseness, the number of operations $\otimes$, $\oplus$ and $+$ required for the MSA remains smaller than that of the min-Chase with $p = 3$.

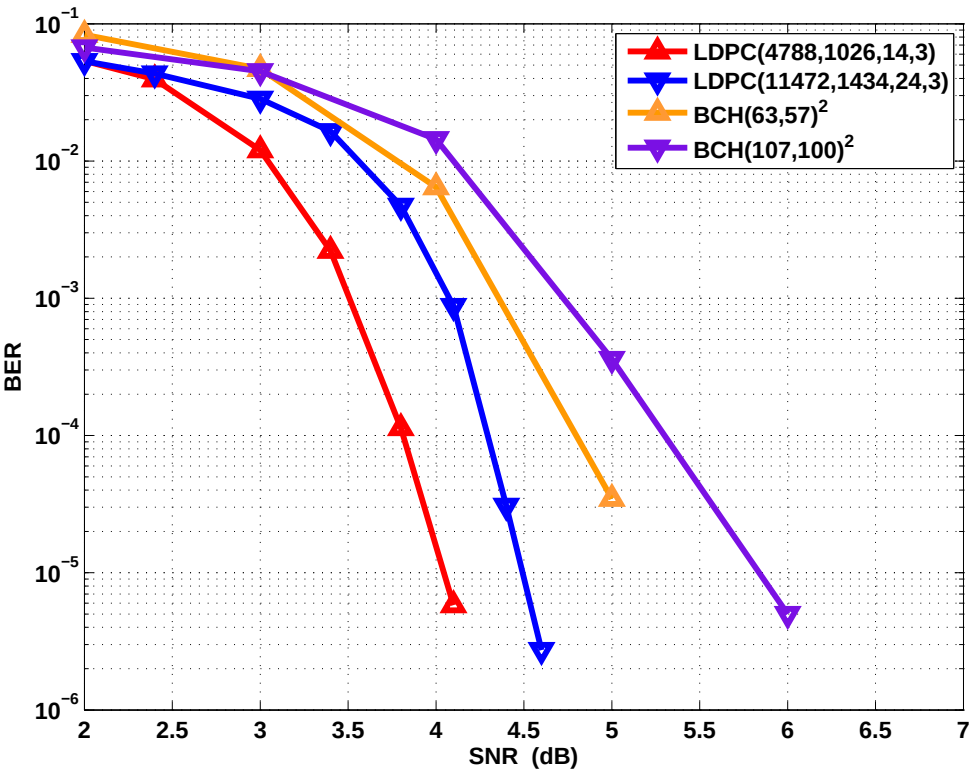


Figure 2.7: Performance comparison between LDPC and Product codes

To complete the comparison, we present in Fig. 2.7, the performance between LDPC codes and Product codes having the same rate and codeword length. The up triangle markers represent codes with $n \approx 4800$ and rate 0.79 whereas the down triangle markers represent codes with $n \approx 11400$ and rate 0.87.
We observe that LDPC codes with the MSA decoding perform better than the Product codes with min-ChaseII algorithm. LDPC $(4788, 1024, 14, 3)$ outperforms BCH$(64, 57)^2$ by 1dB and LDPC $(11472, 1434, 24, 3)$ outperforms BCH$(107, 100)^2$ by 1.5dB at a BER$\approx 10^{-5}$.

2.4 LDPC constructions

From the study made in the previous section, it seems that LDPC are the most appropriate candidates for the next FEC generation in optical communications. They perform better than Product code and their decoding complexity is smaller. In this section we focus on the construction of LDPC codes adapted for high bit-rate transmissions. Indeed, in order to make easier the encoding and the decoding, it is important to have structured LDPC. For example, if the parity-check matrix is quasi-cyclic (QC), the encoding can be implemented with shift-registers and the decoding with parallel circulant architectures.
We are going to present some LDPC constructions proposed in the literature. All are based on algebraic constructions and result to structured parity-check matrix. However their performance is limited due to the construction constraints. We propose at the end of this section, an original random construction of LDPC having the triple advantages to be very flexible, to create QC parity-check matrix and to outperform algebraic constructions.

2.4.1 Algebraic constructions

Most of the codes proposed for optical communications systems are structured codes. Their constructions are based on algebraic objects such as cycle invariant difference set (CIDS), orthogonal arrays (OA) or balanced incomplete block design (BIBD). Combinatorial constructions usually ensure a girth of at least 6. However, these constructions are in general not very flexible and impose some constraints on the design parameters of the code.

2.4.1.1 OA-LDPC

OA-LDPC have been introduced by Djordjevic et. al [65] and are based on orthogonal arrays. An orthogonal array of size n, with m constraints, q levels, strength t, and index λ is denoted λ-OA(n, m, q, t). It is defined as a $m \times n$ matrix $\mathbf{A}$ with entries from a set of q elements such that in any $t \times n$ sub-matrix of $\mathbf{A}$, every $t \times 1$ column vectors are contained λ times.

OA-LDPC construction is based on 1-OA$(q^2, q+1, q, 2)$ in order to be free of size-4 cycles. First all the polynomials $P_k \in GF(q)[X]$ of degree ≤ 1 are determined. There are q^2

polynomials ($1 \leq k \leq q^2$). Let α be the primitive element of $GF(q)$ hence $a_1 = 0, a_2 = 1, a_3 = \alpha \ldots a_q = \alpha^{q-2}$ are all the elements of $GF(q)$. The orthogonal array $\mathbf{A}$ is constructed such as $\mathbf{A}(i,j) = P_j(a_i)$. Finally a last line is added to the matrix in which each entry $\mathbf{A}(q+1,j)$ is equal to the major coefficient of P_j.
Let us call $p^{(1)}_{a_k,i} \ldots p^{(q)}_{a_k,i}$ positions of the element a_k in the i^{th} line of $\mathbf{A}$. The set $B_{a_k,i}$ is defined by associating to every positions $p^{(x)}_{a_k,i}$ a pair $\{\mathbf{A}(k-1, p^{(x)}_{a_k,i}), \mathbf{A}(k, p^{(x)}_{a_k,i})\}$:

$$B_{a_k,i} = \left\{ \left(\mathbf{A}(k-1, p^{(1)}_{a_k,i}), \mathbf{A}(k, p^{(1)}_{a_k,i}) \right) \ldots \left(\mathbf{A}(k-1, p^{(q)}_{a_k,i}), \mathbf{A}(k, p^{(q)}_{a_k,i}) \right) \right\} \tag{2.28}$$

Each pair can be transformed by a linear operation $\{a_u, a_v\} \rightarrow l = v + (u-1)q$ and it results to the new set $\tilde{B}_{a_k,i}$:

$$\tilde{B}_{a_k,i} = \left[l^{(1)}_{a_k,i}, l^{(2)}_{a_k,i} \ldots l^{(q)}_{a_k,i} \right] \tag{2.29}$$

With $1 \leq l^{(x)}_{a_k,i} \leq q^2$. The set $\tilde{B}_{a_k,i}$ is called block and defines the position of the 1 in one line of the parity-check matrix. As we keep the two last lines of $\mathbf{A}$ as index, i can be chosen from 1 to $q-1$, moreover there are q elements in $GF(q)$ so $q.(k-1)$ blocks can be created. Hence, $\boldsymbol{H}$ is at maximum a $q.(k-1) \times q^2$ matrix. An example of OA-LDPC construction is given in Appendix. A

OA-LDPC have quasi-cyclic parity-check matrix. Moreover, this construction can easily produce codes with high rates and good performance. To obtain a larger range of design parameters, Djordjevic proposed in [66] to combine two orthogonal arrays.

2.4.1.2 Array LDPC

In the following we will consider quasi-cyclic parity-check matrix having a block -circulant structure.

$$\boldsymbol{H} = \begin{bmatrix} \mathbf{P}^{i_{1,1}} & \mathbf{P}^{i_{2,1}} & \ldots & \mathbf{P}^{i_{n_{bc},1}} \\ \mathbf{P}^{i_{1,2}} & \mathbf{P}^{i_{2,2}} & & \vdots \\ \vdots & & \ddots & \vdots \\ \mathbf{P}^{i_{1,n_{bl}}} & \mathbf{P}^{i_{2,n_{bl}}} & \ldots & \mathbf{P}^{i_{n_{bc},n_{bl}}} \end{bmatrix} \tag{2.30}$$

where $\mathbf{P}^i$ are $p \times p$ circulant permutation matrices. For example $\mathbf{P}$ can be the permutation of the identity matrix.

$$\mathbf{P} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{2.31}$$

The code is noted (p, n_{bc}, n_{bl}) where n_{bc}, n_{bl} are respectively the number of block-columns and block-lines. The codeword length is $n = p \times n_{bc}$ and the rate is $r \approx 1 - n_{bl}/n_{bc}$.

A cycle of size $2g$ exists in the parity-check matrix if there is a set of index $i_{c_k,v_m} \in \{1 \dots p\}$ such as [67]:

$$i_{c_1,v_1} + i_{c_2,v_2} + \dots + i_{c_g,v_g} = i_{c_2,v_1} + i_{c_3,v_2} + \dots + i_{c_1,v_g} \tag{2.32}$$

where $c_k \in \{1 \dots n_{bl}\}$ and $v_m \in \{1 \dots n_{bc}\}$.
For instance, a cycle of size 4 exists if there are two pairs of indexes which have the same difference:

$$i_{c_1,v_1} - i_{c_2,v_1} = i_{c_1,v_2} - i_{c_2,v_2} \tag{2.33}$$

2.4.1.3 CIDS-LDPC

CIDS-LDPC have been introduced by Milenkovic et. al [57]. The construction is based on algebraic objects called cycle invariant difference sets (CIDS).
A general difference set $S(p,k,\lambda)$ over an additive Abelian group $\mathcal{P}$ of order p, is a set $S = \{s_1, s_2 \dots s_k\}$ of distinct elements from $\mathcal{P}$, such that each nonzero element can be represented as $s_{i_1} = s_{i_2} - s_{i_3}$ in at most λ ways. If $\lambda = 1$, each element is represented as the difference of two other elements in only 1 way and thus according to Eq. 2.33, there are no cycles of size 4.

Let S be a difference set over $\mathbb{Z}_N$, arranged in order. We note C^i the operator that cyclically shift a sequence of i positions. If $\Omega^i = C^i S - S \mod N$ are difference sets for $i \in \{1 \dots m\}$, S is a $(m+1)$-fold cyclic invariant difference set.

Let α be the primitive element of $GF(q^4)$ where q is an odd prime. A q-fold CIDS(q^2–1, q, 1) can be constructed as:

$$S = \left\{a \;\middle|\; 0 \le a \le q^4 - 1, \quad \alpha^a - \alpha \in GF(q)\right\} \tag{2.34}$$

However this results to very large number a, even with small value of q. Therefore it has been proposed to construct the set:

$$\tilde{S} = \left\{a \;\middle|\; 0 \le a \le q^2 - 1, \quad \alpha^a - \alpha \in GF(q)\right\} \tag{2.35}$$

where α is the primitive element of $GF(q^2)$. $\tilde{S}$ is not a CIDS but by permuting and erasing the appropriate elements, a CIDS can be obtained. This later construction results in a larger set of cardinality than that of Eq. 2.34.
CIDS-LDPC have a $(m+1).p \times s_k.p$ parity-check matrix defined by:

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}^{i_1} & \mathbf{P}^{i_2} & \dots & \dots & \mathbf{P}^{i_{s_k}} \\ \mathbf{P}^{i_{s_k}} & \mathbf{P}^{i_1} & & & \vdots \\ \vdots & & \ddots & & \vdots \\ \mathbf{P}^{i_{s_k-m+1}} & \dots & \dots & \dots & \mathbf{P}^{i_{s_k-m}} \end{bmatrix}$$

where $\mathbf{P}$ is a $p \times p$ circulant permutation matrix and the exponent $i_1, i_2 \dots i_{s_k}$ belong to a $(m+1)$-fold CIDS (p,k,λ).

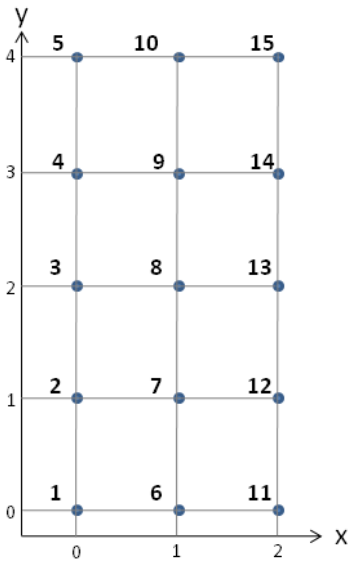


Figure 2.8: Rectangle grid

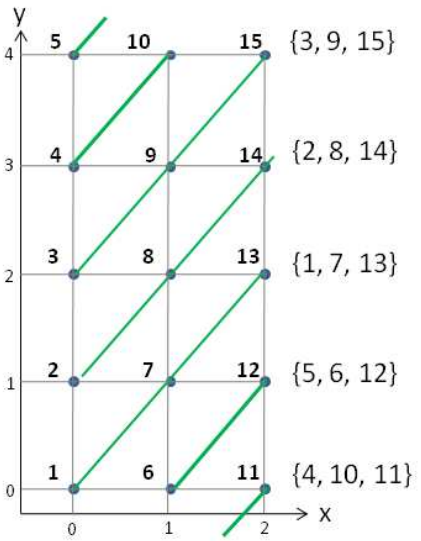


Figure 2.9: Block construction

CIDS-LDPC have usually very good performance. Their parity-check matrix is very sparse, which means that there are few cycles of size 6 compared to the other constructions. The main drawback of this construction is the very limited choice of q. Indeed it has to be an odd prime and the size of the sets $GF(q^2)$ or $GF(q^4)$ grow exponentially. Hence, we can only construct a limited number of codes and it is not possible to choose freely their design parameters (codeword length, code rate...).

2.4.1.4 Lattice LDPC

Lattice LDPC are constructed from balanced incomplete block design (BIBD) obtained from a lattice of points [68]. A BIBD is a pair (V, B) where V is a set of elements called points and B is a collection of subsets of V called blocks. Each block has k points. We note v the size of the set V. t and λ are defined such that every subsets of t points from V appear in exactly λ blocks.

A block corresponds to the positions of the 1 in one column of the parity-check matrix therefore we want $t=2$ and $\lambda=1$ in order to avoid cycles of size 4.

A BIBD is obtained from a rectangle grid of size $m \times k$ as in Fig. 2.8. The points of coordinates (x, y) with $0 \leq x \leq m-1$ and $0 \leq y \leq k-1$ are mapped to the element of V by $(x, y) \longrightarrow y+1+m.(x-1)$. We have $v = m \times k$. The blocks of B are obtained considering the lines of the grid. Every lines are passing through k points and correspond to blocks as in Fig. 2.9. Lines can have slopes from 0 to $m-1$ and in total m^2 blocks can be created, each one having k elements of V. Therefore H is a $(k.m) \times m^2$ matrix. Moreover if there is no integer $i \leq k$ being a factor of m, we have $t=2$ and $\lambda=1$.

2.4.2 LDPC based on Quasi-cyclic PEG

Combinatorial constructions like the ones presented in the previous sections have been proposed for optical transmissions because they produce regular structured parity-check matrix. However their construction criterion does not intend to optimize the performance. Indeed, performance is closely related to the size and number of short-cycles in the Tanner graph but these combinatorial constructions only guarantee size-4 cycle free graph.
By relaxing the constraints (structured parity-check matrix, regular code), codes performing very close to the channel limit have been obtained [69]. They are usually designed based on random constructions with the aim to avoid short-cycles. Let us now present one example of such LDPC.

2.4.2.1 Progressive edge growth algorithm

The progressive edge growth algorithm (PEG) [70] is a random construction of parity-check matrix based on the Tanned graph representation. The idea is to construct the edges (the 1 in the matrix) one by one, in order to avoid short cycles.

We define $\{v_1, v_2 \dots v_{n_c}\}$ the set of variable nodes, $\{c_1, c_2 \dots c_{n_l}\}$ the set of check nodes. $\mathcal{S}_{v_k}$ is the set of check nodes connected to the variable node v_k and $\bar{\mathcal{S}}_{v_k}$ the set of check nodes not connected to v_k. We have $\{v_1 \dots v_{n_c}\} = \mathcal{S}_{v_k} \cup \bar{\mathcal{S}}_{v_k}$.

A given variable node is initially connected to none of the check nodes ($\mathcal{S}_{v_k} = \emptyset$). Then d_v edges are created one by one in order to connect v_k to some of the check nodes. It means that these check nodes are added to the set until $|\mathcal{S}_{v_k}| = d_v$. Introducing a check node c_m in the set corresponds to the creation of an edge E_{c_m,v_k}.
Check nodes are chosen in $\bar{\mathcal{S}}_{v_k}$. However some of them lead to the creation of cycles. Therefore, we have to find the ones not creating cycles, or else, the ones introducing the largest cycles. Note that the first check node introduced in $\mathcal{S}_{v_k}$ can not create a cycle and thus, can be chosen randomly in $\bar{\mathcal{S}}_{v_k}$.

In order to find the most appropriate check nodes, we expand the tree graph from v_k. The idea is to find which are the check nodes already connected to v_k and at which distance they are. The check nodes directly connected to v_k are at a depth 1, then the following one are at the depth 3 etc. We say that c_m is at a depth d if there are d distinct edges between it and v_k. Therefore if an edge E_{c_m,v_k} is built, this will create a cycle of size $d+1$. In Fig. 2.11, an example of the tree expansion of a node from Fig. 2.10 is given.
Before the tree expansion we initialize a set $\mathcal{A}_{v_k}^{(0)} = \bar{\mathcal{S}}_{v_k}$ corresponding to all the check nodes candidates. The expansion is performed by steps and at a step l, it goes to a depth $2l-1$. We denote D_d the set of check nodes at the depth d from v_k. The set $\mathcal{A}_{v_k}^{(l)} = \mathcal{A}_{v_k}^{(l-1)} \backslash D_{2l-1}$ is created. It means that if a check node from $\mathcal{A}_{v_k}^{(l)}$ is chosen, there will be no cycle of size $2l$. The tree expansion stops when $\mathcal{A}_{v_k}^{(l)} = \emptyset$ (there is no check node further than $d = 2l+1$)

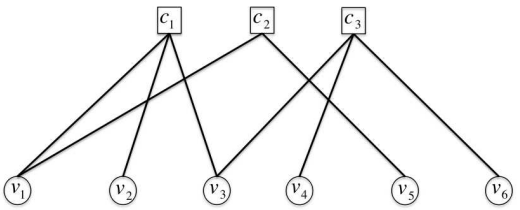


Figure 2.10: Example of Tanner graph

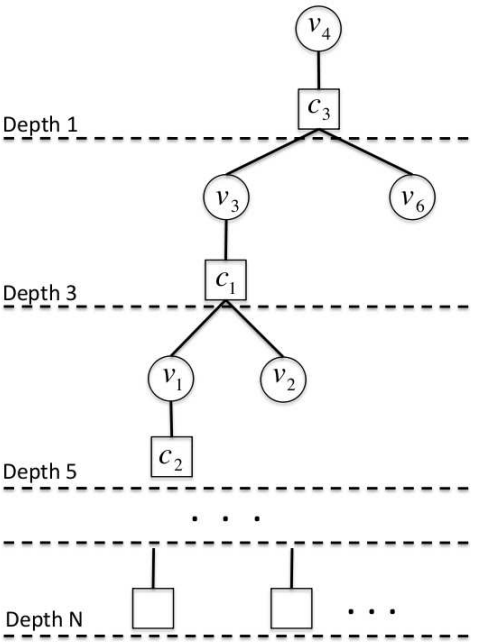


Figure 2.11: Tree expansion

or $\mathcal{A}_{v_k}^{(l)} = \mathcal{A}_{v_k}^{(l-1)}$ (all the check nodes connected to v_k have been found) and c_m is finally chosen in $\mathcal{A}_{v_k}^{(l-1)}$. In order to have an uniform check node degree distribution, the check node having the lowest degree is selected.
Note that PEG algorithm do not always ensure a constant check node degree and thus the parity-check matrix may be irregular. PEG algorithm can be summarized in Algorithm. B.

2.4.2.2 Quasi-cyclic LDPC construction based on PEG algorithm

We propose to modify the PEG algorithm in order to obtain an array LDPC. It means that the parity-check matrix should have the form of Eq. 2.30. PEG is based on a graph structure so we need to introduce the notion of labeled Tanner graph of an array code. We define it as the bipartite graph with n_{bl} block check nodes and n_{bc} block variable nodes where v_k and c_m are connected if there is a permutation matrix $\mathbf{P}^i$ in the k^{th} block column and m^{th} block line of the parity-check matrix. Moreover, every edges have a metric corresponding to the exponent of the associated permutation matrix. An example of labelled Tanner graph is presented in Fig. 2.12.
If there is no null matrix in the parity-check matrix, every variable nodes are connected to every check nodes. Hence, there are a lot of cycles in the labeled Tanner graph but they result in cycles in the regular Tanner graph only if the metrics satisfy Eq. 2.32. Note that if $\mathbf{P}$ are $p \times p$ matrices, a cycle in the labelled Tanner graph corresponds to p cycles in the classic Tanner graph .

PEG algorithm is performed on the labeled graph. For a given variable node v_k, an edge

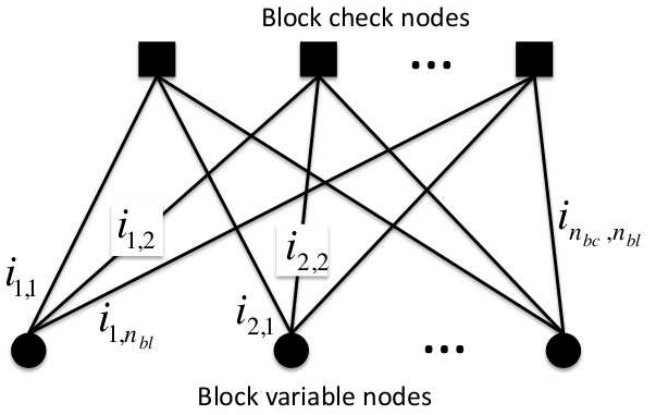


Figure 2.12: Example of labeled Tanner graph

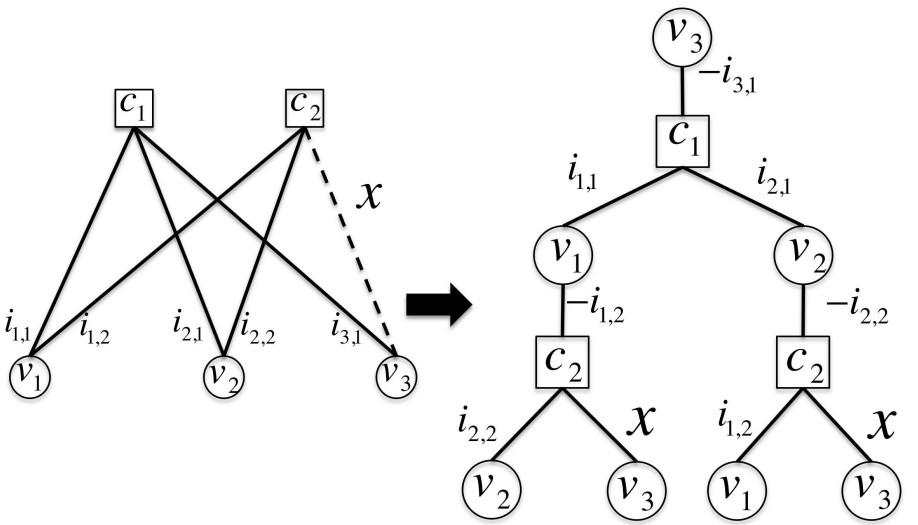


Figure 2.13: Example of Tanner graph

is created with one of the available check node c_m and a metric x is associated. x can be chosen in $[1 \dots p]$ but some of these values may lead to the creation of short cycles (if there is a cycle of metrics satisfying Eq. 2.32).

A tree expansion is realized in order to find the metric x leading to the largest cycles. The principle is depicted in Fig. 2.13 where the tree expansion of v_3 is realized in order to find the metric of the edge E_{c_2,v_3}. In the same way as in the previous section, we define a set $\mathcal{M}_{v_k}^{(l)}$ of the metric which does not create cycles of size $2l$. Initially $\mathcal{M}_{v_k}^{(0)} = [1 \dots p]$. At each new step l of the expansion, the tree is expanded until the depth $2l$ and the cumulated metrics are computed for every paths. Unlike the regular PEG, we here expand the tree until variable nodes. But as the edge E_{c_m,v_k} has been already created, several paths may arrived to v_k passing by E_{c_m,v_k}. These paths have cumulated metrics depending on x. $\mathcal{M}_{v_k}^{(l)}$

is obtained by removing from $\mathcal{M}_{v_k}^{(l-1)}$, the values of x that make the cumulated metric being equal to zero (that produce cycles of size $2l$). The procedure stops if the tree can not be expanded anymore or if $\mathcal{M}_{v_k}^{(l)} = \emptyset$ and finally $i_{m,k}$ is chosen in $\mathcal{M}_{v_k}^{(l-1)}$.
Quasi-cyclic PEG construction is summarized by Algorithm. B. In Fig. 2.13 the tree-graph has been expanded until a depth 4 ($l = 2$). There are size-4 cycles if $x = i_{3,1} - i_{2,1} + i_{2,2}$ or $x = i_{3,1} - i_{1,1} + i_{1,2}$. The current subset is defined as $\mathcal{M}_{v_3}^{(2)} = \mathcal{M}_{v_3}^{(1)} \backslash \{i_{3,1} - i_{2,1} + i_{2,2},\ \ i_{3,1} - i_{1,1} + i_{1,2}\}$.

2.4.2.3 QC-PEG LDPC performance

The QC-PEG algorithm is very flexible and we can chose the design parameters of the code with no constraints but a rate relation. Indeed to have a LDPC code with a codeword length n and a redundancy r, we are free to choose n_{bl} and p such as: $r \times n = n_{bl} \times p$.
In Fig. 2.14, we plot the performance in terms of BER of some LDPC codes proposed in optical communications (dashed lines) and QC-PEG LDPC codes (solid lines). CIDS-LDPC and lattice-based LDPC are both array codes therefore the comparison is made using the same parameters p , n_{bc} and n_{bl}. For OA-LDPC and EG-LDPC [71] the comparison is made choosing equivalent codeword length and redundancy. For every codes, the decoding is performed using the Min-Sum algorithm with 15 iterations.
QC-PEG construction perform better or at least as good as all the algebraic constructions. In our simulations, we observe that a QC-PEG LDPC outperform the CIDS(4608,3477) by 0.5dB, the lattice(3890,2989) by 0.2dB, the OA(8232,7560) by 1dB, the OA(4096,3510) by 1.4dB and the EG($2,2^5$) by 1dB at a BER of 10^{-6}.
This is explained by the fact that our codes have either higher girth or smaller number of short cycles. For instance, our construction achieves a girth-8 when CIDS(4608,3477) and lattice(3890,2989) only have a girth-6. OA(8232,7560) and QC-PEG(243,34,2) both have a girth 6 but the OA-LDPC has one thousand time more size-6 cycles than our code and performs 1 dB worse at 10^{-6}.
The CIDS(4320,3242)* [57] and CIDS(16845,13476)* [72] were obtained from the optimization of CIDS codes. A particular sequence has been extracted from the set of index obtained by the CIDS construction. The QC-PEG performs the same as these two codes until 10^{-6} but it has slightly less short-cycles.

As girth and number of short cycle are the main performance criteria, we represent them in Fig. 2.15 for several QC-PEG code having different redundancies. We construct QC-PEG codes with $n \approx 4000$ and $n_{bl} = 3$ for redundancies from 6% to 50% and we have counted the number of shortest cycles in their Tanner graph. With these parameters, our construction achieves girth at least 8 for redundancies $\geqslant 15\%$. With combinatorial constructions, it is very difficult to obtain such a small girth with these rates.
In Fig. 2.16 we plot the number of shortest cycle of lattice and CIDS LDPC. Lattice construction is flexible and codes with various redundancies can be easily created. In the case of CIDS codes, it is more difficult to construct codes with various redundancies for

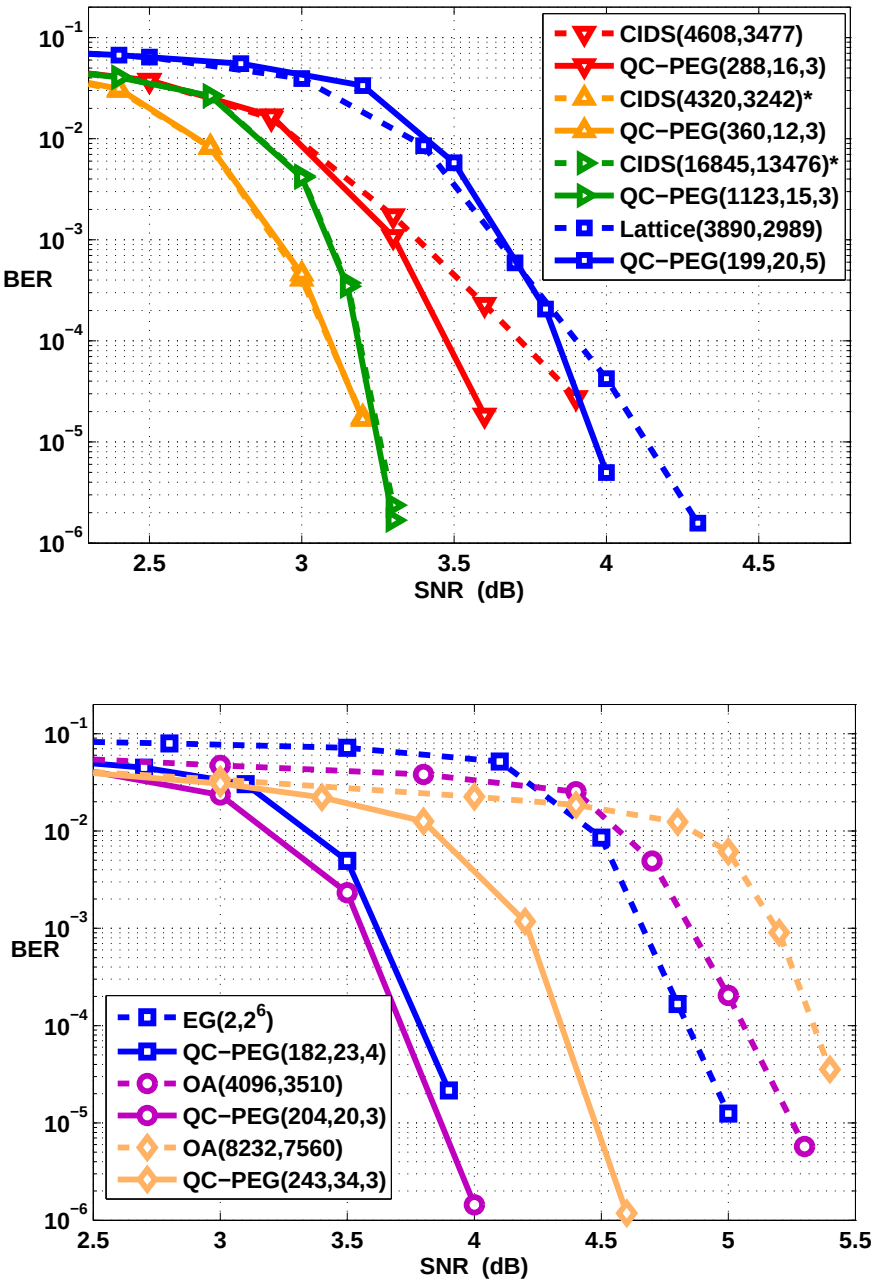


Figure 2.14: Performance comparison between some LDPC codes and the equivalent QC-PEG codes.

given parameters n and r. Here only 3 codes could be constructed based on the CIDS $(q^2-1, q, 1)$ with $q = \{17, 19, 23\}$. From these sets, a first version of CIDS codes have been obtained using the parity-check matrix shape presented in section. 2.4.1.3.

We observe that indeed, algebraic constructions lead to girths equal to 6 (square markers). However higher girths are never obtained even for large redundancies. A second version of

CIDS parity-check matrix has been proposed in [72] having girth equal to 8 (down triangle marker) with our design parameters. In the figure, the black lines represent the QC-PEG codes. We observe that the girth of our codes are usually higher than that of algebraic constructions. Otherwise the number of short cycles is far smaller in QC-PEG LDPC codes.

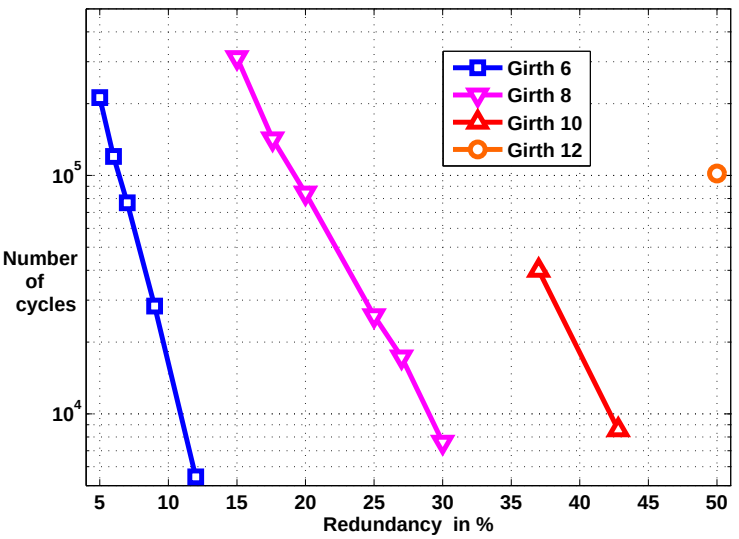


Figure 2.15: Number of cycles of QC-PEG codes for $n \approx 4000$ and $n_{bl} = 3$.

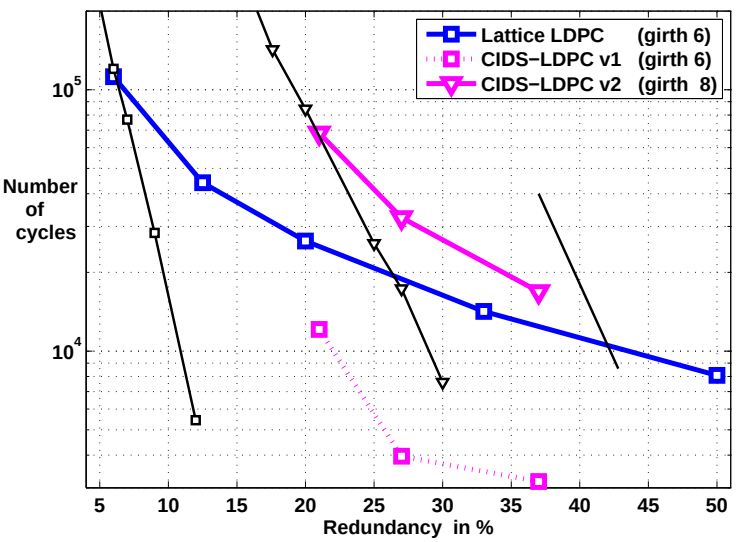


Figure 2.16: Number of cycles of Lattice and CIDS LDPC codes for $n \approx 4000$ and $n_{bl} = 3$.

Conclusion

Forward error correction is a major component of every optical transmissions. However the FEC used in practical systems are not the ones having the best performances. In wireless transmissions, LDPC and Turbo codes are now commonly used (in UMTS, WiMAX, LTE) because of their very good performance. However in optical communication their implementation is a real challenge because of the complexity of their soft-input decoding.

