UNIVERSITÀ DEGLI STUDI DI PARMA

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

Dottorato di Ricerca in Tecnologie dell'Informazione

XX Ciclo

Tommaso Foggi

ELECTRONIC PROCESSING FOR OPTICAL COMMUNICATION SYSTEMS

Dissertazione presentata per il conseguimento

del titolo di Dottore di Ricerca

Gennaio 2008

a Carla, Giannina,

Marino e Michele

Contents

List of Figures

List of acronyms

AFC automatic frequency control

APSK amplitude and phase shift keying

ASE amplified spontaneous emission

AWGN additive white Gaussian noise

BCJR Bahl Cocke Jelinek Raviv

BER bit error rate

DCF dispersion compensation fiber

DD decision directed

DFE decision feedback equalization

DGD differential group delay

DM delay multiply

DSP digital signal processing

DWDM dense wavelength division multiplexing

EDFA erbium-doped fiber amplifier

FED frequency error detector

FFE feed-forward equalization

FIR finite-impulse response

FWHM full width half maximum

FWM four-wave mixing

GVD group velocity dispersion

IF intermediate frequency

IM/DD intensity modulation direct detection

ISI intersymbol interference

LED light-emitting diodes

LMS least mean square

LO local oscillator

MAP maximum a posteriori probability

ML maximum likelihood

MLSD maximum likelihood sequance detection

MMSE minimum mean square error

MSE mean square error

MZM Mach-Zehnder modulator

NRZ nonreturn-to-zero

OOK on-off keying

OSNR optical signal-to-noise ratio

PBS polarization beam splitter

pdf probability density function

PEP pairwise error probability

PIN P-type intrinsic N-type

PLL phase-locked loop

PMD polarization mode dispersion

PMF polarization mantaining fiber

polMUX polarization multiplexing

PSD power spectral density

PSK phase shift keying

PSP principal state of polarization

QAM quadrature amplitude modulation

QPSK quadrature phase shift keyings

RLS recursive least-squares

RSSD reduced-state sequence detection

RZ return-to-zero

SMF single-mode fiber

SOP state of polarization

SPM self-phase modulation

VA Viterbi algorithm

XPM cross-phase modulation

ZF zero forcing

Foreword

The development of optical communications in recent decades has made possible the huge growth of wired communications, leading to new widespread services like the internet. All the services which rely on wired communications are in endless expansion, and the bandwidth demand is always increasing. This fact represents one of the more critical issue from a technological point of view, because, on one hand big efforts deal with the exploitation of the available technology; on the other hand, new solutions to upgrade the existing data rates to higher capacities must be devised.

The optical fiber allows the transmission of outstanding data amounts, incurring in very small attenuation; nevertheless, it presents several drawbacks that make difficult to increase the transmission speed. The most relevant channel impairments which cause a performance degradation, can be classified in linear or nonlinear type. *Group velocity dispersion* (GVD) and *polarization mode dispersion* (PMD) are linear impairments whose impact on system performance becomes very relevant with bit rates beyond 10 Gb/s (nowadays the research is aimed towards 40 or even 100 Gb/s). On the other hand, long-haul transmission systems require quite high signal power levels, which are the cause of nonlinear effects, such as *self-phase modulation* SPM, *cross-phase modulation* (XPM), *four-wave mixing* (FWM), Raman and Brillouin scattering.

Several ways can be followed to upgrade the optical communication systems to higher bit rates and combat the effects of such impairments; basically, it is possible to better exploit the wide bandwidth offered by the optical channel so that the total bit rate can be increased keeping affordable baud rates. *Dense wavelength division multiplexing* (DWDM) is one possible solution, which, together with new advanced fiber production technologies, allows to transmit the data over a broad range of wavelengths, thus reducing the bit rate on the single

channels. A different approach that is recently rising much interest consists of discarding the widely employed *on-off keying* (OOK) amplitude modulation (historically exploited for its simplicity, since in early years of optical communications the great bandwidth amount made spectral efficiency optimization useless), and start to implement more complex and efficient modulation formats as in radio communications. In this perspective, a lot of research activity is also directing to coherent detection techniques, already investigated in the early nineties, but then abandoned because of *intensity modulation direct detection* (IM/DD) cost efficiency, and technological ease. Coherent techniques require a more complex receiver architecture, but allow higher receiver sensitivity and more spectrally efficient modulation formats, therefore it can be guessed that these transmission schemes will probably represent the future of optical communication systems, provided that, nowadays, technological issues are less decisive.

Besides the considerations on the future of optical systems, today an important research field is represented by the investigation of effective penalty reduction techniques at the receiver end. Compensation techniques could be either based on optical processing or electronic processing of the received signal. Signal processing in the electrical domain presents several advantages; it is simple to integrate in the receiver electronics, it is cost effective, and a wide variety of solutions are already known from theory and radio communications. The research activity concerning the typical fiber impairments mentioned above, which are peculiar of optical transmission systems, is getting every year more popular. First works on electrical equalization with fiber optics go back to almost twenty years ago, when high-speed optically-amplified long-haul transmission systems were initially devised. The presence of the photodiode, a nonlinear device that converts the incident light intensity into an electrical current, is responsible for an irreversible transformation of the signal at the receiver. This means that fiber dispersions get worsened, and standard known electronic compensation techniques must be taken into account in this particular environment in order to effectively combat fiber impairments.

The thesis work is organized as follows. The first chapter deals with the desciption of optical communication systems in their fundamental elements, as transmitter and receiver architectures and detection strategies, fiber channel typical impairments and relevant aspects. Chapter two treats the implementation of penalty reduction strategies on IM/DD optical systems, from simple equalization architectures, to more complex and effective sequence detection

strategies, including the development of performance computation methods, which represent a useful tool for system design and analysis. The subject of the third chapter is the novel phase modulation formats, which are not commercial yet, but largely investigated, and in particular a receiver architecture based on IM/DD which performs differential detection, leading to a sequence detection strategy completely insensitive to fiber impairments, is presented here. In chapter four electronic compensation with coherent detection is reported, showing that very simple devices, properly designed, are sufficient to remove any amount of linear dispersion introduced by the fiber. Chapter five draws conclusions on the whole work.

Acknowlegdements.

The work that has been carried out throughout the Ph.D. course and presented in this dissertation was accomplished with the fundamental help and contribution of valuable people I had the opportunity to team up with. A special thank goes to Enrico Forestieri whose kindness and unbounded knowledge of optical theory and mathematics I have always, deeply appreciated. My sincere gratitude to Giulio Colavolpe, who involved me in this great experience, who has always encouraged me, whose enthusiasm for our work has never run out even for one day, a friend before all.

January 21, 2008

Chapter 1

The optical communication system

Optical communication systems are constituted by several peculiar devices all involved in the characterization of the system model upon which compensation techniques are based. A simple schematic of a generic optical communication system is represented in Fig. 1.1. The transmitter and the receiver are interfaces between the optical and the electrical domain and, as will be outlined, they present some important features that deeply influence the signal model. Moreover, the most important characteristic of an optical fiber, that is the carrier frequency of almost 200 THz which allows extremely high data rates, is responsible of a wide range of different impairments that induce their effects on signal propagation. In the following the most important aspects of these elements are described, with a special attention on their effects on the received signal model. Then, a section is dedicated to the basic concepts of electronic processing concerning the techniques that have been investigated in this thesis work, specifically *feed-forward* and *decision-feedback* equalization (FFE and DFE), and *maximum likelihood sequence detection* (MLSD).

Notation.
Some details about the notation used in the remainder of the thesis are now given. In optical communications, especially when dealing with PMD, it is necessary to take into account the vectorial nature of the signals propagating in the fiber. Besides the longitudinal direction of propagation of lightwaves, the optical field is represented in a bidimensional plane orthogonal to the

Figure 1.1: Schematic of an optical communication system.

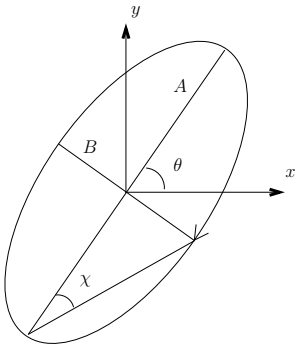

Figure 1.2: Polarization ellipse

longitudinal axis. The evolution of the signal on this plane is related to the degree of polarization of the signal, which we always consider as fully polarized (on the contrary, the noise introduced in optical amplifiers is unpolarized). A common way to represent the polarization of optical signals is the polarization ellipse, shown in Fig. 1.2, where two fundamental parameters can be identified, the angle θ between the x-axis and the major ellipse axis A, called *azimuth*, and the *ellipticity* $\chi = \arctan \varepsilon$, where ε is the *ellipticity angle*, the angle between the major A and minor axis B of the polarization ellipse; these two parameters are sufficient to describe the state of polarization of the optical signal propagating in the fiber. There exist two mathematical models that are historically known [1] to be useful for analytical treatment of polarized signals, the *Jones* and the *Stokes spaces*. In the Jones space (or *calculus*) the polarized signal (only fully polarized light can be described by this model) is represented by a 2x1 vector, whose elements, choosing a proper basis, are the two components of the electrical field, whereas optical elements are expressed by 2x2 matrices. Tab. 1 shows some examples of signal polarizations.

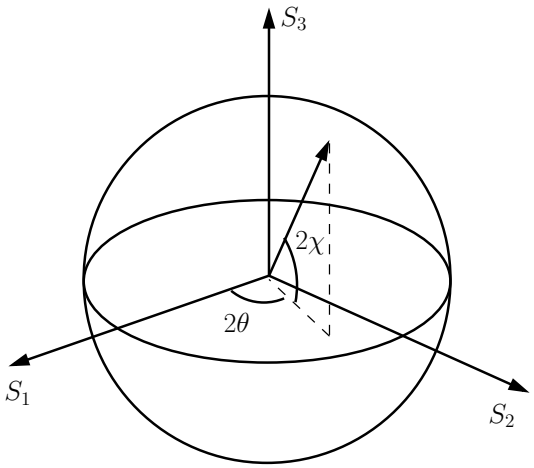

Figure 1.3: The Poincarè sphere.

The other model is the Stokes space, that is a three-dimensional representation that refers to the *Poincarè sphere*, shown in Fig. 1.3. In this representation all possible signal polarizations are mapped on the surface of a sphere (note that partially polarized signals are mapped on the radius inside the sphere, depending on the degree of polarization); four parameters, called Stokes parameters, can be defined and put in relation to polarization ellipse parameters

$$\begin{aligned} S_0 &= I \\ S_1 &= Ip\cos 2\theta \cos 2\chi \\ S_2 &= Ip\sin 2\theta \cos 2\chi \\ S_3 &= Ip\sin 2\chi \end{aligned} \tag{1.1}$$

Linear polarized x-direction	$(1\ 0)^T$
Linear polarized y-direction	$(0\ 1)^T$
Linear polarized at 45° from the x-axis	$1/\sqrt{2}\,(1\ 1)^T$
Right circular polarized	$1/\sqrt{2}\,(1\ \ -\jmath)^T$

Table 1.1: Examples of Jones vectors

where I is the total signal intensity and p is the degree of polarization. The *Poincarè sphere*, therefore, is the space of the vector $u = 1/S_0\,(S_1\ S_2\ S_3)^T$. The Stokes parameters form a vector of four elements that is used in the *Mueller calculus* to study the effects of optical elements on signals. They can also be related to the signal components in the Jones space, once a basis has been defined. Assuming a $(\hat{x}, \hat{y})$ basis

$$\begin{array}{rcl} S_0 & = & |E_x|^2 + |E_y|^2 \\ S_1 & = & |E_x|^2 - |E_y|^2 \\ S_2 & = & 2\mathrm{Re}(E_x E_y^*) \\ S_3 & = & 2\mathrm{Im}(E_x E_y^*). \end{array} \tag{1.2}$$

The two models are more comfortable depending on the specific application. In the remainder of the thesis vectors in the Jones space will be denoted by a cap, e.g. $\hat{e}$, whereas in the Stokes space an arrow will be used, e.g., $\vec{e}$. Moreover, matrices will be denoted by boldface uppercase fonts (e.g, $\mathbf{H}$), whereas vectors of elements in lowercase boldface (e.g. $\mathbf{a}$).

1.1 The transmitter

In an optical communication system at the transmitter end multiple tasks are accomplished; an optical carrier must be generated through a fiber compatible device, then the optical field must be modulated with an information message and finally coupled with the fiber. The most popular optical sources to be mentioned are *light-emitting diodes* (LED) and semiconductor lasers, whose properties, such as efficiency, compactness, reliability, good match with fiber characteristics, possibility of direct modulation at high frequency, led to the current widespread diffusion in optical systems.

1.1.1 Phase noise

As far as the communication system model is concerned, these optical sources present a relevant drawback of a non-negligible spectral linewidth, whose amount depends on the specific kind of source and technology [2]. The effect of the perturbation of the light emitted by a laser is a random phase, that spreads the ideal delta function of the signal spectrum. The phase noise

process $\phi(t)$ is characterized by a Wiener process [3–6]

$$\phi(t) \doteq \int_0^t \dot{\phi}(\tau)d\tau \tag{1.3}$$

where the derivative $\dot{\phi(t)}$ is a zero-mean white Gaussian process with *power spectral density* (PSD) $S_\phi(\omega) = 2\pi\Delta\nu$ (actually, this is an accurate approximation over the bandwidth of interest). To show that this process leads to a signal spectrum of Lorentzian type [7] as depicted in Fig. 1.4, the laser signal can be expressed as

$$s(t) = \sqrt{P_s}e^{\jmath(\omega_0 t+\phi(t)+\theta)} \tag{1.4}$$

where P_s is the signal power, ω_0 the carrier frequency, θ a random phase uniformly distributed over $[0, 2\pi]$ which makes the $s(t)$ wide-sense stationary. The autocorrelation of $s(t)$ can be computed

$$R_s(t_1, t_2) = E[s(t_1)s^*(t_2)] = P_s e^{\jmath\omega_0(t_1-t_2)}E[e^{\jmath\Phi(t_1,t_2)}], \tag{1.5}$$

and $\Phi(t_1, t_2)$ is a zero-mean white Gaussian random variable defined as

$$\Phi(t_1, t_2) \doteq \phi(t_1) - \phi(t_2) = \int_{t_2}^{t_1} \dot{\phi}(u)du \tag{1.6}$$

with variance $\sigma_\Phi^2 = 2\pi\Delta\nu|t_1 - t_2|$. Then, since $E[e^{\jmath\Phi}] = e^{-\sigma_\Phi^2/2}$, (1.5) becomes

$$R_s(t_1, t_2) = P_s e^{\jmath\omega_0(t_1-t_2)}e^{-\pi\Delta\nu|t_1-t_2|} \tag{1.7}$$

and since $s(t)$ is wide-sense stationary $|t_1 - t_2| = |\tau|$ and then

$$R_s(\tau) = P_s e^{\jmath\omega_0\tau}e^{-\pi\Delta\nu|\tau|}. \tag{1.8}$$

The Fourier transform of (1.8) leads to the PSD of the laser linewidth corrupted by phase noise

$$S_s(\omega) = \frac{\frac{2P_s}{\pi\Delta\nu}}{1 + (\frac{\omega-\omega_0}{\pi\Delta\nu})^2}, \tag{1.9}$$

that is a Lorentzian spectrum with *full width half maximum* (FWHM) bandwidth equal to $\Delta\nu$.

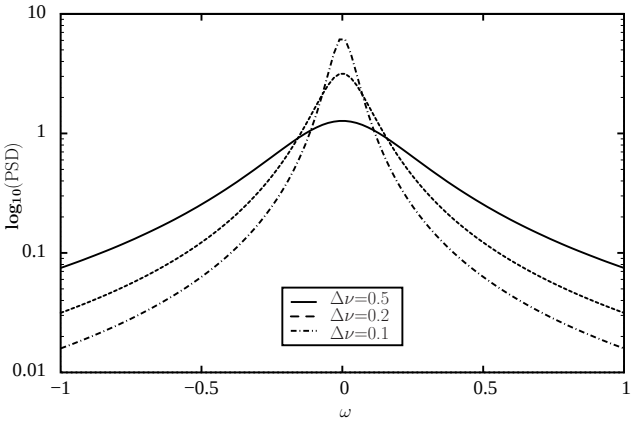

Figure 1.4: Lorentzian functions.

1.1.2 Modulators

The optical beam is modulated with the information bearing electrical signal. The modulation process can be devised in several ways; for low speed (below 10 Gb/s) direct modulation of the optical source is still possible, but it is nowadays obsolete, since current and future system speed are well beyond that limit. Thus, external modulators are used. A good solution for short haul communication systems is represented by the Electro-Absorption Modulator, a sort of reverse-biased P-type intrinsic N-type PIN detector. Although its good performance in terms of chirping in comparison to direct modulation, residual dynamic chirp of this modulator makes long haul transmissions unfeasible. Thus, the most popular modulation device is the *Mach-Zehnder* modulator (MZM), where the optical input is temporarily split into two paths, each properly controlled by the modulating electrical signal so that the resulting optical output is obtained transforming the phase modulation into intensity modulation by recombining the two paths. A perfect match in the two opposite phase modulations would mean a chirp free modulator (no residual phase modulation). MZMs can be dual drive or single drive, the latter of X-cut or Z-cut type; dual drive and X-cut are known to provide a better performance in terms of chirping, whereas Z-cut require a lower drive voltage. In each case, trade-off must be considered between the linearity of the device (the transfer characteristics are usually $\cos^2$ shaped) and its extinction ratio, that is the ra-

tio between its maximum and minimum output powers, expressed in decibels. In OOK systems, the presence of finite extinction ratios turns out to be an important parameter to be accounted for when implementing compensation strategies at the receiver. Mach-Zehnders are also used in more sophisticated transmission systems to realize pulse carvers, needed to obtain return-to-zero (RZ) or duobinary pulses [8, 9] instead of common nonreturn-to-zero (NRZ) pulses. Of particular relevance is their employment with phase modulation formats, together with phase modulators, which present the advantage of producing constant envelope output signals, with respect to MZMs, but at the cost of a chirp that must be removed, precisely, through pulse carvers; on the other hand, MZMs can operate near-perfect phase shifts, but with residual intensity modulation [10].

1.1.3 Chirp

A possible side effect of the modulation process is the chirping of the output signal [2], which, as in the fiber propagation, is a time-dependent carrier frequency change. Frequency chirp is responsible for a spectral broadening of the transmitted pulse, and a phase distortion with effects similar to the ones of GVD, which is treated of in §1.2.1. Depending on the sign of β_2, the dispersion index, the interaction between chirp and dispersion is different, in one case leading also to an initial mutual compensation of the two opposite effects.

This brief description of transmitter end elements was aimed at focusing a few elements which have a clear impact on system performance; the analysis and implementation of signal processing techniques is in fact influenced by either the signal model or laser source impairments.

1.2 The optical fiber

In the beginning of seventies, when the propagation loss in silica waveguides was reduced to 20 dB/Km and then, in the late seventies, to 0.2 dB/Km, lightwave communications experienced an endless expansion that is still ongoing. The pros of optical fibers, intended as communication channels, are countless and decisive:

- Huge bandwidth availability;

- High speed reliable communications;
- Low losses;
- Simple isolation from external interference;
- Simple electronics.

Though, many drawbacks emerge as higher signalling rate are experimented and devised. Well known typical fiber impairments become much more critical when the impulse width is in the order of a few tens of picoseconds.[1] This is the case of GVD and PMD, which have a different origin and nature (being the former deterministic, the latter stochastic), but a similar impact on receiver performance. On the other hand, at higher speed the electronic circuitry sets feasibility issues, just when electronic processing of the signal at the receiver would be recommendable. Moreover, other impairments arise, mostly of non-linear type, but they go beyond the purpose of the thesis work, and are not further addresses.

Then, the whole work is focused on single-mode fiber (SMF) optics, so that the relevant impact that modal dispersion has, for example, on multi-mode fibers is not considered here.

Optical amplification, instead, became fundamental in fiber optics at the beginning of the nineties, allowing the realization of thousand-mile long haul all-optical transmission systems. But, as it is known from amplification theory, noise is introduced, and, in this case, it is due to the amplified spontaneous emission (ASE) phenomenon that takes place in the erbium-doped optical amplifiers. In the following, all these topics are detailed.

1.2.1 Group velocity dispersion

Also known as *chromatic dispersion*, GVD represents maybe the most typical fiber impairment. It is one of the first to be studied and modeled, as its effects on signal propagation are easy to observe because of its deterministic nature. For this reason, it also easy to combat, since simple measures of its magnitude on given pieces of fiber are sufficient to design proper dispersion maps, so that the residual dispersion can be kept below an acceptable threshold. This is made possible by the existence of particular fibers with negative dispersion

[1] in fact, in a common 10 Gb/s system the impulse is about 100 ps wide, whereas in experimental 40 Gb/s systems it is about 25 ps wide.

values (DCF, dispersion compensation fiber), opposite to the standard SMF. Nevertheless, residual dispersions cannot be avoided, especially for high data rate systems, so that it is important to understand the meaning of chromatic dispersion and its effects on the signal model.

Following [2], let us assume a spectral component at frequency ω travelling in a fiber of length L. It spends a time equal to $T_i = L/\nu_g$ to get to the other end, where ν_g is defined as the group velocity

$$\nu_g = \left(\frac{\partial\beta}{\partial\omega}\right)^{-1}, \tag{1.10}$$

where β is the fiber propagation constant. By using $\beta = \tilde{\eta}k_0 = \tilde{\omega}/c$ in (1.10), it is possible to write $\nu_g = c/\tilde{\eta}_g$, where the group index

$$\tilde{\eta}_g = \tilde{\eta} + \omega\,(d\tilde{\eta}/d\omega) \tag{1.11}$$

has been introduced. It can be noticed, then, that the frequency dependence of the group velocity leads to a spectral broadening of the input pulse since every spectral components will arrive at a different time at the output. Being $\Delta\omega$ the spectral width of the pulse, the broadening for a fiber of length L is

$$\Delta T = \frac{dT}{d\omega}\Delta\omega = \frac{d}{d\omega}\left(\frac{L}{\nu_g}\right)\Delta\omega = L\beta_2\Delta\omega. \tag{1.12}$$

The parameter $\beta_2 = d^2\beta/d\omega^2$ is known as the GVD parameter, and it is considered an index of pulse broadening during propagation.

Since it is often used the range of wavelength $\Delta\lambda$ emitted by the optical source, instead of $\Delta\omega$, (1.12) can be written as

$$\Delta T = \frac{d}{d\lambda}\left(\frac{L}{\nu_g}\right) = DL\Delta\lambda, \tag{1.13}$$

where

$$D = \frac{d}{d\lambda}\left(\frac{1}{\nu_g}\right) = -\frac{2\pi c}{\lambda^2}\beta_2, \tag{1.14}$$

and D is called the dispersion parameter and is expressed in ps/(km-nm). Typical values of D for standard SMF in the area of $\lambda \sim 1550$ nm are in the order of 17 ps/(km-nm). From a communication point of view, the effect of GVD can be modeled as a channel transfer function

$$H(\omega) = \exp\left[-\jmath\frac{\beta_2 L_i\omega^2}{2}\right], \tag{1.15}$$

where L_i is the length of the i^{th} piece of fiber, in a multi-span communication system, each possibly characterized by different values of D_i. It can be noticed that the transfer function expressed in (1.15) represents a phase distortion only, while it leaves the amplitude undistorted. This is an important property, that holds for PMD as well, and in the following chapters its implications in relation to amplitude and phase modulation formats will be outlined. A more useful notation can be used [11], in order to make the expression (1.15) independent of the bit rate R_b. The novel parameter γ is then defined as

$$\gamma = 2\lambda_0 R_b^2 D_i/\omega_0, \tag{1.16}$$

where ω_0 is the carrier frequency. The GVD transfer function can now be written as

$$H(\omega) = \exp\left[-\jmath\frac{\gamma}{4}\left(\frac{\omega}{R_b}\right)^2\right]. \tag{1.17}$$

The broadening of the transmitted pulse has the obvious consequence of introducing intersymbol interference (ISI) in the signal pattern, causing a performance degradation. Effective methods for combating the effects of GVD are the design of proper dispersion maps, which depend on the specific system considered, the controlled chirp of transmitted pulses, which can reduce the effects of GVD in a given transmission range, or the use of more advanced modulation formats as the duobinary [8]. The progress of electronics in latest years suggested the possibility of exploiting electronic processing techniques at the receiver to mitigate GVD effects.

1.2.2 Polarization mode dispersion

The polarization mode dispersion is an extremely complex phenomenon due to fiber birefringence, and has been widely and deeply studied over decades because of its stochastic nature. Various analytical approximations that can be used to obtain accurate models, are available [12–18]. In a SMF there exist two orthogonal modes of propagation, whose degeneracy is broken by several factors that contribute to fiber birefringence, like core asymmetries, mechanical stress of the fiber, or imperfections. The degree of birefringence is given by

$$\beta_m = |\tilde{n}_x - \tilde{n}_y| \tag{1.18}$$

where $\tilde{n}_x, \tilde{n}_y$ are the mode indexes of the orthogonally polarized fiber modes. Birefringence leads to a periodic power exchange between the two polarization

components, and the period is called *beat length* $L_b = \lambda/\beta_m$. Typical values of these parameters for $\lambda \sim 1\,\mu$m are $\beta_m \sim 10^{-7}$ and $L_b = 10$ m. The effect of birefringence is that a lightwave maintains its *state of polarization* (SOP) during propagation only if initially linearly polarized along one of the principal axes (*principal states of polarization*, PSPs), otherwise its SOP changes from linear to elliptical and back to linear propagating in a fiber span equal to L_b. Unfortunately, birefringence usually varies through the fiber, because of different conditions of the fiber, production faults, temperature fluctuations, so that a lightwave rapidly assumes a random SOP, and, moreover, each spectral component of a non-monochromatic optical signal can assume different SOPs. Special fibers called *polarization maintaining fibers* (PMF) are used to control the polarization of signals through an intentional high birefringence value, making random changes negligible. Though, in real optical communication systems at high data rate, PMD is a critical impairment mainly for its stochastic nature. A mathematical model in the Jones space is now presented, that helps in the representation and study of its involvements.

PMD is usually mathematically modeled by the PMD vector $\vec{\Omega}$, which can be independent of frequency over the signal bandwidth, and in this case we talk of first-order PMD, or dependent on frequency, and in this case second-order effects arise (or even of higher order). In the second-order approximation the vector is taken as linearly varying with frequency $\vec{\Omega} = \vec{\Omega}_0 + \vec{\Omega}_\omega(\omega - \omega_0)$, with $\vec{\Omega}_\omega$ being the derivative of $\vec{\Omega}$ with respect to the angular frequency, evaluated at the carrier frequency ω_0 [13]

$$\vec{\Omega}_\omega \doteq \left.\frac{\vec{\Omega}(\omega)}{d\omega}\right|_{\omega=\omega_0} = \Delta\tau_\omega \vec{q} + \Delta\tau \vec{q}_\omega \tag{1.19}$$

where

$$\Delta\tau \doteq |\vec{\Omega}(\omega_0)| \tag{1.20}$$

$$\Delta\tau_\omega \doteq \left.\frac{d|\vec{\Omega}(\omega)|}{d\omega}\right|_{\omega=\omega_0} \tag{1.21}$$

$$\vec{q} \doteq \frac{\vec{\Omega}(\omega_0)}{|\vec{\Omega}(\omega_0)|} \tag{1.22}$$

$$\vec{q}_\omega \doteq \left.\frac{d}{d\omega}\frac{\vec{\Omega}(\omega)}{|\vec{\Omega}(\omega)|}\right|_{\omega=\omega_0}. \tag{1.23}$$

The second-order PMD, regardless of the considered approximation (see [15, 17, 19]) depends on these parameters, which are, respectively, the *differential group delay* (DGD) (1.20), the DGD derivative (1.21), and the PMD vector rotation rate (1.23) (we also defined the PMD versor (1.22)).

If constant carrier polarization is assumed, the transfer function of a dispersive fiber can be represented by a 2x2 matrix [19–21]

$$\mathbf{F}(\omega) = \mathbf{R}\mathbf{U}(\omega)\mathbf{R}^{-1}, \tag{1.24}$$

where $U(\omega)$ is a unitary matrix

$$\mathbf{U}(\omega) = \begin{pmatrix} u_1(\omega) & u_2(\omega) \\ -u_2^*(\omega) & u_1^*(\omega) \end{pmatrix}, \tag{1.25}$$

being $|u_1(\omega)|^2 + |u_2(\omega)|^2 = 1$. The expression of $u_1(\omega), u_2(\omega)$ depends on the considered second-order approximation. All of them, though, refer to a reference frame in which, at the carrier frequency, the PMD vector is aligned with the S_1 axis in the Stokes space and such that $\mathbf{U}(0) = \mathbf{I}$, where $\mathbf{I}$ is the identity matrix.

The matrix $\mathbf{R}$ is a random rotation matrix, independent of frequency, representing a change of basis polarization states. Assuming the PMD versor (1.22)(which is uniformly distributed over the Poincarè sphere) coincident with the fast PSP at the carrier frequency. As a consequence, the effect of R is to realize a uniformly distributed signal power splitting between ρ (the ratio of signal power associated to the fast PSP) the PSPs at the carrier frequency. The expression for $\mathbf{R}$ is

$$\mathbf{R} = \begin{pmatrix} r_1 & -r_2^* \\ r_2 & r_1^* \end{pmatrix} \tag{1.26}$$

where

$$\begin{aligned} r_1 &= \cos\theta\cos\varepsilon - \jmath\sin\theta\sin\varepsilon \\ r_2 &= \sin\theta\cos\varepsilon + \jmath\cos\theta\sin\varepsilon \end{aligned} \tag{1.27}$$

θ and ε being independent random variables representing the fast PSP azimuth and ellipticity angle, whose *probability density functions* (pdfs) are

$$p_\theta(x) = \begin{cases} \frac{1}{\pi}, & -\frac{\pi}{2} \le x \le \frac{\pi}{2} \\ 0, & \text{otherwise} \end{cases} \tag{1.28}$$

$$p_\varepsilon(x) = \begin{cases} \cos 2x, & -\frac{\pi}{4} \le x \le \frac{\pi}{4} \\ 0, & \text{otherwise} \end{cases} . \tag{1.29}$$

Thus, the PMD vector at the carrier frequency has the expression $\vec{\Omega}_0 = \Delta\tau\vec{q}$, where

$$\vec{q} = [\cos 2\theta \cos 2\varepsilon, \sin 2\theta \cos 2\varepsilon, \sin 2\varepsilon]^T , \tag{1.30}$$

and the columns of (1.26) are the PSPs at the carrier frequency.

First-order PMD.
In this case, the PMD vector is independent of frequency, i.e., $\vec{\Omega} = \vec{\Omega}_0$, and the matrix $\mathbf{U}$ is diagonal

$$\mathbf{U}(\omega) = \begin{pmatrix} e^{+\jmath\omega\Delta\tau/2} & 0 \\ 0 & e^{-\jmath\omega\Delta\tau/2} \end{pmatrix}. \tag{1.31}$$

If θ_s and ε_s are the azimuth and the ellipticity angle of an input polarized signal so that (as for the PMD versor)

$$\vec{e}_{in} = [\cos 2\theta_s \cos 2\varepsilon_s, \sin 2\theta_s \cos 2\varepsilon_s, \sin 2\varepsilon_s]^T , \tag{1.32}$$

it is then possible to denote 2φ the angle between this input signal and the fast PSP $\vec{q} \cdot \vec{e}_i n = \cos 2\varphi$, and write it in the Stokes space as

$$\cos 2\varphi = \cos 2\theta \cos 2\varepsilon \cos 2\theta_s \cos 2\varepsilon_2 + \sin 2\theta \cos 2\varepsilon \sin 2\theta_s \cos 2\varepsilon_s + \sin 2\varepsilon \sin 2\varepsilon_s . \tag{1.33}$$

Through straightforward manipulations, under these assumptions it can be shown that the input signal, say $s_{in}(t)\hat{e}_{in}$, passing through the fiber and detected by photodiode, gives an output signal $s_{out}(t)$ equal to

$$s_{out}(t) = \rho|s_{in}(t + \Delta\tau/2)|^2 + (1 - \rho)|s_{in}(t - \Delta\tau/2)|^2 \tag{1.34}$$

that is, two replicas of the input signal, respectively delayed by the DGD amount, and scaled proportionally to the power splitting ρ

$$\rho = \frac{1}{2}(1 + \cos 2\varphi) = \cos^2 \varphi$$

$$1 - \rho = \frac{1}{2}(1 - \cos 2\varphi) = \sin^2 \varphi .$$

Thus, in first-order approximation, the photodetected only depends upon the DGD value $\Delta\tau$ and the power splitting $\cos^2 \varphi = \gamma_{ps}$, whereas the orientation of the PMD vector has no impact on the signal; in second- and all-order approximations, though, all PMD parameters contribute to the output signal

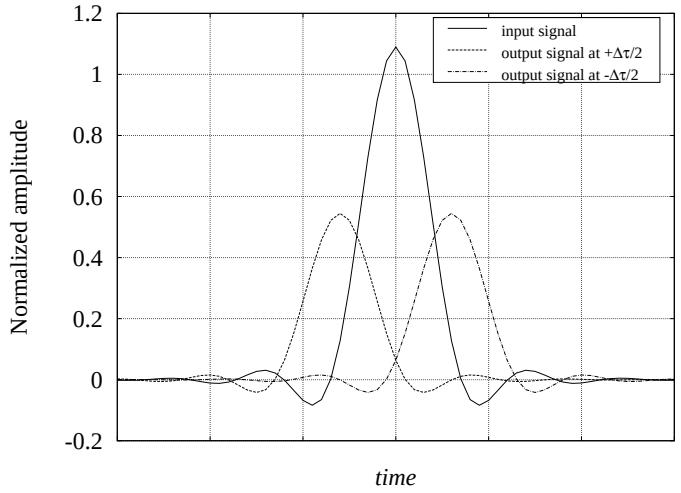

Figure 1.5: Fiber input signal and output signals on two orthogonal polarizations with $\rho = 0.5$ and $\Delta\tau = 1.5T$.

expression [20], and many distorted, and not only delayed, replicas of the input signal become source of ISI and therefore performance penalties. Fig. 1.5 shows a simple example of a signal at the output of a PMD-impaired fiber (before photodetection), with the two half-powered replicas of the input signal.

