Universidade de Aveiro
Departamento de
Electrónica, Telecomunicações e Informática
2013

Ana Patrícia
da Silva Lopes

# Formatos de Modulação Avançados para Redes Ópticas de Acesso

# Advanced Modulation Formats for Optical Access Networks

**Ana Patrícia da Silva Lopes**

# Formatos de Modulação Avançados para Redes Ópticas de Acesso

# Advanced Modulation Formats for Optical Access Networks

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Engenharia Electrónica e Telecomuniações, realizada sob a orientação científica de Dr. António Teixeira e Dr. Mário Lima, Professores do Departamento de Electrónica, Telecomunicaçóes e Informática da Universidade de Aveiro.

**o júri / the jury**

presidente / president

**Professor Doutor Paulo Miguel Nepomuceno Pereira Monteiro**
Professor Associado da Universidade de Aveiro

vogais / examiners committee

**Professor Doutor António Luís Jesus Teixeira**
Professor Associado da Universidade de Aveiro (orientador)

**Professor Doutor Mário José Neves de Lima**
Professor Auxiliar da Universidade de Aveiro (co-orientador)

**Professor Doutor Henrique Manuel de Castro Faria Salgado**
Professor Associado da Universidade do Porto

**agradecimentos / acknowledgements**

Gostaria de começar por agradecer ao Instituto de Telecomunicações pela reunião de todas as condições necessárias à realização desta dissertação.

De igual forma, agradeço aos meus orientadores Prof. Dr. António Teixeira e Prof. Dr. Mário Lima pela oportunidade proporcionada, acompanhamento e conhecimentos transmistidos.

Ao aluno de doutoramento Ali Shahpari e ao Mestre João Prata, o meu obrigado pelos ensinamentos transmitidos em ambiente laboratorial. Um agradecimento também ao Mestre Ricardo Ferreira, bolseiro do Instituto, e ao Doutor Jacklyn Reis, pelos conhecimentos transmitidos.

O meu muito obrigado à minha família e amigos pelo apoio e bons momentos proporcionados ao longo destes anos, em especial, à minha irmã Alzira, à minha madrinha São, à minha avó Célia, por tudo o que fez por mim, ao Rafael, ao Luís, ao Diogo e à Raquel, pela amizade sincera e presença constante.

À minha prima Adriana, o meu profundo agradecimento por ter estado sempre ao meu lado quando precisei e pelos momentos memoráveis que passámos desde a nossa infância. És a minha irmã mais nova.

Ao meu namorado Nuno, o meu enorme agradecimento pelo amor, compreensão e apoio incondicionais e também pelos momentos inesquecíveis. Sem ti, tudo tinha sido mais difícil.

Tudo tenho a agradecer ao meus pais Alcídio e Ana Maria. Obrigado pelo amor, encorajamento e o apoio incondicioanis, a educação e os sacrifícios que fizeram para que eu pudesse chegar até aqui. Não podia ter mais orgulho em vocês.

**Resumo**

Nos últimos anos, as comunicações ópticas têm-se revelado a solução para a crescente procura de elevados ritmos de informação. De facto, nenhuma outra tecnologia (com ou sem fios) permite transmissões a Tbps, ou até dezenas de Tbps, como as comunicações ópticas permitem. Contudo, as actuais redes ópticas passivas não exibem uma capacidade agregada de fibra superior a 10 Gbps e os operadores esperam mais delas. Assim, uma evolução destas redes a médio prazo, a NG-PON2, tem sido discutida, sendo 40 e 10 Gbps as velocidades de download e upload que se prevêem, respectivamente. Não obstante, estes ritmos podem não ser suficientes no futuro, com a emersão dos formatos 3D, dos novos serviços baseados em cloud computing, vídeo conferência, jogos on-line e outras/os aplicações/serviços.

Neste contexto, diversos trabalhos de pesquisa têm sido efectuados. Devido à limitação do espectro e à largura de banda restrita dos dispositivos electrónicos que complementam o sistema de transmissão óptico, elevadas eficiências espectrais são um requisito e consequentemente, a utilização de formatos de modulação avançados. A implementação intrínsica de detecção coerente e processamento digital de sinal proporcionam técnicas de compensação que melhoram o desempenho do sistema de uma forma muito significativa. Uma vez que, o crescente número de bits por símbolo exige uma maior relação sinal-ruído óptica e torna o sistema mais sensível às não linearidades da fibra e ao ruído de fase do laser, estes algoritmos de compensação tornam-se essenciais.

Esta dissertação fornece uma ampla visão dos sistemas ópticos coerentes que inclui, uma extensiva apresentação dos moduladores externos ópticos, um estudo teórico dos formatos de modulação e do formato do impulso, uma descrição de algoritmos de compensação e técnicas de detecção e finalmente, simulações e resultados laboratoriais que permitem concluir qual é o formato de modulação avançado mais indicado para transmissões ópticas a elevados ritmos.

**Abstract**

In the last years, optical fibre communications have revealed themselves as the solution for the increasing demand of high data rates. In fact, no other communication technology (wireline or wireless) enables transmissions at Tbps, or even tens of Tbps, as optical communications do. However, the current passive optical networks do not exhibit aggregate fibre capacity beyond 10 Gbps and operators except more from them. Therefore, a middle-term evolution of these PONs, the NG-PON2, has been discussed, being 40 and 10 Gbps the predicted down- and upstream bit rates, respectively. Nevertheless, these rates upgrades may not be enough in the future, as 3D formats, new cloud computing services, videoconferencing, on-line games and alike applications/services become to emerge.

In this context, much research is being carried on. Because of the limited spectrum and the restrict bandwidth of the electronic devices that complement the optical transmission system, high spectral efficiency is a requirement and consequently, the use of advanced modulation formats. The intrinsic employment of coherent detection and digital signal processing provides compensation techniques that increase hugely the performance of the system. Once the increasing number of bits per symbol requires higher optical signal-to-noise ratio and makes the system more sensitive to fibre nonlinearities and laser phase noise, these compensation algorithms become essential.

This thesis provides an ample overview of coherent optical systems, which includes an extensive presentation of external modulators, a theoretical study of modulation formats and pulse shaping, a description of compensation algorithms and detection techniques and finally, simulations and laboratory results that will provide a conclusion about what is the most suited advanced modulation format for high data rates optical transmissions.

# Contents

# List of Figures

[1] The original context concerns to DPSK, which has the same constellation of BPSK.

# List of Tables

## Acronyms

**16-QAM** 16-Quadrature Amplitude Modulation

**ADC** Analogic-To-Digital Converter

**AMI** Alternate Mark Invertion

**APE** Annealed Proton Exchange

**ASE** Amplified Spontaneous Emission

**ASIC** Application-Specific Integrated Circuit

**ASK** Amplitude Shift Keying

**BER** Bit Error Rate

**BPF** Band-Pass Filter

**BPSK** Binary Phase Shift Keying

**CD** Chromatic Dispersion

**CMA** Constant Modulus Algorithm

**CS-RZ** Carrier-Suppressed Return-To-Zero

**CW** Continuous Wave

**DAC** Digital-to-Analogue Converter

**DB** Duobinary

**DC** Direct Current

**DD-EQ** Decision Directed Equalizer

**DD-MZM** Dual-Drive Mach-Zehnder Modulator

**DFB** Distributed Feedback

**DGD** Differential Group Delay

**DMF** Dispersion-Managed Fibre

**DML** Directly Modulated Laser

**DOF** Degree Of Freedom

**DP-QPSK** Dual Polarization-Quadrature Phase Shift Keying

**DPSK** Differential Phase Shift Keying

**DQPSK** Differential Quadrature Phase Shift Keying

**DSP** Digital Signalling Processing

**EAM** Electro-Absorption Modulator

**EDFA** Erbium Doped Fibre Amplifiers

**EML** Electro-Absorption Modulated Laser

**EML-SOA** Electro-Absorption Modulated Laser-Semiconductor Optical Amplifier

**EOM** Electro-Optic Modulator

**FEC** Forward Error Correction

**FIR** Finite Impulse Response

**F-OAN** Future-Optical Access Network

**FSAN** Full Service Access Network

**FTTx** Fiber-To-The-x

**FWHM** Full-Width at Half-Maximum

**FWM** Four-Wave Mixing

**G-PON** Gigabit-Passive Optical Network

**IF** Intermediate Frequency

**IIR** Infinite Impulse Response

**IM-DD** Intensity-Modulated Direct-Detection

**IQM** IQ Modulator

**ISI** Intersymbol Interference

**ITU-T** International Telecommunications Union-Telecommunications

**$LiNbO_3$** Lithium Niobate

**LO** Local Oscillator

**LPF** Low-Pass Filter

**MLSD** Maximum-Likelihood Sequence Detection

**MZDI** Mach-Zehnder Delayed Interferometer

**MZI** Mach-Zehnder Interferometer

**MZM** Mach-Zehnder Modulator

**NG-PON1** Next Generation-Passive Optical Network 1

**NG-PON2** Next Generation-Passive Optical Network 2

**NRZ** Non-Return-To-Zero

**N-WDM** Nyquist-Wavelength Division Multiplexing

**OCR** Optical Coherent Receiver

**OOK** On-Off Keying

**OPLL** Optical Phase-Locked Loop

**OSNR** Optical Signal-To-Noise Ratio

**OTDM** Optical Time Division Multiplexing

**P2MP** Point-To-Multipoint

**PAM** Pulse Amplitude Modulation

**PAPR** Peak-to-Average Power Ratio

**PDM-16QAM** Polarization Division Multiplexing-16-Quadrature Amplitude Modulation

**PM** Phase Modulator

**PMD** Polarization Mode Dispersion

**Pol-SK** Polarozation-Shift Keying

**PON** Passive optical Network

**PSBT** Phase-Shaped Binary Transmission

**PSD** Power Spectral Density

**PVT** Process, Voltage and Temperature variations

**QCSE** Quantum Confined Stark Effect

**QPSK** Quadrature Phase Shift Keying

**RDE** Radially Directed Equalizer

**RRC** Root-Raised Cosine

**RZ** Return-To-Zero

**SD-MZM** Single-Drive Mach-Zehnder Modulator

**SPM** Self-Phase Modulation

**SSMF** Standard Single Mode Fibre

**Ti** Titanium

**ULAF** Ultra-Large Area Fibre

**UDWDM** Ultra-Dense Wavelength Division Multiplexing

**WDM** Wavelength Division Multiplexing

**XPM** Cross-Phase Modulation

# Chapter 1

# Introduction

We are now observing the transition from *Information Age*, which was characterized by the ample availability of information, to the *Connected Age*, where not only man but also machine are, or can easily be, connected each other in order to simply exchange data or provide different kinds of services. In fact, the connection of an individual to the rest of the world can be accomplished with just one click.

Studies provided by IEEE Communications Society® show that the utilization of the Internet increased 556% worldwide between 2000 and 2012, being Africa and the Middle East the Globe regions where the growth was more significant with a percentage of 3607% and 2640%, respectively. On the other hand, regions where the growth was not that flagrant, such as North America or Europe (153% and 393%, respectively), were the regions where the Internet penetration has experienced the major growth (78.6% and 63.2%) in the same period of the time [34].

Particularly, these last numbers were certainly driven by the incredible technological development verified in the last decade and also in the 90's. The arising of smartphones, tablets and alike smart gadgets together with new applications and web services take us to a new communication paradigm, in which people make use of social networks, video hosting websites, videoconferencing and search engines to interact with other people around the world and to keep themselves informed about whatever happens out there.

Nowadays, more than having connections available anywhere, people want to be connected and get the information they search for as fast as possible. Also, there are other factors that justify the constant increase of network traffic. E.g., it is predictable that the 3D revolution, interactive e-learning, telemedicine services, online games and cloud computing services will hugely contribute to the data rate increase.

And this is how, optical communications (fibre and free-space) reveal themselves as the only communications technology capable of answering for the huge bit rate demands, as Figure 1.1 illustrates.

## 1.1 Motivation

As it is shown in Figure 1.1, optical communications are the only technology between both wireline and wireless solutions, which allow bit rates of order of tens of Terabits. In addition, it is also the only communication technology whose transmission distance can be as long as thousands of kilometres, excepting satellite communications.

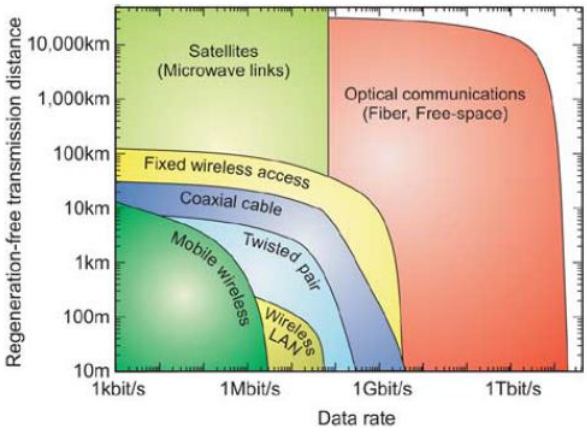



Figure 1.1: Regeneration-free transmission distance versus bit rate for various wireless and wireline communication technologies.[1]

---

Passive optical networks (PONs) are characterized as a point-to-multipoint (P2MP) architecture where all network elements are desirable to be passive, i.e., they do not require power supplies to operate.

PONs have been the most chosen solution for Fiber-to-the-x (x = home, building, etc.) (FTTx) deployment, which have been experienced a great development since 2004. As Figure 1.2 shows, the up- and downstream bit rates of the Gigabit-PON (G-PON) standard is 1.25 and 2.5 Gbps, respectively. Then, the increase need of bandwidth and services support capacity dictates the evolution of the G-PON. From Full Service Access Network (FSAN) and International Telecommunication Union-Telecommunication (ITU-T) point of view, the evolution is divided into two stages: Next Generation-PON1 (NG-PON1) and Next Generation-PON2 (NG-PON2). The former is considered a short-term evolution and it is standardised since 2010, while the latter is considered a medium-term evolution and its standardisation should be finished in 2015. NG-PON1 is also known as XG-PON or 10G-PON, being the up- and downstream bit rates 2.5 and 10 Gbps, respectively. The major goal of NG-PON2 deployment is to increase in four times these up- and downstream bit rates values.

In spite of the present aggregate bit rates in the PONs that are deployed worldwide do not exceed 10Gbps, researches have been developed in order to increase these rates beyond hundreds of Gbps and even Tbps.

Just note some remarkable work reported in the present year of 2013:

- *Schmogrow et al.* [35] reported a 100 km transmission through an ultra-large area fibre (ULAF) of PDM-16QAM at a bit rate of 252Gbps using Nyquist pulse shaping.
- *Huang et al.* [36] demonstrated a real-time 400Gbps ($4 \times 100$ Gbps) transmission of DP-QPSK, using Nyquist pulse shaping, over 3600 km of dispersion-managed fibre (DMF).
- *Shahpari et al.* [37] investigated a bi-directional Terabit+ ($192 \times 10$ Gbps) transmission of Nyquist shaped 16-QAM in a total spectrum of 12.8 nm over 40 km of standard single mode fibre (SSMF).

From these three successful experiments, one can distinguish a common option: the use of

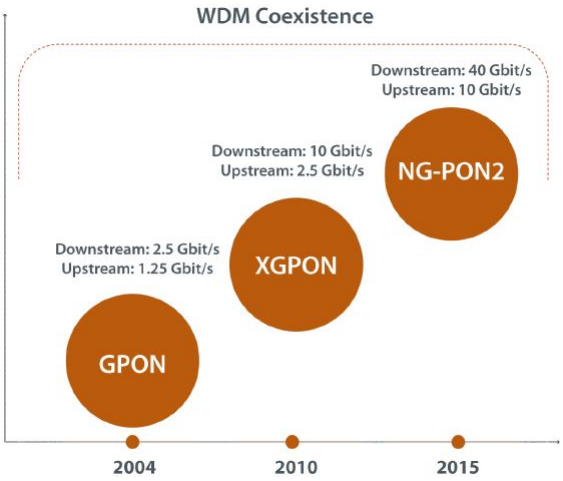



Figure 1.2: Passive Optical Networks evolution.[2]

high-order modulation formats together with Nyquist pulse shaping. Particularly, *Shahpari et al.* developed their work in the context of future Optical Access Networks (F-OAN), proving that the choice of the referred modulation formats will result in occupying narrow bandwidth.

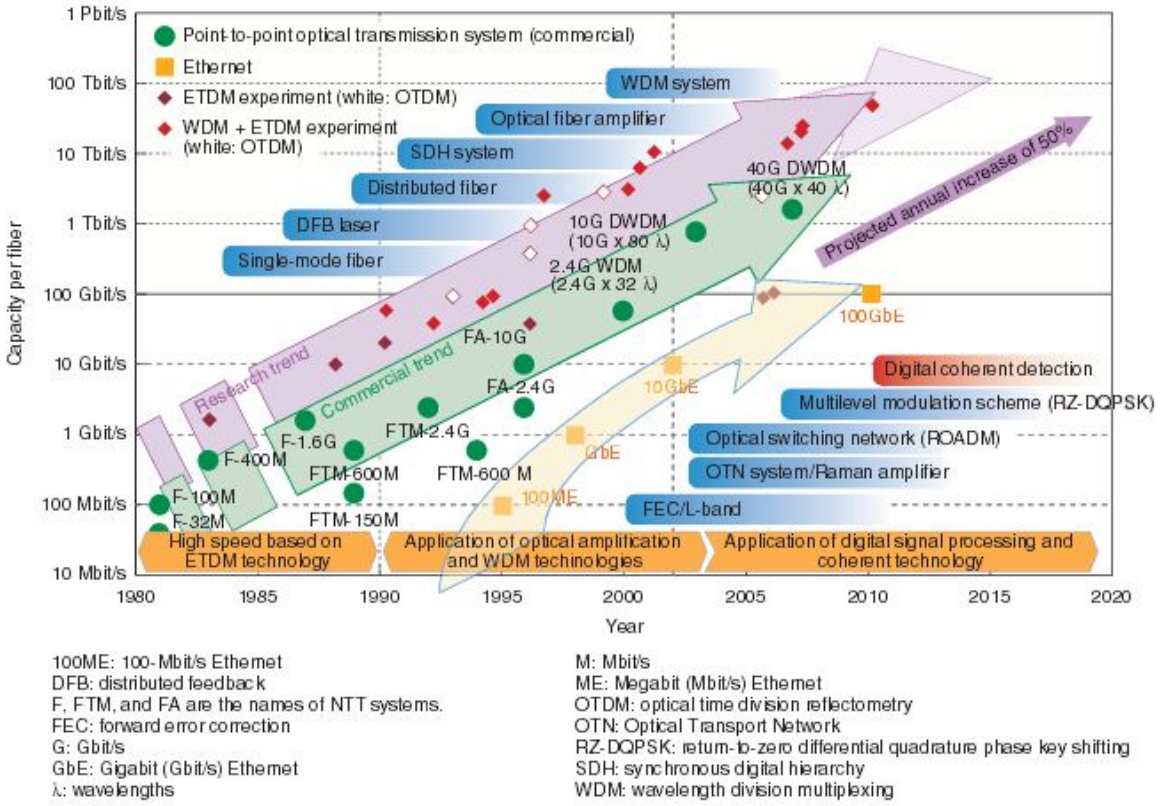
The employment of such modulation formats, like Quadrature Phase Shift Keying (QPSK) and 16-Quadrature Amplitude Modulation (16-QAM), implies the use of coherent detection together with digital signalling processing (DSP), as it will become apparent in the following chapters. In fact, the employment of coherent detection and the intrinsic application of DSP was the reason of the continuous bit rate growth in optical fibre transmissions, as it is illustrated in Figure 1.3.

However, the use of high modulation formats has its cons. As the number of bits per symbol increases, the corresponding constellation becomes larger, which requires a much higher optical signal-to-noise ratio (OSNR) so that a proper transmission can be accomplished. Also, the system sensitivity to fibre nonlinearities and laser phase noise increases with the number of bits per symbol as well. Therefore, the modulation format choice has to take into account these pros and cons.

Finally, Nyquist pulse shaping imposes itself due to its near ideal rectangular spectrum. Nevertheless, its generation is way more complicated when compared, for instance, with Non-Return-to-Zero (NRZ).

## 1.2 Objectives

From what was previously exposed, one can resume the current scenario of optical fibre communications that has encouraged the study of higher modulation formats such as QPSK and 16-QAM. The demand for high data rates in today's world is patent and the operators expect more from the deployment of passive optical networks worldwide. It is predictable that the aggregate 10 and 40 Gbps up- and downstream bit rates, respectively, of the NG-PON2 will become insufficient to answer to the costumers requirements. Therefore, extensive research work has been carried on in order to accomplished a viable PON architecture to




100ME: 100-Mbit/s Ethernet
DFB: distributed feedback
F, FTM, and FA are the names of NTT systems.
FEC: forward error correction
G: Gbit/s
GbE: Gigabit (Gbit/s) Ethernet
λ: wavelengths

M: Mbit/s
ME: Megabit (Mbit/s) Ethernet
OTDM: optical time division reflectometry
OTN: Optical Transport Network
RZ-DQPSK: return-to-zero differential quadrature phase key shifting
SDH: synchronous digital hierarchy
WDM: wavelength division multiplexing

Figure 1.3: Evolution of optical transmission capacity and its trends.[3]

implement in the field as, e.g., the next solution for FTTx. Studies have proved that the investigation of high modulation formats, and consequently, coherent detection together with digital signal processing, is the right way to follow, in order to achieve the mentioned goal.

In this context, several modulation formats are addressed in this thesis, as well as, the line codes and the Nyquist pulse shaping. The intention is to take conclusions about the systems response to each format by comparing the different pros and cons related to each one and distinguishing the most probable efficient solution for future PONs.

Therefore, diverse combinations between multi-level signalling, line code and Nyquist pulse shaping will be studied in terms of achieved bit rate, spectral efficiency/bandwidth saving, number of transmitted channels and spacing between them. A collateral conclusion will be the advantages of both coherent detection and digital signal processing.

## 1.3 Thesis organization

The different topics were organized in such a way that make this document concise and clear in order to provide the best understanding possible.

In the following, the chapters are briefly described:

- **Chapter 2: External Modulators** This chapter contains information about external modulators, whose function is to imprint information into the optical carrier. It begins with a description of the electro-absorption modulators (EAMs) and electro-optic modulators (EOMs) by explaining the related physic effects and comparing the performance of each one of them. Then, the latter is focused more extensively in the remaining of the chapter, since the Mach-Zehnder modulator is the elected external modulator, for the advanced modulation formats generation, which falls in the family of the EOMs.

- **Chapter 3: Coding and Modulation Formats** After presenting the external modulators, the diverse modulation formats are addressed in this chapter. First, the given information concerns about the pulse shaping transmitted in the fibre. Therefore, the two different line codes (Non-Return-to-Zero (NRZ) and Return-to-Zero (RZ)), the partial response signalling (Duobinary) and the Nyquist shaping are exposed and compared between each other. Then, Section 3.4 describes different modulation formats: their generation, spectral efficiency, constellation diagram among other features are discuss in order to characterize each format.

- **Chapter 4: Digital Coherent Optical Receivers** In this chapter, the advantages mentioned previously about the use of coherent receivers and digital signalling processing (DSP) will become apparent. Before giving highlights about the different ADC features that should be take into account to do the right choice of such device, detection techniques are extensively addressed in order to understand the proper detection of each modulation format. Next, DSP algorithms for transmission impairments compensation are exposed followed by a summarized information about forward error correction employment. Finally, the *Optical Coherent Receiver* analyser is presented. It is a digital signal processing software in which the received signal is compensated from transmission impairments.

- **Chapter 5: Simulation and Laboratory Experiments** In this chapter one can find the simulation and laboratory results that were obtained in the context of this

thesis. The first section addresses the developed simulations, which have allowed to study the impact of the transmissions impairments as well as other aspects of UDWM transmissions employing several modulation formats, such as the impact of the number of transmitted channels and the spacing between them. The last section describes the laboratory work, which validated not only the simulation results, but also proved the viability of employing Nyquist shaped advanced modulation formats and the consequent benefits in terms of bandwidth savings and data rates.

- **Chapter 6: Conclusions and Future Work** As the name indicates itself, the main conclusions about the developed work will be presented in this chapter. Moreover, it ends with a perspective of future work.

# Chapter 2

# External Modulators

Optical modulation is the process of imprinting data, which is usually available under the form of an electrical signal, onto an optical carrier by varying one or more of its physical characteristics. The electrical field of the optical carrier emitted by a continuous wave (CW) laser is described by [10]

$$\overrightarrow{E}_{CW}(t) = \sqrt{P}e^{j(\omega_s t+\phi_s)}\overrightarrow{e}_s \tag{2.1}$$

From this expression, one can conclude that power $P(t)$, phase $\phi(t)$, angular carrier frequency $\omega_s$ and polarization $\overrightarrow{e}_s$ are the physical characteristics that can be modified to encode the information. Therefore, intensity, phase, frequency and polarization modulation can be performed. From now on, the vectorial features will be denoted in bold, rather than with an arrow.

Intensity modulation can be accomplished by using two distinct strategies: the direct modulation and the external modulation. In the former, the light intensity is modified by varying the driving current value of the semiconductor laser. In the latter, the light that is emitted by the CW laser has a constant emitted power and the modulation is accomplished by using an external modulator. Figure 2.1 illustrates the concept of these two modulation techniques.

In the following, only external modulation will be studied because it is the option used to generate advanced modulation formats at high bit rates in the context of this thesis.

## 2.1 Electro-Absorption and Electro-Optic Modulators

Electro-absorption modulators (EAM) and Electro-optic modulators (EOM) are the two different types of external modulators. In EAMs, the modulation is achieved due to the variation of the absorption of the semiconductor by applying an external electric field. In EOMs, the application of an external electric field causes a change in the refractive index. Once the phase of the incoming optical carrier is dependent on the refractive index, the phase modulation can be accomplished. Unfortunately, intensity modulation cannot simply be achieved by altering the refractive index. In this case, it is necessary to employ an interferometric structure, which will be presented soon.

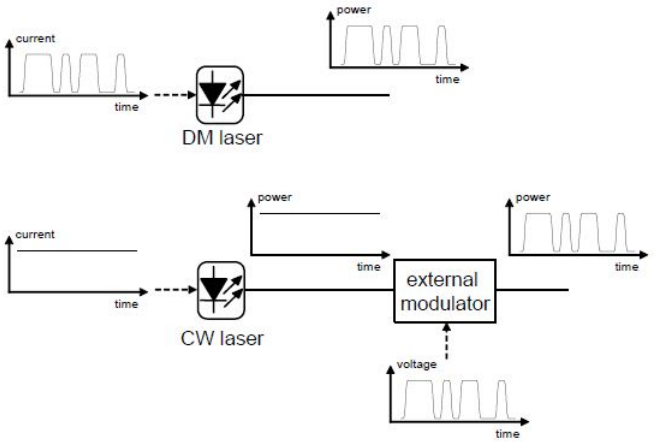



Figure 2.1: Illustration of the concepts of direct (top) and external (bottom) modulation.[4]

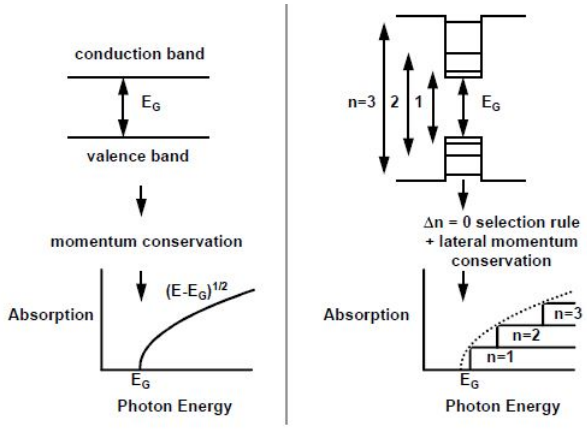



Figure 2.2: Models for absorption in (left) bulk materials and (right) quantum-well materials. [5]

---

### 2.1.1 Electro-Absorption Effect

Ternary and quaternary alloys of III-V semiconductors, such as InGaAs and InGaP, grown in InP substrate are the materials used to produce EAMs operating at 1.3-1.6$\mu$m [33].

In bulk materials, the process of changing the absorption coefficient by applying an external electric field is called Franz-Keldysh effect, while in quantum-well materials, it is known as the quantum confined Stark effect (QCSE) [38]. The main difference between bulk and quantum-well materials is the model for absorption. In the first, an electron is raised from the valence band to the conduction band by absorbing a photon, maintaining the same momentum, and the strength of all transitions are considered similar, although energies differ according to each vertical transition. On the other hand, in the QCSE, only transitions, due to the absorption of a photon, between states of the same quantum number in both bands are allowed and the momentum may not be conserved. Figure 2.2 illustrates the transitions between valence and conduction bands for both bulk and quantum-well materials.

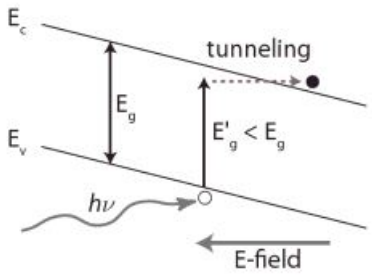



Figure 2.3: Illustration of the Franz-Keldysh effect.[6]

---

The Franz-Keldysh effect was first studied independently in 1958 by Walter Franz and Leonid Keldysh. Figure 2.3 shows an illustration of the Franz-Keldysh effect where the semiconductor is under the influence of a reversed bias. The $E_v$ and $E_c$ represent the valence and conduction bands, respectively. A tilting in the energy bands is achieved by applying an external electric field $E$. Therefore, the valence electron can tunnel to the conduction band more easily, which causes an increase in the absorption of incident photons. The incident photons energy is given by the Planck's Law $h\nu$ and is lower than the bandgap $E_g$. The effective shift of the bandgap is described by [6]

$$\Delta E_g = -\frac{3}{2}\left(\frac{q\hbar E}{\sqrt{m^*}}\right)^{2/3} \tag{2.2}$$

where $q$ represents the electron charge, $\hbar$ is the Planck's constant divided by $2\pi$ and $m^*$ is the electron effective mass [kg].

The QCSE is the process of reducing the transition energy $E_{hp}$ between confined energy levels for electrons and holes by applying an external electric field in the growth direction represented in Figure 2.4 as $F$. If the electrical field increases, the overlap of electron and hole wavefunctions will be reduced. Thus, the absorption will be decreased as well at the transmission energy. On the other hand, the formation of electron-hole pairs (exciton) improves the absorption. The QCSE is traduced in a shift of the absorption edge toward longer wavelengths in the EAM absorption spectra. Comparing thick and thin quantum wells, one can conclude that the former shows higher modulation efficiency and the latter provides stronger absorption [7].

Although quantum well materials offer higher modulation efficiency than the bulk materials due to the use of the lower drive voltages, these last ones are less wavelength sensitive, which is traduced in a larger optical bandwidth. In fact, the high modulation efficiency is not only a consequence of low drive voltages. In EAMs, both optical mode and electrical applied field are confined in a very small area, which not only allows high modulation efficiency but also makes the overall optical loss tolerable.

### Integrated EML-SOA

Typically, the optical propagation loss has values of about 15 to 20 dB/mm. The residual absorption in the active layer, the interband absorption loss induced by free carriers in the highly doped layers and the scaterring loss caused by the roughness of the deep-ridge waveguide sidewalls and defects in the growth material are the main causes of such high propagation

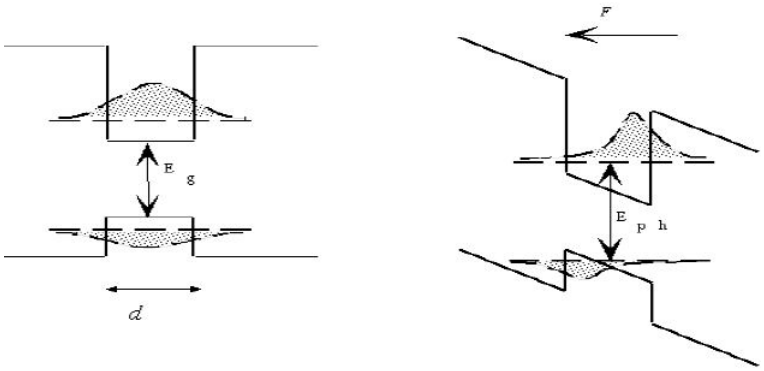



Figure 2.4: Illustration of the quantum confined Stark effect.[7]

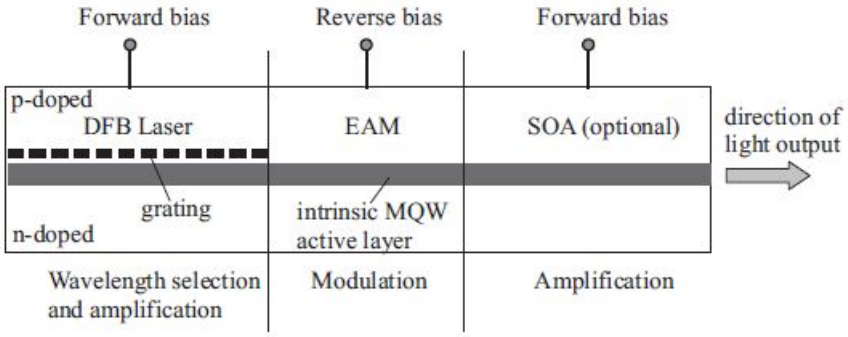



Figure 2.5: Schematic illustration of a monolithically integrated distributed feedback laser, electroabsorption modulator and a semiconductor optical amplifier.[8]

---

loss value [33]. However, the possible integration with other devices makes the EAMs a good solution for many applications. Figure 2.5 illustrates an example of such case.

The electro-absorption modulated laser (EML) is constituted by a distributed feedback (DFB) laser fabricated together with an EAM in the same substrate. The device shown in Figure 2.5 is an EML integrated with a semiconductor optical amplifier (EML-SOA). In this case, the EAM used is a QCSE type modulator. The operator wavelength has to be chosen carefully because it has to guarantee high absorption swing at EAM section as well as gain at laser and SOA sections [8].

### Frequency Chirping on Electro-Absorption External Modulators

Devices integration such as the one mentioned previously are only possible because EAMs show good size/integration relation and low drive voltage. However, the chirp generated can make the long-haul high speed transmission impossible, in spite of being lower than in the case of the direct modulation. In addition, the chirp is dynamic and dependent with the drive voltage [39].

Significant changes in absorption causes concomitant changes in real refractive index, such that electro-absorption is inherently related with electro-refraction through the Kramers-

Kroning relations [8][6].

Equation 2.3 expresses the complex form of the susceptibility $\chi$ and its usual dependence with frequency $\nu$ [8].

$$\chi(\nu) = \chi'(\nu) + i\chi''(\nu) \tag{2.3}$$

$$\chi'(\nu) = \frac{2}{\pi} P \int_0^\infty \frac{\nu' \chi''(\nu')}{\nu'^2 - \nu^2} d\nu' \tag{2.4}$$

$$\chi''(\nu) = \frac{2}{\pi} P \int_0^\infty \frac{\nu \chi'(\nu')}{\nu^2 - \nu'^2} d\nu' \tag{2.5}$$

Equations 2.4 and 2.5 are the Kramers-Kroning relations, which describe the real and imaginary part of the susceptibility, respectively. $P$ represents the Cauchy first value of the integrals.

From the Kramers-Kroning relations, the refractive index change $\Delta n'$ as a function of both wavelength $\lambda$ and voltage $V$ can be deduce, assuming fairly constant carrier density and temperature.