We have shown in this chapter that LDPC codes are the most appropriate codes for high bit-rate optical transmissions because of their lower decoding complexity and their better performance compared to Product codes. We have presented some combinatorial constructions of LDPC codes issued from the literature. These codes have structured parity-check matrices which make the encoding/decoding implementation easier. However their performance is not optimal and it is not always possible to choose freely the design parameters of the codes. Therefore, we have proposed a new construction of LDPC code based on the PEG algorithm but leading to a quasi-cyclic parity-check matrix and outperforming the combinatorial constructions.

The implementation of LDPC codes in 100Gb/s systems can highly improve the transmission quality. The expected coding gain is above 10dB. Moreover, such powerful codes may become a necessity for the development of 400Gb/s transmissions

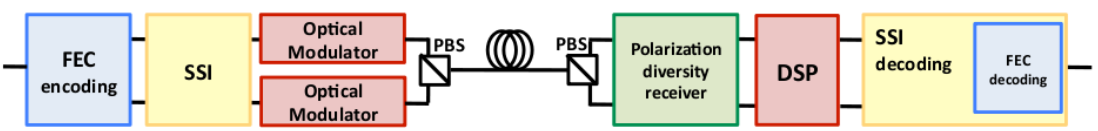
Chapter 3

Structured Symbol Interleaving

Introduction

Phase-shift keying (PSK) is a very efficient modulation format for high bit-rate optical transmission systems, it improves the OSNR by 3dB and thus dramatically increases the transmission distances compared to OOK systems [73][74].
PSK format can be employed in coherent and direct detection systems and both case require a differential encoding of the information. In direct detection systems, due to the lack of absolute phase reference, the phase of the previous modulated symbol is used as a relative phase reference (see Sec. 1.3.1). In coherent detection systems, the carrier phase recovery algorithms such as Viterbi-Viterbi Algorithm [28] introduce a phase ambiguity and differential encoding is necessary to overcome this problem (see Sec. 1.4.2.4).

The FEC techniques have been initially introduced to increase system margins of optical transmission, i.e. to combat optical impairments such as ASE and more recently polarization mode dispersion or nonlinear effects [75]. The FEC initially developed for OOK systems have been directly transposed to phase modulation systems without any adaptation to the differential encoding scheme [76]. Note that differential encoding leads to higher BER [77] which affects the FEC performance.
It is common to use bit interleaving to avoid the burst of errors. However, when differential encoding is considered, classical bit interleaving may efficiently corrects the penalties but not in an optimal way.



In this chapter a FEC codeword construction based on a structured symbol interleaving (SSI) of two or more codewords and the corresponding decoding algorithm is presented .

This coding/decoding scheme corrects the penalties introduced by differential encoding. It also leads to a FEC decoding complexity decrease and a significant redundancy reduction. We here focus on QPSK format but the proposed schemes can be applied for any modulation format using differential encoding. All the ideas presented here are available for direct detection systems and coherent systems.

We first recall in Sec. 3.1 the principles of differential encoding and present an error configurations analysis. We present in Sec. 3.2 the Structured Symbol Interleaving scheme and its performance is compared with the classical interleavers. In Sec. 3.3 we describe the corresponding decoding algorithm and analyze its complexity. Finally we shows in Sec. 3.4 that this coding/decoding scheme let us reduce the redundancy of the FEC codewords.

3.1 Differential encoding

With differential encoding, the information is coded on the transition between two modulated symbols of a constellation. For example, with QPSK modulation, instead of encoding the information on the signal phase, Fig. 3.1(a), the information is encoded on the phase shift between two symbols, Fig. 3.1(b). We call "data symbol" the two bits encoding a phase transition ("00","01") and "QPSK symbol" the modulated symbol belonging to the QPSK constellation (α_1, $\alpha_2 \ldots$).
For example, let us consider that the last emitted QPSK symbol was $\alpha_2 = e^{\frac{3i\pi}{4}}$ and that the next two bits are "11". The data symbol "11" corresponds to a 2 quadrant transition which means a π phase shift. The following emitted QPSK symbol is: $e^{\frac{3i\pi}{4}+i\pi} = e^{\frac{7i\pi}{4}} = \alpha_4$.

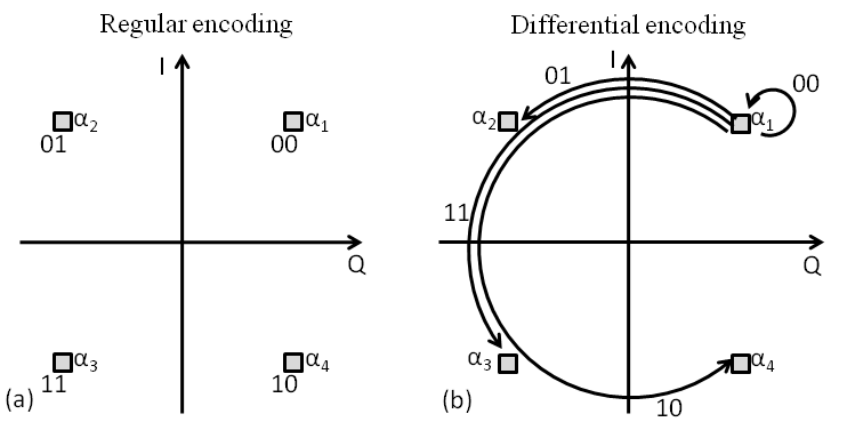


Figure 3.1: (a) Regular encoding of a QPSK modulation with Gray mapping. (b) Differential encoding of a QPSK modulation with Gray mapping.

A "transmission error" is a wrong decision made at the reception of a QPSK symbol. Gray mapping is generally employed to minimize the number of errors. It ensures that only one bit changes between two transitions to consecutive quadrants. As most of the errors are

due to a one quadrant shift at high OSNR, Gray mapping reduces the total number of bit errors.

3.1.1 Error configurations

Figure 3.2: Error configuration on the data bits due to a transmission error.

The major problem of the differential encoding is that there are at least twice as many errors than with a normal encoding. Indeed, as can be seen in Fig. 3.2, one transmission error on a QPSK symbol corrupts two data symbols corresponding to:

- the phase shift between the previous QPSK symbol and the wrong symbol
- the phase shift between the wrong symbol and the next QPSK symbol

Each corrupted data symbol has two bits, so it leads to two or four erroneous bits. Let us assume that an error occurs on a QPSK symbol and that Gray mapping is applied. If the erroneous QPSK symbol belongs to an adjacent quadrant of the emitted symbol quadrant, there will be only two erroneous bits (one per erroneous data symbol). If the erroneous symbol belongs to the opposite quadrant of the emitted symbol quadrant, there will be four erroneous bits (two per erroneous data symbol). At high OSNR, differential encoding produces twice as many errors as classical encoding.

A transmission error corrupts two consecutive data symbols. For instance in Fig. 3.2, the first error changed the transition "11" (+2 quadrants) into "01" (+1 quadrant), it means that the number of quadrant of the transition has been decreased by one quadrant. As a consequence, the transition corresponding to the neighbor erroneous data symbol is increased by one quadrant from "00" (0 quadrant) to "01" (+1 quadrant). The first error is a -1 quadrant error and the second one is a $+1$ quadrant error. In general, n_e consecutive transmission errors affect $n_e + 1$ data symbols. The corresponding errors $Nerr_i$ expressed in quadrant, follow:

$$\sum_{i=1}^{n_e+1} Nerr_i \pmod{M} = 0 \tag{3.1}$$

where M is the constellation size ($M = 4$ for a QPSK constellation). Table 3.1 lists all the quadrant error configurations for one or two consecutive QPSK symbol errors. For a single transmission error ($n_e = 1$), if one of the $Nerr_i$ is known, the value of the second one can be deduced from Eq. 3.1. For example, if $Nerr_2 = +1$, $Nerr_1$ has to be equal

n_e	$Nerr_1$	$Nerr_2$	$Nerr_3$
1	$+1$	-1	
	-1	$+1$	
	$+2$	$+2$	
2	$+1$	$+1$	$+2$
	$+2$	$+1$	$+1$
	-1	-1	$+2$
	$+2$	-1	-1
	$+1$	$+2$	-1
	-1	$+2$	$+1$
	$+2$	0	$+2$
	$+1$	0	-1
	-1	0	$+1$

Table 3.1: Errors configuration for QPSK system with differential encoding

to -1. For n_e consecutive transmission errors, there are some configurations in which the transmission errors can correct themselves and corrupt less than $n_e + 1$ data symbols as can be seen in the last lines of Table. 3.1.

3.2 Structured Symbol Interleaving (SSI)

With a differential encoding scheme, the FEC receives twice as many errors and the decoding is less efficient. However, these errors are not just randomly distributed along the codewords. Indeed, they come by pairs and this is very important knowledge that has to be taken into account.
One can first consider the consecutive errors as a burst of error and use a classical interleaving technique to overcome this problem. Randomly mixing the bits of different codewords so that consecutive bits do not belong to the same codewords is well-known and already applied in deployed long-haul high bit-rate optical transmission systems. However, this technique is not optimal for the differential encoding scheme.

3.2.1 Principle of SSI

The proposed construction does not try to de-correlate the consecutive bits, but on the contrary, to create a certain structure between the codewords. Since a transmission error produces a pair of consecutive erroneous data symbols on a codeword, our idea is to put the two corrupted data symbols on separated codewords. The interleaving is performed after the FEC encoding of the information. The bits of the FEC codewords are the data symbols transmitted by the modulated symbol transitions. Therefore, we propose to interleave the data symbols instead of the bits. Moreover, the interleaving is not random as in the classical

way but is performed alternating the data symbols from different codewords one after the other. We call this construction "structured symbol interleaving" (SSI). For simplicity, we will present the SSI method for the case of a 2 codewords interleaving. The generalization to more codewords is easy to deduce.

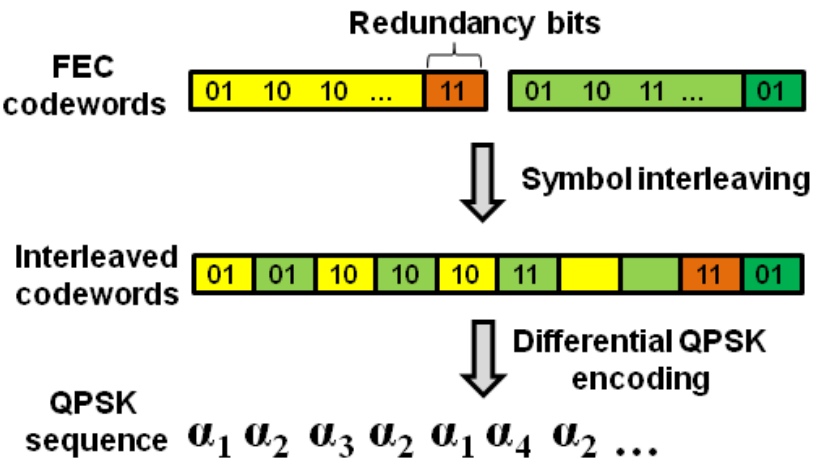


Figure 3.3: Principle of SSI of two FEC codewords

Pairs of codewords are serially mixed alternating two bits (one data symbol for QPSK modulation) of each one. Then, the resulting sequence is differentially modulated into a QPSK signal (see Fig. 3.3). This construction ensures that two adjacent data symbols belong to two different codewords. So if an error happens during the transmission, after the de-interleaving there will be only one erroneous data symbol on each codeword instead of two on a unique codeword (see Fig. 3.4). The proposed scheme is optimal to combat differential encoding pairs of error.

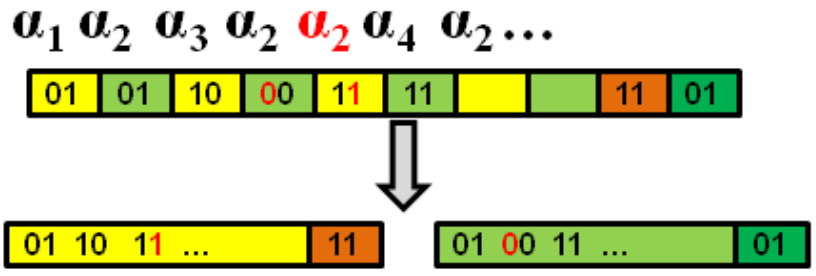


Figure 3.4: Example of error configuration after FEC symbol interleaving

Note that the proposed construction can be applied to any kind of constellation using differential encoding. For instance, differential binary phase shift keying (DPSK) has been demonstrated particularly well-adapted for non coherent optical long haul transmission systems. It is a two state constellation and each state is coded by one bit. So the corresponding construction for a DPSK scheme is a bit interleaving between two codewords.

3.2.2 SSI differential encoding impairment mitigation

Let us consider a FEC of length n with $n-k$ redundancy bits and which can correct t errors. With a differential encoding scheme, a FEC $\mathcal{C}(n,k,\mathrm{t})$ with SSI has the same performance as the code $\mathcal{C}(2n,2k,2\mathrm{t})$ without interleaving. Fig. 3.5 and Fig. 3.6 present the coding gain resulting from the symbol interleaving of two codewords for different families of FECs. As in Chapter 2, the simulations have been realized assuming an AWGN channel.

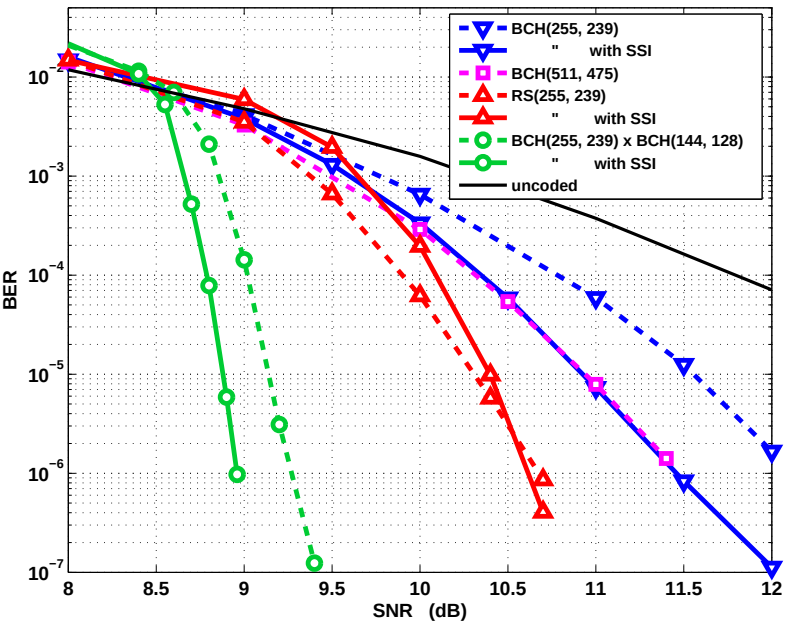


Figure 3.5: Performances comparison of hard decoding FECs with and without SSI

The BCH(255,239) is a binary FEC and the coding gain obtained with the SSI is 0.5dB at a BER$=10^{-6}$. Note that the performance of the BCH (255,239,t$=2$) with SSI is exactly the same as that of the BCH(511,475,t$=4$) without SSI. However, it is more advantageous to use a smaller code combine with the SSI because the encoding and decoding complexity of a small code is lower. A gain of 0.5 dB is also obtained for the BCH(1023,883,t$=14$) with SSI. The product code BCH(255,239)$\times$BCH(144,128) with a hard decoding [48] is more robust to burst of errors thanks to its product constructions, therefore the coding gain is 0.2dB. With non binary FEC such as Reed-Solomon (RS), the efficiency of our construction is reduced. Indeed, these codes are by nature less sensitive to burst of errors and therefore less sensitive to differential encoding impairments. In Fig. 3.6, we plot the performance of Lattice LDPC codes with and without SSI. The coding gains are about 0.1 and 0.2dB at a BER of 10^{-6}.

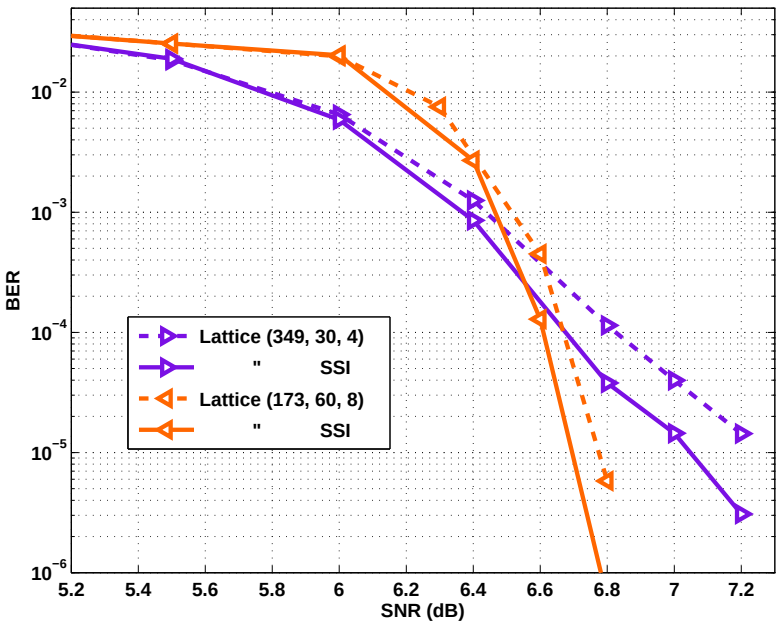


Figure 3.6: Performances comparison of LDPC with and without SSI

3.2.3 SSI Vs classical interleaving

Let us now compare the efficiency of the SSI compare to a classical interleaver. A classical interleaver consists to a random bit mixing of several codeword. Unlike the SSI, the interleaving is realized on the bits and has no structure.

The depth of an interleaver is the number of bits that are mixed. Our SSI scheme has a depth of two codewords, so $2n$ bits. In Fig. 3.7, we compare the performances of the proposed SSI scheme with the classical interleaver scheme with a 2 and 100 codeword depth ($2n$ and $200n$ bits). For a 2 codeword depth the classical interleaver is outperformed by the SSI. In order to obtain the same performance the classical interleaving requires a large depth such as 100 codewords, which induces high implementation complexity. We can easily conclude that the SSI is more adapted to the differential encoding than the classical interleaving.

3.3 Complexity reduction of the FEC decoding

We have seen that errors come in pairs of erroneous data symbols, each one belonging to different codewords (see Fig. 3.3). To recover the data, algebraic decoding of each codeword is performed after the de-interleaving. We call "algebraic decoding" the regular way to decode FEC (see Sec.2.1.2 for BCH and Sec. 2.1.4.2 for LDPC). In this section, we present a reduced complexity decoding algorithm.

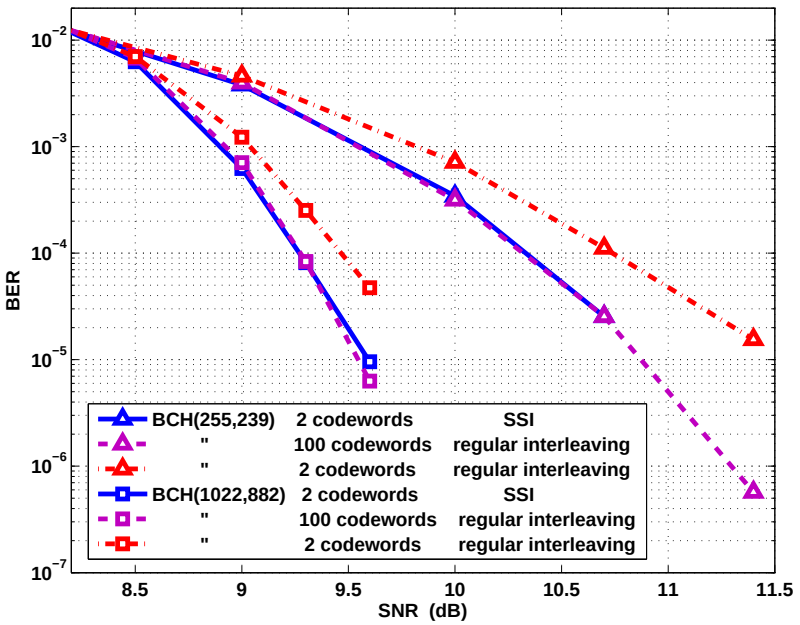


Figure 3.7: Comparison between the proposed SSI and the classical random interleaver

3.3.1 Principles of the decoding complexity reduction

We propose now to reduce the decoding complexity by performing the algebraic decoding on only one of the two codewords and to deduce the decoding of the second one. Indeed, the algebraic decoding gives us the position and the value of the error of one of the erroneous data symbol. As errors come in pairs, we know that the other erroneous data symbol is one of its direct neighbors. Moreover, as explained in Sec. 3.1 in Table 3.1 if we know the error value of the first error we can deduce the correction for the second one. The remaining question is: which neighbor is the erroneous one?

For instance, in Fig. 3.8, the algebraic decoding has corrected an error on the first codeword (in yellow). The correction has changed a "11" transition into a "10" transition: it corresponds to a $+1$ quadrant correction. From Table. 3.1, we know that the correction of the other erroneous data symbol has to be a -1 quadrant correction (under the assumption of a single transmission error). The last step is to determine if the second erroneous data symbol is the left or the right neighbor.
We can correct alternatively both of the neighbors and decide afterwards which one corresponds to the right correction. A pattern of all possible corrections for the second codeword is created. For example, in Fig. 3.8, the pattern has two entries: a codeword corresponding to the correction of the right neighbor and a codeword corresponding to the correction of the left neighbor. This is illustrated in Fig. 3.9). In the pattern, only one of the two corrections can lead to a valid codeword. In one case the erroneous data symbol is corrected

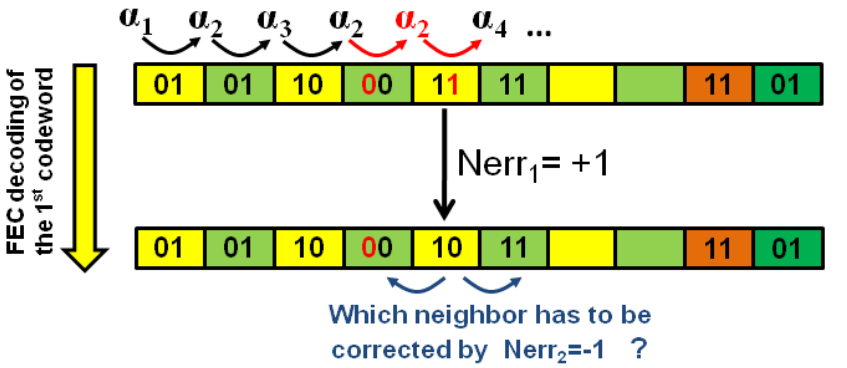


Figure 3.8: The error on the first codeword is detected and corrected using algebraic decoding. The position of the error on the second codeword is one of the neighbors and the correction to apply is known

whereas in the other case, the other neighbor is corrupted and both neighbors are then erroneous.

The right correction is the one giving a valid codeword of the FEC. We can check which one is giving a valid codeword of the FEC by computing the syndrome *synd* as in Eq. 2.3. If the syndrome is null, the codeword is valid.

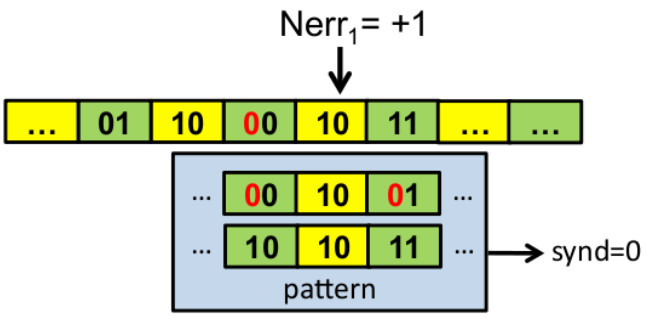


Figure 3.9: decoding of SSI scheme using a pattern

If many errors have been corrected by the algebraic decoding, all the configurations have to be checked. Therefore, the size of the pattern grows exponentially with the number of corrected errors on the first codeword. If there is no valid correction in the pattern, algebraic decoding of the second codeword has to be performed. The decoding algorithm is summarized in Algorithm. 1.

Note that the pattern can consider consecutive error cases, however the size of the pattern increases. The corrections are based on the error configurations listed in Table. 3.1 for $n_e > 1$. For example, if the first decoding corrected a single error on the k^{th} symbol, this can come from an error on one or both of its neighbors (the $k-1^{th}$ and/or the $k+1^{th}$

Algorithm 1 SSI decoding algorithm

```
Algebraic decoding of the first codeword
if the second codeword is valid then
    End
else
    Creation of a correction pattern
    if there is one valid codeword then
        End
    else
        Algebraic decoding of the second codeword
    end if
end if
```

symbol). Let us consider the case where the correction was a $+1$ correction. To consider the single error configuration, we will make a -1 quadrant correction on the $k-1^{th}$ and then a -1 quadrant correction on the $k+1^{th}$ symbol. To consider the two consecutive errors configuration, we will make a $+2$ quadrant correction on the $k-1^{th}$ symbol and a $+1$ quadrant correction on the $k+1^{th}$ symbol (line 5 of Table. 3.1) and the opposite (line 4 of Table. 3.1). So, the pattern would be twice as big.

3.3.2 Evaluation of the decoding complexity

In our algorithm the first codeword is always decoded algebraically and the complexity reduction depends on the second codeword decoding. Syndrome computation represents the most complex operation, as it has to be repeated for all the codeword corrections of the pattern. As the size of the pattern grows exponentially with the number of detected errors, the decoding can become prohibitively complex with powerful FECs. For instance, if the FEC can detect and correct 20 errors, the pattern size may reach 2^{20}.

The size of the correction pattern can be reduced by eliminating some cases. Instead of applying a correction on both neighbor data symbols, only the least reliable one is considered. The least reliable data symbol is the one corresponding to the least reliable QPSK symbol. The reliability of the QPSK symbols can be obtained computing their LLR:

$$L(x_i) = \log \frac{\Pr\{x_i = \alpha_{y_i} | y_i\}}{\Pr\{x_i \neq \alpha_{y_i} | y_i\}} \tag{3.2}$$

where x_i is the transmitted QPSK symbol, y_i the received QPSK symbol and α_{y_i} is its state in the constellation.
The QPSK symbol having the smallest absolute value of LLR is the least reliable. We assume that it is the erroneous one and we only correct the corresponding data symbol. Therefore, the pattern is now reduced to only one codeword and there is only one syndrome

computation to perform. The complexity of this operation can be considered negligible in comparison to an algebraic decoding. Thus, we can decode two codewords with the complexity of a single algebraic decoding.

Sometimes, the most reliable correction is not a valid codeword (the syndrome is not null). Therefore as our correction is wrong, the algebraic decoding of the second codeword has also to be performed. In that case, no complexity gain is obtained. The total decoding complexity of our algorithm depends on the size of the pattern and on how often the decoding of the second codeword is realized.

Let us now present the decoding complexity of the BCH$(1023, 883)$ (which is able to correct up to 14 errors) with our decoding scheme. We compare the cases when all the error configurations are considered (full pattern) and when only the most reliable correction is chosen in the pattern.

In Fig. 3.10, we plot the average number of syndrome computation. It corresponds to the average size of the pattern. We notice that when the full pattern is considered, the number of syndrome computation can be very large. However when only the most reliable correction (corresponding to every least reliable neighbor data symbols) is considered, only one syndrome computation has to be done.

In Fig. 3.11, we plot the probability of that no valid codeword is found in the pattern and consequently that the decoding of the second codeword has to be performed. It can be seen as the average number of algebraic decoding realized for the second codeword. Notice that it decreases exponentially with the SNR which means that at high SNR the correction chosen in the pattern is almost always the right one.

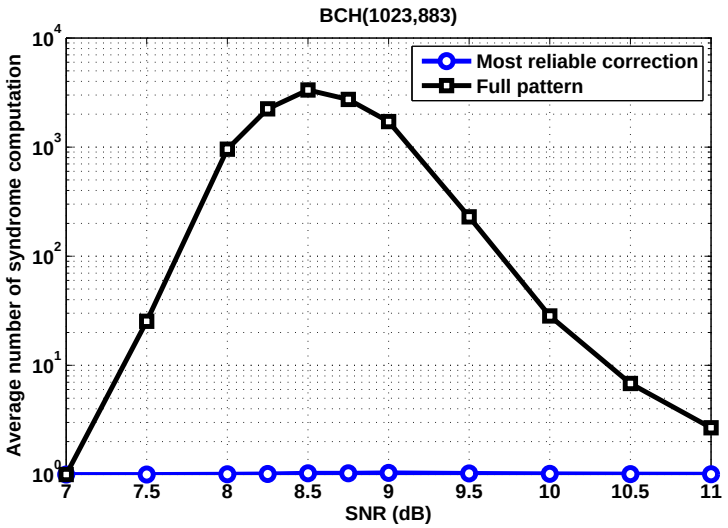


Figure 3.10: Average number of syndrome computation required for the decoding of the second codeword

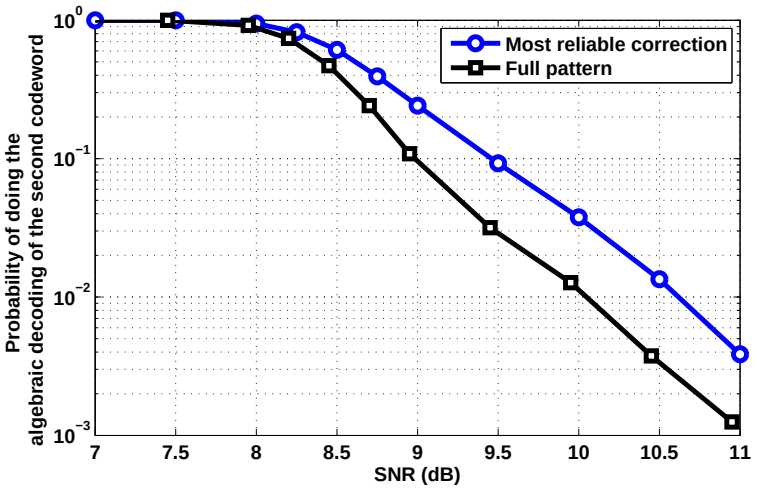


Figure 3.11: Average number of algebraic decoding required for the decoding of the second codeword

The complexity of our algorithm is evaluated by the average number of FEC decoding performed to decode both codewords. We define the total complexity reduction as the ratio between this number and the regular case where both codewords are always decoded. In Fig. 3.12 and Fig. 3.13, the complexity reduction is plotted over the output BER of the FEC. A 50% decoding complexity reduction is reached with the BCH(1022,882) code which means that decoding one codeword is most of the time enough to deduce the correction for the second one. Indeed, the input BER needed to achieve an error free transmission (BER$< 10^{-12}$) with this FEC is high enough to ensure that the probability of having consecutive errors and/or making a mistake on the reliability of the QPSK symbols, is very small. The product code BCH(255,239)$\times$BCH(144,128) with hard decoding, is a more powerful FEC (Fig. 3.5) and works at lower SNR. So, it is more probable that the correction of the second codeword fails and that the algebraic decoding has to be done. As probability of decoding the second codeword is higher, the total reduction reaches 45%. The SSI decoding is also very efficient with the LDPC codes and we observe a 40% reduction with the lattice LDPC $(173, 60, 8)$.

The coding gain resulting from the SSI scheme is the same using the proposed decoding. Hence, we have a complexity gain without any performance loss.

3.4 Redundancy reduction

The previous section has shown how the decoding of the first codeword can lead to the decoding of the second codeword with little effort. As the error correction capability of the second codeword is not really used, it is not necessary to encode it in the same way as the first codeword.

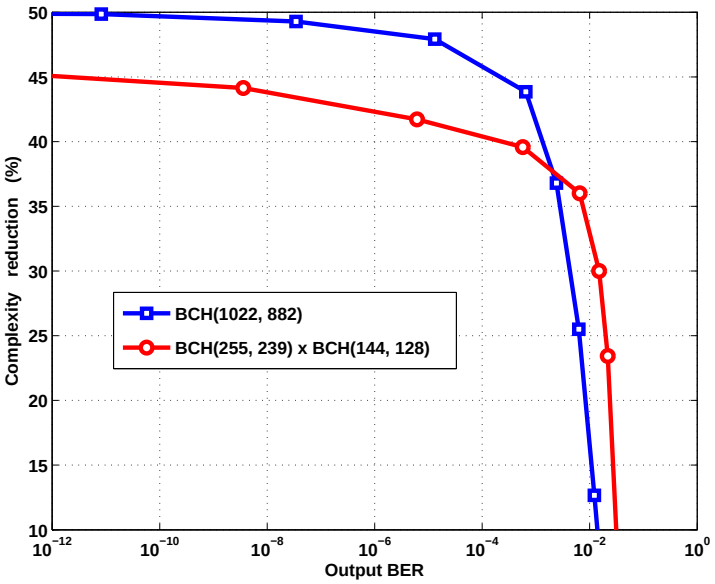


Figure 3.12: Percentage of complexity reduction for BCH and Product codes

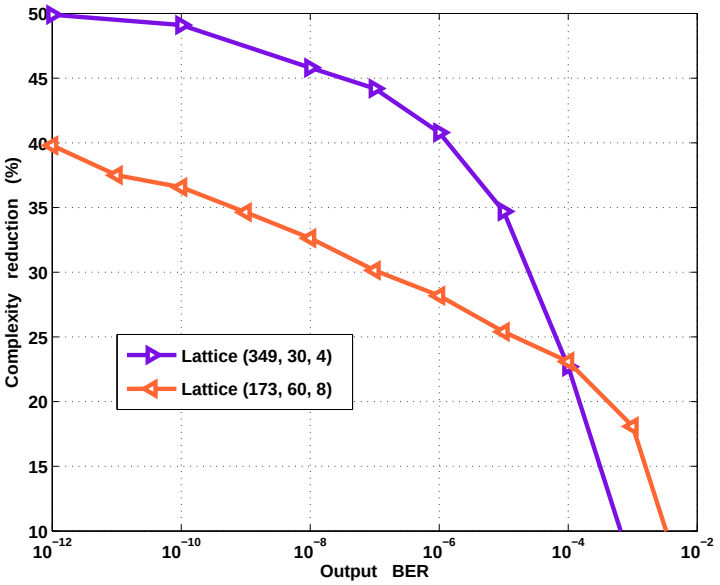


Figure 3.13: Percentage of complexity reduction for Lattice LDPC

We propose to encode the data by two different FECs before the SSI. A FEC with low correction capability, hence low redundancy, is chosen to encode the second codeword. With this configuration the total redundancy is decreased compared to the case using similar FECs. Let us consider two FECs: $\mathcal{C}_1(n_1, k)$ and $\mathcal{C}_2(n_2, k)$ with $n_2 < n_1$. Both FECs here use the same number of information bits k, but different constructions can be imagined in order to be more efficient. The resulted redundancy is:

$$\rho = \frac{(n_1 + n_2) - 2k}{(n_1 + n_2)} \tag{3.3}$$

ρ is less than $1 - \frac{k}{n_1}$ (the total redundancy when the same FEC is used).
The problem of using a less efficient FEC for the second codeword arises when algebraic decoding has to be performed. Indeed, the number of errors on each codeword is the same but the second FEC is not able to correct as many errors as the first one. It means that the second decoding will often fail and this could lead to an error floor on the performance.

Algebraic decoding of the second codeword happens when our algorithm proposes a wrong correction. However, although the correction is not valid, it is close to the right one. Most of the errors have already been removed and the algebraic decoding has only to deal with the remaining errors. Therefore, a less efficient FEC can be used for the second codeword if the number of remaining errors is not too large.