1.2.3 ASE noise

The advent of optical amplification in late eighties had made possible the realization of long haul optical communication systems. Even though modern optical fibers exhibit low propagation loss (0.2 dB/km), links of more than a few hundreds kilometers could not be devised without a re-amplification of the signal. Erbium-doped fiber amplifiers (EDFA) allow optical amplification avoiding electro-optical conversion; unfortunately, in the amplification process, as it happens in electronic amplifiers, a portion of the pump signal power is transferred to bunches of photons not belonging to the input signal, so that a noise signal is generated. This phenomenon is known as *amplified spontaneous emission*. The model of ASE noise has been widely studied [22, 23]; it can be treated as an additive white Gaussian noise (AWGN) [24], with some remarks.

Two independent noise processes take place, one for each polarization.[2] The noise two-sided PSD on each polarization is equal to $N_0 = \eta_{sp}(G-1)h\nu/G$, $\eta_{sp} \geq 1$ being the spontaneous emission parameter, $G \gg 1$ the amplifier gain (this assumption makes the ASE noise dominant over other sources of noise at the receiver, as shot and thermal noise, that therefore will be ignored), $h\nu$ the photon energy.

1.2.4 Filters

A brief section is devoted to the filters that have been used in the work, because in optical communications, besides all theoretical considerations, some constraints cannot be avoided, and the system performance must be analyzed under these constraints, especially in the IM/DD case. Typically, limited sets of optical filters are available. The most common type is the Gaussian filter, of first or higher orders. In particular, having the demand of vestigial symmetry filter shape, when working with phase modulations a fourth-order Gaussian filter has been adopted at the receiver, because it represents a good approximation of the requested filter, as shown in Fig. 1.6. Its frequency response is

$$H(\omega) = \exp[-0.5 \ln 2 \left(\frac{f}{B}\right)^8]. \tag{1.35}$$

Another constraint is represented by the frequency response of the photodiode and the following transimpedance amplifier at the receiver front end, which is usually modeled as a Bessel filter of fifth order. The presence of this filter has been avoided throughout the work, when necessary, assuming that the bandwidth of the electronic devices was sufficiently higher than the data rate of the optical system (as an example, electronics for 20 Gb/s systems used on 10 Gb/s systems would mean, practically, that this filter is not present).

1.3 The Receiver

Optical receivers have a double task: the conversion of the optical signal into an electrical filter, and the detection of the transmitted bit stream. After optical filtering, that is also performed in order to separate multiplexed signals, the optical field is photodetected by a photodiode, that transform the incident

[2] In the remainder, the noise signal will be denoted as $\hat{w}(t) = [w_1(t), w_2(t)]^T$, where $w_1(t), w_2(t)$ are complex-valued processes, one for each polarization.

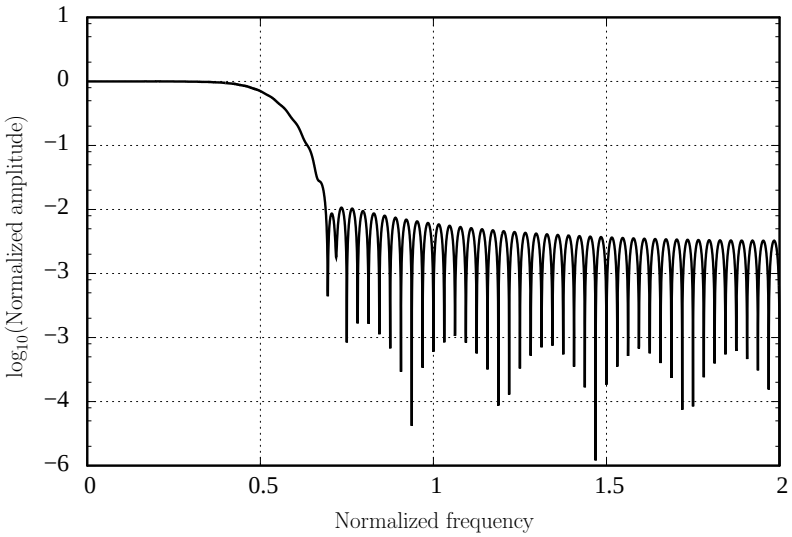

Figure 1.6: Amplitude frequency response of fourth-order Gaussian optical filter

light into a current, subsequently converted into a voltage by a transimpedance amplifier. Detectors are of different type, but typically symbol-by-symbol adaptive threshold detectors are exploited. Here more sophisticated detectors are presented, that have been implemented in the thesis work in order to improve the receiver performance.

1.3.1 Electrical equalization

The first and simplest compensation technique consists of a direct transversal filter, with adaptive taps. Such architecture (the FFE), is usually combined with a feedback tapped delay line, the DFE, which exploits the decisions taken on the received signal to remove additional interference. The optimal adaptation of the tap coefficients is a fundamental task in this kind of compensation; a useful strategy that takes into account both ISI and noise is the minimum mean square error (MSE) based on the minimization of $E\{|y_k - a_{k-D}|^2\}$, where y_k represents the samples of the received signal, a_{k-D} are the corresponding transmitted symbols, properly delayed of D symbol intervals. Other adaptation criteria can be considered, as for instance the minimization of the error

probability, which, though, is too difficult to implement, with respect to the effectiveness and ease of the MSE, in that it admits a closed-form solution [25]. If we consider a vector of input photodetected samples, $\mathbf{z}_k = \{z_n\}_{n=k-N-1}^{k}$, where the time index spans N samples, and a vector of equalizer tap coefficients $\mathbf{c} = \{c_n\}_{n=0}^{N-1}$, the equalizer output can be written

$$y_k = \sum_{i=0}^{N-1} c_i z_{k-i} = \mathbf{z}_k^T \mathbf{c}. \tag{1.36}$$

Then, the MSE becomes

$$\Gamma(\mathbf{c}) = E\{|y_k - a_{k-D}|^2\} = E\{|y_k|^2\} + E\{|a_{k-D}|^2\} - 2E\{\mathrm{Re}[y_k^* a_{k-D}]\}, \tag{1.37}$$

and expanding all the terms

$$E\{|y_k|^2\} = E\{y_k^* y_k\} = E\{\mathbf{c}^H \mathbf{z}_k^* \mathbf{z}_k^T \mathbf{c}\} = \mathbf{c}^H \mathbf{A} \mathbf{c} \tag{1.38}$$

where $(\cdot)^H$ is the transpose conjugate (Hermitian transpose), and A is a $N \times N$ matrix

$$\mathbf{A} = E\{\mathbf{z}_k^* \mathbf{z}_k^T\}. \tag{1.39}$$

Then, when $E\{a_k\} = 0$

$$E\{|a_{k-D}|^2\} = \sigma_a^2 \tag{1.40}$$

$$E\{y_k^* a_{k-D}\} = \mathbf{c}^H \mathbf{b}. \tag{1.41}$$

Finally, the MSE can be expressed as

$$\Gamma(\mathbf{c}) = \mathbf{c}^H \mathbf{A} \mathbf{c} - 2\mathrm{Re}\{\mathbf{c}^H \mathbf{b}\} + \sigma_a^2. \tag{1.42}$$

Now, from (1.42) it is possible to find the coefficients minimizing the MSE. This minimum exists and is unique, since the matrix A is positive definite

$$\mathbf{c}^H \mathbf{A} \mathbf{c} = E\{|y_k|^2\} = E\{|s_k + n_k|^2\} = E\{|s_k|^2\} + \sigma_n^2 \tag{1.43}$$

where s_k is the useful signal component. Thus, $\mathbf{c}^H \mathbf{A} \mathbf{c} > 0$, since $\sigma_n^2 > 0$, $\forall \mathbf{c} \neq 0$. The minimum can be found by setting to zero the gradient of $\Gamma(\mathbf{c})$. Hence, since

$$\nabla_{\mathbf{c}} \left(\mathbf{c}^H \mathbf{A} \mathbf{c}\right) = 2\mathbf{A}\mathbf{c}$$

$$\nabla_{\mathbf{c}} \left(2\mathrm{Re}[\mathbf{c}^H \mathbf{b}]\right) = 2\mathbf{b}$$

we have

$$\nabla_{\mathbf{c}}\Gamma = 2\left(\mathbf{A}\mathbf{c} - \mathbf{b}\right) = 0$$

from which

$$\mathbf{c}_0 = \mathbf{A}^{-1}\mathbf{b}$$

Computing now the minimum MSE

$$\Gamma_0 = \Gamma(\mathbf{c}_0) = \sigma_a^2 - \mathbf{b}^H\mathbf{A}^{-1}\mathbf{b} = \sigma_a^2 - \mathbf{b}^H\mathbf{c}_0. \tag{1.44}$$

This method, besides its theoretical validity, is useful when channel parameters are known to the receiver, or when it is possible to estimate such parameters. An adaptive algorithm for the adaptation of the coefficients can be obtained by using *steepest descent* algorithm which adaptively adjust the channel coefficients by

$$\mathbf{c}_{k+1} = \mathbf{c}_k - \frac{1}{2}\nabla_{\mathbf{c}}\Gamma\Big|_{\mathbf{c}=\mathbf{c}_k}. \tag{1.45}$$

This method allow to iteratively find the minimum MSE following the gradient of Γ in the space of all possible values of the set $\mathbf{c}$,

$$\nabla_{\mathbf{c}}\Gamma = 2\left(\mathbf{A}\mathbf{c} - \mathbf{b}\right) = 2\mathrm{Re}[\mathbf{z}_k^*\left(y_k - a_{k-D}\right)]. \tag{1.46}$$

All the variables involved in its computation in fact are available at the receiver, the vector $\mathbf{a}_k$ is present in the delay line, y_k is the output of the equalizer, a_{k-D} is available if a training sequence is used or can be approximated by the decisions $\hat{a}_{k-D}$. If the process is ergodic the expectation can be substituted by a temporal average, or simply removed to obtain the algorithm known as *least mean square* (LMS) or *stochastic gradient*

$$\mathbf{c}_{k+1} = \mathbf{c}_k - \alpha\mathbf{z}_k^*\left(y_k - a_{k-D}\right), \tag{1.47}$$

where α is a proper step-size. An initial training sequence can be used to make the algorithm converge, and then reliable decisions can be used to track slow channel variations. Similar considerations lead to an analogous algorithm for the estimation of the feedback coefficients $\mathbf{p}$ of a DFE equalizer:

$$\mathbf{p}_{k+1} = \mathbf{p}_k - \alpha\mathbf{a}_k^*\left(y_k - a_{k-D}\right), \tag{1.48}$$

where in this case y_k is the output of both FFE and DFE:

$$y_k = \mathbf{z}_k^T\mathbf{c} + \mathbf{a}_k^T\mathbf{p}\,.$$

1.3.2 Maximum likelihood sequence detection and Viterbi algorithm

From a theoretical point of view it is important to determine what is the performance limit of electronic processing for optical systems. *Maximum likelihood sequence detection* (MLSD) is known to provide the best performance in radio communications [26, 27], thus, even its feasibility in optical communications is not straightforward because of the computational load required for electronic processing, it has been implemented to demonstrate the capability of electronic compensation techniques at the receiver end to cope with fiber impairments. It consists in the maximization of the probability density function (pdf) of the received samples over the set of all possible transmitted symbols

$$\hat{\mathbf{a}} = \arg\max_{\mathbf{a}} p(\mathbf{z}|\mathbf{a}) \tag{1.49}$$

where $\mathbf{a} = \{a_k\}$ ad $\mathbf{z} = \{z_k\}$. If the received signal samples z_k are independent, it is possible to write

$$p(\mathbf{z}|\mathbf{a}) = \prod_k p(z_k|\mathbf{a}) \tag{1.50}$$

and in the hypothesis that the system is causal and with finite memory L we can write

$$p(z_k|\mathbf{a}) = p(z_k|a_k, a_{k-1}, \ldots, a_{k-L}) = p(z_k|a_k, \mu_k) \tag{1.51}$$

having defined $\mu_k = \{a_n\}_{n=k-L}^{k-1}$ as the receiver state. Hence, the implementation of the MLSE strategy can be easily performed by means of the Viterbi algorithm (VA) [25] working on a *trellis* diagram representing all possible states μ_k with branch metric

$$\lambda_k(a_k, \mu_k) = \ln p(z_k|a_k, \mu_k)\,. \tag{1.52}$$

Chapter 2

IM/DD systems

This chapter deals with the implementation of electronic processing techniques in standard, well known, IM/DD optical systems. Thus, the OOK amplitude modulation format is considered. In the linear regime GVD and PMD are the most severe sources of signal distortion and system penalty. Although GVD can be compensated for by DCF in present communications systems, compensation at bit rates higher than 10 Gb/s may be difficult because GVD impact increases with square of the bit rate, and a non-negligible residual dispersion may be left. Moreover, the decreased tolerances and the evolution of the transmission layer to a network layer can make the signal affected by an unpredictable, and even variable, residual GVD that combines with all-order PMD, which being an intrinsically stochastic phenomenon induces penalties difficult to be fully compensate for.

In a first-order approximation (see Chapt. 1), the effect of PMD is just a differential group delay $\Delta\tau$ between the two PSPs of the fiber, causing intersymbol interference. Customarily, PMD in a first-order approximation is independent of frequency. Higher order effects arise when the PMD vector $\vec{\Omega}$ varies with frequency [12]. Second-order effects are mainly signal distortion and broadening. It has been shown that optical compensation is able to recover heavy penalties, due to both first and second-order effects, through various possible implementations, such as the cascade of polarization controllers and polarization maintaining fibers [19], planar lightwave circuits [28], or other optical devices [29].

The mentioned techniques, although very effective, may be still impractical because of optical technology costs. Thus, many efforts have been done

to adapt and develop classical and novel electrical (post-detection) processing techniques in optical communication systems. The first proposal of an electrical equalizer for optical systems was a linear transversal filter to combat ISI arising from chromatic dispersion [30], but also non-linear cancellation was proposed since photodetection implies a non-linear transformation of the signal [31]. Much more recently, comparisons between these methods and optical compensation have been presented, evidencing benefits and drawbacks of both solutions [29, 32, 33]

Besides FFE and DFE equalization, there is an increasing interest in maximum likelihood sequence detection[1], implemented through the VA, due to its potentially optimal performance [34]. In the first work [30], this strategy is already considered, but of course the presence of optical amplifiers could not be accounted for. As a consequence, the ASE noise is not present in [30], and, thus, the statistics of the received signal, necessary to compute the VA branch metrics, are conditionally Gaussian due to shot and thermal noise. As said in Chapt. 1, nowadays all optical systems envisage the presence of optical amplifiers, and, as a consequence, the signal *in the fiber* is impaired by a noise that, in the linear regime, is modeled as AWGN [24]. Since a square-law detector is present at the receiver end, post-detection noise statistics change [35–37], and cannot be considered Gaussian anymore. In the case of the MLSD strategy, assuming Gaussian statistics for the noise after photodetection is neither realistic nor correct, and leads to inaccurate results [38]. Hence, in [29, 33] the statistics of the received signal are approximately measured and updated in real-time during transmission and assuming no decision errors. This method, which always refers to specific constraints, such as sample quantization, memory length, filter kind and parameters, or even absence of filtering, has been compared with classical equalization schemes, showing that a better performance, as expected, can be obtained by using the MLSD strategy [29, 33]. A rigorous detailed description of the MLSD approach can be found in a recent paper [39]. There a novel method for performance evaluation is devised and applied to the case of a synchronous MLSD receiver in the presence of chromatic dispersion only.

The purpose of this chapter is the description of the receiver structure for the optimal MLSD strategy with PMD and GVD; moreover, a reliable method

[1]The term *detection* is preferred instead of the commonly used *estimation* since the estimation theory refers to continuous parameters whereas we are interested in discrete sequences.

for computing the relevant performance for any possible value of the system parameters is provided, under the only constraint of one sample per bit, as in [39,40], or with no constraint on the sampling rate, since oversampling is the only way to guarantee sufficient statistics for this transmission system [41]. In particular, through numerical evaluation [42], a practically exact expression of the received signal statistics in the case of a receiver working with one sample per bit time is derived, so that a look-up table based MLSD receiver could be implemented [43]. In the case of oversampling, since the numerical method cannot be used, as a (neither exact nor approximate) expression for the joint statistics of the samples is not known, a histogram-based receiver is devised. An exhaustive analysis of the correlation of the received samples was carried out, in order to state the best trade-off between the number of samples to be processed and the amount of memory necessary to store a reliable estimate of the probability density functions.

Then, a closed-form approximation of the receiver branch metrics, which entails a negligible performance loss, is derived. Based on the exact branch metric computation, analytical bounds for the system performance are provided, allowing to reach values of *bit error rate* (BER) below 10^{-15}. This analytical method also represents an essential tool for optimizing the receiver parameters without resorting to time-consuming computer simulations. Since PMD is a time-varying phenomenon, the receiver has also to adaptively update some parameters. This aspect is also discussed. A comparison with commonly adopted electronic equalization and optical compensation techniques is also provided, showing that the MLSD approach achieves, as expected, a better performance when compared to other electrical techniques, although optical compensation still provides the best results since, after the irreversible transformation introduced by the photodetector, the receiver post-detection processing is not able to effectively cope with the combined GVD and PMD distortions. The results reported in this chapter are described in [44, 45].

2.1 System model

Fig. 2.1 shows the system model and its low-pass equivalent. A standard NRZ OOK modulated laser beam is launched in a single mode fiber, optically amplified and filtered at the receiver end. The optical amplifier is assumed to have high gain G, so that ASE noise is dominant over thermal and shot noise. The signal is then photodetected, electrically filtered, sampled, and finally pro-

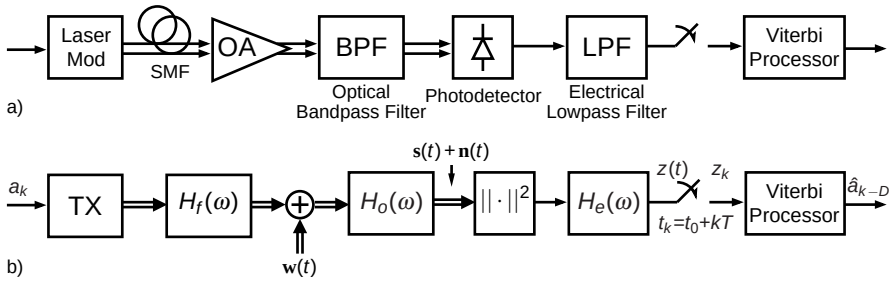

Figure 2.1: (a) Schematic of the system model. (b) Low-pass equivalent.

cessed through the VA, effectively implementing the MAP sequence detection strategy, here equivalent to MLSD, being the transmitted bits $\{a_k\}$ equally likely. In Fig. 2.1, $\mathbf{w}(t) = [w_1(t), w_2(t)]^T$ represents AWGN, where $w_1(t)$, $w_2(t)$ are independent complex noise components accounting for ASE on two orthogonal SOPs. At the optical filter output, the components of the two-dimensional complex vectors $\mathbf{s}(t) = [s_1(t), s_2(t)]^T$ and $\mathbf{n}(t) = [n_1(t), n_2(t)]^T$ represent the useful signal and noise components in each SOP, respectively. The noise components are Gaussian but not white, since they are obtained by filtering the AWGN $\mathbf{w}(t)$. The low-pass equivalent (matrix) transfer functions of fiber, optical and post-detection filters are denoted by $H_f(\omega)$, $H_o(\omega)$, and $H_e(\omega)$, respectively. In a chromatic-dispersion-compensated link, several fiber pieces with alternating sign chromatic dispersion D_i and appropriate length L_i may be used, and as commonly done, we will indicate as "residual dispersion" the quantity $D_r = \sum_i D_i L_i$ in picoseconds per nanometer. Using the dimensionless chromatic dispersion index γ with D_r, the transfer function that will be used in the following is expressed by (1.17).

The fiber Jones matrix is then written as $\mathbf{H}_f(\omega) = H_{GVD}(\omega)\mathbf{R}\mathbf{U}(\omega)\mathbf{R}^{-1}$ (see Chapt. 1).

A closed-form expression of $\mathbf{U}(\omega)$ accounting for all PMD orders is not known, and although several second-order approximations are available (see §1.2.2). To avoid the peculiarities of a specific analytical model, one could use the numerical random waveplate model, but in this case the evaluation of outage probabilities would become exceedingly expensive, so we choose to evaluate outages by using the model in [15] for $\mathbf{U}(\omega)$ as it is also able to partially (but correctly) account for higher PMD orders through second-order parameters only [21].

At the output of the photodiode the detected signal can be described as the sum of two contributions, one for each SOP:

$$z(t) = \|\mathbf{s}(t) + \mathbf{n}(t)\|^2 = |s_1(t) + n_1(t)|^2 + |s_2(t) + n_2(t)|^2 . \tag{2.1}$$

Clearly, after photo-detection the noise becomes signal-dependent and its statistics change. In the following, optical and post-detection filter parameters can be chosen arbitrarily, since the proposed receiver is independent of a particular choice for filter shapes or bandwidths.

2.2 Received signal statistics

In this section, we derive the MLSD detection strategy under the constraints on the above receiver structure and that one sample per bit interval is extracted at the receiver. This receiver cannot, however, be considered as optimal, since the above mentioned received samples do not represent a sufficient statistic for this detection problem. The problem of finding a sufficient statistic will be faced in the next section.

Under the above mentioned constraints, the MLSD detection strategy can be expressed as [35]

$$\hat{\mathbf{a}} = \arg \max_{\mathbf{a}} p(\mathbf{z} \mid \mathbf{a}), \tag{2.2}$$

being $\mathbf{a} = \{a_k\}$ the transmitted bit sequence and $\mathbf{z} = \{z_k\}$, the corresponding received sequence. The synchronous received samples are $z_k = z(t_0 + kT)$, t_0 being a proper time offset and T the bit interval. We assume[2] that, conditional upon the transmitted sequence, the samples $\{z_k\}$ can be taken as independent. Since this is the conditional pdf, this hypothesis depends on the shape of used optical and postdetection filters, irrespective of the presence of ISI due to PMD and GVD. Then the conditional joint pdf of the received samples can be written as

$$p(\mathbf{z} \mid \mathbf{a}) = \prod_k p(z_k \mid \mathbf{a}) \tag{2.3}$$

and, assuming that the system is causal and m is causal and with finite memory L, it is

$$p(z_k \mid \mathbf{a}) = p(z_k \mid a_k, a_{k-1}, ..., a_{k-L}). \tag{2.4}$$

[2]This assumption was numerically validated for the commonly used shape and bandwidth of the optical and electrical filters.

Therefore the optimal MLSD strategy can be implemented by means of the VA using the following branch metrics:

$$\lambda_k(a_k, \mu_k) = \ln p(z_k \mid a_k, \mu_k) \tag{2.5}$$

where $\mu_k = (a_{k-1}, a_{k-2}, ..., a_{k-L})$ is the trellis state. So, the number of states is $S = 2^L$, and hence the receiver complexity increases exponentially with the channel memory L.

A closed-form expression for the pdfs in (2.4) is not known for arbitrary signal formats and filtering. Although the samples at the photodetector output have a non-central chi-square distribution, being the sum of squared Gaussian random variables [35], the presence of the electrical filter modifies these statistics. An appropriate characterization of these pdfs is mandatory, otherwise the performance of the MLSD receiver would be degraded. These pdfs can be evaluated almost exactly by efficient numerical methods and stored in a look-up table that can be addressed, in order to compute the branch metrics, by the received signal samples and by the considered trellis transition.

Undoubtedly, the most efficient numerical method is that using the "saddlepoint" approximation [46], based on the knowledge of the moment generating function $\Psi_{z_k|\mathbf{a}}(s)$ of the samples. This function can be obtained as a closed-form expression by expanding the noise on a proper Karhunen-Loève basis, as shown in [21]. Then, using the saddlepoint approximation, the pdf can be evaluated as

$$p(z_k \mid \mathbf{a}) \simeq \frac{\exp\left[\Phi_{z_k|\mathbf{a}}(s_0)\right]}{\sqrt{2\pi\Phi''_{z_k|\mathbf{a}}(s_0)}}, \tag{2.6}$$

where s_0 is the saddle-point of $\Psi_{z_k|\mathbf{a}}(s)\exp(-sz_k)$ on the real axis,

$$\Phi_{z_k|\mathbf{a}}(s) = \log[\Psi_{z_k|\mathbf{a}}(s)\, e^{-sz_k}] \tag{2.7}$$

and $\Phi''_{z_k|\mathbf{a}}$ is the second derivative of $\Phi_{z_k|\mathbf{a}}$, which is always positive at the saddlepoint. This approach gives a really accurate closed-form approximation for the pdf and, although requiring a search for the saddlepoint, provides an exact method to evaluate the theoretical performance limit for electrical equalization.

When the signal is not distorted, i.e., it is ideal rectangular NRZ, the optical filter has a large bandwidth rectangular-shaped transfer function, and the post-detection filter is an integrate and dump device, the pdf of the received sample is still of chi-square type [37]. Under the hypothesis that the optical

and electrical filters simply influence the number of degrees of freedom and that signal distortion and filtering can be accounted for only through the induced change in signal energy, we can approximate the conditional pdf of a received sample as [38, 39, 47]

$$p(z_k|a_k,\mu_k) \simeq \frac{1}{N_0}\left(\frac{z_k}{s_k}\right)^{(\nu-1)/2} \exp\left(-\frac{z_k+s_k}{N_0}\right) I_{\nu-1}\left(\frac{2\sqrt{z_k s_k}}{N_0}\right) \tag{2.8}$$

where ν, half the number of degrees of freedom, is taken as two times[3] the ratio of the optical and electrical filter noise equivalent bandwidths, i.e., $\nu = 2B_o/B_e$, $s_k \doteq s_R(a_k,\mu_k)$ is the noise-free received sample which depends on the present and past transmitted symbols, according to the channel memory length, and $I_{\nu-1}(x)$ is the modified Bessel function of first kind and order $\nu-1$. Hence, a simplified expression of the branch metrics, discarding irrelevant terms in the maximization, is[4]

$$\lambda(a_k,\mu_k) \simeq -\frac{\nu-1}{2}\ln[s_R(a_k,\mu_k)] - \frac{s_R(a_k,\mu_k)}{N_0} + \ln\left[I_{\nu-1}\left(\frac{2\sqrt{z_k s_R(a_k,\mu_k)}}{N_0}\right)\right]. \tag{2.9}$$

As we will see in the numerical results section, a receiver based on these branch metrics practically attains the same performance of the receiver based on exact branch metrics.

Notice that this same approach is also taken in [38] to compute the performance of optical turbo coded systems but, differently from [38], we use (2.8) both when $a_k = 1$ or $a_k = 0$, as, due to finite extinction ratio, signal distortion and filtering, $s_R(a_k,\mu_k)$ will not be exactly zero even when $a_k = 0$. Clearly, when $a_k = 0$, (2.8) can be used only if $s_R(a_k,\mu_k) > 0$, i.e., when the electrical filter is such that its impulse response either is always positive or its negative values have a negligible impact, as is the case for a Gaussian or Bessel type filters, for example.

The exact (solid), chi-square (dashed), and Gaussian (dotted) pdf approximations are reported in Fig. 2.2 for $E_b/N_0 = 18\,\text{dB}$, where E_b is the received optical signal mean energy per bit. We considered the received samples normalized to the value corresponding to the logical "1". The ratio E_b/N_0

[3] The factor 2 is due to the fact that we consider signal and noise as complex rather than real valued as done in [37–39].

[4] All additive terms in the branch metrics independent of a_k and σ_k can be discarded and, in addition, the branch metrics can be arbitrarily multiplied by a positive constant.

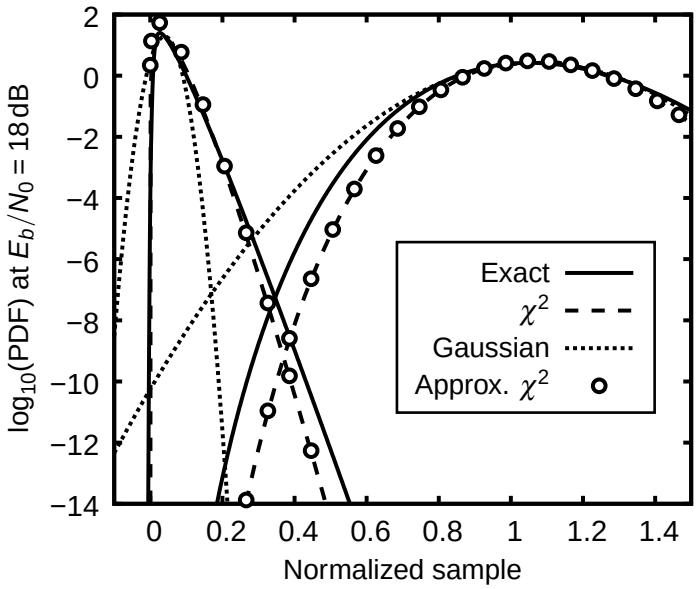

Figure 2.2: pdfs of '1' and '0' for $E_b/N_0 = 18$ dB.

represents the number of detected photons per bit at the input of the optical amplifier, and is related to the optical signal-to-noise ratio (OSNR) by $E_b/N_0 = 2 \cdot \mathrm{OSNR} \cdot WT$, W being the reference measurement bandwidth, usually taken equal to 12.5 GHz (0.1 nm). As it can be seen, the chi-square is a better approximation to the actual pdf than the Gaussian one, which completely fails in estimating the tails, even if in a standard receiver it can predict the BER with good accuracy [37]. A further simplification can be obtained by the crude approximation $I_{\nu-1}(x) \simeq e^x/\sqrt{2\pi x}$ [48], by which (2.8) is approximated as

$$p(z_k|a_k,\mu_k) \simeq \left(\frac{z_k}{s_k}\right)^{\frac{2\nu-1}{4}} \frac{\exp\left(-\frac{\left(\sqrt{z_k}-\sqrt{s_k}\right)^2}{N_0}\right)}{\sqrt{4\pi N_0 z_k}}, \tag{2.10}$$

reported in Fig. 2.2 as open circles. The resulting simplified branch metrics are

$$\lambda(a_k,\mu_k) \simeq 2\sqrt{z_k s_R(a_k,\mu_k)} - s_R(a_k,\mu_k) - \frac{N_0}{2}(\nu-\frac{1}{2})\ln[s_R(a_k,\mu_k)]\,. \tag{2.11}$$

The impact of these approximations in the expression of the branch metrics will be considered in the numerical results section.

As already mentioned, the number of trellis states, and thus the complexity, depends exponentially on the channel memory L. For commonly used optical and electrical filters, and DGD values lower than a bit interval, we verified that $L \leq 2$ in the absence of chromatic dispersion, whereas $L \leq 4$ for the residual dispersion values we took into account. Hence the number of states is at most $S = 16$. In addition, the application of reduced-state sequence detection (RSSD) techniques [49] allows to substantially reduce the number of trellis states. In particular, a reduced state $\mu'_k = (a_{k-1}, a_{k-2}, \ldots, a_{k-L'})$, with $L' < L$, may be defined. The resulting number of states is reduced to $2^{L'} < 2^L$. In order to compute the branch metrics (2.5) in a reduced trellis, the necessary symbols not included in the state definition may be found in the survivor history [49]. We note that, in the limiting case of $L' = 0$, the trellis diagram degenerates and symbol-by-symbol detection with decision feedback is performed. The resulting receiver can be considered as a non linear equalizer with decision feedback.

Since PMD is a time-varying phenomenon, the receiver parameters should be adaptively updated. By using one of the above mentioned approximated closed-form expressions of the branch metrics (2.9) or (2.11), when PMD changes, the receiver has to simply adaptively identify the term $s_R(a_k, \mu_k)$. This can be easily done by using a gradient adaptation algorithm and, as a cost function, the one defining the nonlinear branch metrics.

2.3 Oversampling

Although the MLSD receiver described in the previous section represents the best post-detection technique in the case of synchronous sampling, as already mentioned, one sample per bit time is not a sufficient statistic for the problem at hand. The simpler way to obtain a sufficient statistic, having in mind the practical implementation of the receiver, is through oversampling [41]. In fact, provided that an adequate number of samples per bit is extracted from the signal after photo-detection, the whole information in the electrical signal is preserved and there is no need of a further processing through an electrical filter [41]. As a consequence, this latter filter was removed. In this section, we investigate the performance improvement obtained by adopting the MLSD strategy jointly with the oversampling technique.