Equation 2.6 allows one to verify the relation between the change of the real part of the refractive index with the absorption change $\Delta\alpha$ [8].

$$\Delta n'(\lambda, V) = \frac{\lambda^2}{2\pi^2} P \int_0^\infty \frac{\Delta\alpha(\lambda', V)}{\lambda^2 - \lambda'^2} d\lambda' \tag{2.6}$$

The change of the imaginary part of the refractive index due to the absorption change is simply described by [8]

$$\Delta n'' = \frac{\lambda}{4\pi} \Delta\alpha \tag{2.7}$$

Once the change of the real and imaginary parts of the refractive index are already known, the chirp-parameter $\alpha_H$ can now be deduced, once it is defined as [8]

$$\alpha_H = \frac{\Delta n'}{\Delta n''} \tag{2.8}$$

### 2.1.2 Electro-Optic Effect

The refractive index of some materials, such as Lithium Niobate ($LiNbO_3$), can be modified by applying an external electric field. This phenomenon is known as the linear electro-optic effect or Pockets' effect.

The phase of the incoming optical carrier is related with the refractive index by [9]

$$\phi \approx \frac{2\pi}{\lambda} nL - \pi \frac{V}{V_\pi} \tag{2.9}$$

where

$$V_\pi = \frac{d}{L} \frac{\lambda}{r n^3} \tag{2.10}$$

The parameter $n$, in Equation 2.9, is the refractive index of the material when no electric field is applied. The half-wave voltage $V_\pi$, known as switching voltage as well, is the applied

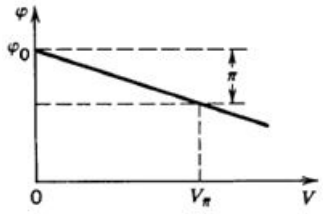



Figure 2.6: Mathematical representation of the linear electro-optic effect. [9]

---

voltage at which the phase experiences a shift of $\pi$. $L$ represents the interaction length, i.e., the quantity of the waveguide material that is in contact with the electrode where the electrical field is applied and $d$ represents the distance between the two electrodes.

From Equations 2.9 and 2.10, one can conclude that it is possible to perform phase modulation by applying an external voltage. The mathematical representation of the linear electro-optic effect is shown in Figure 2.6. As it was already mentioned, the change of the refractive index itself does not allow intensity modulation.

For the EAM case, the frequency chirping was addressed after the electro-absorption effect explanation. However, for the EOM case, this issue will be addressed later, after presenting the dual-drive Mach-Zehnder modulator in order to provide a better understanding of the concept.

### Materials for Electro-Optic Modulators Manufacture

The materials choices are restricted to ternary and quaternary alloys of III-V semiconductors in the case of the EAMs manufacturing. On the other hand, distinct materials have been studied for several years in the production of EOMs. Thus, it is worth to expose some features of each one of these materials.

$LiNbO_3$, III-V semiconductors, polymers and, more recently, silicon are the suited materials to produce EOMs.

The most widely electro-optical material used in external modulators manufacture is Lithium Niobate ($LiNbO_3$), a ferroelectric anistropic crystal with 3-m crystal symmetry. One of the reasons to select this crystal is its combination of high electro-optic coefficients and high optical transparency in the near infrared wavelengths [12]. In addition, the $LiNbO_3$ waveguide can be designed for zero-chirp or adjustable-chirp operation. In fact, zero-chirp and negative-chirp modulators minimize the negative effects related to fibre dispersion [12]. Another advantage shown by the $LiNbO_3$ is its large Curie temperature, known as Curie point as well. It means that above the Curie temperature, the crystal no longer exhibits ferroelectric phases, it becomes to show paraelectric phases [40]. The high Curie temperature makes $LiNbO_3$ a candidate for fabrication of low-loss optical waveguides through indiffusion of metals, normally Titanium (Ti). The resulting single-mode waveguide has a much lower optical propagation loss than the one presented by EAMs. Commonly, the propagation loss is lower than 0.2 dB/cm. Finally, $LiNbO_3$ external modulator is stable at thermal, chemical and mechanical levels and it is compatible with conventional integrated-circuit manufacture [12]. Despite of its pros, the Lithium Niobate has its weaknesses too.

As it was mentioned earlier, the phase modulation is achieved by altering the refractive index. However, the refractive index changes are small and large driving voltages or long

electrode lengths are required. Another problem to face is the mismatch between the electrical and optical propagation constants. This together with electrical attenuation of the electrode limits the bandwidth [12]. Nevertheless, 40 Gbps and beyond [41] can be achieved due to $LiNbO_3$ low-loss waveguides, high electro-optic effect and high optical coupling efficiency with single-mode fibre [33].

Similarly to what happens in EAMs, EOMs can be made of III-V compound materials such GaAs and InP based on bulk or on multi-quantum-wells (MQW) materials. The relatively small EO coefficient, 20 times lower than that of the $LiNbO_3$, is compensated by the large optical refractive indices and the very small region of 2-3$\mu$m where optical guided mode can be confined. At optical wavelength of 1.3-1.6$\mu$m, e.g., the refractive indexes of the InP and GaAs are 3.2 and 3.4, respectively, and the refractive index of the $LiNbO_3$ is 2.2. This leads to an improvement of 3-4 times for the index change in a linear electro-optic modulation application [33].

One advantage of InP EOM is its small spot size, which allows large electric fields to be formed even with small driving voltages. Lower driving voltages leads to lower power consumption and cost of the modulator. In addition, it is possible to perform monolithic integration of this kind of EOM with other optical devices such as lasers and SOAs.

Comparing $LiNbO_3$ and InP EOMs, some differences can be pointed out. One of them is the fact that, in the former case, the refractive index experiences a change in its value and no significant changes in attenuation constant are verified, while both refractive index and attenuation constant are modified in the last case. Another difference is the dependence of the phase change with driving voltage and the wavelength of the optical carrier. In Lithium Niobate EOMs the phase is linear dependent with the driving voltage, while this dependency is nonlinear in InP EOMs. In the case of $LiNbO_3$, the wavelength does not affect the phase change, whereas, in the other case, the phase and amplitude modulation are wavelength dependent [42].

Organic polymers and silicon are not widely used materials in EOMs manufacture, in spite of having great potential in future. Polymer EOMs can not only be integrated with diverse electrical and optical components but also offer very flexible electrode design. In addition, the optical refractive index is near to that of single-mode optical fibre, so that, good match between waveguide and fibre modes can be achieved. Despite these features, there are developments at thermal and photochemical stabilities levels to be perform. Also, power handling capability is much lower than that of the previous mentioned materials. Recently, Silicon has been studied as a material for optical devices mostly because it can lead to low-cost solutions. Note that technology and manufacture infrastructure already exists, which can be used to produce cost-effective devices in a large scale, and the possible integration of optical elements and advanced electrics using bipolar and CMOS technology is very attractive.

### 2.1.3 Comparison of Electro-Absorption and Electro-Optic Modulators

In previous sections, the main characteristics of both EAMs and EOMs were exposed. The most used modulator is the $LiNbO_3$ EOM, in spite of high driving voltages and large device length. Despite their advantages, the utilization of EAMs in long-haul systems is not possible due to frequency chirping. In conclusion, Table 2.1 compares some features of different materials mentioned before.

| TYPE | EOM | | | | EAM |
|---|---|---|---|---|---|
| MATERIAL | $LiNbO_3$ | III-V Compound bulk | III-V Compound MQW | Polymer | InP |
| Optical loss | low | medium | medium | low | high |
| Driving voltage | high | medium | low | high | low |
| Chirp | low | low | low | low | high |
| Footprint | big | big | small | big | small |
| Highest speed | 50 GHz | 40 GHz | 57 GHz | 110 GHz | 50 GHz [1] |
| Commercial use | yes | yes | yes | yes | yes |
| Extra | Bias control circuit | - | - | - | Temperature stabilizing unit |

[1]This value may be outdated, since 60 GHz was already reported [43]

Table 2.1: Comparison of electro-optical and electro-absorption modulators [33].

## 2.2 Phase Modulators

The phase modulator is the simplest EOM. It is an integrated circuit device in which an optical waveguide is embedded in an electro-optic substrate. One of the most used electro-optic substrates is the earlier mentioned crystal, the $LiNbO_3$. The dependence between the phase of optical carrier and the voltage applied to the electrodes is described by Equations 2.9 and 2.10. Once again, one shall emphasize that intensity modulation cannot be achieved. Figure 2.7 shows a simple scheme of a phase modulator.

At the output of the phase modulator, the electrical field of the optical carrier $E_{out}$ is related with the electric field of the incoming optical carrier by [10]

$$E_{out}(t) = E_{in}(t)\, e^{j\phi(t)} \tag{2.11}$$

Substituting Equation 2.9 in Equation 2.11 and ignoring the constant part of $\phi(t)$ [10]

$$E_{out}(t) = E_{in}(t)\, e^{j\pi \frac{V}{V_\pi}} \tag{2.12}$$

This way, the dependence of the modulated carrier with the applied external voltage becomes evident.

## 2.3 Mach-Zehnder Modulator

Nowadays, the most used interferometric structure in optical transmissions is the Mach-Zehnder modulator (MZM). Basically, the MZM is the Mach-Zehnder interferometer (MZI) implementation in an electro-optic crystal.

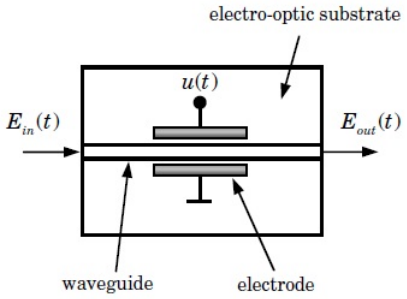



Figure 2.7: Phase modulator. [10]

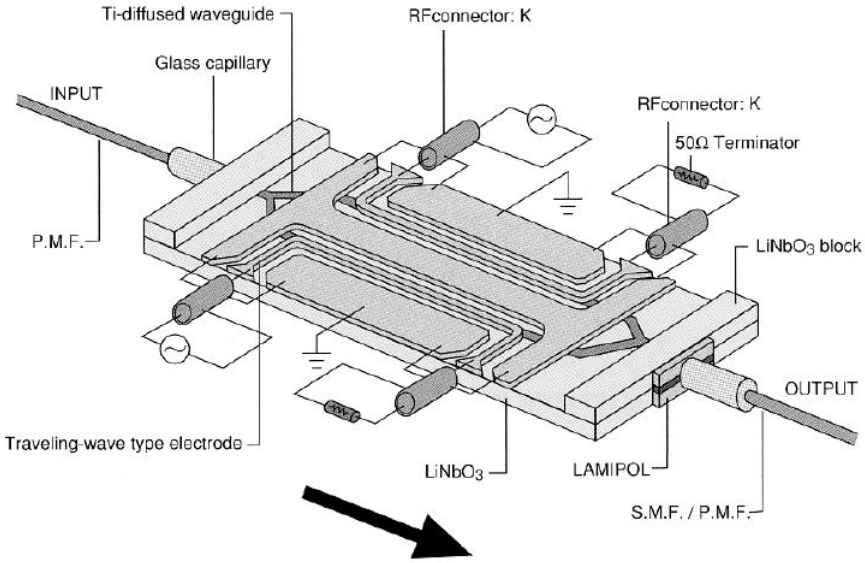



Figure 2.8: Typical layout of a $LiNbO_3$ dual-drive Mach-Zehnder modulator. (Source: Sumitomo Osaka Cement Co., Ltd) [11]

---

The MZI is not only more prominent but also more versatile than other interferometers such as the Jamin interferometer [44]. However, the theory of interference and the interferometer itself are not relevant in the context of this thesis. Therefore, in this section, focus will be given to the operability of the MZM as a whole.

### 2.3.1 Dual-Drive Mach-Zehnder Modulator

A typical layout of a dual-drive Mach-Zehnder modulator (DD-MZM) is shown in Figure 2.8. Once again, the electro-optical material that is chosen as substrate is the crystal $LiNbO_3$. The MZI structure is implemented in the substrate by indiffusing Titanium, so that refractive index is locally raised and waveguides are formed. The modulating signals are applied to gold travelling wave electrodes.

Figure 2.9 shows a simple schematic of a DD-MZM. The incoming light passes through

the input waveguide and it is splitted into two beams in the Y-junction. The optical signals suffer phase shifts $\phi_1(t)$ and $\phi_2(t)$ due to the applied voltages $u_1(t)$ and $u_2(t)$, respectively, according to the Pockets' effect. Keeping in mind the Equation 2.9, the phase shifts are defined as [10]

$$\phi_1(t) = \frac{u_1(t)}{V_{\pi_1}}\pi, \phi_2(t) = \frac{u_2(t)}{V_{\pi_2}}\pi \tag{2.13}$$

where [42]

$$u_1(t) = V_{bias_1} + v_{RF_1}(t), u_2(t) = V_{bias_2} + v_{RF_2}(t) \tag{2.14}$$

A large amount of parameters influences the phase shift: the chosen electro-optical material, the orientation of the applied electric field relatively to the principal axes of the crystal, the polarization of the optical carrier and the geometry and dimensions of the waveguide. In fact, a large length interaction (e.g. $\sim$4 cm for common $LiNbO_3$ modulators) between the optical and electrical radio frequency driving signals is needed to obtain the desired modulation and this interaction is only efficient at high frequencies if travelling electrodes are implemented in the modulator [11].

The fields are then recombined in an output waveguide via another Y-junction. Depending on the relative phase shifts, the interference at the output Y-junction can vary between constructive (relative phase shift is 0) and destructive (relative phase shift is $\pi$), which can be demonstrated mathematically. Neglecting the insertion loss in each arm of the modulator and considering that the power splitting ratios in both input and output Y-junctions are 50/50, Equation 2.15 [10] expresses the relation between the input and output electrical fields of the optical carrier.

$$E_{out}(t) = \frac{1}{2}\left(e^{j\phi_1(t)} + e^{j\phi_2(t)}\right)E_{in}(t) \tag{2.15}$$

It is possible to take some conclusions about the operating modes of the DD-MZM from Equation 2.15. If the phases induced in each arm are similar, i.e., $\phi_1(t) = \phi_2(t) = \phi(t)$ by, e.g., setting $u_1(t) = u_2(t) = u(t)$ and $V_{\pi_1} = V_{\pi_2} = V_\pi$, the DD-MZM is operating in a push-push mode. Therefore, only phase modulation is possible. Note that in these conditions, the Equation 2.15 is turned into Equation 2.11. Thus, other conditions must be found in order to achieve intensity modulation. E.g., by setting $u_1(t) = -u_2(t) = \frac{u(t)}{2}$ and $V_{\pi_1} = V_{\pi_2} = V_\pi$, the phase shifts produced in each arm become symmetrical $\phi_1(t) = -\phi_2(t)$ and in this case, the modulator is working in push-pull mode. Therefore, intensity modulation becomes evident by writing Equation 2.15 in a different form [10]

$$E_{out}(t) = \cos\left(\frac{u(t)}{2V_\pi}\pi\right)E_{in}(t) \tag{2.16}$$

By squaring Equation 2.16, the relation between the input and output powers is obtained [10]

$$\frac{P_{out}(t)}{P_{in}(t)} = \frac{1}{2} + \frac{1}{2}\cos\left(\frac{u(t)}{V_\pi}\pi\right) \tag{2.17}$$

From Equation 2.17, two conclusions can be taken. First, the power transfer function of the DD-MZM operating in push-pull mode is periodic with a period $2V_\pi$. Second, it is

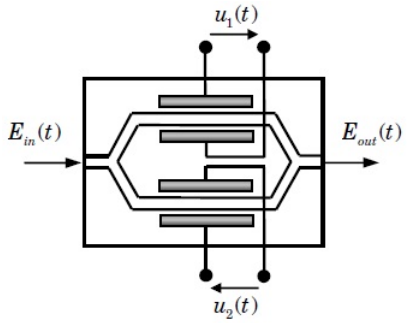



Figure 2.9: Dual-drive Mach-Zehnder modulator. [10]

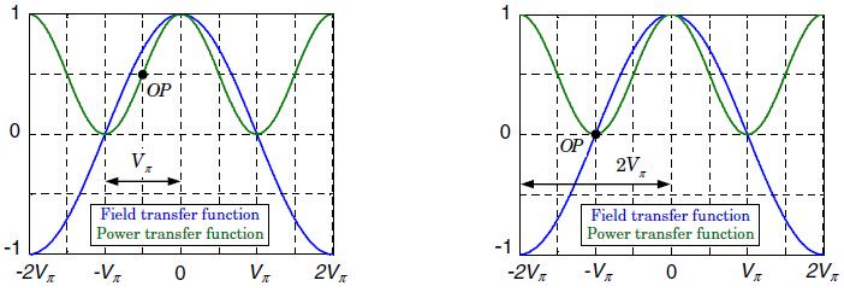



Figure 2.10: Operating the MZM in the quadrature point (left) and the minimum transmission point (right). [10]

---

important to note that phase modulation can be performed. In fact, when the drive voltage equals the half-wave voltage, a phase shift of $\pi$ is induced. However, only binary phase modulation can be performed.

Figure 2.10 shows not only the field and power transfer functions but also the two possible operating points of the DD-MZM, which are the quadrature and the minimum transmission points, depending on the bias voltage $V_{bias_{1/2}}$ ( see Equation 2.14).

The DD-MZM is operating at the quadrature point if $V_{bias_{1/2}} = V_{\pi_{1/2}}/2$, i.e., the DC bias voltage has a value corresponding to a power 3 dB lower than the maximum power of the transfer function. When DD-MZM is operating in the minimum transmission point, the $V_{bias_{1/2}}$ has a value such that the power is the lowest possible.

### 2.3.2 Crystal Cuts

There are three different crystal cuts commercially available: x-, y- and z-cut. Actually, the choice of the crystal axes orientation to the waveguides and the electrodes is critical because it affects both half-wave voltage and modulator chirp and consequently the efficiency of the $LiNbO_3$ external modulators.

As it can be seen in Figure 2.11, in the x-cut configuration, the waveguide is placed between the electrodes and in the z-cut configuration it is found below the electrodes, because the strongest component of the applied electrical field has to be aligned with the axis that provides the highest electro-optic coefficient, which is the z-axis.

The manipulation of x- and y-cut wafers are relatively straightforward. On the other hand, the handling of the z-cut wafers requires additive precautions because electrical charge accumulation on the z-faces are produced due to the crystal's piezoelectric and pyroelectric properties. In addition, in z-cut wafers, charge migrations and the build-up of pyroelectric charges lead to bias drift. Therefore, modulators made on these wafers need special design and packing. Also, a buffer layer is always required in z-cut devices in order to mitigate the attenuation of the optical mode caused by the metal absorption. This happens because the waveguides are placed beneath the electrodes. Conductive buffer layers and charge bleed layers are also produced in z-cut devices to minimize DC drift and pytoelectric charge build-up, respectively.

The buffer layer is not required in x-cut devices because the waveguide is not placed beneath the electrode. Nevertheless, it is employed in order to achieve multi-gigahertz operation. The objective of using the buffer layer is velocity matching of the RF and optical waves. Note the difference between RF dielectric constants of $LiNbO_3$ and optical dielectric constants for x- and z-cut: $\varepsilon_{x,z} = 44, 28$ and $\varepsilon_{x,z} = 4.6, 4.9$, respectively [12].

The overlap between RF and optical fields below the hot electrode is improved by a factor of two, relatively to x-cut, because the RF field flux is more concentrated in this region. However, the overlap of the fields underneath the ground electrode is minimized by a factor of three, relatively to x-cut. These two factors together result in an overall improvement of only about 20% in z-cut $V_\pi$ in comparison to x-cut $V_\pi$ for single-drive modulators [12].

Defining electro-optic efficiency per unit length as $V_\pi L$, the overall winner is the non-buffered x-cut. In fact, this configuration shows about 30% lower $V_\pi L$ than the z-cut waveguide below the hot electrode. But the absence of the buffer layer results in the velocity mismatch of RF and optical waves, which is responsible for bandwidth limitations. In addition, 25 $\Omega$ electrode impedance of nonbuffered x-cut leads to impedance mismatch and RF loss. Finally, electrodes of buffered x- and z-cut can be designed for velocity matching and impedance near to 40 $\Omega$ [12].

The annealed proton exchange (APE) process is used to produce $LiNbO_3$ waveguides. The proton exchange that is performed during this process is only achievable in x- and z-cut wafers because the acid involved in the process chemically ecthes the y-cut wafers. Perhaps, this is one of the reasons for not to consider y-cut configuration one of the most common solutions to produce external modulators. Note that the annealing step in the APE process is fundamental to obtain high-quality electro-optic waveguides [12].

### 2.3.3 Frequency Chirping on Single- and Dual-Drive Mach-Zehnder Modulator

Dual-drive Mach-Zehnder modulator has been introduced ignoring a less complex implementation of the MZM: the single-drive Mach-Zehnder modulator (SD-MZM). As the name indicates, in a SD-MZM the refractive index is only modified in one of the MZM arms through the application of an external electric field.

Equation 2.13 allows one to conclude that phase modulation is performed in this case. Due to the phase dependency with time, changes in the instantaneous frequency of the signal can occur, or equivalently, frequency chirping. Frequency chirping is a critical phenomenon in optical communications, which can even impair the transmission. Normally, it is more relevant in direct modulation or when EAMs are used to perform the external modulation of the optical signal, rather then when EOMs are chosen for the same function.

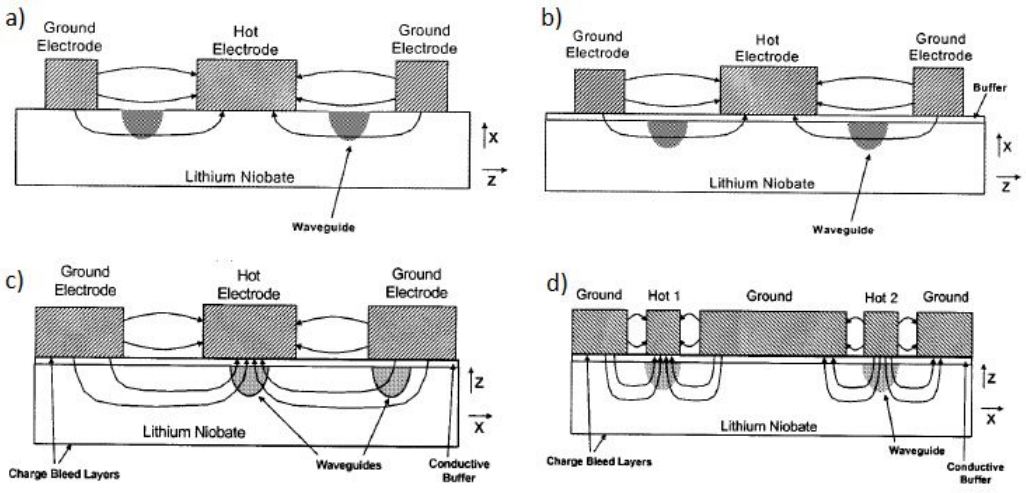



Figure 2.11: Most common electrode configuration for a) nonbuffered x-cut b) buffered x-cut c) buffered single-drive z-cut and d) buffered dual-drive z-cut [12].

---

In external modulation, using electro-optic modulators, the frequency chirping can be expressed in the same manner as in the direct modulation of a semiconductor laser. The instantaneous frequency change, in case of the directly modulated lasers (DMLs) with emitted power $P(t)$, is defined by [11] [45]

$$\delta\nu(t) = \frac{\alpha}{4\pi}\left(\frac{d}{dt}\left[ln\left(P(t)\right)\right] + \kappa P(t)\right) \tag{2.18}$$

Where $\alpha$ is the linewidth enhancement factor and $\kappa$ represents the adiabatic chirp coefficient. The first and second terms are known as the transient and adiabatic chirp, respectively. The transient chirp occurs when the emitted power varies significantly, i.e., when the applied electric signal switches from one level to another one. On the other hand, the adiabatic chirp occurs during the steady states.

In external modulation in which electro-optic modulators are used, the adiabatic chirp is residual and can be neglected. Therefore, the Equation 2.18 is reduced to [11]

$$\delta\nu(t) = \frac{\alpha}{4\pi}\left(\frac{1}{P(t)}\frac{dP(t)}{dt}\right) \tag{2.19}$$

Since, the frequency chirping is defined as [6]

$$\delta\nu(t) = \frac{1}{2\pi}\frac{d\phi(t)}{dt} \tag{2.20}$$

in the EOM case, the $\alpha$-parameter will result in [46]

$$\alpha = \frac{\dfrac{d\phi(t)}{dt}}{\dfrac{1}{2P(t)}\dfrac{dP(t)}{dt}} \tag{2.21}$$

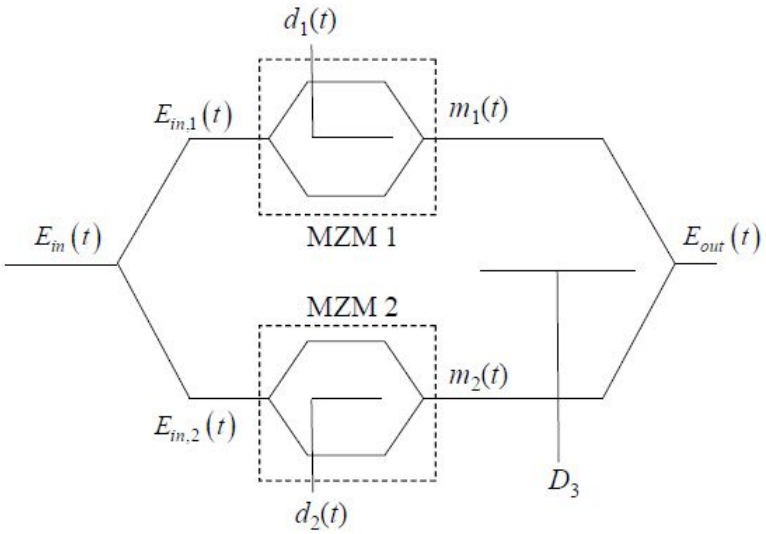



Figure 2.12: Four Phase-Modulator Structure. [13].

---

From Equation 2.15, an expression for the instantaneous phase can be deduced and therefore, its dependence with the driving voltages will become apparent. And that is because the SD-MZM operation is not chirp-free. The chirp induced by the driving signal applied to one arm, that contains the data to be imprinted in the optical carrier, can not be compensated by the driving signal applied to the other one because it does not simply exist.

In DD-MZM the chirp-free operation can be accomplished by balancing the two driving voltages. This occurs when the modulator is working in push-pull mode described earlier. Just remember that, in this mode, a driving signal is applied to one branch and its complementary signal is applied to the other branch. On the other hand, the chirp can be electrically programmable by imbalancing these driving voltages.

## 2.4 Four Phase-Modulators Parallel

The Mach-Zehnder modulators, and in particular, the dual-drive Mach-Zehnder modulator, have been widely studied. These researches have allowed to manufacture other devices based on the integration of not only several DD-MZMs but also DD-MZMs and PMs. One example of several DD-MZMs combination is the Four Phase-Modulators Parallel, while the IQ modulator is an example of the combination of DD-MZMs and a PM. The former will be exposed in this section.

As it can be seen in Figure 2.12, a DD-MZM is implemented in each arm of another DD-MZM, which is equivalent to four phase modulators working in parallel.

The major benefits of such integrated structure are the decrease of the total insertion loss and the reduction of the number of polarization controllers required due to the similar polarization dependence of the inner DD-MZMs that this structure enables.

The inner modulators are x-cut $LiNbO_3$ DD-MZMs operating in push-pull mode. Therefore, and remembering what was explained previously, the output electrical fields of $MZM_1$

and $MZM_2$ are described by [13]

$$m_1(t) = \frac{E_{in_1}}{1+\gamma_1} \exp\left[j\pi \frac{d_1(t)}{2V_{\pi_1}} + \frac{j\phi_1}{2}\right] + \frac{E_{in_1}\gamma_1}{1+\gamma_1} \exp\left[-j\pi \frac{d_1(t)}{2V_{\pi_1}} - \frac{j\phi_1}{2}\right] \tag{2.22}$$

$$m_2(t) = \frac{E_{in_2}}{1+\gamma_2} \exp\left[j\pi \frac{d_2(t)}{2V_{\pi_2}} + \frac{j\phi_2}{2}\right] + \frac{E_{in_2}\gamma_2}{1+\gamma_2} \exp\left[-j\pi \frac{d_2(t)}{2V_{\pi_2}} - \frac{j\phi_2}{2}\right] \tag{2.23}$$

It is important to note that in Equations 2.22 and 2.23, $\phi_1$ and $\phi_2$ do not represent the phase shift caused by the external voltage applied to the electrodes of the upper and lower DD-MZMs, respectively. They represent the residual phases that appear at the output of the each inner modulator due to the different interaction lengths of the two arms of $MZM_1$ and $MZM_2$. $\gamma_1$ and $\gamma_2$ are the non negative scaling factors, which are defined as the quotient of the electrical fields of the lower and upper arm of each inner modulator, i.e., if the electric field of the upper and lower arms of the $MZM_1$ were represented by $E_{in_{11}}$ and $E_{in_{12}}$, the scaling factor $\gamma_1$ would be $\gamma_1 = E_{in_{12}}/E_{in_{11}}$. $d_1(t)$ and $d_2(t)$ are the driving voltages applied to the upper modulator $MZM_1$ and the lower modulator $MZM_2$, respectively, which will be responsible for the phase shifts in each arm of the inner MZMs. It is convenient to recover Equation 2.14 and write these electrical signals as a sum of their DC and AC components [13]

$$d_1(t) = D_1 + x_1(t), d_2(t) = D_2 + x_2(t) \tag{2.24}$$

Comparing the two equations, $D_{1/2}$ is equivalent to $V_{bias_{1/2}}$ and $d_{1/2}$ is the AC electrical signal that carries the data needed to obtain the desired information at the inner DD-MZM output, which is similar to $v_{RF_{1/2}}$ of the Equation 2.14. Also, the information that was given previously about the operating points and the drive of the dual-drive Mach-Zehnder modulators remains valid.

Finally, an expression [13] which describes the $E_{out}(t)$ dependence with $E_{in}(t)$ can be deduced from all the information that has been exposed so far. The outer DD-MZM is also made in a x-cut $LiNbO_3$ design, working in push-pull mode. Therefore, the expression desired will be similar to Equation 2.22.

$$\frac{E_{out}(t)}{E_{in}(t)} = \left[\frac{1}{1+\gamma_3} \exp\left(j\pi \frac{D_3}{2V_{\pi_3}} + \frac{j\phi_3}{2}\right) \frac{m_1(t)}{E_{i_1}} + \frac{\gamma_3}{1+\gamma_3} \exp\left(-j\pi \frac{D_3}{2V_{\pi_3}} - \frac{j\phi_3}{2}\right) \frac{m_2(t)}{E_{i_2}}\right] \tag{2.25}$$

As it can be seen, Equation 2.25 has some similarities with Equations 2.22 and 2.23. $\phi_3$ is the phase difference at the outer DD-MZM output that is caused by the different interaction lengths of its arms and the non negative scaling factor of the outer modulator is denoted by $\gamma_3$. Note that it is necessary to normalize $m_1(t)$ and $m_2(t)$ because the input electric field of each inner modulator is already dependent on $E_{in}(t)$, the incoming electric field of the outer DD-MZM. Basically, the final modulated signal is a result of a combination in a Y-junction of the two modulated signals of each arm of the outer DD-MZM.

But, one question is now imposed: How can the value of the switching voltages $V_{\pi_1}$ an $V_{\pi_2}$ be determined?

In order to answer this question, the scaling factors will be considered ideal, i.e., $\gamma_1 = \gamma_2 = \gamma_3 = 1$ and the device is in the steady state regime. In these conditions, the Equation

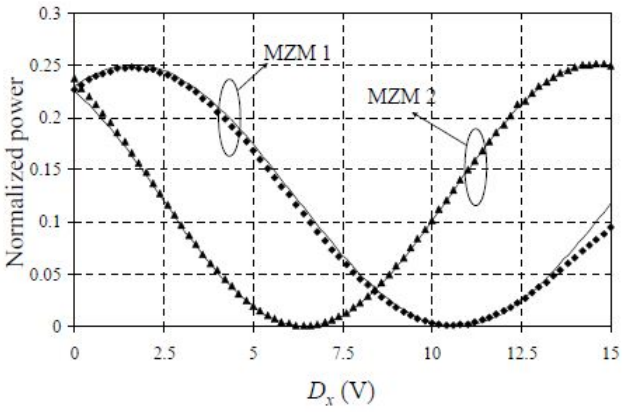



Figure 2.13: Normalized output power as a function of $D_x$. [13].

---

2.25 can be simplified as [13]

$$\begin{aligned} E_{out}(t) = &\left[\exp\left(j\pi\frac{D_3}{2V_{\pi_3}} + \frac{j\phi_3}{2}\right)\cos\left(\pi\frac{D_1}{2V_{\pi_1}} + \frac{\phi_1}{2}\right)\right] E_{in}(t) \\ &+ \left[\exp\left(-j\pi\frac{D_3}{2V_{\pi_3}} - \frac{j\phi_3}{2}\right)\cos\left(\pi\frac{D_2}{2V_{\pi_2}} + \frac{\phi_2}{2}\right)\right] E_{in}(t) \end{aligned} \tag{2.26}$$

If $D_2$ is set to $V_{\pi_2} - \phi_2 V_{\pi_2}/\pi$, the output electrical field of $MZM_2$ is null and Equation 2.26 is simplified to [13]

$$E_{out}(t) = \frac{E_{in}(t)}{2}\exp\left(j\pi\frac{D_3}{2V_{\pi_3}} + \frac{j\phi_3}{2}\right)\cos\left(\pi\frac{D_1}{2V_{\pi_1}} + \frac{\phi_1}{2}\right) \tag{2.27}$$

In these conditions, $D_3$ only influences the phase of the modulated electric field. So, if its value changes, no fluctuations are expected in the power level for a fixed $D_1$. By fixing $D_3$ to a certain value and $D_2$ to $V_{\pi_2} - \phi_2 V_{\pi_2}/\pi$, it is possible to determine both $V_{\pi_1}$ and $\phi_1$.

In practice, when cancelling the output power of $MZM_1$ or $MZM_2$ and $D_3$ is varying, the fluctuations in the output power level of the outer DD-MZM is not null as it is ideally. However, the fluctuations are lower than 1.5 dB and the DC extinction ratio of the inner DD-MZM, whose output power is not null, exceeds 23 dB. Therefore, due to this difference of values, the output power that is desired to be cancelled can be neglected [13]. In addition, these results suggest that considering an ideal scaling factor for the modulators is a good approximation for experimental cases.

The same procedure is valid to find $V_{\pi_2}$ and $\phi_2$.

The transfer function of the outer modulator, when one of the output powers of the inner DD-MZMs is shown in Figure 2.13 and given by [13]

$$T(D_x) = \frac{1}{4}\left[\cos\left(\pi\frac{D_x}{2V_{\pi_x}} + \frac{\phi_x}{2}\right)\right]^2 \tag{2.28}$$

In Equation 2.28 and Figure 2.13, $x$ can be 1 or 2 depending on which inner DD-MZM has a null output power.

To determine $V_{\pi_3}$, both $MZM_1$ and $MZM_2$ must operate in the maximum transmission point, so that variations of the output power of the outer DD-MZM become dependent on $D_3$. In this case, the normalized output power is described by [13]

$$T(D_3) = \left[\cos\left(\pi\frac{D_3}{2V_{\pi_3}} + \frac{\phi_3}{2}\right)\right]^2 \tag{2.29}$$

Note that the factor 1/4 presented in Equation 2.28 do not take place in Equation 2.29 because no power suppression occurs in inner modulators. In fact, the minimum and maximum transmission points of Equation 2.29 occur when both phases of $MZM_1$ and $MZM_2$ are out-of-phase and in-phase, respectively.