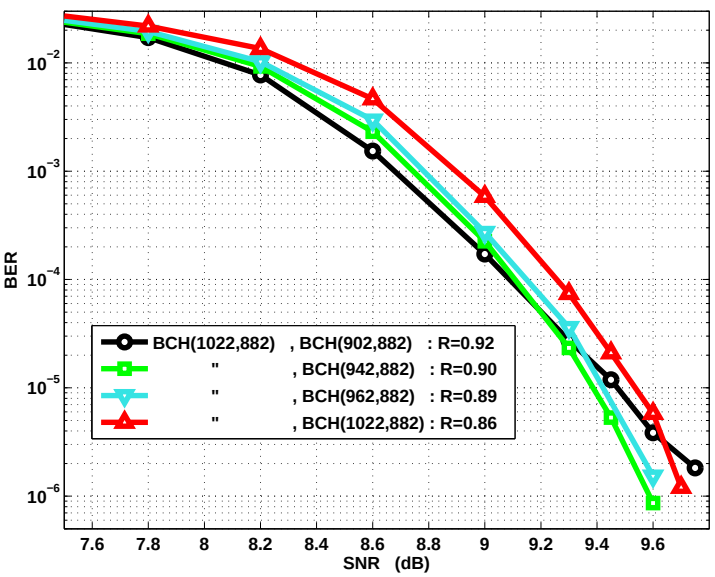


Figure 3.14: Performance of various combinations of FEC with different redundancies

In Fig. 3.14, we plot the performance of some combinations of FECs offering different redundancy reductions. The codes are decoded as described in the previous section. BCH(1022, 882) is used as the FEC of the first codeword. When the same FEC is used

for the second codeword, the total redundancy is 14% (rate $r = 0.86$). However if we use a BCH(962, 882) or a BCH(942, 882) to encode the second codeword, leading respectively to a total redundancy of 11% (rate $r = 0.89$) and 10% (rate $r = 0.90$), no performance degradations are observed until BER= 10^{-6}. However the BCH(902, 882) is not able to correct all the errors left by the first step of the algorithm and its performance is decreased. For the FEC combinations presenting an error floor, a concatenated scheme can be employed where the inner code uses the redundancy reduction proposed technique and the outer codes removes the error floor. In the case of LDPC codes, concatenated schemes are required anyway to remove the decoding error floor. Hence if our scheme is used for LDPC, the outer code would remove at the same time the decoding error floor and the redundancy scheme error floor.
It can be observed in the figure, that FEC with small redundancies perform slightly better. The energy per information bit is equal to $\mathrm{E}_b = \frac{\mathrm{E}_s}{k_c r}$ where E_s is the energy per modulated symbol, $k_c = 2$ is the modulation rate and r is the FEC rate. Therefore FEC with small redundancies, thus high rates correspond to smaller SNR. In Fig. 8, the FEC combinations corresponding to the rates 0.86, 0.89 and 0.90 have the same correction capability on the coded bits however the one with small redundancies correspond to more information bits and so perform better.
In our simulations, we observed no performance degradation using as the FEC of the second codeword, the BCH(942, 882) instead of BCH(1022, 882). Moreover the number of redundancy bits is reduced from 280 (140 redundancy bits on each codeword) to 200 (140 on the first codeword and 60 on the second) which means that the redundancy is reduced by 29%.

Conclusion

We have proposed a coding scheme based on a structured symbol interleaving of the FEC codewords in order to mitigate the differential encoding penalties. We have obtained a significant coding gain for binary FECs. An original decoding of the SSI construction has also been proposed, working with any FEC and offering decoding complexity reduction which can reach 50% without performance degradations. Finally, we have shown that according to the SSI coding/decoding scheme we can chose FEC with smaller redundancy. In our simulations, the rate has been increased from 0.86 to 0.9 with no penalties by choosing a FEC with 29% less redundancy bits.
In Chapter 2, we have seen that the implementation of the new generation of FEC (LDPC) in actual and future high bit-rate optical transmission systems is a major issue. The main challenge is to cope with their very high decoding complexity. Moreover these codes require a significant overhead ($\sim$ 20%) in order to have good performance and show an interest compared to hard decision schemes. Therefore, the proposed SSI coding /decoding scheme can be very profitable to these FEC.

Chapter 4

Polarization-Time coding

Introduction

With polarization multiplexing, the transmission spectral efficiency is doubled by emitting the data simultaneously on two orthogonal polarizations. This technique is very interesting to achieve high bit-rate and is now employed in current 40Gb/s and 100Gb/s optical transmission systems.

Polarization multiplexed transmissions can be seen as a 2×2 multiple-input multiple-output (MIMO) systems. The signals multiplexed on the 2 polarizations at the emission corresponds to multiple inputs, and the signals received on the 2 polarizations corresponds to multiple outputs.
Space-Time (ST) codes have been introduced in wireless communications to exploit all the degrees of freedom of the MIMO channel. In particular, they are very efficient in random fading environments. Space-Time codes can be adapted in optical transmissions and have been referred in the literature as Polarization-Time (PT) codes [78]. The decoding of ST code requires the channel knowledge. Hence, the channel impulse response has to be estimated first by a training sequence. The ST decoding is based on the maximum likelihood criterion which results in a complexity increasing exponentially with the size of the channel impulse response. The single carrier optical transmission described in Chapter 1, is a very dispersive channel due to the multiple dispersion effects occurring in the fiber (CD, PMD). Therefore, the implementation of PT codes on such a channel is impossible

However by using orthogonal frequency division multiplexing (OFDM), the inter-symbol interference can be totally compensated in the frequency domain. Therefore the channel impulse response is minimal and PT coding can be employed.

In wireless transmission, ST codes are effective in a fading environment. However polarization multiplexed transmissions are impaired mainly by PMD which is one of the major source of degradation in high bit-rate systems [79]. PDL is another source of impairments

introduced by all the in-line optical elements of the systems which attenuates in a different manner the two polarization components of the signal. PDL induces polarization dependent optical power fluctuations resulting in unequal optical signal-to-noise ratio on the two components. Note that PDL can be seen as a fading and can not be efficiently mitigated by digital equalization techniques.

In this chapter we investigate the interest of polarization-time coding in optical OFDM transmissions systems. In Sec. 4.1, we describe the optical OFDM systems. In Sec. 4.2 the principles of space-time codes are presented. We discuss about their constructions and their decoding and describe some famous codes in wireless communications. Finally in Sec. 4.3, we focus on polarization-time coding and investigate its performance to combat PDL.

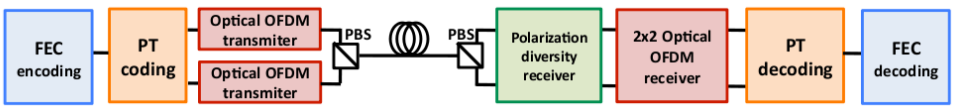


Figure 4.1: Optical OFDM transmission with Polarization-Time coding

4.1 Optical OFDM

Orthogonal frequency division multiplexing is a popular multiplexing technique in digital communication (DSL, LTE, WiMAX) because of its high robustness and its low equalization complexity in frequency selective channels.
The first optical OFDM systems have been realized recently [80]. The interest of this technique in optical fiber transmissions is mainly due to the low equalization complexity of dispersive effects. Indeed optical fiber is not a frequency selective fading media but it is a very dispersive channel where CD and PMD after a long-haul transmission may be very large.
In single carrier transmissions, the complexity of the equalization depends on the amount of dispersion and thus can be a real challenge. Hence, OFDM represents an efficient alternative and is a serious candidate for the future generations of 400Gb/s systems

4.1.1 Principle

The idea of frequency domain multiplexing is to transmit the data over a large number of sub-carriers at low rates. The interest of using low-rate modulations is the better tolerance against inter-symbol interference. Indeed, as the symbol duration is long, the dispersion only affects a small part of the symbol. Therefore a guard interval is generally inserted between the OFDM symbols to suppress the ISI.

In OFDM, the sub-carriers are orthogonal which means that their spectrum can overlap without interfering with each others. Note that the orthogonality increases the spectral efficiency and reduce the complexity of the receiver as separate filters for each sub-carriers are not required. An efficient way to obtain orthogonal subcarriers is to perform the multiplexing by an inverse fast Fourier transform (iFFT). Hence, at the reception because of their orthogonality, subcarriers can be demultiplexed easily, without using any filters, by a FFT.
However, because of dispersive effects, subcarriers orthogonality can be lost at the reception. Therefore, the FFT demultiplexing creates inter carrier interferences (ICI). An elegant solution to this problem is the insertion of a cyclic prefix (CP) at the beginning of the symbol which is a copy of the last part of the symbol. The cyclic prefix acts as a guard interval and prevents ISI. Moreover, it keeps the subcarrier orthogonality (suppresses ICI) by maintaining the periodicity of the OFDM symbol over an extended symbol duration.

4.1.2 Optical OFDM transmitter

The inputs of the OFDM transmitter is a block of N modulated symbols $\mathbf{x} = [x_1, x_2 \dots x_N]$ where x_k corresponds to the data transmitted by the k^{th} subcarrier. The vector is transformed in the time domain by an iFFT:

$$\tilde{x}_m = \frac{1}{\sqrt{N}} \sum_{k=1}^{N} x_k \exp\left(\frac{2\imath\pi km}{N}\right) \tag{4.1}$$

The cyclic prefix is added and the output vector is obtained:

$$\tilde{\mathbf{x}} = [\tilde{x}_{N-\Delta_G+1} \dots \tilde{x}_N, \tilde{x}_1 \dots \tilde{x}_N] \tag{4.2}$$

Δ_G is the cyclic prefix length, it depends on the dispersion amount. To avoid ISI and ICI, the cyclic prefix length must be larger than the total amount of dispersion (due to CD and PMD)[81]:

$$\frac{c}{f_c^2} \cdot |\mathrm{D}_{CD}| \cdot N_{sc} \cdot \Delta f_s + DGD_{\max} \leq \Delta_G \tag{4.3}$$

where f_c is the optical carrier frequency, c the light speed, D_{CD} the total accumulated chromatic dispersion in ps/km, N_{sc} the number of subcarriers, Δf_s the subcarrier channel spacing and $DGD_{\max}$ the maximum budgeted DGD.
Note that training sequence and pilot subcarriers are usually introduced for the channel estimation and the phase recovery. Fig. 4.2 shows an OFDM transmitter.
In a radio frequency (RF) OFDM system, the data is converted by digital to analog convertors (DAC) and then mixed with an RF carrier using an electrical I/Q modulator to obtain the OFDM signal. In optical OFDM, this signal has to be modulated onto an optical carrier. A straight solution is to convert the RF-OFDM signal using a conventional single-ended Mach-Zender modulator as represented in Fig. 4.3(a). However, a mirror image band is created and an additional filtering is required to remove it. To avoid the generation of

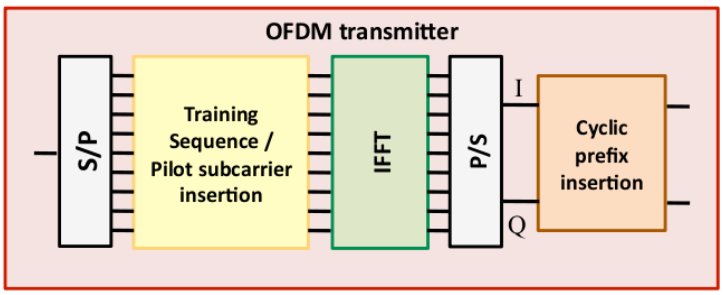


Figure 4.2: OFDM transmitter

the image band we can use an optical I/Q modulator as depicted in Fig. 4.3(b) which corresponds to a super MZ structure (see QPSK transmitter in Fig. 1.7) and to set half of the iFFT entries to zero [82]. Both methods result in a linear relationship between the optical field and the OFDM signal.

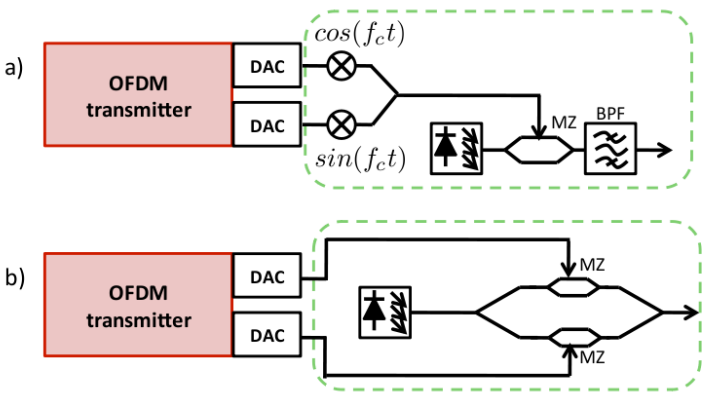


Figure 4.3: Single band ODFM transmitter based on (a) electrical I/Q modulator (b) optical I/Q modulator

In direct detection (DD) optical OFDM [83] the optical carrier is transmitted with the signal whereas in coherent (CO) OFDM [80] the carrier is filtered before the transmission and generated back at the reception by a local oscillator.

Polarization division multiplexing is compatible with OFDM systems. Two independent optical OFDM signals are generated and combined in the fiber by a polarization beam splitter. With polarization multiplexing, the data symbol $\mathbf{X} = [\mathbf{X}_1 \dots \mathbf{X}_N]$ can be represented as a $2{\times}N$ matrix where $\mathbf{x}_k = [x_k^{(1)}, x_k^{(2)}]^T$ are the two modulated symbols transmitted on both polarizations on the k^{th} subcarrier.
The use of PDM in CO-OFDM is straightforward and some practical demonstration can be found in [84] [85]. In DD-OFDM, an optical carrier is transmitted on each polarization

thus, the received components contain a mix of the two original carriers. Hence, they are not perfectly orthogonal to the OFDM data band which causes destructive interferences. A frequency shift of one of the signals can be made in order to avoid polarization mixing between the two carriers [86][87]. This is illustrated in Fig. 4.4.

At the reception, polarization are separated by a PBS and detected by in two separated OFDM receivers. However a joint MIMO processing is required between the receivers to estimate the channel and recover the data.

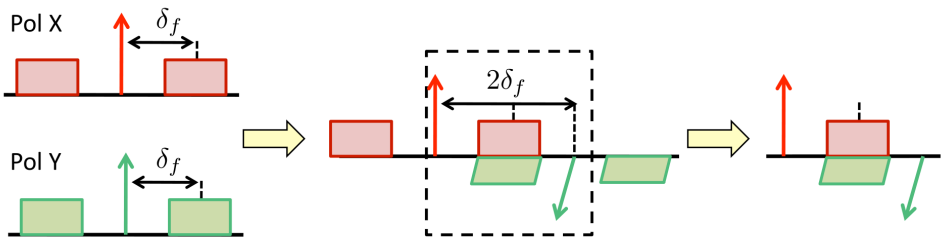


Figure 4.4: frequency shift of OFDM carrier in polarization multiplexed DD-OFDM

4.1.3 Optical OFDM receiver

DD-OFDM receiver is composed of an unique photodiode as a local oscillator is not required (Fig. 4.5 (a)). However a guard interval has to be inserted between the optical carrier and the OFDM band in the spectrum in order to avoid unwanted mixing products from the quadratic receiver. On the other side, CO-OFDM receiver has a local oscillator and thus, is more sensitive to frequency offset and phase noise and may require narrow linewidth laser. In Fig. 4.5 (b) and Fig. 4.5 (c), the coherent receiver for both electrical and optical I/Q demodulator configuration are represented.
At the reception, the signal is mixed with an optical carrier (from the transmitted signal or from the local oscillator) and the mixing product "carrier×signal" is detected. Once converted in the electrical domain, the signal is processed in an OFDM receiver as the one represented in Fig. 4.6.

Before any processing in the frequency domain, we have to make a time and frequency synchronization. Indeed, an OFDM system is very sensitive to time and frequency offsets as it cause the subcarrier orthogonality loss. Various algorithms have been proposed to realize these operations in the time domain [88][89][90]. These methods are based on the autocorrelation with a training sequence
Afterwards, the cyclic prefix is removed and the FFT is performed. As orthogonality has been maintained and the dispersion effects compensated thanks to the cyclic prefix, for each subcarrier there is a linear relationship between the iFFT input $\mathbf{x}_k$ and the FFT

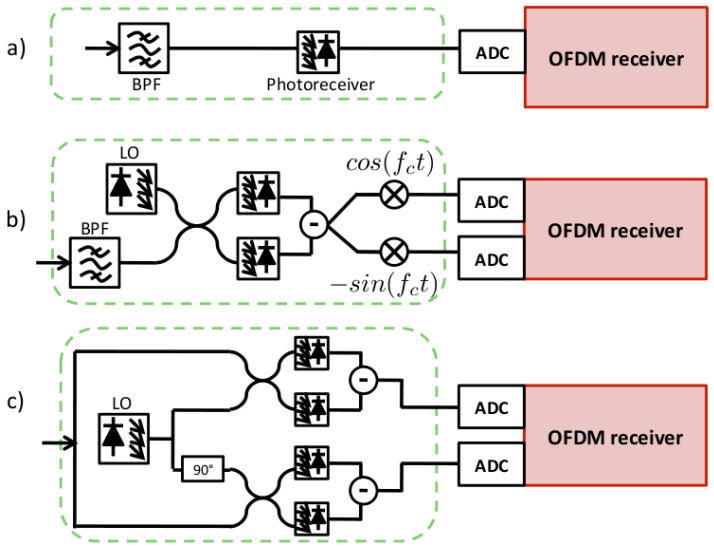


Figure 4.5: Single band receiver for (a) DD-OFDM (b) CO-OFDM with electrical I/Q modulator (c) CO-OFDM with optical I/Q modulator

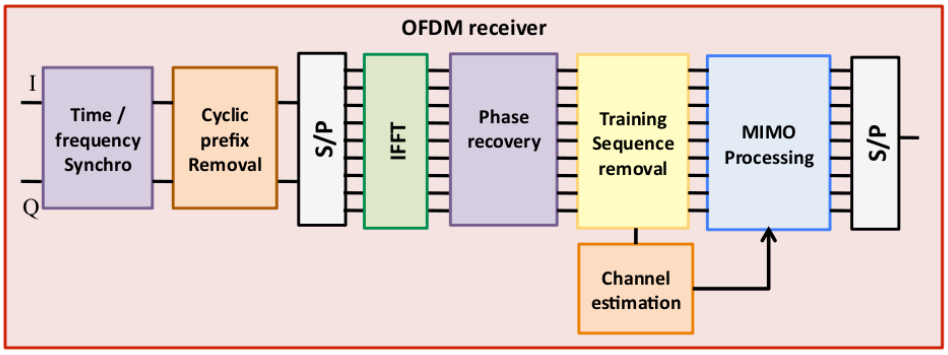


Figure 4.6: PDM-OFDM receiver

output $\mathbf{y}_k$ [91]:

$$\mathbf{y}_k = e^{\imath\Phi}\mathbf{H}_k\mathbf{x}_k + \mathbf{n}_k \qquad 1 \leq k \leq n \tag{4.4}$$

where $\mathbf{H}_k$ is the transfer matrix of the channel for the k^{th} subcarrier and $e^{i\Phi}$ represents the a phase shift caused by the local oscillator that has to be compensated with a carrier phase recovery processing. Phase estimation is usually realized using some pilot subcarriers. In [81], a carrier phase estimation is proposed based on the pilot subcarrier channel estimation and the averaging of the phase over all the pilots. Then, in Eq. 4.4 the phase shift term can be removed and we obtain the relation:

$$\mathbf{y}_k = \mathbf{H}_k\mathbf{x}_k + \mathbf{n}_k \tag{4.5}$$

4.1.3.1 Channel estimation

Channel can be estimated using a training sequence $\mathbf{S}_k$. The channel transfer matrix is a 2×2 matrix thus we need $\mathbf{S}_k$ to be also a 2×2 matrix in order to be able to determine all the coefficients of $\mathbf{H}_k$. The training symbols are sent on both polarizations and during two time slots (two OFDM symbols). For a simple implementation it has been proposed [92] to choose the training sequence $\mathbf{S}_k$ such as:

$$\mathbf{S}_k = \begin{bmatrix} s_k^{(1)}(t_1) & 0 \\ 0 & s_k^{(2)}(t_2) \end{bmatrix} \tag{4.6}$$

A known symbol is first sent on a polarization during the time slot t_1 (on a first OFDM symbol) and then a second symbol on the other polarization during a second time slot t_2 (on a second OFDM symbol). Note that this training sequence is not optimal as only two different symbols are used. In Chapter 5, we discuss about channel estimation by training sequence in the time domain which can also be applied for OFDM systems.

Several estimation algorithms exist in the literature, the simpliest one is based on the least square estimation (LSE) criterion (see Appendix D) which is generally used in practical implementation [85]. From Eq. 4.5 we can easily obtain the estimated transfer matrix $\widehat{\mathbf{H}}_k$:

$$\widehat{\mathbf{H}}_k = \mathbf{Y}_k \left(\mathbf{S}_k^H \left(\mathbf{S}_k \mathbf{S}_k^H \right)^{-1} \right) \tag{4.7}$$

To attenuate the noise effect, many training symbols can be sent and the estimated transfer matrix is obtained by averaging.

4.1.3.2 Symbol detection

All the dispersion has been suppressed thank to the cyclic prefix. Therefore we only have to demultiplex the two polarizations and this operation is realized using a 1-tap butterfly FIR equalizer (see Fig. 1.10 for a multi-tap example). This corresponds to a simple matrix multiplication:

$$\widehat{\mathbf{x}}_k = \mathbf{W}_k \mathbf{y}_k \tag{4.8}$$

where $\mathbf{W}_k$ is a 2×2 matrix corresponding to the equalizer of the k^{th} sub-carrier.

for practical implementation simplicity, a zero-forcing (ZF) equalizer can be used:

$$\mathbf{W}_k = \left(\widehat{\mathbf{H}}_k^H \widehat{\mathbf{H}}_k \right)^{-1} \widehat{\mathbf{H}}_k^H \tag{4.9}$$

In wireless environment, ZF decoding has poor performances because it results in the noise amplification when fading occurs. However, optical fiber channel the fading are caused by PDL and are not as severe.

4.1.4 Optical OFDM channel capacity

The channel capacity is defined as the maximum transmission rate for an arbitrary small error probability. Shannon capacity corresponds to the maximum mutual information between the input and the output of the channel and is expressed in bit/s/Hz. For a channel is varying in the time, we define the ergodic capacity as the mean capacity over the channel realization.

in OFDM, the total capacity is the sum of channel capacity of each sub-carrier. In a polarization multiplexed OFDM transmission, each subcarrier corresponds to a 2×2 MIMO flat channel and its ergodic capacity is equal to [93]:

$$\mathrm{C}_k = \mathbb{E}_{\mathbf{H}_k}\left\{\log_2\left(\det\left[\mathbf{I} + \frac{\rho}{2}\mathbf{H}_k\mathbf{H}_k^H\right]\right)\right\} \tag{4.10}$$

After a singular value decomposition of the matrix $\mathbf{H}_k$, it leads to:

$$\mathrm{C}_k = \sum_{i=1}^{2} \mathbb{E}_{\mathbf{H}_k}\left\{\log_2\left(1 + \frac{\rho}{2}\lambda_i^2\right)\right\} \tag{4.11}$$

where $\rho = k_c\mathrm{SNR}$ (see Sec. 1.2.5) is the signal to noise ratio per symbol on the sub-carrier and λ_i are the singular values of $\mathbf{H}_k\mathbf{H}_k^H$. A 2×2 MIMO channel can be seen as 2 parallel AWGN channels.
Note that we have neglected the nonlinear effects which actually result in a capacity decrease at high SNR (high signal power) [14].

4.2 Space-Time coding

In this section, we introduce the principle of space-time codes. This coding technique has been developed for wireless communication in order to take benefit of all the degrees of freedom of MIMO channel. We will recall construction criterion of these codes and present some of the most popular codes. The construction criterion presented in this section have been derived for wireless transmissions, hence they are no longer available for optical transmissions as the nature of the channel is different.

4.2.1 MIMO channels in wireless transmissions

In wireless communication, multi-antenna transmissions is the main example of MIMO systems. The interest on MIMO systems comes from the early works of Foschini [93] and Telatar [94] which have demonstrated that the channel capacity increases as the number of antenna in the system increases.
MIMO systems offer at the same time, spatial multiplexing (data rate increase) and spatial diversity (performance improvement) which has brought lot of interest on them. In

particular, space-time coding techniques have been developed to enhance the performance of MIMO systems.

4.2.1.1 MIMO system model

We consider a MIMO system with n_t transmit antennas and n_r receive antennas. The channel is supposed to be a block fading Rayleigh channel which means that it remains constant for the duration of T symbols. Modulated symbols are transmitted by n_t antennas and can be represented by a complex matrix $\mathbf{X}$ of size $n_t \times T$. The received signal is represented by a $n_r \times T$ complex matrix $\mathbf{Y}$ corresponding to the signals received on each receive antenna during each time symbol. The channel is represented by a $n_r \times n_t$ complex matrix $\mathbf{H}$ where $h(i,j)$ is the fading coefficient of the path between the i^{th} transmit antenna and the j^{th} receive antenna. The paths are spatially decorrelated and Rayleigh channel model is assumed, so the coefficients are independent and identically distributed (iid) complex gaussian random variables with zero mean and unit variance ($h(i,j) \sim \mathcal{N}(0,1)$). The received symbols can be expressed as:

$$\mathbf{Y} = \mathbf{HX} + \mathbf{N} \tag{4.12}$$

where $\mathbf{N}$ is the additive white gaussian noise.

4.2.1.2 Derivation of the error probability

We supposed perfect channel state information (CSI) at the receiver side. We note $\mathcal{C}$ the set of all possible transmit matrices $\mathbf{X}$. According to the channel model of Eq. 4.12, a maximum likelihood decoder (ML) estimates $\widehat{\mathbf{X}}$ as:

$$\widehat{\mathbf{X}} = \arg\min_{\mathbf{X}\in\mathcal{C}} \|\mathbf{Y} - \mathbf{HX}\|^2 \tag{4.13}$$

The error probability can be defined as:

$$\begin{aligned} P_e &= \Pr\{\widehat{\mathbf{X}} \neq \mathbf{X}\} && (4.14)\\ &= \sum_{\mathbf{X}\in\mathcal{C}} \Pr\{\mathbf{X}\} \cdot \Pr\{\widehat{\mathbf{X}} \neq \mathbf{X} | \mathbf{X}\} && (4.15) \end{aligned}$$

Using the union bound and assuming that the symbols are equiprobable we obtain:

$$P_e = \frac{1}{|\mathcal{C}|} \sum_{\mathbf{X}\in\mathcal{C}} \sum_{\widehat{\mathbf{X}}\in\mathcal{C}, \widehat{\mathbf{X}}\neq\mathbf{X}} \Pr\{\widehat{\mathbf{X}} \longrightarrow \mathbf{X}\} \tag{4.16}$$

where $\Pr\{\mathbf{X} \longrightarrow \widehat{\mathbf{X}}\}$ is the pairwise error probability and can be expressed by [95]:

$$\begin{aligned} \Pr\{\mathbf{X} \longrightarrow \widehat{\mathbf{X}}\} &= \Pr\{\|\mathbf{Y} - \mathbf{H}\widehat{\mathbf{X}}\|^2 \leq \|\mathbf{Y} - \mathbf{HX}\|^2 \mid \mathbf{X}, \mathbf{H}\} && (4.17)\\ &= \Pr\{\|\mathbf{H}(\widehat{\mathbf{X}} - \mathbf{X}) + \mathbf{N}\|^2 \leq \|\mathbf{N}\|^2 \mid \mathbf{X}, \mathbf{H}\} && (4.18)\\ &= \Pr\{\|\mathbf{H}(\widehat{\mathbf{X}} - \mathbf{X})\|^2 + \langle \mathbf{H}(\widehat{\mathbf{X}} - \mathbf{X}), \mathbf{N}\rangle \leq 0 \mid \mathbf{X}, \mathbf{H}\} && (4.19) \end{aligned}$$

$V = \|\mathbf{H}(\widehat{\mathbf{X}} - \mathbf{X})\|^2 + \langle \mathbf{H}(\widehat{\mathbf{X}} - \mathbf{X}), \mathbf{N} \rangle$ is a Gaussian random variable with mean $m_V = \|\mathbf{H}(\widehat{\mathbf{X}} - \mathbf{X})\|^2$ and variance $\sigma_V^2 = 4\sigma_{\mathbf{N}}^2 \|\mathbf{H}(\widehat{\mathbf{X}} - \mathbf{X})\|^2$. So, pairewise error probability is expressed by:

$$\Pr\{\mathbf{X} \longrightarrow \widehat{\mathbf{X}}\} = \mathcal{Q}\left(\frac{m_V}{\sigma_V}\right) \tag{4.20}$$

$$= \mathcal{Q}\left(\frac{\|\mathbf{H}(\widehat{\mathbf{X}} - \mathbf{X})\|}{2\sigma_{\mathbf{N}}}\right) \tag{4.21}$$

where $\mathcal{Q}$ is the Gaussian tail function. By averaging over $\mathbf{H}$, it can be upper bounded by:

$$\Pr\{\mathbf{X} \longrightarrow \widehat{\mathbf{X}}\} \leq \mathbb{E}_{\mathbf{H}}\left[\exp\left(-\frac{\|\mathbf{H}(\widehat{\mathbf{X}} - \mathbf{X})\|^2}{8\sigma_{\mathbf{N}}^2}\right)\right] \tag{4.22}$$

As the coefficients of $\mathbf{H}$ are Gaussian random variables, by averaging over $\mathbf{H}$, the previous expression can be simplified.
Let us first note $\mathbf{A} = (\widehat{\mathbf{X}} - \mathbf{X})(\widehat{\mathbf{X}} - \mathbf{X})^H$. By a singular value decomposition of $\mathbf{A}$, we obtain two unitary matrices $\mathbf{U}$ and $\mathbf{V}$ and a diagonal matrix $\mathbf{D}$, such as $\mathbf{A} = \mathbf{V}\mathbf{D}\mathbf{U}^H$. We note λ_i the diagonal elements of $\mathbf{D}$, they are the singular values of $\mathbf{A}$. $\mathrm{r}_{\mathbf{X}}$ is the rank of $\mathbf{A}$ thus there are $\mathrm{r}_{\mathbf{X}}$ non zero singular values $\lambda_1 \dots \lambda_r$. The Pairewise error probability at high SNR can be bounded [96]:

$$\Pr\{\mathbf{X} \longrightarrow \widehat{\mathbf{X}}\} \leq \left(\prod_{i=1}^{\mathrm{r}_{\mathbf{X}}} \lambda_i\right)^{-n_r} \left(\frac{\mathrm{SNR}}{4}\right)^{-\mathrm{r}_{\mathbf{X}} \cdot n_r} \tag{4.23}$$

4.2.2 Space-Time code construction criteria on Rayleigh fading channel

In Eq. 4.23 it can be seen that MIMO transmission error probability depends on three parameters: $\mathrm{r}_{\mathbf{X}}$, n_r and λ_i. n_r is a given parameter but $\mathrm{r}_{\mathbf{X}}$ and λ_i depend on the transmit matrix $\mathbf{X}$. Therefore we have to find the most adapted transmit matrix $\mathbf{X}$ that minimizes the error probability. These matrices belong to a set $\mathcal{C}$ that we construct and call the code and $\mathbf{X}$ are the codewords. The codewords are constructed from the modulated symbols and $x(i,j)$ are functions of one or many modulated symbols s_i.

$$\mathbf{X} = \begin{bmatrix} f_1(s_1, s_2 \dots) & f_2(s_1, s_2 \dots) & \dots \\ f_3(s_1, s_2 \dots) & f_4(s_1, s_2 \dots) & \\ \vdots & & \end{bmatrix} \tag{4.24}$$

The error probability decreases exponentially function of the SNR and $r \cdot n_r$ determines how fast is the decrease. In a semi logarithmic plot, it represents the asymptotic slope of the error probability function of the SNR. The diversity order of the MIMO system is defined as the power of $\frac{1}{\mathrm{SNR}}$ and is equal to $\mathrm{r}_{\mathbf{X}} \cdot n_r$. $\mathbf{A}$ is a $n_t \times n_t$ matrix and thus $1 \leq \mathrm{r}_{\mathbf{X}} \leq n_t$. The code has to be designed in order to achieve the full (maximum) diversity

$n_t \cdot n_r$ which means having a matrix $\mathbf{A}$ with a full rank $\mathrm{r}_{\mathbf{X}} = n_t$. This is called the *rank criterion.*

When spatial multiplexing is realized, independent modulated symbols are transmitted on each antenna. The codewords are $n_t \times 1$ vectors $\mathbf{X} = [s_1, s_2 \dots s_{n_t}]^t$. It is obvious that $\mathrm{r}_{\mathbf{X}} = 1$ which is not optimal as the maximum rank is $\mathrm{r}_{\mathbf{X}} = n_t$.
However if $\mathbf{X}$ is a $n_t \times T$ matrix, with $T \geq n_t$, we have $\mathrm{r}_{\mathbf{X}} \leq n_t$ which gives us a chance to reach the full rank and thus, reduce the error probability. Note that $T = n_t$ is enough to obtain a full rank and thus full diversity. A codeword $\mathbf{X}$ is a particular construction of transmit symbols sent on n_t antennas and during T symbol times. Therefore the code $\mathcal{C}$ is called a Space-Time code (STC).

In Eq. 4.23, the pairewise error probability also depends on the product of singular values of $\mathbf{A}$, so its determinant. The term corresponds to a coding gain. The error probability is upperbound by the sum of pairewise error probabilities and thus, dominated by the one corresponding to the smallest determinant. Therefore, the space-time code has to be constructed in order to maximize the minimum determinant Δ_{min}:

$$\Delta_{min} = \min_{\mathbf{X} \neq \widehat{\mathbf{X}}} \prod_{i=1}^{\mathrm{r}_{\mathbf{X}}} \lambda_i = \min_{\mathbf{X} \neq \widehat{\mathbf{X}}} \det \mathbf{A} = \min_{\mathbf{X} \neq \widehat{\mathbf{X}}} \det |(\mathbf{X} - \widehat{\mathbf{X}})|^2 \tag{4.25}$$

This is called: *determinant criterion*

Eq. 4.11 shows that the MIMO channel can be seen as $\min(n_t, n_r)$ parallel AWGN channels. So in a $n_t \times n_r$ MIMO transmission system, $\min(n_t, n_r)$ modulated symbols can be transmitted at each symbol time. A Space-Time has a full rate if $\min(n_t, n_r) \cdot T$ modulated symbols are transmitted during T symbol times so if its rate is $r_{\mathrm{STC}} = \min(n_t, n_r)$ symbols per channel use .

We have presented three major design issue for space-time codes in block Rayleigh fading channel: a rate criterion, a rank criterion and a determinant criterion. However in the literature, many others relevant criteria can be found, for example based on the mutual information [97] or the diversity-multiplexing tradeoff [98].