Assuming that n samples per bit time are used, we now resort to the following notation to denote the received samples. The n received samples related to the k-th bit interval will be denoted as $z_{k,i} = z(t_0 + kT + iT/n)$, $i = 0, 1, \ldots, n-1$. As in the previous section, the received sequence will be denoted as $\mathbf{z} = \{z_{k,i}\}$. In addition, we will denote by $\mathbf{z}_k$ the n signal samples related to the k-th bit, i.e., $\mathbf{z}_k = \{z_{k,i}\}_{i=0}^{n-1}$. When samples are spaced less than a bit time, conditionally to the transmitted bit sequence they cannot be considered as independent, in principle. As a consequence, their joint pdf is not given by the product of the marginal pdfs. For this reason we use the chain rule to factorize the joint pdf $p(\mathbf{z}|\mathbf{a})$ necessary to implement the MLSD strategy. Assuming, as in the previous section, that received samples which differ for at least one bit interval are independent, we have [50]

$$p(\mathbf{z}|\mathbf{a}) = \prod_k p(\mathbf{z}_k \mid \mathbf{z}_{k-1}, \mathbf{a}) \prod_k \prod_{i=0}^{n-1} p(z_{k,i} \mid z_{k,i-1}, \ldots, z_{k,0}, \mathbf{z}_{k-1}, \mathbf{a}) \qquad (2.12)$$

and, in the last expression, it is implicitly assumed that, if $i = 0$, terms $z_{k,i-1}, \ldots, z_{k,0}$ disappear. Hence, with an appropriate definition of the receiver state[5] μ_k, the branch metrics of the Viterbi algorithm implementing the MLSD strategy can be expressed as [50]

$$\lambda_k(a_k, \mu_k) = \ln p(\mathbf{z}_k \mid \mathbf{z}_{k-1}, a_k, \mu_k) \sum_{i=0}^{n-1} \ln p(z_{k,i} \mid z_{k,i-1}, \ldots, z_{k,0}, \mathbf{z}_{k-1}, a_k, \mu_k). \qquad (2.13)$$

Since it was not possible to find a way to analytically evaluate the pdfs in (2.13) when the samples are not conditionally independent, reliable estimation of the channel statistics necessary to compute the branch metrics were obtained through simulation, implementing a histogram counter at the receiver. The receiver structure in the case of oversampling ($n > 1$) is shown in Fig. 2.3.

The joint pdfs derived from the above mentioned histograms allowed the evaluation also of the impact of the correlation between samples on the branch metrics and the optimum sampling rate. In the numerical results section, it will be shown that a value of $n = 2$ is practically sufficient to attain the optimal performance. In this case, although the received samples are correlated, by

[5] In principle, in the case of oversampling, the memory of the system will be $M \geq L$. Hence, in this case, the state is defined as $\mu_k = (a_{k-1}, a_{k-2}, \ldots, a_{k-M})$.

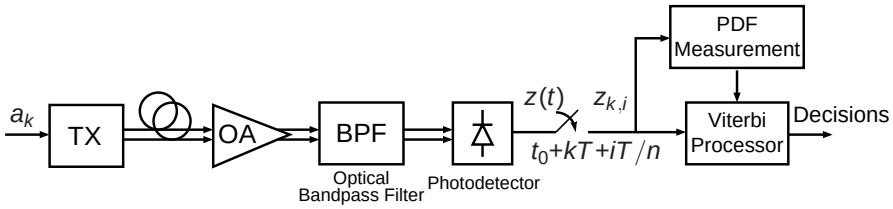

Figure 2.3: System model for the receiver based on oversampling.

considering them as independent in expressing the branch metrics, the same performance that can be obtained with the optimal correlated metrics could be achieved. Hence, the following simplified branch metrics can be adopted with no performance degradation

$$\lambda(a_k, \sigma_k) \simeq \sum_{i=0}^{1} \ln p(z_{k,i} \mid a_k, \mu_k). \tag{2.14}$$

The pdfs appearing in (2.14) can be further simplified by using the approximate expression (2.8) or (2.10).

It is worth mentioning that, although the described branch metrics have been derived having in mind the MAP sequence detection strategy, as demonstrated in [50] the same branch metrics can be used for the BCJR algorithm [51] implementing the MAP symbol detection strategy and employed as component decoder in iterative decoding schemes for optical turbo codes, such as those proposed in [38].

2.4 Channel identification

Some parameters of the distribution of the received samples depend not only on the transmitted pulse shape, filter types and bandwidths, but also on the channel impairments and on the noise power spectral density. Due to the square-law nature of a photodiode, the estimation of the required channel parameters is not straightforward. We propose an effective and simple algorithm which, compared to the ideal case of perfect channel knowledge, leads to a negligible performance loss.

To illustrate the proposed technique for channel and noise PSD estimation, we consider the MLSD receiver working with two samples per bit time, since it

practically represents the optimal electronic processing.[6] In this case, a post-detection electrical filter is not necessary. Hence, the photodetected signal can be expressed as

$$z(t) = |s_1(t) + n_1(t)|^2 + |s_2(t) + n_2(t)|^2 = s_z(t) + n_z(t) \tag{2.15}$$

having defined the useful signal $s_z(t)$ and noise term $n_z(t)$ as

$$s_z(t) \doteq |s_1(t)|^2 + |s_2(t)|^2 \tag{2.16}$$

$$n_z(t) \doteq |n_1(t)|^2 + |n_2(t)|^2 + 2\mathrm{Re}[s_1(t)n_1^*(t) + s_2(t)n_2^*(t)] \tag{2.17}$$

where $\mathrm{Re}(\cdot)$ returns the real part of its argument. The signal-dependent and colored noise $n_z(t)$ is neither Gaussian nor zero-mean. Indeed, it can be shown that its mean value is $\eta \doteq E\{n_z(t)\} = 2N_0B_o$, the couple of samples used by the receiver in the k-th bit interval can be expressed as

$$\{z_{k,i} \doteq z(t_0 + kT + iT/2),\ i = 0, 1\}\ . \tag{2.18}$$

In order to compute all the VA branch metrics at discrete time k, the receiver uses these two samples but also needs to estimate both the noise PSD N_0 (or η, equivalently) and all possible values of the useful components $s_z(t_0 + kT + iT/2)$, $i = 0, 1$, of the received samples. Being L the channel dispersion length, these useful components are functions of the current bit a_k and of the L previous bits $\mu_k \doteq (a_{k-1}, a_{k-2}, \ldots, a_{k-L})$ defining the trellis state. Assuming for the moment that the channel is time-invariant,[7] these functions will be denoted to as $f_i(a_k, \mu_k)$. In order to compute the VA branch metrics, the receiver needs to estimate N_0 (or η) and the two values $f_i(a_k, \mu_k)$, $i = 0, 1$, for each branch defined by the couple (a_k, μ_k).

2.4.1 Estimation of the channel parameters

We tried to estimate the optical ISI channel coefficients whose non linear transformation due to the square-law detection gives the values $f_i(a_k, \mu_k)$, but we did not obtain satisfactory results. Hence, we decided to directly estimate the $2 \times 2^{L+1}$ useful signal components $f_i(a_k, \mu_k)$ to be employed in the expression of the branch metrics.

[6] In any case, the proposed techniques can be straightforwardly extended to the MLSD receiver working with one sample per bit time.

[7] Tracking of a time-varying channel will be discussed in the next section.

We also investigated different strategies to perform the channel noise PSD estimation. Among others, we considered a solution derived from the generalized likelihood strategy [52], consisting of the maximization of the conditional joint pdfs of the received samples given the transmitted bits and the channel parameters. However, we discovered that, due to the presence of different local maxima, the convergence of the resulting algorithm is very difficult. For this reason, we resorted to the classical minimization of the MSE between the received samples and the useful signal components, modified in order to take into account the non zero-mean value of the noise component (2.17). Hence, the updating law for the useful signal components is based on the LMS algorithm which minimizes the mean square value

$$\sum_{i=0}^{1} E\{[z_{k,i} - \hat{\eta} - \hat{f}_i(a_k, \mu_k)]^2\} \tag{2.19}$$

where $\hat{\eta}$ and $\hat{f}_i(a_k, \mu_k)$ are the estimates of η and $f_i(a_k, \mu_k)$, respectively. Notice that for the application of the LMS method it is irrelevant that the noise $\eta_z(t)$ is not Gaussian, signal-dependent, and colored. On the other hand the presence of the mean value η needs to be properly taken into account. For the moment, we assume that the transmitted bits are perfectly known to the receiver. The sequence of transmitted bits will be denoted to as $\{\overline{a}_k\}$ and the corresponding sequence of trellis states as $\{\overline{\mu}_k\}$. Denoting by $\hat{\eta}^{(k)}$ and $\hat{f}_i^{(k)}(a_k, \mu_k)$ the values assumed by the estimates of η and of the useful signal components at discrete time k, following the LMS algorithm [35] the channel coefficients are recursively updated as $(i = 0, 1)$

$$\begin{cases} \hat{f}_i^{(k+1)}(a_k, \mu_k) = \hat{f}_i^{(k)}(a_k, \mu_k) & \text{for } (a_k, \mu_k) \neq (\overline{a}_k, \overline{\mu}_k) \\ \hat{f}_i^{(k+1)}(a_k, \mu_k) = \hat{f}_i^{(k)}(a_k, \mu_k) & \\ \quad +\alpha[z_{k,i} - \hat{\eta}^{(k)} - \hat{f}_i^{(k)}(a_k, \mu_k)] & \text{for } (a_k, \mu_k) = (\overline{a}_k, \overline{\mu}_k) \end{cases} \tag{2.20}$$

where $\alpha > 0$ is a scale factor controlling the amount of adjustment (the so-called "step-size"). Clearly, at every discrete time k, we can update only the estimates $\hat{f}_i(\overline{a}_k, \overline{\mu}_k)$ for the actual couple $(\overline{a}_k, \overline{\mu}_k)$, that is those matching the transmitted sequence. Hence, in practice, the update of a given coefficient is not performed at a fixed rate but at a random rate, depending on the transmitted pattern. The estimates in (2.20) are initialized with the value in the absence of PMD and GVD.

The update of $\hat{\eta}$ cannot be done using the same method. In fact, if we consider the mean square error (2.19), it has different minima. For example, a minimum can be obtain when $\hat{\eta} = \eta$ and $\hat{f}_i(a_k, \mu_k) = f_i(a_k, \mu_k)$, but also when $\hat{\eta} = 0$ and $\hat{f}_i(a_k, \mu_k) = f_i(a_k, \mu_k) + \eta$. Hence, we must resort to a different approach. From (2.15), (2.16), and (2.17), it can observed that, neglecting the terms depending on η^2, it is

$$E\{[z_{k,i} - f_i(\overline{a}_k, \overline{\mu}_k) - \eta]^2\} \simeq f_i(\overline{a}_k, \overline{\mu}_k)\eta\,. \tag{2.21}$$

We use this property to refine the estimates of η by substituting the expectation with a temporal mean. The update of η is performed every N bits (that is when k is a multiple of N), where N is a parameter that can be tuned to reduce fast estimate fluctuations ($N \simeq 100$ is sufficient). Furthermore, the temporal mean is computed only when $\overline{a}_k = 1$, to avoid numerical problems with quantities close to zero. Hence, at time k, with k a multiple of N, we estimate

$$\hat{\eta}^{(k+1)} = \frac{1}{2|\mathcal{N}|} \sum_{\substack{i=0,1 \\ n\in\mathcal{N}}} \frac{[z_{k-n,i} - \hat{\eta}^{(k-n)} - \hat{f}_i^{(k-n)}(\overline{a}_{k-n}, \overline{\mu}_{k-n})]^2}{\hat{f}_i^{(k-n)}(\overline{a}_k, \overline{\mu}_k)} \tag{2.22}$$

where $\mathcal{N} = \{n = 0, 1, \ldots, N-1 | \overline{a}_{k-n} = 1\}$ and $|\mathcal{N}|$ denotes the cardinality of this set. Note that in (2.22) the estimates $\hat{\eta}^{(k-n)}$, $n \in \mathcal{N}$, take the same value since the estimate of η is updated every N bits. From the estimate of η we can then simply derive the estimate of N_0 to be used in the branch metrics.

Up to now, we have considered the case of a time-invariant channel. However, when the channel is (slowly) time-varying, the algorithm will be able to track the channel variations, due to its adaptive nature.

In the operating conditions, the transmitted bits are not available. Hence, we resort to a decision directed approach, that is we use the decisions of the VA entailing no performance loss at least when the bit error rate is lower than 10^{-2}. The decisions of the VA are available with a delay of D symbols which can compromise the capability of the algorithm in tracking the fast channel variations. However, this is not the case of the considered channel since the dispersion we are facing is a slowly-varying phenomenon. For the same reason it is not necessary to update the channel parameters at each bit interval but they can be updated less frequently, depending on the considered channel.

Considering now the initial acquisition of the algorithm, although a training sequence can be adopted to reduce the acquisition time, we verified by

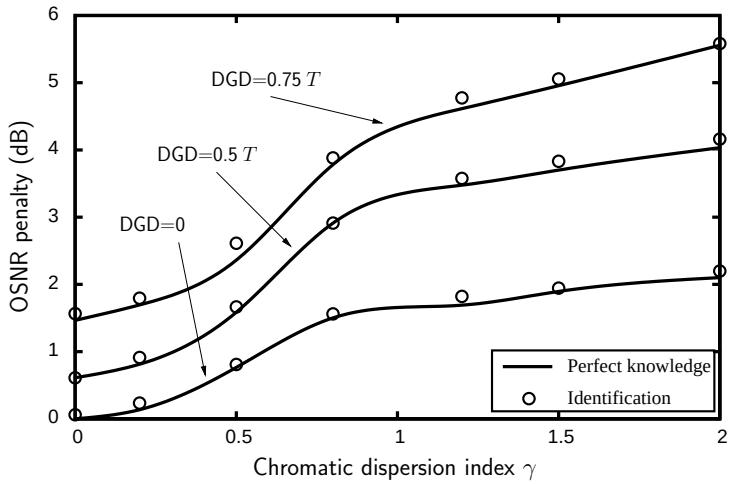

Figure 2.4: OSNR penalty curves at BER=10^{-4}.

computer simulations that the algorithm is able to work in a completely blind manner, as discussed in the following Section.

2.4.2 Numerical results on channel identification

We consider an optical channel affected by first order PMD (with a worst-case power splitting equal to 0.5), and GVD. The optical filter is modeled as a fourth-order Butterworth type with bandpass bandwidth equal to $1.1/T$ and the MLSD receiver is based on a Viterbi algorithm working on, at most, a 128-state trellis. In Fig. 2.4 we show the OSNR penalty, obtained by computer simulations, for increasing values of γ, and for different values of the PMD differential group delay values. The solid lines refer to the ideal receiver and are obtained under the hypothesis of perfect knowledge of the channel parameters knowledge, whereas the circles represent the performance of the receiver when using the described estimation algorithm. It can be seen that no loss in performance is entailed.

We now show that the algorithm is able to work in the decision directed mode in a completely blind manner. We consider a particular case of bad channel impairment, where the chromatic dispersion index γ is equal to 1. At the beginning and for the first 10,000 bits the algorithm is turned off. In this case, for an OSNR (referred to a bandwidth equal to the bit rate) of 13 dB

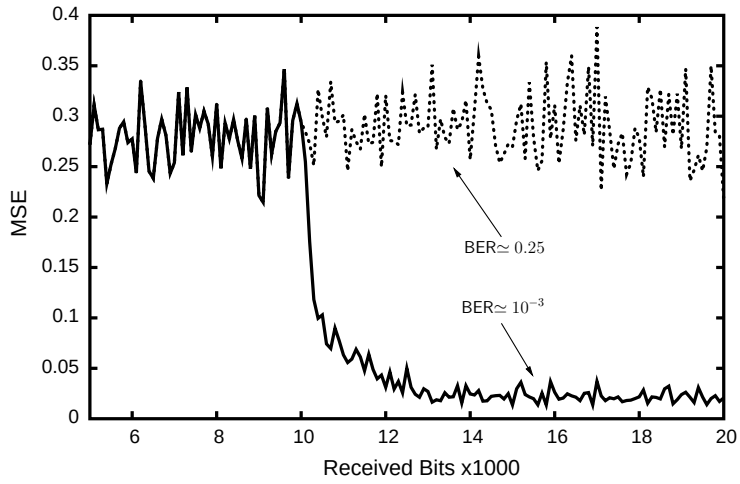

Figure 2.5: MSE transient when the estimation algorithm is turned on.

we have a bit error rate of 0.25. Hence, we do not have reliable decisions. Nonetheless, when the estimation algorithm is turned on, it is still able to converge to a BER of 10^{-3} (the value corresponding to perfect knowledge of the channel parameters) in about $2,000 \div 2,500$ bits, as can be seen from Fig. 2.5 showing the time evolution of the MSE.

2.5 Lower and upper bounds

It is well known that the classical union upper bound on the bit error probability P_b has expression [35]

$$P_b \leq \sum_{\mathbf{a}} P(\mathbf{a}) \sum_{\hat{\mathbf{a}} \neq \mathbf{a}} b(\mathbf{a}, \hat{\mathbf{a}}) P(\mathbf{a} \rightarrow \hat{\mathbf{a}}) \tag{2.23}$$

in which $\mathbf{a} = \{a_k\}$ and $\hat{\mathbf{a}} = \{\hat{a}_k\}$ denote the bit sequences corresponding to the correct and erroneous paths, respectively, $b(\mathbf{a}, \hat{\mathbf{a}})$ is the number of bit errors entailed by the considered error event $(\mathbf{a}, \hat{\mathbf{a}})$, $P(\mathbf{a} \rightarrow \hat{\mathbf{a}})$ is the pairwise error probability (PEP) and $P(\mathbf{a})$ is the a priori probability of sequence $\mathbf{a}$. The PEP $P(\mathbf{a} \rightarrow \hat{\mathbf{a}})$ is the probability that the sum of the branch metrics relative to the erroneous path exceeds the sum of the branch metrics on the correct

path:

$$P(\mathbf{a} \to \hat{\mathbf{a}}) = P\left(\sum_{k=\ell}^{\ell+H} \lambda_k < \sum_{k=\ell}^{\ell+H} \hat{\lambda}_k\right) \tag{2.24}$$

having denoted by $\{\lambda_k\}$ and $\{\hat{\lambda}_k\}$ the branch metrics corresponding to the correct and erroneous path, respectively, and by H the length of an error event starting at discrete time ℓ. In Fig. 2.6, two error events starting at the same instant are considered on a 4-state trellis. The correct bit sequence is the "all zero" sequence, whereas the erroneous sequences are $\hat{\mathbf{a}}_0$ and $\hat{\mathbf{a}}_1$, respectively. The error event has length $H = 2$ in the first case and $H = 3$ in the second one.

From the union bound, a lower and an approximated upper bound for the bit error probability are derived. The lower bound is simply obtained by considering the most likely error event and assuming that only one bit error characterizes it. The approximated upper bound is obtained by truncating the union bound considering a few most frequently occurring error events only.

Let us consider, for example, the case of the receiver working with one sample per bit interval. The event $\{\sum_{k=\ell}^{\ell+H} \lambda_k < \sum_{k=\ell}^{\ell+H} \hat{\lambda}_k\}$ which is involved in the computation of the PEP, can be equivalently expressed as the event that the vector of the received samples $\mathbf{z}_\ell^{\ell+H} = (z_\ell, z_{\ell+1}, \ldots, z_{\ell+H})^T$ belongs to a given domain. Hence, the numerical integration of the joint pdf allows to compute the PEP with high accuracy for values down to 10^{-15}. In this case, the integration can be easily performed since the received samples are independent and their marginal pdf can be obtained by the same method described in the previous section and based on the saddle-point approximation. However, we will shortly see a simpler way for evaluating the PEP which also holds when oversampling is employed.

Indeed, in the case of oversampling receivers, it is not possible to compute the PEP along the previous lines, since the joint pdf of the received samples is not known. In fact, even if it can be approximated as the product of the marginal pdfs for the expression of the branch metrics, this cannot be done for the computation of the PEP if samples are spaced by less than a bit interval. However, we can exploit the fact that, as it can be seen from (2.9) and (2.11), the branch metrics depend only on the square root of the samples in the following way. Using (2.11) and letting

$$y_k = \sqrt{z_k}\,, \tag{2.25}$$

$$d_k = \sqrt{s_k} - \sqrt{\hat{s}_k}\,, \tag{2.26}$$

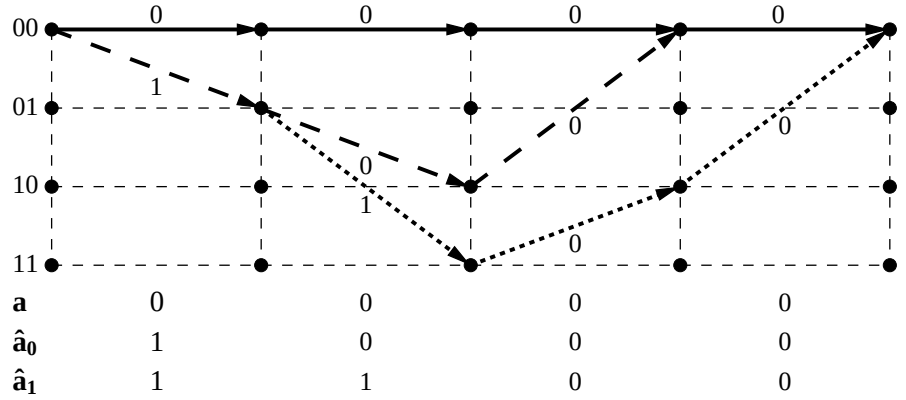

Figure 2.6: Trellis transitions given by wrong sequences $\hat{\mathbf{a}}_0$ and $\hat{\mathbf{a}}_1$.

the PEP in (2.24), neglecting the term $(2\nu-1)N_0\log(s_k/\hat{s}_k)/4$ at high signal-to-noise ratios, can be written as

$$P(\mathbf{a}\to\hat{\mathbf{a}}) = P\left(\sum_k d_k y_k < \frac{1}{2}\sum_k (s_k-\hat{s}_k)\right). \tag{2.27}$$

As shown in Appendix A, the random variable (r.v.) y_k in (2.25) can be approximated as Gaussian with mean and variance, respectively,

$$\eta_{y_k} = \alpha_k\sqrt{s_k}\,, \tag{2.28}$$
$$\sigma^2_{y_k} = N_k/2\,, \tag{2.29}$$

where

$$\alpha_k = \frac{s_k/N_k + (4\nu_k^2-1)/16}{s_k/N_k + (2\nu_k-1)(2\nu_k-3)/16}\,, \tag{2.30}$$

$$N_k = \frac{\sigma^2_{z_k}}{\eta_{z_k}+s_k}\,, \tag{2.31}$$

$$\nu_k = \frac{\eta^2_{z_k}-s_k^2}{\sigma^2_{z_k}}\,, \tag{2.32}$$

η_{z_k} and $\sigma^2_{z_k}$ being the mean and the variance of z_k. Notice that for the over-sampling case, as the post-detection filter is absent, the photodetected samples *are* chi-square distributed, such that (2.31) and (2.32) are simply equal

to $N_k = N_0 B_o$ and $\nu_k = 2$, where B_o is the optical filter noise equivalent bandwidth.

Hence, the r.v.

$$x = \sum_k d_k y_k \tag{2.33}$$

can be approximated, in turn, as Gaussian with mean and variance

$$\eta_x = \sum_k d_k \eta_{y_k} \tag{2.34}$$

$$\sigma_x^2 = \sum_k d_k^2 \sigma_{y_k}^2 + 2 \sum_{\substack{k,\ell \\ k \neq \ell}} \rho_{k\ell} d_k d_\ell \sigma_{y_k} \sigma_{y_\ell}, \tag{2.35}$$

where $\rho_{k\ell}$, the correlation coefficient of y_k and y_ℓ, can be taken to be zero if y_k and y_ℓ correspond to samples spaced by at least a bit time, otherwise it can be computed as shown in Appendix B. Therefore, the PEP can be approximated as

$$P(\mathbf{a} \to \hat{\mathbf{a}}) \simeq \frac{1}{2} \operatorname{erfc} \left(\frac{\eta_x - \frac{1}{2} \sum_k (s_k - \hat{s}_k)}{\sqrt{2}\, \sigma_x} \right), \tag{2.36}$$

which, for the synchronous case ($\rho_{k\ell} \simeq 0\,,\ \forall\, k, \ell$), becomes

$$P(\mathbf{a} \to \hat{\mathbf{a}}) \simeq \frac{1}{2} \operatorname{erfc} \left(\frac{\sum_k \left[(2\alpha_k - 1)\sqrt{s_k} - \sqrt{\hat{s}_k} \right] d_k}{2\sqrt{\sum_k N_k d_k^2}} \right). \tag{2.37}$$

Notice that a similar result was also found in [39] where, however, a different approximation to the χ^2 distribution was used in deriving the branch metrics. It is also worth noting that the results in [39] are only valid for low optical extinction ratios (10 dB or lower) and when the optical filter has a large bandwidth and the post-detection filter is an integrate-and-dump device, whereas our results hold for extinction ratios as high as 20 dB and for almost arbitrary filtering. The accuracy of the lower and upper bounds will be examined in the numerical result section.

2.6 Numerical results

Standard Monte Carlo simulations were carried out in order to evaluate the receiver performance with respect to other electrical and optical equalization

techniques. All BER curves are assessed in terms of E_b/N_0. In all simulations, except those related to the receiver based on oversampling, the optical filter is assumed to be a fourth order Butterworth with 3-dB bandwidth equal to $1.9/T$, whereas the electrical filter is a fifth order Bessel with 3-dB bandwidth equal to $0.75/T$. An optical extinction ratio of 20 dB was assumed. When considering PMD only, the needed $2S$ pdfs at the receiver, one for each trellis transition, were computed using a 32-bit de Bruijn sequence[8], so that all 5-bit interfering patterns were considered, the number of trellis states being thus equal to 16. We verified that these choices are redundant and that a 4-state trellis would be sufficient to describe all interfering patterns. Since our purpose was to achieve the theoretical limit for electrical equalization, a little growth of complexity was tolerated. For the same reason, no quantization of the received signal was taken into account. In the presence of chromatic dispersion, the minimum number of trellis states would be 16, but we used up to 64 states (such large number of states is only useful for large amounts of PMD and/or GVD, when, however, the induced penalty is already several dB's and the achievable benefits are negligible).

2.6.1 Synchronous sampling

The look-up table based MLSD receiver was simulated in different first-order PMD scenarios, with no GVD and DGD equal to 0%, 50%, and 75% of the bit time. The chi-square based MLSD receiver was also tested to verify its performance with respect to the look-up table based one, and in Fig. 2.7 a comparison is presented, showing also the results obtainable by a standard receiver (curves labeled "uncompensated"). As mentioned, the performance of the χ^2 pdf is almost the same as that of the exact one, both giving higher benefits as the DGD increases. Moreover, we found no appreciable difference in the results obtained using the branch metrics in (2.9) or (2.11).

It is interesting to compare the MLSD receiver performance with that of other electrical and optical equalization techniques. In Fig. 2.8, for the same DGD values, we also report the performance of combined (synchronous) feedforward and decision feedback equalizers[9], and that of a 3-stage optical compensator, composed by the cascade of three polarization controllers and

[8] A binary de Bruijn sequence of length 2^n may be obtained from a pseudo-random binary sequence of length $2^n - 1$ by adding a 0 to the longest run of 0's.

[9] The number of taps of both equalizers was optimized, founding no improvement with more than five taps for the FFE and three taps for the DFE.

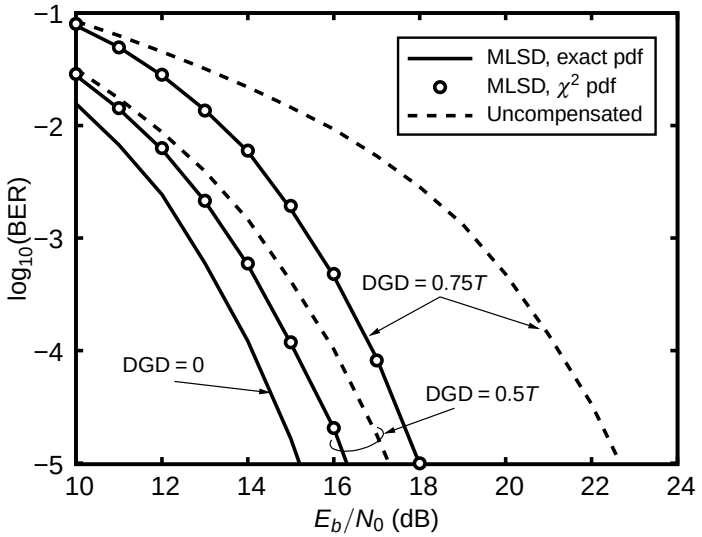

Figure 2.7: Performance of the proposed MLSD receiver with first-order PMD and equal power splitting between PSPs.

polarization maintaining fibers as in [28]. This optical compensator shows a negligible penalty for increasing DGD values (the curves are practically indistinguishable), whereas both electrical processing schemes suffer a larger penalty as the DGD increases. We point out that this is due to the fact that, in this case, we are neglecting higher order PMD effects, and thus the optical compensator is able to perfectly cancel out the ISI in the optical domain, avoiding the beating between ISI and ASE noise, whereas the electrical processing techniques have also to cope with it, and this fact degrades their performance. Yet, MLSD evidences a better performance with respect to electrical equalization for DGD values greater than about $0.5T$, as expected.

The accuracy of the described lower and approximated upper bounds is shown in Fig. 2.9. In this case also, only first-order PMD is considered, with DGD values of 0, $0.5T$, and $0.75T$. It is clear that it is not necessary to resort to time-consuming computer simulations since this tool predicts very well the receiver performance, the higher the accuracy the lower the BER, and can also be used to optimize the system parameters.

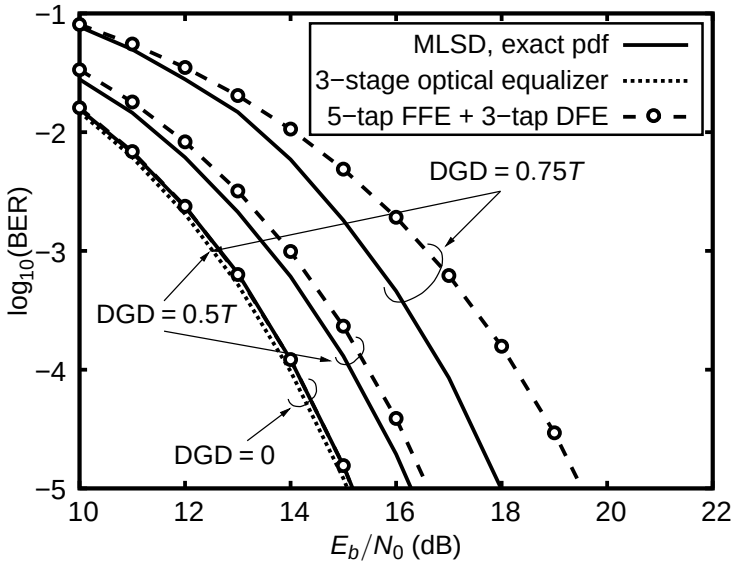

Figure 2.8: Comparison between optical and electrical compensation schemes with first-order PMD and 50% power splitting.

2.6.2 Oversampling

We now consider the receivers based on oversampling. As already mentioned, in the case of oversampling the electrical filter was removed. Hence, the bandwidth of the optical filter was re-optimized, finding that, in this case, its optimum value is $1.2/T$. Accounting for samples correlation, we considered up to 4 samples per bit time, quantized to 4 bits. We found that this quantization level is sufficient to avoid significant penalties, and, at the same time, allows for a reasonable simulation time. Indeed, the complexity required to build up correlated histograms increases exponentially with the number n of samples per bit interval and the number q of quantization bits, as the number of required histogram bins is 2^{qn}. As it can be seen from the simulation curves in Fig. 2.10, no significant improvement is attained for $n > 2$, so an oversampling factor of 2 is assumed in the following.

Since the optimal trade off between computability and efficiency is achieved using just two samples per bit time, it is worth checking the impact of the actual correlation of the samples on the metrics. In fact, if it turns out to be negligible, this could lead again to the possible exploitation of the numerical

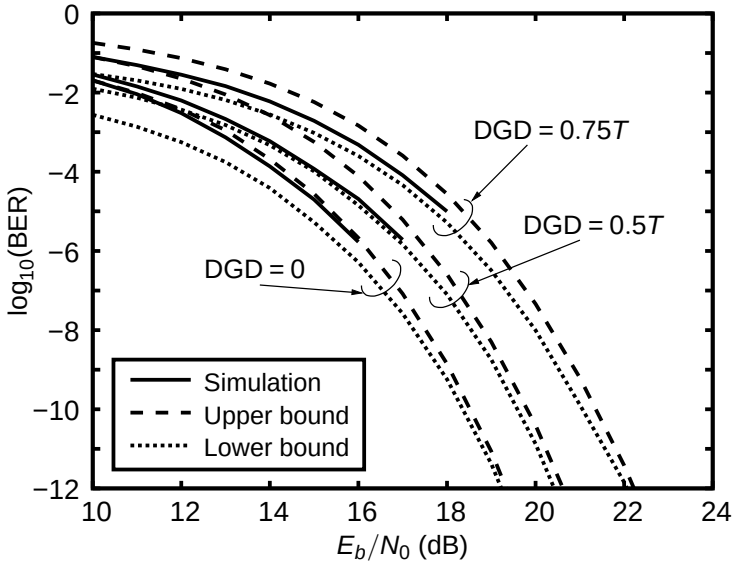

Figure 2.9: Lower and upper bounds of the receiver performance with first-order PMD and 50% power splitting.

method for pdf computation, because the joint pdf could be approximated by the product of the marginal pdfs[10]. The joint pdf accounting for the samples correlation obtained by simulation was compared with the joint pdf analytically evaluated under the hypothesis of sample independence, and this comparison is shown in Fig. 2.11.

The plot in Fig. 2.11 reveals a little amount of correlation. In this figure we compare the joint pdf of two samples in the case of ideal independence, analytically derived as the product of the marginal pdfs, with simulation results showing the actual joint pdf. So Monte Carlo simulations were carried out with both joint and independent histograms in order to evaluate the impact of such correlation on the performance of the receiver. Obviously, proper training sequences of different length were used to build up the histograms, so that the same degree of accuracy was assured. We report in Fig. 2.12 the results obtained by simulation under the two assumptions about samples correlation. As it can be seen, assuming independence leads to a really negligible penalty

[10]We stress the fact that this would be true only for evaluating the metrics, and not in the analytical performance evaluation, which is heavily affected by the correlation value.