A rotation of $\pi/2$ can occur in the output electric field of the outer DD-MZM that influences $\phi_3$. If the minimum transmission point of Equation 2.29 is located between 0 V and both $D_1$ and $D_2$, the inner DD-MZMs are out-of-phase relatively to each other. On the other hand, if the minimum transmission point is located between 0 V and only one of DC voltages ($D_1$ or $D_2$), the constructive interference will change to destructive when $D_1 = D_2 = 0V$. In this situation, the mentioned rotation occurs and it will affect the value of $\phi_3$.

Moreover, if $D_3 = 0V$ and the device is operating above the midpoint, the inner modulators are in-phase for $D_1 = D_2 = 0V$. If the device is working below the midpoint, in these same conditions, $MZM_1$ and $MZM_2$ are out-of-phase.

So far, only the steady state regime was considered. It is now opportune to study an important phenomenon that occurs during the dynamic regime: the frequency chirping or simply chirp.

The introduced chirp can be manipulated in order to achieve desired conditions of the transmission. It is even possible to generate chirp-free signals. Next, an example will be exposed in order to illustrate such case.

Considering that the signal applied to the hot electrode of $MZM_2$ is not a time dependent signal $D_2$ and a non-return to zero (NRZ) signal $d_1(t)$ is applied to $MZM_1$, the frequency chirp, which is described by Equation 2.20, will be dependent on [13]

$$X_1(t) = \pi\frac{d_1(t)}{2V_{\pi_1}} + \frac{\phi_1}{2} \tag{2.30}$$

$$X_2(t) = \pi\frac{D_2}{2V_{\pi_2}} + \frac{\phi_2}{2} \tag{2.31}$$

$$X_3(t) = \pi\frac{D_3}{2V_{\pi_3}} + \frac{\phi_3}{2} \tag{2.32}$$

$X_2$ and $X_3$ allow not only to shift the chirp polarity introduced in the leading and trailing edges but also to generate a chirp-free signal. Chirp-free signals are achieved by set $X_2 = \pi/2 + k_2\pi$ and $X_3 = k_3\pi$, where $k_2$ and $k_3$ are integers [13].

## 2.5 IQ Modulator

Figure 2.14 shows the external modulator which allows to generate any constellation point in the complex IQ-plane: the IQ modulator (IQM). As it can be seen, both I arm (I from in-phase) and Q arm (Q from quadrature) have one DD-MZM operating in the push-pull

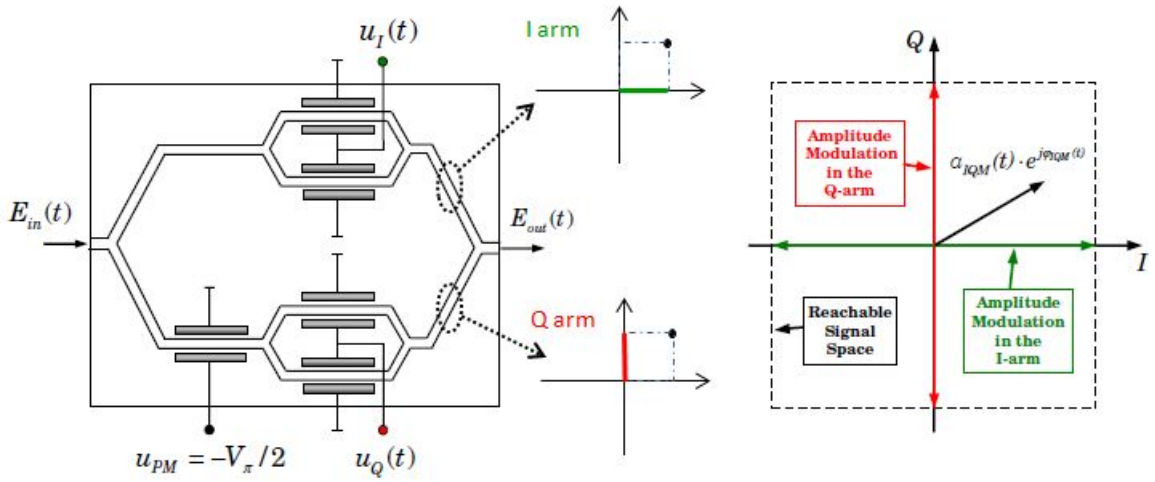



Figure 2.14: IQ modulator (left) complex IQ-plane (right). [10].

---

mode. A phase modulator is implemented in the lower branch so that a relative phase shift of $\pi/2$ is accomplished.

The output electrical field is given by the addition of the both modulated signals in the last Y-junction. Without considering the insertion loss, establishing the driving voltage of the PM to $u_{PM}(t) = -V_\pi/2$ and considering that both DD-MZMs have the same half-wave voltage $V_\pi$ and all split ratios are 50/50, $E_{out}(t)$ is described by [10]

$$E_{out}(t) = \left[\cos\left(\frac{u_I(t)}{2V_\pi}\pi\right) + j\cos\left(\frac{u_Q(t)}{2V_\pi}\pi\right)\right]\frac{E_{in}(t)}{2} \tag{2.33}$$

This way, the in-phase and quadrature components of $E_{out}(t)$ are clearly exposed and, as it can be verified, the first is dependent on the driving voltage of the upper DD-MZM, while the last is imposed by the driving voltage of the lower one. Seeing the output electric field as a complex number, its absolute value (modulus or magnitude) and its phase (or argument) are defined by [10]

$$a_{IQM}(t) = \frac{E_{in}(t)}{2}\sqrt{\cos^2\left(\frac{u_I(t)}{2V_\pi}\pi\right) + \cos^2\left(\frac{u_Q(t)}{2V_\pi}\pi\right)} \tag{2.34}$$

$$\varphi_{IQM}(t) = arg\left[\cos\left(\frac{u_I(t)}{2V_\pi}\pi\right), \cos\left(\frac{u_Q(t)}{2V_\pi}\pi\right)\right] \tag{2.35}$$

# Chapter 3

# Coding and Modulation Formats

The original data to be transmitted, which are under the form of an electromagnetic wave such as the human voice, is not normally adapted to the transmission channel. Thus, the source signal must be processed in accordance to the channel characteristics, so that an efficient transmission can be accomplished. Because of this, the coding and the modulation of the data are essential in the transmission system.

Due to the similarity of the concepts, coding and multi-level signalling can easily be confused. In fact, the modulation plays a rule as or more important than the coding. The high-order modulation formats constitute an huge breakthrough in obtaining transmission rates of 10 Gbps and beyond.

## 3.1 Line Codes

During the optical signal propagation, the pulses enlarge, interfering with their neighbours. In the receiver, the incoming signal is sampled, and the original data is recovered by comparing the sample value with the decision level. Thus, when one symbol overlaps another, it can modify the value of the sample, causing an error in the detection. This problem is known as intersymbol interference (ISI).

The easy way to eliminate the ISI is to increase the bandwidth. However, this solution not only represents a bandwidth waste, but also introduces noise which can be responsible for the increasing of the error probability.

Line codes are one of the solutions that allow a significant decrease of the ISI, without bandwidth waste. Inclusively, it is possible to improve the spectral efficiency. Non-Return-to-Zero (NRZ) and Return-to-Zero (RZ) are two line codes widely used in optical communications. By applying these line codes to the original data bit sequence, the statistic properties of the original signal are changed, in order to adapt the signal to the frequency response of the channel. Independently of being unipolar or bipolar, the difference between NRZ and RZ, in the time domain, is that the value of the non-zero level in NRZ is maintained during the entire bit interval $T_b$ whereas in RZ, the non-zero level has the duration of only a portion of $T_b$.

The properties that must be evaluated when choosing the proper line code for each situation are the following [17]:

- **Favourable power spectral density:** The power spectral density (PSD) of the signal must be adapted to the frequency response of the channel to avoid significant distortion.

E.g., if the channel shows a poor low-frequency response, the chosen line code must have zero power spectral density near DC component.

- **Self-synchronization:** The line code must allow to extract timing information so that the receiver can operate in a proper manner. Long sequences of zeros and ones should be avoided to facilitate the recovery of the clock.

- **Error detection and correction:** The line code should have error detection capability and preferably enable the error correction at the receiver.

- **Efficiency:** The transmission bandwidth must be sufficiently smaller than the channel bandwidth so that ISI does not become a problem. In addition, the transmitted power should be as small as possible.

- **Transparency:** It should be possible to transmit every signal with success independently of the patterns of 1's and 0's. E.g., a long sequence of 0's should not provoke errors in the clock recovery at the receiver.

### 3.1.1 Power Spectral Density of Non-Return-to-Zero

The PSD of the line code allows to evaluate almost every properties mentioned before. Particularly, the system attributes, which are related to the spectrum, are the optical filter bandwidth required for low dispersion and crosstalk, channel spacing required for low crosstalk, sensitivity to chromatic dispersion and robustness against fibre nonlinearities.

In order to deduce the PSD of the NRZ, one must start to study the line code in the time-domain.

Both NRZ and RZ can be express as the convolution of two signals. One of them is described by the rectangular function with an amplitude $A$ and a width of $T_b$ centred at $t = 0s$ and the other is a series of Dirac deltas with different amplitudes $a_k$ centred at $kT_b$, with $k \in \mathbb{Z}$. Representing NRZ by $y(t)$ as the convolution of $p(t)$ and $x(t)$, it results in [17]

$$y(t) = p(t) * x(t) = A rect\left(\frac{t}{T_b}\right) * \sum_{k=-\infty}^{+\infty} a_k \delta(t - kT_b) \tag{3.1}$$

The corresponding PSD to $y(t)$ is defined by [17]

$$S_y(f) = |P(f)|^2 S_x(f) \tag{3.2}$$

Applying the Fourier Transform to $p(t)$, $P(f)$ will be described by [17]

$$P(f) = AT_b \text{sinc}(fT_b) \tag{3.3}$$

On the other hand, $S_x(f)$ is not simply obtained by applying the Fourier Transform. Its deduction is quite a bit complicated. It is possible to demonstrate that [17]

$$S_x(f) = \frac{1}{T} \sum_{n=-\infty}^{+\infty} R_n e^{-j\omega n T_b} \tag{3.4}$$

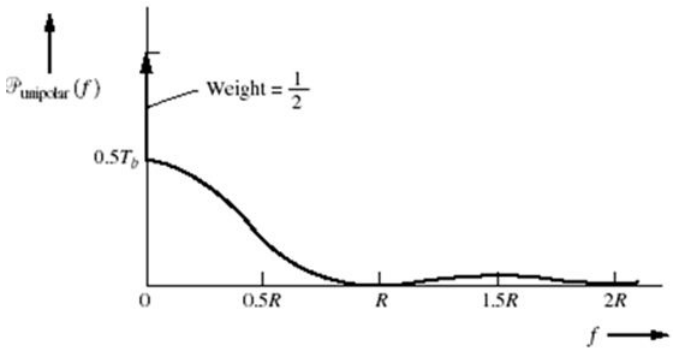



Figure 3.1: Power spectral density of unipolar NRZ. [14]

---

where $R_n$ is the autocorrelation of the data, defined as [47]

$$R_n = E\{a_k a_{k+n}\} = \sum_{i=1}^{N} (a_k a_{k+n})_i \, P_i \tag{3.5}$$

$N$ is the total number of possible bit combinations and $P_i$ represents the probability of having the i-th $a_k a_{k+n}$ product.

For unipolar NRZ, $(a_k, a_{k+n}) \in \{(0,0), (0,A), (A,0), (A,A)\}$ and the two levels are equally probable. Thus, the autocorrelation for $k = 0$ and $k \neq n$ is [47]

$$R_0 = 0 \cdot 0 \cdot \frac{1}{2} + A \cdot A \cdot \frac{1}{2} = \frac{A^2}{2} \tag{3.6}$$

$$R_n = 0 \cdot 0\frac{1}{4} + 0 \cdot A\frac{1}{4} + A \cdot 0\frac{1}{4} + A \cdot A\frac{1}{4} = \frac{A^2}{4} \tag{3.7}$$

The generic equation of $S_x(f)$ becomes [17]

$$S_{x_{unipolarNRZ}} = \frac{A^2}{4T_b} + \frac{A^2}{4T_b} \sum_{n=-\infty}^{+\infty} e^{-j\omega n T_b} = \frac{A^2}{4T_b} + \frac{A^2}{4T_b^2} \sum_{n=-\infty}^{n=\infty} \delta\left(f - \frac{n}{T_b}\right) \tag{3.8}$$

Substituting Equations 3.3 and 3.8 in Equation 3.2, the resulting expression for unipolar NRZ is [47]

$$S_y(f) = \frac{A^2 T_b}{4} \text{sinc}^2(fT_b) \left(1 + \frac{1}{T_b} \sum_{n=-\infty}^{+\infty} \delta\left(f - \frac{n}{T_b}\right)\right), n \in \mathbb{Z} \tag{3.9}$$

Due to the *sinc* function, $S_y(f)$ is null at $f = n/T_b = nR$, where R represents the bit rate, thus, the Dirac deltas that are located in these points will be cancelled. This means that the Dirac delta at the null frequency is the only one that is not cancelled. In fact, the most significant frequency components are located in the main spectral lobe. Figure 3.1 illustrates the PSD of the unipolar NRZ.

The resulting PSD of polar NRZ is very similar to the previous one and its theoretical expression is obtained with the same steps. Equations 3.2 - 3.4 are still valid. However, the autocorrelation is slightly different.

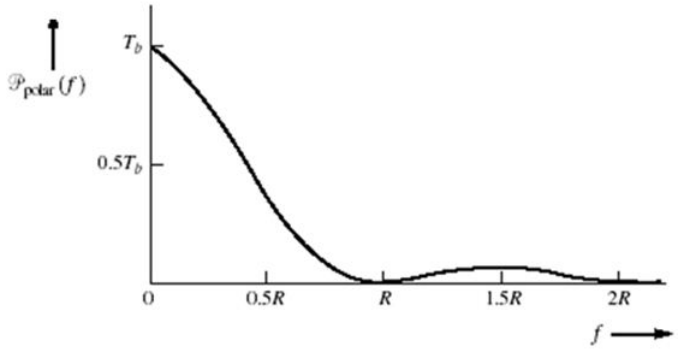



Figure 3.2: Power spectral density of polar NRZ. [14]

---

For polar NRZ, $(a_k, a_{k+n}) \in \{(-A,-A),(-A,A),(A,-A),(A,A)\}$ and the two levels are equally probable. The autocorrelation is then [47]

$$R_0 = (-A)\cdot(-A)\cdot\frac{1}{2} + A\cdot A\frac{1}{2} = A^2 \tag{3.10}$$

$$R_n = (-A)\cdot(-A)\cdot\frac{1}{4} + (-A)\cdot A\cdot\frac{1}{4} + A\cdot(-A)\cdot\frac{1}{4} + A\cdot A\cdot\frac{1}{4} = 0 \tag{3.11}$$

Therefore, the Equation 3.4 is simply reduce to $A^2/T_b$. The resulting PSD is [47]

$$S_{y_{polarNRZ}}(f) = A^2 T_b \text{sinc}^2(fT_b) \tag{3.12}$$

Figure 3.2 shows the PSD of polar NRZ. The main difference between the two PSDs already shown is the absence of Dirac deltas in the last one. Note that the location of the nulls, which are dictated by Equation 3.3, remains at $f = kR, k \in \mathbb{Z} \setminus \{0\}$.

### 3.1.2 Power Spectral Density of Return-to-Zero

As it was mentioned in the previous subsection, the RZ is also the result of the convolution of two signals. The difference is the width of the rectangular function, which in this case is only a portion of the bit interval. E.g., if the duty cycle is 50%, the width will be $T_b/2$. Therefore, the RZ line code is defined in the time-domain by [17]

$$y(t) = Arect\left(\frac{t}{T_b/2}\right) * \sum_{k=-\infty}^{+\infty} a_k \delta(t - kT_b) \tag{3.13}$$

The decrease of the width will cause a change in the Fourier Transform as well [17]

$$P(f) = \frac{AT_b}{2}\text{sinc}\left(f\frac{T_b}{2}\right) \tag{3.14}$$

Nevertheless, Equation 3.2 remains the same and $S_x(f)$ is equal to that of unipolar NRZ because the conditions are the same too. Thus, the PSD for unipolar RZ for 50% duty cycle

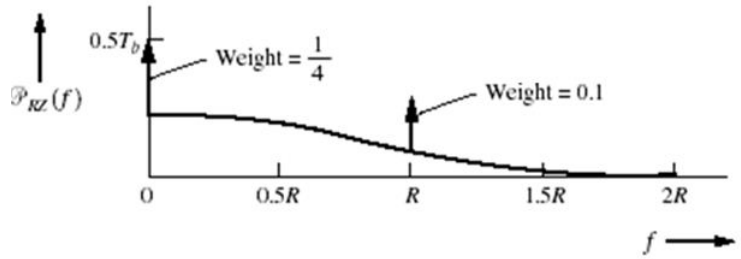



Figure 3.3: Power spectral density of unipolar RZ. [14]

---

is given by [47]

$$S_{y_{unipolarRZ}}(f) = \frac{T_b}{16} mathrmsinc^2\left(f\frac{T_b}{2}\right)\left(1 + \frac{1}{T_b}\sum_{n=-\infty}^{+\infty}\delta\left(f - \frac{n}{T_b}\right)\right), n \in \mathbb{Z} \tag{3.15}$$

From Equation 3.15, it can be conclude that $S_y(f)$ is null at $f = n2/T_b = 2nR$. Thus, the Dirac deltas are located at $f = kR$, where $k$ is an odd integer. Like in the previous case, the most relevant frequency components are located in the main spectra lobe. Figure 3.3 shows the PSD of unipolar RZ.

It is possible to obtain the PSD of polar RZ through the combination of some information given in the deductions of the previous PSDs. Once again, Equation 3.2 is valid. The pulse shape is rectangular with a width of $T_b/2$. So, in this case, $P(f)$ is given by Equation 3.14 as in the case described above. On its side, $S_x(f)$ is equal to that of the polar NRZ because amplitudes $(a_k, a_{k+n})$ can take one of the two values – $A$ or -$A$ – and both have the same probability of occurring, as in the mentioned case. This results in [48]

$$S_{y_{polarRZ}}(f) = \frac{A^2 T_b}{4}\text{sinc}^2(fT_b/2) \tag{3.16}$$

The bipolar RZ, also known as Alternate Mark Invertion (AMI), is a special case. The amplitudes $(a_k, a_{k+n})$ can take one of three values. Logic 1's are coded alternately with $A$ and -$A$, while the null power level corresponds to logic 0's.

Although Equations 3.2 and 3.14 remain valid, the PSD of the bipolar RZ is described by a somewhat different expression, comparing to the already exposed equations. This is because the autocorrelation is affected by the interchange of symmetric levels correspondent to consecutive 1's.

1's and 0's are likely probable, i.e., the products $A \cdot A$ or $(-A) \cdot (-A)$ and $0 \cdot 0$ occur with the same probability of 1/2. Therefore, the autocorrelation for $n = 0$ is [47]

$$R_0 = A^2 \cdot \frac{1}{2} + 0^2 \cdot \frac{1}{2} \tag{3.17}$$

Now, consider the particular case of consecutive and adjacent 1's. This will obligate to distinguish the autocorrelation for $|n| = 1$ from the one for $|n| > 1$. Table 3.1 shows the possible combinations of adjacent bits and the products needed to calculate $R_1$ [47]

$$R_1 = 0 \cdot \frac{1}{4} + 0 \cdot \frac{1}{4} + 0 \cdot \frac{1}{4} + (-A)^2 \cdot \frac{1}{4} = \frac{-A^2}{4} \tag{3.18}$$

| Adjacent Bits | $(0,0)$ | $(0,1)$ | $(1,0)$ | $(1,1)$ |
|---|---|---|---|---|
| $(a_k, a_{k+1})$ | $(0,0)$ | $(0,A)$<br>$(0,-A)$ | $(A,0)$<br>$(-A,0)$ | $(A,A)$<br>$(-A,-A)$ |
| Product | 0 | 0 | 0 | $A^2$ |

Table 3.1: $(a_k, a_{k+1})$ for $R_1$ in th case of bipolar RZ.

Similarly, Table 3.2 shows the possible combinations of $(a_k, a_{k+n})$ values, for $|n| > 1$. Taking into account these combinations and that they occur with the same probability of $1/4$, the autocorrelation is given by [47]

$$R_n = 0 \cdot \frac{1}{4} + 0 \cdot \frac{1}{4} + 0 \cdot \frac{1}{4} + A^2 \frac{1}{8} + (-A)^2 \cdot \frac{1}{8} = 0 \tag{3.19}$$

| Combinations | $(0,0)$ | $(0,1)$ | $(1,0)$ | $(1,1)$ |
|---|---|---|---|---|
| $(a_k, a_{k+1})$ | $(0,0)$ | $(0,A)$<br>$(0,-A)$ | $(A,0)$<br>$(-A,0)$ | $(A,A)$<br>$(A,-A)$<br>$(-A,A)$<br>$(-A,-A)$ |
| Product | 0 | 0 | 0 | $A^2$<br>$-A^2$ |

Table 3.2: $(a_k, a_{k+n})$ for $R_n, n > 1$ in the case of bipolar RZ.

At this moment, all conditions are met to write the $S_x(f)$ and, consequently, the PSD of bipolar RZ [47]

$$S_{x_{bipolar\,RZ}}(f) = \frac{1}{T_b}\left(\frac{1}{2} + \frac{-A^2}{4}e^{j\omega T_b} + \frac{-A^2}{4}e^{-j\omega T_b}\right) = \frac{1}{2T_b}\left(1 - \cos\left(2\pi f T_b\right)\right) \tag{3.20}$$

$$S_{y_{bipolar\,RZ}}(f) = \frac{A^2 T_b}{4}\mathrm{sinc}^2\left(f\frac{T_b}{2}\right)\sin^2\left(\pi f T_b\right) \tag{3.21}$$

Therefore, it is proved that, actually, the PSD of the bipolar RZ is defined by a somewhat different expression comparing to the already exposed equations. As Equation 3.21 proves, this PSD does not show any Dirac delta, and more important, the spectral components near DC have a residual value, as Figure 3.4 shows.

### 3.1.3 Comparison of Non-Return-to-Zero and Return-to-Zero

The previously mentioned line codes are not only used in intensity-modulated direct-detection (IM-DD) systems. However, this will be the type of the systems considered in this subsection because they allow a more straight forward comparison of the two mentioned line codes. In these kind of system, the information is transported in the amplitude of the optical carrier and the receiver is basically a photodiode which converts the incident optical power in electrical power according to the amplitude of the received optical signal.

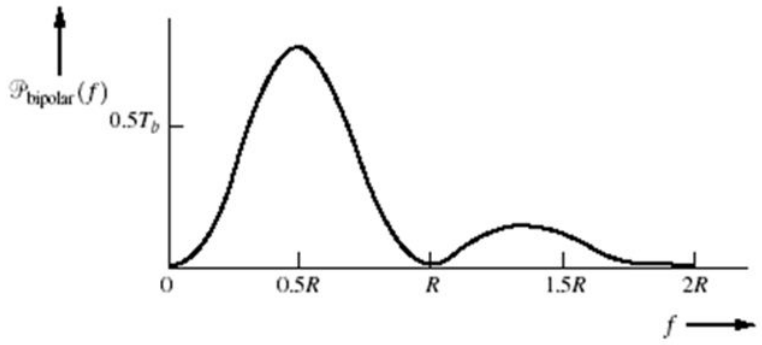



Figure 3.4: Power spectral density of bipolar RZ (know as AMI as well). [14]

---

In the early days of fibre-optical communications, NRZ was probably the most widely used line code. The need of two external modulators to generate RZ pulses in the optical domain, which will be discussed in the next subsection, is one of the reasons. Other reason to choose NRZ over RZ is easily deduced from the previous study of the PSDs, which allows one to conclude that NRZ has a narrowed spectrum than the one that correspondents to RZ. Also in the electrical domain, the required electrical bandwidth in case of using NRZ line code is narrower comparing to RZ case. Moreover, the transceivers employed in systems that use NRZ are less complex than the ones that are implemented in systems that use RZ line code. On the other hand, the larger width of the NRZ pulses can increase the ISI [49].

Nevertheless, the RZ pulse shape allows to increase the robustness against polarization mode dispersion (PMD). In addition, comparing to NRZ, RZ achieves a 1-2 dB advantage in optically pre-amplified receiver sensitivity and, more important, RZ can show a better performance over NRZ in certain cases where chromatic dispersion and fibre nonlinearities are present. However, RZ shows less spectral efficiency.

Carrier-suppressed Return-to-Zero (CS-RZ), a special form of RZ line code, is even more resistant against fibre nonlinearities and transmission impairments.

Table 3.3 summarizes comparisons related with bandwidth, clock recovery, DC component and error detections.

### 3.1.4 Pulse Carving

There are two distinct techniques of imprinting RZ line code in a CW lightwave. It can be generated under the form of an electrical field, which will then drive the external modulator. However, with today's technology, it is not possible to reach bit rates beyond 10 Gbps using this solution. Another option is to carve pulses out of the CW lightwave using an additional modulator, so-called pulse carver, as it is shown in Figure 3.5. This technique enables to reach bit rates of 40 Gbps and beyond [16].

The pulse carver can be either an EAM or a MZM, which is driven with a sinusoidal signal. Note that, multi-gigahertz sinusoidal signal with considerable drive amplitudes are easily generated. If the choice is the EAM, low duty cycle optical pulses can be achieved. Therefore, this solution is suited for optical time division multiplexing (OTDM) systems. However, EAMs show variable absorption characteristics and residual chirp, as it was mentioned in Section 2.1. Because of this, the pulse carver is usually a sinusoidally driven MZM. In this

| **Line Code** | **Advantages** | **Disadvantages** |
|---|---|---|
| Unipolar NRZ | - Relatively easy to generate;<br>- First zero crossing at R. | - DC component:waste of power<br>- Large spectral density near DC;<br>- Poor clock recovery: long sequence of 1's or 0's will cause a loss of clock signal. |
| Polar NRZ | - Relatively easy to generate;<br>- First zero crossing at R. | - Large spectral density near DC;<br>- poor clock recovery: long sequence of 1's or 0's will cause loss of clock signal. |
| Unipolar RZ | - Good clock recovery: periodic impulses at $f = nR$ can be use for clock recovery. | - First zero crossing at 2R;<br>- DC component: waste of power;<br>- Large spectral density near DC. |
| Polar RZ | -Good clock recovery: large sequence of 1's or 0's does not affect the clock recovery;<br>- Null at DC. | - first zero crossing at 2R;<br>- No error detection. |
| Bipolar RZ | - Good clock recovery: possible to convert bipolar RZ to unipolar RZ using full-wave rectification;<br>- First zero crossing at R.<br>- Null at DC;<br>- Single-error-detection: reception of two or more consecutive 1s with same polarity indicates an error. | - Not transparent: long sequence of 0's will cause loss of clock;<br>- The receiver has to distinguish between three logic levels. |

Table 3.3: Comparison of diverse line codes.

case, three carving methods can be employed, as it is shown in Figure 3.6.

- **Duty cycle of 50%** The frequency of the sinusoidal driving signal of the MZM is equal to the bit rate and its peak-to-peak amplitude is $V_\pi$ between the maximum and the minimum transmission point. This way, RZ pulses with a full-width at half-maximum (FWHM) duration equal to 50% of the bit slot are generated.

- **Duty cycle of 33%** The frequency of the sinusoidal driving signal is half bit rate and its peak-to-peak amplitude is $2V_\pi$ between two consecutive minimum transmission points. The resulting signal has a duty-cycle of 33%.

- **Duty cycle of 67%** The frequency and the peak-to-peak amplitude of the driving signal are equal, comparing to the previous case. The difference is the bias point, which in this case, is the minimum transmission point. Therefore, RZ pulses with a duty cycle

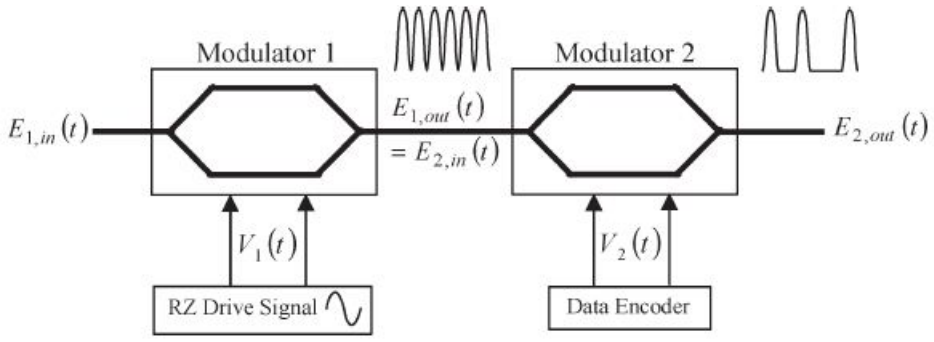



Figure 3.5: Generation of RZ signals using cascade MZMs. *Modulator 1* is the pulse carver and *Modulator 2* modulates the amplitude of the RZ signals according to data. [15]

---

of 67% and with alternating phase are generated. A signal with these characteristics is the previous mentioned CSRZ.

A data symbol sequence $\{I_n\}_{n=-\infty}^{+\infty}$ is generated by the second external modulator represented in Figure 3.5 as *Modulator 2* and it can be drawn from any modulation format, such as OOK or QPSK. This data sequence is modulated onto the periodic RZ pulse train from the first modulator output that is referred in the same figure as *Modulator 1*. Therefore, the electric field at the second external modulator output is given by [15]

$$E_{2,out}(t) = \sum_{n=-\infty}^{+\infty} I_n E_i(t - nT_s) \tag{3.22}$$

where $T_s$ is the symbol interval and $E_i$ is the electric field waveform for isolated 33%, 50% and 67% RZ pulses defined as [15]

$$E_{33}(t) = \begin{cases} \frac{1}{\sqrt{E_{33}}} \sin\left(\frac{\pi}{2}\left[1 + \sin\left(\frac{\pi t}{T_s}\right)\right]\right), & -\frac{T_s}{2} \leq t \geq \frac{T_s}{2} \\ 0, & \text{otherwise} \end{cases} \tag{3.23}$$

$$E_{50}(t) = \begin{cases} \frac{1}{\sqrt{E_{50}}} \sin\left(\frac{\pi}{4}\left[1 + \sin\left(\frac{2\pi t}{T_s}\right)\right]\right), & -\frac{T_s}{2} \leq t \geq \frac{T_s}{2} \\ 0, & \text{otherwise} \end{cases} \tag{3.24}$$

$$E_{67}(t) = \begin{cases} \frac{1}{\sqrt{E_{67}}} \sin\left(\frac{\pi}{2}\cos\left(\frac{\pi t}{T_s}\right)\right), & -\frac{T_s}{2} \leq t \geq \frac{T_s}{2} \\ 0, & \text{otherwise} \end{cases} \tag{3.25}$$

## 3.2 Partial Response Signalling

The duobinary is a specific type of partial response signalling. In this particular type of signalling two consecutive pulses are combined with each other. The signals are correlated in time and the resulting signal has a duration longer than $T_b$, i.e., only a part of the resulting signal is present in the bit interval and that is why this is called partial response signalling.

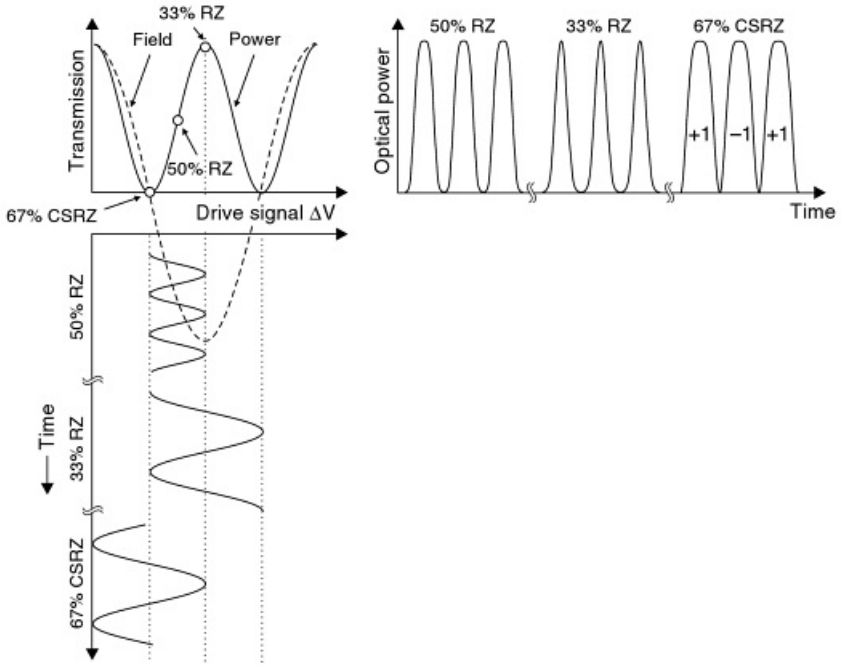



Figure 3.6: 33%, 50% and 67% duty cycle RZ using a sinusoidally driven MZM as pulse carver. (Adapted from [16])

---

Nowadays, duobinary is considered one of the most up-and-coming cost-effective solutions to use the existing 10 Gbps WDM long-haul transmission infrastructures to deploy 40 Gbps technology. This means that the design of the transmitter, receiver and transmission lines does not need to be modified, it is only necessary to replace 10 Gbps channels by 40 Gbps ones. In fact, a spectral efficiency of 0.8 bit/s/Hz at 40 Gbps per channel was already reported [18]. High spectral efficiency and high bit rates are achievable due to low spectral occupancy and high tolerance to chromatic dispersion.

In the time domain, one can define duobinary as [17]

$$y_{db}(t) = \Sigma_{k=-\infty}^{+\infty} x_k h(t - kT), k \in \mathbb{Z} \tag{3.26}$$

Where [17]

$$h(t) = \frac{4\cos\left(\frac{\pi(t - T/2)}{T}\right)}{\pi\left(1 - \frac{4(t - T/2)^2}{T^2}\right)} \tag{3.27}$$

And $x_k$ represents the k-th bit of the data sequence. Equation 3.27 is much more simple in discrete time domain [17]

$$h(nT) = \begin{cases} 1, & n = 0, 1 \\ 0, & \text{otherwise} \end{cases} \tag{3.28}$$

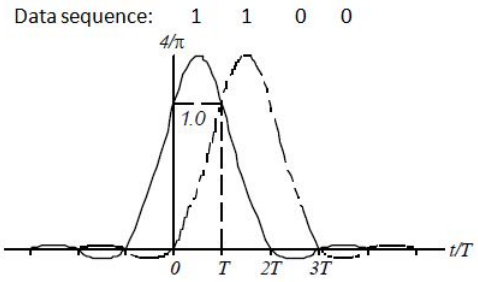



Figure 3.7: Duobinary pulses shape and the consequent ISI. (Adapted from [17])

---

Thus, from Equations 3.26 - 3.28, one can conclude that ISI exists. However, the existing ISI is introduced in a controlled manner. Note that only the *k-1* adjacent neighbour interferes in the $k$ data bit, as it is shown in Figure 3.7. This way, the $k$ data bit in the receiver input is defined as [17]

$$y_k = x_k + x_{k-1} \tag{3.29}$$

This means that $y_k$ can take only three distinct values. E.g., if $x_k \in \{-1, 1\}$, $y_k$ will be -2, 0 or 2. Therefore, the received signal, i.e. after the demodulation, would have an amplitude of 2 or null and a phase of $\pi$ or 0 rad. Normally, a $\pi$ phase shift takes place between two groups of 1's when the number of 0's in-between is odd.