4.2.3 Space-Times block codes

A Space-Time block code (STBC) codewords are $n_t \times T$ matrices $\mathbf{X}$ functions of N_s modulated symbols $s_1, s_2 \dots s_{N_s}$. T is the number of symbol time T_s during which a ST codeword is transmitted. Note that T_s will be called in the following a channel use (cu). The rate of an STBC is expressed in symbol per channel use and defined as:

$$R_{\mathrm{STBC}} = \frac{N_s}{T} \qquad (symb/cu) \tag{4.26}$$

The codeword matrix represents the symbols transmitted on each antenna a_i and during each symbol time t_j. Here is an example of a codeword transmitted over two symbol times in a 2×2 MIMO system:

$$\begin{aligned} \mathbf{X} &= \begin{bmatrix} x_{a_1}(t_1) & x_{a_1}(t_2) \\ x_{a_2}(t_1) & x_{a_2}(t_2) \end{bmatrix} \\ &= \begin{bmatrix} f_1(s_1 \dots s_{N_s}) & f_3(s_1 \dots s_{N_s}) \\ f_2(s_1 \dots s_{N_s}) & f_4(s_1 \dots s_{N_s}) \end{bmatrix} \end{aligned} \tag{4.27}$$

A STBC is said to be complex linear if the functions f_i are linear. If we can define a complex matrix $\mathbf{G}_{\mathbb{C}}$ called generator matrix of the code, such as:

$$\begin{aligned} \text{vec}_{\mathbb{C}}(\mathbf{X}) &= \mathbf{G}_{\mathbb{C}}\mathbf{S} \\ \begin{bmatrix} x_{a_1}(t_1) \\ x_{a_2}(t_1) \\ \vdots \\ x_{a_{n_t}}(t_T) \end{bmatrix} &= \mathbf{G}_{\mathbb{C}} \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_{N_s} \end{bmatrix} \end{aligned} \tag{4.28}$$

where $\text{vec}_{\mathbb{C}}(\mathbf{X})$ is the vectorization of the matrix $\mathbf{X}$ into a $n_t \cdot T \times 1$ complex vector.
A STBC is said to be real linear if it has a real $(2n_t \cdot T) \times (2N_s)$ generator matrix $\mathbf{G}_{\mathbb{R}}$ such as:

$$\begin{aligned} \text{vec}_{\mathbb{R}}(\mathbf{X}) &= \mathbf{G}_{\mathbb{R}}\mathbf{S}_{\mathbb{R}} \\ \begin{bmatrix} \text{Re}\,(x_{a_1}(t_1)) \\ \text{Re}\,(x_{a_2}(t_1)) \\ \vdots \\ \text{Im}\,(x_{a_1}(t_1)) \\ \text{Im}\,(x_{a_2}(t_1)) \\ \vdots \end{bmatrix} &= \mathbf{G}_{\mathbb{R}} \begin{bmatrix} \text{Re}(s_1) \\ \text{Re}(s_2) \\ \vdots \\ \text{Im}(s_1) \\ \text{Im}(s_2) \\ \vdots \end{bmatrix} \end{aligned} \tag{4.29}$$

$\text{vec}_{\mathbb{R}}(\mathbf{X}) = [\text{Im}\,(\text{vec}_{\mathbb{C}}(\mathbf{X}))\ \ \text{Im}\,(\text{vec}_{\mathbb{C}}(\mathbf{X}))]^t$. If the modulated symbols belong to a QAM or a PSK constellation, $\mathbf{S}_{\mathbb{R}} = \mathbf{S}_{\mathbb{Z}} \in \mathbb{Z}^{N_s}$.

4.2.3.1 Alamouti code

Alamouti code was designed in order to bring spatial diversity in 2×1 MIMO system. The inputs are blocks of two modulated symbols $\mathbf{S} = [s_1 \ \ s_2]^T$. During a first time, they are transmitted each one on an antenna. The Alamouti codeword matrix is written as:

$$\mathbf{X}_{\mathcal{A}} = \begin{bmatrix} s_1 & -s_2^* \\ s_2 & s_1^* \end{bmatrix} \tag{4.30}$$

The codeword matrix is orthogonal ($\mathbf{X}_{\mathcal{A}}\mathbf{X}_{\mathcal{A}}^H = (|s(1)|^2 + |s(1)|^2)\mathbf{I}$), so the Alamouti code is said an orthogonal code

The received signal is:

$$[y_1(t_1) \;\; y_1(t_2)] = [h(1) \;\; h(2)] \begin{bmatrix} s_1 & -s_2^* \\ s_2 & s_1^* \end{bmatrix} + [n(1) \; n(2)] \tag{4.31}$$

which can be rewritten as:

$$\begin{bmatrix} y_1(t_1) \\ y_1(t_2)^* \end{bmatrix} = \begin{bmatrix} h(1) & h(2) \\ h(2)^* & -h(1)^* \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} + \begin{bmatrix} n(1) \\ n(2)^* \end{bmatrix} \tag{4.32}$$

We have an equivalent channel model $\tilde{\mathbf{Y}} = \tilde{\mathbf{H}}\mathbf{S} + \mathbf{n}$ where $\tilde{\mathbf{H}}$ is orthogonal and has the same structure as the Alamouti codewords. This can be simplified into:

$$\tilde{\mathbf{H}}^H \tilde{\mathbf{Y}} = \|\mathbf{H}\|^2 \mathbf{S} + \mathbf{N} \tag{4.33}$$

And so, the maximum likelihood decoding rule expressed in Eq. 4.13 is obtained by a simple threshold decision (ZF decoding).
Alamouti code is very famous for the low complexity of its decoding. ML performance is obtained by linear processing. Moreover this code is optimal in 2×1 MIMO systems as it achieves full diversity and full rate.
However in a 2×2 MIMO scheme, Alamouti code does not achieve the full rate anymore. Indeed the rate of the code is $R_{\mathcal{A}} = 1$ symb/cu whereas the possible maximum rate is $\min(n_t = 2, n_r = 2) = 2$ symb/cu.

4.2.3.2 Golden code

The Golden code [99] is the best 2×2 STBC for Rayleigh fading channels. It has various properties as for instance, being full diversity ($\mathrm{r}_{\mathcal{G}} = n_t$), full rate ($R_{\mathcal{G}} = min(n_t, n_r)$) or having an uniform average energy per antenna.
The Golden codeword matrix is:

$$\mathbf{X}_{\mathcal{G}} = \frac{1}{\sqrt{5}} \begin{bmatrix} \alpha\,(s_1 + \theta s_2) & \alpha\,(s_3 + \theta s_4) \\ i\overline{\alpha}\,(s_3 + \overline{\theta} s_4) & \overline{\alpha}\,(s_1 + \overline{\theta} s_2) \end{bmatrix} \tag{4.34}$$

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 modulated symbols. The Golden code generator matrix is:

$$\mathbf{G}_{\mathbb{C},\,\mathcal{G}} = \frac{1}{\sqrt{5}} \begin{bmatrix} \alpha & \alpha\theta & 0 & 0 \\ 0 & 0 & i\overline{\alpha} & i\overline{\alpha}\overline{\theta} \\ 0 & 0 & \alpha & \alpha\theta \\ \overline{\alpha} & \overline{\alpha}\overline{\theta} & 0 & 0 \end{bmatrix} \tag{4.35}$$

The minimum determinant of the Golden code, so the coding gain, is equal to $\frac{1}{5}$ which is the largest possible value among all 2×2 STBC.

4.2.3.3 Silver code

The Silver code is the name given to the code proposed in [100] because its performance is slightly smaller than that of Golden code. The author construction is based on a layered structure of two alamouti codes. Silver code satisfies the rank and the rate criteria.
The Silver codeword matrix is:

$$\mathbf{X}_{\mathcal{S}} = \mathbf{X}_{\mathcal{A}}(s_1,\ s_2) + \mathbf{T}\ \ \mathbf{X}_{\mathcal{A}}(z_1,\ z_2) \tag{4.36}$$

$$= \begin{bmatrix} s_1 & -s_2^* \\ s_2 & s_1^* \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} z_1 & -z_2^* \\ z_2 & z_1^* \end{bmatrix} \tag{4.37}$$

where z_1 and z_2 are obtained by:

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

Silver code is a real linear STBC and its generator matrix can be found in [101].
The minimal determinant of the Silver code is equal to $\frac{1}{7}$ which is slightly inferior to that of the Golden code. In Fig. 4.7, we plot the performance of the Silver and the Golden code. We can observe that Golden code performs better than the Silver code. However because of its particular structure, its decoding complexity is smaller than the Golden code when exhaustive search is realized. The Sezginer-Sari (SS) code [102] construction is based on the same structure as the Silver code. In Fig. 4.8, we represent the codeword symbol constellation of four STBC.

4.2.4 Space-Time block code decoding

In Sec. 4.2.1.2 , we have derived the error probability of a MIMO transmission on a flat Rayleigh fading channel under the assumption of a ML decoding. In particular, we have outlined a spatial diversity coding gain in Eq. 4.23 which is only obtained when a ML decoding is performed. For example, ZF decoding (Sec. 4.1.3.2) only achieves an order of diversity of $n_t - n_r + 1$ for $n_r \geq n_t$.
ML decoding can be realized by an exhaustive search. In that case, the norm $\|\mathbf{Y} - \mathbf{HX}\|^2$ has to be computed for all possible codewords. The number of computations depends the constellation size M. If the code is full rate, $n_t \cdot n_r$ symbols are transmitted in each codeword and thus an exhaustive search requires $M^{n_t \cdot n_r}$ norm computations. This can become prohibitively high for large number of antennas or/and high spectral efficiency modulations.

Lattice decoder

By considering the linear property of real linear STBC, the decoding can be realized using the lattice decoding presented in [103], with less complexity than the exhaustive search.

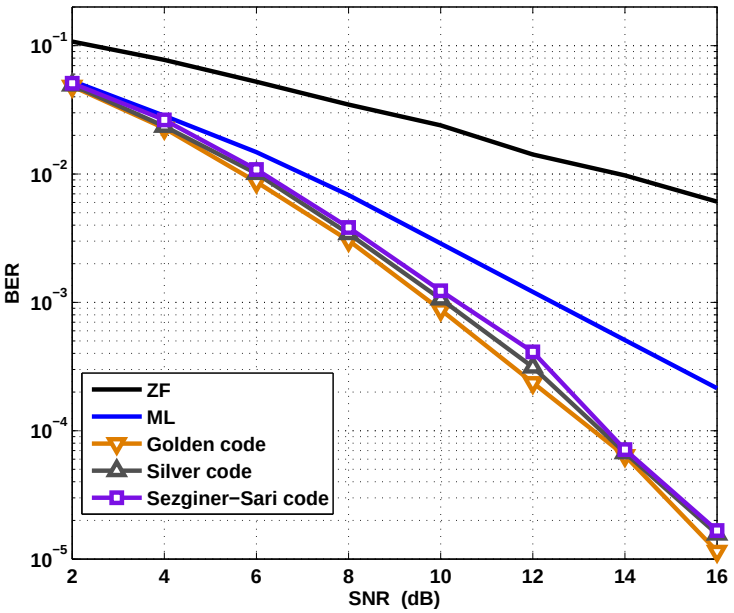


Figure 4.7: Golden, Silver and Sari-Sezginer code performance on Rayleigh fading channel with ML decoding

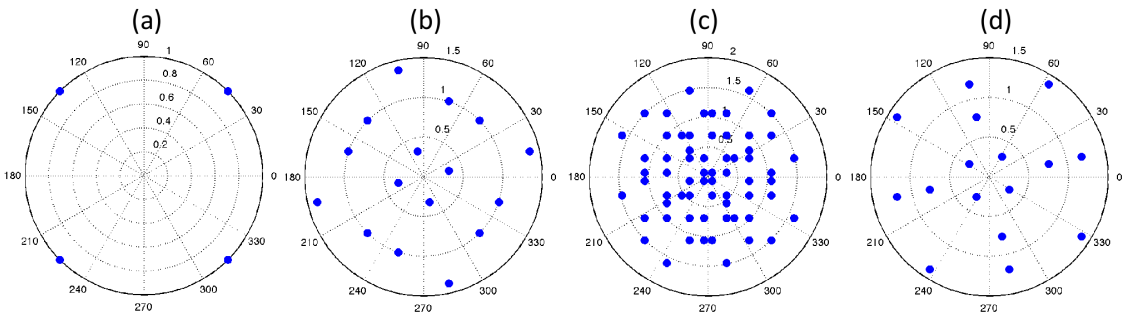


Figure 4.8: Constellations of STBC: (a) Alamouti code, (b) Golden code, (c) Silver code, (d) Sezginer-Sari code

A lattice Λ in $\mathbb{R}^n$ is a discrete subgroup of $\mathbb{R}^n$ which spans the real vector space $\mathbb{R}^n$:

$$\Lambda = \left\{ \sum_{i=1}^{n} a_i v_i \quad | \quad a_i \in \mathbb{Z} \right\}$$

where $v_1, v_2 \dots v_n$ is a base of $\mathbb{R}^n$. Thus, using a matrix form we have $\Lambda = \{\mathbf{Va}\}$, where $\mathbf{V} = [v_1 \dots v_n]$ and $\mathbf{a} = [a_1 \dots a_n]^T \in \mathbb{Z}^n$.

By vectorizing the vectors in $\mathbb{R}$ and using the generator matrix representation of real linear

STBC, Eq. 4.12 can be rewritten:

$$\mathbf{Y} = \mathbf{HX} + \mathbf{N}$$

$$\mathrm{vec}_{\mathbb{C}}(\mathbf{Y}) = \begin{bmatrix} \mathbf{H} & 0 \\ 0 & \mathbf{H} \end{bmatrix} \mathrm{vec}_{\mathbb{C}}(\mathbf{X}) + \mathrm{vec}_{\mathbb{C}}(\mathbf{N}) \tag{4.38}$$

$$\mathrm{vec}_{\mathbb{R}}(\mathbf{Y}) = \begin{bmatrix} \mathrm{Re}(\mathbf{H}) & 0 & -\mathrm{Im}(\mathbf{H}) & 0 \\ 0 & \mathrm{Re}(\mathbf{H}) & 0 & -\mathrm{Im}(\mathbf{H}) \\ \mathrm{Im}(\mathbf{H}) & 0 & \mathrm{Re}(\mathbf{H}) & 0 \\ 0 & \mathrm{Im}(\mathbf{H}) & 0 & \mathrm{Re}(\mathbf{H}) \end{bmatrix} \mathrm{vec}_{\mathbb{R}}(\mathbf{X}) + \mathrm{vec}_{\mathbb{R}}(\mathbf{N}) \tag{4.39}$$

$$\mathbf{Y}_{\mathbb{R}} = \mathbf{H}_{\mathbb{R}}\mathbf{G}_{\mathbb{R}}\mathbf{S}_{\mathbb{R}} + \mathbf{N}_{\mathbb{R}} \tag{4.40}$$

$$\mathbf{Y}_{\mathbb{R}} = \mathbf{M}_{\mathbb{R}}\mathbf{S}_{\mathbb{Z}} + \mathbf{N}_{\mathbb{R}} \tag{4.41}$$

The last equation is obtained by assuming that the modulation format is a QAM (i.e $\mathbf{S}_{\mathbb{Z}}$). If the Space-Time code is full rank, $\mathbf{G}_{\mathbb{R}}$ is a base of $\mathbb{R}^n$ with $n = 2n_tT$. As the coefficients of $\mathbf{H}_{\mathbb{R}}$ are all independent, its lines can not be dependent and thus the matrix $\mathbf{M}_{\mathbb{R}} = \mathbf{H}_{\mathbb{R}}\mathbf{G}_{\mathbb{R}}$ forms also a base of $\mathbb{R}^n$. Therefore $\Lambda = \{\mathbf{M}_{\mathbb{R}}\mathbf{S}_{\mathbb{Z}}\}$ is a lattice.

The ML decoding rule (Eq. 4.13) can be reinterpreted as a lattice closest point search:

$$\widehat{\mathbf{S}} = \underset{\mathbf{S}\in\mathcal{Z}^n}{\arg\min} \|\mathbf{Y}_{\mathbb{R}} - \mathbf{M}_{\mathbb{R}}\mathbf{S}_{\mathbb{Z}}\|^2 \tag{4.42}$$

$$\widehat{\mathbf{U}}_{\Lambda} = \underset{\mathbf{U}\in\Lambda}{\arg\min} \|\mathbf{Y}_{\mathbb{R}} - \mathbf{U}_{\Lambda}\|^2 \tag{4.43}$$

The received point $\mathbf{Y}_{\mathbb{R}}$ is not a lattice point as it has been transmitted over a noisy channel. Lattice decoding consists of looking for the closest point to $\mathbf{Y}_{\mathbb{R}}$ in a finite region centered on the received point. The search region could be a sphere. The decoder looks for all lattice points inside this sphere and select the closest one. As the number of considered lattice point is limited, the decoding complexity is reduced. The Schnorr-Euchner decoder [104] and the Sphere decoder [105] are two lattice search algorithms searching the closest point inside the sphere. They follow different strategies but result in equivalent performance and decoding complexity. In Appendix. C, we briefly describe the Sphere decoder presented in [105] [95].

4.3 Polarization-Time coding

In Sec. 4.1.2, we have seen that polarization multiplexed optical OFDM system can be considered as linear 2×2 MIMO system. The equivalent transmission model for each

subcarrier is expressed by Eq. 4.5. Then in Sec. 4.2.2, we have shown that Space-Time code can bring both coding and diversity gain in MIMO Rayleigh fading channels. The straightforward idea is to apply space-time coding techniques to polarization multiplexed optical transmissions. Here the polarizations have the same role as the antennas in wireless communications. So this technique has been referred as polarization-time (PT) coding in the literature.

However, optical fiber channel is not a Rayleigh fading channel and advantages of Space-Time coding may not be the same in the context of optical fiber transmissions. For example, spatial (or polarization) diversity derived from the error probability expression have no obvious definition here.

Originally, PT coding has been proposed to mitigate polarization scattering induced by XPM in WDM systems [78]. Alamouti code was used and it was observed that PT coding significantly reduces the impact of the neighbor channels. However many details were untold as for example the channel model or how the decoding was performed.

In [106] Alamouti code has been also proposed but in order to mitigate PMD in coherent optical OFDM. PT-OFDM scheme seems to perform better than for example, turbo equalization or regular polarization diversity OFDM. However authors have observed that when combined with FEC, the advantages of polarization-time coding compared to the regular scheme, is lost.

Alamouti code has been proposed against PMD also in DD-OFDM systems in [107]. PMD induced a power fading at the reception that PT coding can mitigate. However no details are given about the fading channel model.

4.3.1 PT coding against PMD

Many questions arise about the real efficiency of PT coding to mitigate PMD in OFDM systems. First, PMD can be very easily removed in the frequency domain by OFDM (with direct or coherent detection), when cyclic prefix is used. Without cyclic prefix, the channel impulse response in Eq. 4.5 is not large and Space-Time decoding can not be performed with a reasonable complexity. Moreover, PMD as CD, are dispersion effects and can not be considered as fadings.

We consider an optical transmission scheme considering only PMD effects. The channel transfer matrix has the form of Eq. 1.14. We have not taken into account the chromatic dispersion as it is a deterministic effect and can be easily managed by OFDM cyclic prefix. We can notice that the transfer matrix of the PMD is unitary ($\mathbf{H}_k\mathbf{H}_k^H = \mathbf{I}$) which means that it does not introduce any fading. The channel capacity expressed by Eq. 4.10 can be

rewritten:

$$\begin{aligned} \mathrm{C}_k &= \sum_{i=1}^{2} \mathbb{E}_{\mathbf{H}_{\mathrm{PMD}}} \left\{ \log_2 \left(1 + \frac{\rho}{2} \lambda_i^2 \right) \right\} \\ \mathrm{C}_k &= 2 \log_2 \left(1 + \frac{\rho}{2} \right) \end{aligned} \tag{4.44}$$

Although PMD can be a very limiting impairment for transmissions, it does not reduce the capacity of the channel. Indeed the capacity corresponds to two parallel gaussian channels ($\mathbf{Y}$=$\mathbf{X}$+$\mathbf{N}$). On such non fading channels, space-time coding technique brings no benefits hence PT coding is useless against PMD in optical OFDM systems.

4.3.2 PT coding against PDL

Polarization dependent loss is a fading effect introduced by the in-line components of the transmission. The channel transfer matrix of a component with PDL is given in Eq. 1.15. We can see that the attenuation concerns only one of the channels but none of them can be in fading at the same time (due to the optical amplifiers). The coefficient of the matrix are not independent and identically distributed gaussian variables as in the case of a Rayleigh fading channel. As ST codes are designed for Rayleigh fading channels, their performance may be very different in an optical fiber transmission. Nevertheless we will show that they are still very efficient in order to mitigate PDL fading effects.

4.3.2.1 A simple model of PDL

We consider an optical OFDM transmission scheme including only PDL effects.

$$\mathbf{Y} = \mathbf{H}_{\mathrm{PDL}} \mathbf{X} + \mathbf{N} \tag{4.45}$$

where $\mathbf{H}_{\mathrm{PDL}}$ corresponds to the transfer matrix expressed in Eq. 1.15:

$$\mathbf{H}_{\mathrm{PDL}} = \mathbf{R}_1 \underbrace{\begin{bmatrix} \sqrt{1-\gamma} & 0 \\ 0 & \sqrt{1+\gamma} \end{bmatrix}}_{\mathbf{D}} \mathbf{R}_2$$

The channel transfer function depends on two random rotation matrices corresponding to the angles φ_1, φ_2 and on the PDL value $\mathbf{\Gamma}_{dB}$. φ_1 and φ_2 are chosen uniformly in $[0\ 2\pi]$. $\mathbf{\Gamma}_{dB}$ has a deterministic value expressed in dB.

Unlike PMD, PDL does reduce the channel capacity:

$$\mathrm{C}_k = \log_2 \left(1 + \frac{\rho}{2} \sqrt{1-\gamma} \right) + \log_2 \left(1 + \frac{\rho}{2} \sqrt{1+\gamma} \right) \tag{4.46}$$

where γ corresponds to the PDL attenuation expressed in Eq. 1.16. This can be seen as the capacity of two parallel AWGN channels with their own SNR $\rho_1 = \rho\sqrt{1-\gamma}$ and $\rho_2 = \rho\sqrt{1+\gamma}$ which are function of the PDL fading.

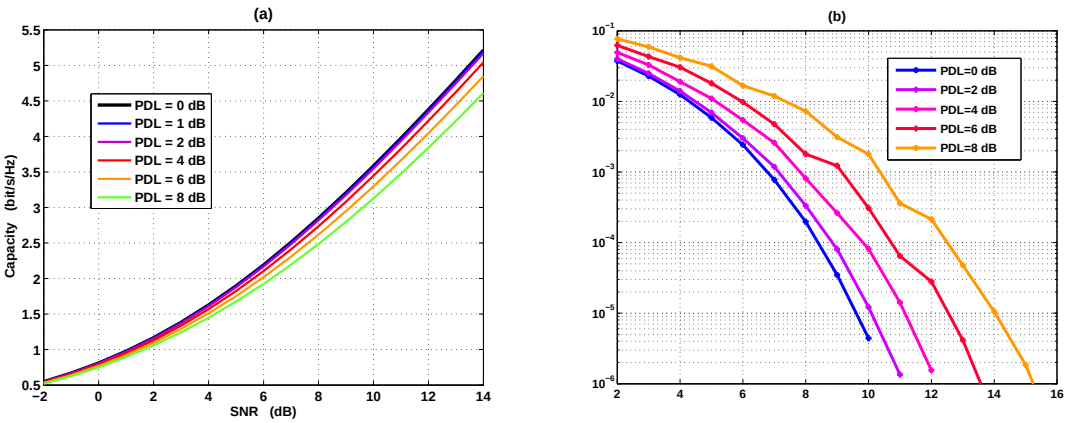


Figure 4.9: (a) Ergodic channel capacity, (b) BER

In Fig. 4.9 (a), we represent the channel ergodic capacity for one subcarrier function of the SNR for different values of PDL. In Fig. 4.9 (b), we plot the BER of one subcarrier for 5 PDL values. We observed that PDL reduces the channel capacity and increases the BER. For instance for a PDL of 6dB, there is a 2.2dB penalty at a BER of 10^{-3} compared to the case without PDL. We can also remark that even in presence of PDL, the BER follows a $\mathcal{Q}$-function of the SNR. This has been demonstrated in [10]. Therefore the error probability can be expressed as:

$$P_e \sim \mathcal{Q}\left(\frac{\|\mathbf{H}_{\text{PDL}}\left(\mathbf{X}_1 - \mathbf{X}_2\right)\|}{2\text{N}_0}\right) \tag{4.47}$$

Note that this case is totally different from the Rayleigh fading channel where BER has an asymptotic behavior and spatial diversity is defined as its slope. Here, talking about polarization diversity in the same way that we defined the spatial diversity, has no sense.

4.3.2.2 Design criterion of PT codes

Considering an optical transmission scheme without PDL effects, we have a gaussian channel model ($\mathbf{Y}=\mathbf{X}+\mathbf{N}$). Minimizing the error probability is equivalent to maximizing the minimal distance of the constellation. However Eq. 4.47 shows that in presence of PDL, the BER also depends on the minimal distance:

$$d_{min} = \min_{\mathbf{X}_1,\mathbf{X}_2\in\mathcal{C}} \left(\|\mathbf{H}_{\text{PDL}}\left(\mathbf{X}_1 - \mathbf{X}_2\right)\|\right) \tag{4.48}$$

As outlined in [10], d_{min} corresponds to the smallest distance between a pair of constellation points in the re-normalized space (re-normalized by PDL). The PT code having the best performance on the optical fiber channel is the one having the highest minimum distance d_{min}. For each PDL value, the transfer matrix $\mathbf{H}_{\text{PDL}}$ is function of the random angles φ_1 and φ_2 thus, the minimum distance of the code is only function of φ_2. Indeed the re-normalized space corresponds to a rotation of φ_2 of the QAM constellation followed by

an attenuation that reduce the minimal distance. Hence the last rotation of φ_1 has not effect.

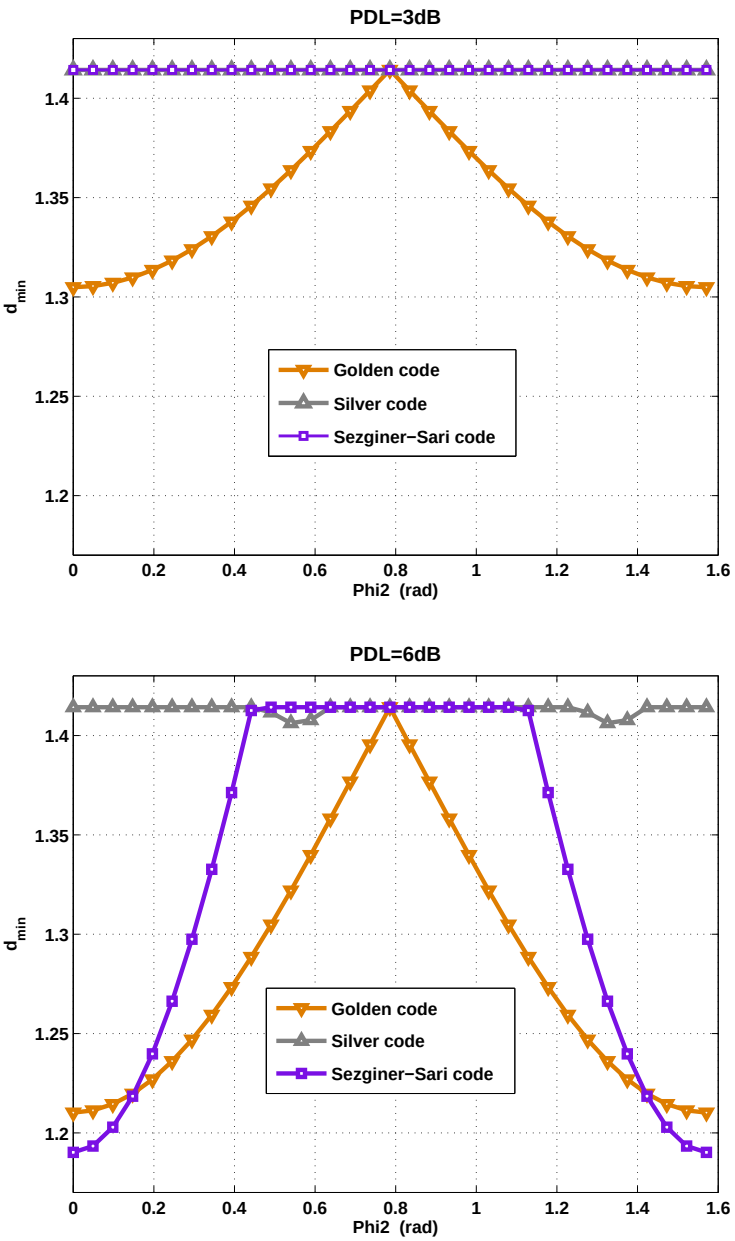


Figure 4.10: Minimal distance between pair of codewords of Golden code, Silver code and Sezginer-Sari code.

In Fig. 4.10, the minimal distance d_{min} of the Golden code, the Silver code and the Sezginer-Sari [102] code are represented in function of the rotation angle φ_2. QPSK modulation has been considered.

We can observe that for PDL = 3dB, that the minimal distance of the Silver and the

Sezginer Sari code is constant which is not the case of the d_{min} of the Golden code. When PDL is increased to 6dB, Sezginer-Sari code becomes dependent of the angled and has its minimal distance reduced for some rotations. However Silver code is still quite insensitive to the rotation angle variation.

For these three PT codes, d_{min} shows a $\frac{\pi}{2}$ periodicity. As the rotation matrix angle is uniformly distributed over $[0, 2\pi]$, we can deduce the average minimum distance of each code by averaging d_{min} on $\left[0, \frac{\pi}{2}\right]$.

$$\mathbb{E}_{\mathbf{H}}\left[d_{min}\right] \tag{4.49}$$

In Fig. 4.11, we can see the decrease of the average minimal distance of the three PT codes in function of the PDL. A very interesting result appears in this figure. Indeed the average d_{min} of the Golden code is smaller than the one of Silver and Sezginer-Sari code. The average minimal distance can not be directly related to the error probability but it gives a good idea of the behavior of the codes.

The error probability depends on d_{min}, so according to Fig. 4.11, Silver and Sezginer-Sari codes must outperform the Golden code. Note that both are quasi orthogonal Alamouti layer structure codes their average minimal distance remains constant for small values of PDL.

Unlike in 2×2 MIMO Rayleigh fading channel, here the Silver code is the best PT code.

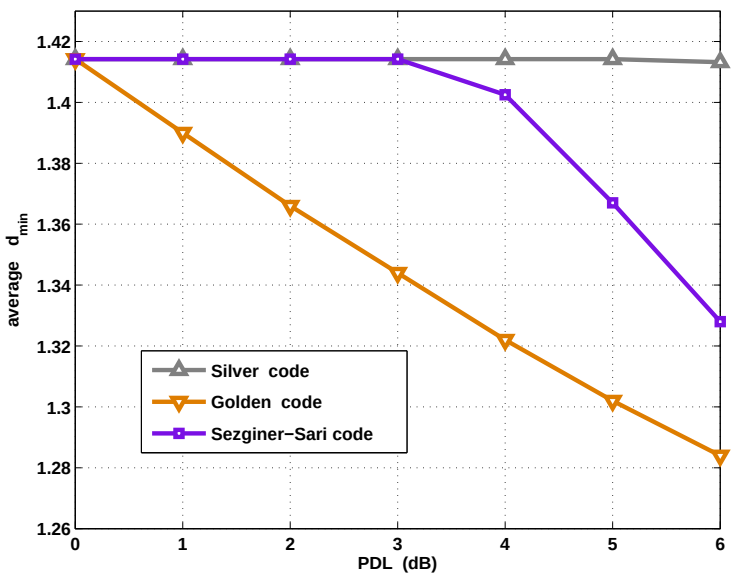


Figure 4.11: Average minimal distance of Golden code, Silver code and Sezginer-Sari code.

Let us now find an expression for d^2_{min}.

$$\|\mathbf{H}_{\text{PDL}}(\mathbf{X}_1 - \mathbf{X}_2)\|^2 = \left\|\mathbf{H}_{\text{PDL}}\bar{\mathbf{X}}\right\|^2 \tag{4.50}$$

$$= \left\|\mathbf{R}_1\mathbf{D}\mathbf{R}_2\bar{\mathbf{X}}\right\|^2 = \|\mathbf{D}\underbrace{\mathbf{R}_2\bar{\mathbf{X}}}_{\mathbf{U}}\|^2 \tag{4.51}$$

we note $u_{i,j}$ the coefficients of the matrix $\mathbf{U}$ and $\mathbf{v}_i = [u_{i,1}\ u_{i,2}]$.

$$\mathbf{DU} = \begin{bmatrix} \sqrt{1-\gamma}u_{1,1} & \sqrt{1-\gamma}u_{1,2} \\ \sqrt{1+\gamma}u_{2,1} & \sqrt{1+\gamma}u_{2,2} \end{bmatrix} \tag{4.52}$$

$$\|\mathbf{DU}\|^2 = (1-\gamma)(\|u_{1,1}\|^2 + \|u_{1,2}\|^2) + (1+\gamma)(\|u_{2,1}\|^2 + \|u_{2,2}\|^2) \tag{4.53}$$

$$= \|\mathbf{v}_1\|^2 + \|\mathbf{v}_2\|^2 + \gamma\left(\|\mathbf{v}_1\|^2 - \|\mathbf{v}_2\|^2\right) \tag{4.54}$$

We note that a code with $\|\mathbf{v}_1\|^2 = \|\mathbf{v}_2\|^2$ for every pairs of points $\{\mathbf{X_1},\ \mathbf{X}_2\} \in \mathcal{C}$, is insensitive to the PDL. For instance the Alamouti code satisfies this property. Let us note the difference between two pairs of Alamouti codewords:

$$\bar{\mathbf{X}}_{\mathcal{A}} = \mathbf{X}_{2,\mathcal{A}} - \mathbf{X}_{1,\mathcal{A}} = \begin{bmatrix} \bar{s}_1 & -\bar{s}_2^* \\ \bar{s}_2 & \bar{s}_1^* \end{bmatrix} \tag{4.55}$$

We can show that:

$$\|\mathbf{DU}_{\mathcal{A}}\|^2 = 2\left(|\bar{s}_1|^2 + |\bar{s}_2|^2\right) \tag{4.56}$$

which means that the minimal distance is independent of the PDL.

4.3.2.3 PDL mitigation with PT coding

Let us now evaluate the performance of PT code against PDL. We consider the system presented in Fig. 4.1 where a QPSK modulation is used. Fig. 4.12 shows the BER in function of the SNR for a PDL of 6dB.

We can clearly see the efficiency of PT coding against PDL. Without PT codes, the SNR degradation induced by PDL is about 3dB at BER= 10^{-5} but using PT codes, the SNR penalties are only 1dB with the worst code. We can even see that the Silver code almost totally mitigate the PDL impairments and has its BER very close to the case without PDL. Note that with no PDL, these codes do not introduce any penalties as they are by construction redundancy free. Therefore, the use of a Polarization-Time code is always profitable

These simulation results are in agreement with the previous section analysis. Indeed the Silver and Sezginer-Sari codes perform better than the Golden code. This result is very interesting because on Rayleigh fading channel, it is the opposite and Golden code has the best performance (see Fig. 4.7). We deduce that space-time code performance and polarization-time code performance are not based on the same criteria. For example, it is obvious that the determinant criteria is not valid here and that d_{min} defined in the previous

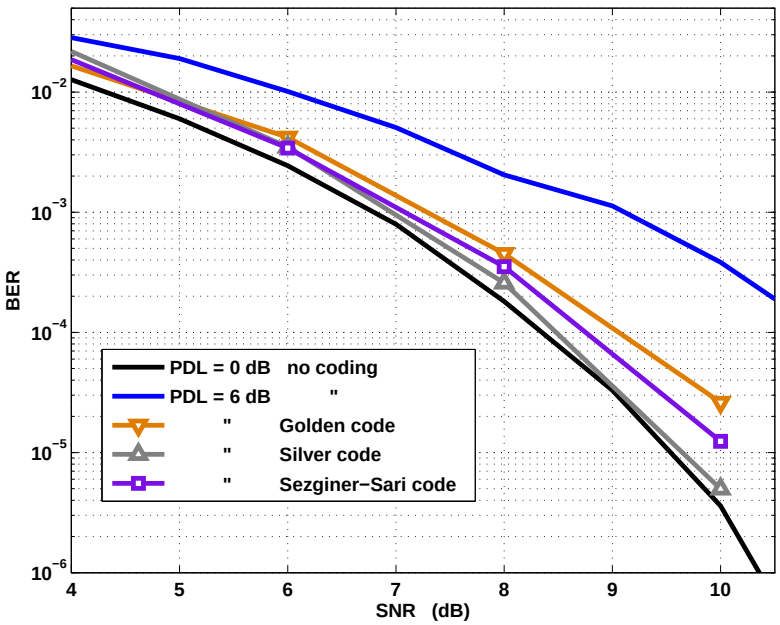


Figure 4.12: Performance of PT codes.

section is the relevant criterion. For practical values of PDL, Silver has its average minimal distance constant which explains why its performance is so close and almost equal to the PDL free error probability.