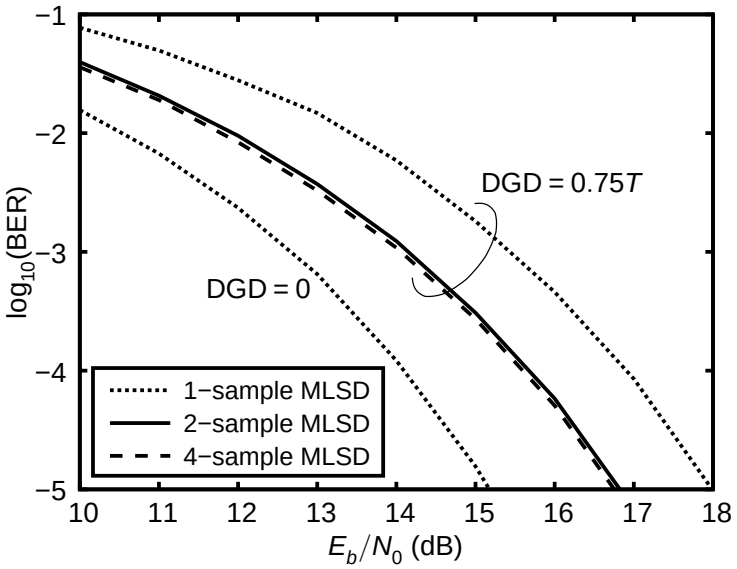

Figure 2.10: Simulation results for the histogram-based oversampling receiver for $n = 1, 2, 4$. First-order PMD and 50% power splitting.

(almost unnoticeable in the graphs), thus we can state that two samples can be considered as independent for evaluating the metrics.

To perform a fair comparison, a histogram-based synchronous receiver was implemented as well, with same quantization, bin width, and training sequence length as the oversampling receiver. After filter bandwidth re-optimization, its performance was evaluated using the reference DGD values 0, 0.5 and 0.75 T. Final results are shown in Fig. 2.13, evidencing that oversampling leads to a significant improvement for increasing values of DGD.

2.6.3 Outage probability

The method that we developed in order to obtain lower and upper bounds allows the computation of the outage probability, defined here as the probability that the BER exceeds a given value for a fixed signal to noise ratio margin, or, equivalently, as the probability that the penalty exceeds the same margin when the BER is fixed at the same value [21]. In computing the outage probability, we consider a reference BER of 10^{-12} and a 3 dB margin, such that Monte Carlo simulations are infeasible. So, we evaluated the penalty by

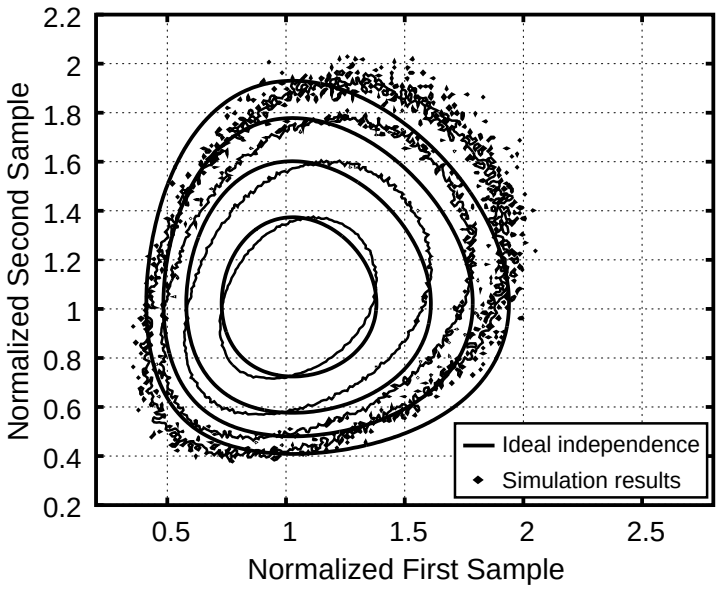

Figure 2.11: Joint pdf of two $T/2$ spaced samples at $E_b/N_0 = 12\,\text{dB}$. Starting from the inner, contours are at $10^0, 10^{-1}, 10^{-2}, 10^{-3}$.

using the approximated upper bound previously discussed.

We initially considered first-order PMD only. In this case, once verified that the penalty contour at 3 dB vs DGD and power splitting is U-shaped, the outage probability can be approximated as [21]

$$P_{\text{out}} \simeq \exp\left[-\frac{4}{\pi}\left(\frac{\Delta\tau_{3\text{dB}}}{\langle\Delta\tau\rangle}\right)^2\right] \tag{2.38}$$

where $\langle\Delta\tau\rangle$ is the mean DGD, whereas $\Delta\tau_{3\text{dB}}$ is the instantaneous DGD giving 3 dB penalty at 50% power splitting, as shown in Fig. 2.14, where the penalty vs DGD is reported. Note that, if the post-detection filter has a highly non-symmetric impulse response, the isopenalty curves may not be U-shaped and, in this case, (2.38) may not be accurate. We verified that this is not the case with the electrical fifth-order Bessel filter we use. The outage probability due to first-order PMD is shown in Fig. 2.15(a). As it can be seen, a synchronous MLSD receiver does fairly better than a FFE combined with a DFE, allowing for a mean DGD value about 30% higher for $P_{\text{out}} < 10^{-6}$, whereas the oversampling MLSD receiver improves this value to about 60%.

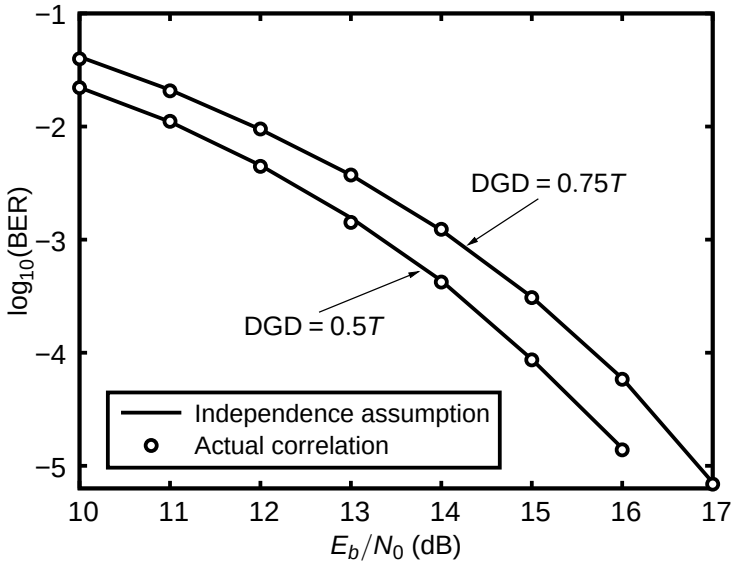

Figure 2.12: Simulation results for independence assumption or actual correlation on noise statistics when taking 2 samples per bit (histogram-based receiver).

There is no closed form approximation for the outage probability due to second-order PMD and we analytically evaluated it as described in [21] using the model in [15] for the fiber Jones matrix, as said. Results are shown in Fig. 2.15(b), where we also superimpose the outage probability due to first-order PMD only. From the comparison it is possible to notice that the impact of second-order PMD is almost negligible, being the relevant penalty due to first-order PMD only. This conclusion is also corroborated by comparison with the outage probability obtainable by exact first-order compensation in the optical domain [21], also shown in Fig. 2.15(b). As in our experience the analytical second-order models for the fiber PMD are always more pessimistic than the random waveplate model, which accounts for all PMD orders, we are confident that our results remain valid also when taking into account higher PMD orders.

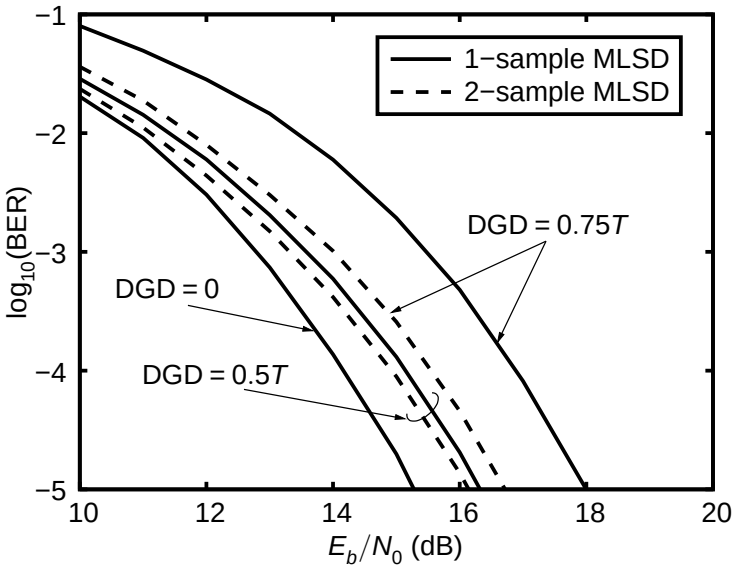

Figure 2.13: Comparison between synchronous and oversampling receivers. First-order PMD and 50% power splitting.

2.6.4 Impact of GVD

We verified that also in the presence of GVD an oversampling factor of 2 is sufficient to give optimal performance, the only difference with PMD being that a larger number of trellis states is required. Anyway, as already discussed, 16 states turned out to be sufficient in most cases. GVD is quantified through the dimensionless chromatic dispersion index γ so that the results are independent of the bit rate R_b. As an example, $\gamma = 0.1$ corresponds to a residual dispersion of 392 ps/nm at 10 Gb/s, 24 ps/nm at 40 Gb/s, and 1.5 ps/nm at 160 Gb/s.

As can be seen from Fig. 2.16, an oversampling MLSD receiver outperforms standard electrical equalization techniques, as well as the synchronous one but for values of the chromatic dispersion index higher than about 0.3. It is interesting to note that the synchronous receiver penalty shows a well defined local minimum around $\gamma = 0.4$ which can be explained analyzing the Euclidean distances between correct and corresponding wrong patterns on the trellis diagram. Indeed, as shown in Fig. 2.17, the distance of the most likely error event, which determines the receiver performance, is smaller when $\gamma = 0.3$

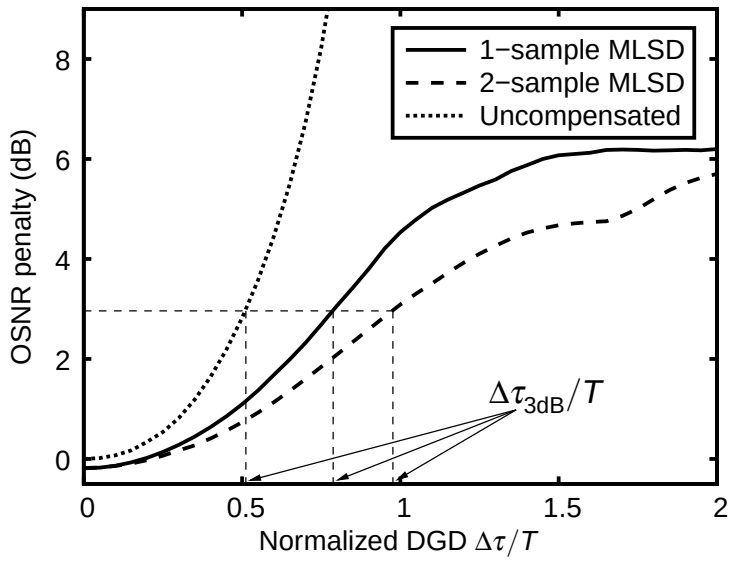

Figure 2.14: OSNR penalty at BER = 10^{-12}vs normalized DGD. First-order PMD and 50% power splitting.

with respect to $\gamma = 0.4$. When the chromatic dispersion index changes from $\gamma = 0.2$ to $\gamma = 0.4$ we also notice a change in the optimum sampling time from $t = T/2$ to $t = 0$. Now, with reference to Fig. 2.17, the maximum eye opening occurs at $t = T/2$ (the eye is closed for $\gamma > 0.6$), and, for $\gamma = 0.4$, if the same samples used in the MLSD receiver (for which the optimum sampling time is $t = 0$) were used in a standard receiver, the BER would be close to 1/2. This is due to the nature of MLSD which performs *sequence* rather than *bit-by-bit* detection, such that a sequence of samples not necessarily is to be taken at the maximum eye opening times and even when the eye is closed it still can be correctly detected.

We next consider the combined effect of GVD and PMD. In Fig. 2.18 we report the maximum value of the mean DGD for which the outage probability is less than 10^{-6} as a function of the chromatic dispersion index. The synchronous receiver curve shows a dip around $\gamma = 0.4$ which can be explained by the above considerations, while the maximum mean DGD tolerable by the oversampling receiver decreases monotonically.

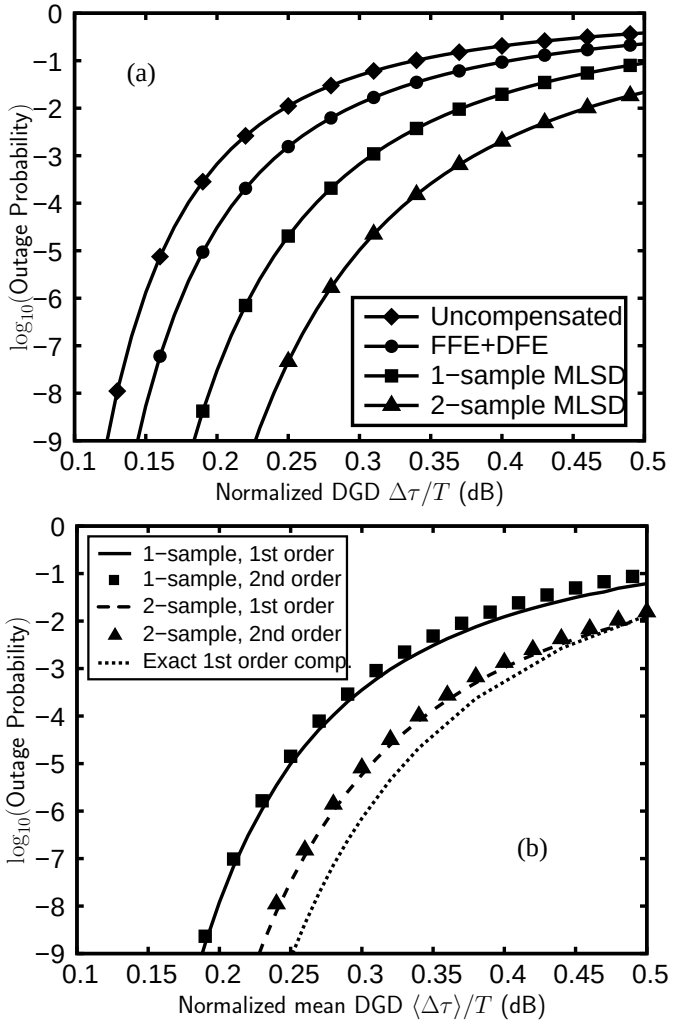

Figure 2.15: Outage probability due to first-order PMD (a), and to first- and second-order PMD (b).

Appendix A

In this Appendix we derive an accurate approximation for the pdf of the square root of the samples of the photodetected signal.

Even though (2.10) proved to be adequate for deriving branch metrics for

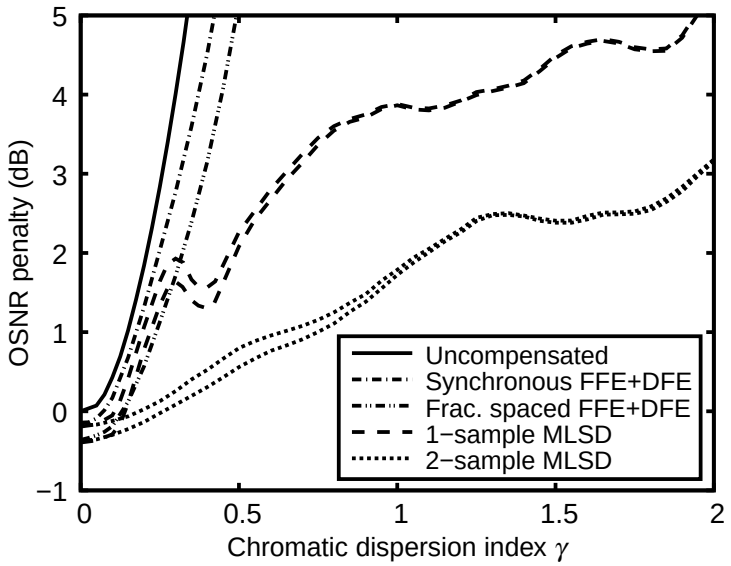

Figure 2.16: OSNR penalty at BER $= 10^{-12}$ as a function of chromatic dispersion index. Both upper and lower bounds are reported for MLSD.

the VA, it is not adequate for performance evaluation. Indeed, as it can be seen from Fig. 2.2, it can quite accurately predict the correct threshold, but would give a BER smaller than about two orders of magnitude with respect to the true value. However, we can use the functional form (2.10) to fit the actual pdf of the sample z_k with very high accuracy as

$$p_{z_k}(z) \simeq \frac{c_k}{\sqrt{4\pi N_k z}} \left(\frac{z}{s_k}\right)^{\frac{2\nu_k-1}{4}} \exp\left(-\frac{\left(\sqrt{z}-\sqrt{s_k}\right)^2}{N_k}\right) \tag{A.1}$$

where N_k and ν_k are as in (2.31) and (2.32), and c_k is the normalization constant

$$c_k = \frac{2\sqrt{\pi}\,(s_k/N_k)^{(2\nu_k-1)/4}\, e^{s_k/N_k}}{\Upsilon(1,1) + 2\sqrt{\frac{s_k}{N_k}}\,\Upsilon(3,3)}\,, \tag{A.2}$$

having defined

$$\Upsilon(n,m) \triangleq \Gamma\left(\frac{2\nu_k+n}{4}\right)\, {}_1F_1\left(\frac{2\nu_k+n}{4}; \frac{m}{2}; \frac{s_k}{N_k}\right), \tag{A.3}$$

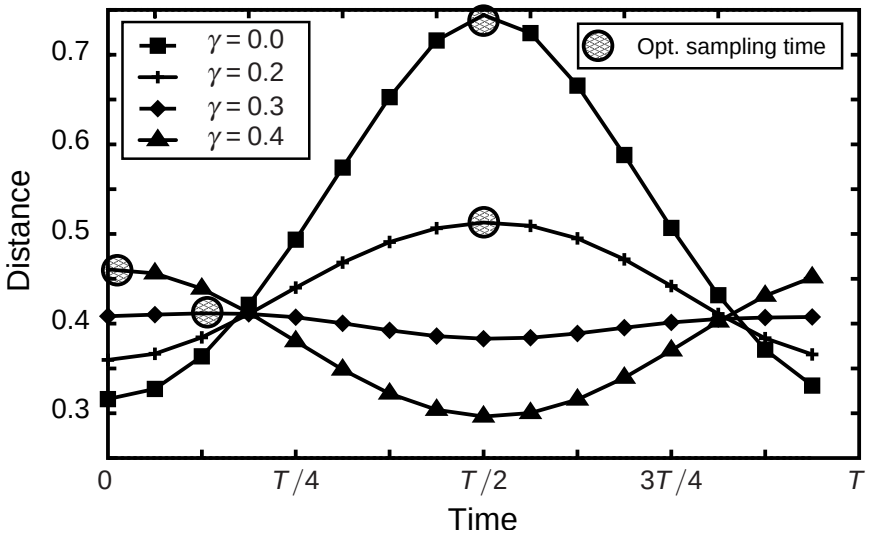

Figure 2.17: Euclidean distance between correct and corresponding wrong patterns on the trellis diagram vs sampling time.

and ${}_1F_1$ being the confluent hypergeometric function [48]. Using the following approximation (less accurate for $x \approx 1$, but still valid)

$$ {}_1F_1(a;b;x) \simeq \frac{\Gamma(b)}{\Gamma(a)}\, x^{a-b} e^x \left(1 + \frac{(a-b)(a-1)}{x}\right),\ x > 1, \qquad \text{(A.4)} $$

it can be seen that c_k depends on the ratio s_k/N_k in a quite simple manner, as

$$ c_k \simeq \frac{s_k/N_k}{s_k/N_k + (2\nu_k - 1)(2\nu_k - 3)/16}\,, \qquad \text{(A.5)} $$

and its value is next to 1 for $s_k/N_k > 3 \div 4$ when $\nu_k < 3$.

From (A.1) it turns out that the pdf of $y_k = \sqrt{z_k}$ can be approximated as

$$ p_{y_k}(y) = 2\sqrt{z}\, p_z(z) \simeq \frac{c_k}{\sqrt{\pi N_k}} \left(\frac{y}{\sqrt{s_k}}\right)^{\nu_k - \frac{1}{2}} e^{-(y-\sqrt{s_k})^2/N_k},\ y \geq 0, \qquad \text{(A.6)} $$

whose tails (only the right-hand one for small s_k/N_k values) are Gaussian.

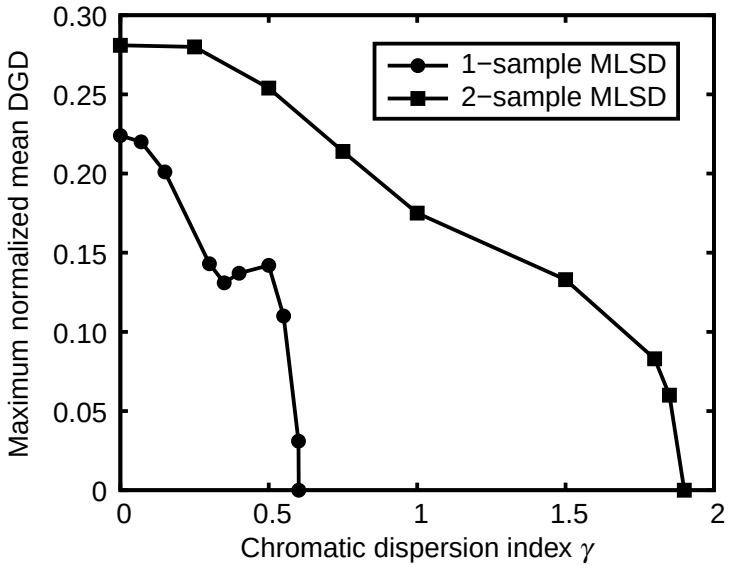

Figure 2.18: Maximum mean DGD giving outage probability less than 10^{-6} vs chromatic dispersion index. First- and second-order PMD is considered.

The mean value of y_k turns out to be

$$\begin{aligned} \eta_{y_k} &\simeq \sqrt{N_k}\,\frac{\Upsilon(3,1)+2\sqrt{\frac{s_k}{N_k}}\,\Upsilon(5,3)}{\Upsilon(1,1)+2\sqrt{\frac{s_k}{N_k}}\,\Upsilon(3,3)} \\ &\simeq \sqrt{s_k}\,\frac{s_k/N_k+(4\nu_k^2-1)/16}{s_k/N_k+(2\nu_k-1)(2\nu_k-3)/16}, \end{aligned} \tag{A.7}$$

such that (A.6) can be approximated as

$$p_{y_k}(y) \simeq \frac{1}{\sqrt{\pi N_k}} \exp\left(-\frac{(y-\eta_{y_k})^2}{N_k}\right). \tag{A.8}$$

We point out again that, for $s_k/N_k \ll 1$, (A.8) is a good approximation only for the right-hand tail of (A.6), but this is suitable to our purposes.

Appendix B

In this Appendix we derive an accurate approximation for the correlation coefficient of the square root of the samples of the phodetected signal.

Writing signal and noise on the two reference polarizations in the optical domain in terms of their real (in phase) and imaginary (quadrature) components

$$\begin{array}{rcll} s_i(t) & = & s_{i1}(t) + j s_{i2}(t) & \\ n_i(t) & = & n_{i1}(t) + j n_{i2}(t) & \end{array} \qquad i = 1, 2 \tag{B.1}$$

and letting

$$x_{ij}(t) \doteq s_{ij}(t) + n_{ij}(t) \qquad i, j = 1, 2 \tag{B.2}$$

the photodetector output in (2.1) can be written as

$$z(t) = x_{11}^2(t) + x_{12}^2(t) + x_{21}^2(t) + x_{22}^2(t).$$

The r.v. $x_{ij}(t)$ is Gaussian with mean $s_{ij}(t)$ and variance $\sigma^2 = N_0 B_o/2$, and the r.v.'s $x_{ij}(t_1)$ and $x_{ij}(t_2)$ are jointly Gaussian, with joint pdf

$$p_{x_{ij}}(x, y) = \frac{\exp\left\{-\frac{(x-\eta_{ij1})^2 - 2\rho(x-\eta_{ij1})(y-\eta_{ij2}) + (y-\eta_{ij2})^2}{2(1-\rho^2)\sigma^2}\right\}}{2\pi\sigma^2\sqrt{1-\rho^2}} \tag{B.3}$$

where $\eta_{ijk} = s_{ij}(t_k)$, and $\rho = R(t_1 - t_2)/R(0)$ is their correlation coefficient, being $R(\tau) = \mathcal{F}^{-1}\left\{|H_o(\omega)|^2\right\}$.

Our aim is to evaluate the correlation coefficient ρ_z of $\sqrt{z(t_1)}$ and $\sqrt{z(t_2)}$

$$\rho_z = \frac{E\{\sqrt{z(t_1)z(t_2)}\} - E\{\sqrt{z(t_1)}\}E\{\sqrt{z(t_2)}\}}{\sigma_{\sqrt{z_1}}\sigma_{\sqrt{z_2}}}, \tag{B.4}$$

where $\sigma_{\sqrt{z_i}}$ is the standard deviation of $\sqrt{z(t_i)}$. As it turns out to be a quite involved task, we resort to an approximate evaluation. Observing that

$$\frac{1}{2}\sum_{k=1}^{4}|a_k| \le \sqrt{\sum_{k=1}^{4} a_k^2} \le \sum_{k=1}^{4}|a_k|\,, \tag{B.5}$$

and letting

$$y(t) \doteq |x_{11}(t)| + |x_{12}(t)| + |x_{21}(t)| + |x_{22}(t)|, \tag{B.6}$$

we have that

$$\sqrt{z(t_i)} = \kappa_i \, y(t_i), \tag{B.7}$$

where $1/2 \leq \kappa_i \leq 1$. Under the hypothesis that the r.v.'s κ_i, $i = 1, 2$, are independent of each other and of $y(t_i)$, we see that ρ_z is equal to the correlation coefficient ρ_y of $y(t_1)$ and $y(t_2)$, and even if this hypothesis does not hold, it is apparent that ρ_z is close to ρ_y. Hence, instead of (B.4), we will evaluate

$$\rho_y = \frac{E\{y(t_1)y(t_2)\} - E\{y(t_1)\}E\{y(t_2)\}}{\sigma_{y_1}\sigma_{y_2}}, \tag{B.8}$$

assuming that $\rho_z \simeq \rho_y$.

As, for $ij \neq k\ell$, $x_{ij}(t)$ and $x_{k\ell}(t)$ are independent,

$$\begin{aligned} &E\{y(t_1)y(t_2)\} - E\{y(t_1)\}E\{y(t_2)\} = \\ &\quad \sum_{ij} \left[E\{|x_{ij}(t_1)x_{ij}(t_2)|\} - E\{|x_{ij}(t_1)|\}E\{|x_{ij}(t_2)|\}\right], \end{aligned} \tag{B.9}$$

and, while $E\{|x_{ij}(t_k)|\}$ is straightforward to evaluate

$$E\{|x_{ij}(t_k)|\} = \eta_{ijk} \, \mathrm{erf}\left(\frac{\eta_{ijk}}{\sqrt{2}\,\sigma}\right) + \sqrt{\frac{2}{\pi}}\,\sigma \exp\left(-\frac{\eta_{ijk}^2}{2\sigma^2}\right), \tag{B.10}$$

the evaluation of $E\{|x_{ij}(t_1)x_{ij}(t_2)|\}$ requires some more work. Letting, for reasons to be soon explained,

$$\eta_1 \doteq \begin{cases} \eta_{ij1} & \text{if } |\eta_{ij1}| < |\eta_{ij2}| \\ \eta_{ij2} & \text{otherwise} \end{cases} \tag{B.11}$$

and

$$\eta_2 \doteq \begin{cases} \eta_{ij2} & \text{if } |\eta_{ij1}| < |\eta_{ij2}| \\ \eta_{ij1} & \text{otherwise} \end{cases} \tag{B.12}$$

it can be shown that

$$E\{|x_{ij}(t_1)x_{ij}(t_2)|\} = \iint_{-\infty}^{\infty} |xy| \, p_{x_{ij}}(x, y) \, dxdy = \int_{-\infty}^{\infty} |x| \, f(x) \, dx \tag{B.13}$$

where

$$f(x) = \frac{\sqrt{1-\rho^2}}{\pi} \exp\left(-\frac{(x-\eta_1)^2}{2\sigma^2}\right) \left\{\exp\left[-g^2(x)\right] + \sqrt{\pi}\, g(x) \, \mathrm{erf}\left[g(x)\right]\right\}, \tag{B.14}$$

and

$$g(x) = \frac{\rho(x-\eta_1)+\eta_2}{\sigma\sqrt{2(1-\rho^2)}}. \tag{B.15}$$

Notice that, due to (B.3), the result of the integral in (B.13) remains unchanged if we swap η_{ij1} and η_{ij2}, and we exploited this fact in (B.11) and (B.12) because the function $f(x)$ can be approximated as $f(x) \simeq |f_a(x)|$, where

$$f_a(x) = \frac{\rho(x-\eta_1)+\eta_2}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{(x-\eta_1)^2}{2\sigma^2}\right), \tag{B.16}$$

with an accuracy increasing with the ratio $|\eta_2|/\sigma$. Letting now

$$\xi \doteq \eta_1 - \eta_2/\rho \tag{B.17}$$

we have

$$\int_{-\infty}^{\infty} |x|\, f(x)\, dx \simeq \int_{-\infty}^{\infty} |x\, f_a(x)|\, dx = \int_{-\infty}^{\infty} x\, f_a(x)\, dx - 2\,\mathrm{sgn}(\xi) \int_0^{\xi} x\, f_a(x)\, dx. \tag{B.18}$$

The first integral in the last equation in (B.18) gives

$$\int_{-\infty}^{\infty} x\, f_a(x)\, dx = \rho\sigma^2 + \eta_1\eta_2, \tag{B.19}$$

while the second one gives

$$\begin{aligned}\int_0^{\xi} x\, f_a(x)\, dx &= \frac{1}{2}\left(\rho\,\sigma^2 + \eta_1\,\eta_2\right)\left[\mathrm{erf}\left(\frac{\eta_1}{\sqrt{2}\,\sigma}\right) - \mathrm{erf}\left(\frac{\eta_2}{\sqrt{2}\,\rho\,\sigma}\right)\right] \\ &\quad + \frac{\sigma}{\sqrt{2\pi}}\left[\eta_2 \exp\left(-\frac{\eta_1^2}{2\sigma^2}\right) - \rho\,\eta_1 \exp\left(-\frac{\eta_2^2}{2\rho^2\sigma^2}\right)\right].\end{aligned}$$

When both η_{ij1} and η_{ij2} are of the same order or smaller than σ, (B.18) loses accuracy and we must add a correction term to it. Using the approximation $f(x) - |f_a(x)| \simeq f_e(x)$, where

$$f_e(x) = \frac{\sqrt{1-\rho^2}}{\pi} \exp\left(-\frac{(x-\eta_1)^2}{2\sigma^2} - \sqrt{\pi}\,|g(x)| - \frac{\pi-2}{2}\,g^2(x)\right),$$

the correction term to be added to (B.18) is approximated with great accuracy by

$$\begin{aligned}\int_{-\infty}^{\infty} |x|\, f_e(x)\, dx \;\; = \;\; & \varepsilon(\xi)\Big\{2e^{-\upsilon_1^2(\xi)} - e^{-\upsilon_2^2(\xi)} \\ & + \sqrt{\pi}\,\upsilon_1(\xi)\,[2\,\mathrm{erf}\,\upsilon_1(\xi) - \mathrm{erf}\,\upsilon_2(\xi) - \mathrm{sgn}\,\xi]\Big\} \\ & + \varepsilon(-\xi)\Big\{e^{\upsilon_2^2(-\xi)} + \sqrt{\pi}\,\upsilon_1(-\xi)\,[\mathrm{erf}\,\upsilon_2(-\xi) + \mathrm{sgn}\,\xi]\Big\},\end{aligned} \tag{B.20}$$

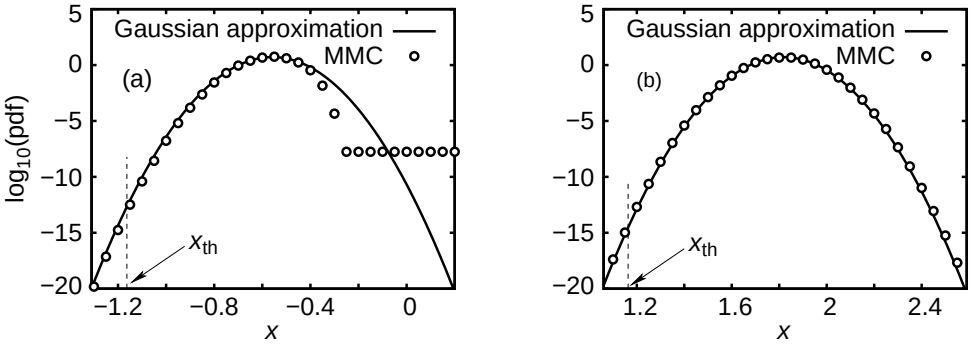

Figure 2.19: pdf of the r.v. x in (2.33) when (a) $\mathbf{a} = \{1,1,1,0,1,1,1,0\}$ and $\hat{\mathbf{a}} = \{1,1,1,1,1,1,1,0\}$, or (b) sequences $\mathbf{a}$ and $\hat{\mathbf{a}}$ are exchanged. Samples are spaced by $T/2$ and $E_b/N_0 = 20\,\mathrm{dB}$.