Duobinary can fit into two different categories: the standard duobinary and phase-shaped binary transmission (PSBT). Although duobinary shows absence of power during logic 0's, PSBT exhibits small bounces of power, as it can be seen in Figure 3.8. The $\pi$ phase shifts, that occur in the middle of these bounces, limit significantly the effect of ISI in PSBT. On the other hand, they also limit the eye diagram opening and consequently this degrades the back-to-back OSNR sensitivity of PSBT when compared with NRZ. In the case of duobinary, the $\pi/2$ phase shifts that occur before and after an even number of consecutive 0's, does not limit the impact of ISI, only the $\pi$ phase shifts that occur before and after an odd number of consecutive 0's contribute for the limitation of the impact of ISI. In addition, the similarity of NRZ and duobinary intensity profiles suggests that amplified spontaneous emission (ASE) noise produces identical effects in both cases.

Figure 3.9 shows two possible configurations to generate standard duobinary using the delay-and-add filtering method, and three configurations to produce PSBT using the low-filtering technique. After differential pre-coding, the delay-and-add filtering method is performed in the electrical domain for "electrical" duobinary and it is accomplished in the optical domain for "optical" duobinary. In the optical domain, the filtering requires a constructive port of a Mach-Zehnder 1-bit delayed interferometer (MZDI) and it is employed after a differential phase shift keying transmitter (DPSK). In electrical domain, the filtering consists in adding the signal and its replica delayed of 1-bit duration. After the differential pre-coding, the "electrical" PSBT is performed by the electrical fifth-order Bessel low-pass filter (LPF), whose cut-off frequency is of about 11.2 GHz, 28% of the bit rate (40 Gbps). As in the "optical" duobinary case, the DPSK transmitter is also needed to generate the "optical" PSBT. The filtering is achievable through an optical Gaussian band-pass filter (BPF), whose 3-dB bandwidth is of about 22.4 GHz, 52% of the bit rate. The "optimum" PSBT is a combination

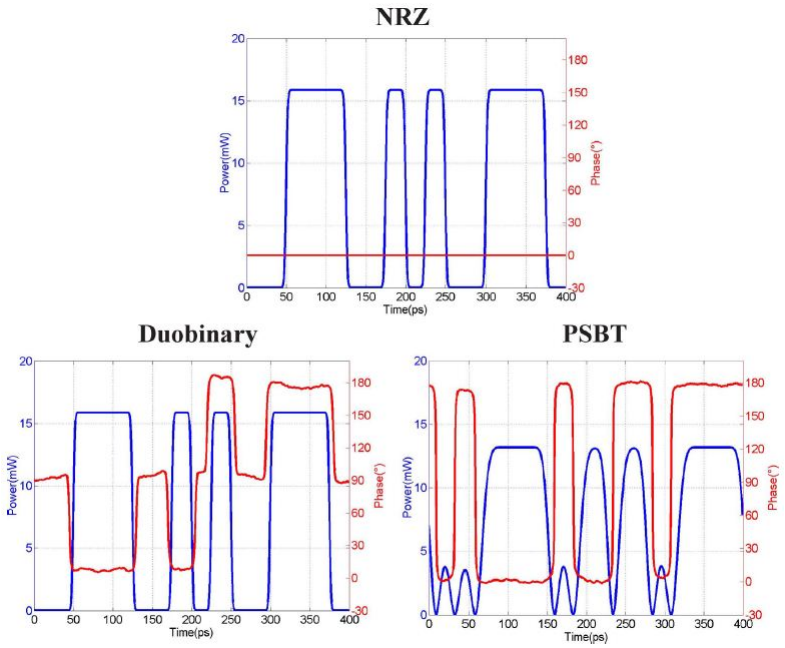



Figure 3.8: Intensity (blue) and phase (red) of NRZ, standard duobinary e PSBT corresponding to sequence '1110010100111'. [18]

---

of the "electrical" and "optical" PSBT. The fifth-order Bessel LPF is employed after the precoder and its cut-off frequency is of about 15.2 GHz. The second-order Gaussian BPF, whose band-pass filter was increased to 32.8 GHz, is implemented after the DPSK transmitter.

The eye diagrams and the PSDs, that are represented in Figure 3.9, allow to take some conclusions about the performance of the different configurations. Particularly, one can see that the eye diagrams corresponding to standard duobinary are very similar to that of NRZ and their PSDs are not very different from that of NRZ also. Comparing all the PSDs, one can verify that the "Optimum" PSBT has the narrowest spectrum.

### 3.2.1 Comparison of Duobinary and Non-Return-to-Zero

Other simulations realized owing *VPI TransmissionMaker*$^{TM}$ by *Tan et al.* show that [18]:

- **Back-to-back OSNR sensitivity** At $BER = 10^{-9}$, "Optimum" PSBT exhibit the best back-to-back OSNR sensitivity. The sensitivity curves of "electrical" and "optical" duobinary are perfectly coincident and identical to the one of NRZ. The worst performances belong to "electrical" and "optical" PSBT due to the energy bounces located in 0's slots.

- **Robustness to residual CD and first-order PMD** The highest robustness to residual CD was shown by "electrical" and "optical" PSBT because of the $\pi$ phase shifts that were mentioned earlier. It was also verified that both standard duobinary and NRZ show very identical resilience to residual CD. In addition, PSBT pulses keep their shape and integrity when a residual CD of 170 ps/nm is applied. Considering the PMD,

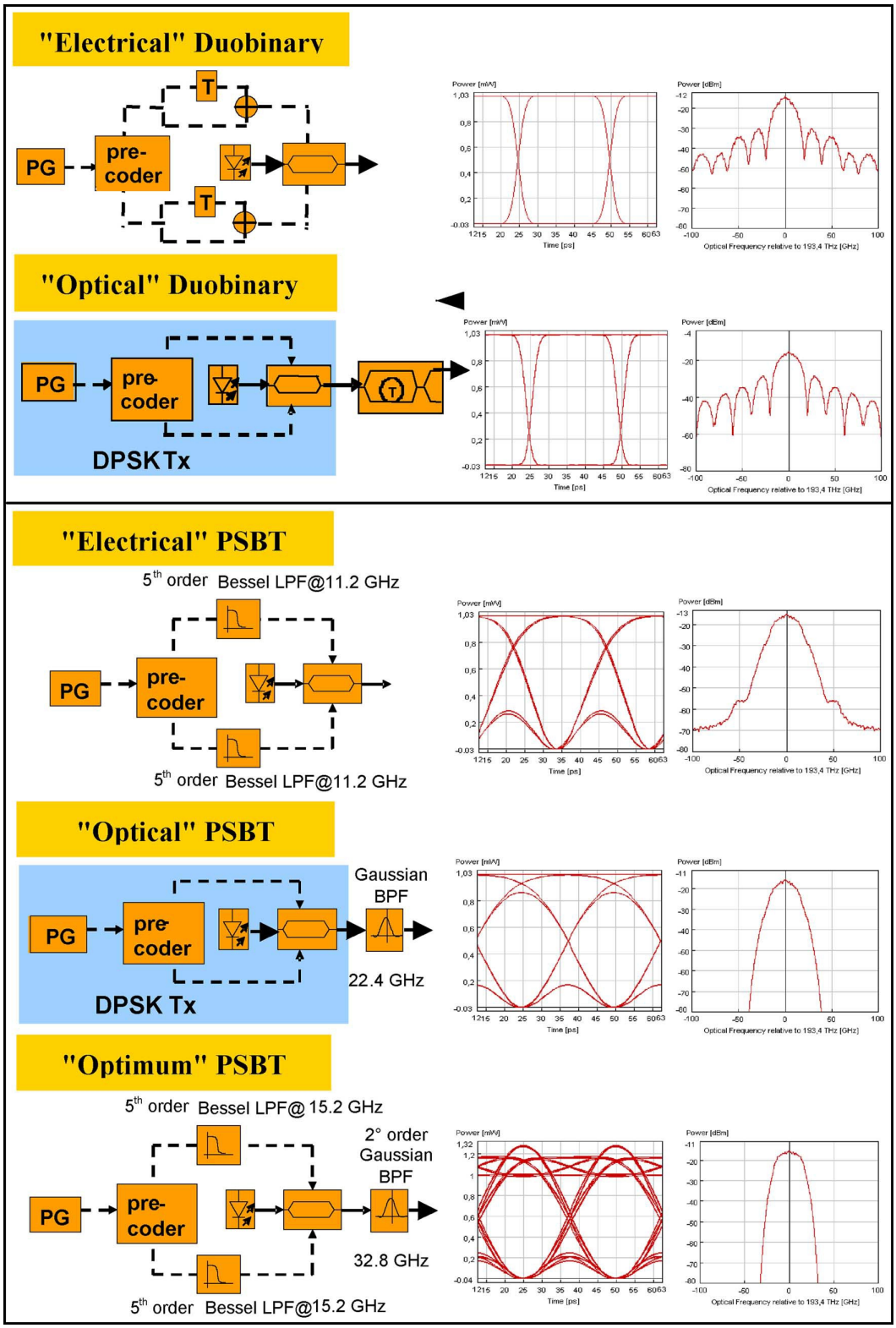



Figure 3.9: Configuration of standard duobinary and PSBT transmitters and their correspondent eye diagram and spectrum. [18]

both duobinary and PSBT presented 1 dB OSNR penalty for a differential group delay (DGD) (which some authors consider to be PMD magnitude) of 6 ps. It shows that duobinary and PSBT have worse robustness to PMD than NRZ (8 ps) due to their narrower spectrum, or equivalently, their wider temporal pulse width. But this behaviour against PMD can be reversed if residual CD is presented.

- **Robustness to Intrachannel Transmission Impairments** The minimum BER of "electrical", "optical" duobinary and "optical" PSBT was slightly better than that of NRZ, what means that they show more resilience against to intrachannel nonlinearities than NRZ. "Optimum" PSBT shown a minimum BER 2 decades better than that of NRZ, which means the first has more tolerance to intrachannel nonlinearities than the last. In spite of its worse BER, "electrical" PSBT shows better robustness to intrachannel transmission impairments than NRZ.

- **Robustness to WDM Transmission Impairments** Because of their narrow spectrum, all PSBT formats are considerably less affected than NRZ and both standard duobinary formats by the reduction of the channel spacing from 100 to 50 GHz. The BER of NRZ suffered a decrease of 6 decades, while the BER of the both "electrical" and "optical" duobinary suffered a degradation of 3.5 dB. In opposition of what one could suppose, the interchannel nonlinearities is not the main reason of the degradation of the PSBT performance. The "Optical" PSBT was the one of the three PSBT formats whose transmission quality suffered the more, owing the presence of multiplexers and de-multiplexers, which distort its optimal optical filtering function rather than intercahnnel nonlinearities.

## 3.3 Nyquist Shaping

In spite of the high bit rates that are achieved with duobinary employment, Tbps super channels require even more advanced pulse shapes. In fact, these super channels are likely to emerge within this decade, allowing, e. g., foreseeable 3D media revolution and new cloud services [50].

Digital signal processing together with finite impulse response (FIR) filters and digital-to-analogue converters (DACs) enable the generation of a variety of pulse shapes, generally known as Nyquist pulses. One of them is the root-raised cosine (RRC). This particular pulse shape enables a spectral efficiency improvement and a better out-of-band attenuation when compared with the pulse shapes that were mentioned in the previous sections. Other pulse shapes can be generated using the same technique to improve other aspects such as the peak-to-average power ratio (PAPR) or the sensitivity to timing jitter.

Zero ISI is imposed in Nyquist shaping, whereas in other cases like duobinary, ISI is introduced in a controlled manner as it was exposed in Section 3.2. Therefore, the Nyquist's first criterion stipulates that the contributions of the adjacent received waveforms at the sampling instants $T, 2T, 3T, ..., nT$ must be null. In time domain, this criterion results in [19]

$$p\left(kT\right) = \begin{cases} 1, & k = 0 \\ 0, & k \neq 0 \end{cases} \tag{3.30}$$

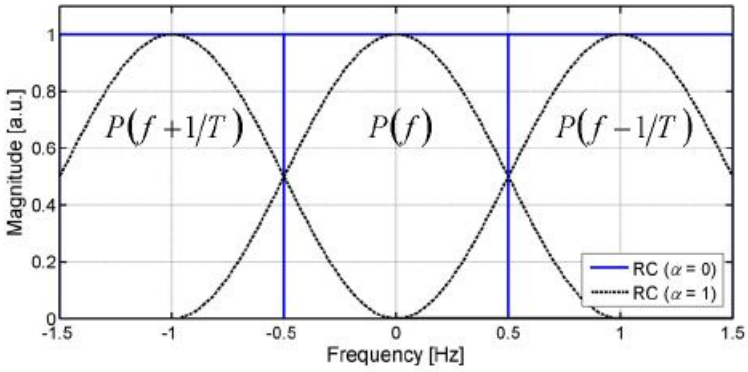



Figure 3.10: Nyquist pulse shaping power spectral density for $\alpha = 0$ (solid line) and $\alpha = 1$ (dashed line). [19]

---

where $k \in \mathbb{Z}$ and $T$ is the symbol interval. In the frequency domain, the impulse response of a Nyquist pulse shaping filter is expressed by [19]

$$\sum_{k=-\infty}^{+\infty} = P\left(f - k/T\right) = T \tag{3.31}$$

which means that the sum of the delayed pulses spectra by $k/T, k \in \mathbb{Z}$ must be $T$, or in a general way, a constant value.

Nyquist also demonstrated that $2W$ independent symbols per second can be transmitted through a channel whose bandwidth is $W$. This together with Nyquist's first criterion, results in the Nyquist pulse without excess bandwidth, i. e., with a roll-off factor $\alpha$ of 0 [19]

$$P\left(f\right) = \begin{cases} T, & |f| < 1/\left(2T\right) \\ 0, & |f| \geq 1/\left(2T\right) \end{cases} \tag{3.32}$$

Figure 3.10 shows that Nyquist pulses with an excess bandwidth of 100%, i. e., $\alpha = 1$ , also satisfies the Nyquist's first criterion. However, the bandwidth in this case is twice the minimum-bandwidth Nyquist pulse, which corresponds to $\alpha = 0$.

Instead of using the roll-off factor $\alpha \in [0, 1]$, some authors use the variable $\beta \in [0, r/2]$, where $r = 1/T$ is the symbol rate. In this case, the minimum required transmission bandwidth is interpreted as [20]

$$B = r/2 + \beta \tag{3.33}$$

For $\beta = 0$, the impulse response decay rate is very slow (Figure 3.11(b)), resulting in a high PAPR and a bigger sensitivity to synchronization errors. A faster decay rate is obtained by increasing the roll-off factor. Thus, Equation 3.32 has to be modified in order to comprise

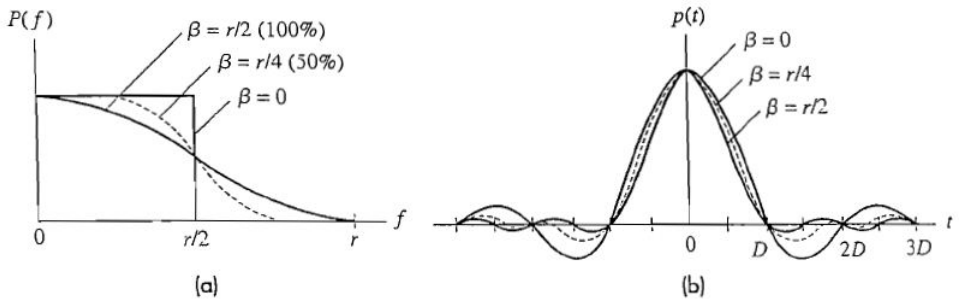



Figure 3.11: Nyquist pulse shaping (a) frequency domain and (b) time domain for $\beta = 0, r/4, r/2$, where $D = T$. [20]

---

all possible roll-off factor values [20]

$$P(f) = \begin{cases} T, & |f| < 1/(2T) - \beta \\ T\cos^2\left[\dfrac{\pi}{4\beta}\left(|f| - \dfrac{1}{2T} + \beta\right)\right], & \dfrac{1}{2T} - \beta \leq |f| < \dfrac{1}{2T} + \beta \\ 0, & |f| \geq \dfrac{1}{2T} + \beta \end{cases} \tag{3.34}$$

In the time domain, this Nyquist pulse is described by [20]

$$p(t) = \frac{\cos(2\pi\beta t)}{1 - (4\beta t)^2}\mathrm{sinc}(rt) \tag{3.35}$$

Figure 3.11 illustrates the Nyquist pulse shaping in both frequency and time domain. As it can be seen, the impulse response decays more rapidly, as well as the trailing oscillations, as $\beta$ increases.

For an excess bandwidth of 100%, or equivalently, $\beta = r/2$, the Nyquist spectrum defined by Equation 3.31 is reduced to the raised cosine spectrum [20]

$$P(f) = T\cos^2\left(\frac{\pi f}{2r}\right) = \frac{T}{2}\left[1 + \cos\left(\frac{\pi f}{r}\right)\right], |f| \leq r \tag{3.36}$$

And the corresponding pulse shape is [20]

$$p(t) = \frac{\mathrm{sinc}(2rt)}{1 - (2rt)^2} \tag{3.37}$$

In this particular case, the half-amplitude width is precisely the symbol interval $T$ and additional zero crossings appear at $t = \pm 1.5T, \pm 2.5T, ....$ Thus, a polar signal constructed with raised cosine pulses will have zero crossings halfway between the pulses centres when the amplitude modifies polarity, as it is shown in Figure 3.12. These zero crossings simplify the clock recovery. However, note the wider bandwidth comparing to that of $\beta = 0$.

Since the impulse response decay rate is important in respect to system's sensitivity to synchronization errors, research was conducted in order to increase it [19]. These works result in an approximation to the root-raised cosine pulse, which is simply defined as the square-root of Equation 3.37.

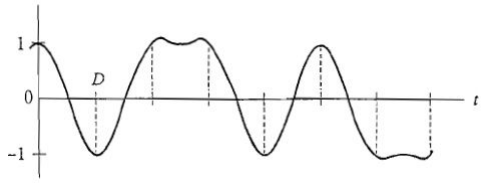



Figure 3.12: Nyquist pulse shaping sequence 10110100 for $\beta = r/2$ [20]

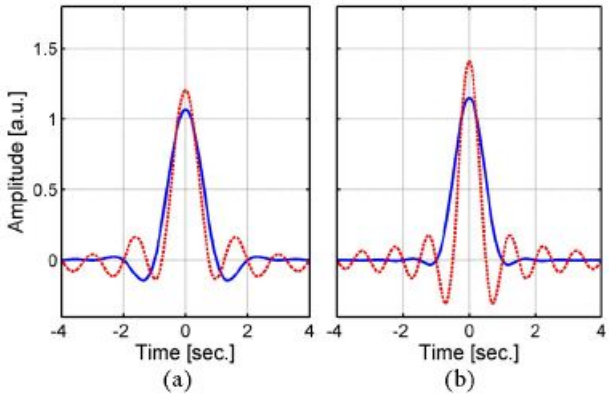



Figure 3.13: Impulse response of root-raised cosine (solid line) and its approximation (dashed line) for excess bandwidth of (a) 50% and (b) 100%. [19]

---

For excess bandwidth between 0% and 100%, this approximation is defined by [19]

$$\sqrt{p\left(t\right)} = \left(1 - \sqrt{1/2}\right)\left(1 - \alpha\right)\operatorname{sinc}\left(\left(1 - \alpha\right)\pi t\right) + \sqrt{1/2}\left(1 + \alpha\right)\operatorname{sinc}\left(\left(1 + \alpha\right)\pi t\right) \tag{3.38}$$

Figure 3.13 shows the desired impulse response decay rate increase, though it is a slightly change relatively to the original root-raised cosine pulse shape.

However, the bandwidth can be the decision factor instead of the impulse decay rate. In that case, the roll-off factor should be null, corresponding to the minimum-bandwidth Nyquist pulse $B = r/2$. For this particular value, the pulse shape is known as the sinc-shaped Nyquist pulse [21].

### 3.3.1 Nyquist Pulses Generation

The exact minimum-bandwidth Nyquist pulse $B = r/2$, where $r$ is the symbol rate, requires a sinc-function infinitely extended in time. However, in practice, Nyquist pulses' length has to be finite in time. This is achieved by using finite impulse response (FIR) filters. Oversampling by a factor of $q$ is employed too in order to separate the baseband spectrum from its periodic repetitions.

In a FIR filter of order $R$, a series of $R$ delay elements $T_s/q$, where $T_s$ represents the symbol interval, are located in-between the $R + 1$ taps. Therefore, if $R = 0$, the "one-tap" filter reproduces the filter input. Note that a too small filter order can deteriorate the signal in a very significant way because of the presence of quantization noise.

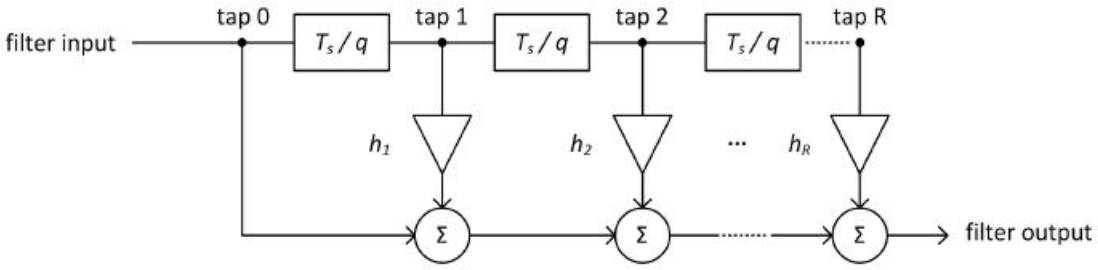



Figure 3.14: FIR filter of order R, where $h_r$ era the filter coefficients. [21]

---

The filter output is the sum result of the input filter and $R$ tapped, weighted signals, as it is shown in Figure 3.14.

The PSDs depicted in Figure 3.15(b)-(d), are the convolution of the sinc-shaped spectrum of the rectangular time window with the rectangular spectrum of an infinitely extended temporal sinc-pulse. In this figure, the FIR filter of order $R$ corresponds to the rectangular time window represented in the left column by the green solid line.

As it can be verified in Figure 3.15(b)-(d), the PSDs evolve towards an ideal rectangle as the filter order increases. In fact, for $R = 1024$, the ideal spectrum is almost achieved. However, strong ringing at the steep spectral slopes occur due to Gibb's phenomenon. In spite of leading to a smoother spectrum and a stronger suppression of side lobes, non-rectangular window functions, such as Hann or Hamming windows, lead also to a wider spectrum, which results in spectral efficiency reduction. This is why rectangular windows are chosen.

### 3.3.2 Comparison of Nyquist Shaping and Non-Return-to-Zero

The principal advantage of Nyquist pulse shaping when comparing with NRZ is the narrower bandwidth, which is half the bandwidth required for NRZ formats, as it is illustrated in Figure 3.15. Consequently, the spectral efficiency in the Nyquist shaping case is higher than that in the NRZ case. In fact, this benefit make the Nyquist pulses a promising suitable solution to be employed in WDM networks.

A particular case is the Nyquist-Wavelength Division Multiplexing (N-WDM), where the Nyquist pulse shape has a near zero roll-off factor, allowing the decrease of the WDM channel frequency spacing from twice the symbol rate for NRZ formats to the symbol rate for Nyquist pulse shaping. E.g., for an optical transmission of 16-QAM formats at 10 Gbps, 5 GHz is the minimum channel spacing required for NRZ line code employment, while 2.5 GHz is the minimum channel spacing required for Nyquist pulse shaping.

Regarding to signal reception, no significant differences are verified comparing both cases. E. g., for advanced modulation formats like M-ary QAM, the Nyquist signal reception is similar to that of NRZ signal. As it will be discussed further, the incoming signal is down-converted to the baseband by a coherent receiver. Next, the signal is sampled by digital-to-analogue converters (DACs), and finally, it is processed in the digital domain.

Furthermore, the clock recovery is essential for receiver synchronization. For NRZ signals, this is usually accomplished by squaring the signals before performing the fast Fourier transform over several received symbols [21]. This procedure enables one to extract the clock signal from the discrete frequency components that are located at the symbol rate. However, this

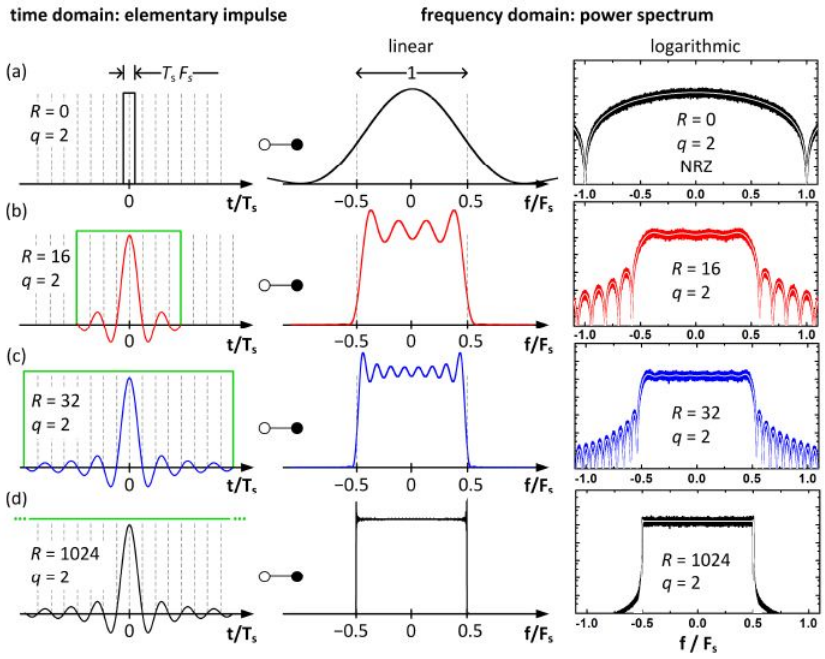



Figure 3.15: Impulse responses (left column) and transfer functions in a linear scale (middle column) and in a logarithm scale (right column) of FIR filters with various orders R. All pulses in these plots have been q-fold oversampled with $q = 2$ [21]

---

technique is not effective in the Nyquist shaping case, once the mentioned peaks disappear. Therefore, alternative algorithms for clock recovery are required [21] [51].

## 3.4 Multi-level Signalling

Optical modulation formats are essential to design flexible and cost-effective WDM fibre networks operating at bit rates beyond 40 Gbps.

The physical characteristic of the CW lightwave which is modified to convey the digital data and the number of symbols considered to represent the binary transmit information are the key features to classify diverse optical modulation formats.

As it was already exposed in Chapter 2, the physical quantities, which can be modified to carry binary data through an optical fibre, are the amplitude, the phase and the polarization. The first two features have been widely studied and employed, while the last one has been discarded itself in some way. One of the reasons not to give much attention to Polarization-Shift Keying (Pol-SK), without combining it with other physical characteristic modulation, is the required active polarization management at the receiver. Another reason is the fact that Pol-SK does not offer the significant receiver sensitivity improvement over intensity modulation that justify the additional receiver complexity that is associated with Pol-SK [1]. Nevertheless, Polarization Division Multiplexing (PDM) has been used in last years to double the system capacity, or equivalently, the spectral efficiency. This is accomplished because the information is conveyed in two orthogonal states of polarization.

On their side, amplitude and/or phase modulation result in a multi-level signalling, in

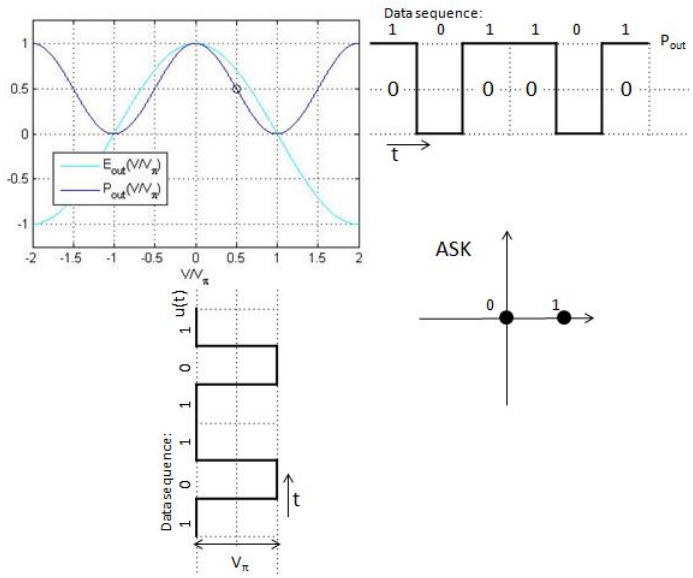



Figure 3.16: Driving the DD-MZM in push-pull mode at the quadrature point to obtain ASK and its corresponding constellation diagram.

---

which $N$ data bits are encoded on $2^N = M$ symbols, whose symbol rate is $r = R/N$, being R the bit rate. Each symbol $b_k$, with $k \in \mathbb{N}_0$, is represented as a complex number in the complex IQ-plane. Note that lower symbol rates allow not only to increase the robustness against dispersive signal distortions as CD and PMD, but also to simplify digital electric signal processing. Moreover, multi-level signalling enables high spectral efficiencies.

On the other hand, the system tolerance to noise is reduced as $M$ grows.

### 3.4.1 Amplitude Shift Keying

Multi-level Amplitude Shift Keying (M-ASK), known as multi-level Intensity Modulation (IM) as well, is the most simple multi-level format, in which the information is encoded in the amplitude of the optical carrier, resulting in different intensity levels.

2-ASK, commonly known as binary On-Off Keying (OOK), is the employed ASK format in commercial optical transmission systems. In fact, multi-level intensity modulations for $M > 2$ have not shown benefits comparing to 2-ASK, mainly due to their considerable back-to-back receiver sensitivity penalty. Thus, for now on, binary ASK will be simply mentioned as ASK, unless otherwise indicated.

Figure 3.16 shows a digram which exemplifies the principle of Amplitude Shift-Keying (ASK) modulation using a DD-MZM. The modulator is working in push-pull mode, so that intensity modulation becomes possible. The peak-to-peak amplitude of the driving voltage is $V_\pi$ and the operating point is the quadrature point. In this case, the output phase is always equal to 0 rad because the minimum transmission point is never crossed.

Putting together this information about the generation of ASK and the information exposed in Subsection 3.1.4 about RZ pulse carving, a RZ-ASK transmitter can be idealized. Therefore, a pulse carver is required to generate RZ pulses followed by a DD-MZM for intensity modulation, as it is illustrated in Figure 3.17. The IS block represents the electrical

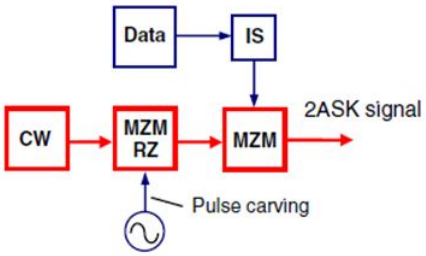



Figure 3.17: Simplified RZ-ASK transmitter scheme. (Adapted from [10])

---

impulse shaper because the electrical driving fields have to be overshoot free with specific rise times. These features are accomplished by filtering a rectangular input time function with a non-causal linear time invariant filter with the specific Gaussian shaped impulse response [10].

However, a lack of interest in OOK has been verified in the last years because its spectral efficiency is not enough to reach bit rates as high as 40 Gbps and beyond. In fact, the OOK spectral efficiency does not exceed 1 bit/s/Hz per polarization, once only one bit is transmitted per symbol.

Additionally, OOK formats show lower resilience to transmission impairments such as CD and PMD because, for a fixed bit rate, nonbinary modulation formats employ a lower symbol rate than OOK (or any binary modulation format), which reduces the signal bandwidth and consequently the pulse spreading caused by chromatic dispersion. Another reason is that nonbinary modulation formats show longer symbol's duration than OOK formats. Therefore, the first can often tolerate greater pulse spreading caused by CD and PMD [22].

Figure 3.18 shows the relation between the spectral efficiency and SNR requirement as the number of constellation points increases, for a BER of $10^{-9}$. From it, one can conclude that OOK (denoted as Pulse Amplitude Modulation (PAM) in the figure) exhibits the largest SNR requirement for a given spectral efficiency. An additional conclusion is the decrease of the system tolerance to noise (higher SNR requirement) as the number of bits per symbol increases, independently of the modulation format used.

### 3.4.2 Phase Shift Keying

M-Phase Shift Keying (M-PSK) refers to modulation formats in which the phase is the only optical carrier's physical feature that is modified in order to convey the information, being $M$ the number of constellation symbols.

In this subsection only Binary- and Quaternary-Phase Shift Keying (BPSK and QPSK) will be addressed, because these are the M-PSK formats studied in the context of this thesis. Actually, for $M > 4$, the employment of M-QAM formats is more suitable than that of M-PSK, since the former show lower SNR requirements than the latter, as Figure 3.18 illustrates.

#### Binary-Phase Shift Keying

Once BPSK symbols encode only one bit per symbol, they do not show advantages in terms of spectral efficiency when compared with OOK formats. However, a 3 dB sensitivity advantage over OOK can be exploited.

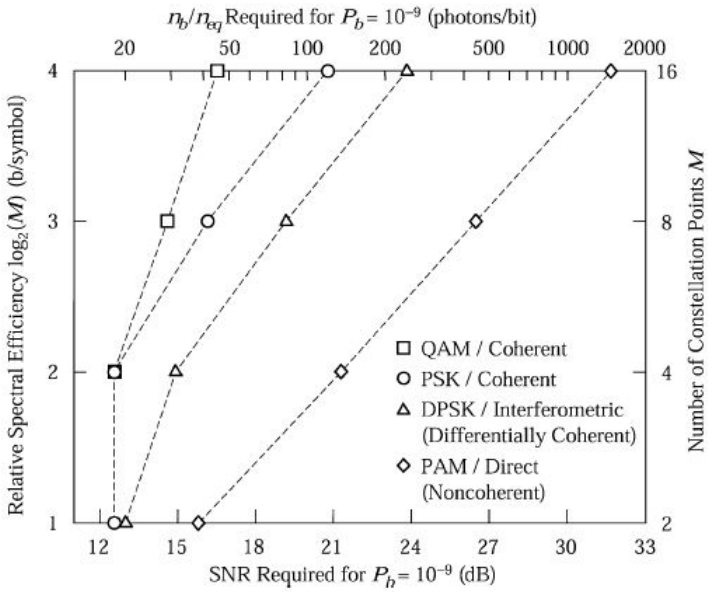



Figure 3.18: Spectral efficiency versus SNR requirement for various modulation formats and detection techniques for a BER of $10^{-9}$. Only one polarization is assumed. [22]

---

Figure 3.19 shows the OOK and BPSK's constellations, which allows to understand the reason of the mentioned 3 dB sensitivity advantage. As it can be seen, the distance between BPSK symbols, which are expressed in terms of optical field, is increased by $\sqrt{2}$ comparing to the distance between OOK symbols, for the same average power. Note that the power average in the case of OOK is [23]

$$P_{avg} \approx \frac{1}{2}x^2 + \frac{1}{2}0^2 = \frac{1}{2}x^2 \tag{3.39}$$

and for BPSK [23]

$$P_{avg} \approx \frac{1}{2}\left(\frac{x}{\sqrt{2}}\right)^2 + \frac{1}{2}\left(\frac{x}{\sqrt{2}}\right)^2 = \frac{1}{2}x^2 \tag{3.40}$$

This means that the average optical power needed in BPSK is half of that required in OOK to achieve the same symbols distance. Such sensitivity improvements can be useful to extend transmission's distances, to reduce optical power requirements or to relax component specifications.

Other advantages can be pointed out. Together with balanced detection, BPSK offers large tolerance to power fluctuations in the receiver decision circuit due to the fact that the decision threshold is independent of the input power. Also, BPSK is more resilient than OOK to some transmission nonlinearities because the optical power is more uniformly distributed than in OOK formats, which results in a decrease of bit-pattern-dependent nonlinear effects.

