Maximum-Likelihood Sequence Detection With Closed-Form Metrics in OOK Optical Systems Impaired by GVD and PMD

Tommaso Foggi, Enrico Forestieri, *Member, IEEE*, Giulio Colavolpe, and Giancarlo Prati, *Fellow, IEEE*

Abstract—This paper thoroughly investigates the maximum-likelihood sequence detection (MLSD) receiver for the optical ON–OFF keying (OOK) channel in the presence of both polarization-mode dispersion and group velocity dispersion (GVD). A reliable method is provided for computing the relevant performance for any possible value of the system parameters, with no constraint on the sampling rate. With one sample per bit time, a practically exact expression of the statistics of the received samples is found, and therefore the performance of a synchronous MLSD receiver is evaluated and compared with that of other electronic techniques such as combined feedforward and decision-feedback equalizers (FFE and DFE). It is also shown that the ultimate performance of electronic processing can be obtained by sampling the received signal at twice the bit rate. An approximate accurate closed-form expression of the receiver metrics is also identified, allowing for the implementation of a practically optimal MLSD receiver.

Index Terms—Chromatic dispersion, electrical equalization, intersymbol interference (ISI), maximum-likelihood sequence detection (MLSD), optical transmission systems, polarization mode dispersion (PMD).

I. INTRODUCTION

IN FIBER-OPTICS communication systems, sophisticated receivers based on electronic processing have recently been receiving much attention in conjunction with the design of high-speed links (40 Gb/s and beyond), mainly because electronic techniques, although less effective, would allow for the implementation of cost-effective integrated receivers as compared with optical processing techniques.

As known, in the linear regime, group velocity dispersion (GVD) and polarization mode dispersion (PMD) are the most severe sources of signal distortion and system penalty. Although GVD can be compensated for by dispersion-compensating fibers in present communication systems, compensation at bit rates higher than 10 Gb/s may not prove to be adequate because GVD tolerance decreases with the square of the bit rate, leaving a nonnegligible residual dispersion. 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 PMD, which is an intrinsically stochastic phenomenon whose penalties are difficult to fully compensate for.

In a first-order approximation, the effect of PMD is just a differential group delay (DGD) $\Delta\tau$ between the two principal states of polarization (PSPs) of the fiber, causing intersymbol interference (ISI). Customarily, PMD is described by a vector $\vec{\Omega}$ in Stokes space, which in a first-order approximation is taken as frequency independent. Higher order effects arise when the PMD vector $\vec{\Omega}$ varies with frequency. In a common second-order approximation, $\vec{\Omega}$ is assumed to be a linear function of frequency, i.e., $\vec{\Omega} = \vec{\Omega}_0 + \vec{\Omega}_1(\omega - \omega_0)$, where $\vec{\Omega}_1$ is the derivative of $\vec{\Omega}$ evaluated at carrier frequency ω_0 [1]. 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 [2], planar lightwave circuits (PLCs) [3], or other optical devices [4].

The above techniques, although very effective, may still be impractical because of their cost, due to optical technologies. Thus, many efforts have been done to apply classical and novel electrical (postdetection) processing techniques to optical communication systems. The first proposal of an electrical equalizer for optical systems, to the authors' knowledge, was a linear transversal filter to combat ISI arising from chromatic dispersion [5], but nonlinear cancellation was also proposed since photodetection implies a nonlinear transformation of the signal [6]. Much more recently, comparisons between these methods and optical compensation have been presented, evidencing benefits and drawbacks of both solutions [4], [7], [8].