ξ being as in (B.17) and

$$\varepsilon(\xi) = \frac{r^2\sigma^2\sqrt{1-\rho^2}}{\pi}\exp\left(-\frac{\eta_2\mu(\xi)}{\sigma^2} + \frac{(\pi-2)\eta_2^2 + \pi\rho^2\sigma^2}{2\sigma^2\,[2-(4-\pi)\rho^2]}\right), \tag{B.21}$$

$$\upsilon_1(\xi) = \frac{\eta_1 - \rho\,\mu(\xi)}{\sqrt{2}\,r\,\sigma}, \tag{B.22}$$

$$\upsilon_2(\xi) = \frac{\eta_2 - \rho^2\mu(\xi)}{\sqrt{2}\,\rho\,r\,\sigma}, \tag{B.23}$$

$$r = \sqrt{\frac{2(1-\rho^2)}{2-(4-\pi)\rho^2}}, \tag{B.24}$$

$$\mu(\xi) = \frac{(\pi-2)\eta_2 - \mathrm{sgn}(\xi)\sigma\sqrt{2\pi(1-\rho^2)}}{2-(4-\pi)\rho^2}. \tag{B.25}$$

Notice that, when η_1 and η_2 are chosen as in (B.11) and (B.12), the correction term (B.20) can always be safely added to (B.18), otherwise it could cause an accuracy loss if either $|\eta_{ij1}|$ or $|\eta_{ij2}|$ are greater than σ.

The last ingredients we need for evaluating ρ_y are σ_{y_1} and σ_{y_2}, but they pose no problems. Indeed, the second order moment of $|x_{ij}(t_k)|$ is the same as that of $x_{ij}(t_k)$, i.e., $\sigma^2 + \eta_{ijk}^2$, and, taking into account (B.10), we have

$$\sigma_{y_k}^2 = 4\sigma^2 + \sum_{ij} \eta_{ijk}^2 - \sum_{ij} \left[\eta_{ijk} \operatorname{erf}\left(\frac{\eta_{ijk}}{\sqrt{2}\,\sigma} \right) + \sqrt{\frac{2}{\pi}}\,\sigma \exp\left(-\frac{\eta_{ijk}^2}{2\sigma^2} \right) \right]^2 .$$

In order to check our hypotheses, we evaluated by the multicanonical Monte Carlo (MMC) technique [53] the pdf of the r.v. x in (2.33) and compared it with a Gaussian pdf whose mean and variance are as in (2.34) and (2.35), respectively. In Fig. 2.19(a) we report the results obtained by oversampling when the correct sequence is taken as $\mathbf{a} = \{1, 1, 1, 0, 1, 1, 1, 0\}$ and the wrong sequence as $\hat{\mathbf{a}} = \{1, 1, 1, 1, 1, 1, 1, 0\}$, and in Fig. 2.19(b) the same quantity but with the role of the two previous sequences interchanged. The threshold $x_{\text{th}} = (1/2) \sum_k (s_k - \hat{s}_k)$ giving the PEP $P(\mathbf{a} \to \hat{\mathbf{a}}) = P(x < x_{\text{th}})$ is also shown.

Chapter 3

Multilevel modulations with the interferometric front end

Multilevel signaling formats based on amplitude modulations have been investigated in [54] and, more recently, other schemes, still based on multilevel modulations, have been proposed or have been brought back to the attention of the scientific community. In particular, we are referring to systems using differentially encoded quaternary phase-shift keying (QPSK) signals demodulated by an interferometric IM/DD receiver, recently proposed in [55] (see also [10] and references therein) and extended to higher-order modulations in [56, 57], or to systems using combined amplitude-phase modulations demodulated by coherent techniques [2, 58]. In both cases, the use of multilevel modulations allows to reduce, for a given bit rate, the signaling rate, thus reducing the impact of GVD and PMD. In addition, the use of a different and more sophisticated (with respect to a simple photodetector) front end gives hope that with a proper electronic processing the impact of PMD and GVD can be completely mitigated. If this is certainly true when synchronous coherent techniques are employed [2, 58], in the case of interferometric IM/DD receivers some attempts to devise a more effective electronic processing are described in [59–61] where, by resorting to *multi-symbol differential detection* methods, commonly adopted in wireless communications [62–64], the authors try to improve the performance over the conventional symbol-by-symbol receiver.

In this chapter, the use of high-order modulations with interferometric front end in optical transmission systems is discussed. First of all, without

constraints on the receiver front end, we identify the optimal processing to be performed on the received signal from a theoretical point of view, showing that this optimal processing is able to perfectly compensate for PMD and GVD. The implementation aspects are then discussed. The results represent an evolution of those in [65,66] which in turn evolve from the multi-symbol differential schemes in [62–64] (see [65] for a comprehensive discussion and performance comparison among these receivers). The state-complexity reduction of the VA is then described along with other side (although very important) aspects such as the channel estimation and the generalization to different channel encoders. The possibility of employing the described detection schemes in the case of *transmit* polarization diversity (often referred to as *polarization multiplexing*), hence further reducing the signaling rate for a given bit rate, will be also discussed. Regarding the receiver front end, it will be shown that we can employ a (slightly modified) interferometric IM/DD front end. Although the proposed receiver allow for a perfect compensation of the dispersion effects when its complexity (in particular on the number of the VA trellis states) is not constrained, in the numerical results we have also considered the effect of a limited receiver complexity on the system performance. The results presented in this chapter have been published in [67].

3.1 System model

In the considered system, a sequence $\mathbf{c} = (c_0, c_1, \ldots, c_{K-1})^T$ of K complex symbols belonging to an M-ary complex alphabet $\mathcal{C}$ is obtained, through a proper differential encoding rule [68], from a sequence $\mathbf{a} = (a_1, \ldots, a_{K-1})^T$ of $K-1$ complex symbols belonging to the same alphabet.[1] Without loss of generality, in the numerical results we will consider classical phase-shift keying (PSK) signals, for which the standard differential encoding rule is employed, and square quadrature amplitude modulations (QAMs) for which quadrant differential encoding is adopted [68] (see also [65, Section V-A] for a concise description). However, our derivations can be also applied to other alphabets, for example amplitude- and phase shift keying (APSK) modulations, whose signal constellations are composed of more concentric rings of PSK points.

We consider the receiver as composed of an analog part, the opto-electronic (O/E) front end, devoted to signal demodulation and conversion from the op-

[1] In the following, $(\cdot)^T$ denotes transpose, $(\cdot)^H$ transpose conjugate, and $(\cdot)^*$ complex conjugate.

tical to the electrical domain, and a digital part devoted to electronic processing. The O/E front end may be based on interferometric IM/DD or coherent techniques, as explained later in Chapt. 4. The transfer function of the fiber channel accounting for GVD and PMD is described in Chapt. 1.

The low-pass equivalent of the signal at the receiver is denoted by

$$\mathbf{x}(\mathrm{t}) = [x_1(t), x_2(t)]^T,$$

being $x_1(t)$ and $x_2(t)$ the received signal components on the above mentioned orthogonal SOPs. We can express each component of $\mathbf{x}(t)$ as

$$x_i(t) = s_i(t, \mathbf{c})e^{j\theta_i(t)} + w_i(t), \quad i = 1, 2 \tag{3.1}$$

where $s_i(t, \mathbf{c})$ is the useful signal, distorted by GVD and PMD, and $\theta_i(t)$ is a time-varying phase uncertainty accounting for the laser phase noise and for the uncertainties due to channel propagation. Although the source of phase noise is the same for both components, in general $\theta_1(t)$ and $\theta_2(t)$ differ for an unknown phase shift, that we may assume uniformly distributed in the interval $[0, 2\pi)$. The useful signal components can be expressed as

$$s_i(t, \mathbf{c}) = \sum_{k=0}^{K-1} c_k p_i(t - kT), \quad i = 1, 2 \tag{3.2}$$

where $p_i(t)$ is the received pulse on the i-th signal component and T the symbol interval. Without loss of generality, we may assume that the pulse $p_i(t)$ has its support in the interval $[0, (L_i + 1)T)$, where L_i is a suitable integer, and we will denote $L = \max_{i=1,2}\{L_i\}$. When $L \geq 1$, ISI arises. Although in general the ISI causes a performance degradation [25], this is not the case of GVD and PMD. In fact, since the fiber Jones transfer matrix $\mathbf{H}_f(\omega)$ is unitary, it is $\mathbf{H}_f(\omega)\mathbf{H}_f^H(\omega) = \mathbf{H}_f(\omega)\mathbf{H}_f^{-1}(\omega) = \mathbf{I}$. Hence, the linear distortions introduced by a dispersive fiber, regarded as a two-input/two-output system, can be considered as the two-dimensional extension of what in a single input/single output system we call a "phase distortion". This means that an ideal receiver could, in principle, estimate $\mathbf{H}_f(\omega)$ and filter the received signal with $\mathbf{H}_f^{-1}(f) = \mathbf{H}_f^H(f)$, thus perfectly compensating the distortions introduced by the channel, without modifying the noise statistics. Hence, the performance in the absence of distortions (the back-to-back case) could be attained. However, although the channel estimate is in general feasible, since

GVD is a static phenomenon and PMD is slowly-varying, the implementation of the inverse filter[2] poses some complexity issues. In addition, when the phase noise is rapidly-varying, its estimate is not so trivial. In order to circumvent these problems, in the next sections we show an equivalent electronic processing and discuss its implementation aspects.

3.2 Optimal receiver processing

Not considering for now the implementation aspects of the receiver front end, a possible way of extracting sufficient statistics from the received signal $\mathbf{x}(t)$ is by means of proper analog prefiltering and sampling at the Nyquist rate [41]. In the following, we will assume that η samples per symbol interval are extracted from the signal. This number of samples depends on the bandwidth of the received useful signal. As an example, when its low-pass equivalent has a bandwidth $B_s \leq 1/2T$, $\eta = 1$ is sufficient, when $1/2T < B_s \leq 1/T$, $\eta = 2$ is sufficient, and so on. Without loss of generality, we assume that the analog prefilter has no effect on the useful signal and that its low-pass equivalent has a vestigial symmetry around $\eta/2T$ [41]. This latter condition ensures that the noise samples are independent and identically distributed complex Gaussian random variables with mean zero and variance $\sigma^2 = N_0\eta/T$ [41]. Denoting by $\mathbf{x}'(t) = [x'_1(t), x'_2(t)]^T$ and $\mathbf{w}'(t) = [w'_1(t), w'_2(t)]^T$ the received signal and noise after the analog prefilter, we define ($i = 1, 2$, $k = 0, 1, \ldots, K-1$, $n = 0, 1, \ldots, \eta - 1$)

$$\begin{aligned} x_{i,k\eta+n} &\triangleq x'_i(kT + nT_c) \\ s_{i,k\eta+n}(\mathbf{c}) &\triangleq s_i(kT + nT_c, \mathbf{c}) \\ \theta_{i,k\eta+n} &\triangleq \theta_i(kT + nT_c) \\ w_{i,k\eta+n} &\triangleq w'_i(kT + nT_c) \end{aligned} \tag{3.3}$$

[2]The inverse filter can be approximated by an optical compensator (e.g., see [69] and references therein).

where $T_c = T/\eta$ is the sampling interval. Due to the limited support of the pulse $p_i(t)$, it is $s_{i,k\eta+n}(\mathbf{c}) = s_{i,k\eta+n}(\mathbf{c}_{k-L_i}^k)$.[3] From (3.1) and (3.2), we have

$$\begin{aligned} x_{i,k\eta+n} &= s_{i,k\eta+n}(\mathbf{c}_{k-L_i}^k)e^{j\theta_{i,k\eta+n}} + w_{i,k\eta+n} \\ &= \sum_{\ell=0}^{L_i} c_{k-\ell} p_i(\ell T + nT_c)e^{j\theta_{i,k\eta+n}} + w_{i,k\eta+n}\,. \end{aligned} \tag{3.4}$$

Let us define $\mathbf{p}_{i,n} = [p_i(nT_c), p_i(T+nT_c), \ldots, p_i(L_iT+nT_c)]^T$, $i = 1, 2$, $n = 0, 1, \ldots, \eta-1$. Assuming for the moment that the phase noise is constant during the whole transmission, i.e., $\theta_i(t) = \theta_i$, with θ_i modeled as a random variable uniformly distributed in the interval $[0, 2\pi)$, and that vectors $\mathbf{p}_{i,n}$ are perfectly known to the receiver, the sequence $\hat{\mathbf{c}}$, detected according to the MLSD strategy, can be expressed as

$$\hat{\mathbf{c}} = \operatorname*{argmax}_{\mathbf{c}} \Lambda(\mathbf{c}) \tag{3.5}$$

where the sequence metric $\Lambda(\mathbf{c})$ has expression

$$\Lambda(\mathbf{c}) = \sum_{i=1}^{2} \left[\left| \sum_{k=0}^{K-1} \sum_{n=0}^{\eta-1} x_{i,k\eta+n} s_{i,k\eta+n}^*(\mathbf{c}_{k-L_i}^k) \right| - \frac{1}{2} \sum_{k=0}^{K-1} \sum_{n=0}^{\eta-1} |s_{i,k\eta+n}(\mathbf{c}_{k-L_i}^k)|^2 \right]. \tag{3.6}$$

This strategy is an extension of that in [65, eqn. (18)] to the case of multiple samples per symbol interval and of two received signals which, conditionally to the sequence $\mathbf{c}$, are independent. In addition, the approximation $\ln I_0(x) \simeq x$, where $I_0(x)$ is the zero-th order modified Bessel function of the first kind, has been used and the terms irrelevant for the detection process have been discarded.

As demonstrated by means of information theory arguments [70–72], the strategy (3.5)–(3.6) attains the same performance of the ideal receiver which perfectly knows the channel phases θ_i in addition to vectors $\mathbf{p}_{i,n}$. Hence, this receiver as well is insensitive to phase distortions as PMD and GVD. Unfortunately, there are a few problems related to the implementation of the strategy (3.5)–(3.6). By means of the same manipulations described in [65],

[3] In the following, given two integers k_1 and $k_2 > k_1$, we define $\mathbf{c}_{k_1}^{k_2} \doteq (c_{k_1}, c_{k_1+1}, \ldots, c_{k_2})^T$. Obviously, $\mathbf{c} = \mathbf{c}_0^{K-1}$.

the sequence metric $\Lambda(\mathbf{c})$ can be equivalently computed as

$$\Lambda(\mathbf{c}) = \sum_{k=0}^{K-1} \Delta_k(\mathbf{c}_0^k) \tag{3.7}$$

where

$$\begin{aligned} \Delta_k(\mathbf{c}_0^k) &= \sum_{i=1}^{2} \left[\left| \sum_{n=0}^{\eta-1} x_{i,k\eta+n} s_{i,k\eta+n}^*(\mathbf{c}_{k-L_i}^k) + q_{i,k-1}(\mathbf{c}_0^{k-1}) \right| \right. \\ &\quad \left. - \left| q_{i,k-1}(\mathbf{c}_0^{k-1}) \right| - \frac{1}{2} \sum_{n=0}^{\eta-1} \left| s_{i,k\eta+n}(\mathbf{c}_{k-L_i}^k) \right|^2 \right] \end{aligned} \tag{3.8}$$

and

$$q_{i,k-1}(\mathbf{c}_0^{k-1}) = \sum_{m=0}^{k-1} \sum_{n=0}^{\eta-1} x_{i,m\eta+n} s_{i,m\eta+n}^*(\mathbf{c}_{m-L_i}^m)\,. \tag{3.9}$$

The difficulty inherent in the incremental metric (3.8) is its unlimited memory. In fact, $\Delta_k(\mathbf{c}_0^k)$ depends on the entire previous code sequence [65]. This implies that the maximization of the sequence metric cannot be implemented by means of the VA working on a properly defined trellis diagram [34].

Other problems must be also considered. Indeed, in order to obtain the MLSD strategy (3.5)–(3.6), constant channel phases θ_i and perfectly known vectors $\mathbf{p}_{i,n}$ have been assumed. These problems will be faced in the next section.

3.3 Practical implementation

3.3.1 Rectangular window

A possible solution for the above mentioned problem of unlimited memory of the incremental metric $\Delta_k(\mathbf{c}_0^k)$ in (3.8) is suggested in [65], where it is proposed to truncate $q_{i,k-1}(\mathbf{c}_0^{k-1})$ by using a rectangular window. In other words, $q_{i,k-1}(\mathbf{c}_0^{k-1})$ in (3.9) is substituted by

$$q_{i,k-1}^{(N)}(\mathbf{c}_{k-N-L_i}^{k-1}) = \sum_{m=k-N}^{k-1} \sum_{n=0}^{\eta-1} x_{i,m\eta+n} s_{i,m\eta+n}^*(\mathbf{c}_{m-L_i}^m)\,. \tag{3.10}$$

The design integer parameter N is called *implicit phase memory* [65]. The resulting sequence metric can be approximated as

$$\Lambda(\mathbf{c}) \simeq \sum_{k=0}^{K-1} \lambda_k^{(N)}(\mathbf{c}_{k-N-L}^k) \tag{3.11}$$

with

$$\begin{aligned}\lambda_k^{(N)}(\mathbf{c}_{k-N-L}^k) &= \sum_{i=1}^{2} \left[\left| \sum_{n=0}^{\eta-1} x_{i,k\eta+n} s_{i,k\eta+n}^*(\mathbf{c}_{k-L_i}^k) + q_{i,k-1}^{(N)}(\mathbf{c}_{k-N-L_i}^{k-1}) \right| \right. \\ &\quad \left. - \left| q_{i,k-1}^{(N)}(\mathbf{c}_{k-N-L_i}^{k-1}) \right| - \frac{1}{2} \sum_{n=0}^{\eta-1} \left| s_{i,k\eta+n}(\mathbf{c}_{k-L_i}^k) \right|^2 \right] . \end{aligned} \tag{3.12}$$

In this case, the maximization of the approximated sequence metric (3.11) can be performed by means of the VA and $\lambda_k^{(N)}(\mathbf{c}_{k-N-L}^k)$ assumes the meaning of branch metric on a trellis diagram whose state is defined as $\mu_k^{(N)} = \mathbf{c}_{k-N-L}^{k-1}$. Hence, the number of states depends exponentially on $L+N$. However, techniques for state-complexity reduction, such as those described in Section 3.3.3, can be used in order to limit the number of states without excessively reducing the value of N.

In the case of constant channel phases θ_i, although in principle the performance of the optimal detection strategy (3.5)–(3.6) is obtained only when $N \to \infty$ [73], it is sufficient to use small values of N (a few units) to obtain a performance very close to the optimal one [65]. On the other hand, this new algorithm requires approximately constant channel phases θ_i in a window of N symbol intervals only [65]. Hence, it can be used when the channel phases are time-varying—the smaller the value of N, the greater the robustness to the phase noise. Finally, from (3.12) notice that, although more samples per symbol interval are used as a sufficient statistic, the VA works at the symbol rate.

3.3.2 Exponential window

Instead of a rectangular window, in [66] an exponentially decaying window is employed. In particular, a trellis state $\mu_k = \mathbf{c}_{k-L}^{k-1}$ is defined and $q_{i,k-1}(\mathbf{c}_0^{k-1})$ in (3.9) is substituted by a complex quantity $q_{i,k-1}^{(\alpha)}(\mu_k)$ estimated based on

per-survivor processing [74] and computed recursively as

$$q_{i,k-1}^{(\alpha)}(\mu_k) = \alpha q_{i,k-2}^{(\alpha)}(\mu_{k-1}) + \sum_{n=0}^{\eta-1} x_{i,(k-1)\eta+n} s_{i,(k-1)\eta+n}^{*}(\mathbf{c}_{k-1-L_i}^{k-1}) \tag{3.13}$$

where $0 \leq \alpha \leq 1$ is a *forgetting factor*. Therefore, the resulting sequence metric is

$$\Lambda(\mathbf{c}) \simeq \sum_{k=0}^{K-1} \lambda_k^{(\alpha)}(c_k, \mu_k) \tag{3.14}$$

with

$$\begin{aligned} \lambda_k^{(\alpha)}(c_k, \mu_k) &= \lambda_k^{(\alpha)}(\mathbf{c}_{k-L}^{k}) = \sum_{i=1}^{2} \left[\left| \sum_{n=0}^{\eta-1} x_{i,k\eta+n} s_{i,k\eta+n}^{*}(\mathbf{c}_{k-L_i}^{k}) + q_{i,k-1}^{(\alpha)}(\mu_k) \right| \right. \\ &\quad \left. - \left| q_{i,k-1}^{(\alpha)}(\mu_k) \right| - \frac{1}{2} \sum_{n=0}^{\eta-1} \left| s_{i,k\eta+n}(\mathbf{c}_{k-L_i}^{k}) \right|^2 \right] \end{aligned} \tag{3.15}$$

and its maximization is performed by means of the VA working at a symbol rate. When the channel phases are constant, for $\alpha \to 1$ the performance of this algorithm tends to that of the optimal one [66], hence ensuring in this case no performance degradation due to PMD and GVD, too. In addition, this algorithm also works well in the presence of a time-varying phase noise, although it is less robust than that described in Section 3.3.1. To this purpose, the value of α must be optimized for the phase noise at hand. Since for the phase noise of the commonly used lasers the robustness of this algorithm is sufficient, in the numerical results we will only consider it.

We would like to point out a different equivalent expression for the branch metric (3.15). To illustrate the main idea, we limit ourself to the case $\eta = 2$ that will be considered in the numerical results. The branch metric (3.15) can be expressed as

$$\begin{aligned} \lambda_k^{(\alpha)}(c_k, \mu_k) &= \sum_{i=1}^{2} \left\{ \frac{1}{|x_{i,2k}|} \left[\left| x_{i,2k}^{*} x_{i,2k+1} s_{i,2k+1}^{*}(\mathbf{c}_{k-L_i}^{k}) \right| \right. \right. \\ &\quad \left. + |x_{i,2k}|^2 s_{i,2k}^{*}(\mathbf{c}_{k-L_i}^{k}) + x_{i,2k}^{*} q_{i,k-1}^{(\alpha)}(\mu_k) - \left| x_{i,2k}^{*} q_{i,k-1}^{(\alpha)}(\mu_k) \right| \right] \\ &\quad \left. - \frac{1}{2} \left| s_{i,2k+1}(\mathbf{c}_{k-L_i}^{k}) \right|^2 - \frac{1}{2} \left| s_{i,2k}(\mathbf{c}_{k-L_i}^{k}) \right|^2 \right\} \end{aligned} \tag{3.16}$$

with

$$q_{i,k-1}^{(\alpha)}(\mu_k) = \alpha q_{i,k-2}^{(\alpha)}(\mu_{k-1}) + x_{i,2k-1}s_{i,2k-1}^*(\mathbf{c}_{k-1-L_i}^{k-1}) + x_{i,2k-2}s_{i,2k-2}^*(\mathbf{c}_{k-1-L_i}^{k-1})\,.$$

Defining now

$$z_{i,\ell} \doteq x_{i,\ell}x_{i,\ell-1}^* \tag{3.17}$$

and taking into account that

$$x_{i,\ell}x_{i,\ell-2}^* = \frac{z_{i,\ell}z_{i,\ell-1}}{|x_{i,\ell-1}|^2} \tag{3.18}$$

we can express

$$\begin{aligned}\lambda_k^{(\alpha)}(\mathbf{c}_{k-L}^k) = \sum_{i=1}^{2}\Bigg\{ &\frac{1}{|x_{i,2k}|}\left[\left|z_{i,2k+1}s_{i,2k+1}^*(\mathbf{c}_{k-L_i}^k)\right.\right. \\ &\left.\left.+|x_{i,2k}|^2 s_{i,2k}^*(\mathbf{c}_{k-L_i}^k) + g_{i,k-1}^{(\alpha)}(\mu_k)\right| - \left|g_{i,k-1}^{(\alpha)}(\mu_k)\right|\right] \\ &-\frac{1}{2}\left|s_{i,2k+1}(\mathbf{c}_{k-L_i}^k)\right|^2 - \frac{1}{2}\left|s_{i,2k}(\mathbf{c}_{k-L_i}^k)\right|^2\Bigg\}\,.\end{aligned} \tag{3.19}$$

where $g_{i,k-1}^{(\alpha)}(\mu_k) \doteq x_{i,2k}^* q_{i,k-1}^{(\alpha)}(\mu_k)$ can be computed recursively as

$$\begin{aligned}g_{i,k-1}^{(\alpha)}(\mu_k) = &\alpha g_{i,k-2}^{(\alpha)}(\mu_{k-1})\frac{z_{i,2k}^* z_{i,2k-1}^*}{|x_{i,2k-1}|^2|x_{i,2k-2}|^2} + z_{i,2k}^* s_{i,2k-1}^*(\mathbf{c}_{k-1-L_i}^{k-1}) \\ &+\frac{z_{i,2k}^* z_{i,2k-1}^*}{|x_{i,2k-1}|^2}s_{i,2k-2}^*(\mathbf{c}_{k-1-L_i}^{k-1})\end{aligned} \tag{3.20}$$

In other words, the branch metric can be equivalently computed by using samples $\{z_{i,k\eta+n}\}$ and $\{|x_{i,k\eta+n}|\}$ instead of samples $\{x_{i,k\eta+n}\}$. This equivalent expression will be exploited in the receiver implementation described in Section 3.4. Similar considerations hold for the branch metric (3.12), in this case as well the branch metric can be expressed as a function of $\{z_{i,k\eta+n}\}$ and $\{|x_{i,k\eta+n}|\}$.

3.3.3 Complexity reduction

The state-complexity of the detection schemes described in Sections 3.3.1 and 3.3.2 can be limited by employing the well-known RSSD technique [49, 75, 76].

Following this technique, a reduced number of symbols is considered in the trellis state definition, hence reducing the number of trellis states. More complex techniques based on set partitioning may also be employed [49, 75, 76]. In order to compute the branch metrics (3.12), (3.15), or (3.19) in the reduced trellis, the necessary symbols not included or not completely specified in the state definition may be found in the survivor history. We note that, in the limiting case of a degenerate trellis diagram with only one state, symbol-by-symbol detection with decision feedback is performed. By using the RSSD technique, the number of states becomes a degree of freedom to trade performance against complexity. As we will see in the numerical results, when the number of trellis states is lower than M^L, a performance loss must be expected.

3.3.4 Channel estimation

The algorithms described in Sections 3.3.1 and 3.3.2 require the knowledge of vectors $\mathbf{p}_{i,n}$ up to a constant phase term. For this estimation problem, conventional LMS and recursive least-squares (RLS) algorithms [25] cannot be employed [77]. However, the noncoherent LMS and RLS techniques proposed in [77] can be straightforwardly extended to this case of two conditionally independent received signals and multiple samples per symbol intervals. In addition, they prove to be very robust against phase variations.

Denoting by $\hat{\mathbf{p}}_{i,n}^{(k)}$ the estimate of vector $\mathbf{p}_{i,n}$ at the k-th symbol interval and assuming, although not necessary since the PMD is a slowly-varying phenomenon, that this estimate is updated at each symbol interval, we now extend the update rule for the noncoherent LMS in [77] to our case. Without taking into account the decision delay of the VA, the channel estimate is updated as

$$\hat{\mathbf{p}}_{i,n}^{(k+1)} = \hat{\mathbf{p}}_{i,n}^{(k)} + \delta \left[\frac{q_{i,k}^{(\cdot)}}{|q_{i,k}^{(\cdot)}|} x_{i,k\eta+n}^{*} - s_{i,k\eta+n}^{*}(\hat{\mathbf{c}}_{k-L_i}^{k}) \right] \hat{\mathbf{c}}_{k-L_i}^{k} \tag{3.21}$$

where δ is the adaptation step-size and $q_{i,k}^{(\cdot)}$ can be either $q_{i,k}^{(N)}$ in (3.10) or $q_{i,k}^{(\alpha)}$ in (3.13). Since the VA provides decisions with a delay $3L \div 5L$, for rapidly-varying channels tentative decisions [25] or per-survivor processing [74] are usually adopted. However, in this case of a slowly-varying channel, the more reliable VA final decisions can be used without affecting the receiver performance.

It is worth mentioning that the recursive relation (3.21) to update the channel coefficients can be also equivalently expressed as a function of samples $\{z_{i,k\eta+n}\}$ and $\{|x_{i,k\eta+n}|\}$. As an example, in the case $\eta = 2$ we can express

$$\begin{aligned}
\hat{\mathbf{p}}_{i,0}^{(k+1)} &= \hat{\mathbf{p}}_{i,0}^{(k)} + \delta \left[\frac{g_{i,k}^{(\cdot)}}{|g_{i,k}^{(\cdot)}|} \frac{z_{i,2k+2} z_{i,2k+1}}{|x_{i,2k+2}||x_{i,2k+1}|^2} - s_{i,2k}^{*}(\hat{\mathbf{c}}_{k-L_i}^{k}) \right] \hat{\mathbf{c}}_{k-L_i}^{k} \\
\hat{\mathbf{p}}_{i,1}^{(k+1)} &= \hat{\mathbf{p}}_{i,1}^{(k)} + \delta \left[\frac{g_{i,k}^{(\cdot)}}{|g_{i,k}^{(\cdot)}|} \frac{z_{i,2k+2}}{|x_{i,2k+2}|} - s_{i,2k+1}^{*}(\hat{\mathbf{c}}_{k-L_i}^{k}) \right] \hat{\mathbf{c}}_{k-L_i}^{k} \,. \qquad (3.22)
\end{aligned}$$

3.3.5 Application to different channel encoders

The receivers proposed in [65, 66] represent a generalization, to larger observation windows and to modulation formats other than M-ary PSK, of the classical differential receiver [25]. In addition, these receivers can be adopted to decode not only differentially encoded symbols but also more powerful channel coding schemes provided that they are *noncoherently non-catastrophic* or known symbols are inserted to remove the phase ambiguities [65]. The VA branch metric in this case remains unchanged.

The same branch metric can be used for the algorithm by BCJR [51], implementing the MAP symbol detection strategy and employed as a component decoder in iterative decoding schemes for turbo codes [78], and also for message-passing algorithms used to decode low-density parity-check codes [79].

3.3.6 Polarization multiplexing (or transmit polarization diversity)

The receivers described so far process independently the signals on two orthogonal SOPs before the VA. The VA branch metric is then computed as a sum of two independent contributions, one for each SOP. In other words, polarization diversity is adopted at the receiver end. Obviously, this corresponds to doubling the receiver front end and we wonder whether it is possible to use this complexity increase in a more profitable way. That is to say whether polarization multiplexing (polMUX), or in other words transmit polarization diversity, can be also employed at the transmitter end in order to double the spectral efficiency while keeping the same receiver structure. The answer to this question is affirmative. In fact, the described receiver structure can be

adopted "as is" in the case of two independent data sequences $\mathbf{c}^{(1)} = \{c_k^{(1)}\}$ and $\mathbf{c}^{(2)} = \{c_k^{(2)}\}$ transmitted on two orthogonal polarizations. A single VA with branch metrics (3.12) or (3.15) can still be adopted, the only difference being the fact that the signal terms appearing in the branch metrics are now $s_{i,k\eta+n}(\mathbf{c}_{k-L_i}^{(1)k}, \mathbf{c}_{k-L_i}^{(2)k})$, i.e., they are a function of symbols of both sequences. Hence, a supertrellis must be built taking into account both sequences. Provided that a sufficient number of trellis states is considered, the described receiver is able to separate both signals and compensate phase distortions as GVD and PMD with no performance loss.

3.4 Interferometric IM/DD front end

Up to now, we described the proposed electronic processing taking no interest in the O/E front end. We will show that at least a couple of front end families can be equivalently employed, an interferometric one that will be described in this section, and a coherent one that will be addressed at the beginning of next chapter.

A (slightly modified) interferometric IM/DD front end, originally proposed in [55] for differentially encoded QPSK, can be also adopted. As already mentioned, due to the need of performing polarization diversity, the front end processing must be doubled. Hence, after the optical filter a polarization beam splitter (PBS) splits the signals on two orthogonal SOPs.

Given the signal on the i-th SOP at the input of the interferometric IM/DD front end, the samples $\{z_{i,k\eta+n}\}$ are obtained at the output. As a consequence, an additional photodetector must be also employed to obtain the samples $\{|x_{i,k\eta+n}|^2\}$ which are necessary, along with samples $\{z_{i,k\eta+n}\}$, to compute the VA branch metric in the form (3.19). With respect to the interferometric IM/DD front end used in the receivers for differentially encoded QPSK, the delay is T_c and not T, whereas the phase shifts are 0 and $-\pi/2$ instead of $\pi/4$ and $-\pi/4$. This is because, at the sampling time ℓT_c, we want to obtain the output samples $\mathrm{Re}[z_{i,\ell}]$ and $\mathrm{Re}[e^{-j\pi/2}z_{i,\ell}] = \mathrm{Im}[z_{i,\ell}]$, instead of $\mathrm{Re}[e^{j\pi/4}z_{i,\ell}]$ and $\mathrm{Re}[e^{-j\pi/4}z_{i,\ell}]$. Notice that, in this case, the analog prefiltering before sampling, mentioned in Section 3.2, must be performed in the optical domain. Hence, more attention must be devoted to the implementation of this filter. The overall receiver is shown in Fig. 3.1.