In optical transmission systems, the FEC requirement is to provide an output BER $< 10^{-12}$ for an input BER $> 10^{-3}$. The SNR penalties at BER $= 10^{-3}$ are evaluated for different PDL values on Fig. 4.13. There is a penalty about 1dB for a PDL of 3.5dB. This is approximatively the order of degradation observed experimentally by [108] in a OOK transmission. With a PDL of 6dB, the SNR penalty reaches 2.3dB. With the use of Silver code the penalties are reduced to only 0.05dB for a PDL of 3dB and 0.3dB for a PDL of 6dB. It corresponds to a reduction of about 90% of the SNR penalties.
In Fig. 4.14, PT coding is combined with FEC. We use a lattice LDPC code $\mathcal{C}(349, 30, 4)$. In the figure, the solid lines correspond to the use of the FEC and the triangle markers correspond to the use of the Silver code. There is coding gain of $\sim$ 7dB at a BER of 10^{-5} using a FEC and a Silver compared to the case with no coding. The FEC brings 4.5dB and the PT code, 2.3dB. There is less than 1dB of difference between the case with and without PDL when FEC and PT coding are employed.

The Alamouti code presented in Sec. 4.2.3.1 has not been considered because of its low rate. However, it has been shown in [109] that for very large values of PDL, Alamouti code performs better than Golden and Silver codes.

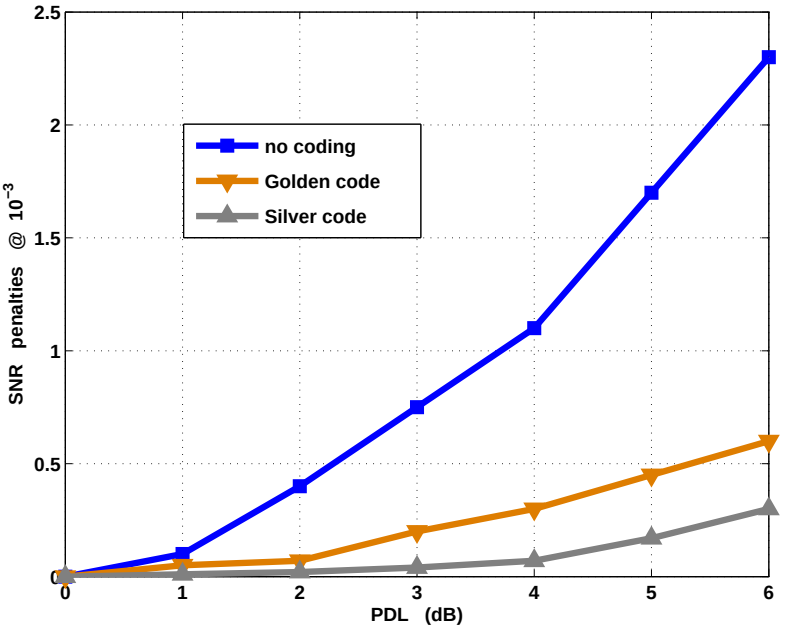


Figure 4.13: SNR penalties introduced by PDL at BER = 10^{-3} with and without PT coding.

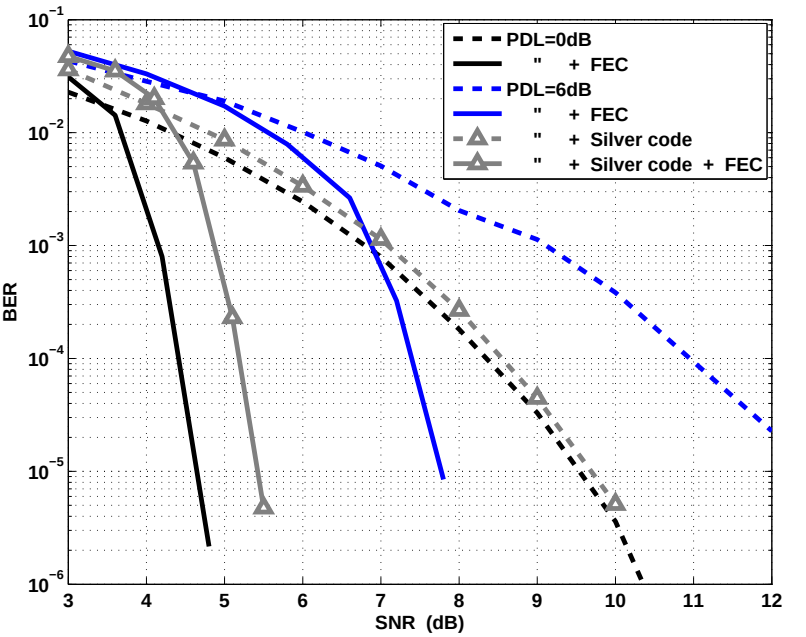


Figure 4.14: Performance of PT codes with FEC.

Alamouti symbol rate is half the symbol rate of the Silver and the Golden code. Hence, the comparison between theses codes can be realized only at an equivalent symbol rate R_s. As QPSK modulation is used for the full-rate codes ($R_s = 4$), hence a 16QAM has to be considered to obtain the same symbol rate with the Alamouti code. Using a 16QAM results in higher BER. For instance, in a AWGN channel, there is about a 4dB penalty at 10^{-5} between QPSK and 16QAM formats.
However, Alamouti code have the interesting property to have its minimal distance (defined in the previous section) being constant for every values of PDL (according to our model). So it is insensitive to PDL which confirms the conclusion in [109]. Therefore for very high PDL values, Golden and Silver code can result in more than 4dB of penalty and thus have lower performance than the Alamouti code. This observation confirms that the minimal distance is a key criterion of the performance of PT coding in optical fiber transmissions. In Fig. 4.15, the performance of the Alamouti code are plotted and compared with that of the Silver code. Unlike the Silver, there is no performance degradation caused by the increase of the PDL. However, as a 16-QAM is considered for the Alamouti code, it performs about 3.5dB less than the Silver code.

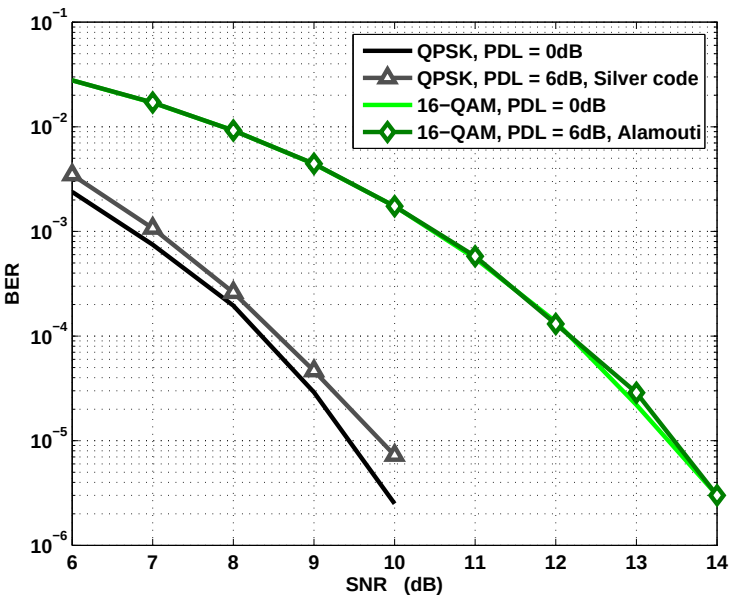


Figure 4.15: Performance of the Alamouti code (16QAM)

4.3.2.4 A second PDL model

In [110] authors have shown that PDL is a stochastic process following a Gaussian distribution. We propose a second model for the channel matrix where $\mathbf{\Gamma}_{dB}$ is a gaussian $\sim \mathcal{N}(0, \mathrm{PDL})$. $\mathbf{R}_1$ and $\mathbf{R}_2$ are still rotation matrices with uniformly distributed angles. This model is quite different compared to the the previous one. Indeed fading are not

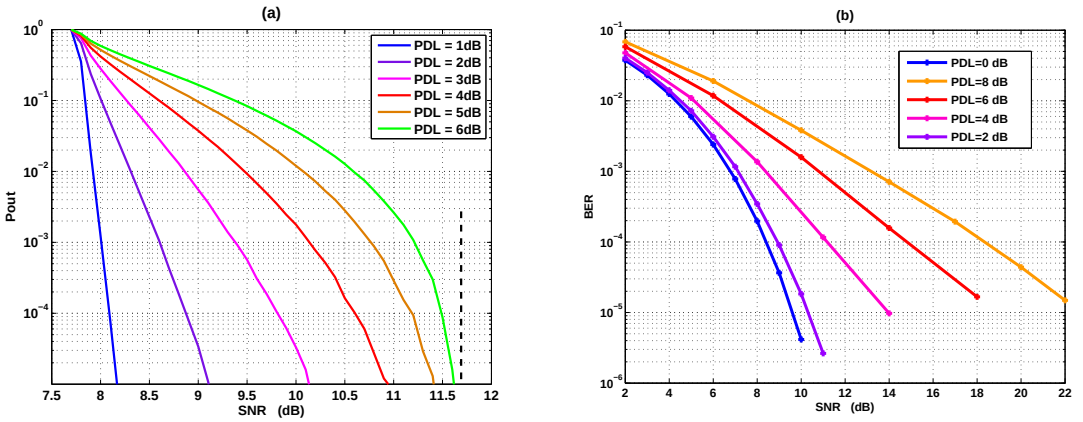


Figure 4.16: (a) Outage probability (b) BER

constant anymore but random variables.

As the channel is random, we have to use the outage probability instead of the ergodic capacity to describe the channel. The system is said to be in outage if the ergodic capacity C is less than a given spectral efficiency R_s bit/cu. When this event happens, it becomes impossible to decode correctly the information. The outage probability corresponds to the probability that the system is in outage.

$$P_{out} = \Pr\{\mathrm{C}(\mathbf{H_{PDL}}) \leq R_s\} \tag{4.57}$$

The outage probability is often represented in function of the SNR for a given rate R_s. Without PDL, the system is either always on outage ($P_{out} = 1$) or never on outage ($P_{out} = 0$). Indeed, capacity is a constant value and at a certain SNR, it becomes higher than the target rate R_s. However, with PDL, the capacity is stochastic and the outage probability decreases in function of the SNR.

In Fig. 4.16(a), the outage probability is represented for a rate $R_s = 4$ which corresponds to a PDM-QPSK format. In the figure, we can clearly see three regions. For small SNR, the system is always in outage, so $P_{out} = 1$. For high value of PDL, the system is never in outage and $P_{out} = 0$. However there is an intermediary region where the outage probability decreases in function of the SNR. The value of the PDL just determines how fast is the decrease. Note that the behavior is very different than in Rayleigh fading fading where there is always a possibility (even very small) to be in outage. Here, the system can be seen as two parallel channels having there own fading, however, unlike the Rayleigh fading case, these fading are correlated. Indeed, if one channel is under fading, the other is not and both channel can not be under total fading at the same time. So for high SNR values, the capacity of the single channel not under fading can be enough to reach the target rate R_s and the system will never be in outage.

In Fig. 4.16(b) the error probability is represented in function of the SNR. It can be seen

that unlike the previous model, the slope of the curves are functions of the PDL. Hence, this model of PDL is more severe as it results in a SNR penalty and a smaller error probability decrease.

Fig. 4.17 shows the performance of PT coding with this second PDL model. We can observe that the slopes of the error probability with and without PT coding are different. With PT coding, PDL causes only a small SNR degradation but the BER decrease is the same as without PDL. Once again, Silver outperforms Golden code and the BER with and without PDL are almost identical. The performance of PT coding is summarized in Fig. 4.18. With this model of PDL, we observe a penalty of 3.8dB for a PDL of 6dB. Using the Golden or the Silver code, the penalties are reduced to only 1dB which corresponds to a 3dB coding gain.

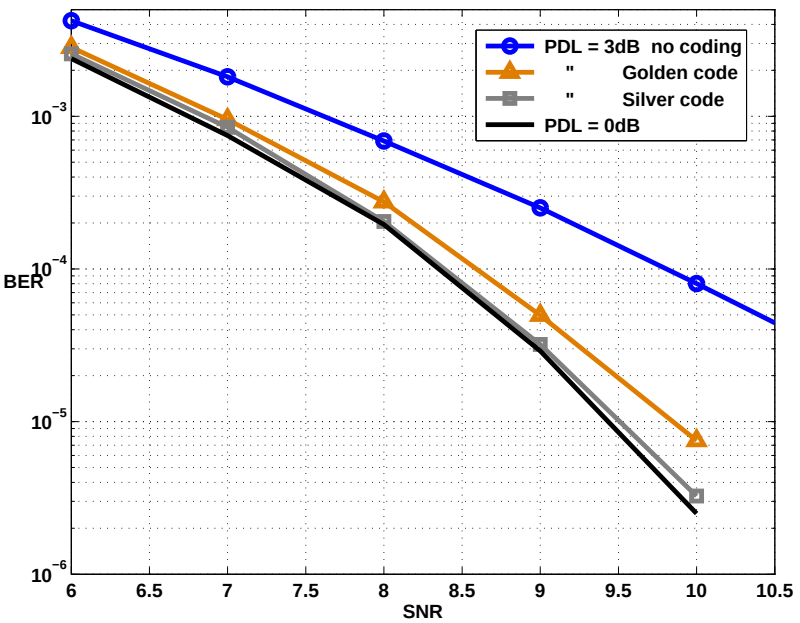


Figure 4.17: BER with and without PT coding.

Conclusion

In this chapter, we have introduced Polarization-Time coding to mitigate polarization dependent loss in optical fiber transmissions. This technique requires optical OFDM (with direct or coherent detection) in order to make the coding/decoding in the frequency domain. The proposed polarization-time codes have been selected among the ones designed in wireless communication and we have considered three famous full-rate codes: the Golden code, the Silver code and the Sezginer-Sari code. In Rayleigh fading channel, the code having the best performance is the Golden code. However we have proved that the performance in

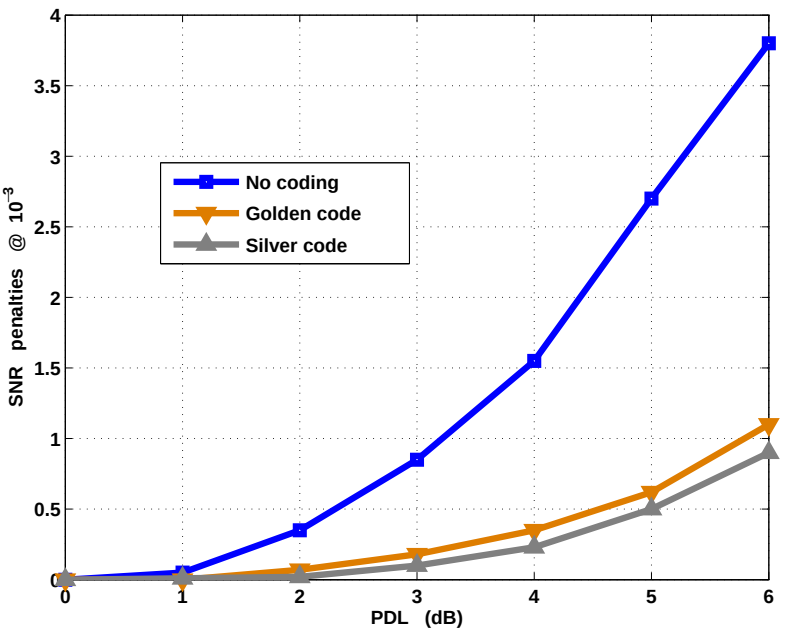


Figure 4.18: SNR penalties introduced by PDL at BER = 10^{-3} with and without PT coding.

optical fiber transmissions are quite different. Indeed the relevant criterion is the minimal distance between pair of received codeword. Silver code and Sezginer-Sari code have higher minimal distances than the Golden code and we have observed by simulations that they both outperform the Golden code.
PT coding appears to be a very efficient technique to mitigate PDL impairments. In particular the Silver code can remove quasi entirely the effect of PDL. Silver code has its average minimal distance being constant for practical values of PDL and d_{min} starts to decrease for PDL values higher than 6dB. This property is due to its quasi-orthogonal layered structured. Indeed the Sezginer-Sari code shows the same behavior but d_{min} is constant only until PDL= 3dB.
Moreover, We have shown that PT coding can be very efficiently combined with FEC. We observed that the FEC coding gain is added to PT coding gain. Therefore the FEC and the SSI scheme proposed in Chapter 2 and Chapter 3 and PT coding techniques are compatible.

Polarization-time coding represents a very interesting solution for future generation of transmissions systems. By mitigating the PDL, the quality of the transmission can be improved and distances increased.

Chapter 5

Equalization in multi-mode fiber transmissions

introduction

The optical fiber is a waveguide that allows the propagation of multiple modes. A mode is a certain spatial distribution of the electrical field in the transversal plane. Each mode corresponds to a solution of the wave equation and depends on the refractive index profile of the fiber. Hence, each mode propagates at its own velocity which can causes interferences between the different modes.
Single mode fibers (SMF) are designed in order to let the propagation of only one mode, called the fundamental mode. All the energy is focused in the core of the fiber. This kind of fiber is usually chosen for long haul transmissions. SMF are characterized by a smaller optical aperture (compared to multi-mode fibers) which makes the assembly cost higher. Therefore multi-mode fibers (MMF) have been widely adopted in access networks where the costs are a main issue [111].

Recently, MMF have attracted a lot of attention because of its multiplexing potential. Indeed, it is possible to excite different modes separately and thus multiplex the data over multiple modes. This technique is called mode group diversity multiplexing (MGDM) and demonstration at 6Gb/s [112] and 10.7Gb/s [112] have been realized over few meters. The main challenge of this kind of transmissions is to deal with the coupling between modes and the large differential mode delay (DMD). Therefore an application in long transmissions over hundreds or thousands of kilometers is only possible if these two impairments are compensated.
Efficient MIMO equalization techniques already exist in optics to deal with PMD and birefringence in polarization multiplexed systems. However, in MGDM systems, the DMD can be particularly severe and causes interference between hundreds of symbols. Moreover multiplexing of tens of modes should be considered which makes the equalization complexity,

one of the main issue.

In this chapter we will discuss about the MIMO equalization in multi-mode fiber transmissions. After presenting the MMF channel model in Sec. 5.1, we will investigate the possible methods of equalization. We will present the principle of equalization based on the channel estimation and show its performance in MMF systems in Sec. 5.2. Finally, in Sec. 5.3 we will introduce OFDM as a solution to reduce the equalization complexity.

5.1 MMF channel model

Multi-core fibers are a new design of optical fibers having multiple core waveguides as depicted in Fig. 5.1. They are a particular case of multi-mode fibers as having the energy propagating in one of the core corresponds to a specific field spatial distribution and thus a mode. In these kind of fibers, MGDM corresponds to a spatial multiplexing. Indeed a signal can be transmitted in each core. However because of the coupling, a part of the signal transmitted in one core is transferred on the others ones. Moreover depending on the fiber design, the propagation velocity can be different from a core to another and results in a differential mode delay. One of the advantage of multi-core fiber compared to regular multi-mode fibers is the control of the fiber design that helps to know and choose the amount of DMD and coupling between the modes.

The DMD can vary from few to hundreds of symbols. Note that we prefer to express the delay in symbols rather than in seconds as we are concerned by the ISI. The optical channel variation is very slow compared to the symbol duration, typically around 10kHz (100μs). At 100Gb/s (50Gsymb/s with QPSK modulation), the channel is constant over 5000000 symbols. Hence, optical channel is considered as a quasi-static channel.

Transmission on a MMF fiber can be seen as a MIMO systems as represented in Fig. 5.2. If there are N_{mode} transmitted modes the channel is equivalent to a $N \times N$ MIMO channel. If polarization multiplexing is realized on each mode, we have so $N = 2N_{mode}$ otherwise $N = N_{mode}$. In the following we will consider $N = 6$, which corresponds to either 6 modes or 3 modes with polarization multiplexing. We note T_s the symbol duration on a mode. For example, in a 100Gb/s QPSK transmission over 6 channels, the mode symbol rate is $\frac{100}{26}$Gb/s which corresponds to $T_s = 120$ps.

In order to model the optical channel, we only consider the effects of the coupling and the DMD. We extend the PMD transfer matrix function given in Eq. 1.14 to a $N \times N$ matrix.

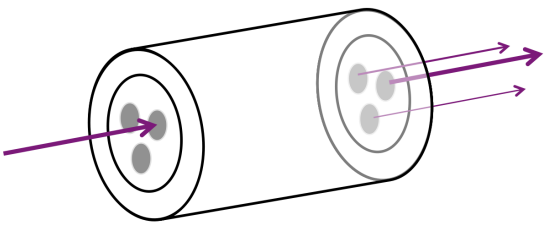


Figure 5.1: Multi-core fiber

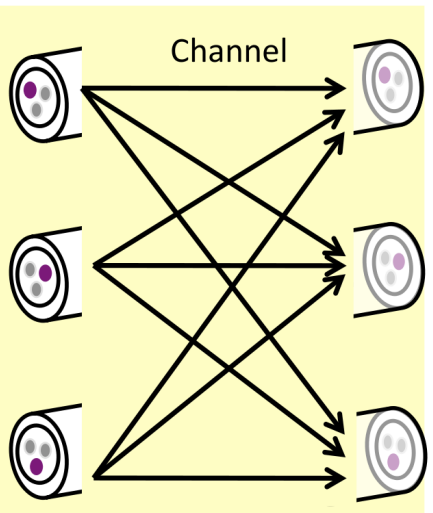


Figure 5.2: MIMO representation of a MMF transmission

For instance in the case of $N=6$ we use the following transfer function:

$$\mathbf{H}_{\mathrm{DMD}}(\omega) = \mathbf{R}_1 \begin{bmatrix} \exp(\frac{\imath\omega a_1\tau}{2}) & 0 & 0 & 0 & 0 & 0 \\ 0 & \exp(\frac{\imath\omega a_2\tau}{2}) & 0 & 0 & 0 & 0 \\ 0 & 0 & \exp(\frac{\imath\omega a_3\tau}{2}) & 0 & 0 & 0 \\ 0 & 0 & 0 & \exp(\frac{-\imath\omega a_4\tau}{2}) & 0 & 0 \\ 0 & 0 & 0 & 0 & \exp(\frac{-\imath\omega a_5\tau}{2}) & 0 \\ 0 & 0 & 0 & 0 & 0 & \exp(\frac{-\imath\omega a_6\tau}{2}) \end{bmatrix} \mathbf{R}_2 \quad (5.1)$$

$\mathbf{R}_1$ and $\mathbf{R}_2$ are two $N\times N$ random unitary matrices corresponding to the coupling between modes. τ is the maximum delay between the modes and a_i are uniformly distributed values over $[-1 \;\; 1]$. DMD is a deterministic delay and depends on the fiber design. However in our model we have considered random delays (a_i random) as we do not have an accurate knowledge of the fiber design and thus of the DMD.

If polarization multiplexing is considered, the lines have to be taken by pair and $a_{2i+1} = a_{2i+2}$ in absence of PMD. In order to take into account the effect of PMD, the transfer function should be multiplied by a diagonal bloc matrix where each block is a 2×2 matrix having the form of Eq. 1.14.

$$\mathbf{H} = \mathbf{H}_{\mathrm{DMD}} \begin{bmatrix} \mathbf{H}_{\mathrm{PMD}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_{\mathrm{PMD}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{H}_{\mathrm{PMD}} \end{bmatrix} \quad (5.2)$$

The transfer matrix $\mathbf{H}$ is defined in the frequency domain. A frequency domain multiplication corresponds to a convolution product in the time domain and can be modeled by a FIR filter (see Fig. 1.9). Therefore a $N \times N$ MIMO channel can be represented by

N^2 L-tap filters $\mathbf{h}_{i,j} = [h_{i,j}(1) \quad h_{i,j}(2) \dots h_{i,j}(L)]^T$ corresponding to the channel impulse response (CIR) in the time domain between the i^{th} and the j^{th} modes (see Fig. 5.3).

We note $x_i(k)$ and $y_i(k)$ respectively the transmitted and received symbols on the i^{th} channel at the time kT_s. The received symbols can be expressed by:

$$y_j(k) = \sum_{i=1}^{N} \sum_{l=1}^{L} h_{ij}(l) x_i(k-l+1) + n_j \tag{5.3}$$

This equation can be rewritten into a linear system:

$$\mathbf{Y} = \mathbf{HX} + \mathbf{n}$$

$$\begin{bmatrix} \mathbf{y}(k) \\ \mathbf{y}(k-1) \\ \mathbf{y}(k-2) \\ \vdots \end{bmatrix} = \begin{bmatrix} \mathbf{H}_1 & \mathbf{H}_2 & \dots & \mathbf{H}_L & 0 & 0 & \dots \\ 0 & \mathbf{H}_1 & \mathbf{H}_2 & \dots & \mathbf{H}_L & 0 & 0 \dots \\ 0 & 0 & \mathbf{H}_1 & \mathbf{H}_2 & \dots & \mathbf{H}_L & 0 \quad 0 \dots \\ & & & \ddots & & & \end{bmatrix} \begin{bmatrix} \mathbf{x}(k) \\ \mathbf{x}(k-1) \\ \mathbf{x}(k-2) \\ \vdots \end{bmatrix} + \mathbf{n} \tag{5.4}$$

where $\mathbf{y}(k) = [y_1(k) \; y_2(k) \; \dots y_N(k)]^T$, $\mathbf{x}(k) = [x_1(k) \; x_2(k) \; \dots x_N(k)]^T$. Each sub-matrix $\mathbf{H}_i$ represents the i^{th} taps of every channels:

$$\mathbf{H}_i = \begin{bmatrix} h_{1,1}(i) & h_{2,1}(i) & \dots & h_{N,1}(i) \\ h_{1,2}(i) & h_{2,2}(i) & & \vdots \\ \vdots & & \ddots & \\ h_{1,N}(i) & \dots & & h_{N,N}(i) \end{bmatrix} \tag{5.5}$$

If we assume that the signal and the noise energies are equal on each channel, their auto covariance matrices can be expressed by:

$$R_{\mathbf{x}} = \sigma_{\mathbf{x}}^2 \mathbf{I}_N \tag{5.6}$$

$$R_{\mathbf{n}} = \sigma_{\mathbf{n}}^2 \mathbf{I}_N \tag{5.7}$$

where σ_x^2 and σ_n^2 are the respectively the modulated symbol energy and the noise variance. In the following we suppose that noise and signal are uncorrelated ($\mathbb{E}\mathbf{xn} = 0$).

5.1.1 Fractionally spaced equalization

Until now we have assumed that the signal is sampled at 1 sample per T_s which is theoretically accurate but practically difficult to realize as the signal bandwidth is often larger than $f_s = 1/T_s$. Therefore sampling at f_s is not enough to satisfy Nyquist theorem and oversampling is necessary. In general the oversampling factor r_{ov} is chosen as an integer in order to ease the implementation. So r_{ov} samples of the signal are taken every T_s. The equalizer has to deal with a larger number of sample which increases the complexity.

The channel and the equivalent discrete channel model are represented on Fig.5.4(a) and (b). The i^{th} channel outputs after oversampling are $y_i(k\frac{T_s}{r_{ov}})$ and we note the samples

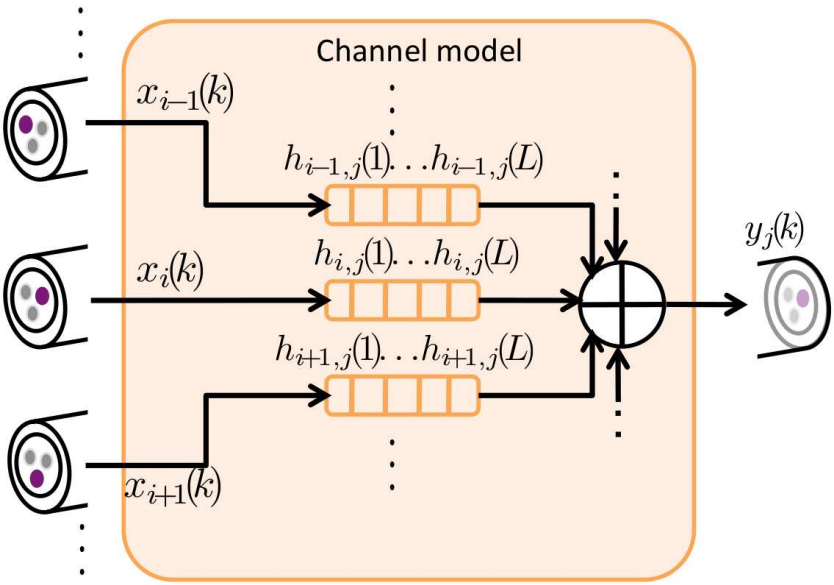


Figure 5.3: FIR channel model of a MMF transmission

$y_i(k)$ for clarity. The modulated symbols $u_i(k'T_s)$ (noted $u_i(k')$) are sent at each T_s so we consider as zeros all the samples between.

$$x_i(k) = \begin{cases} u_i(\frac{k}{r_{ov}}) & \text{if } k = 0 \bmod r_{ov} \\ 0 & \text{else} \end{cases} \tag{5.8}$$

where k is the index of the oversampling rate and k' of the symbol rate. As in Eq.5.3 the outputs can be expressed by:

$$y_j(k) \quad = \quad \sum_{i=1}^{N} \sum_{l=1}^{r_{ov}L} h_{ij}(l) x_i(k-l+1) \quad + n_j \tag{5.9}$$

The outputs can be grouped by sampling time p:

$$y_j(k\frac{Ts}{r_{ov}}) = y_j(k'T_s + p\frac{Ts}{r_{ov}}) = y_j(r_{ov}k'\frac{Ts}{r_{ov}} + p\frac{Ts}{r_{ov}}) \qquad \begin{array}{l} k' \in \mathbb{N} \\ 0 \leq p \leq (n_{ov}-1) \end{array}$$

If we drop once again the notation $\frac{Ts}{r_{ov}}$, we can rewrite the previous expressions as:

$$\begin{aligned} y_j(r_{ov}k'+p) \quad &= \quad \sum_{i=1}^{N} \sum_{l=1}^{r_{ov}L} h_{ij}(l) x_i(r_{ov}k'+p-l+1) \quad + n_j \\ y_j(r_{ov}k'+p) \quad &= \quad \sum_{i=1}^{N} \sum_{l'=1}^{L-1} h_{ij}(r_{ov}l'+p+1) x_i(r_{ov}(k'-l')) \quad + n_j \end{aligned} \tag{5.10}$$

This shows that the received samples only depend on a fraction of the CIR coefficients. Therefore the system can be seen as r_{ov} parallel MIMO channels as depicted on Fig.5.4(c).

For instance, in the case of an oversampling of 2 samples per T_s, the even received samples $y_j^0(k') = y(2k')$ only depend on the odd channel taps $h_{ij}^1(l') = h_{ij}(2l'+1)$ and the other way around.

At the reception, the samples are grouped by sampling time p. Then the channel estimation is realized independently for each group in order to estimate $\mathbf{H}^p$, the CIR experienced by the samples $y_j^p(k') = y_j(k'T_s + p\frac{T_s}{n_{ov}})$. Each equalizers has to correct only a fraction of the CIR so equalization of an oversampled signal is often referred as fractionally spaced equalization (FSE).

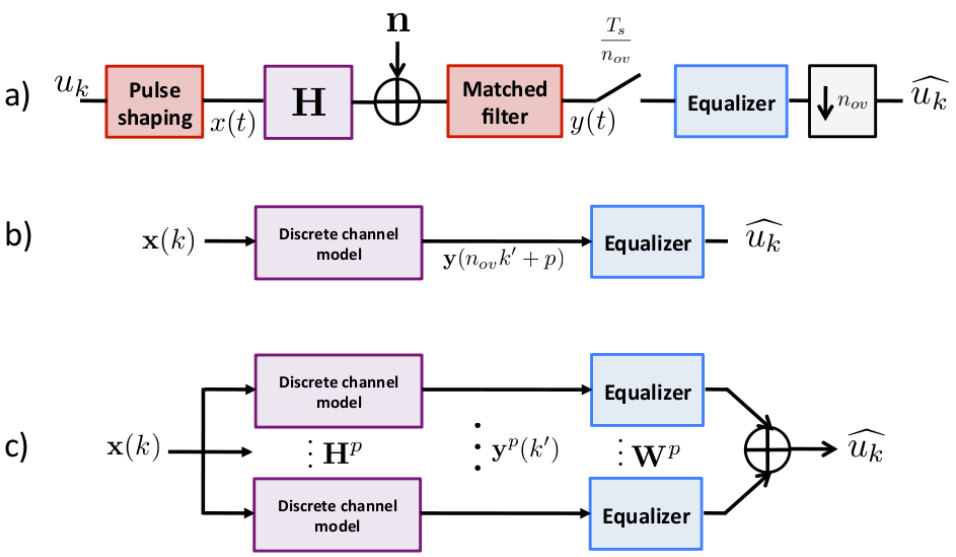


Figure 5.4: a) Transmission channel with oversampling b) Discrete channel model c) Parallel discrete channel model

Eq. 5.10 can be rewritten in a matrix representation as:

$$\mathbf{Y}_{\text{FSE}} = \mathbf{H}_{\text{FSE}}\mathbf{X} + \mathbf{n}$$

$$\begin{bmatrix} \mathbf{y}^0(k) \\ \mathbf{y}^0(k-1) \\ \mathbf{y}^0(k-2) \\ \vdots \\ \mathbf{y}^p(k) \\ \mathbf{y}^p(k-1) \\ \mathbf{y}^p(k-2) \\ \vdots \end{bmatrix} = \begin{bmatrix} \mathbf{H}_1^0 & \mathbf{H}_2^0 & \dots & \mathbf{H}_L^0 & 0 & 0 & \dots \\ 0 & \mathbf{H}_1^0 & \mathbf{H}_2^0 & \dots & \mathbf{H}_L^0 & 0 & 0 \dots \\ 0 & 0 & \mathbf{H}_1^0 & \mathbf{H}_2^0 & \dots & \mathbf{H}_L^0 & 0 \dots \\ & & & \ddots & & & \\ \mathbf{H}_1^p & \mathbf{H}_2^p & \dots & \mathbf{H}_L^p & 0 & 0 & \dots \\ 0 & \mathbf{H}_1^p & \mathbf{H}_2^p & \dots & \mathbf{H}_L^p & 0 & 0 \dots \\ 0 & 0 & \mathbf{H}_1^p & \mathbf{H}_2^p & \dots & \mathbf{H}_L^p & 0 \dots \\ & & & \ddots & & & \end{bmatrix} \begin{bmatrix} \mathbf{x}(k) \\ \mathbf{x}(k-1) \\ \mathbf{x}(k-2) \\ \vdots \end{bmatrix} + \mathbf{n} \tag{5.11}$$

where $\mathbf{y}^p(k) = [y_1^p(k) \dots y_N^p(k)]^T$.

In the following, we are going to focus on the channel estimation and equalization principles. In a matter of clarity, they will be presented in the case of no oversampling. However in

our numerical simulation, an oversampling factor of 2 has been used.

5.2 Equalization by channel estimation

It exists two main equalization methods:

1. Blind adaptive equalizer such as CMA or DD-LMS which are currently used in polarization multiplexed systems. However for a large number of channels ($N \geq 4$) the convergence of such techniques may be slow or problematic.

2. Channel estimation with a training sequence before to realize the equalization. This reduces the transmission effective bit-rate due to the introduction of an overhead but it does not suffers of convergence problems. Note that as the channel is varying very slowly, there is no need to update the equalizer very often. Hence the introduced overhead is very limited and the complexity is smaller than with adaptive equalization.

According to the property of the MMF channel (slow variation, large number of modes), we will focus in this chapter on the equalization by channel estimation. Note that as the channel is estimated, the equalization can be seen as a decoding.