Besides feedforward and decision-feedback equalizations (FFE and DFE), there is an increasing interest in maximum-likelihood sequence detection[1] (MLSD), implemented through the Viterbi algorithm (VA) [9], [10], due to its potentially optimal performance [9]. In the pioneering work [5], this strategy was already considered, but of course, the presence of optical amplifiers (OAs) could not be accounted for. As a consequence, the amplified spontaneous emission (ASE) noise is not present in [5], and the statistics of the received signal, necessary to compute the VA branch metrics, are conditionally

[1]We prefer the term "detection" instead of the commonly used "estimation" because estimation theory refers to continuous parameters whereas we are interested in discrete sequences.
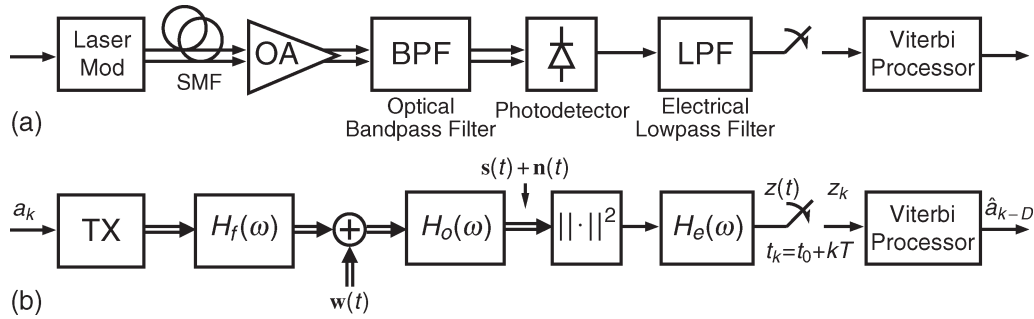
Fig. 1. (a) Schematic of the system model. (b) Lowpass equivalent.

Gaussian due to shot and thermal noise. Nowadays, all optical systems envisage the presence of OAs, and, as a consequence, the signal in the fiber is impaired by a noise that, in the linear regime, can be modeled as additive white Gaussian noise (AWGN) [11]. Since a square-law detector is present at the receiver end, postdetection noise statistics change [10], [12], [13] and can no longer be considered Gaussian. In the case of the MLSD strategy, assuming Gaussian statistics for noise after photodetection is neither realistic nor correct and leads to inaccurate results [14]. Hence, in [4] and [8], 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 the 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 [4], [8]. A rigorous detailed description of the MLSD approach can be found in a recent paper [15], where 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.

In this paper, we consider the optimal MLSD strategy with PMD and GVD in the sense that our aim is to describe the receiver structure and to provide a reliable method for computing the relevant performance for any possible value of the system parameters, under the only constraint of one sample per bit, as in [15] and [16], or no constraint on the sampling rate, since oversampling is a possible way to guarantee sufficient statistics for this transmission system [17] (see also [18] for applications to optical channels). In particular, through numerical evaluation [19], we derived a practically exact expression of the received signal statistics in the case of a receiver working with one sample per bit time so that a look-up-table-based MLSD receiver could be implemented [20]. In the case of oversampling, since our numerical method cannot be used as we do not know an (neither exact nor approximate) expression for the joint statistics of the samples, we resorted to a histogram-based receiver. An exhaustive analysis of the correlation of the received samples was carried out to state the best tradeoff between the number of samples to be processed and the amount of memory necessary to store a reliable estimate of the probability density functions (pdfs).

Finally, 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 system performance are provided, allowing us to reach very low values of bit error rate (BER). 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 also has 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, 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 postdetection processing is not able to effectively cope with the combined GVD and PMD distortions.

This paper is organized as follows. Section II presents the system model and a brief background on PMD. Section III explains the numerical approach adopted to obtain the pdf of the received signal, which is necessary to compute the VA branch metrics, and the derivation of an accurate approximated closed-form expression. Section IV is about the derivation of a sufficient statistic based on oversampling, whereas Section V explains the computation of lower and upper bounds. Finally, in Section VI, numerical results are presented in the presence of both PMD and GVD, and in Section VII conclusions are drawn.