Since amplitude and polarization parameters are unchanged, two types of external modulators can be used for modulation: the phase modulator and the Mach-Zehnder modulator.

Although phase changes occur relatively fast, they are not instantaneous. Therefore, the PM will introduce frequency chirping across bit transitions. In addition, any imperfections in

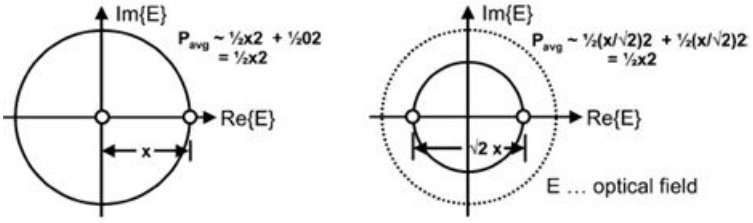



Figure 3.19: Binary OOK (left) and binary PSK (right) constellations. [a] [23].

[a]The original context concerns to DPSK, which has the same constellation of BPSK.

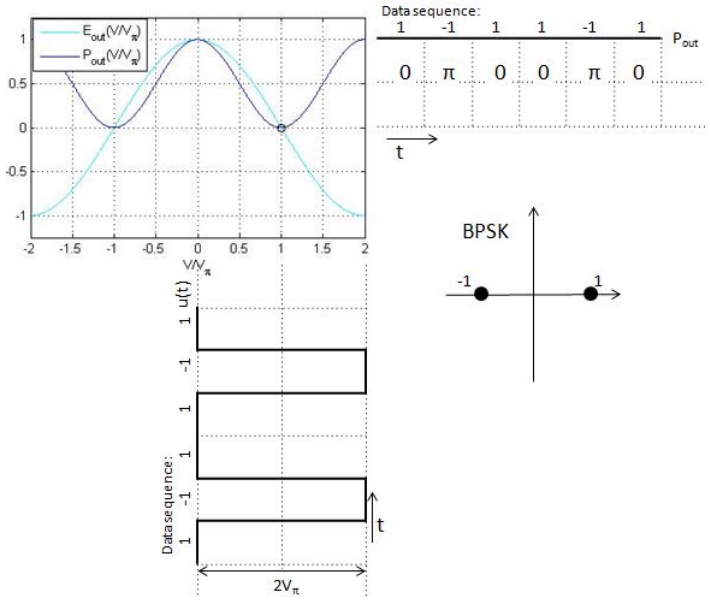



Figure 3.20: Driving the DD-MZM in the push-pull mode at the minimum transmission point to obtain BPSK and its corresponding constellation diagram.

---

the driving voltages get directly mapped onto the optical phase of the optical carrier, which can degrade the PM performance.

In opposition to what happens in the PM's case, the verified drive-waveform imperfections do not interfere with the information, in the MZM's case, since they are only translated in amplitude changes. Additionally, the nonlinear transfer function of the MZM mitigates the effects of drive-waveform overshoots or of limited drive-signal rise time.

An example of generation of Binary Phase Shift Keying is illustrated in Figure 3.20. The DD-MZM is operating in push-pull mode. However, both the operating point and peak-to-peak amplitude of the driving voltage have to be different of that of ASK formats' generation. In order to perform the phase shift of $\pi$, the modulator operates in the minimum transmission point with a peak-to-peak voltage of $2V_\pi$. This way, a phase skip of $\pi$ is obtained when crossing the minimum transmission point and the output power's level is maintained constant.

Independently of the used external modulator, a pulse-carver is required to achieve RZ-

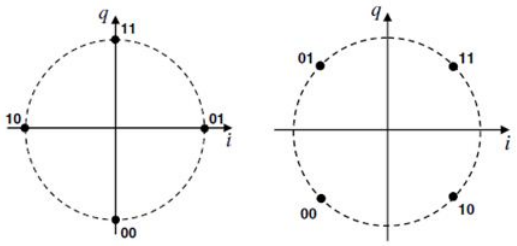



Figure 3.21: QPSK constellation diagram with Gray-coded bit mappings. (Adapted from [10])

---

BPSK formats. However, this device can introduce phase distortions (e.g., chirp), which can cause negative impacts on BPSK receiver.

### Quaternary-Phase Shift Keying

In the previous cases, i.e., 2-ASK and BPSK, only one bit per symbol is transmitted, which means that the symbol rate equals the bit rate and the required bandwidth to guarantee the absence of ISI is twice the bit rate. In the Quadrature Phase Shift Keying's case, two bits per symbol are transmitted. Thus, the symbol rate is half the bit rate and the needed bandwidth to ensure the absence of ISI is the bit rate, what makes QPSK more efficient and more robust against chromatic dispersion [52].

Figure 3.21 shows two possible constellation diagrams for QPSK formats with Gray-coded bit mapping, which mitigates the noise effects. As it can be deduced from the constellation diagrams, a different phase corresponds to each 2-bit combination and the power level is the same, independently of the symbol. This is represented in Table 3.4.

| Symbol | $Phase^{(1)}$ | $Phase^{(2)}$ |
|---|---|---|
| 11 | $\pi/2$ | $\pi/4$ |
| 10 | $\pi$ | $3\pi/4$ |
| 00 | $3\pi/2$ | $5\pi/4$ |
| 01 | $2\pi$ | $7\pi/4$ |

Table 3.4: Phases corresponding to each symbol. $Phase^{(1)}$ and $Phase^{(2)}$ refer to the constellation's points illustrated in the left and in the right sides of the Figure 3.21, respectively.

Any point of both constellation diagrams that are illustrated in Figure 3.21 is achieved using the IQM exposed in Section 2.5, because neither PM nor MZM itself provide both I and Q components of the QPSK symbols.

In spite of the reduction that is observed in the needed bandwidth to transmit the information without ISI, the QPSK format has its drawbacks. One of them is the increase of the OSNR requirements to guarantee the success of the transmission. Another one is the complexity that is verified in the receiver. As it was said, the information is imprinted in the phase of the optical carrier. Therefore, a local oscillator, whose frequency and phase are matched to the frequency and phase of the received QPSK signal, is employed in the receiver. The synchronization between these two signals is essential to recover the data from the optical

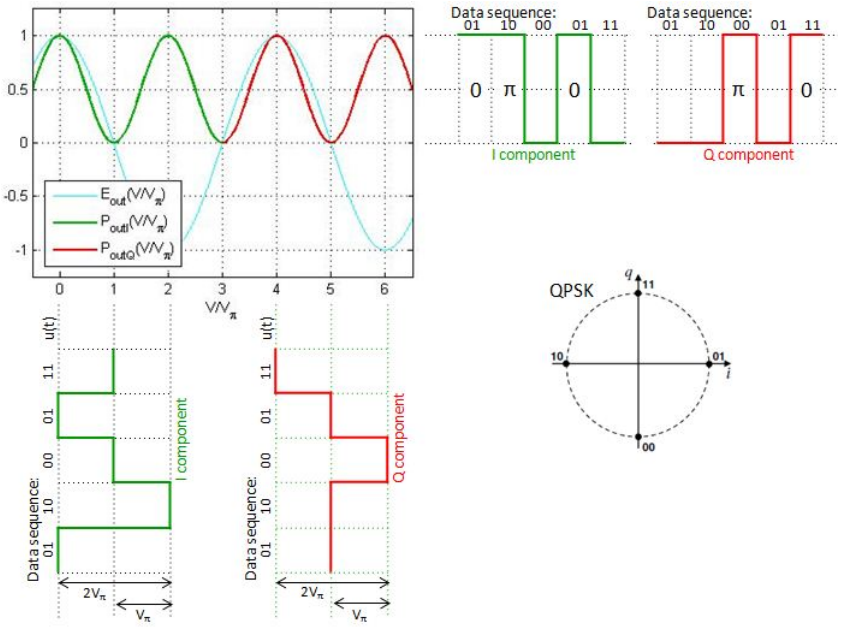



Figure 3.22: Principle of QPSK modulation using an IQM, corresponding to the constellation diagram represent in the left side of Figure 3.21

---

carrier but it is hard to achieve at optical frequencies. The resulting complexity can make the QPSK formats less attractive than other formats such as differential quadrature phase shift keying (DQPSK), since the local oscillator is discarded in these cases. On the other hand, the differential formats require pre-coding, i.e., an additional block has to be added to the transmitter [23] [53].

Figures 3.22 and 3.23 show the principle of QPSK modulation according to the constellations represented in the left and right sides of Figure 3.21, respectively. In both cases, all DD-MZMs are driven at the minimum transmission point. However, the peak-to-peak amplitude of the driving signal is different: in the first case, it is $2V_\pi$ and in the second, it is $1.28V_\pi$. Another difference that can be verified through an analysis of Figures 3.22 and 3.23 is the number of possible values of the I and Q components. In the case in which the symbols overlap the I and Q axes, the amplitude of each component can be null or unitary, while, in the other case, the amplitude of each component is always $\sqrt{2}/2$.

As in any modulation format, the RZ line code can also be employed in QPSK transmitters. It only requires a pulse carver. Figure 3.24 illustrates one example of such transmitter with one particularity: it uses both polarizations to transmit the QPSK signal. Nevertheless, all information that was exposed in Subsection 3.1.4 remains valid.

### 3.4.3 Quadrature Amplitude Modulation

Although M-Quadrature Amplitude Modulation (QAM), particularly Square 16-QAM, has been studied in RF communications for decades, this format was only considered a candidate for future high-capacity and high spectrally efficient optical systems in 2006 [25]. Since then, the increasing interest for this modulation format has been justified by the research

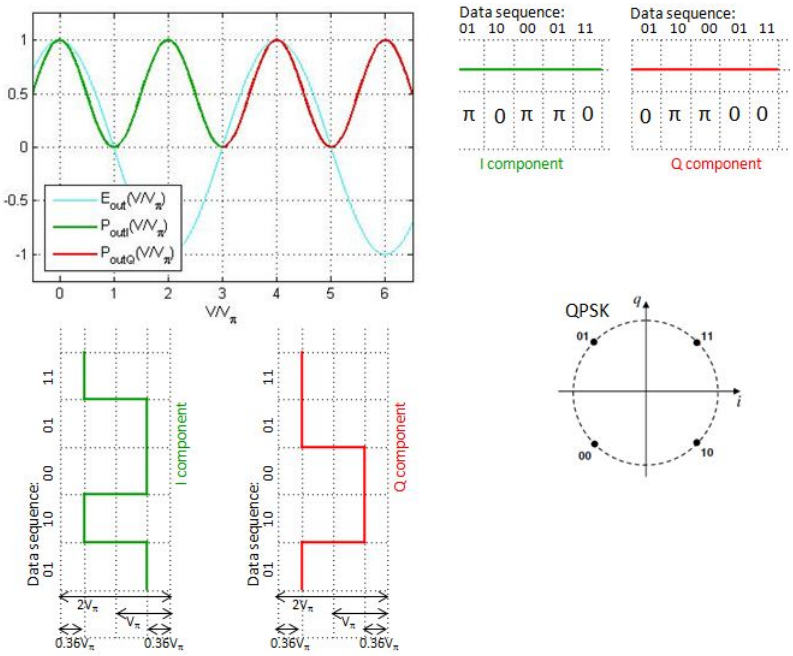



Figure 3.23: Principle of QPSK modulation using an IQM, corresponding to the constellation diagram represent in the right side of Figure 3.21

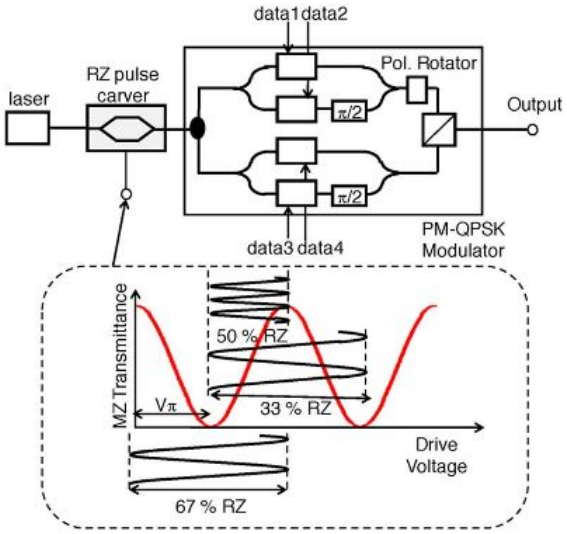



Figure 3.24: Schematic diagram of PM-RZ-QPSK transmitter. [24]

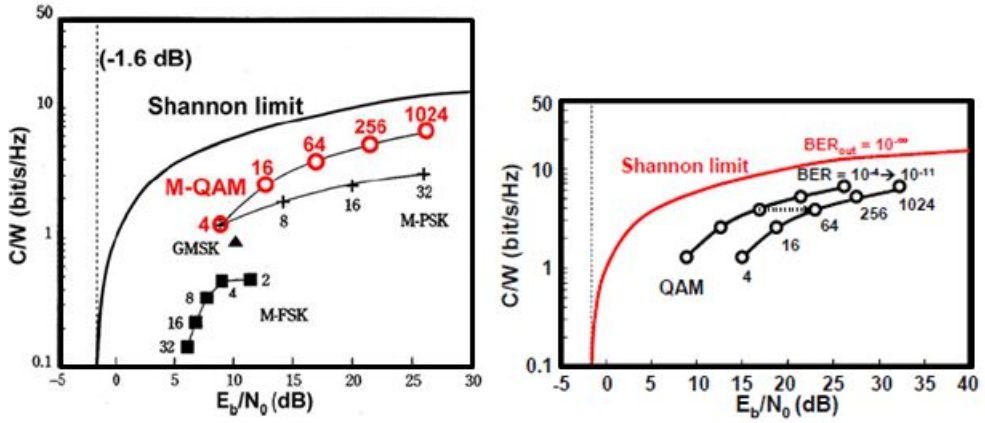



Figure 3.25: Spectral efficiency of (left) M-QAM, M-PSK formats and the Shannon limit at $BER = 10^{-4}$ and (right) M-QAM formats and the Shannon limit at $BER = 10^{-4}$ and $BER = 10^{-11}$. [25] [26]

---

developments in coherent optical systems together with digital signal processing.

Despite giving special emphasis to the 16-QAM formats in this subsection, other QAM formats with a larger number of symbols, such as 128, have already been reported, with a spectral efficiency of 10 bit/s/Hz (for this case), which is relatively high when compared with the spectral efficiency of 16-QAM, 4 bit/s/Hz [54]. However, only 16-QAM will be addressed in the following, since it was the studied M-QAM format in the performed simulations and laboratory experiments.

In 16-QAM formats, 4 bits are encoded in each symbol. Thus, the minimum channel spacing that can be applied in an UDWDM scenario is four times lower than that used if the chosen modulation format was OOK, which allows to increase significantly the number of channels. The same reduction would occur if the 16-PSK was employed instead of 16-QAM. However, the SNR requirement would be greater as well, as Figure 3.18 shows. That is why the Quadrature Amplitude Modulation is chosen over Phase Shift Keying for $M > 4$.

Moreover, the spectral efficiency that is shown by M-QAM formats approaches the Shannon limit closer than QPSK, as it is shown in Figure 3.25 (left) illustrates. This limit is defined as [25]

$$\frac{C}{W} = log_2\left(1 + \frac{E_b}{N_0}\frac{C}{W}\right) \tag{3.41}$$

For both cases, the approach to the Shannon limit implies an increase of the SNR per bit, defined as $E_b/N_0$. For the same SNR and M, the spectral efficiency of the M-QAM is larger than that of the M-PSK.

Figure 3.25 (right) represents the spectral efficiency evolution of M-QAM for two distinct values of $BER$, namely, $BER = 10^{-4}$ and $BER = 10^{-11}$.

There are several options to implement the 16-QAM transmitter. One of them is a tandem-QPSK transmitter that is constituted by cascading IQM, MZMs and PMs or cascading two IQMs. In spite of being driven by binary electrical signals, which simplifies the drive electronics, the connection of these discrete external modulators leads to a large insertion loss. Another solution is to integrate several external modulators in parallel structure, such as two

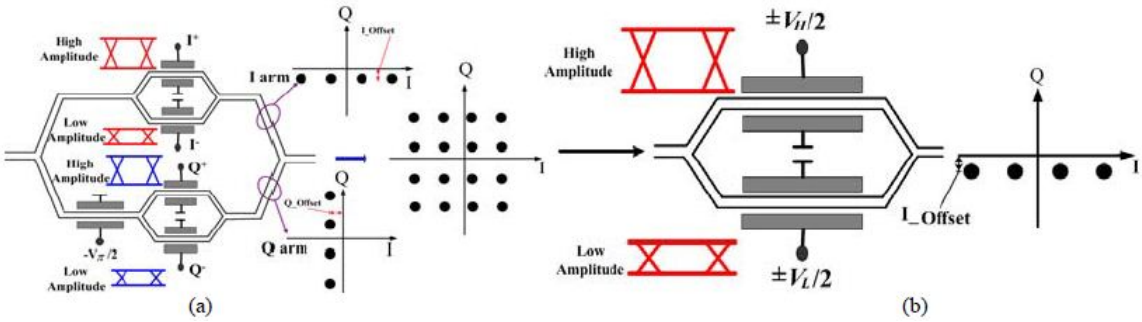



Figure 3.26: (a) Principle of 16-QAM using dual-drive IQM driven by binary signals. (b) four-level amplitude- and phase- shift keying generation with DD-MZM using binary signals with different amplitudes. [27]

---

IQMs or some MZMs. Although it does not suffer from the previous problem, it is hard to manufacture these kind of complex structures. A single-drive IQM, which is simply presented in Section 2.5 as IQ modulator, can also be use to generate 16-QAM formats by driving the modulator with four-level signals. However, these signals are more difficult to obtain and process than binary signals, not to mention that expensive digital-to-analogue converters (DACs)are usually used to generate four-level signals. Another option is the DD-MZM but a look-up table and a high-resolution DAC are needed to generate arbitrary waveforms signals [27]. Finally, the dual-drive IQM can be used to generate 16-QAM formats as well. Comparing to single-drive IQM, the advantage of the dual-drive IQM is that four-level drive signals are not need, although it requires four different drive signals with two different amplitudes, as it is shown in Figure 3.26.

In spite of having the same structure of single-drive IQM presented in Section 2.5, the dual-drive IQM operation is slightly different. Therefore, it will be addressed in the following, instead of that of the single-drive IQM.

A DD-MZM is located in each arm of the DD-IQM. Both DD-MZMs are driven by two binary signals with different amplitudes, which results in two four-level Amplitude- and Phase-Shift Keying (4-APSK) in each branch of the IQM, modulating the I and Q component of the optical carrier. Thus, the two 4-APSK signals are combined at the output Y-junction to generate the Square 16-QAM.

Each DD-MZM is operating in push-pull mode and it is biased at the minimum transmission point. The driving signal of the upper branch has an amplitude of $V_H$, while the lower one is driven by a signal with an amplitude of $V_L$, with $V_H > V_L$. In this way, the four states that are illustrated near the upper and lower arms in Figure 3.26(a) are generated corresponding to the driving voltage levels $\pm V_H/2$ and $\pm V_L/2$. Neglecting insertion loss, the relation between the output and the input electrical field of each DD-MZM of the DD-IQM

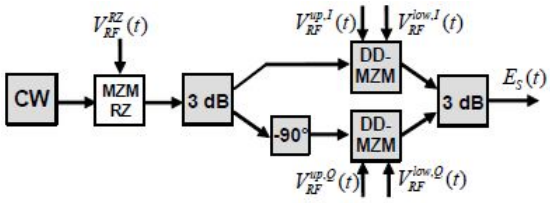



Figure 3.27: 16-QAM generation using an IQM together with a pulse carver. [25]

---

is given by [27]

$$\frac{E_{out}(t)}{E_{in}(t)} = \frac{1}{2}\left(e^{\left(j\frac{\pm V_H\pi}{2V_\pi}+\frac{\pi}{2}\right)} + e^{\left(-j\frac{\pm V_L\pi}{2V_\pi}+\frac{\pi}{2}\right)}\right) \tag{3.42}$$
$$= -\sin\left(\frac{\pm V_H \pm V_L}{4V_\pi/\pi}\right)\cos\left(\frac{\pm V_H \mp V_L}{4V_\pi/\pi}\right) - j\sin\left(\frac{V_H+V_L}{4V_\pi/\pi}\right)\sin\left(\frac{V_H-V_L}{4V_\pi/\pi}\right)$$

The extra phase offset is deduced from Equation 3.42

$$I_{offset} = |Im\left(\frac{E_{out}(t)}{E_{in}(t)}\right)| = \sin\left(\frac{V_H+V_L}{4V_\pi/\pi}\right)\sin\left(\frac{V_H-V_L}{4V_\pi/\pi}\right) \tag{3.43}$$

The 4-APSK signals move away from the axis due to this extra phase offset and, as it can be verified in Equation 3.43, the offset value depends on both the driving voltage and the switching voltage. Additionally, the driving voltages must be adjust in order to satisfy [27]

$$\tan\left(\frac{\pi(V_H+V_L)}{4V_\pi}\right) = 3\tan\left(\frac{\pi(V_H-V_L)}{4V_\pi}\right) \tag{3.44}$$

An offset from the origin of the complex (IQ) plane occurs in the resulting 16-QAM constellation diagram, and consequently, some residual carrier will be present in the generated 16-QAM signals. However, this offset can be eliminated by a normalization process at the receiver digital signal processing (DSP) unit [27].

The RZ line code is also achievable in the QAM formats' case by employing an additional pulse carver, as it is shown in Figure 3.27. Its principle of operation was already explained in Subsection 3.1.4.

# Chapter 4

# Digital Coherent Optical Receivers

The coherent detection technique, together with analogue-to-digital converters and digital signal processing, composes the digital coherent optical receiver.

The digital coherent optical receiver can be seen as an assemblage of four subsystems, which are:

- Optical front end;
- Analogue-to-digital converter;
- Digital demodulator;
- Outer receiver.

The first one is basically the coherent detection itself. Independently of being heterodyne, homodyne or intradyne, the function of the optical front end is to map the optical signal into a set of electrical signals. The resulting electrical signals are then sampled and quantized using an analogue-to-digital converter (ADC). On its side, the digital demodulator holds many functions in order to transform the digital samples generated in the previous subsystem into a set of signals at the symbol rate. Finally, the outer receiver, which includes forward error correction (FEC), produces the best estimation of the original data sequence.

## 4.1 Detection Techniques

The detection method has to be chosen according to the modulation format of the transmitted signal, i.e., the choice is dictated by the physical characteristic of the optical carrier that is modified in order to encode the information.

The employed detection techniques in optical systems are:

- Noncoherent detection;
- Coherent detection;
- Differentially coherent detection;
- Hybrid detection.

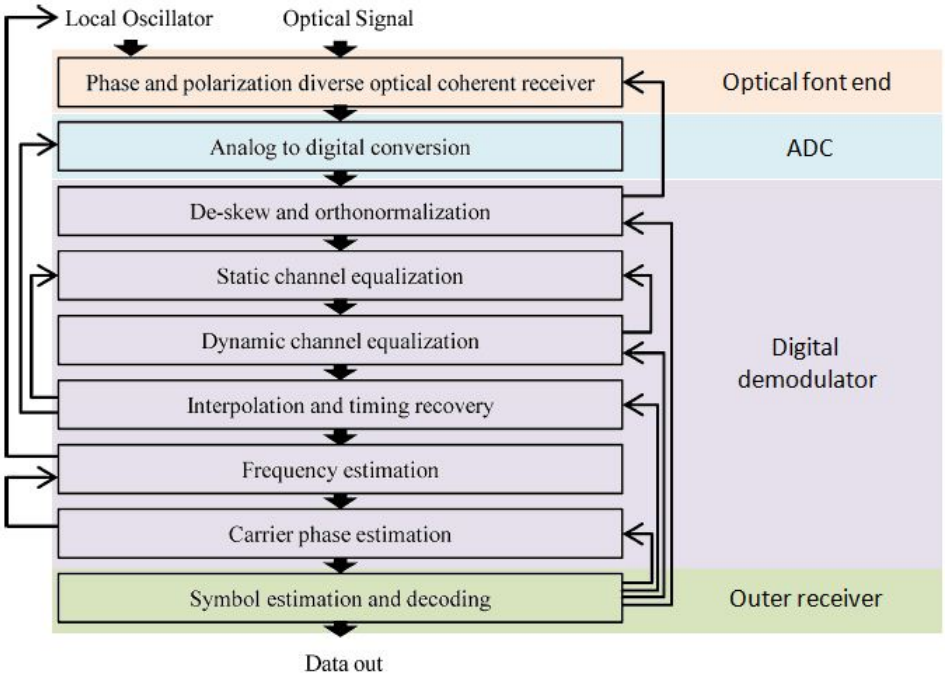



Figure 4.1: Scheme of a general digital coherent optical receiver. (Adapted from [28])

---

Noncoherent detection, known as direct detection as well, is the one that implies the simplest receiver of all. However, it is only applied in system whose modulation format encodes the information in the power of the optical carrier. In this way, another detection method has to be applied in case of high-order modulation formats, where both amplitude and phase of the optical carrier are modified in order to convey the information. The most promising option is the coherent detection, as it will be seen in Subsection 4.1.2. Differentially coherent detection is suited to differential formats M-DPSK, where only the phase is retrieved from the incoming signal. Nevertheless, the hybrid of noncoherent and differentially coherent detection, which requires polarization-shift keying (PolSK) formats, can be use to recover the information from both amplitude and differential phase.

Because M-DPSK and (PolSK) formats are not studied in the context of this thesis, both differentially and hybrid detection will not be addressed here.

### 4.1.1 Noncoherent Detection

Noncoherent detection was widely used in the early days of optical communications because OOK was basically the modulation format employed in optical systems back then, the IM-DD systems mentioned in Subsection 3.1.3. However, in the late 1980s and early 1990s, coherent detection took advantage over noncoherent in order to investigate other modulation formats like DPSK. Then, the interest in OOK formats, and consequently noncoherent detection, rose with the appearance of Erbium doped fibre amplifiers (EDFAs) and its employment in long-haul WDM systems. Nowadays, the research is focusing itself in high-order modulation formats rather than in optical-amplifiers repeaters for increasing the transmission bit rate, which contributes for the decreasing of interest in noncoherent detection.

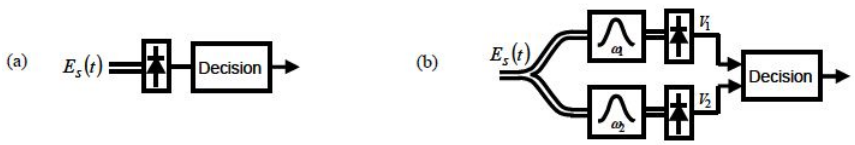



Figure 4.2: Illustration of noncoherent detection principle for (a) ASK and (b) BFSK. [29]

---

The noncoherent detection, known as direct detection as well, is simply implemented with a photodiode, as it is shown in Figure 4.2. Also, it can be seen that the same principle can be used to recover information from a Binary Frequency Shift Keying (BFSK), although it is necessary to separate the two frequencies, $\omega_1$ and $\omega_2$, which are used to carry the data.

In case of direct detection, only one degree of freedom (DOF) per polarization is available, which reduces the spectral and power efficiency. Another disadvantage is the loss of phase information during the down-conversion process that prevents full equalization of linear channel impairments like CD and PMD by linear filters [29]. The performance is suboptimal comparing to equalization processes that use the full electric field, even when maximum-likelihood sequence detection (MLSD) is employed.

### 4.1.2 Coherent Detection

Unlike direct detection, coherent detection is a synchronous process that requires an additional laser, which is used as local oscillator (LO) in the receiver. In fact, the coherent detection principle relies in the product of the received optical carrier and the LO.

The electrical field of the received signal is defined by [16]

$$E_s(t) = A_s(t)\, e^{(j\omega_s t)} \tag{4.1}$$

where $A_s(t)$ represents the complex amplitude and $\omega_s$ is the angular frequency of the received optical carrier. The power of the optical field of the received signal is given by $P_s = |A_s|^2/2$.

The electrical field of the local oscillator is given by [16]

$$E_{LO}(t) = A_{LO}(t)\, e^{(j\omega_{LO} t)} \tag{4.2}$$

where $A_{LO}(t)$ and $\omega_{LO}$ are the amplitude and the angular frequency of the LO, respectively. The power of the LO oscillator is defined as $P_{LO} = |A_{LO}|^2/2$ and it is always constant.

Figure 4.3 shows a representation of the coherent detection principle where balanced detection is employed. By applying the balanced detection, the DC component is removed and the signal photocurrent is increased. When both signal and LO are copolarized, the resulting electric fields are [16]

$$E_1 = \frac{1}{\sqrt{2}}(E_s + E_{LO}) \tag{4.3}$$

$$E_2 = \frac{1}{\sqrt{2}}(E_s - E_{LO}) \tag{4.4}$$

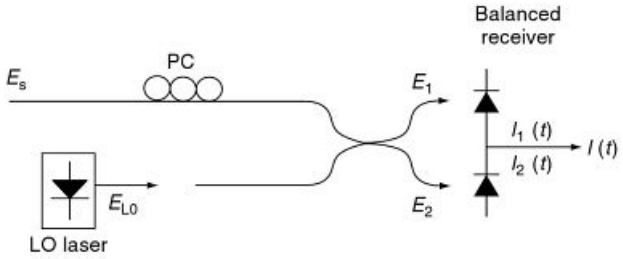



Figure 4.3: Illustration of coherent detection principle. [16]

---

And the resulting photocurrents are [16]

$$I_1(t) = \frac{R}{2}\left[P_s + P_{LO} + 2\sqrt{P_s P_{LO}}\cos\left(\omega_{IF}t + \theta_{sig}(t) - \theta_{LO}(t)\right)\right] \tag{4.5}$$

$$I_2(t) = \frac{R}{2}\left[P_s + P_{LO} - 2\sqrt{P_s P_{LO}}\cos\left(\omega_{IF}t + \theta_{sig}(t) - \theta_{LO}(t)\right)\right] \tag{4.6}$$

where $\omega_{IF}$ represents the intermediate frequency (IF), which is defined as $\omega_{IF} = \omega_s - \omega_{LO}$, $\theta_{sig}$ and $\theta_{LO}$ are the phases of the received signal and the LO, respectively, and $R$ is the responsitivity of the photodetector define by [16]

$$R = \frac{e\eta}{\hbar\omega_s} \tag{4.7}$$

where $e$ is the electron charge, $\eta$ is the quantum efficiency of the photodiode and $\hbar$ represents the Planck's constant divided by $2\pi$.

Subtracting the two mentioned photocurrents, the resulting balanced detector output is [16]

$$I(t) = I_1(t) - I_2(t) = 2R\sqrt{P_s P_{LO}}\cos\left(\omega_{IF}t + \theta_{sig}(t) - \theta_{LO}(t)\right) \tag{4.8}$$

Note that $\theta_{LO}(t)$ includes only the phase noise of the LO that varies with the time and the sum of the frequency component because it will be averaged out to zero due to the limited bandwidth of the photodiode [16].

### 4.1.3 Coherent Detection Advantages

As it was already said, the spectral efficiency has become an essential topic of research due to the increasing demand of high-capacity WDM transmissions. In this context, the coherent detection shows the following advantages [16]:

- The LO provides a signal gain, whereas the LO shot-noise overwhelms the thermal noise of the receiver. In this way, the shot-noise limited receiver sensitivity is achieved with sufficient LO power.

- Closely spaced WDM channels can be separated at the electrical stage because the frequency resolution at the IF or baseband stage is widely large.

Besides these two advantages that make the coherent detection attractive to improve the spectral efficiency, this type of detection shows another benefit comparing to noncoherent reception. Once the phase detection is available in the former, the receiver sensitivity can be improved in relation to ID-MM systems. Note that the distance between symbols, which are expressed as phasors, is extended by the use of the phase information.

### 4.1.4 Heterodyne and Homodyne Coherent Detection

Depending on the value of the IF $\omega_{IF}$, the coherent detection is classified as heterodyne or homodyne. In the former, the IF is established as $\omega_{IF} \gg \omega_b/2$, where $\omega_b$ is the modulation bandwidth of the carrier determined by the bit rate, and in the latter, the IF is null.

Considering the PSK modulation in order to understand the impact of the IF value, the receiver output is described by [16]

$$I(t) = 2R\sqrt{P_s P_{LO}}\cos\left(\omega_{IF}t + \theta_s(t) + \theta_n(t)\right) \tag{4.9}$$

For the heterodyne detection case, where $R$ is the responsitivity, the signal power $P_s$ is constant and the phase of the received signal that is defined as $\theta_{sig}(t) = \theta_s(t) + \theta_{sn}(t)$ is dependent on both phase modulation $\theta_s(t)$ and phase noise $\theta_{sn}(t)$.

In the homodyne detection case, the receiver output is [16]

$$I(t) = 2R\sqrt{P_s P_{LO}}\cos\left(\theta_{sig}(t) - \theta_{LO}(t)\right) \tag{4.10}$$

In this case, an optical phase-locked loop (OPLL) is used in order to extract the emitted signal's phase $\theta_s(t)$ correctly, i.e., to recover the symbol, the local oscillator's phase $\theta_{LO}(t)$ must track the transmitted phase noise $\theta_{sn}(t)$ so that the phase noise defined as $\theta_n(t) = \theta_{sn}(t) - \theta_{LO}(t)$ becomes null. The implementation of this function is complicated and increases the complexity of the homodyne coherent detection.

From Equations 4.9 and 4.10, one can conclude that it is possible to extract the complex amplitude on $\exp(j\omega_{IF}t)$ for heterodyne detection, resulting in [16]

$$I_c(t) = 2R\sqrt{P_s P_{LO}}e^{j(\theta_s(t)+\theta_n(t))} \tag{4.11}$$

While, in homodyne detection, it is not possible to recover all the information of the signal complex amplitude because only the in-phase component with the respect to the LO phase is given by Equation 4.10. Nevertheless, both components can be extracted with an additional local oscillator, implementing the phase-diversity homodyne detection.

### 4.1.5 Phase-Diversity Homodyne Detection

As it was said in the previous subsection, homodyne detection itself does not allow to recover the information from both in-phase and quadrature components. However, applying an additional LO, whose phase is shifted by $\pi/2$, all the information of the signal complex amplitude can be extracted. As such, the phase-diversity homodyne detection is implemented as Figure 4.4 shows.