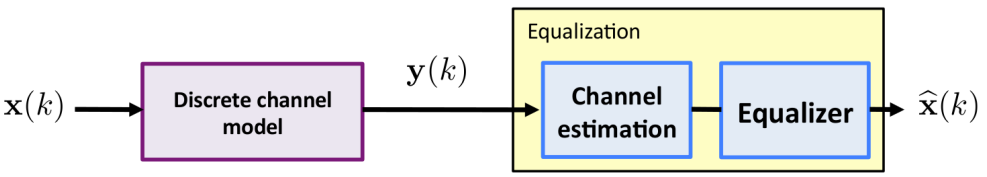


Figure 5.5: Equalization principle

We consider the optical transmission channel represented in 5.5. $\mathbf{x}(k)$, $\mathbf{y}(k)$ are respectively, the transmitted and received symbols at the sampling rate kT_s/r_{ov}. The discrete channel takes into account the pulse shaping filter, the DMD and the matched filter at the reception. In our numerical simulations, we uses a raised cosine pulse shaping [4]:

$$p(t) = \sin c(t/T_s)\frac{\cos\left(\pi\rho_{off}t/T_s\right)}{1-4\rho_{off}^2t^2/T^2} \tag{5.12}$$

where $\rho_{off} \in [0\ 1]$ is the roll-off factor that characterize the pulse decay. A large value for ρ_{off} results in a fast decay and thus a short pulse which reduces the ISI caused by sampling timing error. However it leads to larger spectrum bandwidth.
The equalizations is realized based on the channel knowledge therefore the first step is the estimation of the discrete channel.

5.2.1 MIMO channel estimation

The principle of channel estimation is to transmit periodically a training sequence known at the transmitter and the receiver in order to estimate the channel impulse response. When the channel is slowly varying, training sequences are usually sent by block before the data. In Sec. 5.1, it has been seen that the channel can be interpreted as FIR filters. The channel estimation corresponds to the computation of these filter taps.

We first rewrite the channel model of Eq. 5.4 into a new form. $s_i(k)$ denotes the training symbol emitted on the i^{th} mode. N_s is the number of training symbols per mode and L is the length of the CIR. Let us define:

$$\mathbf{h_i} = [\mathbf{h}_{1,i}^T \ \mathbf{h}_{2,i}^T \ \dots \ \mathbf{h}_{N,i}^T]^T \tag{5.13}$$

$$\mathbf{S}_i = \begin{bmatrix} s_i(L) & s_i(L-1) & \dots & s_i(1) \\ s_i(L+1) & s_i(L) & \dots & s_i(2) \\ & & \ddots & \\ s_i(N_s) & \dots & \dots & s_i(N_s-L+1) \end{bmatrix} \tag{5.14}$$

Eq. 5.4 can be rewritten into an equivalent linear channel model:

$$\mathbf{y}_i = \mathbf{S}\ \mathbf{h_i} + \mathbf{n} \tag{5.15}$$

where $\mathbf{S} = [\mathbf{S_1} \ \ \mathbf{S_2} \dots \mathbf{S_N}]$, $\mathbf{n}$ is the noise vector and $\mathbf{y}_i = [y_i(L) \ \dots y_i(N_s)]^T$ are the symbols received on the i^{th} mode.

The channel estimation is based on the least-square error (LSE) criterion presented in Appendix D. Indeed, although mean-square error (MSE) criterion is more accurate, it requires the knowledge of the channel statistic which is not always available at the receiver. Hence we search the optimal value of $\widehat{\mathbf{h}}_\mathbf{i}$ that minimizes $\|\mathbf{y}_i - \mathbf{S}\widehat{\mathbf{h}}_\mathbf{i}\|^2$. The solution is given by Eq. D.7:

$$\widehat{\mathbf{h}}_\mathbf{i} = (\mathbf{S}^H\mathbf{S})^{-1}\mathbf{S}^H\ \mathbf{y}_i \qquad 1 \leq i \leq N \tag{5.16}$$

$\widehat{\mathbf{h}}_\mathbf{i}$ has NL coefficients (N L-tap FIR filter in parallel as depicted in Fig. 5.3). However we can see it as an unique L-tap filter where a tap is the N coefficients $\widehat{\mathbf{h}}_\mathbf{i}(k) = [\widehat{h}_{1,i}(k) \ \dots \ \widehat{h}_{N,i}(k)]$ (see Fig. 5.6).

In Fig. 5.7, we represent the channel estimation mean square error defined as:

$$\text{MSE} = \|\mathbf{h}_{1:N} - \widehat{\mathbf{h}}_{1:N}\|^2 \tag{5.17}$$

In function of the SNR. The MSE describes the accuracy of the channel estimation. In our simulations, the DMD is 600ps so $5T_s$ at 100Gb/s. The pulse shape is a raised cosine with a roll-off factor $\rho_{off} = 0.2$ in (a) and $\rho_{off} = 0.5$ in (b).
The training sequence length is defined in number of symbols per estimated taps (noted symb/tap). We use in our numerical simulation a training sequence of 100 symbols per

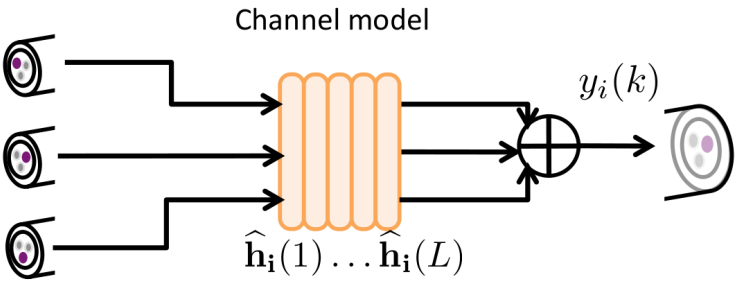


Figure 5.6: Channel model

estimated taps, which means that if the channel is estimated over 11 taps, there are 1100 training symbols transmitted on each modes. There are 6 modes so the total training sequence length is 6600 symbols. Note that as we use an oversampling factor of 2, one tap of the filter represents only half a symbol duration. Therefore if the ISI occur among 5 symbols, at least 10 taps are necessary (5 taps for the odd samples and 5 taps for the even samples).

In Fig. 5.7, we have represented the Cramer-Rao lower bound (CRB) [113] which is a general bound on the minimum MSE estimation of a random parameter. It depends on the size of the training sequence. The LSE estimator reaches this bound. However we observe that the MSE tends to this bound when the number of estimated taps increases. Indeed, because of the pulse shape and the DMD there are some ISI and the CIR is large. Consequently, the estimation has to be realized over a number of taps larger than the discrete CIR in order to reach the CRB.

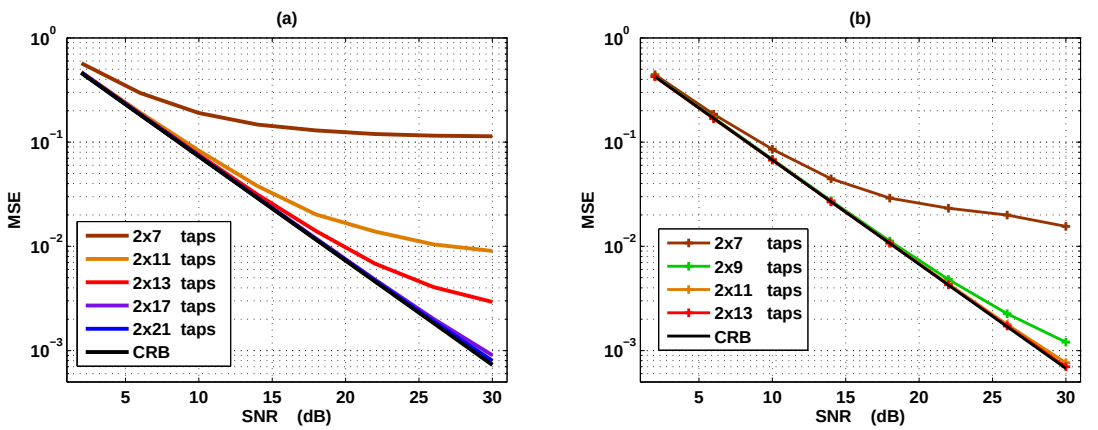


Figure 5.7: MSE function of the SNR for various number of estimated taps and a training sequence of 100 symb/tap. The pulse shape is a raised cosine with: (a) $\rho_{off} = 0.2$ (b) $\rho_{off} = 0.5$

The raised cosine has an infinite response but if the roll off factor is large enough, the response decreases very fast and introduces some significant ISI only over a limited number of symbols. The discrete channel impulse response is actually larger than just the dispersion delay and strongly depends on the pulse shape. In Fig. 5.7 (a), a raised cosine with a roll-off factor $\rho_{off} = 0.2$ is employed and estimating the channel on 13 taps is not enough. On the other hand, with a roll-off factor $\rho_{off} = 0.5$ the discrete CIR is about 9 taps long. Therefore a shorter pulse shape results in a shorter discrete channel and thus less taps to estimate.

Fig. 5.8 shows the decrease of the MSE when the number of training symbols increases. The MSE has been estimated for a SNR of 20dB and a training sequence of 100 symb/tap.

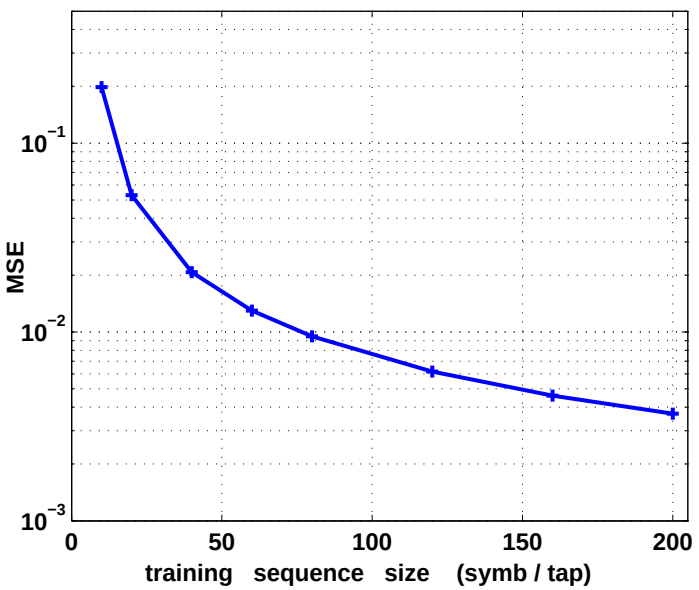


Figure 5.8: MSE function of the training sequence size. The CIR is estimated over 2×11 taps.

5.2.2 Finite-length MIMO MMSE equalizer

A $N \times N$ MIMO equalizer is a bench of N^2 N_W-tap FIR filters as depicted in Fig. 5.9. $w_{i,j}(k)$ corresponds to the k^{th} tap of the FIR filter between the i^{th} input channel and the j^{th} output channel. The equalizer can be seen as a N_W-tap filter $\mathbf{W} = [\mathbf{W}_1 \dots \mathbf{W}_{N_W}]$ with:

$$\mathbf{W}_k = \begin{bmatrix} w_{1,1}(k) & w_{2,1}(k) & \dots & w_{N,1}(k) \\ w_{1,2}(k) & w_{2,2}(k) & & \\ \vdots & & \ddots & \\ w_{1,N}(k) & \dots & & w_{1,N}(k) \end{bmatrix} \tag{5.18}$$

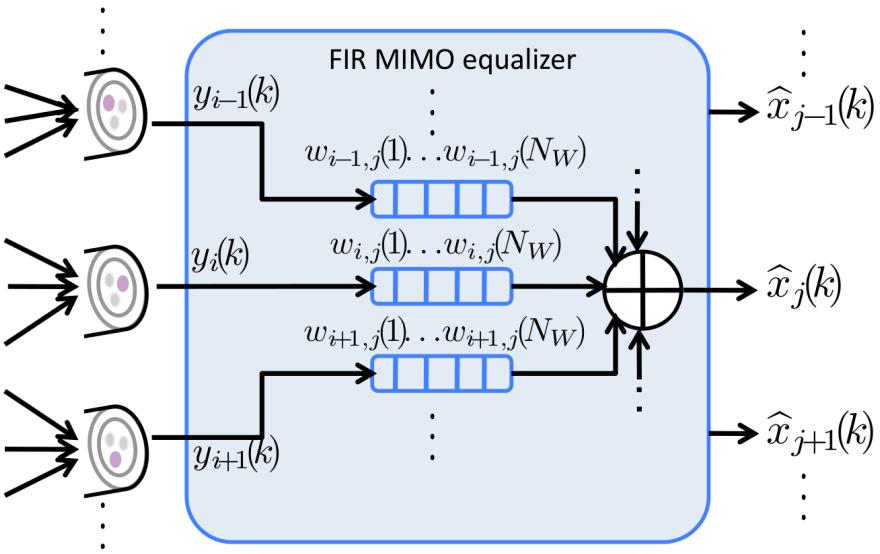


Figure 5.9: MIMO FIR equalizer

The equalizer estimates $\mathbf{x}(k)$ or a delayed version of it, $\mathbf{x}(k-\Delta)$ as:

$$\begin{aligned} \widehat{\mathbf{x}}(k-\Delta) &= \mathbf{W}\ \mathbf{Y}_{k:k-N_W+1} \\ &= [\mathbf{W}_1\ \mathbf{W}_2 \ldots \mathbf{W}_{N_W}] \begin{bmatrix} \mathbf{y}(k) \\ \mathbf{y}(k-1) \\ \mathbf{y}(k-2) \\ \vdots \\ \mathbf{y}(k-N_W+1) \end{bmatrix} \end{aligned} \tag{5.19}$$

where $\mathbf{y}(k) = [y_1(k)\ y_2(k) \ldots y_{N_W}(k)]$ and Δ is the delay .

Finding the linear equalizer (or estimator) minimizing the mean square error (MMSE) $\mathbb{E}\|\widehat{\mathbf{x}}(k-\Delta)-\mathbf{x}(k-\Delta)\|^2$ corresponds to the problem defined by Eq. D.2 and the optimal solution is given by Eq. D.4:

$$\mathbf{W}_{\text{MMSE}} = R_{\mathbf{x}(k-\Delta)\mathbf{Y}}\ R_{\mathbf{Y}}^{-1} \tag{5.20}$$

From Eq. 5.4, we have $\mathbf{Y} = \widehat{\mathbf{H}}\mathbf{X} + \mathbf{n}$ with $\mathbf{X} = [\mathbf{x}(k) \ldots \mathbf{x}(k-L-N_W+2)]^T$ and $\widehat{\mathbf{H}}$ is the channel transfer matrix whose coefficient have been estimated by the channel estimation.

Hence, the auto-covariance matrices can be computed as follows:

$$\begin{aligned} R_{\mathbf{x}(k-\Delta)\mathbf{Y}} &= \mathbb{E}\ \mathbf{x}(k-\Delta)\mathbf{Y}^H \\ &= \mathbb{E}\left[\mathbf{x}(k-\Delta)\mathbf{X}^H\ \widehat{\mathbf{H}}^H\right] \\ &= \mathbb{E}\left[\mathbf{x}(k-\Delta)[\mathbf{x}(k)^H \dots \mathbf{x}(k-L+N_W)^H]\ \widehat{\mathbf{H}}^H\right] \\ &= [\underbrace{\mathbf{0}\dots\mathbf{0}}_{\Delta}\ R_{\mathbf{x}}\ \mathbf{0}\dots]\ \widehat{\mathbf{H}}^H \end{aligned} \tag{5.21}$$

$$\begin{aligned} R_{\mathbf{Y}} &= \mathbb{E}\ \mathbf{Y}\mathbf{Y}^H \\ &= \widehat{\mathbf{H}} R_{\mathbf{x}} \widehat{\mathbf{H}}^H + R_{\mathbf{n}} \end{aligned} \tag{5.22}$$

Finally the MMSE filter is expressed by:

$$\mathbf{W}_{\text{MMSE}} = \sigma_{\text{x}}^2 \mathbf{\Phi}_{\mathbf{\Delta}} \widehat{\mathbf{H}}^H \left(\sigma_{\text{x}}^2 \widehat{\mathbf{H}}\widehat{\mathbf{H}}^H + \sigma_{\text{n}}^2 \mathbf{I}\right)^{-1} \tag{5.23}$$

with $\mathbf{\Phi}_{\mathbf{\Delta}} = [0\ \ 0\dots 0\ \ \mathbf{I}\ \ 0\dots]$ being a $N \times N(N_W + L - 1)$ zeros bloc matrix with the identity at the position $\Delta+1$.

Fig. 5.10 presents the performance of the MMSE equalizer for different filter length Nini. We can see that the performance converges when the training sequence length increases. However it does not converge to the optimal performance (represented in black in the plot and corresponding to the maximum likelihood detection). Indeed the filter has only 2×7 taps which is smaller than the discrete CIR. Therefore, by estimating the channel with more taps, we obtain better performance.

In Fig. 5.11, we represent the BER function of the training sequence length at a SNR of 10dB. The equalizer length (thus the estimation) corresponds to 2×7, 2×11 and 2×19 taps. A larger filter performs closer to the optimal performance. We observe that performance converges when the training sequence length increases. A length of 100 or 200 samp/tap are enough to ensure near convergence performance.

5.2.3 Experimental data analysis

In this section, we evaluate the performance of the channel estimation and the MMSE equalization on experimental data coming from a polarization multiplexed transmission. The transmitter scheme is represented on Fig. 5.12. The data are generated by 4 pseudo-random binary sequences (PRBS) at 7Gb/s. The length of each PRBS is $2^{15}-1$ bits. The data are then multiplexed into a single bit sequence at 28Gb/s corresponding to the in-phase signal. The quadrature signal is obtained by cyclically shift the in-phase signal by 50 bits. The I/Q components are transformed into an optical QPSK signal by the optical I/Q modulator. The bits zero are converted into a negative amplitude and the bits one

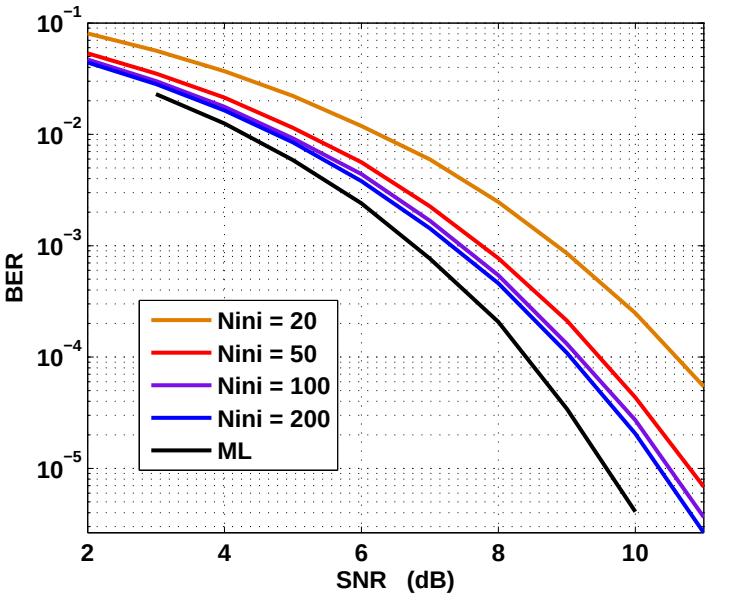


Figure 5.10: Performance of MMSE equalizer in the case of $6T_s$ DMD, $\rho_{off} = 0.2$ and a 2×7-taps filter

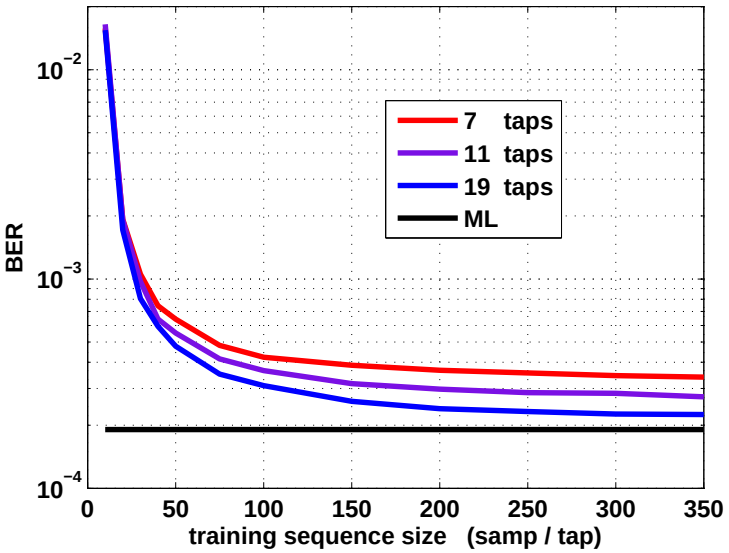


Figure 5.11: BER at SNR=10dB in function of the equalizer length

into a positive amplitude. For instance '01' is transformed into the QPSK symbol $-1+\imath$. The optical signal is separated in two by a 3dB coupler, then one branch is delayed by a certain number of symbol duration and finally the two optical signals are combined by a polarization beam splitter.

The signal is received by an optical coherent receiver and digitalized by four ADC corre-

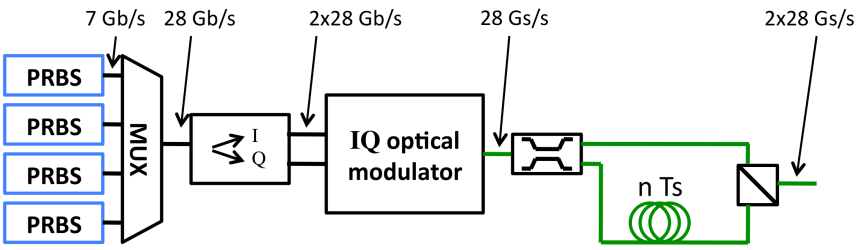


Figure 5.12: transmitter experimental setup

sponding to the in-phase and quadrature components of each polarization. The sampling rate of the ADC is 50GHz, which corresponds approximatively to 1.7 samples per symbol duration.

To ease the equalization, the signal is resampled into an integer number of samples per symbol duration. This can be performed either in the time domain or in the frequency domain as depicted in Fig. 5.13. Sampling at 1.7 samp/T_s is enough to satisfy Nyquist theorem which means that the analog signal can be perfectly reconstructed from the samples. Therefore the samples corresponding to 2 samp/Ts sampling can all be computed using the time convolution by a *sinc* function. However, re-sampling can be achieved in a more efficient manner in the frequency domain. The sampled signal spectrum is computed by an FFT and corresponds to the optical signal spectrum periodically repeated with a $1.7/T_s$ spacing. Zeros are then added in order to increase the distance until having a $2/T_s$ spacing between the spectrums. Finally, the re-sampled signal is obtained by an iFFT.

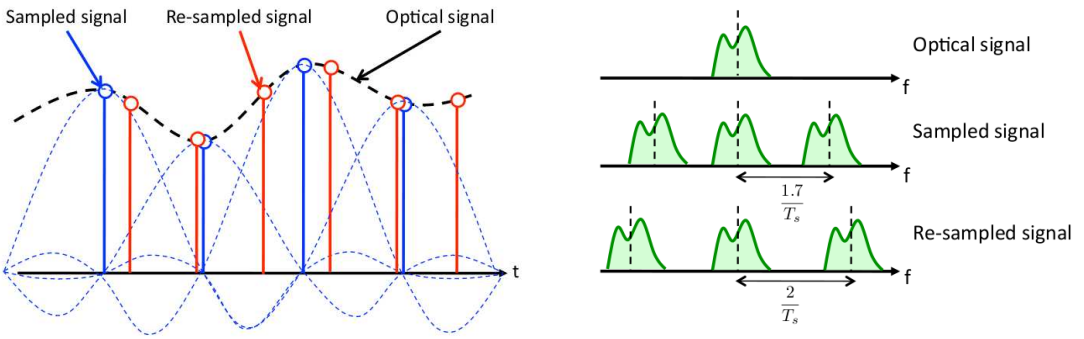


Figure 5.13: Re-sampling principle: (left) in the time domain, (right) in the frequency domain

In order to realize the channel estimation, the knowledge of the training sequence is required. We assume that first received symbols corresponds to the training sequence. However in practice, we do not know which part of the PRBS sequence, the first received symbols are. Therefore we have to synchronize the received signal with the PRBS. This is realized by searching in transmitted sequence which part corresponds to the first received

symbols. Each time, we compute $\widehat{\mathbf{h}}$ using Eq. 5.16 and synchronization corresponds to the minimal value of $\|\mathbf{y}-\mathbf{S}\widehat{\mathbf{h}}\|$.

Note that the channel estimation presented in Sec. 5.2.1 supposes that the channel is constant during the transmission of the training sequence. However due to the frequency offset of the laser, there is a phase shift which makes the estimation impossible if the offset is not compensated before. As presented in [114], a joint channel and frequency offset estimation can be performed based on training sequences. However in our simulations the initial frequency offset has been obtained by first performing the CMA (which is not sensitive to phase shifts) and then evaluating the offset on the equalized data as in [27].

After the channel estimation, MMSE equalization is performed. The PMD and the birefringence can be compensated by the equalizer because they are varying very slowly however the phase shift due to the frequency jitter and the phase noise can not. Therefore, the carrier phase has to be recovered and tracked using for instance the Viterbi-Viterbi algorithm and the frequency offset compensated as discussed in [115].

Fig. 5.14 represents the MMSE equalizer performance compared to the CMA performance. The training sequence length is 100 symbols per taps and the equalizer length is 2×7 taps. Note that as it is a back-to-back experiment (no DMD) and as NRZ pulse shape is employed (no ISI), a shorter equalizer may have been chosen. We can see the the MMSE equalizer slightly outperforms the CMA.

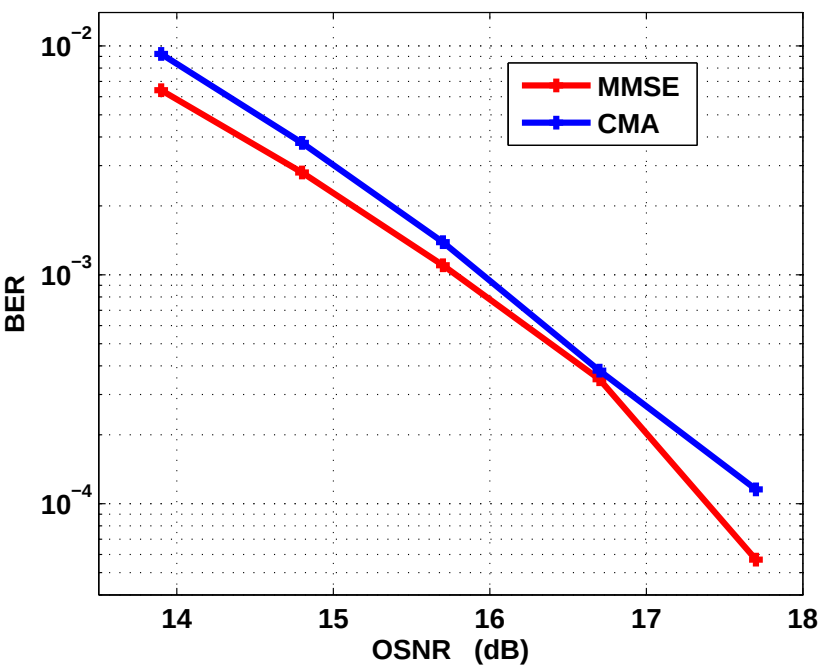


Figure 5.14: Performance of MMSE equalization on a PDM-QPSK transmission

We use the back-to-back experimental data as input signals of a MMF transmission simulation (see Fig. 5.15). In order to generate the signals on other modes, the experimental data have been delayed by 144 and 341 T_s. As we consider a polarization multiplexing, three modes correspond to a 6×6 MIMO system.

In the simulation the equalizer length is 2×7 taps. We can see in Fig. 5.16 that for small values of DMD, the equalizer performs very close to the back-to-back performance which means that the ISI is almost perfectly compensated. However, for DMD$\geq 5.5T_s$ the performance is degradated.

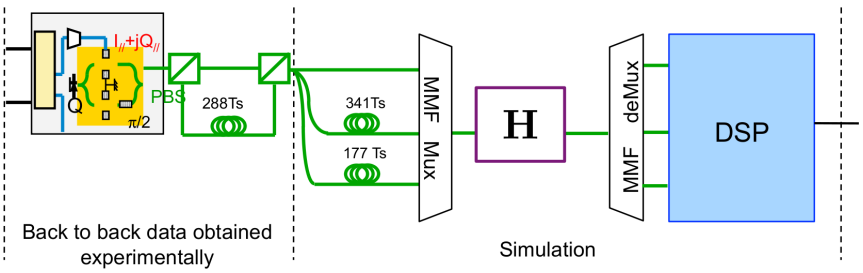


Figure 5.15: Simulation of an MMF transmission based on experimental back-to-back data

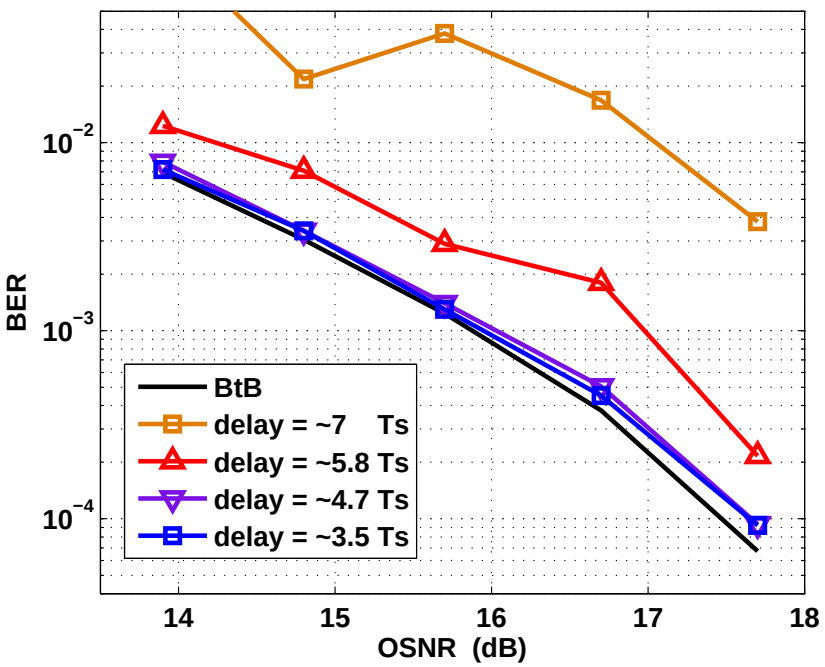


Figure 5.16: Performance of MMSE equalization in a MMF transmission based on experimental data

5.2.4 Complexity and overhead

The equalization complexity includes the CIR estimation (Eq. 5.16) and the equalizer coefficient computation(Eq. 5.23).

In Fig. 5.10 we have observed that a training sequence with 120 symbols per taps gives enough accuracy to the channel estimation. However a tap actually corresponds to $N=6$ FIR coefficients. Therefore we define $k_{tr} = 20$, the number of training symbols per FIR coefficient. The total number of training symbols is equal to $k_{tr}N^2L$.
The channel estimation is a matrix multiplication between $\mathbf{A} = (\mathbf{S}^H\mathbf{S})^{-1}\mathbf{S}^H$ and the received symbols $\mathbf{y}_i$. $\mathbf{A}$ is a $(NL)\times(k_{tr}NL-L+1)$ matrix and is perfectly known at the receiver (it only depends on the training sequence). Therefore the channel estimation complexity (in multiplication number) is:

$$\sim \mathcal{O}\left(k_{tr}^2 L^3 N^4\right) \tag{5.24}$$

At 100Gb/s, the channel is considered constant over 5000000 symbols. Some of the symbols belong to the training sequence, thus are not data, and introduce an overhead. Fig. 5.17 shows the overhead introduced by the training sequence and the complexity of the channel estimation function of the CIR length and the number of channels.
For a $N = 14$ (14 modes or 7 polarization multiplexed modes) and a training sequence length of 200 symb/taps, the overhead is about 15% whereas with $N=6$ it corresponds to only 3%. In the same way, we observe that the channel estimation complexity can be very large for large values of N.

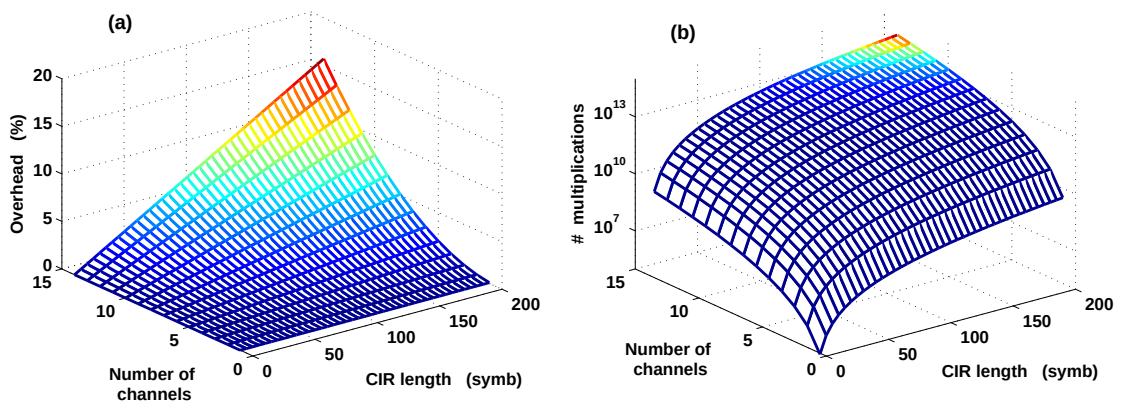


Figure 5.17: (a) Overhead introduced by training sequence at 100Gb/s, (b) CIR computation complexity

The FIR equalizer coefficients are computed following Eq. 5.23. $\mathbf{H}$ is a $(NN_W)\times(N(N_W+L-1))$ matrix thus, the complexity is $\sim \mathcal{O}\left(N^3N_W^3 + N_W^2L\right)$ for $\mathbf{H}\mathbf{H}^H$, $\sim \mathcal{O}\left(N^3N_W^3\right)$ for the matrix inversion and $\sim \mathcal{O}\left(N^3N_W^2\right)$ for the $\mathbf{\Phi_\Delta}\mathbf{H}^H$ multiplication.

The size of the equalizer has to be at least as large as the CIR length so, $N_W = r_{ov}L$. The total complexity of the MMSE filter coefficient computation is equal to:

$$\sim \mathcal{O}\left(N^3\left[(2f_{ov}^3 + f_{ov}^2)L^3 + f_{ov}^2L^2\right]\right) \tag{5.25}$$

and is represented in Fig. 5.18 for $f_{ov} = 2$.

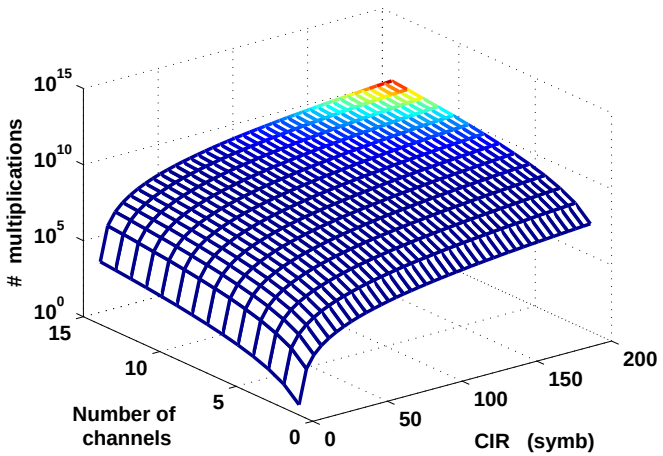


Figure 5.18: MMSE filter coefficient computation complexity

Once the filter coefficients found, the modulated symbols can be estimated as presented in Fig. 5.9 which requires NL multiplications for each output.