II. SYSTEM MODEL

Fig. 1 shows the system model and its low-pass equivalent. A standard nonreturn-to-zero (NRZ) ON–OFF-keying (OOK)-modulated laser beam is launched in a single-mode fiber (SMF), optically amplified and filtered at the receiver end. The OA is assumed to have a high gain G so that the ASE noise is dominant over thermal and shot noise. The signal is then photodetected, electrically filtered, sampled, and finally processed through the VA, effectively implementing the maximum *a posteriori* (MAP) sequence detection strategy; here, it is equivalent to MLSD, since the transmitted bits $\{a_k\}$ are equally likely. In Fig. 1, $\mathbf{w}(t) = [w_1(t), w_2(t)]^T$ represents AWGN, where $w_1(t)$ and $w_2(t)$ are independent complex noise components accounting for ASE on two orthogonal states of polarization (SOPs), each with two-sided power spectral density (PSD) equal to N_0. When $G \gg 1$, $N_0 = n_{\mathrm{sp}} h\nu (G-1)/G \simeq n_{\mathrm{sp}} h\nu$, where $n_{\mathrm{sp}} \geq 1$ is the spontaneous emission

parameter, and $h\nu$ is the photon energy. At the optical filter output, the components of the two-dimensional (2-D) 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 the fiber, optical, and postdetection filters are denoted by $H_f(\omega)$, $H_o(\omega)$, and $H_e(\omega)$, respectively.

The well-known transfer function of a fiber affected by GVD only is expressed as $H_{\mathrm{GVD}}(\omega) = \exp(-j\beta_2 L\omega^2/2)$, where $\beta_2 = -\lambda_0 D/\omega_0$ is the *chromatic dispersion parameter* [21], D being the fiber chromatic dispersion (usually expressed in picoseconds per nanometer per kilometer) and λ_0 the wavelength, while L is the fiber length. 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. To make our results independent of the bit rate R_b, we will use the dimensionless *chromatic dispersion index* γ [22], which is defined as

$$\gamma = 2\lambda_0 R_b^2 D_r/\omega_0. \tag{1}$$

Using the chromatic dispersion index γ, the transfer function of a GVD-compensated link can be written as $H_{\mathrm{GVD}}(\omega) = \exp(j\gamma(\omega/R_b)^2/4)$.

Accounting now for PMD, the fiber Jones matrix is written as $H_f(\omega) = H_{\mathrm{GVD}}(\omega)RU(\omega)R^{-1}$, where $U(\omega)$ describes the fiber PMD on the basis of its PSPs at the carrier frequency that, without loss of generality, can be taken to be coincident at the input and output. R is a random rotation matrix, independent of frequency, representing a change of basis SOPs. Thus, its columns represent the PSPs at the carrier frequency on the basis of the reference SOPs, which is often taken as two linear orthogonal polarizations aligned along the S_1 axis in Stokes space. Assuming, without loss of generality, that the carrier SOP remains unchanged, $U(\omega)$ is such that $U(0) = I$.

A closed-form expression of $U(\omega)$ accounting for all PMD orders is not known, and although several second-order approximations are available (see [23] and references therein), they account differently for higher PMD orders [24]. 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 [25] for $U(\omega)$ as it is also able to partially (and correctly) account for higher PMD orders through second-order parameters only [24].

At the output of the photodiode, the detected signal can be described as the sum of two contributions (one for each SOP), i.e.,

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

Clearly, after photodetection, noise becomes signal-dependent and its statistics change. In the following, optical

and postdetection filter parameters can be chosen arbitrarily since the proposed receiver is independent of a particular choice for filter shapes or bandwidths.

III. RECEIVED SIGNAL STATISTICS

In this section, we derive the MLSD detection strategy under the constraints of 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 aforementioned 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 aforementioned constraints, the MLSD detection strategy can be expressed as [10]

$$\hat{\mathbf{a}} = \arg\max_{\mathbf{a}} p(\mathbf{z}|\mathbf{a}) \tag{3}$$

where $\mathbf{a} = \{a_k\}$ is the transmitted bit sequence, and $\mathbf{z} = \{z_k\}$ is the corresponding received sequence. The synchronous received samples are $z_k = z(t_0 + kT)$, t_0 being a proper time offset and $T = 1/R_b$ 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/or GVD. Then, the conditional joint pdf of the received samples can be written as