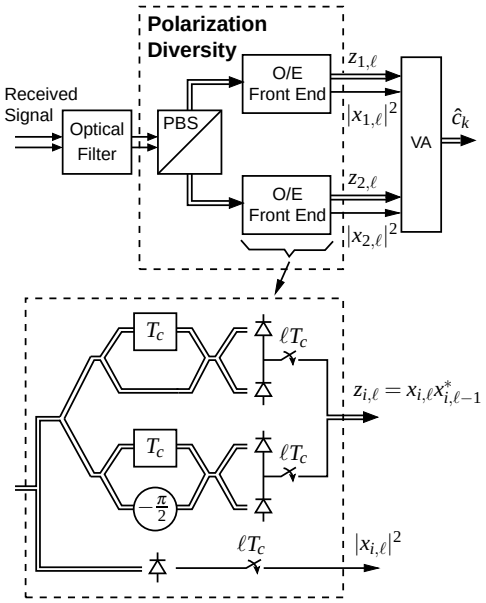

Figure 3.1: Receiver using the extended balanced front end.

3.5 Numerical results

We performed standard Monte Carlo simulations to evaluate the performance of the proposed receivers. The considered modulation formats are QPSK and 8-PSK with the standard differential encoding rule and 16-QAM with quadrant differential encoding. Gray mapping is employed to map bits onto M-ary symbols.

A NRZ pulse filtered through an electrical baseband Gaussian filter with -3dB bandwidth equal to $1/T$ is adopted at the transmitter. At the receiver, we use an optical 4th-order Gaussian filter with -3dB bandwidth equal to $2/T$ and the described modified interferometric IM/DD front end. Although the receive filter is only an approximation of the ideal analog prefilter, we verified that the related performance loss is less than one-tenth of dB. The receiver works using two samples per symbol interval and is based on the branch metrics (3.19). When employed, channel estimation is based on the updating rules (3.22).

In Figs. 3.2-3.5 we show BER curves versus E_b/N_0, E_b being the received signal energy per information bit. In all these figures, the presence of GVD,

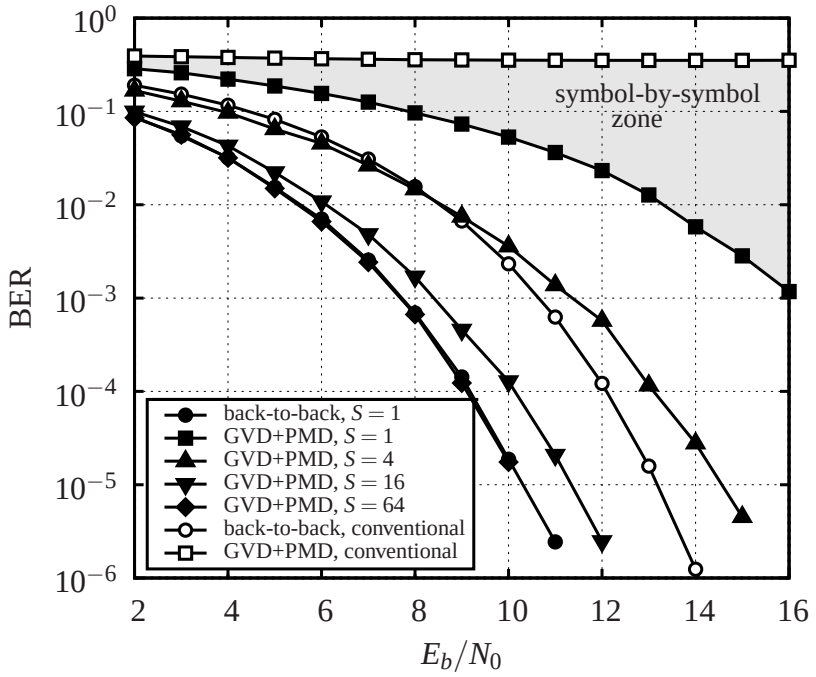

Figure 3.2: Performance of the proposed algorithm for a QPSK transmission.

2nd order PMD, and phase noise is considered. As already mentioned, phase noise can be characterized by its linewidth $\Delta\nu T_b$ normalized to the bit rate. In the reported BER results, the following values are considered: $\gamma = 1.6$, $\rho = 0.5$, $\Delta\tau = 3T_b$, $\Delta\tau_\omega = 0.4T_b^2$, $q_\omega = 0.4T_b$, and $\Delta\nu T_b = 0.25 \cdot 10^{-3}$, corresponding to $\Delta\nu = 10$ MHz for a bit rate $1/T_b = 40$ Gb/s.

In Fig. 3.2, QPSK is considered. The performance of the proposed algorithm is shown for a different number of trellis states. Indeed, thanks to the RSSD technique this value can be chosen arbitrarily. Note that $S = 1$ means that a symbol-by-symbol receiver with decision feedback is considered. Although the channel estimation algorithm is employed in a completely blind manner (that is no training symbols are used in the acquisition phase), no perceptible difference has been observed with respect to the case of perfect knowledge of the received pulse. For this reason, in the remaining figures a perfect knowledge of the channel is assumed. As a comparison, the performance of the proposed algorithm in the absence of GVD and PMD (the back-to-back case)

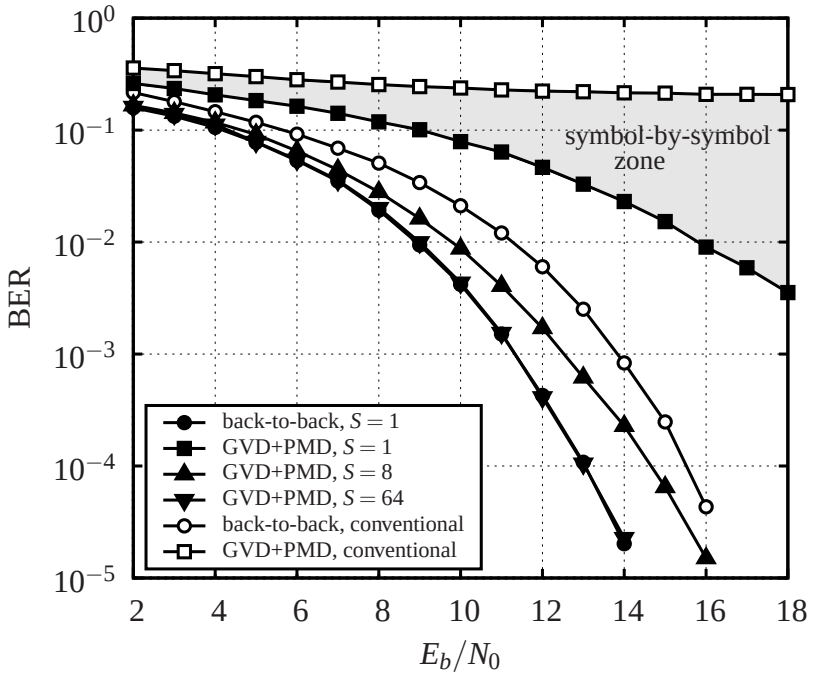

Figure 3.3: Performance of the proposed algorithm for an 8-PSK transmission.

and that of the conventional symbol-by-symbol receiver [55] in the absence and in the presence of GVD and PMD are also shown. Note that in the back-to-back case the amount of ISI is very limited. As a consequence, the proposed receiver with $S = 1$ (solid circles) practically attains the optimal performance. Also note that with respect to the conventional receiver (hollow circles), the proposed algorithm exhibits a gain of more than 2 dB in the back-to-back case. This is in line with the results in [65]. In addition, in the presence of GVD and PMD, the performance of the conventional receiver rapidly degrades whereas the proposed algorithm is able to perfectly compensate for both PMD and GVD, provided that a sufficient number of trellis states is adopted. As already mentioned, the proposed receiver with $S = 1$ is a symbol-by-symbol receiver with decision feedback. This can be considered as the best non-linear decision feedback equalizer we can design. Hence, in the presence of GVD and 2nd-order PMD with the described parameters, the region between the curves with solid and hollow squares represents the "symbol-by-symbol zone", that

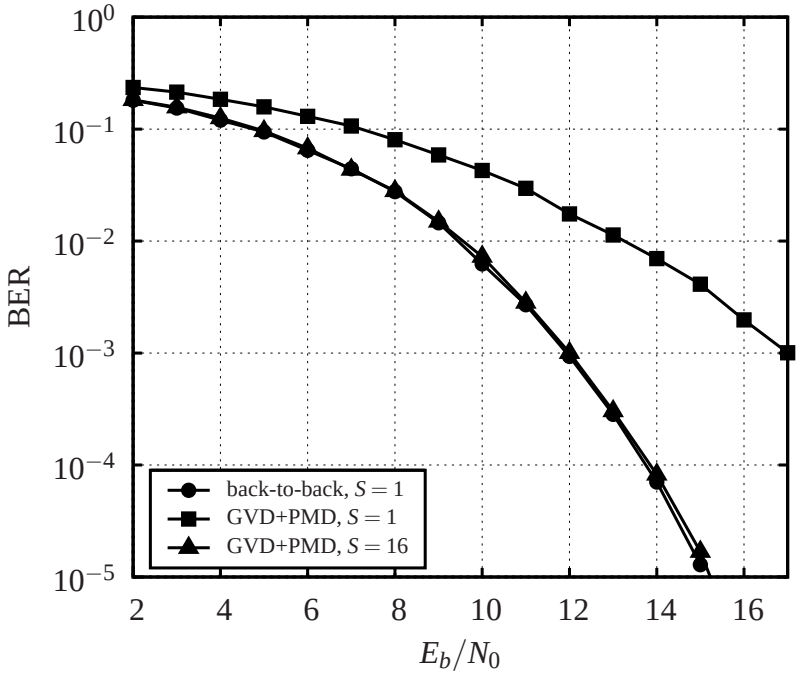

Figure 3.4: Performance of the proposed algorithm for a 16-QAM transmission.

is no matter what electronic processing is used, when we do not employ the VA, the performance curve will fall in that zone.

Similar considerations hold for the 8-PSK and the 16-QAM formats considered in Fig. 3.3 and Fig. 3.4, respectively. In the 16-QAM case, a conventional receiver is not defined, so no comparison is performed.

Fig. 3.5 deserves a more detailed comment. In this case, we consider a QPSK transmission with transmit polarization diversity and we show the overall BER performance. As can be seen, provided that a sufficient number of states is adopted, the performance of the back-to-back case, which coincides with that in Fig. 3.2, can be attained. This means that the proposed receiver is not only able to perfectly compensate PMD and GVD but is also able to separate both signals.

To assess the robustness to GVD and PMD of the proposed schemes with limited complexity, in Figs. 3.6(a)-(b) we show, for the QPSK modulation, the

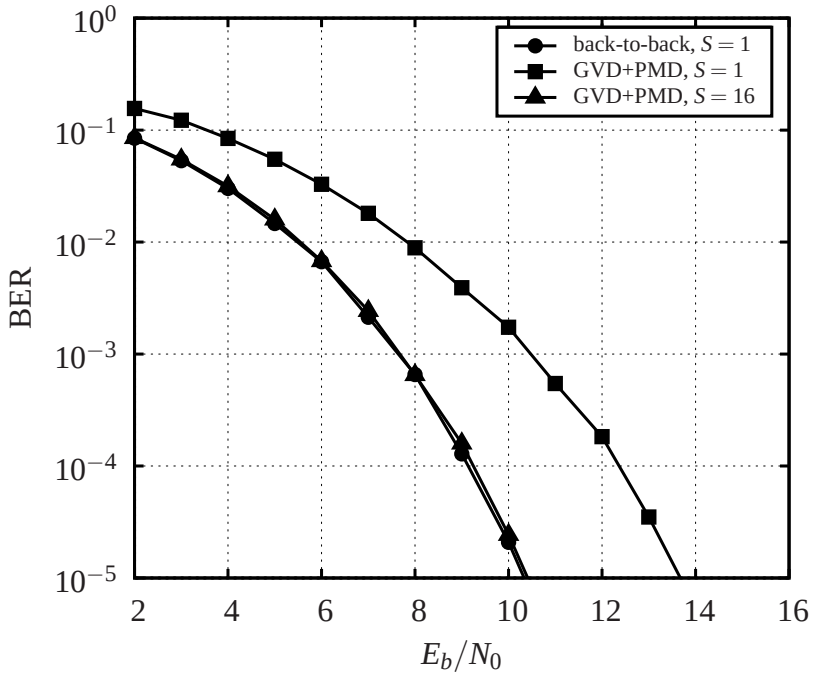

Figure 3.5: Performance of the proposed algorithm for a QPSK transmission with transmit polarization diversity.

values of E_b/N_0 necessary to obtain a BER of 10^{-4} for a different amount of dispersion. In Fig. 3.6(a), the presence of GVD only is considered. On the contrary, in Fig. 3.6(b) only 1st order PMD, with $\rho = 0.5$ and different values of $\Delta\tau$, is present. The robustness of the conventional receiver is also shown. The proposed receiver, even in a symbol-by-symbol configuration ($S = 1$) is able to guarantee a significant performance improvement with respect to the conventional one. In addition, by increasing the number of trellis states of the VA, the robustness is greatly increased. It is worth mentioning the particular behavior in the presence of 1st order PMD. In fact, it can be observed that, when the VA trellis complexity is not sufficient to guarantee a perfect compensation, the performance penalty is limited to at most 3 dB. This is due to the following reason. The receiver with polarization diversity is able to perfectly resolve the slow and fast PSP in the 1st-order PMD approximation. When the number of trellis states is not sufficient to describe the ISI associated to the

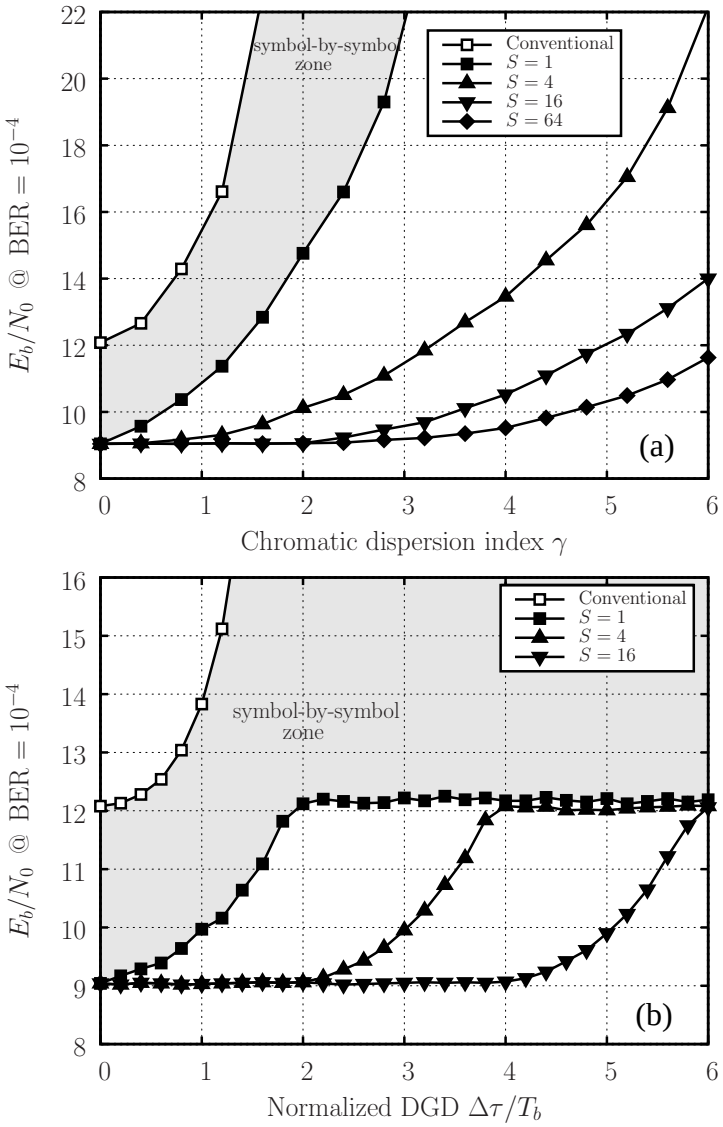

Figure 3.6: Values of E_b/N_0 necessary to obtain a BER of 10^{-4} versus the value of $\Delta\tau/T_b$, (a) when GVD only is present, (b) when 1st order PMD only is present. The considered modulation format is a QPSK.

slow PSP, this PSP is perfectly canceled out by decision feedback implicit in the RSSD technique. Hence, half of the received signal power is canceled out, thus producing the 3 dB loss. In general, when a different power splitting is observed, the loss is lower, that is $\rho = 0.5$ is the worst case. In any case, this asymptotic loss can be predicted by considering the amount of power canceled out. A similar behaviour is observed in the case of a 16-QAM constellation, as can be seen in Fig. 3.7(a), referring to the case of GVD only, and in Fig. 3.7(b) devoted to the case of 1st order PMD only.

Finally, in Fig. 3.8, the robustness of the proposed receivers to GVD and 1st order PMD (with $\rho = 0.5$) is reported by showing the contour curves corresponding to an E_b/N_0 penalty of 4 dB with respect to the back-to-back case and for a BER of 10^{-4} . As can be observed, the improvement with respect to the conventional receiver is impressive. In addition, irrespective of the number of adopted trellis states, there is a given amount of chromatic dispersion that is always tolerated for any amount of instantaneous DGD. This is due to the already observed property that, in the proposed receivers, the PMD produces a loss of at most 3 dB. Moreover, the proposed receivers show the interesting property that small values of DGD can even improve the robustness against GVD, as it was observed for duobinary modulation in [80].

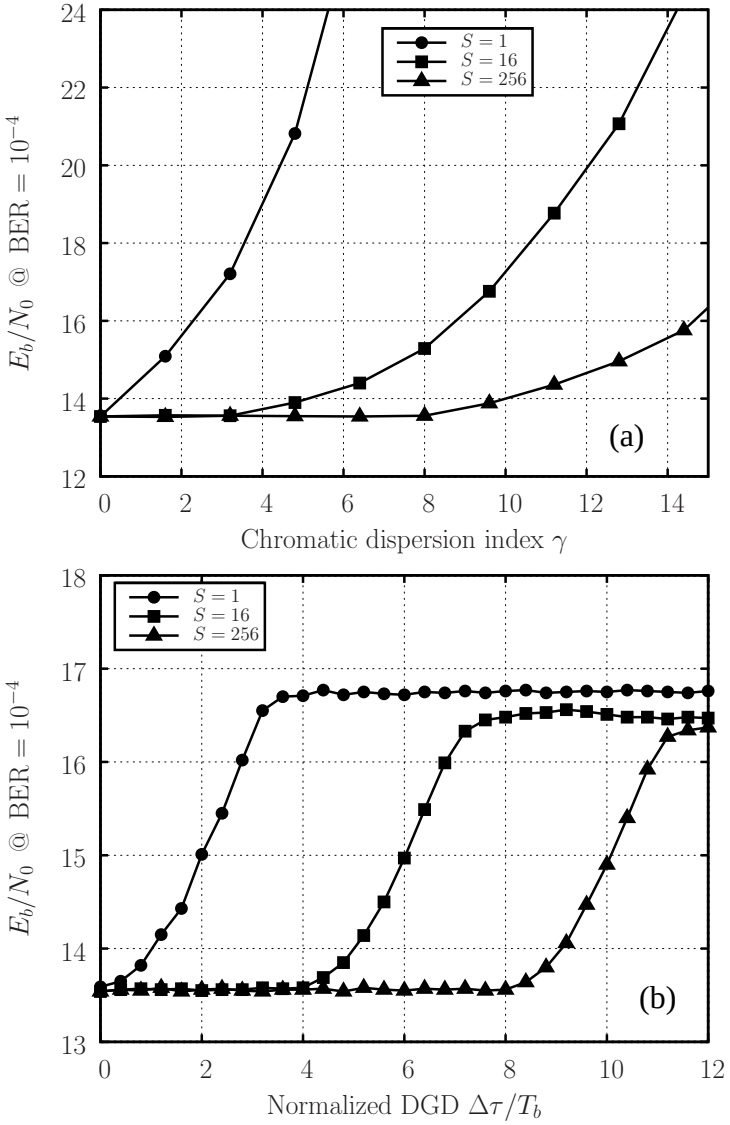

Figure 3.7: Values of E_b/N_0 necessary to obtain a BER of 10^{-4} versus the value of $\Delta\tau/T_b$, (a) when GVD only is present, (b) when 1st order PMD only is present. The considered modulation format is a 16-QAM.

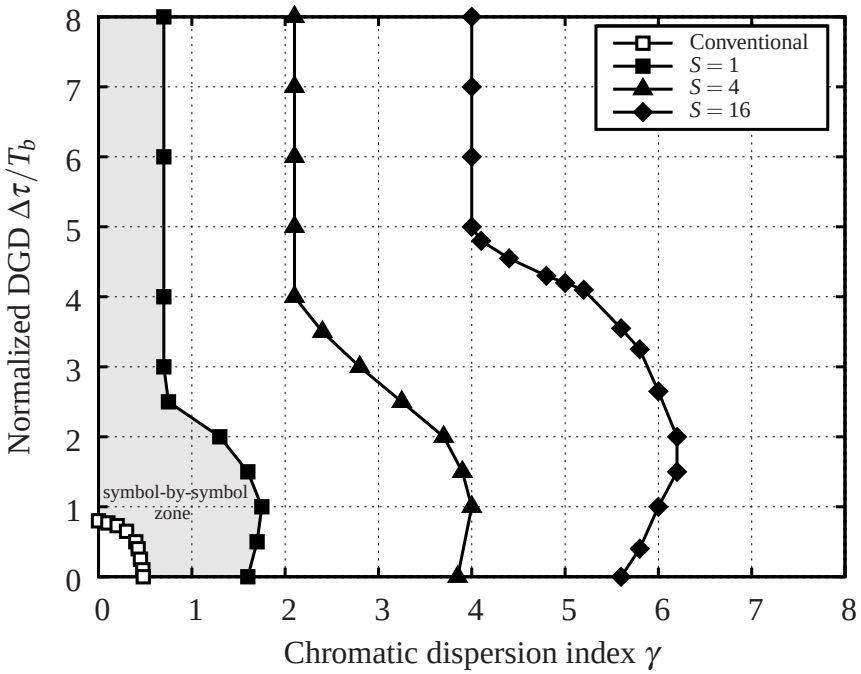

Figure 3.8: Contour curves corresponding to an E_b/N_0 penalty of -4 dB for a BER=10^{-4}, c versus $\Delta\tau/T_b$ and γ. The QPSK modulation is considered.

Chapter 4

Multilevel modulations with a coherent front end

In recent years the research on optical communications has rediscovered the field of coherent[1] detection, that was already investigated in early nineties, but lately abandoned when the advent of optical amplifiers [81] drove the development and diffusion of IM/DD systems; in fact, coherent detection, although sensitive to both the amplitude and phase of the received optical signal [2], suffered from more demanding technology issues. Nowadays, the need for advanced modulation formats with higher spectral efficiency makes coherent detection a viable solution for optical systems.

The effectiveness of post-detection processing in coherent optical systems was known since these receivers were firstly devised; in [82] it was demonstrated how simple fractionally-spaced equalizers, in bot homodyne and heterodyne systems, could significantly reduce linear fiber impairments with any modulation formats. More recently, theoretical and experimental works have been carried out to explore the post-processing capability of improving the

[1]There is some confusion between the terminology used in the optical community and that used in the context of wireless communication systems. In wireless systems, the term "coherent" refers to the knowledge of the phase of the received signal, that is a coherent receiver is designed assuming that the phase of the received signal is known or separately estimated. On the contrary, a noncoherent receiver is designed assuming that the channel phase is unknown and modeled as a random variable or a stochastic process. This terminology is employed in [65, 66, 70–73, 77–79]. In optical systems, the term "coherent" refers to the coherence of the optical carrier whereas the terms "synchronous" or "asynchronous" refer to a processing which assumes or not the knowledge of the channel phase.

performance of next generation optical communication systems with coherent detection. In particular, homodyne detection allows for simple and effective post-detection processing [83], but the demand for a stable and accurate local oscillator locked to the optical carrier still entails problems dealing with the feasibility of optical PLL. In [84], experimental results on digital signal processing (DSP) with a digital PLL demonstrate GVD compensation, under the hypothesis of known channel parameters at the receiver; feedforward carrier recovery, polarization control with algorithms tolerant to phase noise are presented in [85, 86], whereas in [87, 88] phase diversity techniques are experimentally validated with M-PSK; then, in [89], finite-impulse response (FIR) filtering and carrier and phase estimation demonstrate GVD compensation and polarization multiplexing; in [90] two-dimensional fractionally-spaced FFEs, similar to those in [82], is implemented, with closed-form expression for tap coefficients and in-depth examination of different oversampling rates.

Our work was firstly focused on sequence detection, since the solution proposed in Chapt. 3 can be adopted with a coherent front end. Though, it is not necessary to resort to complex strategies based on the VA, in this case, because equalization techniques provides the same results with lower complexity and simpler architectures. We investigated a fractionally-spaced feedforward FIR structure, where the signal received by a coherent front end with polarization diversity is processed by a two-dimensional filter whose purpose is to invert the fiber Jones matrix transfer function. The tap coefficient adaptation is based on the minimum mean square error (MMSE) criterion, implementing a filter matched to the received pulse. In order to improve the robustness of the receiver against phase noise, the detection strategy is derived from the noncoherent metric of Chapt. 3, in a symbol-by-symbol version, which exploits the information brought by previous samples, providing excellent phase noise tolerance with a simpler solution than those in [86, 87]. A proper updating algorithm for the MMSE criterion is therefore presented, which permits a faster convergence to the optimal configuration and reliable tracking of channel slow-time variations. The proposed algorithm acts practically as a PLL, in addition resolving timing jitters. It also shows a non-negligible tolerance to the frequency offsets between the optical carrier and the local oscillator; nevertheless, an automatic frequency control (AFC) is to be included in the processing unit to keep such offsets within an acceptable range.

4.1 Coherent front end with a noncoherent metric

An alternative implementation of the front end in §3.3 is that used in *coherent systems* based on *homodyne* or *heterodyne* techniques [2]. As illustrated in Fig. 4.1 for the homodyne case, after a preliminary optical filtering, two orthogonal SOPs are split through a PBS. They are then separately combined with the optical field of a local oscillator laser (LO) in a 2×4 90° hybrid [58] and detected with two balanced photodetectors. In this way the two received signals, one for each SOP, are converted in the electrical domain, in practice performing a frequency conversion. When the local-oscillator frequency coincides with that of the received signal, homodyne detection is performed. Otherwise, in heterodyne schemes a second frequency conversion stage in the electrical domain is necessary [2, 58].

Since the receivers described in the previous chapter do not require the knowledge of the channel phases θ_i, that is they represent two ways of performing asynchronous processing (with optimal performance), it is not necessary to track the channel phases with an optical (in the case of homodyne detection) or an electrical (in the case of heterodyne detection) phase-locked loop (PLL) but only the frequency must be tracked by an automatic frequency control (AFC), thus simplifying the receiver implementation. Regarding the analog prefiltering before sampling, mentioned in Section 3.2, it can be performed either in the optical or in the electrical domain, in this latter case either at baseband or at intermediate frequency.

4.2 Receiver based on a linear processing

4.2.1 System model and receiver structure

The system model and its low-pass equivalent are shown in Fig. 4.2. In the considered system, polarization multiplexing is employed. To this end, two independent sequences $\{a_{i,k}, k = 1, 2, \ldots, K-1\}$, $i = 1, 2$, of $K-1$ complex symbols belonging to an M-ary complex alphabet undergo separate differential encoding [68], thus generating two sequences[2] $\{d_{i,k}, k = 0, 1, \ldots, K-1\}$, $i = 1, 2$, of K complex symbols each, belonging to the same alphabet. For

[2] In Chapt. 3 the encoded sequence was denoted with symbols $\{c_k\}$. It is now referred to with $\{d_k\}$ because we reserve letter c to denote equalizer tap coefficients.

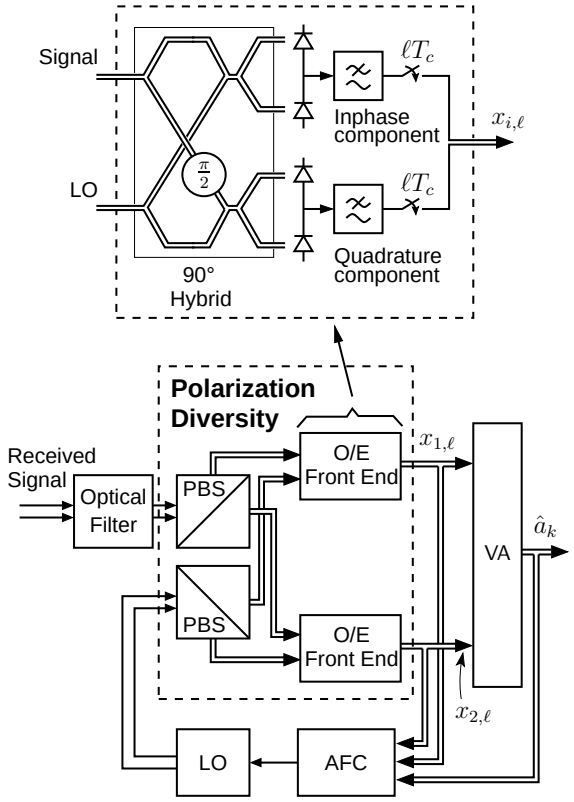

Figure 4.1: Receiver using coherent homodyne detection.

conciseness, we will denote

$$\begin{aligned} \mathbf{a}_k &= [a_{1,k}, a_{2,k}]^T \\ \mathbf{d}_k &= [d_{1,k}, d_{2,k}]^T . \end{aligned} \tag{4.1}$$

Without loss of generality, in the numerical results we will consider classical PSK signals, for which the standard differential encoding rule is employed, and square QAMs for which quadrant differential encoding is adopted [68] (see also [65, Section V-A] for a concise description). However, our derivations can be also applied to other alphabets, for example amplitude- and phase shift keying (APSK) modulations, whose signal constellations are composed of more concentric rings of PSK points.

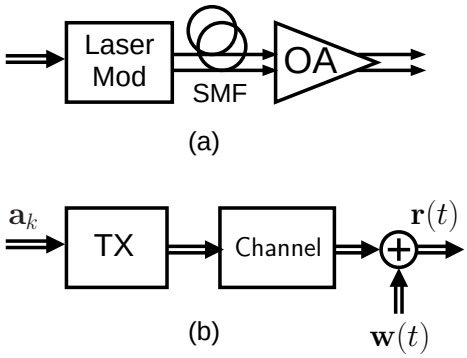

Figure 4.2: (a) System model. (b) Low-pass equivalent.

These two symbol streams are launched, after linear modulation, on two orthogonal SOPs of a single-mode fiber. We can express the low-pass equivalent of the transmitted signal components as

$$[s_1(t), s_2(t)]^T = \sum_{\ell} \mathbf{d}_\ell p(t - \ell T) = \sum_{\ell} \mathbf{P}(t - \ell T)\mathbf{d}_\ell \tag{4.2}$$

where T is the symbol interval, $p(t)$ the transmitted pulse and

$$\mathbf{P}(t) = p(t)\mathbf{I}\,.$$

having denoted by $\mathbf{I}$ the 2×2 identity matrix. Without loss of generality, we suppose that $p(t) \otimes p^*(-t)|_{kT} = \delta_k$, where δ_k is the Kronecker delta and $\otimes$ denotes "convolution", that is we suppose that the transmitted pulse, after its corresponding matched filter, satisfies the condition for the absence of ISI. This ensures that in the back-to-back case, the optimal detector is symbol-by-symbol.

The two-dimensional impulse response of the SMF is represented by the 2×2 matrix $\mathbf{H}(t)$, accounting for both GVD and PMD and a possible constant unknown phase shift due to the phase uncertainty of the transmit and receive lasers.[3] Since its entrywise Fourier transform is a unitary matrix we have

$$\mathbf{H}(t) \otimes \mathbf{H}^H(-t) = \mathbf{H}^H(-t) \otimes \mathbf{H}(t) = \delta(t)\mathbf{I} \tag{4.3}$$

having denoted by $\delta(t)$ the Dirac delta function.[4]

[3]The phase noise will be taken into account later.

[4]In (4.3), the integral of a matrix $\mathbf{A}(t)$ is defined as a new matrix whose entries are the integrals of the entries of the original matrix $\mathbf{A}(t)$.

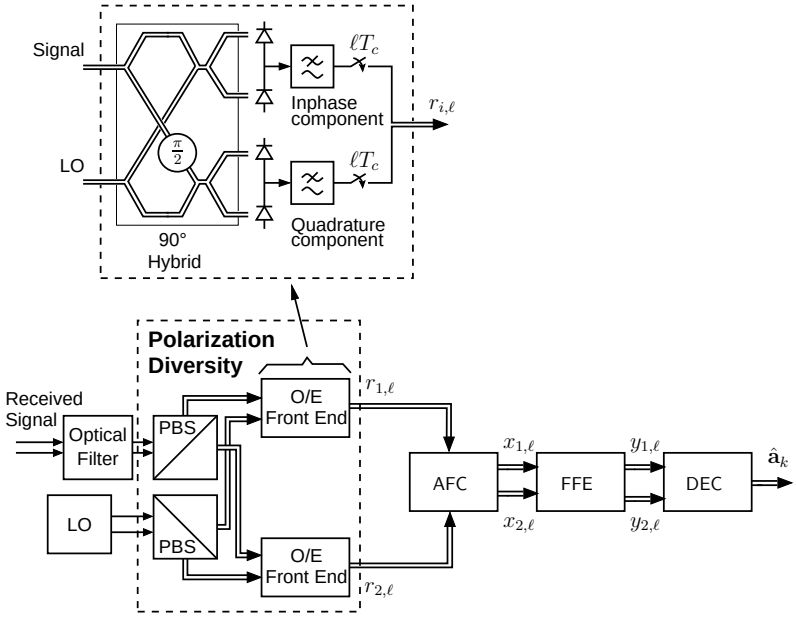

Figure 4.3: Receiver structure.