The required four outputs $E_1, E_2, E_3$ and $E_4$ are generated by a 90° optical hybrid from the two inputs $E_s$ and $E_{LO}$ [16].

$$E_1 = \frac{1}{2}\left(E_s + E_{LO}\right) \tag{4.12}$$

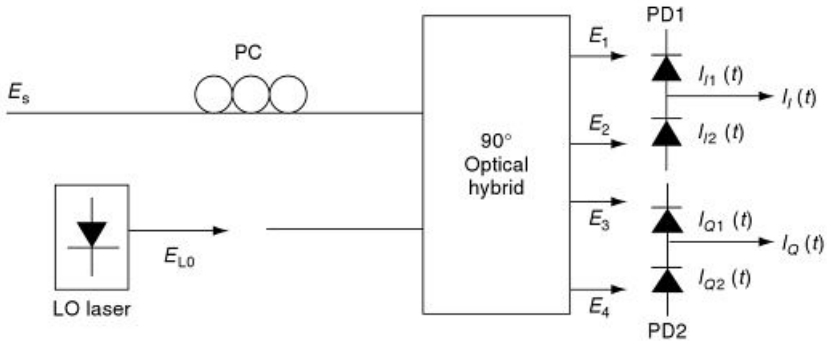



Figure 4.4: Illustration of phase-diversity homodyne detection principle. [16]

---

$$E_2 = \frac{1}{2}\left(E_s - E_{LO}\right) \tag{4.13}$$

$$E_3 = \frac{1}{2}\left(E_s + jE_{LO}\right) \tag{4.14}$$

$$E_4 = \frac{1}{2}\left(E_s - jE_{LO}\right) \tag{4.15}$$

These four 90° optical hybrid outputs fall into the balanced photodetectors, resulting in the following output photocurrents [16]

$$I_I\left(t\right) = I_{I1}\left(t\right) - I_{I2}\left(t\right) = R\sqrt{P_s P_{LO}}\cos\left(\theta_{sig}\left(t\right) - \theta_{LO}\left(t\right)\right) \tag{4.16}$$

$$I_Q\left(t\right) = I_{Q1}\left(t\right) - I_{Q2}\left(t\right) = R\sqrt{P_s P_{LO}}\sin\left(\theta_{sig}\left(t\right) - \theta_{LO}\left(t\right)\right) \tag{4.17}$$

In the case of PSK formats, the photocurrents that are described by Equations 4.16 and 4.17 become [16]

$$I_I\left(t\right) = R\sqrt{P_s P_{LO}}\cos\left(\theta_s\left(t\right) + \theta_n\left(t\right)\right) \tag{4.18}$$

$$I_Q\left(t\right) = R\sqrt{P_s P_{LO}}\sin\left(\theta_s\left(t\right) + \theta_n\left(t\right)\right) \tag{4.19}$$

The $E_s\left(t\right)$ and $\theta_s\left(t\right)$ can be retrieved from $I_I\left(t\right)$ and $I_Q\left(t\right)$ using digital signal processing (DSP), which is implemented in the postprocessing circuit.

Although both heterodyne and phase-diversity homodyne detection show the capability of recovering all information of the signal complex amplitude, the last has the advantage of shifting the incoming signal to the baseband.

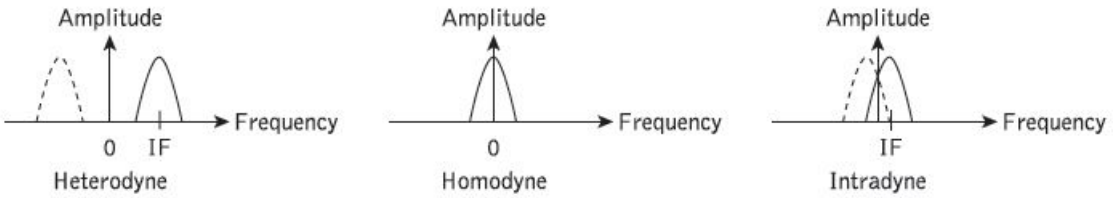



Figure 4.5: Spectral comparison between heterodyne, homodyne and intradyne PSK detection. [30]

### 4.1.6 Intradyne Detection

Intradyne coherent detection is an alternative to both exposed coherent detection techniques. In spite of having a similar detection scheme to that of phase-diversity homodyne detection, the OPLL employment can be avoided in the intradyne detection, since the LO exhibits a frequency near to that of the incoming signal, resulting in a low IF. Therefore, the incoming signal is shifted to a frequency much lower than the data transmission rate, whereas it is shifted to the baseband in homodyne detection, as it is shown in Figure 4.5.

Once the intermediate frequency has a low value, the baseband filter that is implemented in the postprocessing circuit is slightly wider than that used in the homodyne detection. However, the requirements on the automatic frequency control for LO are identical to those of the heterodyne detection [30].

## 4.2 Analogue-to-Digital Converter

The analogue-to-digital converter (ADC) is used to convert the analogue received signal into a set of digital signals. Thus, the ADC can be seen as a combination of a sampler and a quantizer. In the sampler stage, samples of the analogue signal are taken, transforming the continuous-time analogue signal into a discrete-time analogue signal. Next, the quantizer makes the range of the sample amplitude discrete, i. e., the value of the sample is now confined to a finite set of equally spaced values dictated by the bits of resolution of the ADC.

*Flash*, *flash with track and hold* and *time-interleaved* ADCs are implementations that allow to realize high-speed ADCs. Flash converters quantize the signal by comparing the sample value with all the thresholds. The number of comparators is $2^N - 1$ in the case of a N-bit full-flash ADC, which results in an exponential increase of current consumption with the number of bits. So, for medium-high resolutions, this solution is not power effective. However, the power consumption can be reduce using interpolation [55]. In order to improve their AC performance, a track and hold (T&H or T/H) circuit is implemented as a front end element in a *flash ADC*. The T&H circuit is constituted by a sampling switch followed by a buffer, being the performance of the circuit highly dependent on the type of the chosen switch [56]. The time-interleaved ADC is defined as *"an ADC that cycles through a set of N sub-ADCs, such that the aggregate sample-rate is N times the sample-rate of the individual sub-ADCs"* [31], where the sub-ADCs tend to be similar. Nowadays, the aggregate sample-rate is of the order of tens of GHz, a value quite larger than that reported in 1980, 2.5 MHz, by *Black Jr. et. all* [57]. In spite of being composed by various sub-ADCs, the time-interleaved ADC is more area efficient than the flash ADC and as it can be implemented with CMOS

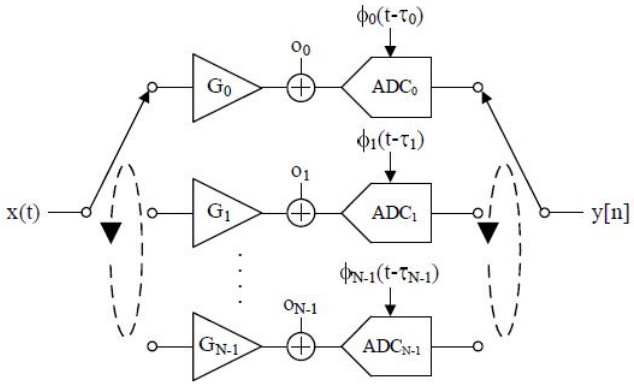



Figure 4.6: Scheme of an non-ideal time-interleaved ADC. [31]

---

technology, it can be integrated with the following DSP blocks in a single application-specific integrated circuit (ASIC). Time-varying errors such as offset, gain and timing skew (delay) are generated in each sub-ADC due to process, voltage and temperature (PVT) variations, layout nonidealities and even current flows, as it is represented in Figure 4.6. However, their impacts are well-known and they can be compensated by applying calibration or numerous algorithms.

## 4.3 Digital Demodulator

As it was already said, the digital demodulator comprises various subsystems in order to compensate an huge range of different fibre transmissions' impairments. Table 4.1 contains the list of their functions and the respective objectives.

| Function | Objective |
|---|---|
| De-skew | Align the digital signal temporally |
| Orthogonalization | Maximize independence between the signals |
| Normalization | Ensure that signals have correct amplitude |
| Digital equalization | Correct for channels impairments |
| Timing recovery | Determine the timing error |
| Interpolation | Correct for timing error |
| Carrier phase estimation | Compensates for the carrier phase error |

Table 4.1: Developed functions in the digital demodulator subsystem. [28]

### 4.3.1 De-skew and Orthogonalization

The path length mismatches between the signals. So, the required synchronization in time of digital signals between each other is accomplished by employing de-skew algorithms after cross correlating the signals in order to measure or estimate the timing delays. Usually, these time delays are not an integer of samples, therefore, some authors represent them as

$\tau = iT_{ADC} + \mu T_{ADC}$, where $i = \lfloor \tau / T_{ADC} \rfloor$ is the basepoint delay and $\mu T_{ADC}$ represents the fractional delay.

Also, the 90° optical hybrid is not ideal. Thus, orthogonalization algorithms are required to restore an angle of $\pi/2$ between the components. Considering the independent zero mean and of unit variance transmitted signals $t_1$ and $t_2$, the corresponding received signals are $r_1 = t_1 \cos(\theta) + t_2 \sin(\theta)$ and $r_2 = t_2 \cos(\theta) + t_1 \sin(\theta)$, respectively. This results in an angle of $\pi/2 - 2\theta$ between the components, rather than the desirable $\pi/2$. The correlation matrix $R$ is then [28]

$$R = \begin{bmatrix} \langle r_1^2 \rangle & \langle r_1 r_2 \rangle \\ \langle r_2 r_1 \rangle & \langle r_2^2 \rangle \end{bmatrix} = \begin{bmatrix} 1 & \sin(2\theta) \\ \sin(2\theta) & 1 \end{bmatrix} \tag{4.20}$$

Gram-Schmidt and Löwdin orthogonalization algorithms are used to perform the orthogonalization between the received signals.

### Gram-Schmidt Orthogonalization

In this algorithm, the first received vector is the reference to create a set of mutually orthogonal signals. Considering two received signals $\mathbf{r} = [r_1, r_2]^T$, the algorithm outputs $\mathbf{g} = [g_1, g_2]^T$ are the product of $G\boldsymbol{r}$, where G is [28]

$$G = \begin{bmatrix} 1 & 0 \\ -\dfrac{\langle r_1 r_2 \rangle}{\langle r_1^2 \rangle} & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -\sin(2\theta) & 1 \end{bmatrix} \tag{4.21}$$

After orthogonalization, the vectors are normalized to the unit power. This way, the second vector is multiplied by $\cos(2\theta)$.

As it can be seen in Figure 4.7, the first vector remains unchanged, while the second one suffers a rotation of $2\theta$.

The Gram-Schmidt orthogonalization has been shown to be effective in coherent systems. However, it increases the quantization noise for the displaced vector. Next, it will be presented an orthogonalization algorithm that distributes the quantization noise equally.

### Löwdin Orthogonalization

The purpose of this algorithm is to generate a set of vectors that are in a least-mean square sense, closest to the original vector [28]. The Löwdin orthogonalization outputs have to be created in such way that $L^2 R = I$, being $L$ the transformation matrix $L = R^{-1/2}$. E. g., in the case of $2 \times 2$ signals [28]

$$L = \begin{bmatrix} \dfrac{\cos(\theta)}{\cos(2\theta)} & -\dfrac{\tan(2\theta)}{2\cos(\theta)} \\ -\dfrac{\tan(2\theta)}{2\cos(\theta)} & \dfrac{\cos(\theta)}{\cos(2\theta)} \end{bmatrix} \tag{4.22}$$

As it can be seen in Figure 4.7, both vectors $\mathbf{l}_1$ and $\mathbf{l}_2$ are equally displaced from the corresponding original vectors by an angle of $\theta$. This symmetric process leads to an equal distribution of the quantization noise for the both vectors.

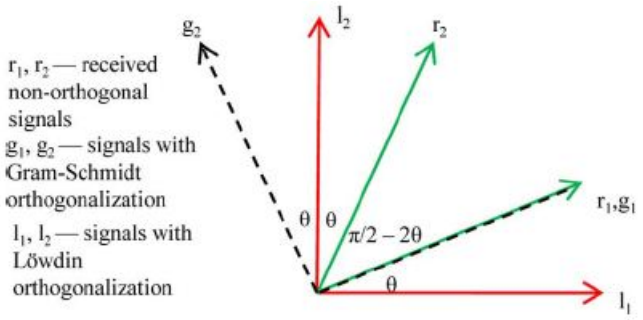



Figure 4.7: Vectorial representation of the Gram-Schmidt and Löwdin orthogonalization algorithms. [28]

---

### 4.3.2 Static- and Dynamic-Channel Equalization

Linear and nonlinear, static and time-varying effects, are compensate by performing equalization, which can be divided into static and dynamic equalization. The first one commonly requires large static filters, while the last one requires a set of relatively short adaptive filters [28].

#### Static-Channel Equalization

Both linear and nonlinear effects can be compensated by applying static-channel equalization.

Considering the received pulse $\mathbf{A}(z,t) = [A_x(z,t), A_y(z,t)]^T$, it can be shown that [28]

$$\mathbf{A}(z=0,t) = \exp\left(-\widehat{D}L_{total}\right)\mathbf{A}(z=L_{total},t) \tag{4.23}$$

Therefore, the original information can be recovered from the dispersed pulse.

In Equation 4.23, $\widehat{D}$ is dependent on the group delay dispersion $\beta_2$, which is approximately $-21ps^2/km$ for SSMF, and it is defined by [28]

$$\widehat{D} = j\frac{\beta_2}{2}\frac{\partial^2}{\partial t^2} \tag{4.24}$$

It can also be proved that a finite impulse filter (FIR) can be used to recover the original data from the received signal by sampling it every $T_{ADC}$ seconds. The total number of the FIR taps is $N = \lfloor|\rho|\rfloor$, with $\rho = 2\pi\beta_2 L_{total}/T_{ADC}^2$. The tap weights are given by [28]

$$h_{cd}[k] = \frac{1}{\sqrt{\rho}}\exp\left(-j\frac{\pi}{\rho}\left[k - \frac{N-1}{2}\right]^2\right), k \in [0, N-1] \tag{4.25}$$

A shorter infinite impulse response (IIR) filter can be used for the same aim. However, the FIR filter is often chosen because it can be implemented efficiently in the frequency domain. Other options, such as the subband equalizer structure, have been studied [28]

For long distances, nonlinearities become relevant. If a constant dispersion is considered and the step size is such that $L\alpha \gg 1$, being $\alpha$ the polarization independent loss, it can be

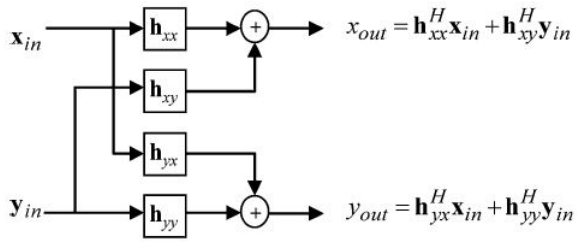



Figure 4.8: MIMO dynamic compensation equalizer. [28]

---

shown that [28]

$$\mathbf{A}(z=0) \approx \exp\left(j\frac{8}{9}\gamma\frac{P_0}{\alpha}\right)\exp\left(-\widehat{D}L\right)\mathbf{A}(z=L) \tag{4.26}$$

In order to recover the original information, Equation 4.26 demonstrates that the linear compensation already exposed (see Equation 4.24) also took place in this situation, followed by an instantaneous nonlinear phase shift.

### Dynamic-Channel Equalization

Figure 4.8 represents a set of four FIR filters, which is employed in order to compensate time varying channel impairments such as PMD. Its function is performing the inverse Jones matrix of the dynamic channel in such way that [28]

$$x_{out}[k] = \mathbf{h}_{xx}^H[k]\,\mathbf{x}_{in}[k] + \mathbf{h}_{xy}^H[k]\,\mathbf{y}_{in}[k] \tag{4.27}$$

$$y_{out}[k] = \mathbf{h}_{yx}^H[k]\,\mathbf{x}_{in}[k] + \mathbf{h}_{yy}^H[k]\,\mathbf{y}_{in}[k] \tag{4.28}$$

Being $N$ the number of the taps of the FIR filter: $\mathbf{h}_{xx}$, $\mathbf{h}_{yx}$, $\mathbf{h}_{xy}$ and $\mathbf{h}_{yy}$ are vectors of length $N$ representing the tap weights and $\mathbf{x}_{in}$ and $\mathbf{y}_{in}$ are the sliding blocks of $N$ samples, such that [28]

$$\mathbf{x}_{in}[k] = [x_{in}[k], x_{in}[k-1], ..., x_{in}[k-N]] \tag{4.29}$$

$$\mathbf{y}_{in}[k] = [y_{in}[k], y_{in}[k-1], ..., y_{in}[k-N]] \tag{4.30}$$

As an first approach, it will be consider that the modulation format is the PDM-QPSK and that the baud rate is such that the channel is considered to be flat and the polarization dependent loss is negligible. In this conditions, the received signal $\mathbf{E}_{rx}[k] = [E_{rx}^x[k], E_{rx}^y[k]]^T = R\mathbf{E}_{tx}[k]$ results in [28]

$$\mathbf{E}_{rx}[k] = e^{j\vartheta[k]}\begin{bmatrix}\cos(\theta) + e^{jn\pi/2}e^{-j\phi}\sin(\theta) \\ -e^{j\phi}\sin(\theta) + e^{jn\pi/2}\cos(\theta)\end{bmatrix} \tag{4.31}$$

From the received signal, one can conclude that the success of the polarization demultiplex is dependent on the determination of $\theta$ and $\phi$. Constant modulus algorithm (CMA), decision directed and radius directed are some of the techniques used to compensate time varying

impairments.

**1) Constant Modulus Algorithm (CMA)**

Initially, the CMA was developed in order to perform blind equalization for QPSK signals, which is a 2-D modulation format. However, this algorithm has been extensively applied to polarization division multiplexing formats, which are considered to be 4-D formats.

The original information of the PMD-QPSK signal is recovered by minimizing the cost functions $\varepsilon_x^2 = \left(1 - |E_{rx}^x|^2\right)^2$ and $\varepsilon_y^2 = \left(1 - |E_{rx}^y|^2\right)^2$. As it can be shown that [28]

$$\varepsilon_x^2 = \varepsilon_y^2 = \sin^2(2\theta)\cos^2\left(\phi - \frac{n\pi}{2}\right) \tag{4.32}$$

Only $\varepsilon_x^2$ is tracked.

Note that depending on the value of $n$, Equation 4.32 results in:

$$\varepsilon_x^2 = \begin{cases} \sin^2(2\theta)\cos^2(\phi), & n \in \{0, 2, 4, ...\} \\ \sin^2(2\theta)\sin^2(\phi), & n \in \{1, 3, 5, ...\} \end{cases} \tag{4.33}$$

And consequently, the control signal is defined by [28]

$$\langle \varepsilon_x^2 \rangle = \frac{1}{2}\sin^2(2\theta) \tag{4.34}$$

Depending on the value of $\theta$, the two polarizations can be divided once [28]

$$E_{rx}^x \propto \begin{cases} E_{tx}^x & \theta = m\pi \\ E_{tx}^y & \theta = \pi/2 + m\pi \end{cases} \tag{4.35}$$

Therefore, an initial estimate of $\theta$ has to be found, as well as a procedure to update its value as time tends to infinite, which is denoted as

$$\theta_\infty = lim_{k\to\infty}\theta[k] \tag{4.36}$$

This procedure has to take into account if $\theta_\infty$ is unique, how sensitive $\theta_\infty$ is to the initial conditions, if the algorithm converges, how fast it converges and if the rate of convergence depends on the initial conditions [28].

The most used algorithm to achieve $\theta_\infty$ is the stochastic gradient descent algorithm. It starts by determining the minima [28]

$$\frac{d\langle \varepsilon_x^2 \rangle}{d\theta} = 0 \tag{4.37}$$

And then, this value is updated following [28]

$$\theta[k+1] = \theta[k] - \mu\frac{d\langle \varepsilon_x^2 \rangle}{d\theta} \tag{4.38}$$

Where $\mu$ dictates the rate of the convergence of the algorithm. Considering that $n$ of Equation 4.33 is equally likely to be even or odd and sufficient data points were taken, the update rule becomes [28]

$$\theta[k+1] = \theta[k] - \mu\sin(4\theta[k]) \tag{4.39}$$

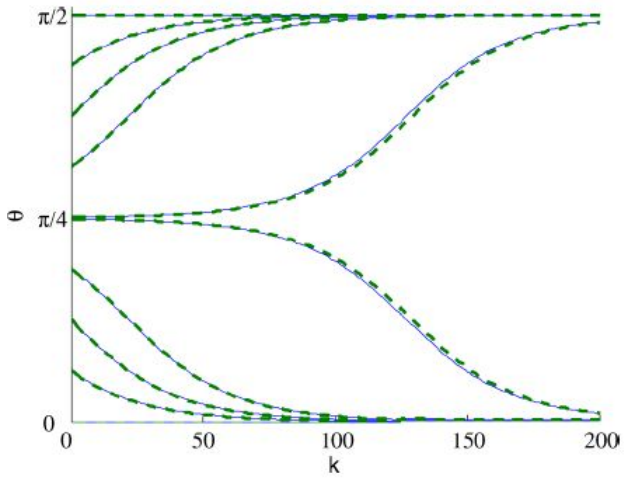



Figure 4.9: Evolution of $\theta$ by using the stochastic gradient descendent algorithm for $\mu = 0.01$. [28]

---

Figure 4.9 illustrates this particular update for $\mu = 0.01$. From it, one conclude that the final solution is very dependent on the initial conditions, namely, near $\theta = \pi/4$. Note that too very close initial estimate near $\pi/4$ leads to two distinct solutions, $\theta = 0$ or $\theta = \pi/2$. Also, it can be verified the time of convergence increases significantly when the bifurcation occurs.

Often, a closed solution is not achievable, which requires an approximation or a linearisation of the equations about an operating point. This linearisation hides the presence of multiple minima that can lead the two outputs lock on to the same source, being strongly affected by the choice of initial tip weights. Therefore, a 2-D control surface was proposed [28].

The resulting control surface $J$ is depicted in Figure 4.10 and defined as [28]

$$\begin{aligned} J = &-\frac{1}{2}\sin^2(2\alpha)\sin^2(2\theta)\cos^2(\phi-\psi) \\ &-\frac{1}{2}\sin(4\alpha)\cos(2\theta)\sin(2\theta)\cos(\phi-\psi) \\ &+\frac{1}{2}\left(1-\cos^2(2\alpha)\cos^2(2\theta)\right) \end{aligned} \tag{4.40}$$

As it can be seen, two optima are well defined, leading to a correct polarization channels de-multiplexing. From Equation 4.40, it can be concluded that the two optima regions are divided by line defined by the maximum of the surface control, which is $J = 1/2$, resulting in [28]

$$\theta = \frac{\pi}{2} - \frac{1}{2}arg(\Theta) \tag{4.41}$$

Where [28],

$$\Theta = -\cos(2\alpha) + j\cos(\phi-\psi)\sin(2\alpha) \tag{4.42}$$

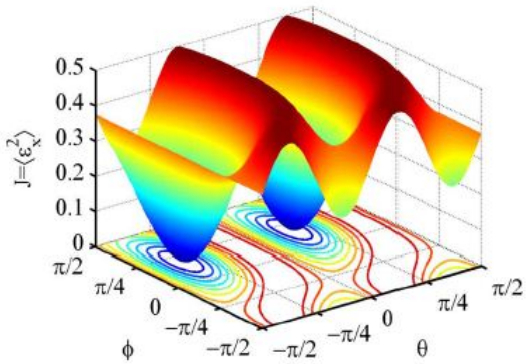



Figure 4.10: Control surface $J$ with $\alpha = \pi/6$ and $\psi = \pi/5$. [28]

---

**2)Decision Direct**

In this technique the equalizer output is feed into the decision circuit, resulting in the following cost function [28]

$$\varepsilon_x^2 = |\frac{csgn\left(E_{eq}^x\right)}{\sqrt{2}} - E_{eq}^x|^2 \tag{4.43}$$

Where [28]

$$\mathsf{csgn}\,(x) = \begin{cases} 1+j & Re\,(x) > 0, Im\,(x) > 0 \\ 1-j & Re\,(x) > 0, Im\,(x) < 0 \\ -1+j & Re\,(x) < 0, Im\,(x) > 0 \\ -1-j & Re\,(x) < 0, Im\,(x) < 0 \end{cases} \tag{4.44}$$

The corresponding control surface is depicted in Figure 4.11. As it can be seen, nonoptimal local minima occur, which can provoke the equalizer to stuck in these points, causing the equalizer to fail in the convergence to the right output. Note that the control surface of CMA algorithms do not exhibit these nonoptimal local minima, so that, the decision directed equalizer (DD-EQ) can be preconditioned by using the CMA. This way, the tap weights become closer to the desired minima (e.g. $-\pi/4 < \theta < \pi/4$) and the DD-EQ will converge to the correct output.

**3) Training Based**

The previously mentioned algorithms perform blind equalization based on the characteristics of the modulation formats without knowing the information that was transmitted. Another option is to use a training sequence to train the equalizer, so that, it learns the transmitted data completely. In this case, the equalizer role is to minimize the function [28]

$$\varepsilon_x^2 = |x - E_{eq}^x|^2 \tag{4.45}$$

Where $x$ represents the original data.

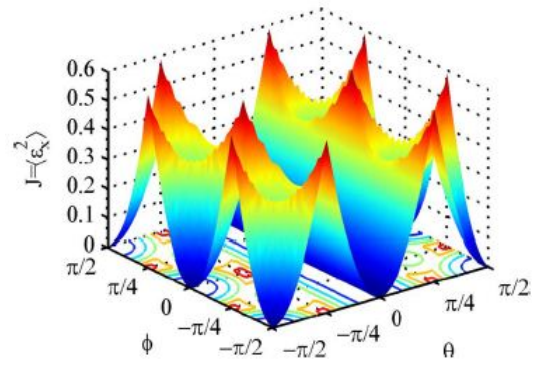



Figure 4.11: Control surface for the decision directed equalizer. [28]

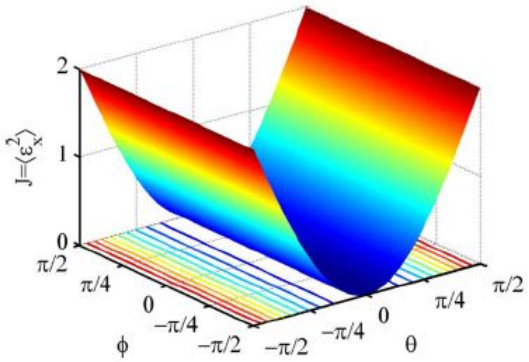



Figure 4.12: Control surface using training sequences. [28]

---

The resulting control surface is illustrated in Figure 4.12. As it can be seen, its shape is much smoother than the previous shown control surfaces, which indicates that the convergence will be better than in the previous cases. However, it requires symbols redundancy in order to train the equalizer.

**4) Control Surfaces for PDM-16-QAM**

Once the PDM-16-QAM does not have constant modulus, it is considered to be constituted by three shells, as Figure 4.13 shows. Even though, the CMA cannot be used by itself because the error term does not tend to zero at the optimum, which causes the equalizer to adapt erroneously in each interaction. Hence, a new surface control has to be found in order to ensure the correct equalizer outputs.

The radially directed equalizer (RDE) overcomes this problem.

The error function is defined by [28]

$$\varepsilon_x^2 = \left(R_0^2 - r^2\right)^2 \tag{4.46}$$

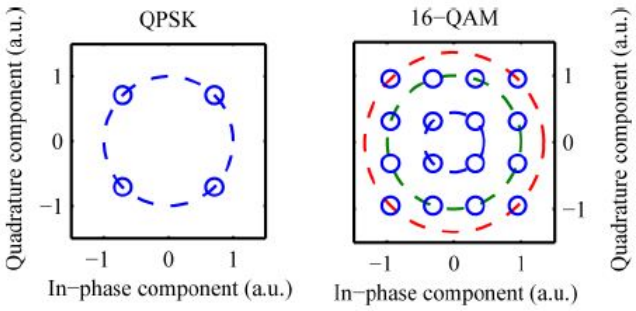



Figure 4.13: Constellation diagrams normalized to unit power for QPSK (with unit circle shown) and 16-QAM (with circles of radius $\sqrt{0.2}$, 1 and $\sqrt{1.8}$ shown). [28]

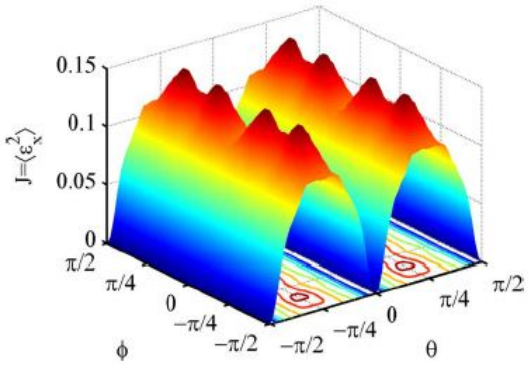



Figure 4.14: Control surface for PDM-16-QAM with RDE. [28]

---

Where $r = |E_{eq}^{x}|$ and $R_0$ is given by [28]

**if** $r^2 < 0.6$ **then**
  point belongs to group 1 and $R_0^2 = 0.2$;
**else**
  **if** $r^2 > 1.4$ **then**
    point belongs to group 3 and $R_0^2 = 1.8$;
  **else**
    point belongs to group 2 and $R_0^2 = 1.0$;
  **end**
**end**

with the signal normalized to unit power on each polarization.

Although the minimum value of the control surface is null, its behaviour is only good over a small region, as it can be seen in Figure 4.14. This way, the RDE can be preconditioned by using the CMA in order to bring the equalizer closer to the region of convergence. Once the CMA has converged, switch to the RDE [28].

Also, the DD-EQ can be employed to perform the equalization of 16-QAM formats. However, CMA or RDE are required to precondition the equalizer.

Once again, using training sequences will result in a smoother control surface, very similar to that of Figure 4.12.

### 4.3.3 Interpolation and Time Recovering

Given the samples $x[i]$ at $t = iT_{ADC} + \mu T_{ADC}$ (see Subsection 4.3.1), the interpolation subsystem has the function of obtaining the samples $y[k]$ at multiples of the symbol period, i.e., $t = t_0 + kT_{sym}$. The procedure is based on interpolation followed by resampling.

First, the continuous-time approximation $y(t)$ is obtained thanks to the interpolation functions $\varphi_i(t)$ [28]

$$y(t) = \sum_i x[i]\,\varphi_i(t - [iT_{ADC} + \mu T_{ADC}]) \tag{4.47}$$

Then, $y(t)$ is resampled at $t = kT_{sym} + \varepsilon T_{sym}$, resulting in [28]

$$y[k] = \sum_i x[i]\,\varphi_i(kT_{sym} + \varepsilon T_{sym} - [iT_{ADC} + \mu T_{ADC}]) \tag{4.48}$$

Where $k$ is an integer and $0 \leq \varepsilon < 1$.

Once $y[k]$ is a linear combination of the inputs, it can be obtained using a FIR filter.

Next, this resulting resampled signal is used for digital time recovering, which is the same as estimating the ratio $w_k = T_{sym}/T_{ADC}$.

Two algorithms can be implemented: non-data-aided and data-aided. The latter will not be considered in the following because the former makes the interaction between the digital demodulator subsystems more simple.

The goal of this algorithm is to maximize the square modulus $|y(m_k, \mu_k)|^2$. Note that $y(m_k, \mu_k)$ is an alternative notation of $y[k]$, where $\mu_k$ represents a fractional delay and $m_k = \lfloor kw_k + \mu_k \rfloor$. The maximization can be achieved by minimizing the error signal defined as [28]

$$e(m_k) = \frac{d|y(m_k, \mu_k)|^2}{dt} \tag{4.49}$$

The signal error is later used to update the $w_k$ value [28]

$$w_{k+1} = w_k + \sum_{i=0}^{N-1} c[i]\,e(m_{k-i}) \tag{4.50}$$

From Equation 4.50, one can conclude that $w_{k+1}$ value results on the error signal being filtered by a FIR filter length $N$ with coefficients $c[i]$.

### 4.3.4 Frequency Estimation

The task of this subsystem is to estimate the carrier frequency deviation $\Delta f$. Note that, even if the laser that is used in the transmitter is also used as the receiver local oscillator, $\Delta f$ would not probably be null because the signal can deviate itself from the original frequency due to nonlinear effects. Figure 4.15 illustrates the frequency and phase offset effects in 16-QAM formats.

In addition, the function of this block not only minimizes the amount of the phase that the carrier recovery subsystem has to track, but also improves the performance of the latter subsystem. In fact, an uncompensated frequency offset results in a higher phase-error variance.

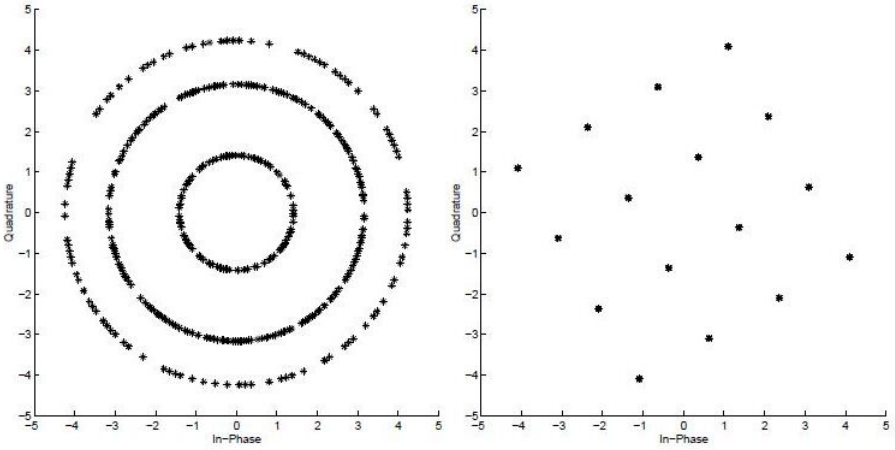



Figure 4.15: 16-QAM constellation diagram with (left) frequency and (right) phase offset. [28]

---

In the following, the input signal will be considered as [28]

$$x_{in}[k] = x_{sym}[k] \exp\left(j\left[\phi[k] + 2\pi\Delta f k T_{sym}\right]\right) \tag{4.51}$$

Both differential phase-based methods and spectral methods are essentially feedforward techniques employed in order to estimate $\Delta f$. However, the former can only be use in QPSK systems, being the latter employed in both QPSK and 16-QAM systems.

The differential phase-based methods are based on the principle that [28]

$$\left(x_{in}[k]\, x_{in}^{*}[k-1]\right)^4 \propto \exp\left(4j\Delta\phi[k]\right) \tag{4.52}$$

And $4\Delta\phi$ has a circular Gaussian distribution, such that [28]

$$f(4\Delta\phi) = \frac{\exp\left(\kappa\cos\left(4\Delta\phi - 8\pi\Delta f T_{sym}\right)\right)}{2\pi I_0(\kappa)} \tag{4.53}$$

Where $\kappa$ is related to the linewidth of the laser.

The probability density function, which is described by Equation 4.53, is used in the maximum likelihood technique so the frequency deviation can be estimate[28]

$$\Delta f = \frac{1}{8\pi T_{sym}} \arg\left\{\sum_{k=1}^{N}\left(x_{in}[k]\, x_{in}^{*}[k-1]\right)^4\right\} \tag{4.54}$$

Also, $\Delta f$ can be estimate using an iterative process described by [28]

$$\Delta f[k] = \Delta f[k-1](1-\mu) + \mu\frac{\arg\left\{\left(x_{in}[k]\, x_{in}^{*}[k-1]\right)^4\right\}}{8\pi T_{sym}} \tag{4.55}$$

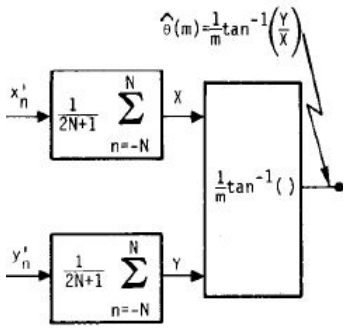



Figure 4.16: Illustration of V&V algorithm. [32]

---

In spectral methods, the frequency shift is determined by observing the peak in the spectrum of $x_{in}^4$ [28]

$$\Delta f = \arg\max_{\Delta f} \sum_{k=1}^{N} x_{in}^4 [k] \exp\left(-8j\pi k \Delta f T_{sym}\right) \tag{4.56}$$

As the number of points may be insufficient to accomplish a correct estimate of $\Delta f$, an iterative process can be employed to improve the estimate value.

Once $\Delta f$ is known, Equation 4.51 is multipled by $exp\left(-2j\pi\Delta f k T_{sym}\right)$, so that, the next step is to determine the carrier phase in order to recover the original symbol $x_{in}[k] = x_{sym}[k]$.