$$p(\mathbf{z}|\mathbf{a}) = \prod_k p(z_k|\mathbf{a}) \tag{4}$$

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

$$p(z_k|\mathbf{a}) = p(z_k|a_k, a_{k-1}, \dots, a_{k-L}). \tag{5}$$

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

$$\lambda_k(a_k, \mu_k) = \ln p(z_k|a_k, \mu_k) \tag{6}$$

where $\mu_k = (a_{k-1}, a_{k-2}, \dots, a_{k-L})$ 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 (5) is not known for arbitrary signal formats and filtering. Although the samples at the photodetector output have a noncentral chi-square distribution, being the sum of squared Gaussian random variables [10], 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 by the received signal samples and by the considered trellis transition to compute the branch metrics.

---

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

Undoubtedly, the most efficient numerical method is that using the "saddlepoint" approximation [26], 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 noise on a proper Karhunen–Loève basis, as shown in [24]. Then, using the saddlepoint approximation, the pdf can be evaluated as

$$p(z_k|\mathbf{a}) \simeq \frac{\exp\left[\Phi_{z_k|\mathbf{a}}(s_0)\right]}{\sqrt{2\pi \Phi''_{z_k|\mathbf{a}}(s_0)}} \qquad (7)$$

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

$$\Phi_{z_k|\mathbf{a}}(s) = \log\left[\Psi_{z_k|\mathbf{a}}(s)e^{-sz_k}\right] \qquad (8)$$

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 postdetection filter is an integrate and dump device, then the pdf of the received sample is still of chi-square type [13]. 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 [14], [15], [27]

$$p(z_k|a_k,\mu_k) \simeq \frac{1}{N_0}\left(\frac{z_k}{s_k}\right)^{(\nu-1)/2}$$
$$\times \exp\left(-\frac{z_k+s_k}{N_0}\right) I_{\nu-1}\left(\frac{2\sqrt{z_k s_k}}{N_0}\right) \qquad (9)$$

where ν, which is 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., $\nu = 2B_o/B_e$; $s_k \triangleq s_R(a_k,\mu_k)$ is the noise-free received sample that 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 the first kind and order $\nu-1$. Hence, a simplified expression of the branch metrics, discarding irrelevant terms in the maximization, is[4]

$$\lambda(a_k,\mu_k) \simeq -\frac{\nu-1}{2}\ln\left[s_R(a_k,\mu_k)\right] - \frac{s_R(a_k,\mu_k)}{N_0}$$
$$+ \ln\left[I_{\nu-1}\left(\frac{2\sqrt{z_k s_R(a_k,\mu_k)}}{N_0}\right)\right]. \qquad (10)$$

---

[3]The factor 2 is because we consider signal and noise as complex rather than real valued as done in [13]–[15].

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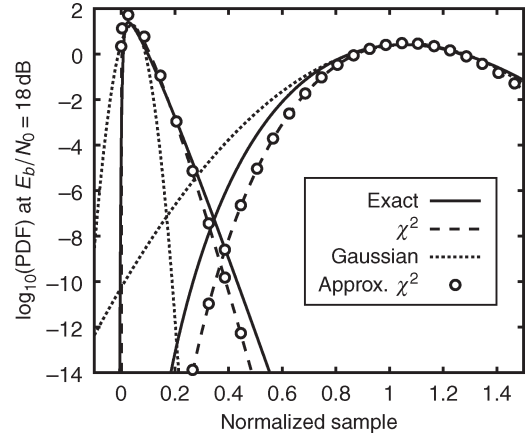

Fig. 2. PDFs of "1" and "0" for $E_b/N_0 = 18$ dB.

As we will see in Section VI, 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 [14] to compute the performance of optical turbo-coded systems, but differently from [14], we use (9) both when $a_k = 1$ and $a_k = 0$ 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 $a_k = 0$, (9) can be used only if $s_R(a_k,\mu_k) > 0$, i.e., when the electrical filter is such that either its impulse response is always positive or its negative values have a negligible impact, as is the case for Gaussian- or Bessel-type filters, for example.