We consider the receiver as composed of an analog part, the opto-electronic (O/E) front end, devoted to signal demodulation and conversion from the optical to the electrical domain, and a digital part devoted to electronic processing (see Fig. 4.3). After a preliminary optical filtering, two orthogonal SOPs are split through a polarization beam slitter. They are then separately combined with the optical field of a local oscillator laser in a 2×4 90° hybrid [58] and detected with two balanced photodetectors. In this way the two received signals, one for each SOP, are converted in the electrical domain, in practice performing a frequency conversion. We suppose, since this is commonly considered feasible from a technological point of view, that the frequency offset f_D between the incoming signal and the LO laser is at most equal to the symbol rate. This allows to use, at the receiver, a free-running LO laser without resorting to a complex optical PLL delegating to the electronic processing the task of a fine frequency recovery. In other words, an intradyne scheme is

implemented [91].[5] Hence, the received signal can be expressed as

$$\mathbf{r}(t) = [r_1(t), r_2(t)]^T = \sum_{\ell} \mathbf{Q}(t - \ell T)\mathbf{d}_\ell e^{j2\pi f_D t} + \mathbf{w}(t) \tag{4.4}$$

where $f_D \leq 1/T$ is the above mentioned frequency offset between the incoming signal and the local oscillator, $\mathbf{Q}(t)$ is a 2×2 matrix given by $\mathbf{Q}(t) = \mathbf{H}(t) \otimes \mathbf{P}(t)$, and $\mathbf{w}(t) = [w_1(t), w_2(t)]^T$ collects the noise signal components on the above mentioned orthogonal SOPs.

Without loss of generality, we assume that the further processing is fully digital. To this purpose, a possible way of extracting sufficient statistics from the received signal $\mathbf{r}(t)$ is by means of sampling at the Nyquist rate [41]. In the following, we will assume that η samples per symbol interval are extracted from the signal, that is the sampling interval is $T_c = T/\eta$. This number of samples depends on the bandwidth of the received useful signal and the value of f_D. We assume that the optical filter has no effect on the useful signal and that its low-pass equivalent has a vestigial symmetry around $\eta/2T$ [41]. This latter condition ensures that the noise samples are independent and identically distributed complex Gaussian random variables with mean zero and variance $\sigma^2 = N_0\eta/T$ [41]. The samples of $\mathbf{r}(t)$ at discrete-time instants $\ell T_c = kT + nT_c = (k\eta + n)T_c$, $k = 0, 1, \ldots, K-1$, $n = 0, 1, \ldots, \eta - 1$, will be denoted as $\mathbf{r}_\ell = \mathbf{r}(\ell T_c) = [r_{1,\ell}, r_{2,\ell}]^T$.

The fine frequency recovery is then performed by means of an electrical AFC loop which performs closed-loop frequency estimation and compensation assuming that neither data nor clock information is available. The samples at its output will be denoted as $\mathbf{x}_\ell = [x_{1,\ell}, x_{2,\ell}]^T$. As will be demonstrated in the next section, an adaptive two-dimensional fractionally-spaced FFE of sufficient length is then able to perfectly compensate for GVD and PMD, thus allowing a simple classical symbol-by-symbol detection. In the presence of phase noise, a more robust symbol-by-symbol detection strategy with decision-feedback will be also described.

[5] An equivalent heterodyne scheme may be also conceived. In this case, the LO laser will perform the conversion from the optical to the electrical domain at an intermediate frequency (IF). A further LO in the electrical domain is then necessary. The frequency discriminator described later can still be adopted, provided that the frequency offset between the incoming signal converted to IF and the electrical LO is at most equal to the symbol rate.

4.2.2 Adaptive two-dimensional fractionally-spaced FFE

The adaptive two-dimensional fractionally-spaced FFE will be now described, deferring the discussion on the frequency estimation and compensation to Section 4.2.5. The reason is that the algorithm for frequency estimation will be chosen according to the robustness requirements of the detection blocks following it. In the hypothesis that the AFC block perfectly compensates the frequency offset f_D, the samples $\mathbf{x}_{k\eta+n}$ can be expressed as

$$\mathbf{x}_{k\eta+n} = \sum_{\ell} \mathbf{Q}_{\eta(k-\ell)+n}\mathbf{d}_\ell + \mathbf{w}_{k\eta+n} = \sum_{\ell=0}^{L_s-1} \mathbf{Q}_{\ell\eta+n}\mathbf{d}_{k-\ell} + \mathbf{w}_{k\eta+n} \tag{4.5}$$

having assumed that $\mathbf{Q}_\ell = \mathbf{Q}(\ell T_c)$ has a length of L samples or equivalently that $\mathbf{Q}(t)$ has a length of $L_s = \lceil L/\eta \rceil$ symbols. Samples $\mathbf{w}_{k\eta+n}$ after the frequency compensation are statistically equivalent to the samples of $\mathbf{w}(t)$ in (4.4). Hence, the two components $w_{1,k\eta+n}$ and $w_{2,k\eta+n}$ of $\mathbf{w}_{k\eta+n}$ are independent and identically distributed complex Gaussian random variables each with mean zero and variance $\sigma^2 = N_0\eta/T$, i.e., $E\{|w_{1,k\eta+n}|^2\} = E\{|w_{2,k\eta+n}|^2\} = \sigma^2 = N_0\eta/T$. Let us now suppose to filter the discrete two-dimensional signal $\mathbf{x}_{k\eta+n}$ with a two-dimensional filter with impulse response $\mathbf{C}_\ell$, $\ell = -L_c+1, \ldots, -1, 0$.[6] The output of this fractionally-spaced FFE is

$$\mathbf{y}_{k\eta+n} = \sum_{\ell=-(L_c-1)}^{0} \mathbf{C}_\ell \mathbf{x}_{k\eta+n-\ell}\,. \tag{4.6}$$

Assuming now that we know $\mathbf{Q}_\ell$, since from (4.3) it is

$$\mathbf{Q}_{-\ell}^H \otimes \mathbf{Q}_\ell = \delta_\ell \mathbf{I} \tag{4.7}$$

it is sufficient to choose $L_c = L$ and $\mathbf{C}_\ell = \mathbf{Q}_{-\ell}^H$ and the FFE output at symbol time will be

$$\mathbf{y}_{k\eta} = \mathbf{d}_k + \mathbf{w}'_{k\eta} \tag{4.8}$$

where $\mathbf{w}'_{k\eta}$ *is statistically equivalent* to $\mathbf{w}_{k\eta}$ since the FFE channel impulse response satisfy the condition (4.7) and hence it does not colour the noise. Hence, a symbol-by-symbol detector on each symbol stream is sufficient in this

[6] Without loss of generality, this FIR filter is supposed to be anticausal. In this way, we can avoid to consider the delay of $L-1$ samples at its output, thus simplifying the notation.

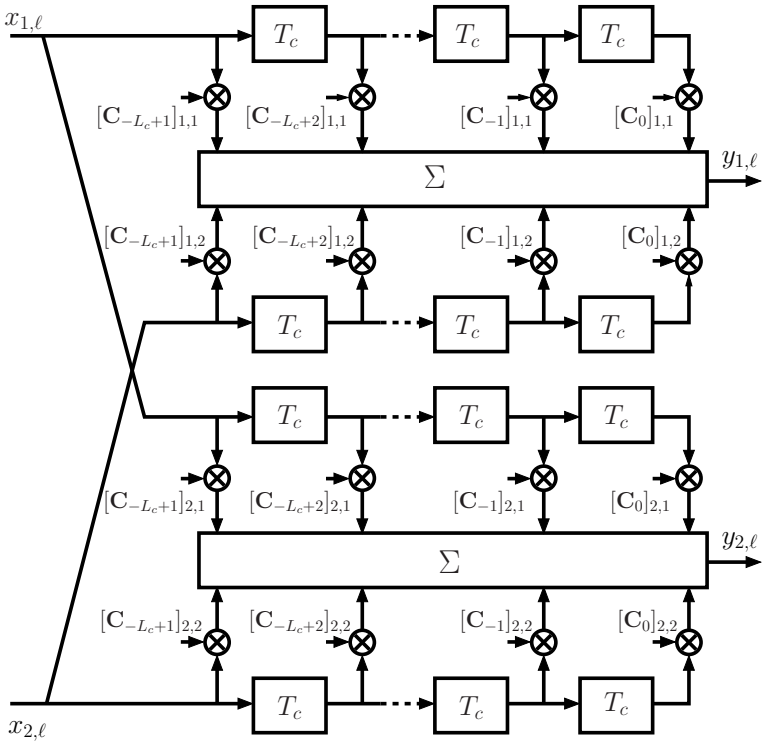

Figure 4.4: FFE structure.

case to obtain a decision on the transmitted symbols. This is not surprising, since we are implementing a two-dimensional matched filter and the discrete-time channel impulse response $\mathbf{Q}(t)$ satisfy the condition for the absence of ISI (see (4.7)). We can also state that in the back-to-back case, by filtering each of the two components of $\mathbf{x}_{k\eta+n}$ with a filter matched to $p(t)$, we would obtain an output statistically equivalent to $\mathbf{y}_{k\eta}$. Hence, the described receiver structure is able to attain a *perfect compensation of GVD and PMD*, provided that the AFC is able to estimate and compensate the frequency offset and a perfect knowledge of $\mathbf{Q}_\ell$ is available. Notice that this FFE also performs the phase compensation which is implicit in the knowledge of the channel impulse response $\mathbf{Q}_\ell$. Hence, an explicit phase estimation is not necessary. The FFE structure is shown in Fig. 4.4 where $[\mathbf{C}_\ell]_{k,n}$ denotes the (k,n) entry of $\mathbf{C}_\ell$.

In a practical implementation, the channel must be estimated, or in other

words, an adaptive FFE is necessary. This can be done in a decision-directed (DD) fashion with the algorithm described in Section 4.2.3. However, it must be noticed that when a DD algorithm is employed the channel will be estimated except for a couple of ambiguities. The first one affects each signal component and is related to the rotation symmetry angle ϕ of the employed alphabet. As an example, $\phi = \pi/2$ for QPSK and QAM alphabets. The second kind of ambiguity is related to a possible exchange of the transmitted streams. In other words, in the case of a DD channel estimation the FFE output at symbol time will be

$$\mathbf{y}_{k\eta} = \mathbf{R}\mathbf{d}_k + \mathbf{w}'_{k\eta} \tag{4.9}$$

where

$$\mathbf{R} = \begin{bmatrix} e^{j\phi m_1} & 0 \\ 0 & e^{j\phi m_2} \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}^i \tag{4.10}$$

where m_1 and m_2 are suitable integers and $i = 0, 1$—$i = 1$ means that the two polarizations have been exchanged. The first kind of ambiguity does not represent a problem since, as mentioned in Section 4.2.1, the two streams are differentially encoded. Regarding the second kind of ambiguity, this problem can be faced by inserting some known (pilot) symbols in the transmitted sequence. However, since a training sequence is usually inserted to ease the convergence of the FFE, these known symbols will also allow to correctly align the two streams at the transmission beginning. One may wonder if this is sufficient, that is, once the two streams are correctly aligned, it is possible to avoid the further insertion of known symbols without a subsequence exchange of the two polarizations due to the noise or to channel variations. In order to investigate this aspect, we run several computer simulations in which, after the initial training period that allows the FFE to reach convergence, simultaneous variations of the azimuth and ellipticity angles of the PSPs up to 10° have been injected every one hundred symbol intervals and never observed an exchange of the two polarization. Hence, one may conjecture that, unless a very strong channel variation occurs, and in this case a new training period is necessary for the FFE also, there is no risk to exchange the two polarizations even in the absence of other known symbols. Notice that this is not true for the first kind of ambiguity in the sense that although after the training sequence the FFE is able to converge ensuring $m_1 = m_2 = 0$ in (4.10), due to the ASE noise and the laser phase noise a phase slip is highly likely. Hence, the differential encoding described in Section 4.2.1 cannot be avoided.

4.2.3 Filter adjustment

The filter adjustment can be performed by resorting to the zero-forcing (ZF) criterion. However, we must force to zero the two-dimensional channel impulse response after the FFE, namely $\mathbf{C}_\ell \otimes \mathbf{Q}_\ell$, only at the time instant of the form $k\eta$ with $k \neq 0$ by neglecting its values at time instants $k\eta + n$ with $n \neq 0$. Hence, the MMSE criterion must be preferred since i) it converges to the two-dimensional matched filter when the FFE as a length L_c of at least L samples, as shown in Appendix A and ii) it guarantees a better performance with respect to the ZF criterion when the FFE as a length lower than $L_c < L$. In the Appendix A, a method to compute the expression of the optimal MMSE coefficients $\mathbf{C}_\ell$ is provided for each value of L_c.

The MMSE criterion is based on the minimization of

$$J = E\{||\mathbf{y}_{k\eta} - A\mathbf{d}_k||^2\} = E\{||\mathbf{e}_k||^2\} \tag{4.11}$$

having defined $\mathbf{e}_k = \mathbf{y}_{k\eta} - A\mathbf{d}_k$, where $A = 1 + \frac{\sigma_d^2}{\sigma_d^2}$ and $\sigma_d^2 = E\{|d_{i,k}|\}$. For the orthogonality principle [92], this is equivalent to the following conditions

$$E\{\mathbf{e}_k \mathbf{x}_{k\eta-\ell}^H\} = 0 \quad , \quad \ell = -(L-1), \ldots, -1, 0$$

Therefore, the filter adjustment is made by using the following updating rule (decision-directed stochastic gradient with a symbol-time adjustment) [25]

$$\mathbf{C}_\ell^{(k+1)} = \mathbf{C}_\ell^{(k)} - \alpha \hat{\mathbf{e}}_k \mathbf{x}_{k\eta-\ell}^H \,. \tag{4.12}$$

4.2.4 Noncoherent detection strategy and noncoherent filter adjustment

In order to increase the receiver robustness towards phase noise, a symbol-by-symbol noncoherent detection strategy can be adopted. In particular, one of the noncoherent strategies in [65, 66] can be employed. According to them, symbols $d_{i,k}$, $i = 1, 2$, are detected as

$$\hat{d}_{i,k} = \underset{d_{i,k}}{\operatorname{argmax}} \left[\left| y_{i,k\eta} d_{i,k}^* + q_{i,k} \right| - \frac{|d_{i,k}|^2}{2} \right] \tag{4.13}$$

where

$$q_{i,k} = \sum_{n=1}^{N} y_{i,(k-n)\eta} \hat{d}_{i,k-n}^* \tag{4.14}$$

N being a design integer parameter, in the case of the strategy in [65], whereas $q_{i,k}$ is recursively computed as

$$q_{i,k} = \beta q_{i,k} + y_{i,(k-1)\eta}\hat{d}^*_{i,k-1} \tag{4.15}$$

$0 \leq \beta \leq 1$, in the case of the strategy in [66].

When a noncoherent detection strategy is adopted, it is more convenient to substitute the described filter adjustment rule (4.16) with the following one

$$\hat{\mathbf{C}}_\ell^{(k+1)} = \hat{\mathbf{C}}_\ell^{(k)} - \alpha(\mathbf{y}_{k\eta} - A\mathbf{q}_k \odot \hat{\mathbf{d}}_k)\mathbf{x}^H_{k\eta-\ell} \tag{4.16}$$

where $\mathbf{q}_k = (q_{1,k}, q_{2,k})^T$ and $\odot$ denotes the Hadamard product (entrywise product or Schur product). This update rule is the extension to the case of two polarizations of that described in [93] and converges to the same minimum of the rule (4.12), although more robust in the presence of phase noise.

4.2.5 Frequency estimation and compensation

Frequency estimation and compensation must be performed by using a closed-loop non-data-aided and non clock-aided algorithm. The AFC scheme is shown in Fig. 4.5. We denote by $\hat{f}_{D,\ell}$ the frequency estimate at time ℓT_c. Using this estimate, the AFC input is derotated by a phase ϕ_ℓ related to $\hat{f}_{D,\ell}$ by the recursive equation

$$\phi_{\ell+1} = \phi_\ell + 2\pi\hat{f}_{D,\ell}T_c$$

obtaining the signal $\mathbf{x}_\ell = \mathbf{r}_\ell e^{-j\phi_\ell}$ sent at the FFE input. The phasor $e^{-j\phi_\ell}$ is obtained through the look-up table shown in the figure. The signal $\mathbf{x}_\ell$ is also sent to the input of the frequency error detector (FED) which computes the error signal ϵ_ℓ used to update the frequency estimate according to the recursion

$$\hat{f}_{D,\ell} = \hat{f}_{D,\ell-1} + \gamma\epsilon_\ell$$

where γ is a proper step-size.

Several classical FEDs can be chosen according to the desired performance/complexity trade-off. In our simulations, we considered the delay& multiply (DM) FED and a FED derived from the maximum likelihood (ML) criterion [94]. Both FEDs suffice to our purposes although, as known, the DM FED has a lower complexity but a worse performance. In this case, the error signal ϵ_ℓ is obtained as [94]

$$\epsilon_\ell = \mathrm{Im}\{x_{1,\ell}x^*_{1,\ell-1} + x_{2,\ell}x^*_{2,\ell-1}\}$$

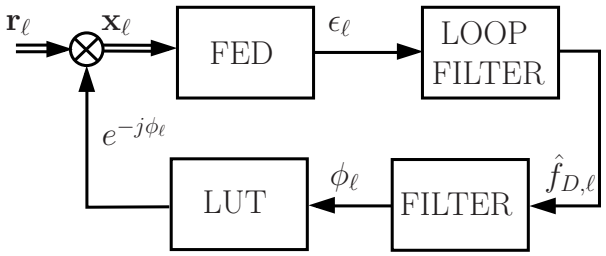

Figure 4.5: AFC structure.

whereas in the case of the ML-based FED, by following the same manipulations in [94] and using also (4.7), we obtain

$$\epsilon_\ell = \mathrm{Im}\{x'_{1,\ell} x''^{*}_{1,\ell} + x'_{2,\ell} x''^{*}_{2,\ell}\}$$

where $x'_{i,\ell}$ and $x''_{i,\ell}$ are obtained by filtering $x_{i,\ell}$ with two FIR filters with impulse responses $p_{-\ell} = p(-\ell T_c)$ and $\ell p_{-\ell}$, respectively.[7]

4.3 Numerical results

The performance of the proposed receivers is evaluated through Monte Carlo simulations. BER curves were carried out for differential QPSK with standard differential encoding and 16-QAM with quadrant differential encoding. In both cases, polarization multiplexing will be considered.

A nonreturn-to-zero (NRZ) pulse filtered through an electrical baseband Gaussian filter with -3dB bandwidth equal to $0.7/T$ is adopted at the transmitter. At the receiver, we use an optical 4th-order Gaussian filter with -3dB bandwidth equal to $2/T$ and the described coherent intradyne front end.

Figs. 4.6-4.7 show BER curves versus E_b/N_0, as in previous chapters. In all these figures, the presence of GVD and 2nd order PMD is accounted for as in Chapt. 3, whereas in Figs. 4.8-4.9 phase noise is considered. Phase noise is characterized by its linewidth $\Delta\nu$ normalized to the bit rate T_b, whereas the mismatch between the real frequency offset and its estimate will be denoted Δf. In the reported BER results, the following GVD and PMD values are considered: $\gamma = 1.6$, $\rho = 0.5$, $\Delta\tau = 3T_b$, $\Delta\tau_\omega = 0.4T_b^2$, $q_\omega = 0.4T_b$.

[7]For the commonly used transmitted waveforms, these FIR filters have a very small number of taps.

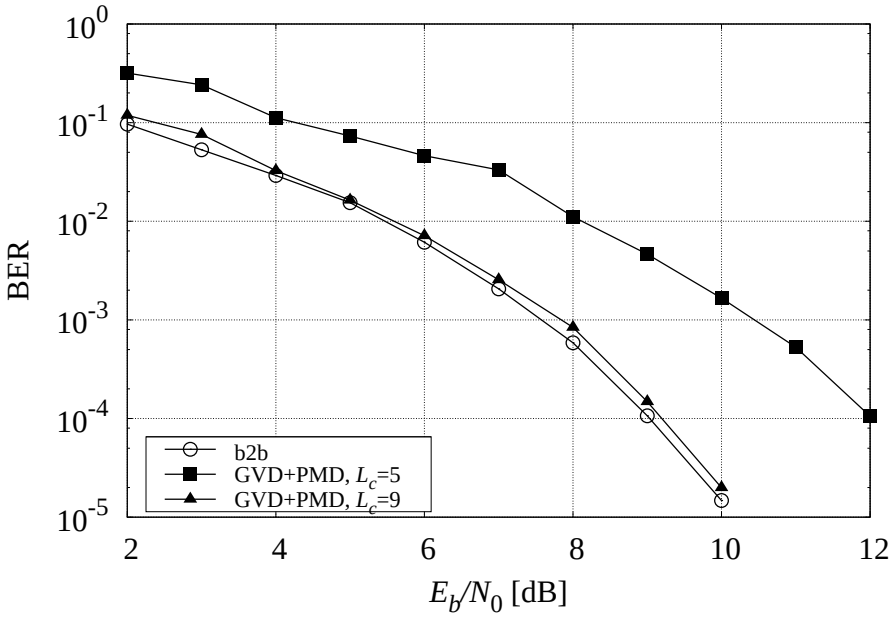

Figure 4.6: Performance of the proposed receiver for a QPSK transmission.

In Fig. 4.6, results for QPSK are reported. The performance of the proposed receiver is shown for a different number of equalizer taps, denoted by L_c. In this case, the updating algorithm for the equalizer taps cannot be employed in a completely blind manner (that is without training symbols in the acquisition phase), because the ambiguity between the two data sequences cannot be resolved by this algorithm. As a comparison, the performance of the proposed algorithm in the absence of GVD and PMD (the back-to-back case) is also shown. Note that the amount of ISI is significant, and in the absence of the equalizer the receiver would not work. Instead, the proposed receiver with $L_c = 9$ (squares) practically attains the optimal performance. In general, the proposed algorithm is able to perfectly compensate for both PMD and GVD, provided that a sufficient number of equalizer taps is adopted. Similar considerations hold for the 16-QAM format considered in Fig. 4.7, and it must be noted that the same results can be obtained with any modulation formats.

An exhaustive analysis of phase noise tolerance is then presented, in order to assess the performance of the detection strategy (4.13)-(4.14) compared to

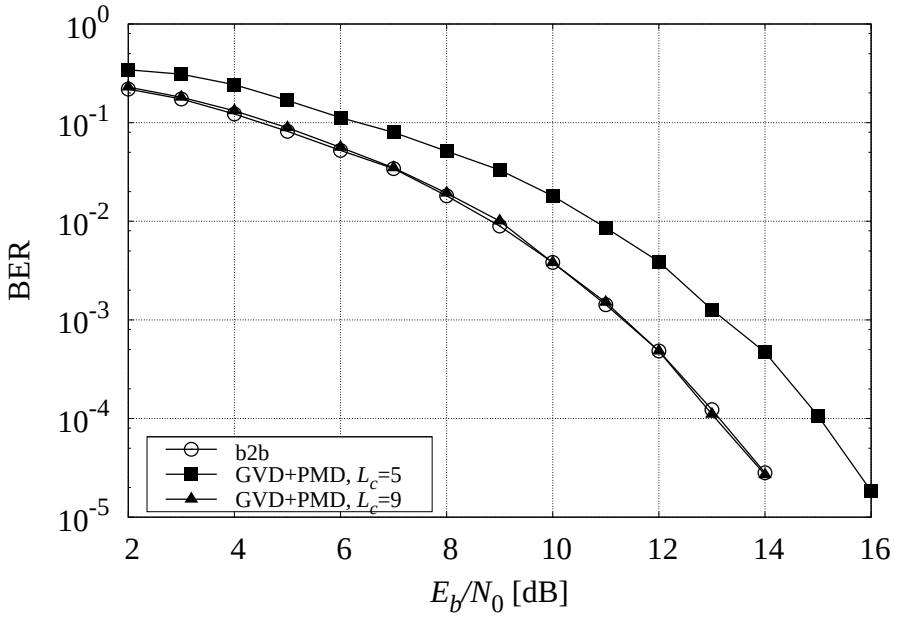

Figure 4.7: Performance of the proposed receiver for a 16-QAM transmission.

the standard symbol-by-symbol receiver. In Fig. 4.8 the performance of the standard symbol-by-symbol receiver with a different amount of phase noise is shown, with a fixed number of equalizer taps. It can be noticed that a significant penalty occurs. Also notice that in order to obtain satisfactory results, the equalizer step size must be tuned according to the phase noise amount. From Fig. 4.9, instead, the performance of this receiver is compared to that of detection strategy (4.13)-(4.14), in terms of penalty curves. It is clear that a significant improvement is provided by this novel detection strategy.

Finally, some results are provided on the implemented AFC necessary to avoid the presence of an optical PLL in the proposed receiver. As shown in Fig. 4.10, the fractionally-spaced FFE itself has a non-negligible tolerance to frequency offset, but still insufficient to hold a frequency mismatch of tens percent. Thus, two different algorithms were devised to implement a digital frequency control, as explained in § 4.2.5. The S-curve $S(f_D) = E\{f_D\}$ when the AFC loop is open [94] is plotted in Fig. 4.11 for the DM and the ML frequency error detectors. It can be observed that the acquisition range in the

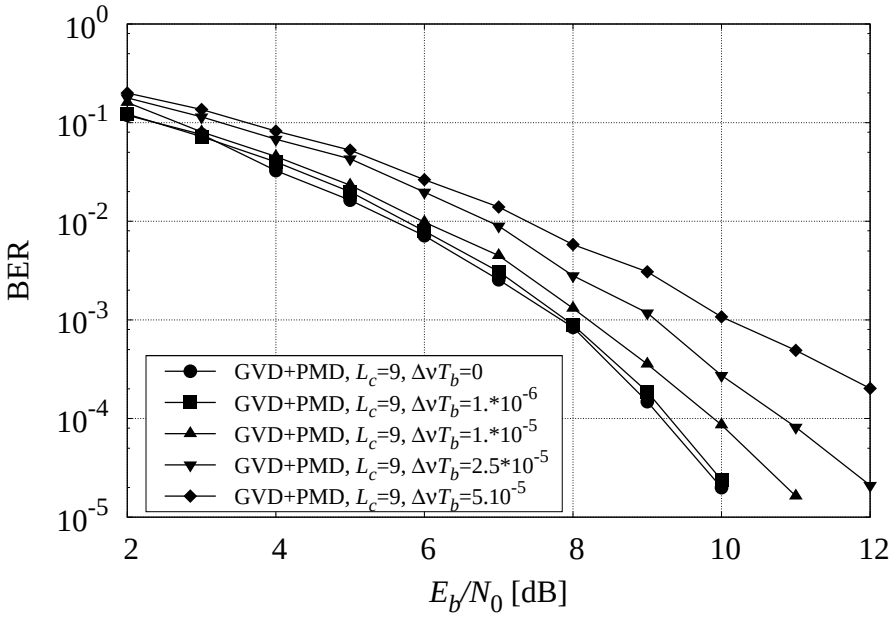

Figure 4.8: Performance of the standard symbol-by-symbol receiver for a QPSK transmission in presence of phase noise, referred to a 100 Gb/s rate.

DM case is on the order of the symbol rate in the case of the DM FED and slightly lower for the ML FED. Then, the standard deviation and the bias of the frequency estimator are given in Fig. 4.12 and Fig. 4.13, respectively. From these figures it can be noticed that the performance of the AFC is sufficient, for the considered value of the equivalent bandwidth $BeqT$ of the loop normalized to the symbol rate, to reduce the frequency error well below the robustness required in Fig. 4.10.

Appendix A

In this appendix, we show how to compute the coefficients of the FFE following the MMSE criterion. We also show that when $L_c = L$ the FFE implements the matched filter.

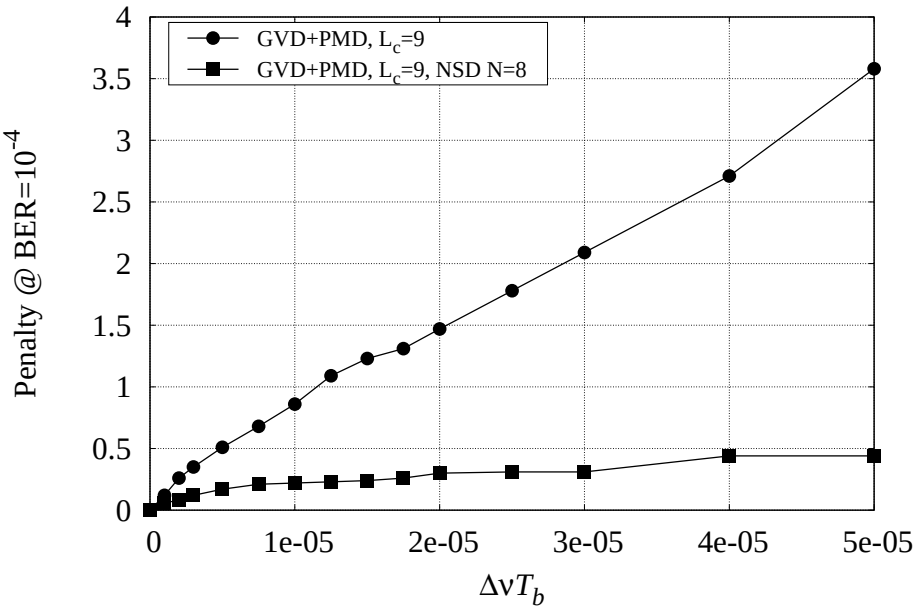

Figure 4.9: Values of E_b/N_0 necessary to obtain a BER of 10^{-4} versus the value of $\Delta f_D T_b$. The considered modulation format is a QPSK, refereed to a 100Gb/s rate.