### 4.3.5 Carrier Phase Estimation

Viterbi and Viterbi (V&V) is the most used algorithm in phase carrier estimation. It was idealized for M-PSK modulation formats. However, its application can be extended to 16-QAM formats, considering that this format is constituted by three shells, as it is illustrated in Figure 4.13.

Figure 4.16 represents the steps involved in V&V algorithm.

First, and in opposition to what was considered in the previous subsection, the $n$-th received symbol will be denoted as $x'_n + jy'_n$, being $x'_n$ the in-phase component and $y'_n$ the quadrature component.

Next, the M-PSK modulation has to be removed by raising each symbol component to the M-th power and adding all the resulting products. In the case of QPSK, it results in [32]

$$x = \frac{1}{2N+1} \sum_{n=-N}^{N} {x'_n}^4, y = \frac{1}{2N+1} \sum_{n=-N}^{N} {y'_n}^4 \tag{4.57}$$

For QPSK, the estimate carrier phase is then [32]

$$\widehat{\theta} = \frac{1}{4} \arctan\left(\frac{y}{x}\right) \tag{4.58}$$

Note that $\widehat{\theta}$ corresponds to $\phi[k]$ of Equation 4.51. Since the frequency shift was already cancelled, the original symbol is recovered by multiplying $x_{in}[k]$ by $exp\left(-j\widehat{\theta}\right)$.

## 4.4 Outer Receiver

In the digital demodulator all the transmission impairments were compensated by performing orthogonalization, both static and dynamic channel equalization, interpolation followed by timing recovery and finally, frequency and carrier phase estimation. This way, all conditions are joint so that signal decoding can be accomplished by the outer receiver.

Forward error correction (FEC) is the main function performed by the outer receiver.

The use of FEC codes is only possible if redundancy is introduced in the original data sequence, which is known as FEC-encoding. The original data sequence is grouped in words of $k$ symbols. Then, each word is transformed in another one of $n$ symbols by adding $n-k$ redundant symbols (parity-data). The receiver exploits this redundancy so it can perform two tasks [58]:

- **Error-detection** Verify the association between input- and parity-data;
- **Error-correction** Take actions accordingly.

Unfortunately, some errors can not be corrected. The minimum distance of the code $d_{min}$ is the minimum number of code-symbols that the codewords can differ between each other. Therefore, the code can detect $d_{min}-1$ errors and correct only $(d_{min}-1)/2$.

Additionally, the FEC codes can be divided according distinct categories [58]:

- **Outband vs Inband** In inband FEC, the parity-data is introduced in the superfluous overhead of the client-signal. In case of absence of superfluous overhead, the frame has to be increased in order to add redundancy. This is the so-called outband FEC.
- **Hard vs Soft-Decision** The receiver performs hard-decision, if it only distinguishes between two logic levels corresponding to 1 and 0. On the other hand, if the receiver recognizes intermediate levels, it is said to perform soft-decision. The latter is not common in optical communications due to technological complications related to high transmission rates. Also, the maximum unregenerated spans requires hard-decision receivers with high sensitivity.
- **Block vs Convolutional** Block-codes are constituted by finite- and constant-length codewords. On its side, the convolutional-codes apply over input-data, continuously. Input code-symbols fall into the same codeword until a decision is made.
- **Serial vs Parallel-Concatenation** In order to improve the FEC effectiveness, FEC codes are concatenated. If the input-data is successively encoded by an outer $(h,k)$ code and then by an inner $(n,k)$ code, generating a concatenated $(n,k)$ code-scheme, the concatenation is serial. On the other hand, the concatenation is parallel if the input-data is encoded twice (or more), once in the arrival order and another in the permuted order.

No further details will be explore about FEC codes neither about FEC architectures, given the context of this thesis. However, FEC codes can have an important role in the correct information decoding.

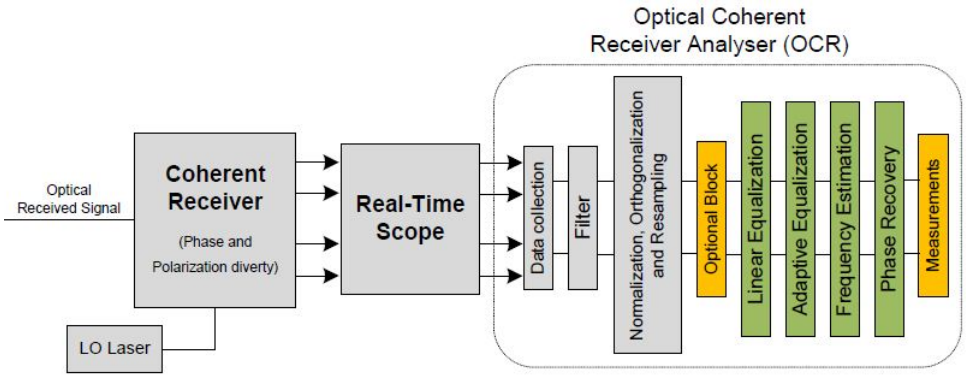



Figure 4.17: *Optical Coherent Receiver* analyser block diagram. (Kindly provided by Ricardo Ferreira)

## 4.5 Optical Coherent Receiver Analyser

The *Optical Coherent Receiver* (OCR) analyser is, as the name indicates itself, a digital signal processing software, whose interface has been developed based on MATLAB® GUI, in Instituto de Telecomunicações, Pólo de Aveiro. It is employed in optical systems that use multi-level signalling, such as QPSK and 16-QAM, which requires coherent detection and can inclusively be transmitted through the same channel with two polarizations. It also evaluates the performance of the transmission system in terms of EVM, BER and eye-diagram characteristics. Another particular feature is the ability of processing information from a source file instead of the oscilloscope outputs.

The DSP that is performed by the OCR analyser allows the compensation of transmission impairments such as CD and PMD. The diagram, which is shown in Figure 4.17, summarizes the implemented subsystems, which are very alike to that exposed previously. This scheme is enabled in the user's interface in such way that, by clicking in its blocks, the user can both adapt the DSP algorithms to the correct modulation format and store the processed data in the most convenient manner.

### 4.5.1 Implemented Digital Signalling Processing Subsystems

First of all, important parameters that are related to both transmitted modulation format and information's source as to be specified. This function is provided by the data collection subsystem. As it was already mentioned, the source information is not restricted to the oscilloscope output. The data can also be stored in a *.mat*, *.out* or *.wfm* file.

Next, the received signal is converted from heterodyne to intradyne by using the Nyquist-Shaping Technique:

- A first order bandpass Gaussian filter is applied to the signal;
- The signal is down-converted;
- A lowpass filter is applied to the down-converted signal.

The low-pass filter can be a Gaussian, Bessel or Butterworth. Also, the user is allowed to chose the order and the cutoff frequencies of both filters, as well as, the central frequency of the bandpass filter.

Note that this subsystem is particularly important in case of multi-channel optical systems, once the distinct channels are filtered here.

At this point, amplitude normalization, interpolation and orthogonalization of the signal is required. The signal is normalized to the unit amplitude before employing the interpolation. Remember from Subsection 4.3.3 that interpolation is needed in order to resample the signal and consequently, to recover the clock. Interpolation with low-pass filter, linear, cubic spline and piecewise cubic hermite interpolation are the options available to the user. The Gram-Schmidt orthogonalization is the algorithm used to restore an angle of $\pi/2$ between the components.

This block is followed by another one, which can optionally be or not used. Basically, this option allows the user to add new code, new functions of digital signal processing. An example is the use of training sequences instead of using blind techniques.

Next, the equalization of the signal is performed. The CD is compensated with linear equalization, while ISI is minimized using dynamic equalization, in particularly, employing the CMA that was addressed in Subsection 4.3.2.

Finally, the frequency and phase of the optical carrier have to be estimated. The frequency estimation is performed using the method, which was described in Subsection 4.3.4. On its side, the phase recovery is accomplished using the Viterbi and Viterbi algorithm, that was already exposed in Subsection 4.3.5.

At the end of this process, EVM, BER, quadrature skew and eye diagram parameters can be determined. Note that the EVM can be calculated in relation to the ideal constellation, the average of the received constellation or the exact transmitted symbols, the last one, being the one that gives the most accurate results. However, in many practical experiments, it is not possible to access the original transmitted data. In addition, the BER calculation requires the knowledge of the transmitted sequence. In this case, both bitwise or cross-correlation can be used to performed the required synchronization between the transmitted data and the received one.

Figure 4.18 shows the main view of the user's interface. As it can be seen, it allows the control of several parameters that are related to the data to be processed.

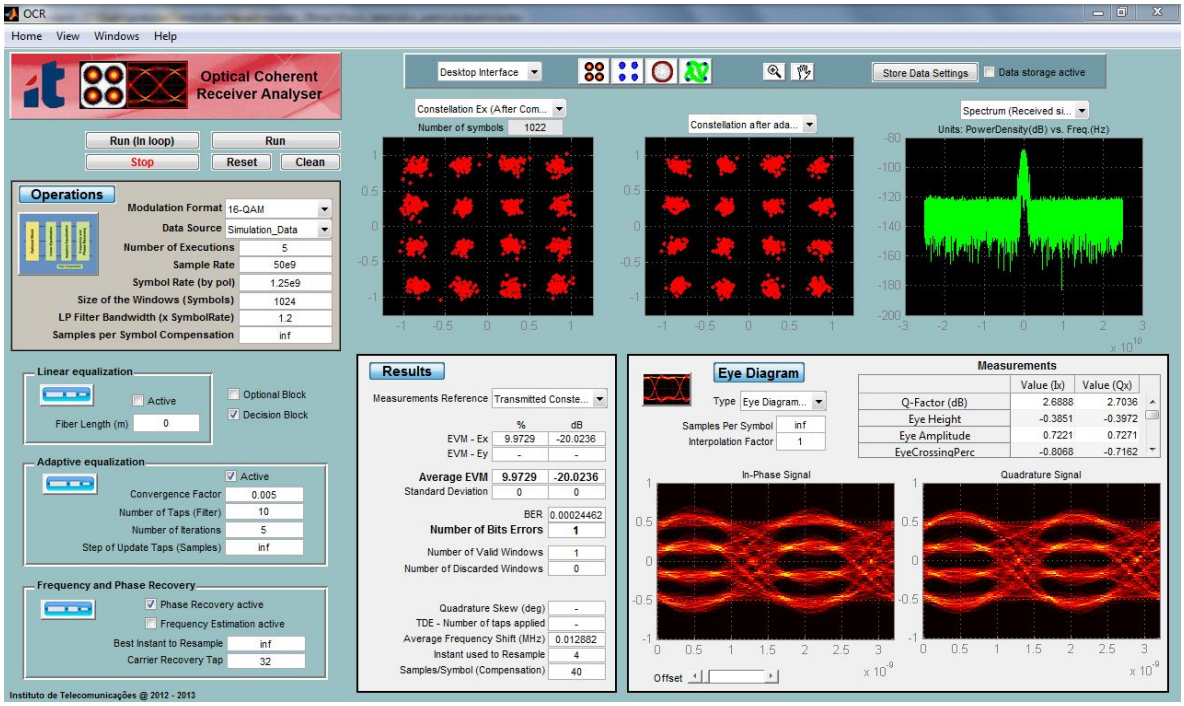



Figure 4.18: *Optical Coherent Receiver* (OCR) analyser user's interface.

# Chapter 5

# Simulation and Laboratory Experiments

From the information that was exposed in the previous chapters, one can conclude that advanced modulation formats together with coherent detection and digital signal processing are the key features of high data rate optical transmissions.

In this way, the simulation and laboratory experiments, that will be described in this chapter, were developed in order to demonstrate the advantages of using high modulation formats, mainly in terms of spectral efficiency. Also, it will be studied the influence of other parameters, which are involved in an UDWDM transmission, such as the transmitted number of channels and the spacing between them. Independently of what is the studied feature, the system performance will be evaluated in terms of error vector magnitude (EVM).

In spite of not being an advanced modulation format, the BPSK will be addressed in the following, so that, the advantages of employing higher formats as, e.g., QPSK and 16-QAM become more clear.

## 5.1 Simulation Results

The simulations were developed in the Optical Coherent Receiver (OCR) addressed in Section 4.5. Note that the available option of simulating optical fibre transmissions was not addressed in the mentioned section because it is inserted in a chapter that is exclusively dedicated to digital coherent optical receivers.

Additionally, one must notice that 4098 symbols were used in each simulation.

### 5.1.1 Simulation Setup

Figure 5.1 illustrates the used setup to simulate the transmission of all the following modulation formats: NRZ-BPSK, NRZ-QPSK, NRZ-16-QAM, Nyquist shaped BPSK, Nyquist shaped QPSK and Nyquist shaped 16-QAM. Note that the last three formats were generated with a zero roll-off factor (see Section 3.3).

The information, which is generated by the bit pattern generator (BPG), is encoded in the phase, in the case of BPSK and QPSK, and in both the phase and amplitude, in the case of the 16-QAM, of the optical carrier that is provided by each one of the continuous wave (CW) lasers that are wavelength multiplexed (WDM), using an IQ modulator as it was

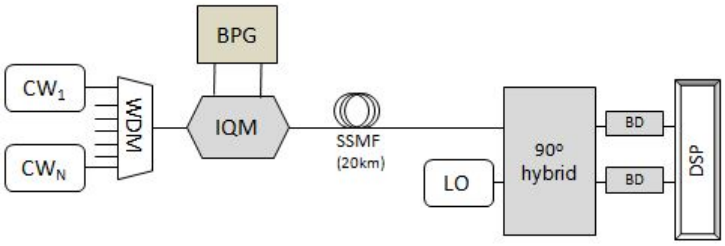



Figure 5.1: Simulation setup.

explained in Chapter 3. The same random sequence is transmitted in every channel. Then, the modulated optical carrier is transmitted over a SSMF of 20km with an attenuation coefficient of 0.2 dB/km, a dispersion of 17 ps/nm/km and a nonlinear coefficient of 1.4 mW/m.

Once the phase is modified in order to imprinted the data in the carrier, an optical coherent receiver is required to detect each symbol and map it correctly, according to the information exposed in Chapter 4 about the digital coherent optical receivers. As it is referred in this chapter, the coherent receiver requires a local oscillator (LO), which is combined with the received signal in the 90° optical hybrid. The resulting signal is then converted to the electrical domain by the balanced detectors (BD). Finally, each I and Q component is sampled by the ADC, so that, the digital signal processing (DSP) can be performed, as it is explain in Section 4.3. Note that in the Figure 5.1, the analogue-to-digital conversion is not explicit, being also represented by the DSP block.

### 5.1.2 Discussion of Results

#### Analysis of the Spacing Between Channels

Figure 5.3 and 5.4 show the evolution of the EVM for different modulation formats and different channel spacing. NRZ-BPSK/QPSK/16-QAM and Nyquist shaped BPSK/QPSK/16-QAM were the simulated modulation formats at 2.5 Gbaud. The spacing between the 8 channels is a value from 2.5 GHz to 10 GHz.

As it can be seen, the minimum spacing value for NRZ formats is twice the one of the Nyquist shaped formats. These values can be explained based on the spectrum of each format, which is represented in Figure 5.2. From it, one can verify that the bandwidth of the NRZ signal is twice the symbol rate, while in the other case, the bandwidth is equal to the symbol rate. That is why the minimum channel spacing of the former is twice the one of the latter, in this particular case, 5 GHz and 2.5 GHz, respectively.

Back to Figure 5.3, it can be verified that, in the back-to-back (B2B) configuration, the system response is identical in all modulation formats due to the absence of nonlinear effects. In this way, the degradation of the system performance as the power at the fibre input per channel decreases is caused only by the reduction of the OSNR.

Actually, it is inevitable to refer the nonlinear effects, when analysing an optical fibre transmission. For the presented case, the nonlinear effects that should be taken into account are:

- **Self-Phase Modulation (SPM)** Signals with a time varying amplitude cause temporary changes in the refractive index of the propagation medium, resulting in temporary

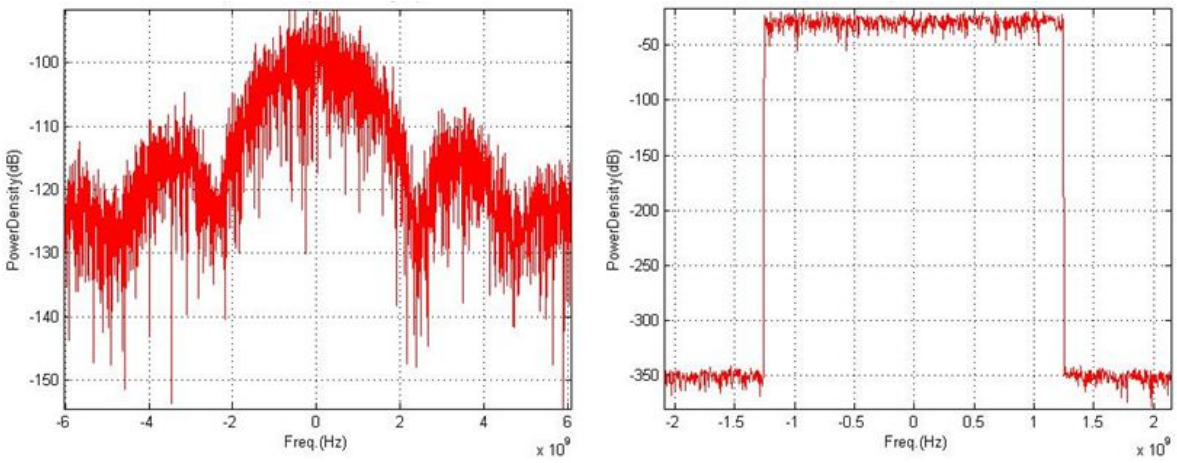



Figure 5.2: Spectrum of (left) NRZ and (right) Nyquist shaped formats at 2.5 Gbaud.

---

shifts of the signal phase, and consequently, in frequency chirping. This self-induced phenomenon traduces itself in a symmetric enlargement of the signal spectrum, maintaining the pulse shape unalterable.

- **Cross-Phase Modulation (XPM)** In a multi-channel scenario, the power variations that occur in a given channel will be translated in phase fluctuations in another co-propagating channel. The resulting enlargement of the spectrum may be symmetric or asymmetric, causing, in this last case, a distortion in the pulse shape.

- **Four-Wave Mixing (FWM)** Three or more co-propagating channels can interact between each other in such a way that a new co-propagating channel is created. This phenomenon is independent of the symbol rate and its effect increases as the spacing between channels decreases.

So, the system response to the propagation of a single-channel in a 20km SSMF is very similar, in the case of BPSK and QPSK formats, to that of the back-to-back (B2B) configuration, aside a 4dB penalty in the input optical power per channel, which was caused by the attenuation that was introduced by the fibre.

For high input optical power per channel, the EVM increases due to the appearance of nonlinearities. A closer look to Figure 5.3 allows to verify that the minimum value of EVM occurs for lower input fibre's power per channel as the spacing between channels decreases. This happens because of the fact that the FWM becomes more significant with the reduction of the spacing between channels. Once again, the system's response to the BPSK and QPSK is very similar. It is worse in the case of the propagation of 16-QAM formats because, in contrast to the other cases, the amplitude of the signal is time varying, provoking the emergence of SPM and also of XPM.

Inclusively, the system performance when propagating a 16-QAM signal, in the case of transmitting single-channel, shows a particularity: the EVM increases for high input power per channel. This happens due to the fact that the amplitude of the 16-QAM signal is not constant, resulting in the appearance of SPM for high power levels. Since the modulus of BPSK and QPSK symbols is constant, this phenomenon is not observed for these formats.

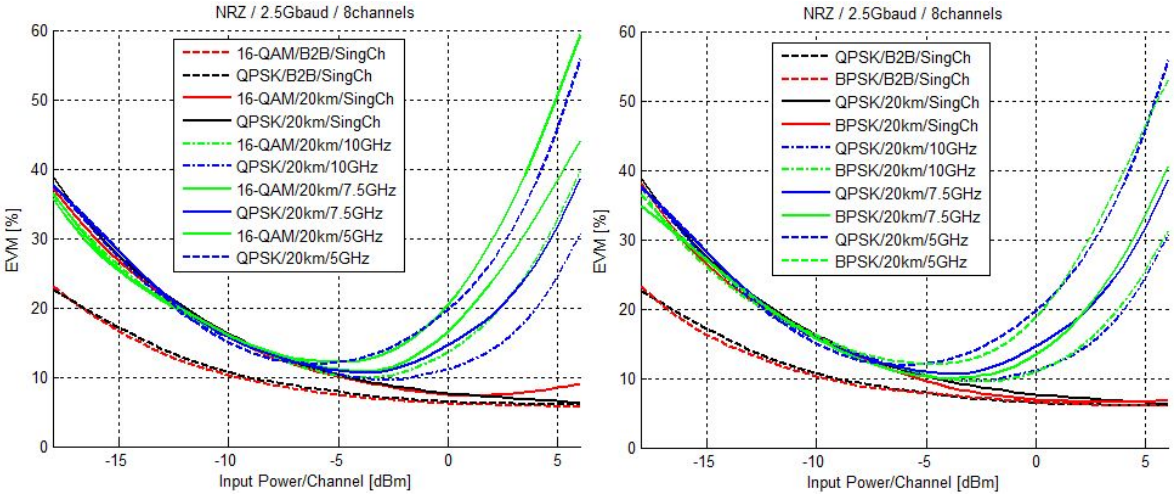



Figure 5.3: Evolution of EVM for (left) NRZ-QPSK/16-QAM and (right) NRZ-QPSK/BPSK at 2.5 Gbaud for several values of spacing between channels. Legend:modulation format/fibre length/spacing between channels.

---

For low input optical power per channel, the increase of the EVM is caused by the degradation of the OSNR.

Note that Figure 3.18 shows that the SNR which is required when transmitting 16-QAM is higher than the one that is required when transmitting QPSK in order to achieve the same BER for both formats. In Figures 5.3 and 5.4 this difference of the SNR is not observed because the system response is evaluated in terms of EVM, rather than in terms of BER.

As Figure 5.4 shows, alike conclusions can be taken in the case of employing Nyquist shaped formats.

From these comparisons, the 16-QAM does not seem more attractive than the QPSK format or even the BPSK. However, one should not forget that a symbol rate of 2.5 Gbaud for 16-QAM formats traduces itself on a bit rate of 10 Gbps, while the bit rates that correspond to 2.5 Gbaud are 5 Gbps and 2.5 Gbps, in case of QPSK and BPSK, respectively.

In addition, if the Nyquist shaping is applied instead of the NRZ line code, the occupied bandwidth is halved. This way, the bandwidth that is required to transmit a NRZ-QSPK signal is twice the one needed to transmit a Nyquist shaped QPSK, which means that it is possible to achieved twice the symbol rate, occupying the same bandwidth.

Comparing two cases in which neither the pulse shape nor the multi-level signalling is maintained, the bandwidth saving becomes even more evident. E.g., for a given symbol rate of 2.5 Gbaud, the corresponding bit rate is 2.5 Gbps and 10 Gbps, when transmitting NRZ-BPSK and Nyquist shaped 16-QAM, respectively. For the mentioned symbol rate, the required bandwidth to transmit a NRZ-BPSK signal is 5 GHz, while, in the other transmission, it is 2.5 GHz. It means that the employment of Nyquist shaped 16-QAM allows to achieve four times the bit rate that was accomplished employing NRZ-BPSK with the half of the occupied bandwidth. If one compares NRZ-QPSK, instead of NRZ-BPSK, with the Nyquist shaped 16-QAM, it will be verified that the latter allows to achieve twice the bit rate that is verified in the former transmission, requiring half the bandwidth.

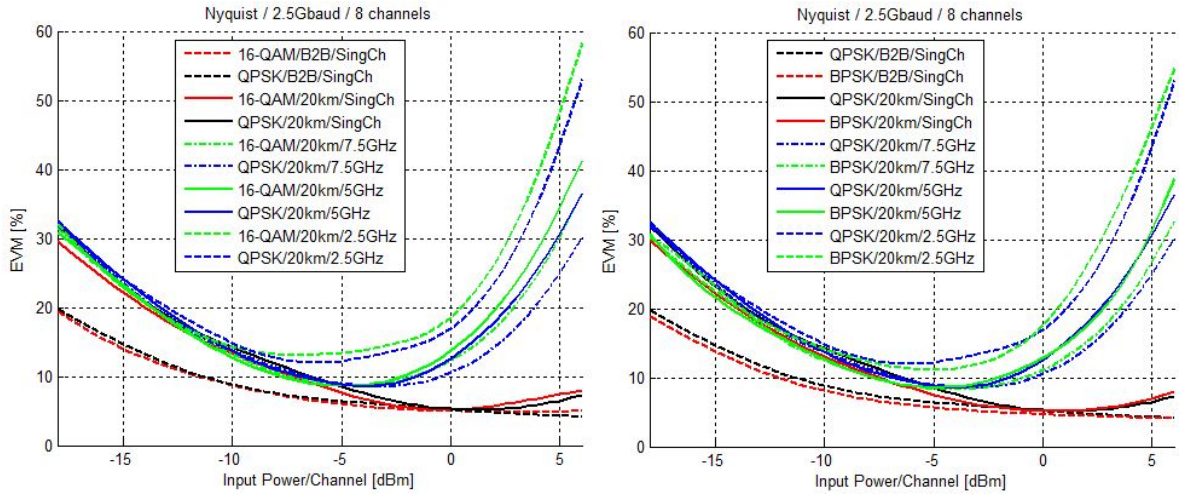



Figure 5.4: Evolution of EVM for (left) Nyquist shaped QPSK/16-QAM and (right) Nyquist shaped QPSK/BPSK at 2.5 Gbaud for several values of spacing between channels. Legend:modulation format/fibre length/spacing between channels.

---

### Analysis of the Number of Channels

Figure 5.5 shows the evolution of the EVM's value according to different number of channels for a spacing between them of 2.5 and 5 GHz, when transmitting a NRZ-QPSK signal at 1.25 Gbaud.

From it, one can conclude that the increase of the number of channels causes the degradation of the system performance, independently of the spacing between channels, for high input power per channel. This behaviour is justified by the fact that the impact of the fibre nonlinearities that were described in the previous subsection is more significant as the transmitted power in the fibre increases.

Moreover, one can verify once again the increasing effects of the FWM as the channels get closer between each other.

For lower input power per channel, the system response in the case of the multi-channel scenario is very similar of that in the case of the transmission of a single-channel in a 20km SSMF because the nonlinear effects are not triggered for such low power levels, being the degradation of the OSNR the reason of the observed decrease of the EVM.

### Analysis of the Symbol Rate

Figure 5.6 shows the evolution of the EVM, when transmitting a NRZ-QPSK signal, in each one of the 8 channels, at 2.5 and 1.25 Gbaud for different spacing between channels.

Identically to what was verified in the previous simulations, the system has a similar performance when transmitting single- or multi-channel for low input power per channel, independently of the adopted symbol rate. The increase of the EVM in this power range results from the OSNR degradation.

For higher power levels, nonlinear effects become significant. The four-wave mixing is the dominant nonlinearity, since, theoretically, the self- and cross-phase modulation do not manifest in the fibre transmission of modulation formats whose symbols have constant modulus,

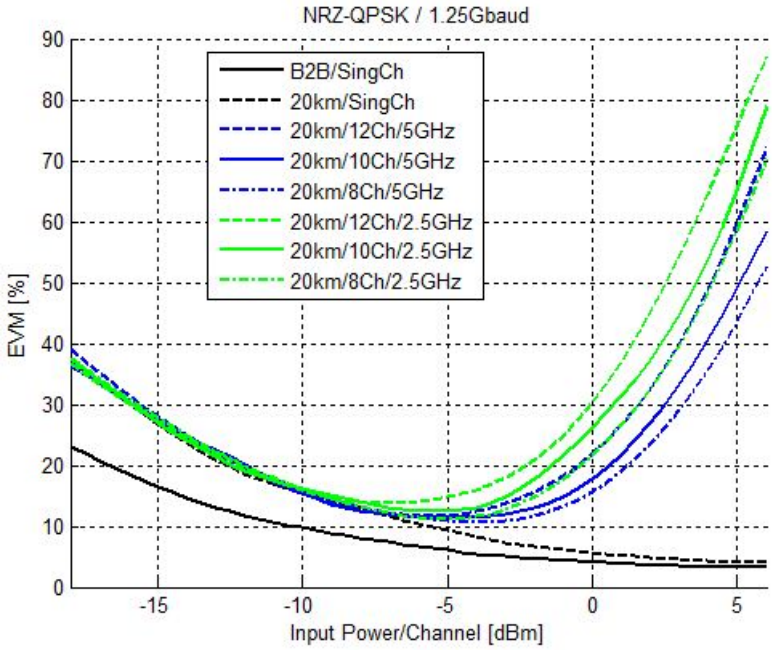



Figure 5.5: Evolution of the EVM for different number of channels, when transmitting NRZ-QPSK formats at 1.25 Gbaud. Legend: fibre length/number of channels/spacing between channels.

---

as QPSK and BPSK. Actually, the FWM is responsible for the higher values of EVM that were verified in the case of the lowest symbol rate.

For each symbol rate, the tested spacing between channels were either twice or four times the symbol rate. Also, the relation between the sample rate and the symbol rate was the same for all cases that were compared. In spite of maintaining these relations, the system's performance was not similar for large input power, which proves that FWM depends mostly on the spacing between channels (in the case of transmitting three or more co-propagating channels) and the level of the transmitted power in the fibre. That is why, for a given relation channel spacing/symbol rate, the observed EVM in the case of 1.25 Gbaud is larger than that in the case of 2.5 Gbaud, i.e., because the channels are closer to each other.

## 5.2 Laboratory Experiments

Due to the limitation of resources and schedule, it was only possible to perform the transmission of NRZ-QPSK and Nyquist shaped 16-QAM formats.

Results will prove that, in spite of being relatively novel, Nyquist shaped formats are practicable and its employment provides considerable bandwidth savings.

All error vector magnitude values were obtained for 1024 recovered symbols, over 10 independent measurements.

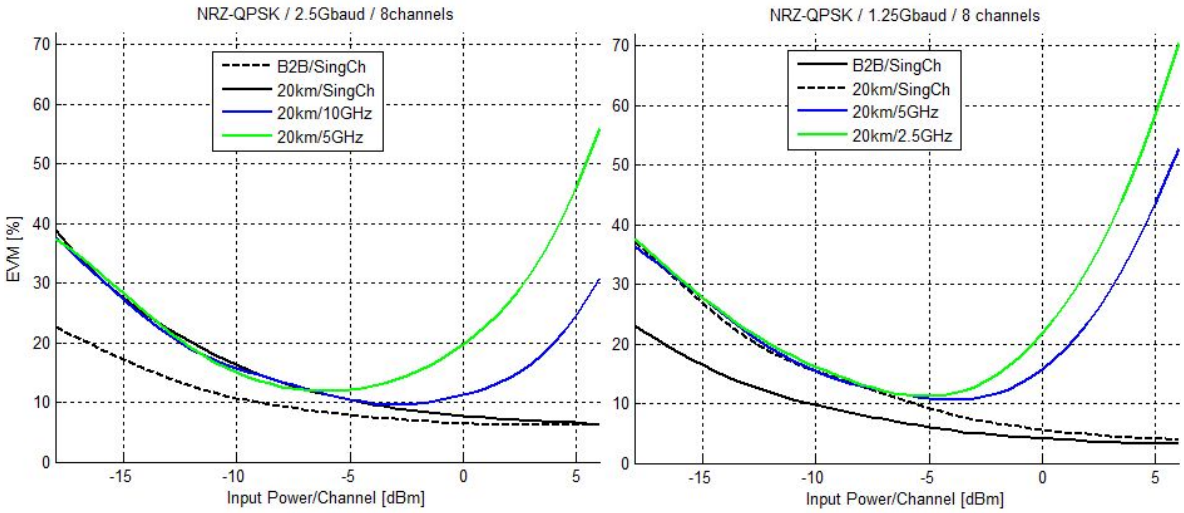



Figure 5.6: Evolution of the EVM for NRZ-QPSK transmission at (left) 2.5 Gbaud and (right) 1.25 Gbaud. Legend: fibre length/spacing between channels.

### 5.2.1 Laboratory Setup

Figure 5.7 illustrates the used experimental setup to analyse the transmission of both NRZ-QPSK and Nyquist shaped 16-QAM.

Actually, the setup is not exactly the same for both cases. The difference between them is the equipment that was used to generate the two distinct modulation formats. Once the generation of 16-QAM formats requires two four-level drive signals, the bit pattern generator (SHF 12100B included in SHF 56Gbps BERT) that drives the IQ modulator (IQM), together with the signal generator ($S\&R^{®}$ SMR50), in order to generate a NRZ-QPSK signal, has to be substituted. This way, an arbitrary waveform generator (AWG) (Agilent M8190A) took its place in the Nyquist shaped 16-QAM transmission.

In fact, it was not only the multi-level signalling that have determined the use of the AWG, but also the pulse shaping. The bit pattern generator that was used to generate NRZ-QPSK does not enable the generation of Nyquist shaped formats. E.g., the transmission of Nyquist shaped QPSK would require the use of the AWG, instead the utilization of the mentioned bit pattern generator.

The UDWDM channel group was created by a Mach-Zehnder modulator (MZM), which was driven by two radio-frequency signals whose phase relation was of about $3\pi/2$. This channel group is then filtered by a wave shaper (WS)(Finisar 4000S) tuned to $\sim$ 80 GHz bandwidth.

Because of the introduced losses of the remaining part of the system, the wave shaper is followed by an Erbium-doped fibre amplifier (IPG EAD-500-C3-W). This amplifier adds noise to the signal, so that, it has to be filtered after the amplification.

In order to observe the UDWDM channel group in the optical spectrum analyser (APEX AP2441A), a 50/50 coupler was employed after the filter. Note that this visualization is indispensable for the system operator, since it allows to verify the channels alignment.

Then, the input channels power is modified by acting in the optical attenuator that was placed between the mentioned 50/50 coupler and a 1/99 coupler. The employment of this

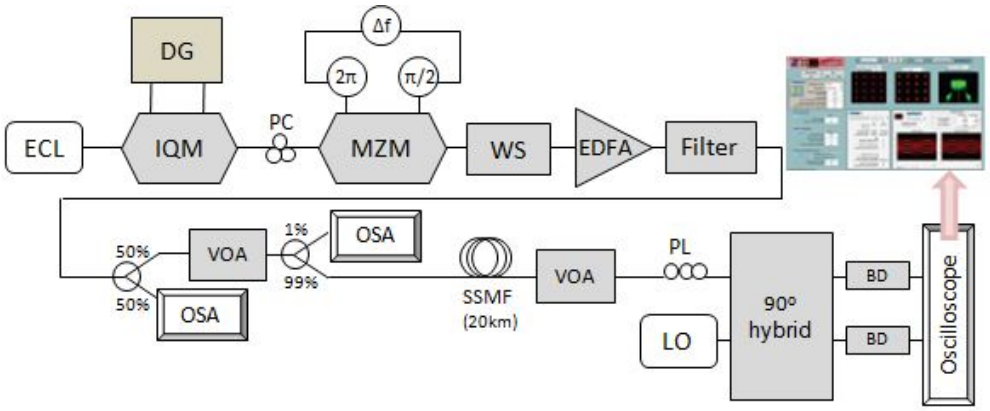



Figure 5.7: Experimental setup.

coupler is required to measure the input power in another optical spectrum analyser (EXFO FTB-500).

The optical attenuator that is located between the fibre and the polarization locker is important so that it limits the power at the coherent receiver's input, whose maximum input power is -6 dBm. The coherent receiver is constituted by a 90° optical hybrid and a pair of balanced detectors (BDs).