The exact (solid), chi-square (dashed), and Gaussian (dotted) pdf approximations are reported in Fig. 2 for $E_b/N_0 = 18$ 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 represents the number of detected photons per bit at the input of the OA and is related to the optical signal-to-noise ratio (OSNR) by $E_b/N_0 = 2 \cdot \text{OSNR} \cdot WT$, with W being the reference measurement bandwidth usually taken equal to 12.5 GHz (0.1 nm). As 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 [13]. A further simplification can be obtained by the crude approximation $I_{\nu-1}(x) \simeq e^x/\sqrt{2\pi x}$, by which (9) is approximated as

$$p(z_k|a_k,\mu_k) \simeq \left(\frac{z_k}{s_k}\right)^{\frac{2\nu-1}{4}} \frac{\exp\left(-\frac{(\sqrt{z_k}-\sqrt{s_k})^2}{N_0}\right)}{\sqrt{4\pi N_0 z_k}} \qquad (11)$$

reported in Fig. 2 as open circles. The resulting simplified branch metrics are

$$\lambda(a_k,\mu_k) \simeq 2\sqrt{z_k s_R(a_k,\mu_k)} - s_R(a_k,\mu_k)$$
$$-\frac{N_0}{2}\left(\nu - \frac{1}{2}\right)\ln\left[s_R(a_k,\mu_k)\right]. \qquad (12)$$

The impact of these approximations in the expression of the branch metrics will be considered in Section VI.
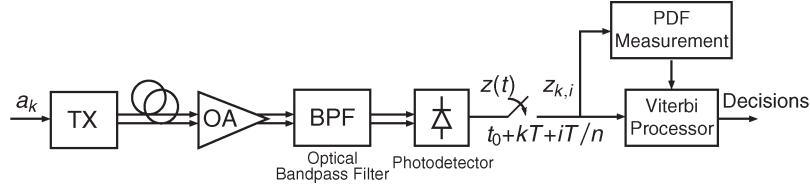
Fig. 3. System model for the receiver based on oversampling.

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 [28] 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 $2^{L'} < 2^L$. To compute the branch metrics (6) in a reduced trellis, the necessary symbols not included in the state definition may be found in the survivor history [28]. 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 nonlinear equalizer with decision feedback.

Since PMD is a time-varying phenomenon, the receiver parameters should be adaptively updated. By using one of the aforementioned approximated closed-form expressions of the branch metrics (10) or (12), 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.

IV. OVERSAMPLING

Although the MLSD receiver described in the previous section represents the best postdetection 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 [17]. In fact, provided that an adequate number of samples per bit is extracted from the signal after photodetection, the whole information in the electrical signal is preserved and there is no need of a further processing through an electrical filter [17]. 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 kth 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 kth 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 the received samples that differ for at least 1 bit interval are independent, we have [29]

$$p(\mathbf{z}|\mathbf{a}) = \prod_k p(\mathbf{z}_k|\mathbf{z}_{k-1}, \mathbf{a})$$

$$= \prod_k \prod_{i=0}^{n-1} p(z_{k,i}|z_{k,i-1}, \ldots, z_{k,0}, \mathbf{z}_{k-1}, \mathbf{a}) \quad (13)$$

and, in the last expression, it is implicitly assumed that, if $i = 0$, the 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 VA implementing the MLSD strategy can be expressed as [29]

$$\lambda_k(a_k, \mu_k) = \ln p(\mathbf{z}_k|\mathbf{z}_{k-1}, a_k, \mu_k)$$

$$= \sum_{i=0}^{n-1} \ln p(z_{k,i}|z_{k,i-1}, \ldots, z_{k,0}, \mathbf{z}_{k-1}, a_k, \mu_k). \quad (14)$$

We did not succeed in finding a way to analytically evaluate the pdfs in (14), when the samples are not conditionally independent, and we had to resort to simulation, implementing a histogram counter at the receiver to obtain a reliable estimation of the channel statistics necessary to compute the branch metrics. The receiver structure in the case of oversampling ($n > 1$) is shown in Fig. 3.