5.3 Optical MMF OFDM

Realizing the equalization in the time domain can result in a high complexity when mode dispersion is large and when many modes are employed. Using OFDM can reduce the complexity because only 1-tap equalizers are necessary in the frequency domain. Therefore, the computation of the equalizer coefficients is reduced. However, OFDM introduces a cyclic prefix that increases the overhead. A description of optical OFDM system has been given in Sec. 4.1.

5.3.1 OFDM equalization complexity

We denote N_{FFT} the size of the FFT and the cyclic prefix length is assumed to be superior to the discrete CIR. Channel estimation is realized in the time domain and then converted in the frequency domain by:

$$\widetilde{\mathbf{h}}_{i,j} = \mathbf{F}\mathbf{h}_{i,j} \tag{5.26}$$

where $\mathbf{F}$ is a $N_{FFT} \times L$ Fourier transform unitary matrix and $\widetilde{\mathbf{h}}_{i,j} = [\tilde{h}_{i,j}(1) \ldots \tilde{h}_{i,j}(N_{\mathrm{FFT}})]^T$. So converting the N^2 CIR in the frequency domain requires $\sim \mathcal{O}\left(N^2 N_{\mathrm{FFT}} L\right)$

Once the channels for each sub-carrier are known, we have to compute the MMSE equalizer coefficients using Eq. 5.23:

$$\mathbf{W}_i = \sigma_{\mathrm{x}}^2 \widetilde{\mathbf{H}}_i^H \left(\sigma_{\mathrm{x}}^2 \widetilde{\mathbf{H}}_i \widetilde{\mathbf{H}}_i^H + \sigma_{\mathrm{n}}^2 \mathbf{I} \right)^{-1} \qquad i = 1 \dots N_{\mathrm{FFT}} \tag{5.27}$$

where $\widetilde{\mathbf{H}}_i$ is now a $N \times N$ matrix:

$$\widetilde{\mathbf{H}}_i = \begin{bmatrix} \tilde{h}_{1,1}(i) & \tilde{h}_{2,1}(i) & \dots & \tilde{h}_{N,1}(i) \\ \tilde{h}_{1,2}(i) & \tilde{h}_{2,2}(i) & & \vdots \\ \vdots & & \ddots & \\ \tilde{h}_{1,N}(i) & \dots & & \tilde{h}_{N,N}(i) \end{bmatrix} \tag{5.28}$$

Therefore, computing all the filters requires $\sim \mathcal{O}\left(3 N_{\mathrm{FFT}} N^3\right)$. Fig. 5.19 shows the complexity of the filter coefficient computation in an OFDM system with $N_{\mathrm{FFT}} = 1024$. We have introduced the complexity of performing the FFT (to convert the CSI in the frequency domain) which explains the dependence to the CIR length. In light grey, the complexity of the MMSE equalizer coefficient computation in the time domain is represented. We can observe that OFDM technique results in a lowest computation complexity. For instance, with $N = 3$ and 200 symb/tap, OFDM requires 1000 times less multiplications.

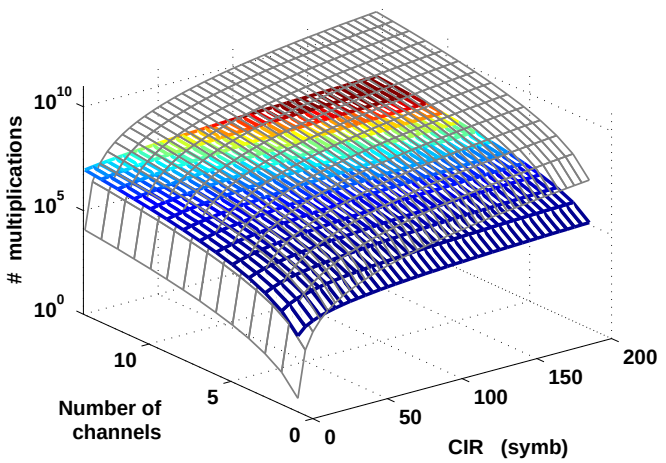


Figure 5.19: MMSE filter coefficient computation complexity

Finally the estimated symbols are obtained by a simple matrix multiplication, giving N symbols by N^2 multiplications thus each one at the cost of only N multiplications. Note that in the time domain NL operations were required.

5.3.2 OFDM with large dispersion

An OFDM symbol length corresponds to the size of the FFT plus the cyclic prefix. The size of the cyclic prefix has to be at least as long as the dispersion delay in order to keep

the orthogonality between the subcarriers. Therefore if the dispersion is large, the CP introduces a significant redundancy.
For example, let us consider a FFT over 512 symbols. We know that mode dispersion can be more than 100 symbols which means that the cyclic prefix would introduce at least 20% of redundancy. This may not be suitable for all kinds of applications.

On the other hand, the cyclic prefix can be set to a given size. If the dispersion delay is larger, it results to inter-carrier and inter-symbol interferences. A solution to this problem is called channel shortening, also known as time-domain equalizer (TEQ). It consists in a receiver front-end FIR filter which reduces the channel impulse response experienced by the receiver [116][117]. Thus, the received signal is first equalized in the time domain in order to partially reduce the CIR to a length smaller than the cyclic prefix. Then, the CP is removed and FFT is applied to convert it in the frequency domain (see Fig. 5.20).

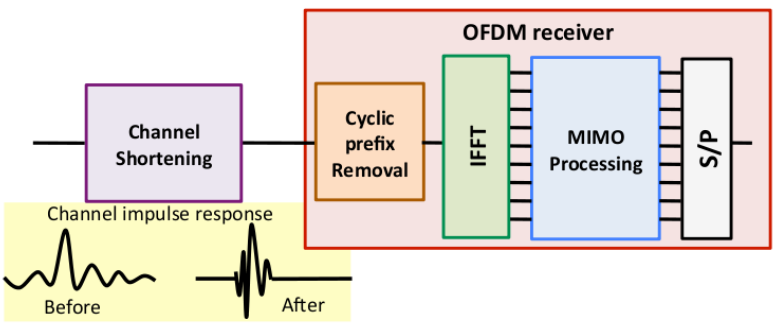


Figure 5.20: Channel shortening before the OFDM receiver

The introduction of a multi-tap filter increases the complexity of the equalization but as the CP still removes a part of the dispersion, the equalizers only deals with a fraction of the total dispersion and thus the filter size is smaller than in single carrier transmissions. Therefore, optical OFDM, even for large delay transmissions has a lower receiver complexity compared to the single carrier case.

5.3.2.1 Maximum shortening signal to noise ratio (MMSNR)

MMSNR channel shortening was originally proposed by Melsa *et. al* in [117]. The original idea was to minimize the energy of the CIR outside a window of length $\nu + 1$, with ν corresponding to the size of the cyclic prefix. However this may lead to a complicated solving problem thus alternatively, it has been proposed to maximize the energy inside the window [118].

We consider the shortener filter $\mathbf{W} = [\mathbf{W}_1 \dots \mathbf{W}_{N_W}]$. We want the equivalent channel $\mathbf{H}^{\text{eq}}$

to have an impulse response called target impulse response (TIR), of length $\nu+1$ thus:

$$\begin{aligned}\mathbf{H}^{\text{eq}} &= \mathbf{W}\ \mathbf{H} \\ &= [\mathbf{W}_1 \dots \mathbf{W}_{N_W}] \begin{bmatrix} \mathbf{H}_1 & \mathbf{H}_2 & \dots & \mathbf{H}_L & 0 & 0 & \dots \\ 0 & \mathbf{H}_1 & \mathbf{H}_2 & \dots & \mathbf{H}_L & 0 & 0 & \dots \\ & & & & \ddots & & & \\ & & 0 & 0 & \mathbf{H}_1 & \mathbf{H}_2 & \dots & \mathbf{H}_L \end{bmatrix} \\ &= [\underbrace{\mathbf{0} \dots \mathbf{0}}_{\Delta}\ \ \mathbf{H^{\text{eq}}}_1 \dots \mathbf{H^{\text{eq}}}_{\nu+1}\ \ \mathbf{0} \dots]\end{aligned} \tag{5.29}$$

The Δ front zeros $N \times N$ matrices correspond to the delay of the shortener equalizer. We can define $\mathbf{H}_{win}$ as the block columns of $\mathbf{H}$ corresponding to the window and $\mathbf{H}_{wall}$ the other ones. This can be written as:

$$\mathbf{H}_{win} = \mathbf{H}\ \mathbf{G}_{\Delta} \tag{5.30}$$

$$\mathbf{H}_{wall} = \mathbf{H}\ \overline{\mathbf{G}}_{\Delta} \tag{5.31}$$

with $\mathbf{G} = \text{diag}([\mathbf{0}_{\Delta N}\ \ \mathbf{I}_{(\nu+1)N}\ \ \mathbf{0}_{(N_W+L-\Delta-\nu)N}])$ and $\overline{\mathbf{G}}_{\Delta} = \mathbf{I} - \mathbf{G}_{\Delta}$.

The expression of the energy inside and outside the window can be written as:

$$\varepsilon_{win} = \text{trace}\{\mathbf{W}\ \underbrace{\mathbf{H}_{win}\mathbf{H}_{win}^H}_{\mathbf{B}}\ \mathbf{W}^H\} \tag{5.32}$$

$$\varepsilon_{wall} = \text{trace}\{\mathbf{W}\ \underbrace{\mathbf{H}_{wall}\mathbf{H}_{wall}^H}_{\mathbf{A}}\ \mathbf{W}^H\} \tag{5.33}$$

The problem of maximizing the energy inside the window and keep the energy outside contained can be formulated by:

$$\mathbf{W}^{opt} = \arg\max_{\mathbf{W}}\ \text{trace}\left\{\mathbf{W}\ \mathbf{B}\ \mathbf{W}^H\right\} \quad \text{subject to} \quad \mathbf{W}\ \mathbf{A}\ \mathbf{W}^H = \mathbf{I} \tag{5.34}$$

MMSNR method does not take into account the noise but it can be extended to the noisy case [119] by redefining the problem with $\mathbf{A} = \mathbf{H}_{wall}\mathbf{H}_{wall}^H + \sigma_n^2\mathbf{I}$.
As presented in [118], the solution to the problem is given by:

$$\mathbf{W}^{opt} = \mathbf{L}^{-1}\ \mathbf{U} \tag{5.35}$$

where $\mathbf{L}$ is obtained by the Cholesky decomposition of the matrix $\mathbf{A}$, $\mathbf{A} = \mathbf{L}^H\ \mathbf{L}$ and $\mathbf{U}$ are the N eignvectors corresponding to the N largest eignvalues of the matrix $(\mathbf{L}^H)^{-1}\ \mathbf{H}\mathbf{H}^H\ \mathbf{L}^{-1}$.

Fig. 5.21 presents an example of channel shortening. The CIR was originally 11taps long and after the filter it has been reduced to only 5 taps. Therefore a CP corresponding to $5T_s$ would be enough to ensure no interference in the frequency domain.

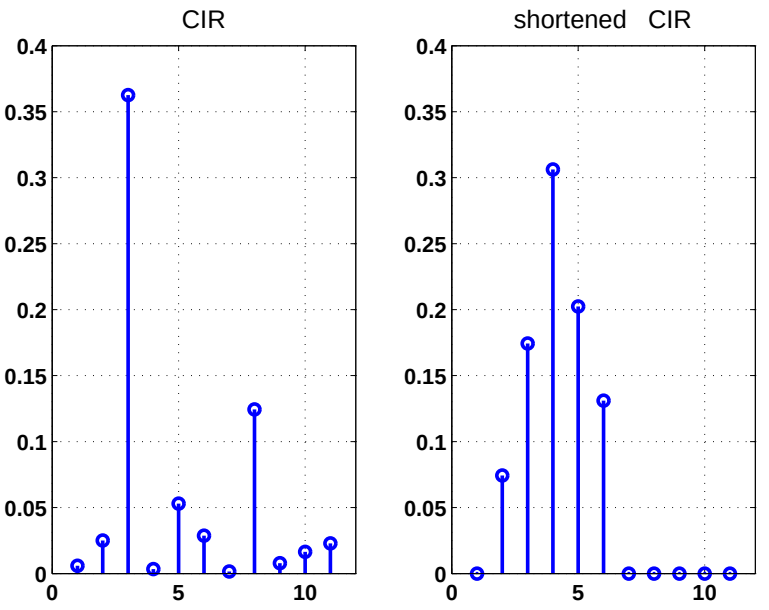


Figure 5.21: CIR and TIR

5.3.2.2 Minimum mean-square error (MMSE)

MMSE channel shortening is a second method to reduce the CIR. We keep calling $\mathbf{H}_{win}$ the target impulse channel and $\mathbf{W}$ the shortening filter. The MMSE shortening equalizer intends to minimize the mean square error between the output of $\mathbf{W}$ and the output of the virtual short channel (with a delay Δ). This scheme is represented in Fig. 5.22 where $\mathbf{X}_\Delta = [\mathbf{x}(k-\Delta)^T \dots \mathbf{x}(k-\Delta-N_{win}+1)]^T$, $\mathbf{X} = [\mathbf{x}(k)^T \dots \mathbf{x}(k-L-N_W+2)]^T$ and the error can be expressed as:

$$\varepsilon = \mathbf{H}_{win}\mathbf{X}_\Delta - \mathbf{W}\mathbf{Y} \tag{5.36}$$

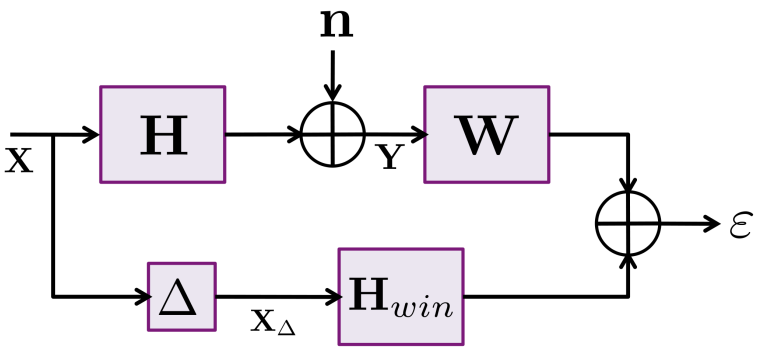


Figure 5.22: Scheme of the MMSE shortener equalizer

We note $\beta = \mathbf{H}_{win}\mathbf{X}_\Delta$. The auto covariance matrix and the covariance matrix between β

and $\mathbf{Y}$ can be written as:

$$R_\beta = \mathbf{H}_{win} \underbrace{\sigma^2_{\mathbf{x}} \mathbf{I}_{N_{win}}}_{R_{\mathbf{x}}} \mathbf{H}^H_{win} \tag{5.37}$$

$$\begin{aligned} R_{\beta\mathbf{Y}} &= \mathbb{E}\beta\mathbf{Y}^H \\ &= \mathbb{E}\ \mathbf{H}_{win}\mathbf{X}_\Delta\mathbf{X}^H\mathbf{H}^H \\ &= \mathbf{H}_{win}\underbrace{\sigma^2_{\mathbf{x}}\mathbf{\Phi}_\Delta\mathbf{H}^H}_{R_{\mathbf{xy}}} \end{aligned} \tag{5.38}$$

where $\mathbf{\Phi}_\Delta = [\mathbf{0}\ldots\mathbf{0}\ \mathbf{I}\ \mathbf{0}\ldots]$ has Δ leading zero matrices.

The problem now corresponds the the minimization of $\mathbb{E}|\varepsilon|^2$ with $\mathbf{W}$ being an estimator of $\mathbf{H}_{win}\mathbf{X}$. The optimal value $\mathbf{W}^{opt}$ and the corresponding error auto covariance are given by Eq. D.4 and Eq. D.5:

$$\begin{aligned} \mathbf{W}^{opt} &= R_{\beta\mathbf{Y}}R^{-1}_{\mathbf{Y}} \\ &= \mathbf{H}_{win}\sigma^2_{\mathbf{x}}\mathbf{\Phi}_\Delta\mathbf{H}^H\left(\sigma^2_{\mathbf{x}}\mathbf{H}\mathbf{H}^H + \sigma^2_{\mathbf{n}}\mathbf{I}\right)^{-1} \end{aligned} \tag{5.39}$$

$$\begin{aligned} R_e &= R_\beta - R_{\beta\mathbf{Y}}R^{-1}_{\mathbf{Y}}R_{\mathbf{Y}\beta} \\ &= \mathbf{H}_{win}\underbrace{\left(R_{\mathbf{x}} - R_{\mathbf{xy}}R^{-1}_{\mathbf{Y}}R_{\mathbf{xy}}\right)}_{\mathbf{R}}\mathbf{H}^H_{win} \end{aligned} \tag{5.40}$$

We are in front of an optimization problem with two parameters $\mathbf{H}_{win}$ and $\mathbf{W}$. $\mathbf{W}^{opt}$ has been expressed in function of $\mathbf{H}_{win}$ by following the MMSE criterion. Therefore we now have to find the optimal expression of the shorten channel $\mathbf{H}_{win}$ in order to minimize the error auto covariance trace. We can add some constraints to the TIR. As proposed in [116], we impose to the TIR to have an unit norm ($\mathbf{H}_{win}\mathbf{H}^H_{win} = \mathbf{I}$) and thus the problem becomes:

$$\mathbf{H}^{opt}_{win} = \arg\min_{\mathbf{R}}\ \text{trace}\left\{\mathbf{W}\ \mathbf{R}\ \mathbf{W}^H\right\} \quad \text{subject to} \quad \mathbf{H}_{win}\ \mathbf{H}^H_{win} = \mathbf{I} \tag{5.41}$$

The optimal solution is obtained with $\mathbf{H}^{opt}_{win}$ being the eigenvectors corresponding to the N minimal eigenvalues of $\mathbf{R}$ [120].

Conclusion

In this chapter, we have investigated the issue of digital equalization in multi-mode fiber transmissions. Mode group diversity multiplexing in MMF presents a great potential to increase the capacity of future generation of optical systems. Indeed, several streams of data can be simultaneously transmitted on the different modes supported by the fiber.

However, the main challenge is to deal the severe inter-symbol interferences caused by the coupling between the modes and the large differential propagation delay.

Digital signal processing is already efficiently employed in polarization multiplexed system to mitigate polarization impairments. Unlike PDM systems, we propose to realized the equalization based on the channel estimation. Training sequence are transmitted before the data in order to estimate the channel and the equalizer coefficients are computed afterward. We have shown that this technique has very good performance compared to CMA and does not have convergence problem. Moreover as the optical channel varies very slowly, it doesn't introduce too much overhead.

When the DMD is large, the complexity of equalization in the time domain can be very high and consequently, OFDM becomes an interesting alternative. We have shown that OFDM results in a lower equalization complexity but introduces an extra overhead. Therefore, we have presented two methods of pre-filtering in order to reduce the channel impulse response and thus, allows smaller cyclic prefix length.

Conclusions & Perspectives

In this work, we have proposed new digital signal processing techniques in order to improve the quality of actual and future generations of optical transmission systems. By "quality", we mean several criteria such as the performances (SNR improvement) of the transmission, the complexity of the receiver or the transmission bit-rate.

With the come-back of coherent detection and the development of high-speed electronics, soft-decision FEC have appeared to be a promising solution to improve the performance of optical systems. More than 2 dB coding gain could be obtained compared to the concatenated FEC schemes currently used. However, their implementation at very high bit-rates is a very challenging issue.

LDPC codes and product codes are the two candidate families of soft decoding FEC. We have investigated their decoding algorithms and show that LDPC codes is the most appropriate family in terms of both performance and decoding complexity. The FEC are expected to have high rates (≥ 0.85), good performance ($\sim$ 10dB coding gain, low error floor) and to ease the practical implementation (short codeword length, structured parity-check matrix). To satisfy these criteria, most of the LDPC proposed in the literature are based on combinatorial constructions which result in quasi-cyclic codes. However, due to construction constraints, the performance are not optimal and the design parameters (rate, codeword length, sparseness of the parity-check matrix) can not be chosen freely.

We have proposed an original construction of QC-PEG LDPC codes based on the progressive edge growth algorithm. The construction intends to minimize the number of short cycles in the parity-check matrix thus improve the performance and reduce the error floor level. In order to obtain a quasi-cyclic structure, we have introduced and based our algorithm on the labeled Tanner graph. Moreover, the design parameters can be chosen with no constraint. QC-PEG LDPC performs better than algebraic constructions and results in codes having higher girth and less short cycles.

Powerful FEC are necessary to deal with the SNR degradation caused by the transmission impairments, especially at high bit-rates. Another source of degradation comes from the differential encoding of the information. However, this type of encoding is necessary due to the lack of phase reference in direct detection systems and to the phase ambiguity caused

by the phase recovery in coherent detection systems. In order to mitigate this impairment, we have proposed an original coding/decoding scheme named structure symbol interleaving. By interleaving codewords by blocs of bits, we have demonstrated a performance improvement. This codeword structure also leads to a very reduced decoding complexity. The complexity reduction can reach 50%. Moreover we have shown that by using two different FEC, the total complexity can be reduced with no performance degradation.
Our coding/decoding scheme can be very profitable for the use of LDPC in optics. Indeed LDPC are characterized by a high decoding complexity and large redundancies and both could be reduced by the SSI scheme.

Then, we have investigated the use of space-time coding in optical transmissions. They are used in polarization multiplexed systems by modulating jointly the two polarizations. If the advantages of these techniques are now obvious in wireless transmissions, it was not the case for optical transmissions. We have demonstrated for the first time that ST coding can very efficiently mitigate polarization dependent loss. However, they can not be employed in single carrier transmissions due to an excessive decoding complexity and OFDM is required. We have analyzed the performance of several ST codes and show that they differ from the wireless case. In particular, we have observed that the determinant construction criterion is not relevant and outline a new criterion based on the minimal distance. The numerical simulations confirm this analysis and show that Silver code performs better than Golden code against PDL.
ST coding is the only solution that can really deal with PDL. We have observed that Silver code can compensate almost 90% of the PDL impairments. Therefore the implementation of ST code in future 100-400Gb/s OFDM systems can be very profitable.

Finally, we have studied MIMO equalization in multi-mode fiber transmission. MMF transmissions are very promising because mode multiplexing can be realized in order to increase the channel capacity compared to classical transmissions on SMF. However, due to the large number of modes and the severe ISI, the equalization is very challenging. We have proposed to estimate the channel by a training sequence before to compute the MMSE equalizer. We observed by numerical simulations that the ISI can be very efficiently mitigated when the channel has been well estimated.
When the CIR is very large, we have proposed to realize the equalization in the frequency domain, by using OFDM, in order to decrease the complexity. To avoid excessively long cyclic prefix, we have also presented channel shortening techniques.

Perspectives

The principal perspectives of this work are:

- The experimental validation of the concepts presented in the thesis.

- We are interested in evaluating the performance of the SSI scheme in a real long-haul transmissions systems and we want to observe the influence of the linear and nonlinear impairments on the results. We would like to propose a specific coding scheme based on the QC-PEG LDPC presented in the chapter 2 and taking advantages of the complexity and the redundancy reduction.

- An analysis of the QC-PEG LDPC error floor level can be realized in order to compare with the other LDPC codes proposed in the literature, in particular the CIDS-LDPC.

- We have started a collaboration with Karlsruhe Institute of Technology (KIT) in the frame of the project EUROFOS in order to realize a PDM-OFDM transmission using polarization-time coding. We want to validate the results obtained by numerical simulations and compare the performance of Silver code with other codes.

• The continuation of the work about space-time coding techniques in optical transmission systems

- In Chapter 4, we have observed that space-time codes are a very efficient solution to mitigate PDL in optical OFDM systems. We have shown that their performance is very different than in wireless systems and derived a construction criterion based on the minimal distance.

  In a future work, we want to design new polarization-time codes adapted to optical systems that are PDL insensitive and hence performing better that the Silver code.

- Space-time code can be employed not only in polarization multiplexed transmission but in all kinds of optical MIMO systems. A WDM transmission can be seen as a MIMO system where the transmitted wavelength are the outputs and the received wavelength are the outputs. Due to the transmission impairments, mainly nonlinear effects, the wavelength interact and can be received with unequal SNR. For instance, a QPSK channel can be impaired due to the co-propagation with a neighbor OOK channel. Therefore using space-time coding techniques may improve the performance of WDM systems.

Appendix A

OA-LDPC construction

Let us build an OA-LDPC based on the 1-OA$(4^2, 4+1, 4, 2)$. We note α the primitive element of $GF(4)$ and its primitive polynomial is X^2+X+1. The elements of $GF(4)$ are $[a_1=0\ a_2=1\ a_3=\alpha\ a_4=\alpha^2]$.

We determine all the polynomials of $GF(4)[X]$ with degree up to 1:

$$\begin{array}{llll} P_1=0 & P_2=1 & P_3=\alpha & P_4=\alpha^2 \\ P_5=X+0 & P_6=X+1 & P_7=X+\alpha & P_8=X+\alpha^2 \\ P_9=\alpha X+0 & P_{10}=\alpha X+1 & P_{11}=\alpha X+\alpha & P_{12}=\alpha X+\alpha^2 \\ P_{13}=\alpha^2 X+0 & P_{14}=\alpha^2 X+1 & P_{15}=\alpha^2 X+\alpha & P_{16}=\alpha^2 X+\alpha^2 \end{array}$$

We create the orthogonal array such that $\mathbf{A}(i,j) = P_j(a_i)$. A fifth line is added corresponding to the major coefficient of P_j

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & \alpha & \alpha^2 & 0 & 1 & \alpha & \alpha^2 & 0 & 1 & \alpha & \alpha^2 & 0 & 1 & \alpha & \alpha^2 \\ 0 & 1 & \alpha & \alpha^2 & 1 & 0 & \alpha^2 & \alpha & \alpha & \alpha^2 & 0 & 1 & \alpha^2 & \alpha & 1 & 0 \\ 0 & 1 & \alpha & \alpha^2 & \alpha & \alpha^2 & 0 & 1 & \alpha^2 & \alpha & 1 & 0 & 1 & 0 & \alpha^2 & \alpha \\ 0 & 1 & \alpha & \alpha^2 & \alpha^2 & \alpha & 1 & 0 & 1 & 0 & \alpha^2 & \alpha & \alpha & \alpha^2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & \alpha & \alpha & \alpha & \alpha & \alpha^2 & \alpha^2 & \alpha^2 & \alpha^2 \end{bmatrix}$$

Then we compute the blocks $B_{a_k,i}$:

$$\begin{aligned} B_{a_1,1} &= \{(0,0),(\alpha^2,1),(1,\alpha),(\alpha,\alpha^2)\} \\ B_{a_2,1} &= \{(1,0),(\alpha,1),(0,\alpha),(\alpha^2,\alpha^2)\} \\ B_{a_3,1} &= \{(\alpha,0),(1,1),(\alpha^2,\alpha),(1,\alpha^2)\} \\ B_{a_4,1} &= \{(\alpha^2,0),(0,1),(\alpha,\alpha),(0,\alpha^2)\} \\ &\vdots \end{aligned}$$

These block are converted into positions by transforming the pairs as: $(a_u, a_v) \rightarrow l = v + (u-1)q$

$$\begin{aligned} \tilde{B}_{a_1,1} &= \{1, 8, 10, 15\} \\ \tilde{B}_{a_2,1} &= \{2, 7, 9, 16\} \\ \tilde{B}_{a_3,1} &= \{3, 6, 12, 14\} \\ \tilde{B}_{a_4,1} &= \{4, 5, 11, 13\} \\ &\vdots \end{aligned}$$

Finally each block determines the positions of the 1 in the lines of $\boldsymbol{H}$:

$$\boldsymbol{H} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ & & & & & & & & \vdots & & & & & & & \end{bmatrix}$$

Appendix B

PEG & QC-PEG algorithms

Algorithm 2 PEG algorithm

for $k = 1$ to n_c **do**
  for $m = 1$ to n_l **do**
    if m=0 **then**
      • c is randomly chosen in $\bar{\mathcal{S}}_{v_k}$ among the check nodes having the smallest degrees
      • $\bar{\mathcal{S}}_{v_k} = \bar{\mathcal{S}}_{v_k} \backslash c, \quad \mathcal{S}_{v_k} = \mathcal{S}_{v_k} \cup c$
    else
      • $\mathcal{A}_{v_k}^{(0)} = \bar{\mathcal{S}}_{v_k}, \qquad l = 1$
      while $\mathcal{A}_{v_k}^{(l)} \neq \emptyset$ and $\mathcal{A}_{v_k}^{(l)} \neq \mathcal{A}_{v_k}^{(l-1)}$ **do**
        • Expand the tree until a depth $2l-1$
        • $\mathcal{A}_{v_k}^{(l)} = \mathcal{A}_{v_k}^{(l-1)} \backslash D_{2l-1}$
        • $l = l + 1$
      end while
      • c is chosen among the check nodes of $\mathcal{A}_{v_k}^{(l-1)}$ having the smallest degrees
      • $\bar{\mathcal{S}}_{v_k} = \bar{\mathcal{S}}_{v_k} \backslash c, \quad \mathcal{S}_{v_k} = \mathcal{S}_{v_k} \cup c$
    end if
  end for
end for

Algorithm 3 QC-PEG algorithm

for $k = 1$ to n_{bc} **do**
  for $m = 1$ to n_{bl} **do**
    • $\mathcal{S}_{v_k} = \mathcal{S}_{v_k} \cup c_m, \quad \bar{\mathcal{S}}_{v_k} = \bar{\mathcal{S}}_{v_k} \backslash c_m$
    • Set $i_{c_m,v_k} = x$
    • $\mathcal{M}_{v_k}^{(0)} = [1, 2, \dots p]$, $l = 1$
    while $S_x^l \neq \emptyset$ **do**
      • Expand the tree-graph from v_m to depth of $2l$
      • Compute the cumulated metrics of each paths
      • Find the values $x_1, x_2 \dots$ that make the metrics depending of x equal to zero.
      • $\mathcal{M}_{v_k}^{(l)} = \mathcal{M}_{v_k}^{(l-1)} \backslash \{x_1, x_2 \dots\}$
      • $l = l + 1$
    end while
    • Select i_{c_m,v_k} in $\mathcal{M}_{v_k}^{(l-1)}$
  end for
end for

Appendix C

Sphere decoder

We consider the lattice $\Lambda = \{\mathbf{M}_{\mathbb{R}}\mathbf{S}_{\mathbb{Z}}\}$ and we note $\mathbf{U}_{\Lambda}$ the lattice points. $\mathbf{S}_{\mathbb{Z}} = [s_1 \dots s_n]$ are QAM symbols. For simplicity, we drop the index $_{\mathbb{R},\mathbb{Z},\Lambda}$.
We look for the lattice points $\mathbf{U}$ inside the n-sphere of radius $\sqrt{\rho_{sph}}$ centered on the received symbol $\mathbf{Y}$. Hence the coordinates u_i of $\mathbf{U}$ are bounded by the dimension of the sphere and so are the coordinates s_i. As $\mathbf{U}$ is inside the sphere, the distance $d(\mathbf{Y},\mathbf{U})$ between the lattice point and $\mathbf{Y}$ is inferior to $\sqrt{\rho_{sph}}$.

$$d(\mathbf{Y},\mathbf{U})^2 \leq \rho_{sph} \tag{C.1}$$

The distance can be expressed as:

$$\begin{aligned} d(\mathbf{Y},\mathbf{U})^2 &= \|\mathbf{Y}-\mathbf{U}\|^2 \\ &= \|\mathbf{Y}^T-\mathbf{U}^T\|^2 \\ &= \|\mathbf{z}^T\mathbf{M}^T-\mathbf{S}^T\mathbf{M}^T\|^2 \\ &= \|(\mathbf{z}^T-\mathbf{S}^T)\mathbf{M}^T\|^2 \\ &= (\mathbf{z}^T-\mathbf{S}^T)\mathbf{M}^T\mathbf{M}(\mathbf{z}-\mathbf{S}) \end{aligned} \tag{C.2}$$

where $\mathbf{z} \triangleq \mathbf{M}^{-1}\mathbf{Y}$ is called the ZF point. $\mathbf{M}$ is a square integer matrix so it can be rewritten using the QR decomposition into $\mathbf{M} = \mathbf{QR}$ where $\mathbf{Q}$ is an orthogonal matrix and $\mathbf{R}$ an upper triangular matrix.

$$\begin{aligned} d(\mathbf{Y},\mathbf{U})^2 &= (\mathbf{z}^T-\mathbf{S}^T)\mathbf{R}^T\mathbf{R}(\mathbf{z}-\mathbf{S}) \\ &= \|\mathbf{R}\ (\mathbf{z}-\mathbf{S})\|^2 \\ &= \sum_{i=1}^{n}\left(R_{ii}(z_i-s_i)+\sum_{j=i+1}^{n}R_{ij}(z_j-s_j)\right)^2 \end{aligned}$$

if we note $q_{ii} = R_{ii},\ i = 1 \dots n$ and $q_{ij} = R_{ij}/R_{ii}$, Eq. C.1 can be expressed as:

$$\sum_{i=1}^{n} q_{ii}\left((\rho_i-s_i)+\sum_{j=i+1}^{n}q_{ij}(\rho_j-s_j)\right)^2 \leq \rho_{sph} \tag{C.3}$$

We can observe that the bounds on a coordinate s_i depend on the values of $s_{j>i}$. From Eq. C.3; we can be bound each coordinate s_i, $i = 1 \dots n$:

$$\left\lceil -\sqrt{\frac{\Theta_i}{q_{ii}}} + \Omega_i \right\rceil \leq s_i \leq \left\lfloor \sqrt{\frac{\Theta_i}{q_{ii}}} + \Omega_i \right\rfloor \tag{C.4}$$

where Θ_i and Ω_i correspond to:

$$\Omega_i = z_i + \sum_{j=i+1}^{n} q_{ij}(z_j - s_j) \tag{C.5}$$

$$\Theta_n = \rho_{sph} \tag{C.6}$$

$$\Theta_{i-1} = \Theta_i - q_{ij}\left(\Omega_i - s_i\right)^2 \tag{C.7}$$

we note $\mathrm{b}_i = [b_i^{inf}\ b_i^{sup}]$ the upper and lower bound of s_i.

The sphere decoder searches the vectors $\mathbf{S}$ satisfying these bounds for each of its coordinate. It starts by the n^{th} coordinate whose value is chosen inside the range determined by the bounds $b_n^{inf} \leq s_n \leq b_n^{sup}$. Then the bounds and the value of the $n-1^{th}$ coordinate are computed. We chose in the same way all s_i until s_1 and so having the n coordinates of a lattice point inside the sphere. The distance of this point from the center of the sphere is equal to:

$$d(\mathbf{Y}, \mathbf{U})^2 = \rho_{sph} - T_1 + q_{11}\left(\Omega_1 - s_1\right)^2 \tag{C.8}$$

When a lattice point inside the sphere is found, the sphere radius is adjusted $\rho_{sph} = d(\mathbf{Y}, \mathbf{U})^2$ and the procedure is repeated in order to find a new lattice point inside this smaller sphere.

All the points of the lattice are not inside the sphere so it means that at a certain step i the bounds can give no choice ($b_i^{inf} > b_i^{sup}$). Therefore we have to go up to the step $i-1$ and change the values of s_{i-1} if it is possible, in order to have the new bounds at the i^{th} level such that $b_i^{inf} \leq b_i^{sup}$. Note that the lattice point search can be seen as a search in a tree graph of depth n.

The decoding complexity of the Sphere decoder has been evaluated to [105]:

$$\mathcal{O}\left(n^2 \times \left(1 + \frac{n-1}{4\lambda_{min}\rho_{sph}}\right)^{4\lambda_{min}\rho_{sph}}\right) \tag{C.9}$$

where λ_{min} is the minimum singular value of the matrix $\mathbf{M}\mathbf{M}^T$.