Finally, the resulting electrical signal is converted to the digital domain by a 100GSa/s real-time oscilloscope (Tektronix MS072004C). The oscilloscope was connected to the *Optical Coherent Receiver* (OCR) analyser, presented in Section 4.5, via Ethernet, where the digital signal processing was performed in order to compensate fibre's impairments.

### 5.2.2 Laboratory Results

#### NRZ-QPSK Transmission

Figure 5.8 illustrates the evolution of the error vector magnitude that was obtained for the transmission of NRZ-QPSK formats at different symbol rates, namely, 0.625, 1.25 and 2.5 Gbaud.

As it can be seen, the relations between the channels spacing and the symbol rate was not maintained for all cases, as it was desired. In fact, the spacing between channels was dictated by the success of the hardest task that was inherent to the transmission: the alignment of the UDWDM channel group.

All channels shall have the same power, so that, they interfere between each other in the same way. This is accomplished by acting on the phase of the two RF signals that drive the MZM and on the bias voltage of the referred modulator. Figure 5.9 shows two UDWDM channel groups, at the input fibre, that were obtained after performing the described procedure, which has revealed itself very time-consuming. From this figure, one can also verify that the number of channels is dependent on the spacing between them, for the same available transmission bandwidth, and indirectly on the symbol rate because the minimum spacing between channels must not be lower than twice the symbol rate in case of the NRZ format.

Regarding system performance, one can conclude that it has met the expectations.

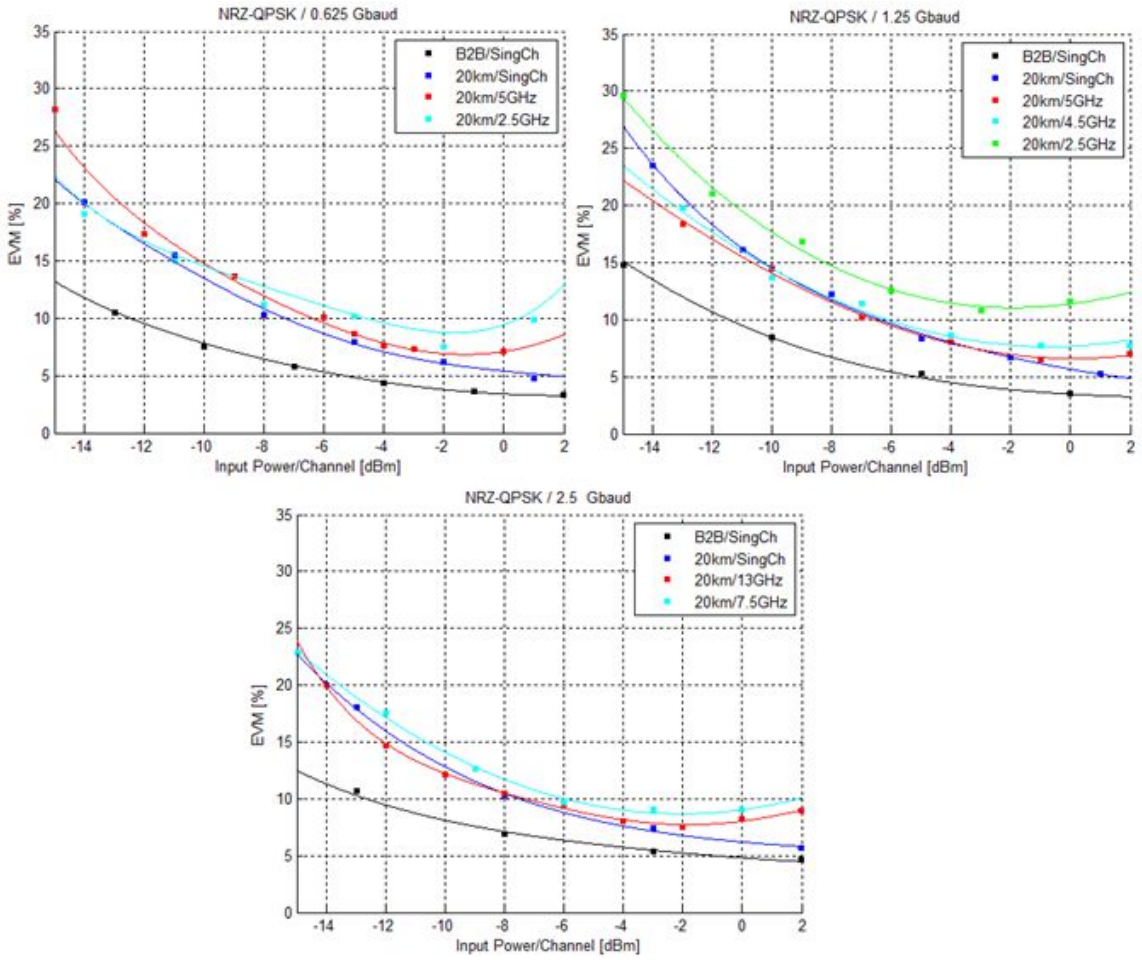



Figure 5.8: Evolution of EVM for NRZ-QPSK formats transmission at (upper left) 0.625 Gbaud (upper right) 1.25 Gbaud and (bottom) 2.5 Gbaud.

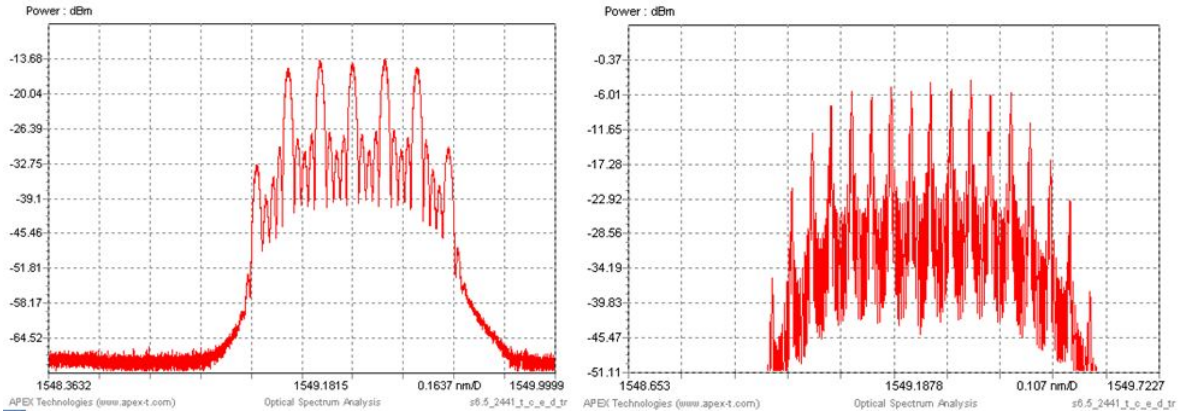



Figure 5.9: Generated UDWDM channel group, employing NRZ-QPSK, at (left) 2.5 Gbaud with a spacing between channels of 13 GHz and (right) 0.625 Gbaud with a spacing between channels of 5 GHz.

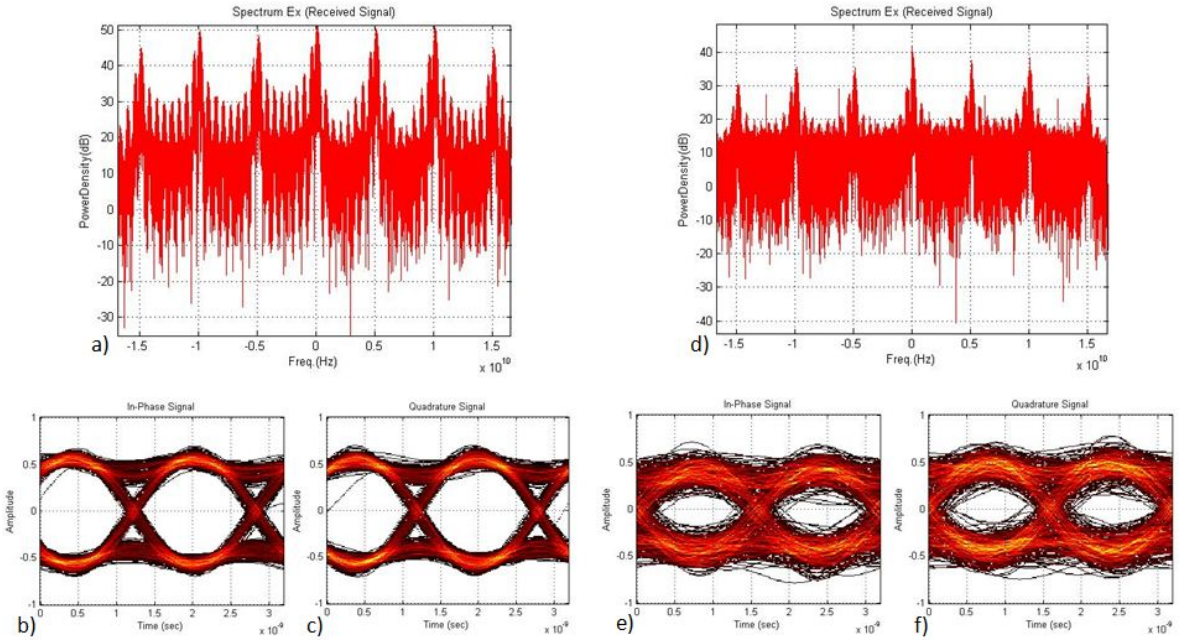



Figure 5.10: Spectrum and eye-diagrams of the received NRZ-QPSK signal (@0.625 Gbaud, $\Delta f = 5GHz$) for an input power per channel of (a)(b)(c) -3 dBm and (d)(e)(f) -15 dBm.

---

The analysis of Figure 5.8 allows to verify a penalty of about 4 dB of the transmission through the 20km SSMF in relation to the B2B configuration for the three examined symbol rates. Also, it can be seen that, for all tested symbol rates, the EVM increases as the input power per channel decreases. This increase is justified by the degradation of the OSNR, which can be verified not only in the received spectrum, but also in the eye-diagram, as it is shown in Figure 5.10.

On the other hand, the growth of the EVM for higher input fibre's power per channel is not related with OSNR. Instead, this increase occurs owing to the emergence of nonlinear effects, which were already addressed in Subsection 5.1.2.

Theoretically, self- and cross-phase modulation do not manifest, since the modulus of the NRZ-QPSK symbols is constant. Therefore, the four-wave mixing, is the fibre nonlinearity that causes the increase of the EVM that was obtained for this power range. That is why the system's response is degraded as the spacing between channels is reduce, as it can be seen in Figure 5.8.

### Nyquist shaped 16-QAM Transmission

Figure 5.11 represents the evolution of the error vector magnitude that was obtained after the detection and compensation of Nyquist shaped 16-QAM symbols, which were transmitted at 2.5 and 3 Gbaud.

Unlike the experiments that were described in the previous subsection, the generated spacing between channels was equal for both symbol rates, more precisely, 5 GHz. Unfortunately, it was not possible to test another symbol rate/channel spacing combinations due to the limited schedule and the very time-consuming task of aligning all channels.

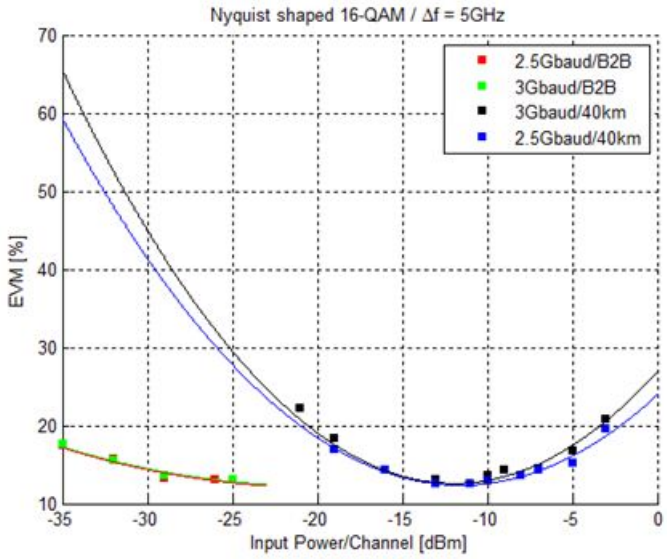



Figure 5.11: Evolution of EVM for Nyquist shaped 16-QAM transmission at 2.5 and 3 Gbaud. Legend: symbol rate/fibre length.

---

The two UDWDM channel groups are illustrated in Figure 5.12. The difference in relation to the previous case is evident. In this case, each channel spectrum has a near-perfect rectangular spectrum, while, in the previous case, each channel spectrum was sinc shaped. This results in bandwidth saving, since a NRZ format occupies twice the bandwidth for a Nyquist shaped format, as it was seen in Subsection 5.1.2.

Focusing on the developed transmissions of both NRZ-QPSK and Nyquist shaped 16-QAM at 2.5 Gbaud, one can verify that the achieved bit rate in the latter transmission is twice the former, occupying half of the bandwidth that NRZ-QPSK signal requires. Namely, 2.5 Gbaud traduces itself in 5 Gbps and 10 Gbps in case of transmitting QPSK and 16-QAM, respectively, and the required bandwidth is twice the symbol rate, i.e. 5 GHz, when transmitting the NRZ format, and equal to the symbol rate, i.e. 2.5 GHz, when transmitting the Nyquist shaped format. This shows the evident benefit of using Nyquist shaped formats in terms of bandwidth saving. In other words, for the same available transmission bandwidth, the employment of Nyquist shaped 16-QAM provides bit rates four times higher the ones that are achieved employing NRZ-QPSK.

Moreover, the difference between the two modulation formats is observed in the corresponding eye-diagrams, which are illustrated in Figures 5.10 and 5.13. This was expected, since the pulse shape in the time domain is substantially different, as it was demonstrated in Chapter 3.

Figure 5.11 allows to verify the expected system behaviour in the case of transmitting the UDWDM channel group over 40km of SSMF, which is similar to that of the NRZ-QPSK transmission in many ways.

Just like in the NRZ-QPSK transmissions, for low input power, the increase of the EVM results from the degradation of the optical signal-to-noise ratio, as it can be seen in Figure 5.13.

Additionally, nonlinear effects are triggered in the UDWDM transmission employing

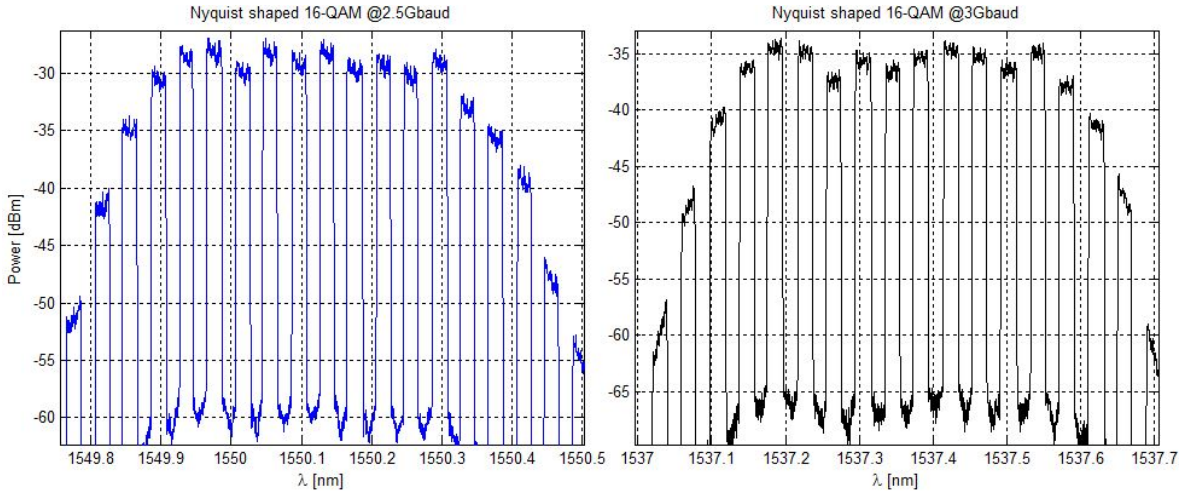



Figure 5.12: Obtained UDWDM channel group, employing Nyquist shaped 16-QAM, at (left) 2.5 and (right) 3 Gbaud with a spacing between channels of 5 GHz, at the input fibre.

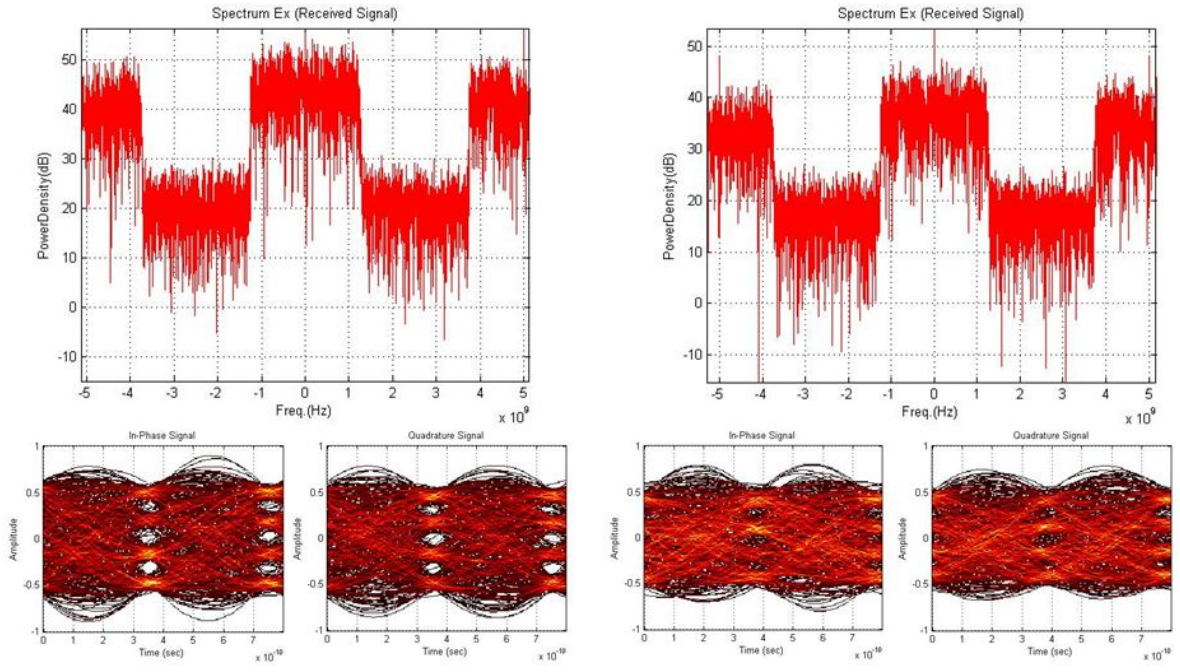



Figure 5.13: Spectrum and eye-diagrams of the received Nyquist shaped 16-QAM signal (@2.5 Gbaud, $\Delta f = 5GHz$) for an input power per channel of (a)(b)(c) -10 dBm and (d)(e)(f) -19 dBm.

Nyquist shaped 16-QAM for high power levels, as in the previous case. However, in this case, self- and cross-phase modulation become relevant, besides four-wave mixing. Owing to the fact that 16-QAM, and in this particular case, Nyquist shaped 16-QAM signals have a time varying amplitude, SPM and XPM effects will emerge for the referred power range. Since the spacing between channels is kept to 5 GHz in both transmissions, the FWM will affect them in a similar way. This is the reason why the evolution of the error vector magnitude is identical for both studied symbol rates.

# Chapter 6

# Conclusions and Future Work

The new communication paradigm, in which not only individuals but also machines are connected between each other, exchanging greats amounts of data, is pushing the network capacity to its limit.

It is well-known that optical communications is the only communication technology that provides bit rates of order of Terabits. However, the current optical networks (XG-PONs), were projected to allow the upload and the download of information at 2.5 and 10 Gbps, respectively, and the ones that are under standardisation today (NG-PON2), will allow up- and downstream rates of 10 and 40 Gbps, which are still far from the maximum fibre's capacity of transmission.

With the purpose of enlarging the referred rates, many researches have been carried on with successful results, achieving transmissions at bit rates of hundreds of Gbps or even Tbps. All of them show a common feature: the employment of high-order modulation formats together with coherent detection and digital signal processing.

In fact, such rates require spectral efficiencies that OOK formats cannot provide. Therefore, the utilization of advanced modulation formats, such as QPSK and 16-QAM, becomes imperative.

Once the information is convey in phase, in the case of M-PSK formats, and in the phase and amplitude, in the case of M-QAM formats, optical coherent reception is required because direct detection techniques only recover information from the amplitude of the optical carrier. The coherent detection together with digital signal processing have enabled a huge breakthrough in fibre communications.

Nevertheless, it is not only the multi-level signalling that dictates the occupied bandwidth, but also the pulse shape, as the developed simulations and laboratory experiments have shown. It was verified that the required bandwidth to transmit NRZ formats is twice the one needed to transmit Nyquist shaped formats.

Particularly, the laboratory experiments have shown that employing Nyquist shaped 16-QAM, the bit rate is doubled with half of the bandwidth that is occupied by the NRZ-QSPK. E.g., for a symbol rate of 2.5 Gbaud, the corresponding bit rate is 5 Gbps, when employing NRZ-QPSK formats, and 10 Gbps, when using Nyquist shaped 16-QAM, being the occupied bandwidth 5 GHz in the first case and 2.5 GHz in the last one.

Moreover, the experimental work have allowed to verify that, in spite of being relatively recent, the employment of Nyquist shaped formats is practicable and results in a significant bandwidth saving, which becomes even more evident when combining Nyquist shaping with

high-order multi-level signalling.

In relation to the modulation formats' generation, the performed simulations, as well as the laboratory experiments, have shown that the IQ modulator is a very powerful external modulator, since it can be used to generate many different modulation formats, such as NRZ-BPSK or Nyquist shaped 16-QAM, by applying the proper drive signals.

Additionally, the experimental procedure has revealed another application of the Mach-Zehnder modulator, besides the external modulation of an optical carrier. As it was observed, it also allows the generation of UDWDM channel groups, when it is driven by two RF signals with a phase relation of $3\pi/2$.

It was also possible to study some parameters that are related to UDWDM transmissions of advanced modulation formats. Namely, it was verified that the increase of the channels number and the reduction of the spacing between them results in a degradation of the system's response, for high power's levels, due to the fibre's nonlinearities. As the spacing between the channels decreases, the four-wave mixing has a bigger impact on the transmission and it is even more evident as the number of channels increases, since it causes an increase of power in the fibre. The simulations and laboratory results have shown as well that 16-QAM formats are more affected by the fibre's nonlinearities than BPSK and QPSK formats. In fact, theoretically, BPSK and QPSK are not affect by neither self- nor cross-phase modulation because their symbols have a constant modulus.

The discussed results indicate the combination of Nyquist shaping and high-order multi-level signalling as a viable answer to increase the current fibre transmission data rates.

However, the use of an arbitrary waveform generator is not a cost-effective solution to generate such advanced modulation formats. This is how Field Programmable Gate Array (FPGA) arises as an alternative option to achieve Nyquist shaped formats. Since the operations' principles of the two devices are substantially different, a lot of research has to be done in this context.

# Bibliography


[1] P. J. Winzer and R. Essiambre, "Advanced Optical Modulation Formats," *Proceedings of the IEEE*, vol. 94, pp. 952–985, May 2006.

[2] "Evolution of FTTH Networks for NG-PON2," White paper, February 2013.

[3] S. Matsuoka, "Ultrahigh-speed Ultrahigh-capacity Transport Network Technology for Cost-effective Core and Metro Networks," August 2011.

[4] C. Peucheret, "Direct and External Modulation of Light," November 2009. Lecture notes, Experimental Course in Optical Communication, Department of Photonics Engineering, Technical University of Denmark.

[5] D. A. B. Miller, "Optical physics of quantum wells." AT&T Bell Laboratories.

[6] H. Zappe, *Fundamentals of Micro-Optics*. Cambridge University Press, 2010.

[7] J. Piprek, Y. Chiu, and J. E. Bowers, "Analysis of Multi-Quantum Well Electroabsorption Modulators."

[8] B. K. Saravanan, *Frequency Chirping Properties of Electroabsorption Modulators Integrated with Laser Diodes*. PhD thesis, University of Ulm, April 2006.

[9] B. E. A. Saleh and M. C. Teich, *Fundamentals of Photonics*. John Wiley & Sons, Inc., 1991.

[10] M. Seimetz, *High-Order Modulation for Optical Fiber Transmission*. Springer, 2009.

[11] C. Peucheret, "Generation and Detection of Optical Modulation Formats," March 2012. Lecture notes, Technical University of Denmark, Department of Photonics Engineering.

[12] E. L. Wooten, K. M. Kissa, A. Yi-Yan, E. J. Murphy, D. A. Lafaw, P. F. Hallemeier, D. Maack, D. V. Attanasio, D. J. Fritz, G. J. McBrien, and D. E. Bossi, "A Review of Lithium Niobate Modulators for Fiber-Optic Communications Systems," *IEEE JOURNAL OF SELECTED TOPICS IN QUANTUM ELECTRONICS*, vol. 6, January/February 2000.

[13] D. Fonseca, *Optical Single Sideband Transmission Systems*. PhD thesis, Instituto Superior Técnico, March 2008.

[14] H. Bilgekul, "Chapter 3: Line Codes and Their Spectra." Communication Systems I, Department of Electrical and Electronic Engineering.

[15] E. Ip and J. M. Kahn, "Power Spectra of Return-to-Zero Optical Signals," *JOURNAL OF LIGHTWAVE TECHNOLOGY*, vol. 24, March 2006.

[16] I. P. Kaminow, T. Li, and A. E. Willner, *Optical Fiber Telecommunications V B: Systems and Networks*, vol. 2. Elsevier Inc., USA, 2008.

[17] J. F. da Rocha, "Sistemas de Comunicação Digital," 2010. Apontamentos de Sistemas de Comunicação I, Departmento de Eletrónica, Telecomunicações e Informática.

[18] A. Tan and E. Pincemin, "Performance Comparison of Duobinary Formats for 40-Gb/s and Mixed 10/40-Gb/s Long-Haul WDM Transmission on SSMF and LEAF Fibers," *JOURNAL OF LIGHTWAVE TECHNOLOGY*, vol. 27, February 2009.

[19] B. Châtelain, C. Laperle, K. Roberts, M. Chagnon, X. Xu, A. Borowiec, F. Gagnon, and D. V. Plant1, "A Family of Nyquist Pulses for Coherent Optical Communications," *OPTICS EXPRESS*, vol. 20, April 2012.

[20] A. B. Carlson, P. B. Crilly, and J. C. Rutledge, *Communications Systems: An Introduction to Signals and Noise in Electrical Communication.* McGraw-Hill, fourth ed., 2002.

[21] R. Schmogrow, M. Winter, M. Meyer, D. Hillerkuss, S. Wolf, B. Baeuerle, A. Ludwig, B. Nebendahl, S. Ben-Ezra, J. Meyer, M. Dreschmann, M. Huebner, J. Becker, C. Koos, W. Freude, and J. Leuthold, "Real-time Nyquist Pulse Generation Beyond 100 Gbit/s and Its Relation to OFDM," *OPTICS EXPRESS*, vol. 20, January 2012.

[22] J. M. Kahn and K.-P. Ho, "Spectral Efficiency Limits and Modulation/Detection Techniques for DWDM Systems," *IEEE JOURNAL OF SELECTED TOPICS IN QUANTUM ELECTRONICS*, vol. 10, March/April 2004.

[23] A. H. Gnauck and P. J. Winzer, "Optical Phase-Shift-Keyed Transmission," *JOURNAL OF LIGHTWAVE TECHNOLOGY*, vol. 23, January 2005.

[24] H. S. Chung, S. H. Chang, and K. Kim, "Mitigation of Imperfect Carver-Induced Phase Distortions in a Coherent PM-RZ-QPSK Receiver," *JOURNAL OF LIGHTWAVE TECHNOLOGY*, vol. 28, December 2010.

[25] M. Seimetz, "Performance of Coherent Optical Square-16-QAM-Systems based on IQ-Transmitters and Homodyne Receivers with Digital Phase Estimation," in *Optical Fiber Communication Conference, 2006 and the 2006 National Fiber Optic Engineers Conference. OFC 2006*, March 2006.

[26] M. Nakazawa, T. Hirooka, M. Yoshida, and K. Kasai, "Ultrafast Coherent Optical Transmission," *Selected Topics in Quantum Electronics, IEEE Journal of*, vol. 18, pp. 363–376, January/February 2012.

[27] S. Yan, D. Wang, Y. Gao, C. Lu, L. A.P.T., L. Liu, and X. Xu, "Generation of Square or Hexagonal 16-QAM signals using a Single Dual Drive IQ Modulator Driven by Binary Signals," in *Optical Fiber Communication Conference and Exposition (OFC/NFOEC), 2012 and the National Fiber Optic Engineers Conference*, pp. 1–3, 2012.

[28] S. J. Savory, "Digital Coherent Optical Receivers: Algorithms and Subsystems," *IEEE JOURNAL OF SELECTED TOPICS IN QUANTUM ELECTRONICS*, vol. 16, September/October 2010.

[29] E. Ip, A. P. T. Lau, D. J. F. Barros, and J. M. Kahn, "Coherent Detection in Optical Fiber Systems," *OPTICS EXPRESS*, vol. 16, January 2008.

[30] J. M. Senior, *Optical Fiber Communications: Principles and Practice*. Prentice Hall Europe, Pearson Education Limited, third ed., 2009.

[31] M. El-Chammas, "Time-Interleaved ADCs: Theory and Design," December 2011. Time-interleaved Tutorial.

[32] A. j. Viterbi and A. M. Viterbi, "Nonlinear Estimation of PSK-modulated Carrier Phase with Application to Burst Digital Transmission," *Information Theory, IEEE Transactions on*, vol. 29, no. 4, pp. 543–551, 1983.

[33] H. Chen, *Development of an 80 Gbit/s InP-based Mach-Zehnder Modulator*. PhD thesis, Faculty IV - Tachnical University of Berlin, 2007.

[34] "Internet Growth vs Internet Penetration by Geographic Region." `http://www.comsoc.org/blog/ieee-comsoc-infographic-internet-growth-vs-population-\penetration-geographical-region`. IEEE Communications Society, Last access on June 2013.

[35] R. M. Schmogrow, M. Meyer, P. C. Schindler, A. Josten, S. Ben-Ezra, C. Koos, W. Freude, and J. Leuthold, "252 Gbit/s Real-Time Nyquist Pulse Generation by Reducing the Oversampling Factor to 1.33,," in *Optical Fiber Communication Conference/National Fiber Optic Engineers Conference 2013, OSA Technical Digest*, 2013. paper OTu2I.1.

[36] Y. Huang, D. Qian, F. Yaman, T. Wang, E. Mateo, T. Inoue, Y. Inada, Y. Toyota, T. Ogata, M. Sato, Y. Aono, and T. Tajima, "Real-Time 400G Superchannel Transmission using 100-GbE based 37.5-GHz Spaced Subcarriers with Optical Nyquist Shaping over 3,600-km DMF link," in *Optical Fiber Communication Conference/National Fiber Optic Engineers Conference 2013, OSA Technical Digest*, 2013. paper NW4E.1.

[37] A. Shahpari, J. D. Reis, R. Ferreira, D. M. Neves, M. Lima, and A. Teixeira, "Terabit+ (192 x 10 Gb/s) Nyquist Shaped UDWDM Coherent PON with Upstream and Downstream over a 12.8 nm Band," in *Optical Fiber Communication Conference/National Fiber Optic Engineers Conference 2013, OSA Technical Digest*, 2013. paper PDP5B.3.

[38] S. Zhang, *Traveling-wave Electroabsorption Modulators*. PhD thesis, University of California, Santa Barbara, April 1999.

[39] S. C. Technologies, "Tutorial Note Number 5 - Modulation Schemes."

[40] Z. Lua, K. Zhao, and X. Li, "Photovoltaic Effect in Ferroelectric LiNbO3 Single Crystal," 2011.

[41] "DP-QPSK 100Gbps LN Modulator." http://jp.fujitsu.com/group/foc/en/services/optical-devices/100gln/. Fujitsu Optical Components, Last access on July 2013.

[42] Y. Jiang, *High-Speed Electronic Signal Processing for Pre-Compensation in Optical Communications*. PhD thesis, Queen's University, October 2011. Kingston, Ontario, Canada.

[43] "60G-R-EAM-1550 60 GHz Reflective Electroabsorption Modulator (EAM)." http://www.ist-iphobac.org/news/60GHz-REAM/60G_R_EAM_1550_A.pdf. Preliminary datasheet, CPI Technologies, Last access on July 2013.

[44] M. Born and E. Wolf, *Principles of Optics*. Pergamon Press, Oxford, sixth ed., 1993.

[45] N. Suzuki and T. Ozeki, "Simultaneous Compensation of Laser Chirp, Kerr Effect, and Dispersion in lO-Gb/s Long-Haul Transmission Systems," *JOURNAL OF LIGHTWAVE TECHNLOGY*, vol. 11, September 1993.

[46] F. Koyama and K. Iga, "Frequency chirping of external modulation and its reduction," *Electronics Letters*, vol. 21, pp. 1065–1066, July 1985.

[47] A. Sarkar, "Line Coding." Digital Communication, Chapter 4.

[48] P. Siripongwutikorn, "Baseband Transmission (1): Line Coding and Power Spectrum Analysis." Lecture of Principles of Communications.

[49] S. Zhang, "Advanced Optical Modulation Formats in High-speed Lightwave System."

[50] J. Leuthold, R. Schmogrow, D. Hillerkuss, and W. F. C. Koos, "Super Channels Based on Nyquist Multiplexing," in *Opto-Electronics and Communications Conference (OECC), 2012 17th*, pp. 30–32, 2012.

[51] N. Stojanovic, N. Gonzalez, C. Xie, Y. Zhao, B. Mao, J. Qi, and L. MBinh, "Timing Recovery in Nyquist Coherent Optical Systems," in *Telecommunications Forum (TELFOR), 2012 20th*, pp. 895–898, 2012.

[52] L. Zhao, H. Shankar, and A. Nachum, "40G QPSK and DQPSK Modulation."

[53] R. Griffin and A. Carter, "Optical Differential Quadrature Phase Shift Key (oOQPSK) for High Capacity Optical Transmission," in *Optical Fiber Communication Conference and Exhibit, 2002. OFC 2002*, pp. 367–368, March 2002.

[54] M. Nakazawa, "Optical Quadrature Amplitude Modulation (QAM) with Coherent Detection up to 128 States," in *Optical Fiber Communication - incudes post deadline papers, 2009. OFC 2009. Conference on*, pp. 1–3, March 2009.

[55] V. Ferragina, N. Ghittori, and F. Maloberti, "Low-power 6-bit Flash ADC for High-speed Data Vonverters Architectures," in *Circuits and Systems, 2006. ISCAS 2006. Proceedings. 2006 IEEE International Symposium on*, 2006.

[56] H. Agarwal and A. Rana, "Front End Track and Hold Circuit for High Speed Analog to Digital Converters," in *Communication Systems and Network Technologies (CSNT), 2012 International Conference on*, pp. 785–789, 2012.

[57] W. C. B. Jr. and D. Hodges, "Time interleaved converter arrays," *Solid-State Circuits, IEEE Journal of*, vol. 15, no. 6, pp. 1022–1029, 1980.

[58] A. Tychopoulos, O. Koufopavlou, and I. Tomkos, "FEC in Optical Communications - A Tutorial Overview on the Evolution of Architectures and the Future Prospects of Outband and Inband FEC for Optical Communications," *Circuits and Devices Magazine, IEEE*, vol. 22, no. 6, pp. 79–86, 2006.