The output of the FFE at discrete-time 0 has expression

$$\mathbf{y}_0 = \sum_{\ell=-(L_c-1)}^{0} \mathbf{C}_\ell \mathbf{x}_{-\ell} = \sum_{i=0}^{L_c-1} \mathbf{C}_{-i} \mathbf{x}_i$$

where from (4.5)

$$\mathbf{x}_i = \sum_{m=} \mathbf{Q}_{i-m\eta} \mathbf{d}_m + \mathbf{w}_i \,.$$

The mean square error to be minimized is

$$J = E\{||\mathbf{y}_0 - A\mathbf{d}_0||^2\} \,.$$

For the orthogonality principle [92], coefficients $\mathbf{C}_\ell$ must satisfy the condition

$$E\{(\mathbf{y}_0 - A\mathbf{d}_0)\mathbf{x}_\ell^H\} = 0 \quad , \quad \ell = 0, 1, \ldots, L_c - 1$$

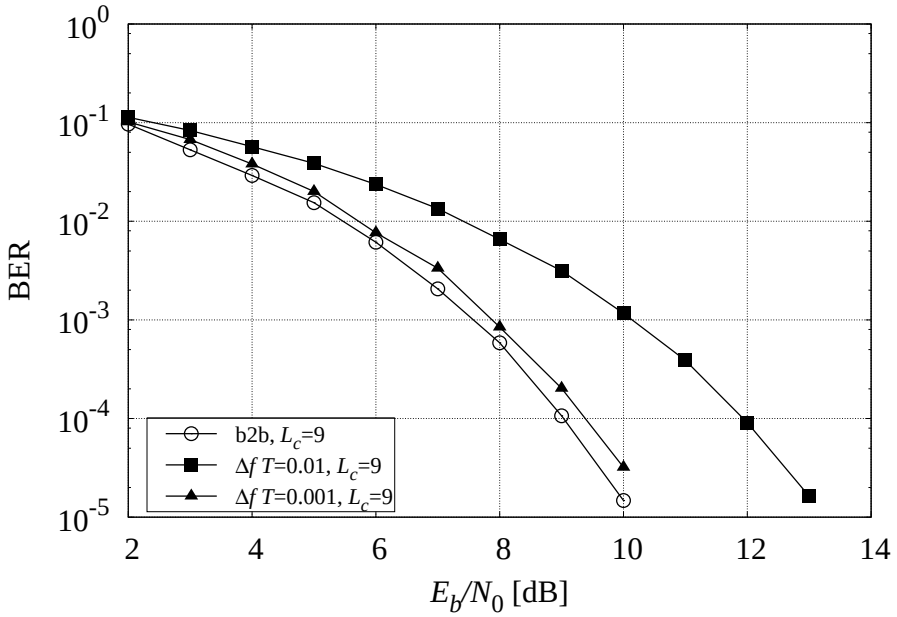

Figure 4.10: Performance for a QPSK transmission in presence of a frequency offset.

or, equivalently

$$E\{\mathbf{y}_0\mathbf{x}_\ell^H\} = AE\{\mathbf{d}_0\mathbf{x}_\ell^H\}$$

or

$$\sum_{i=0}^{L-1}\mathbf{C}_{-i}E\{\mathbf{x}_i\mathbf{x}_\ell^H\} = AE\{\mathbf{d}_0\mathbf{x}_\ell^H\}\,. \tag{A.1}$$

Let us compute the term

$$\begin{aligned} E\{\mathbf{x}_i\mathbf{x}_\ell^H\} &= \sum_m\sum_n \mathbf{Q}_{i-m\eta}E\{\mathbf{d}_m\mathbf{d}_n^H\}\mathbf{Q}_{\ell-n\eta}^H + E\{\mathbf{w}_i\mathbf{w}_\ell^H\} \\ &= \sigma_d^2\sum_m \mathbf{Q}_{i-m\eta}\mathbf{Q}_{\ell-m\eta}^H + \sigma^2\delta_{i-\ell}\mathbf{I} \end{aligned}$$

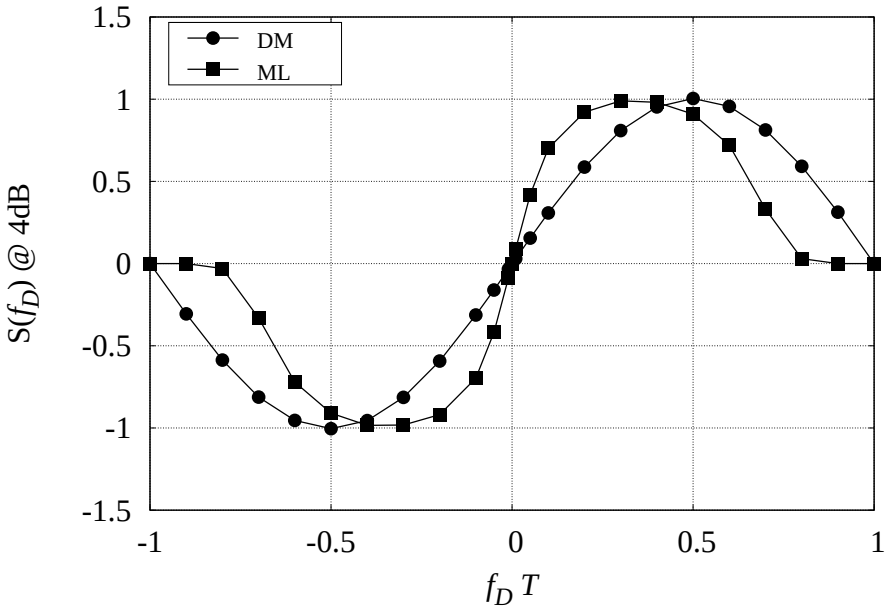

Figure 4.11: S-curve for a QPSK transmission.

since

$$E\{\mathbf{d}_m\mathbf{d}_n^H\} = \sigma_d^2\delta_{m-n}\mathbf{I}$$
$$E\{\mathbf{w}_i\mathbf{w}_\ell^H\} = \sigma^2\delta_{i-\ell}\mathbf{I}\,.$$

Similarly

$$E\{\mathbf{d}_0\mathbf{x}_\ell^H\} = \sigma_d^2\mathbf{Q}_\ell^H$$

As a consequence, by substituting in (A.1) and remembering the definition of A, we have

$$\sigma_d^2\sum_{i=0}^{L_c-1}\mathbf{C}_{-i}\sum_m\mathbf{Q}_{i-m\eta}\mathbf{Q}_{\ell-m\eta}^H + \sigma^2\mathbf{C}_{-\ell} = (1+\frac{\sigma^2}{\sigma_d^2})\sigma_d^2\mathbf{Q}_\ell^H$$

that is

$$\sigma_d^2\sum_{i=0}^{L_c-1}\mathbf{C}_{-i}\sum_m\mathbf{Q}_{i-m\eta}\mathbf{Q}_{\ell-m\eta}^H + \sigma^2\mathbf{C}_{-\ell} = \sigma_d^2\mathbf{Q}_\ell^H + \sigma^2\mathbf{Q}_\ell^H\,. \tag{A.2}$$

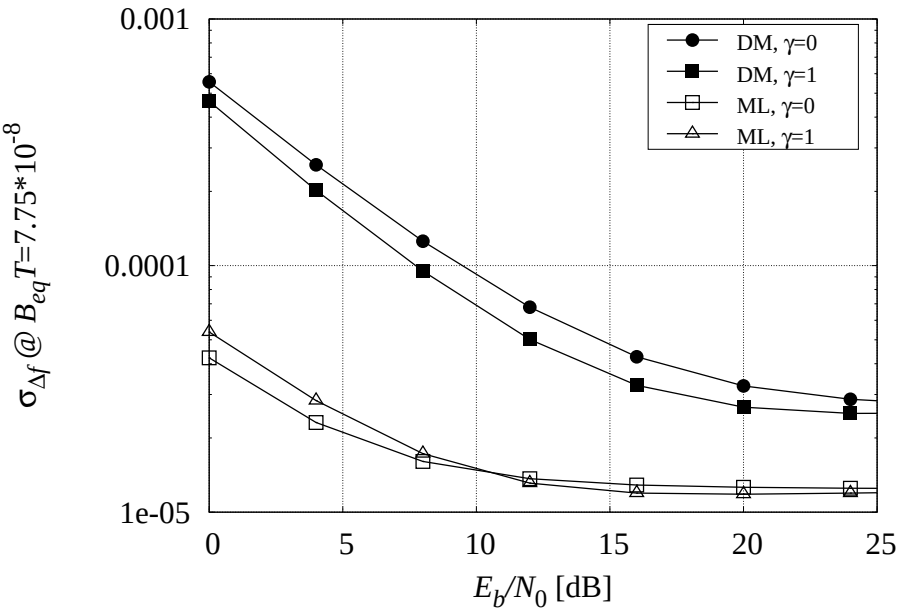

Figure 4.12: Standard deviation of the frequency estimator in presence of GVD.

Hence, a linear system with $4L_c$ equations and $4L_c$ complex unknown results, whose solution provides the desired coefficients $\mathbf{C}_\ell$. When $L_c = L$, it results $\mathbf{C}_\ell = \mathbf{Q}_{-\ell}^H$. In order to prove this statement, it is sufficient to verify that when $\mathbf{C}_\ell = \mathbf{Q}_{-\ell}^H$, the equation (A.2) is satisfied. In this case in fact

$$\sigma_d^2 \sum_{i=0}^{L-1} \mathbf{Q}_{-i}^H \sum_m \mathbf{Q}_{i-m\eta} \mathbf{Q}_{\ell-m\eta}^H = \sigma_d^2 \mathbf{Q}_\ell^H$$

or equivalently

$$\sum_m \left[\sum_{i=0}^{L-1} \mathbf{Q}_{-i}^H \mathbf{Q}_{i-m\eta} \right] \mathbf{Q}_{\ell-m\eta}^H = \mathbf{Q}_\ell^H \, .$$

since $\sum_{i=0}^{L-1} \mathbf{Q}_{-i}^H \mathbf{Q}_{i-m\eta} = \delta_m \mathbf{I}$, the equality is satisfied.

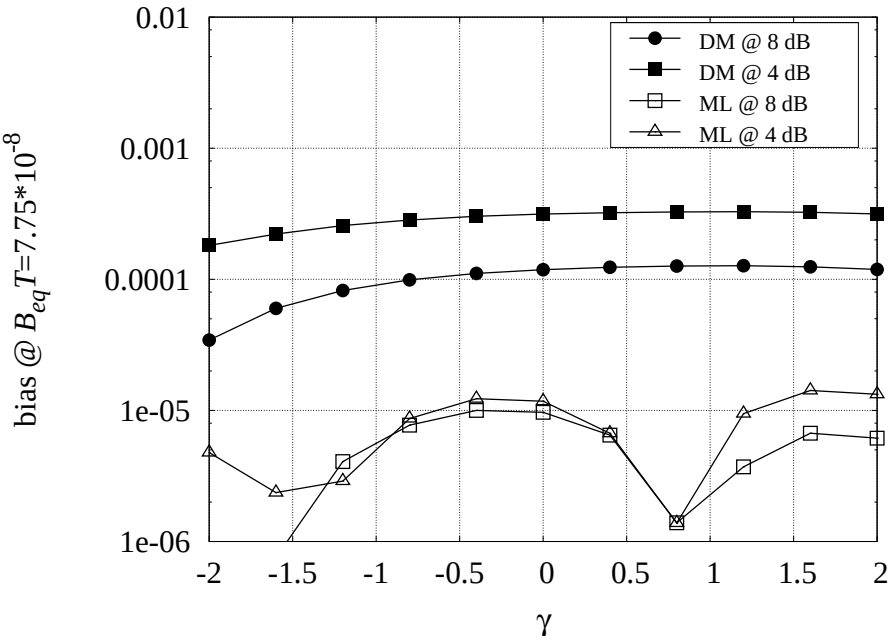

Figure 4.13: Bias of the frequency estimator as a function of GVD.

Chapter 5

Conclusions

The work presented in this thesis deals with post-detection signal processing in linearly impaired optical communication systems. Chromatic dispersion and polarization mode dispersion are the most severe sources of performance degradation in high-speed optical communications. Being chromatic dispersion a static impairment, dispersion maps can be devised in order to reduce the effects of pulse broadening. Though, at current data rates (beyond 10 Gb/s), residual dispersion could induce relevant penalties on system performance. On the other hand, polarization mode dispersion is a stochastic phenomenon, so that its effects could be hardly compensated if non-adaptive devices are envisaged. Optical compensators have demonstrated the capability of completely recover the penalties due to such linear impairments, but their feasibility and implementation costs prevent the diffusion of this kind of solutions. Then, since the early nineties, the research was aimed towards electrical equalization of fiber impairments, because of the reduced cost, the simple hardware implementation, the possibility of exploit techniques known from literature with few adjustments.

The implementation of feedforward and decision-feedback equalizers was envisaged as soon as first long-haul optically amplified fiber communication systems were devised. In intensity modulation direct detection systems, which are the only ones currently employed, the presence of the photodiode introduces a nonlinear transformation in the received signal which makes the effects of GVD and PMD worse. In such scenario, it is of deep interest the analysis of advanced electrical equalization strategies in order to state the ultimate performance of post-processing techniques. Thus, maximum likelihood sequence

detection has been considered in this work, since it is known to provide the best performance in classical radio and wireless channels. Nevertheless, specific fiber channel features make the implementation of MLSD a complex task, especially for the presence of the square-law detector that changes the statistics of the received signal, upon which this detection strategy is based, and making this channel a novel and unexplored field of research. In Chapt. 2 it is reported our contribution to this research area. We proposed a novel and simple branch metric cost function derived from approximations on the exact probability density function of the photo-detected signal, whose validity has been demonstrated through the comparison with the exact solution derived by numerical evaluation. Then, an analysis on sufficient statistics through oversampling was carried out, finding that by means of two samples per symbol interval, taken as independent, are sufficient to attain the best performance. Starting from the proposed metric, a novel evaluation method was derived, that allows to computing of bit error rates as low as desired without resorting to time-consuming Monte Carlo simulations, and under any configurations of system parameters. This method is especially useful to obtain results on outage probability in presence of PMD, and also in the presence of both PMD and GVD, for values that is not even possible to reach with standard simulations. These results show that a big impact on system performance is due to first-order PMD, and that the MLSD strategy, although effectively in reducing receiver penalties, is not able to completely cancel the effects of GVD and PMD as optical compensators do, even if it clearly outperforms other electrical equalization techniques as FFE and DFE.

More recently, great interest has arisen for high-order modulation formats, in that they allow a better exploitation of the channel bandwidth, together with requirements less demanding on electronic speed and a superior tolerance against fiber impairments. In Chapt. 3 it is described how a particular interferometric front end can be modified in order to realize a MLSD strategy with phase and phase-amplitude modulation formats able to perfectly compensate the effects of linear fiber impairments. The proposed solution is derived from the noncoherent sequence detection, it is basically based on the observation of past received samples to improve the receiver performance, and it is adapted to work with the oversampled interferometric receiver that intrinsically decodes differentially encoded modulations. The receiver architecture performs polarization diversity in order to attain the optimal compensation, in this way allowing also the possibility of exploiting polarization multiplexing

at the transmitter (i.e., two independent data streams are launched on two orthogonal polarizations). The receiver complexity (i.e. the number of trellis states) is addressed in the presented work, since it is closely related to the feasibility and effectiveness of the proposed detection strategy; complexity reduction techniques can be adopted to trade between the tolerable amount of power penalty due to fiber impairments and the achievable integrated hardware complexity. Nevertheless, with a small number of states the proposed receiver does not suffer the penalty (almost 2.5 dB in case of DQPSK) due to differential detection with respect to coherent detection. Moreover, this particular receiver offers a relevant robustness against phase noise which is known to heavily affect coherent optical systems with high-order modulation formats.

Finally, in Chapt. 4, coherent optical communications systems are treated. In these systems, an optical sinusoid provided by a local oscillator at the receiver is mixed with the received optical signal to shift its spectrum to lower frequencies; if the signal spectrum is shifted to baseband the system is homodyne, whereas if the the signal spectrum is shifted to an intermediate frequency it is heterodyne, or intradyne if the intermediate frequency if smaller than the signal bandwidth (actually, this is the case of our proposed receiver, since we do not employ an optical PLL and the frequency offset is compensated in the electrical domain). We focused on receivers in which the local oscillator light is nominally matched to the wavelength of the optical signal carrier, so that a baseband signal can be processed in the electrical domain. Coherent systems were studied and developed before the advent of optical amplifiers, but then abandoned because IM/DD systems became more convenient. Nowadays, technology improvements make this solution interesting for the development of high-order modulation formats in optical communications. The branch metric and the signal processing presented in Chapt. 3 can be modified to work with a coherent front end, leading to the same results. Though, in this case a simpler processing is possible, as explained in Chapt. 4. In fact, linear impairments as GVD and PMD, being only phase distortions, can be completely compensated with proper feedforward equalization. If polarization diversity is realized, and a matrix equalizer, using minimum mean square error algorithm to adapt tap coefficients, is devised (i.e. a transversal finite-impulse response filter 2×2 which implements the inverse of the channel 2×2 transfer function), it is demonstrated that back-to-back performance can always be attained, given a sufficient number of equalizer taps. Moreover, the proposed fractionally-

spaced equalizer makes the receiver insensitive to timing errors and robust against phase noise, which is a critical source of performance degradation. Then, in order to improve the receiver tolerance, we proposed a detection strategy based on a noncoherent metric, that allows minimum penalties due to phase noise. Furthermore, with the proposed solution there is no need for optical phase-locked loop, which are expensive and complex optical devices, but simple electrical automatic frequency controllers are sufficient to recover a frequency offset due to potential mismatches between the optical carrier and the local oscillator.

Bibliography

[1] E. Collett, *Polarized light. Fundamentals and applications.* New York, NJ: Dekker, 1992.

[2] G. Agrawal, *Fiber-optic communications systems*, 3rd ed. John Wiley & Sons, 2002.

[3] L. Kazovsky, "Phase- and polarization-diversity coherent optical techniques," *IEEE J. Lightwave Tech.*, vol. 7, pp. 279–292, Feb. 1989.

[4] J. Salz, "Coherent lightwave communications," in *AT & T Tech. J.*, vol. 64, Dec. 1985, pp. 2153–2209.

[5] L. L. Jeromin and V. W. S. Chan, "Modulation design for heterodyne optical communication system," in *in Proc. IEEE Global Commun. Conf.*, 1983, pp. 412–415.

[6] G. J. Foschini, L. J. Greenstein, and G. Vannucci, "Noncoherent detection of coherent lightwave signals corrupted by phase noise," *IEEE Trans. Commun.*, vol. 36, pp. 306–314, Mar. 1988.

[7] J. R. Barry and E. A. Lee, "Performance of coherent optical receivers," in *Proceedings of the IEEE*, vol. 78, Aug. 1990, pp. 1369–1394.

[8] K. Yonenaga, S. Kuwano, S. Norimatsu, and N. Shibata, "Optical duobinary transmission system with no receiver sensitivity degradation," *Electronics Letters*, vol. 31, Feb. 1995.

[9] D. Penninckx, M. Chbat, L. Pierre, and J.-P. Thiery, "The phase-shaped binary transmission (PSBT): a new technique to transmit far beyond the chromatic dispersion limit," *IEEE Photon. Technol. Lett.*, vol. 9, pp. 259–261, Feb. 1997.

[10] A. H. Gnauck and P. J. Winzer, "Optical phase-shift-keyed transmission," *IEEE J. Lightwave Tech.*, vol. 23, no. 1, pp. 115–130, Jan. 2005.

[11] A. F. Elrefaie, R. E. Wagner, D. A. Atlas, and D. G. Daut, "Chromatic dispersion limitations in coherent lightwave transmission systems," *IEEE J. Lightwave Tech.*, vol. 6, no. 5, pp. 704–709, May 1988.

[12] G. J. Foschini and C. D. Poole, "Statistical theory of polarization dispersion in single mode fibers," *IEEE J. Lightwave Tech.*, vol. 9, no. 11, pp. 1439–1456, Nov. 1991.

[13] G. J. Foschini, R. M. Jopson, L. E. Nelson, and H. Kogelnik, "The statistics of PMD-induced chromatic fiber dispersion," *IEEE J. Lightwave Tech.*, vol. 17, pp. 1560–1565, Sept. 1999.

[14] ——, "Probability densities of second-order polarization mode dispersion including polarization dependent chromatic fiber dispersion," *IEEE J. Lightwave Tech.*, vol. 12, pp. 293–295, Mar. 2000.

[15] H. Kogelnik, L. E. Nelson, J. P. Gordon, and R. M. Jopson, "Jones matrix for second-order polarization mode dispersion," *Optics Lett.*, vol. 25, no. 1, pp. 19–21, 2000.

[16] F. Bruyère, "Impact of first- and second-order PMD in optical digital transmission systems," *Opt. Fiber Technol.*, vol. 2, pp. 269–280, 1996.

[17] A. Eyal, W. K. Marshall, M. Tur, and A. Yariv, "Representation of second-order polarization mode dispersion," *Electon. Lett.*, vol. 35, pp. 1658–1659, Sept. 1999.

[18] C. D. Poole and R. E. Wagner, "Phenomenological approach to polarization dispersion in long single-mode fibers," *Electron. Lett.*, vol. 22, pp. 1029–1030, 1986.

[19] E. Forestieri and L. Vincetti, "Exact evaluation of the Jones matrix of a fiber in the presence of polarization mode dispersion of any order," *IEEE J. Lightwave Tech.*, vol. 17, no. 12, pp. 1898–1909, Dec. 2001.

[20] C. Francia, F. Bruyère, D. Penninckx, and M. Chbat, "PMD second-order effects on pulse propagation in single-mode optical fibers," *IEEE Photon. Technol. Lett.*, vol. 10, pp. 1739–1741, Dec. 1998.

[21] E. Forestieri and G. Prati, "Exact analytical evaluation of second-order PMD impact on the outage probability for a compensated system," *IEEE J. Lightwave Tech.*, vol. 22, no. 4, pp. 988–996, Apr. 2004.

[22] R. Olshansky, "Noise figure for erbium-doped optical fibre amplifiers," *Electr. Lett.*, vol. 24, pp. 1363–1365, Oct. 1988.

[23] K. Kikuchi, "Generalised formula for optical-amplifier noise and its application to erbium-doped fibre amplifiers," *Electr. Lett.*, vol. 26, pp. 1851–1853, Oct. 1990.

[24] P. S. Henry, "Error-rate performance of optical amplifiers," in *Tech. Dig. OFC'89*, Feb. 1989, paper THK3.

[25] J. G. Proakis, *Digital Communications*, 4th ed. New York: McGraw-Hill, 2001.

[26] G. D. Forney, Jr., "Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference," *IEEE Trans. Inform. Theory*, vol. 18, pp. 284–287, May 1972.

[27] G. Ungerboeck, "Adaptive maximum-likelihood receiver for carrier-modulated data-transmission systems," *IEEE Trans. Commun.*, vol. 22, pp. 624–636, May 1974.

[28] M. Secondini, E. Forestieri, and G. Prati, "PLC optical equalizer for chromatic and polarization-mode dispersion compensation based on MSE control," *IEEE Photon. Technol. Lett.*, vol. 16, no. 4, pp. 1173–1175, Apr. 2004.

[29] F. Buchali and H. Bulow, "Adaptive PMD compensation by electrical and optical techniques," *IEEE J. Lightwave Tech.*, vol. 22, pp. 1116–1126, Apr. 2004.

[30] J. H. Winters and R. D. Gitlin, "Electrical signal processing techniques in long-haul fiber-optic systems," *IEEE Trans. Commun.*, vol. 38, pp. 1439–1453, Sept. 1990.

[31] J. H. Winters and S. Kasturia, "Adaptive nonlinear cancellation for high-speed fiber-optic systems," *IEEE J. Lightwave Tech.*, vol. 10, pp. 971–977, July 1992.

[32] H. Bulow and G. Thielecke, "Electronic PMD mitigation—from linear equalization to maximum-likelihood detection," in *Proc. OFC'01*, vol. 3, 2001, pp. WDD34–1–WDD34–3.

[33] H. F. Haunstein, W. Sauer-Greff, A. Dittrich, K. Sticht, and R. Urbansky, "Principles for electronic equalization of polarization-mode dispersion," *IEEE J. Lightwave Tech.*, vol. 22, pp. 1169–1182, Apr. 2004.

[34] G. D. Forney, Jr., "The Viterbi algorithm," *Proc. IEEE*, vol. 61, pp. 268–278, Mar. 1973.

[35] J. G. Proakis, *Digital Communications*, 3rd ed. New York: McGraw-Hill, 1996.

[36] D. Marcuse, "Derivation of analytical expression for the bit-error probability in lightwave systems with optical amplifiers," *IEEE J. Lightwave Tech.*, vol. 8, no. 12, pp. 1816–1823, Dec. 1990.

[37] P. A. Humblet and M. Azizoğlu, "On the bit error rate of lightwave systems with optical amplifiers," *IEEE J. Lightwave Tech.*, vol. 9, no. 11, pp. 1576–1582, Nov. 1991.

[38] Y. Cai, J. M. Morris, T. Adali, and C. R. Menyuk, "On turbo code decoder performance in optical-fiber communication systems with dominating ASE noise," *IEEE J. Lightwave Tech.*, vol. 21, no. 3, pp. 727–734, Mar. 2003.

[39] O. E. Agazzi, M. R. Hueda, H. S. Carrer, and D. E. Crivelli, "Maximum-likelihood sequence estimation in dispersive optical channels," *IEEE J. Lightwave Tech.*, vol. 23, no. 2, pp. 749–763, Feb. 2005.

[40] A. J. Weiss, "On the performance of electrical equalization in optical fiber transmission systems," *IEEE Photon. Technol. Lett.*, vol. 15, no. 9, Sept. 2003.

[41] H. Meyr, M. Oerder, and A. Polydoros, "On sampling rate, analog prefiltering, and sufficient statistics for digital receivers," *IEEE Trans. Commun.*, vol. 42, pp. 3208–3214, Dec. 1994.

[42] E. Forestieri, "Evaluating the error probability in lightwave systems with chromatic dispersion, arbitrary pulse shape and pre- and postdetection

filtering," *IEEE J. Lightwave Tech.*, vol. 18, no. 11, pp. 1493–1503, Nov. 2000.

[43] T. Foggi, G. Colavolpe, E. Forestieri, and G. Prati, "Adaptive electronic processing in optical PMD-impaired systems," in *Optical Networks and Technology*, K.-I. Kitayama, F. Masetti-Placci, and G. Prati, Eds. New York: Springer, 2004, pp. 499–506.

[44] T. Foggi, E. Forestieri, G. Colavolpe, and G. Prati, "Maximum likelihood sequence detection with closed-form metrics in OOK optical systems impaired by GVD e PMD," *IEEE J. Lightwave Tech.*, vol. 24, no. 8, pp. 3073–3087, Aug. 2006.

[45] ——, "Channel estimation algorithms for MLSD in optical cammunication systems," *IEEE Photon. Technol. Lett.*, vol. 18, no. 19, pp. 1984–1986, Nov. 2006.

[46] C. W. Helstrom, "Approximate evaluation of detection probabilities in radar and optical communications," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 14, pp. 630 – 640, July 1978.

[47] B. Chan and J. Conradi, "On the non-gaussian noise in erbium-doped fiber amplifiers," *J. Lightwave Technol.*, vol. 15, no. 4, pp. 680–687, Apr. 1997.

[48] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions.* New York: Dover, 1972.

[49] M. V. Eyuboğlu and S. U. Qureshi, "Reduced-state sequence estimation with set partitioning and decision feedback," *IEEE Trans. Commun.*, vol. 38, pp. 13–20, Jan. 1988.

[50] G. Ferrari, G. Colavolpe, and R. Raheli, "A unified framework for finite-memory detection," *IEEE J. Select. Areas Commun.*, vol. 23, pp. 1697–1706, Sept. 2005.

[51] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inform. Theory*, vol. 20, pp. 284–287, Mar. 1974.

[52] H. L. Van Trees, *Detection, Estimation and Modulation Theory - Part I.* New York, NJ: John Wiley & Sons, 1968.

[53] R. Holzlöhner and C. R. Menyuk, "Use of multicanonical monte carlo simulations to obtain accurate bit error rates in optical communications systems," *Opt. Lett.*, vol. 28, p. 1894, 2003.

[54] S. Walklin and J. Conradi, "Multilevel signaling for increasing the reach of 10 Gb/s lightwave systems," *J. Lightwave Technol.*, vol. 17, no. 11, pp. 2235–2248, Nov. 1999.

[55] R. A. Griffin and A. C. Carter, "Optical differential quadrature phase-shift key (oDQPSK) for high capacity optical transmission," in *Proc. Optical Fiber Commun. Conf. (OFC'02)*, Anaheim, CA, Feb. 2002, pp. 367–368.

[56] S. Hayase, N. Kikuchi, K. Sekine, and S. Sasaki, "Proposal of 8-state per symbol (binary ask and QPSK) 30-Gbit/s optical modulation / demodulation scheme," in *Proc. European Conf. on Optical Commun. (ECOC'03)*, Sept. 2003, pp. 1008–1009, paper Th.2.6.4.

[57] M. Ohm and J. Speidel, "Optimal receiver bandwidths, bit error probabilities and chromatic dispersion tolerance of 40Gbit/s optical 8-DPSK with NRZ and RZ impulse shaping," in *Proc. Optical Fiber Commun. Conf. (OFC'05)*, Anaheim, CA, Mar. 2005, paper OFG5.

[58] L. G. Kazowsky, S. Benedetto, and A. Willner, *Optical Fiber Communication Systems.* Norwood, MA: Artec House, 1996.

[59] S. Calabró, D. van den Borne, S. L. Jansen, G. D. Khoe, and H. de Waardt, "Improved detection of differential phase shift keying through multi-symbol phase estimation," in *Proc. European Conf. on Optical Commun. (ECOC'05)*, vol. 3, Sept. 2005, pp. 737–738, paper We4.P.118.

[60] M. Nazarathy and E. Simony, "Multichip differential phase encoded optical transmission," *IEEE Photon. Technol. Lett.*, vol. 17, no. 5, pp. 1133–1135, May 2005.

[61] X. Liu, "Receiver sensitivity improvement in optical DQPSK and DQPSK/ASK through data-aided multi-symbol phase estimation," in

Proc. European Conf. on Optical Commun. (ECOC'06), Sept. 2006, paper We2.5.6.

[62] D. Divsalar and M. Simon, "Multiple-symbol differential detection of MPSK," *IEEE Trans. Commun.*, vol. 38, pp. 300–1308, Mar. 1990.

[63] F. Edbauer, "Bit error rate of binary and quaternary DPSK signals with multiple differential feedback detection," *IEEE Trans. Commun.*, vol. 40, no. 3, pp. 457–460, Mar. 1992.

[64] H. Leib, "Data-aided noncoherent demodulation of DPSK," *IEEE Trans. Commun.*, vol. 43, no. 2,3,4, pp. 722–725, Feb./Mar./Apr. 1995.

[65] G. Colavolpe and R. Raheli, "Noncoherent sequence detection," *IEEE Trans. Commun.*, vol. 47, pp. 1376–1385, Sept. 1999.

[66] R. Schober and W. H. Gerstacker, "Metric for noncoherent sequence estimation," *IEE Electronics Letters*, vol. 35, no. 25, pp. 2178–2179, Nov. 1999.

[67] G. Colavolpe, T. Foggi, E. Forestieri, and G. Prati, "Multilevel optical systems with MLSD receivers insensitive to GVD and PMD," to appear on *IEEE J. of Lightwave Tech.*

[68] W. J. Weber, "Differential encoding for multiple amplitude and phase shift keying systems," *IEEE Trans. Commun.*, vol. 26, no. 3, pp. 385–391, Mar. 1978.

[69] E. Forestieri, G. Colavolpe, and G. Prati, "Novel MSE adaptive control of optical PMD compensators," *IEEE J. Lightwave Tech.*, vol. 20, no. 12, pp. 1997–2003, Dec. 2002.

[70] M. Peleg and S. Shamai (Shitz), "On the capacity of the blockwise incoherent MPSK channel," *IEEE Trans. Commun.*, vol. 46, pp. 603–609, May 1998.

[71] G. Colavolpe and R. Raheli, "The capacity of the noncoherent channel," *European Trans. Telecommun.*, vol. 12, no. 4, pp. 289–296, July/August 2001.

[72] R. Nuriyev and A. Anastasopoulos, "Capacity and coding for the block-independent noncoherent AWGN channel," *IEEE Trans. Inform. Theory*, vol. 51, no. 3, pp. 866–883, Mar. 2005.

[73] G. Colavolpe and R. Raheli, "Theoretical analysis and performance limits of noncoherent sequence detection of coded PSK," *IEEE Trans. Inform. Theory*, vol. 46, no. 4, pp. 1483–1494, July 2000.

[74] R. Raheli, A. Polydoros, and C. Tzou, "Per-survivor processing: A general approach to MLSE in uncertain environments," *IEEE Trans. Commun.*, vol. 43, pp. 354–364, February-April 1995.

[75] A. Duel-Hallen and C. Heegard, "Delayed decision feedback estimation," *IEEE Trans. Commun.*, vol. 37, pp. 428–436, May 1989.

[76] P. R. Chevillat and E. Eleftheriou, "Decoding of trellis-encoded signals in the presence of intersymbol interference and noise," *IEEE Trans. Commun.*, vol. 36, pp. 669–676, July 1989.

[77] R. Schober and W. H. Gerstacker, "Noncoherent adaptive channel identification algorithms for noncoherent sequence estimation," *IEEE Trans. Commun.*, vol. 49, no. 2, pp. 229–234, Feb. 2001.

[78] G. Colavolpe, G. Ferrari, and R. Raheli, "Noncoherent iterative (turbo) decoding," *IEEE Trans. Commun.*, vol. 48, no. 9, pp. 1488–1498, Sept. 2000.

[79] G. Colavolpe, "On LDPC codes over channels with memory," *IEEE Trans. Wireless Commun.*, vol. 5, no. 7, pp. 1757–1766, July 2006.

[80] N. Kaneda, X. Liu, Z. Zheng, X. Wei, M. Tayahi, M. Movassaghi, and D. Levy, "Improved polarization-mode-dispersion tolerance in duobinary transmission," *IEEE Photon. Technol. Lett.*, vol. 15, no. 7, pp. 1005–1007, July 2003.

[81] C. R. Giles and E. Desurvire, "Modeling erbium-doped fiber amplifiers," *IEEE J. Lightwave Tech.*, vol. 9, pp. 271–283, Feb. 1991.

[82] J. H. Winters, "Equalization in coherent lightwave systems using a fractionally spaced equalizer," *IEEE J. Lightwave Tech.*, vol. 8, pp. 1487–1491, Oct. 1990.

[83] L. G. Kazovsky, "Homodyne phase-shift-keying systems: past challenges and future opportunities," *IEEE J. Lightwave Tech.*, vol. 24, pp. 4876–4884, Dec. 2006.

[84] M. G. Taylor, "Coherent detection method using DSP for demodulation of signal and subsequent equalization of propagation impairments," *IEEE Photon. Technol. Lett.*, vol. 16, pp. 674–676, Feb. 2004.

[85] R. Noè, "PLL-free synchronous QPSK polarization multiplex/diversity receiver concept with digital I-Q baseband processing," *IEEE Photon. Technol. Lett.*, vol. 17, pp. 887–889, Apr. 2005.

[86] E. Ip and J. M. Kahn, "Feedforward carrier recovery for coherent optical communications," *IEEE J. Lightwave Tech.*, vol. 25, pp. 2675–2692, Sept. 2007.

[87] S. Tsukamoto, K. Katoh, and K. Kikuchi, "Coherent demodulation of optical multilevel phase-shift-keying signals using homodyne detection and digital signal processing," *IEEE Photon. Technol. Lett.*, vol. 18, pp. 1131–1133, May 2006.

[88] D.-S. Ly-Gagnon, S. Tsukamoto, K. Katoh, and K. Kikuchi, "Coherent detection of optical quadrature-phase-shift-keying signal with carrier phase estimation," *IEEE J. Lightwave Tech.*, vol. 24, pp. 12–21, Jan. 2006.

[89] G. Charlet, J. Renaudier, M. Salsi, H. Mardoyan, P. Tran, and S. Bigo, "Efficient mitigation of fiber impairments in an ultra-long haul trasmission of 40 gb/s polarization-multiplexed data, by digital signal processing in a coherent receiver," in *Proc. Optical Fiber Commun. Conf. (OFC'07)*, Anaheim, CA, Mar. 2007, paper PDP17.

[90] E. Ip and J. M. Kahn, "Digital equalization of chromatic dispersion and polarization mode dispersion," *IEEE J. Lightwave Tech.*, vol. 25, pp. 2033–2043, Aug. 2007.

[91] F. Derr, "Coherent optical QPSK intradyne system: concept and digital receiver realization," *IEEE J. Lightwave Tech.*, vol. 10, no. 9, pp. 1290–1296, Sept. 1992.

[92] A. Papoulis, *Probability, Random Variables and Sthocastic Processes.* New York, NY: McGraw-Hill, 1991.

[93] R. Schober and W. H. Gerstacker, "Adaptive noncoherent DFE for MDPSK signals transmitted over ISI channels," *IEEE Trans. Commun.*, vol. 48, no. 7, pp. 1128–1140, July 2000.

[94] U. Mengali and A. N. D'Andrea, *Synchronization Techniques for Digital Receivers (Applications of Communications Theory).* New York, NJ: Plenum Press, 1997.