Appendix D

Linear estimators

We consider two random complex vectors $\mathbf{x} = [x(1)\ x(2)\ldots]^T$ and $\mathbf{y} = [y(1)\ y(2)\ldots]^T$ with zero mean values ($\mathbb{E}\mathbf{x} = \mathbb{E}\mathbf{y} = 0$) and we note:

$$\begin{aligned} R_x &= \mathbb{E}\mathbf{x}\mathbf{x}^H \\ R_y &= \mathbb{E}\mathbf{y}\mathbf{y}^H \\ R_{xy} &= R_{yx}^H = \mathbb{E}\mathbf{x}\mathbf{y}^H \end{aligned}$$

The general problem is to find an unbiased linear estimator K giving an approximation of $\mathbf{x}$:

$$\widehat{\mathbf{x}} = K\ \mathbf{y} \tag{D.1}$$

D.1 Mean-square-error criterion

The MSE estimator K_o minimizes the mean quadratic error between $\mathbf{x}$ and its estimate $\widehat{\mathbf{x}}$ thus:

$$K_o = \arg\min_K \ \mathbb{E}\ \|\widehat{\mathbf{x}} - \mathbf{x}\|^2 \tag{D.2}$$

Let us define the mean square error auto-covariance:

$$\begin{aligned} J(K) &= \mathbb{E}\ \|\mathbf{x} - \widehat{\mathbf{x}}\|^2 \\ &= \mathbb{E}\ (\mathbf{x} - \widehat{\mathbf{x}})(\mathbf{x} - \widehat{\mathbf{x}})^H \\ &= \mathbb{E}\mathbf{x}\mathbf{x}^H - \mathbb{E}\mathbf{x}\mathbf{y}^H K^H - K\mathbb{E}\mathbf{y}\mathbf{x}^H + \mathbb{E}K\mathbf{y}^H\mathbf{y}K^H \\ &= R_x - R_{xy}K^H - KR_{yx} + KR_yK^H \end{aligned} \tag{D.3}$$

The minimum of $J(K)$ can be found by differentiation with respect to K:

$$\nabla_K J(K) = -R_{yx} + R_yK^H$$

Finally, the estimator minimizing the mean square error is given by:

$$\begin{aligned} K_o^H &= R_y^{-1}\, R_{yx} \\ K_o &= R_{xy}\, R_y^{-1} \end{aligned} \tag{D.4}$$

The minimal value of the error auto-covariance is expressed by:

$$\begin{aligned} J(K_o) &= R_x - R_{xy}R_y^{-1}R_{yx} - R_{xy}R_y^{-1}R_{yx} + R_{xy}R_y^{-1}R_{yx} \\ J(K_o) &= R_x - R_{xy}R_y^{-1}R_{yx} \end{aligned} \tag{D.5}$$

D.2 Least-square-error criterion

Let us now redefine the estimator as $\widehat{\mathbf{x}} = \mathbf{y}\ K$. The LSE estimator K_o minimizes the quadratic error between $\mathbf{x}$ and its estimate $\widehat{\mathbf{x}}$ thus:

$$K_o = \min_K \quad \|\mathbf{x} - \widehat{\mathbf{x}}\|^2 \tag{D.6}$$

Let $J(K)$ denote the cost function:

$$\begin{aligned} J(K) &= \|\mathbf{x} - \widehat{\mathbf{x}}\|^2 \\ &= (\mathbf{x} - \mathbf{y}K)^H(\mathbf{x} - \mathbf{y}K) \\ &= \mathbf{x}^H\mathbf{x} - k^H\mathbf{y}^H\mathbf{x} - \mathbf{x}\mathbf{y}K + K^H\mathbf{y}^H\mathbf{y}K \end{aligned}$$

By differentiation with respect to K^H we obtain:

$$\nabla_{K^H} J(K) = -\mathbf{y}^H\mathbf{x} + \mathbf{y}^H\mathbf{y}K$$

which is null, thus minimizing the cost function for:

$$K_o = (\mathbf{y}^H\mathbf{y})^{-1}\mathbf{y}^H\mathbf{x} \tag{D.7}$$

Bibliography

[1] www.cisco.com. Cited page 1

[2] www.telegeography.com. Cited page 1

[3] L. Zehnder. "Ein neuer Interferenzrefraktor". *Zeitschr. für Intrumentenkunde*, 11:275–285, 1891. Cited page 13

[4] John G. Proakis. *Digital communications*. McGraw-Hill, New York :, 1983. Cited pages 16 and 111

[5] G. P. Agrawal. *Fiber-optic communication systems*. John Wiley & Sons, Inc, 1992. Cited pages 17 and 18

[6] S. J. Savory. "Chromatic dispersion compensation in digital coherent systems". In *Optical Communications, 2008. ECOC 2008. European Conference on*, pages 1 – 2, 09 2008. Cited page 19

[7] M. Karlsson. Probability density functions of the differential group delay in optical fiber communication systems. *Lightwave Technology, Journal of*, 19(3):324 –331, mar. 2001. Cited page 19

[8] A. O. Lima, Jr. I. T. Lima, C. R. Menyuk, and T. Adali. "Comparison of penalties resulting from first-order and all-order polarization mode dispersion distortions in optical fiber transmission systems". *Opt. Lett.*, 28(5):310–312, 2003. Cited page 19

[9] Francisco A. Garcia, Darli A. Mello, and Helio Waldman. "Feedforward carrier recovery for polarization demultiplexed signals with unequal signal to noise ratios". *Opt. Express*, 17(10):7958–7969, 2009. Cited page 20

[10] Mark Shtaif. Performance degradation in coherent polarization multiplexed systems as aresult of polarization dependent loss. *Opt. Express*, 16(18):13918–13932, Sep 2008. Cited pages 20 and 95

[11] R. J. Mears, L. Reekie, I. M. Jauncey, and D. N. Payne. Low-noise erbium-doped fibre amplifier operating at 1.54μm. *Electronics Letters*, 23(19):1026 –1028, sep. 1987. Cited page 20

[12] E. Desurvire, J. R. Simpson, and P. C. Becker. High-gain erbium-doped traveling-wave fiber amplifier. *Opt. Lett.*, 12(11):888–890, 1987. Cited page 20

[13] J. P. Gordon, Walker L. R., and LouiseII W. H. Quantum statistics of masers and attenuators. *Phys Rev.*, 130:806–812, 1963. Cited page 21

[14] R.-J. Essiambre, G. Kramer, P.J. Winzer, G.J. Foschini, and B. Goebel. "Capacity Limits of Optical Fiber Networks". *Lightwave Technology, Journal of*, 28(4):662 –701, feb.15, 2010. Cited pages 21, 22, and 84

[15] Ivan P. Kaminow, Tingye Li, and Alan E. Willner. *Optical Fiber Telecommunications V B (Fifth Edition)*. Elsevier, 2008. Cited pages 22 and 25

[16] J. Bromage. Raman amplification for fiber communications systems. *Lightwave Technology, Journal of*, 22(1):79 – 93, jan. 2004. Cited page 23

[17] Dirk van den Borne. *Robust Optical Transmission systems*. Phd dissertation, Universiteit Eindhoven, 2008. Cited page 24

[18] H. Bulow, F. Buchali, and A. Klekamp. Electronic dispersion compensation. *Lightwave Technology, Journal of*, 26(1):158 –167, jan 2008. Cited page 25

[19] R.I. Killey, P.M. Watts, V. Mikhailov, M. Glick, and P. Bayvel. "Electronic dispersion compensation by signal predistortion using digital Processing and a dual-drive Mach-Zehnder Modulator". *Photonics Technology Letters, IEEE*, 17(3):714 –716, march 2005. Cited page 25

[20] P.M. Watts, V. Mikhailov, S. Savory, P. Bayvel, M. Glick, M. Lobel, B. Christensen, P. Kirkpatrick, Song Shang, and R.I. Killey. "Performance of single-mode fiber links using electronic feed-forward and decision feedback equalizers". *Photonics Technology Letters, IEEE*, 17(10):2206 – 2208, oct. 2005. Cited page 25

[21] Jr. Forney, G. "Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference". *Information Theory, IEEE Transactions on*, 18(3):363 – 378, may 1972. Cited page 26

[22] Joan M. Gene, Peter J. Winzer, S. Chandrasekhar, and Herwig Kogelnik. "Joint PMD and Chromatic Dispersion Compensation Using an MLSE". In *Optical Communications, 2006. ECOC 2006. European Conference on*, pages 1 –2, 24-28 2006. Cited page 26

[23] Chunmin Xia and W. Rosenkranz. "Electrical dispersion compensation for different modulation formats with optical filtering". In *Optical Fiber Communication Conference, 2006 and the 2006 National Fiber Optic Engineers Conference. OFC 2006*, page 3 pp., 5-10 2006. Cited page 26

[24] Seb J. Savory. "Digital filters for coherent optical receivers". *Opt. Express*, 16(2):804–817, 2008. Cited page 28

[25] V.A.J.M. Sleiffer, D. van den Borne, M.S. Alfiad, S.L. Jansen, and H. de Waardt. Dispersion management in long-haul 111-gb/s polmux-rz-dqpsk transmission systems. In *LEOS Annual Meeting Conference Proceedings, 2009. LEOS '09. IEEE*, pages 569 –570, oct. 2009. Cited page 28

[26] Tianhua Xu, Gunnar Jacobsen, Sergei Popov, Jie Li, Evgeny Vanin, Ke Wang, Ari T. Friberg, and Yimo Zhang. "Chromatic dispersion compensation in coherent transmission system using digital filters". *Opt. Express*, 18(15):16243–16257, 2010. Cited page 28

[27] U. Mengali and A. N. D'andrea. *synchronization techniques for digital receivers*. Plenum Press, New York, 1997. Cited pages 29 and 119

[28] D.-S. Ly-Gagnon, S. Tsukamoto, K. Katoh, and K. Kikuchi. "Coherent detection of optical quadrature phase-shift keying signals with carrier phase estimation". *Lightwave Technology, Journal of*, 24(1):12 – 21, jan. 2006. Cited pages 29 and 61

[29] A. Hocquenghem. "Codes correcteurs d'erreurs". *Chiffres*, 2:147–156, 1959. Cited page 36

[30] R.C Bose and C.R. Ray-Chaudhuri. "On a class of error-correcting binary group codes". *IRE Trans. on Information Theory.*, 3:68–79, 1960. Cited page 36

[31] W.W. Peterson. "encoding and error-correction procedures for the Bose-Chaudhuri codes". *IRE Trans. on Information Theory.*, 6:459–470, September 1960. Cited page 36

[32] Massey J. L. "Shift-register synthesis and BCH decoding". *Information Theory, IEE Transactions on*, 10:122 –127, January 1969. Cited page 36

[33] Chien R. T. "Cyclic decoding procedures for Bose-Chaudhuri-Hocquenghem codes". *Information Theory, IEE Transactions on*, 10:357 –363, Oct 1964. Cited page 36

[34] Jr. Forney, G. On decoding bch codes. *Information Theory, IEEE Transactions on*, 11(4):549 – 557, October 1965. Cited page 37

[35] R. Gallager. "Low-density parity-check codes". *Information Theory, IEE Transactions on*, 8(1):21 –28, january 1962. Cited page 37

[36] D.J.C. MacKay and R.M. Neal. "Near Shannon limit performance of low density parity check codes". *Electronics Letters*, 33(6):457 –458, mar 1997. Cited page 37

[37] T. and R. Urbanke. *Modern Coding Theory*. Cambridge university Press, 2008. Cited page 39

[38] Xiaofei Huang. "Single-Scan Min-Sum Algorithms for Fast Decoding of LDPC Codes". In *Information Theory Workshop, 2006. ITW '06 Chengdu. IEEE*, pages 140 –143, oct. 2006. Cited page 39

[39] G. D. Forney. *Concatenated codes*. Cambridge, MA: MIT Press, 1966. Cited page 39

[40] R.M. Pyndiah. "Near-optimum decoding of product codes: block turbo codes". *Communications, IEEE Transactions on*, 46(8):1003 –1010, aug 1998. Cited page 40

[41] S.A. Hirst, B. Honary, and G. Markarian. "Fast Chase algorithm with an application in turbo decoding". *Communications, IEEE Transactions on*, 49(10), oct 2001. Cited page 41

[42] W.D. Grover. "Forward error correction in dispersion-limited lightwave systems". *Lightwave Technology, Journal of*, 6(5):643 –654, may 1988. Cited page 42

[43] P. Moro and D. Candiani. "565 Mb/s optical transmission system for repeaterless sections up to 200 km". In *Communications, 1991. ICC '91, Conference Record. IEEE International Conference on*, pages 1217 –1221 vol.3, jun 1991. Cited page 42

[44] P.M. Gabla, J.L. Pamart, R. Uhel, E. Leclerc, J.O. Frorud, F.X. Ollivier, and S. Borderieux. "401 km, 622 Mb/s and 357 km, 2.488 Gb/s IM/DD repeater less transmission experiments using erbium-doped fiber amplifiers and error correcting code". *Photonics Technology Letters, IEEE*, 4(10):1148 –1151, oct 1992. Cited page 42

[45] J.L. Pamart, E. Lefranc, S. Morin, G. Balland, Y.C. Chen, T.M. Kissell, and J.L. Miller. "Forward error correction in a 5 Gbit/s 6400 km EDFA based system". *Electronics Letters*, 30(4):342 –343, feb 1994. Cited page 42

[46] Telecommunication Standardization Sector. "forward error correction for submarine systems". *International Telecommunication Union*, 2000. Cited page 42

[47] J.F. Marcerou, F. Pitel, G. Vareille, O. Ait Sab, D. Lesterlin, and J. Chesnoy. "From 40 to 80 10 Gbit/s DWDM transmission for ultra long-haul terrestrial transmission above 3000 km". In *Optical Fiber Communication Conference and Exhibit, 2001. OFC 2001*, volume 1, pages ME3–1 – ME3–3 vol.1, march 2001. Cited page 42

[48] O. Ait Sab and V. Lemaire. "Block turbo code performances for long-haul DWDM optical transmission systems". In *Optical Fiber Communication Conference, 2000*, volume 3, pages 280 –282 vol.3, 2000. Cited pages 42 and 66

[49] A. Puc, F. Kerfoot, A. Simons, and D.L. Wilson. "Concatenated FEC experiment over 5000 km long straight line WDM test bed". In *Optical Fiber Communication Conference, 1999, and the International Conference on Integrated Optics and Optical Fiber Communication. OFC/IOOC '99. Technical Digest*, volume 3, pages 255 –258 vol.3, 1999. Cited page 42

[50] G. Vareille, O. Ait Sab, G. Bassier, J.P. Collet, B. Julien, D. Dufournet, F. Pitel, and J.F. Marcerou. "1.5 terabit/s submarine 4000 km system validation over a deployed line with industrial margins using 25 GHz channel spacing and NRZ format over NZDSF". In *Optical Fiber Communication Conference and Exhibit, 2002. OFC 2002*, pages 293 – 295, mar 2002. Cited page 43

[51] T. Mizuochi, K. Ishida, K. Kinjo, T. Kobayashi, S. Kajiya, K. Shimizu, T. Tokura, K. Motoshima, and K. Kasahara. "1.7 Tbit/s (85×22.8 Gbit/s) transmission over 9180 km using symmetrically collided transmission methodology". *Electronics Letters*, 38(21):1264 – 1265, oct 2002. Cited page 43

[52] T. Tsuritani, K. Ishida, A. Agata, K. Shimomura, I. Morita, T. Tokura, H. Taga, T. Mizuochi, N. Edagawa, and S. Akiba. "70-GHz-spaced 40×42.7 Gb/s transpacific transmission over 9400 km using prefiltered CSRZ-DPSK signals, all-Raman repeaters, and symmetrically dispersion-managed fiber spans". *Lightwave Technology, Journal of*, 22(1):215 – 224, jan. 2004. Cited page 43

[53] J.-X. Cai, M. Nissov, A.N. Pflipetskii, A.J. Lucero, C.R. Davidson, D. Foursa, H. Kidorf, M.A. Mills, R. Menges, P.C. Corbett, D. Sutton, and N.S. Bergano. "2.4 Tb/s (120×20 Gb/s) transmission over transoceanic distance using optimum FEC overhead and 48 % spectral efficiency". In *Optical Fiber Communication Conference and Exhibit, 2001. OFC 2001*, volume 4, pages PD20–1 – PD20–3 vol.4, march 2001. Cited page 43

[54] D. Ouchi, K. Kubo, T. Mizuochi, Y. Miyata, H. Yoshida, H. Tagami, K. Shimizu, T. Kobayashi, K. Shimomura, K. Onohara, and K. Motoshima. "A fully integrated block turbo code FEC for 10 Gb/s optical communication systems". In *Optical Fiber Communication Conference, 2006 and the 2006 National Fiber Optic Engineers Conference. OFC 2006*, page 3 pp., march 2006. Cited page 43

[55] I.B. Djordjevic and B. Vasic. "MacNeish-Mann theorem based iteratively decodable codes for optical communication systems". *Communications Letters, IEEE*, 8(8):538 – 540, aug. 2004. Cited page 43

[56] B. Vasic and I.B. Djordjevic. "A forward error correction scheme for ultra long haul optical transmission systems based on low-density parity-check codes". In *Communications, 2003. ICC '03. IEEE International Conference on*, volume 2, pages 1489 – 1493 vol.2, may 2003. Cited page 43

[57] O. Milenkovic, I.B. Djordjevic, and B. Vasic. "Block-circulant low-density parity-check codes for optical communication systems". *Selected Topics in Quantum Electronics, IEEE Journal of*, 10(2):294 – 299, march-april 2004. Cited pages 43, 52, and 57

[58] I.B. Djordjevic, S. Sankaranarayanan, S.K. Chilappagari, and B. Vasic. "Low-density parity-check codes for 40-Gb/s optical transmission systems". *Selected Topics in Quantum Electronics, IEEE Journal of*, 12(4):555 –562, july-aug. 2006. Cited page 43

[59] I.B. Djordjevic, M. Cvijetic, Lei Xu, and Ting Wang. "Using LDPC-Coded Modulation and Coherent Detection for Ultra Highspeed Optical Transmission". *Lightwave Technology, Journal of*, 25(11):3619 –3625, nov. 2007. Cited page 43

[60] I.B. Djordjevic, B. Vasic, and M.A. Neifeld. "LDPC-Coded OFDM for Optical Communication Systems with Direct Detection". *Selected Topics in Quantum Electronics, IEEE Journal of*, 13(5):1446 –1454, sept.-oct. 2007. Cited page 43

[61] Y. Miyata, W. Matsumoto, H. Yoshida, and T. Mizuochi. "Efficient FEC for Optical Communications using Concatenated Codes to Combat Error-floor". In *Optical Fiber communication/National Fiber Optic Engineers Conference, 2008. OFC/NFOEC 2008. Conference on*, pages 1 –3, feb. 2008. Cited page 44

[62] T. Mizuochi, Y. Konishi, Y. Miyata, T. Inoue, K. Onohara, S. Kametani, T. Sugihara, K. Kubo, H. Yoshida, T. Kobayashi, and T. Ichikawa. "Experimental Demonstration of Concatenated LDPC and RS Codes by FPGAs Emulation". *Photonics Technology Letters, IEEE*, 21(18):1302 –1304, sept.15, 2009. Cited page 44

[63] John Zweck, Ivan T. Lima Jr., Yu Sun, Aurenice O. Lima, Curtis R. Menyuk, and Gary M. Carter. "Modeling Receivers in Optical Communication Systems With Polarization Effects". *Opt. Photon. News*, 14(11):30–35, 2003. Cited page 45

[64] M. Kuschnerov, S. Calabro, K. Piyawanno, B. Spinnler, M.S. Alfiad, A. Napoli, and B. Lankl. Low complexity soft differential decoding of qpsk for forward error correction in coherent optic receivers. In *Optical Communication (ECOC), 2010 36th European Conference and Exhibition on*, pages 1 –3, sept. 2010. Cited page 45

[65] I.B. Djordjevic and B. Vasic. "Iteratively decodable codes from orthogonal arrays for optical communication systems". *Communications Letters, IEEE*, 9(10):924 – 926, oct. 2005. Cited page 50

[66] I.B. Djordjevic, H.G. Batshon, M. Cvijetic, Lei Xu, and Ting Wang. "PMD Compensation by LDPC-Coded Turbo Equalization". *Photonics Technology Letters, IEEE*, 19(15):1163 –1165, aug.1, 2007. Cited page 51

[67] M.P.C. Fossorier. "Quasi-cyclic low-density parity-check codes from circulant permutation matrices". *Information Theory, IEEE Transactions on*, 50(8):1788 – 1793, aug. 2004. Cited page 52

[68] B. Vasic and I.B. Djordjevic. "Low-density parity check codes for long-haul optical communication systems". *Photonics Technology Letters, IEEE*, 14(8):1208 – 1210, aug 2002. Cited page 53

[69] Sae-Young Chung, Jr. Forney, G.D., T.J. Richardson, and R. Urbanke. On the design of low-density parity-check codes within 0.0045 db of the shannon limit. *Communications Letters, IEEE*, 5(2):58 –60, February 2001. Cited page 54

[70] Xiao-Yu Hu, E. Eleftheriou, and D.-M. Arnold. "Progressive edge-growth Tanner graphs". In *Global Telecommunications Conference, 2001. GLOBECOM '01. IEEE*, volume 2, pages 995 –1001 vol.2, 2001. Cited page 54

[71] Shu Lin, Heng Tang, and Yu Kou. "On a class of finite geometry low density parity check codes". In *Information Theory, 2001. Proceedings. 2001 IEEE International Symposium on*, page 2, 2001. Cited page 57

[72] Lyubomir L. Minkov, Ivan B. Djordjevic, Lei Xu, Ting Wang, and Franko Kueppers. "Evaluation of large girth LDPC codes for PMD compensation by turbo equalization". *Opt. Express*, 16(17):13450–13455, 2008. Cited pages 57 and 59

[73] A.H. Gnauck and P.J. Winzer. "Optical phase-shift-keyed transmission". *Lightwave Technology, Journal of*, 23(1):115 – 130, jan. 2005. Cited page 61

[74] P.J. Winzer, G. Raybon, Haoyu Song, A. Adamiecki, S. Corteselli, A.H. Gnauck, D.A. Fishman, C.R. Doerr, S. Chandrasekhar, L.L. Buhl, T.J. Xia, G. Wellbrock, W. Lee, B. Basch, T. Kawanishi, K. Higuma, and Y. Painchaud. "100-Gb/s DQPSK Transmission: From Laboratory Experiments to Field Trials". *Lightwave Technology, Journal of*, 26(20):3388 –3402, oct.15, 2008. Cited page 61

[75] T. Mizuochi. "Recent progress in forward error correction and its interplay with transmission impairments". *Selected Topics in Quantum Electronics, IEEE Journal of*, 12(4):544 –554, july-aug. 2006. Cited page 61

[76] T. Mizuochi, Y. Miyata, T. Kobayashi, K. Ouchi, K. Kuno, K. Kubo, K. Shimizu, H. Tagami, H. Yoshida, H. Fujita, M. Akita, and K. Motoshima. "Forward error correction based on block turbo code with 3-bit soft decision for 10-Gb/s optical communication systems". *Selected Topics in Quantum Electronics, IEEE Journal of*, 10(2):376 – 386, march-april 2004. Cited page 61

[77] Ezra Ip, Alan Pak Tao Lau, Daniel J. F. Barros, and Joseph M. Kahn. "Coherent detection in optical fiber systems". *Opt. Express*, 16(2):753–791, 2008. Cited page 61

[78] Yan Han and Guifang Li. "Polarization Diversity Transmitter and Optical Nonlinearity Mitigation Using Polarization-Time Coding". In *Optical Amplifiers and Their Applications/Coherent Optical Technologies and Applications*, page CThC7. Optical Society of America, 2006. Cited pages 77 and 93

[79] Gabriel Charlet, Jeremie Renaudier, Massimiliano Salsi, Haik Mardoyan, Patrice Tran, and Sebastien Bigo. "Efficient Mitigation of Fiber Impairments in an Ultra-Long Haul Transmission of 40Gbit/s Polarization-Multiplexed Data, by Digital Processing in a Coherent Receiver". In *Optical Fiber Communication Conference and Exposition and The National Fiber Optic Engineers Conference*, page PDP17. Optical Society of America, 2007. Cited page 77

[80] W. Shieh and C. Athaudage. "Coherent optical orthogonal frequency division multiplexing". *Electronics Letters*, 42(10):587 – 589, 11 2006. Cited pages 78 and 80

[81] Simin Chen, Qi Yang, Yiran Ma, and W. Shieh. "Real-Time Multi-Gigabit Receiver for Coherent Optical MIMO-OFDM Signals". *Lightwave Technology, Journal of*, 27(16):3699 –3704, aug.15, 2009. Cited pages 79 and 82

[82] B.J.C. Schmidt, A.J. Lowery, and J. Armstrong. "Experimental Demonstrations of Electronic Dispersion Compensation for Long-Haul Transmission Using Direct-Detection Optical OFDM". *Lightwave Technology, Journal of*, 26(1):196 –203, jan.1, 2008. Cited page 80

[83] Arthur Lowery and Jean Armstrong. "Orthogonal-frequency-division multiplexing for dispersion compensation of long-haul optical systems". *Opt. Express*, 14(6):2079–2084, 2006. Cited page 80

[84] Qi Yang, Yan Tang, Yiran Ma, and W. Shieh. "Experimental Demonstration and Numerical Simulation of 107-Gb/s High Spectral Efficiency Coherent Optical OFDM". *Lightwave Technology, Journal of*, 27(3):168 –176, feb.1, 2009. Cited page 80

[85] S.L. Jansen, I. Morita, T.C.W. Schenk, and H. Tanaka. "121.9-Gb/s PDM-OFDM Transmission With 2-b/s/Hz Spectral Efficiency Over 1000 km of SSMF". *Lightwave Technology, Journal of*, 27(3):177 –188, feb.1, 2009. Cited pages 80 and 83

[86] Dayou Qian, N. Cvijetic, Junqiang Hu, and Ting Wang. "108 Gb/s OFDMA-PON With Polarization Multiplexing and Direct Detection". *Lightwave Technology, Journal of*, 28(4):484 –493, feb.15, 2010. Cited page 81

[87] A. Al Amin, H. Takahashi, I. Morita, and H. Tanaka. "100-Gb/s Direct-Detection OFDM Transmission on Independent Polarization Tributaries". *Photonics Technology Letters, IEEE*, 22(7):468 –470, april 2010. Cited page 81

[88] T.M. Schmidl and D.C. Cox. "Robust frequency and timing synchronization for OFDM". *Communications, IEEE Transactions on*, 45(12):1613 –1621, dec 1997. Cited page 81

[89] Hlaing Minn, V.K. Bhargava, and K.B. Letaief. "A robust timing and frequency synchronization for OFDM systems". *Wireless Communications, IEEE Transactions on*, 2(4):822 – 839, july 2003. Cited page 81

[90] A.B. Awoseyila, C. Kasparis, and B.G. Evans. "Robust time-domain timing and frequency synchronization for OFDM systems". *Consumer Electronics, IEEE Transactions on*, 55(2):391 –399, may 2009. Cited page 81

[91] J. Armstrong. "OFDM for Optical Communications". *Lightwave Technology, Journal of*, 27(3):189 –204, feb.1, 2009. Cited page 82

[92] Sander L. Jansen, Itsuro Morita, Tim C. Schenk, and Hideaki Tanaka. Long-haul transmission of 16×52.5 gbits/s polarization-division- multiplexed ofdm enabled by mimo processing (invited). *J. Opt. Netw.*, 7(2):173–182, 2008. Cited page 83

[93] G. J. Foschini and M. J. Gans. "On limits of wireless communications in a fading environment when using multiple antennas". *Wireless Personal Communications*, 6:311–335, 1998. Cited page 84

[94] Emre Telatar. Capacity of Multi-antenna Gaussian Channels, 1999. Cited page 84

[95] Ghaya Rekaya-Ben Othman. *Nouvelles constructions algébriques de codes spatio-temporels atteignant le compromis multiplexage-diversité*. Phd dissertation, Ecole National supérieur des télécommunications, 2004. Cited pages 85 and 92

[96] V. Tarokh, N. Seshadri, and A.R. Calderbank. "Space-time codes for high data rate wireless communication: performance criterion and code construction". *Information Theory, IEEE Transactions on*, 44(2):744 –765, mar 1998. Cited page 86

[97] B. Hassibi and B.M. Hochwald. "High-rate codes that are linear in space and time". *Information Theory, IEEE Transactions on*, 48(7):1804 –1824, jul 2002. Cited page 87

[98] Lizhong Zheng and D.N.C. Tse. "Diversity and multiplexing: a fundamental tradeoff in multiple-antenna channels". *Information Theory, IEEE Transactions on*, 49(5):1073 – 1096, may 2003. Cited page 87

[99] J.-C. Belfiore, G. Rekaya, and E. Viterbo. The golden code: a 2×2 full-rate space-time code with non-vanishing determinants. In *Information Theory, 2004. ISIT 2004. Proceedings. International Symposium on*, pages 310 – 310, june 2004. Cited page 89

[100] A. Tirkkonen and A. Hottinen. "Improved MIMO performance with non-orthogonal space-time block codes". In *Global Telecommunications Conference, 2001. GLOBECOM '01. IEEE*, volume 2, pages 1122 –1126 vol.2, 2001. Cited page 90

[101] E. Biglieri, Yi Hong, and E. Viterbo. "On Fast-Decodable Space-Time Block Codes". In *Communications, 2008 IEEE International Zurich Seminar on*, pages 116 –119, 12-14 2008. Cited page 90

[102] S. Sezginer and H. Sari. "A High-Rate Full-Diversity 2x2 Space-Time Code with Simple Maximum Likelihood Decoding". In *Signal Processing and Information Technology, 2007 IEEE International Symposium on*, pages 1132 –1136, 15-18 2007. Cited pages 90 and 96

[103] O. Damen, A. Chkeif, and J.-C. Belfiore. "Lattice code decoder for space-time codes". *Communications Letters, IEEE*, 4(5):161 –163, may 2000. Cited page 90

[104] E. Agrell, T. Eriksson, A. Vardy, and K. Zeger. "Closest point search in lattices". *Information Theory, IEEE Transactions on*, 48(8):2201 – 2214, aug 2002. Cited page 92

[105] E. Viterbo and J. Boutros. "A universal lattice code decoder for fading channels". *Information Theory, IEEE Transactions on*, 45(5):1639 –1642, jul 1999. Cited pages 92 and 138

[106] Ivan B. Djordjevic, Lei Xu, and Ting Wang. "Alamouti-type polarization-time coding in coded-modulation schemes with coherent detection". *Opt. Express*, 16(18):14163–14172, 2008. Cited page 93

[107] Wei-Ren Peng, Kai-Ming Feng, and Sien Chi. "Joint CD and PMD compensation for direct-detected optical OFDM using polarization-time coding approach". In *Optical Communication, 2009. ECOC '09. 35th European Conference on*, pages 1 –2, 20-24 2009. Cited page 93

[108] Zinan Wang, Chongjin Xie, and Xiaomin Ren. Pmd and pdl impairments in polarization division multiplexing signals with direct detection. *Opt. Express*, 17(10):7993–8004, May 2010. Cited page 99

[109] Eado Meron, Alon Andrusier, Meir Feder, and Mark Shtaif. Use of space–time coding in coherent polarization-multiplexed systems suffering from polarization-dependent loss. *Opt. Lett.*, 35(21):3547–3549, 2010. Cited pages 99 and 101

[110] Junfeng Jiang, D. Richards, S. Oliva, P. Green, and Rongqing Hui. "In-situ monitoring of PMD and PDL in a traffic-carrying transatlantic fiber-optic system". In *Optical Fiber Communication - incudes post deadline papers, 2009. OFC 2009. Conference on*, pages 1 –3, 22-26 2009. Cited page 101

[111] R.E. Freund, C.-A. Bunge, N.N. Ledentsov, D. Molin, and C. Caspar. High-speed transmission in multimode fibers. *Lightwave Technology, Journal of*, 28(4):569 –586, feb 2010. Cited page 105

[112] S. Schollmann, S. Soneff, and W. Rosenkranz. 10.7 gb/s over 300 m gi-mmf using a 2 ÃŮ 2 mimo system based on mode group diversity multiplexing. In *Optical Fiber Communication and the National Fiber Optic Engineers Conference, 2007. OFC/NFOEC 2007. Conference on*, pages 1 –3, march 2007. Cited page 105

[113] L. Berriche, K. Abed-Meraim, and J.C. Belfiore. Cramer-rao bounds for mimo channel estimation. In *Acoustics, Speech, and Signal Processing, 2004. Proceedings. (ICASSP '04). IEEE International Conference on*, volume 4, pages iv–397 – iv–400 vol.4, May 2004. Cited page 113

[114] M. Morelli and U. Mengali. "Carrier-frequency estimation for transmissions over selective channels". *Communications, IEEE Transactions on*, 48(9):1580 –1589, sep. 2000. Cited page 119

[115] Andreas Leven, Noriaki Kaneda, Ut-Va Koc, and Young-Kai Chen. "Frequency Estimation in Intradyne Reception". *Photonics Technology Letters, IEEE*, 19(6):366 –368, mar. 2007. Cited page 119

[116] N. Al-Dhahir and J.M. Cioffi. "Optimum finite-length equalization for multicarrier transceivers". *Communications, IEEE Transactions on*, 44(1):56 –64, jan 1996. Cited pages 124 and 127

[117] P.J.W. Melsa, R.C. Younce, and C.E. Rohrs. "Impulse response shortening for discrete multitone transceivers". *Communications, IEEE Transactions on*, 44(12):1662 –1672, dec 1996. Cited page 124

[118] T. Islam and K. Hasan. "On MIMO Channel Shortening For Cyclic-Prefixed Systems". In *Wireless Communications, Networking and Mobile Computing, 2008. WiCOM '08. 4th International Conference on*, pages 1 –4, 12-14 2008. Cited pages 124 and 125

[119] R.K. Martin, K. Vanbleu, Ming Ding, G. Ysebaert, M. Milosevic, B.L. Evans, M. Moonen, and Jr. Johnson, C.R. "Unification and evaluation of equalization structures and design algorithms for discrete multitone modulation systems". *Signal Processing, IEEE Transactions on*, 53(10):3880 –3894, oct. 2005. Cited page 125

[120] N. Al-Dhahir. "FIR channel-shortening equalizers for MIMO ISI channels". *Communications, IEEE Transactions on*, 49(2):213 –218, feb 2001. Cited page 127

Publication

Conference

- Mumtaz, S.; Othman, G.R.-B.; Jaouen, Y.; , "PDL mitigation in PolMux OFDM systems using Golden and Silver Polarization-Time codes," *Optical Fiber Communication (OFC), collocated National Fiber Optic Engineers Conference, 2010 Conference on (OFC/NFOEC)* , vol., no., pp.1-3, 21-25 March 2010

- Mumtaz, S.; Othman, G.-B.; Jaouen, Y.; , "Space-Time Codes for Optical Fiber Communication with Polarization Multiplexing," *Communications (ICC), 2010 IEEE International Conference on* , vol., no., pp.1-5, 23-27 May 2010

- Mumtaz, S.; Ben-Othman, G.R.; Jaouen, Y.; Charlet, G.; , "Efficient interleaving of FEC codewords for optical PSK systems," *Optical Communication, 2009. ECOC '09. 35th European Conference on* , vol., no., pp.1-2, 20-24 Sept. 2009

Journal

- Mumtaz, S.; Rekaya-Ben Othman, G.; Jaouen, Y.; "Efficient coding/decoding scheme for PSK optical systems with differential encoding" *submitted to IET Optoelectronics*

Patent

- S. Mumtaz, Ghaya Rekaya-Ben Othman, Yves Jaouën et Bruno Thedrez; "méthode et système de transmission WDM à codage chromato-temporel" nř FR 10/58204, Octobre 2010

- S. Mumtaz, G. Rekaya-Ben Othman et Y. Jaouën ;"Procédé et dispositif de modulation mettant en oeuvre une modulation différentielle, procédé et dispositif de démodulation, signal et produits programme d'ordinateur correspondants" nř FR 09/52207, Avril 2009