The joint pdfs derived from the aforementioned histograms allowed us to also evaluate the impact of the correlation between samples on the branch metrics and the optimum sampling rate. In Section VI, we will see that a value of $n = 2$ is practically sufficient to attain optimal performance. In this case, although the received samples are correlated, by considering them as independent in expressing the branch metrics, we obtain the same performance that can be obtained with the optimal correlated metrics. Hence, a simplified branch metrics can be adopted with no performance degradation

$$\lambda(a_k, \mu_k) \simeq \sum_{i=0}^{1} \ln p(z_{k,i}|a_k, \mu_k). \quad (15)$$

The pdfs appearing in (15) can be further simplified by using the approximate expression (9) or (11).

---

[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 $\mu_k = (a_{k-1}, a_{k-2}, \ldots, a_{k-M})$.
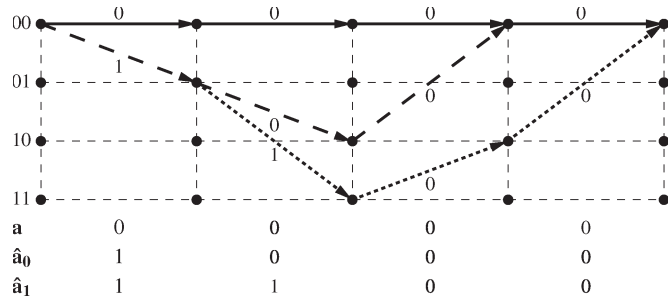
Fig. 4.   Trellis transitions given by wrong sequences $\hat{\mathbf{a}}_0$ and $\hat{\mathbf{a}}_1$.

It is worth mentioning that, although the described branch metrics have been derived having in mind the MAP sequence detection strategy, as demonstrated in [29], the same branch metrics can be used for the algorithm by Bahl, Cocke, Jelinek, and Raviv (BCJR) [30], implementing the MAP symbol detection strategy and employed as a component decoder in iterative decoding schemes for optical turbo codes, such as those proposed in [14].

V. LOWER AND UPPER BOUNDS

It is well known that the classical union upper bound on the bit error probability P_b has the expression [10]

$$P_b \leq \sum_{\mathbf{a}} P(\mathbf{a}) \sum_{\hat{\mathbf{a}} \neq \mathbf{a}} b(\mathbf{a}, \hat{\mathbf{a}}) P(\mathbf{a} \to \hat{\mathbf{a}}) \qquad (16)$$

where $\mathbf{a} = \{a_k\}$ and $\hat{\mathbf{a}} = \{\hat{a}_k\}$ denote bit sequences corresponding to correct and erroneous paths, respectively, $b(\mathbf{a}, \hat{\mathbf{a}})$ is the number of bit errors entailed by the considered error event $(\mathbf{a}, \hat{\mathbf{a}})$, $P(\mathbf{a} \to \hat{\mathbf{a}})$ is the pairwise error probability (PEP), and $P(\mathbf{a})$ is the *a priori* probability of sequence $\mathbf{a}$. The PEP $P(\mathbf{a} \to \hat{\mathbf{a}})$ 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, i.e.,

$$P(\mathbf{a} \to \hat{\mathbf{a}}) = P\left(\sum_{k=\ell}^{\ell+H} \lambda_k < \sum_{k=\ell}^{\ell+H} \hat{\lambda}_k\right) \qquad (17)$$

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

From the union bound, we derive a lower and an approximated upper bounds for the bit error probability. The lower bound is simply obtained by considering the most likely error event and assuming that only 1 bit of 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 us 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 saddlepoint 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, although 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 can be seen from (10) and (12), the branch metrics depend only on the square root of the samples. Using (12) and letting

$$y_k = \sqrt{z_k} \qquad (18)$$
$$d_k = \sqrt{s_k} - \sqrt{\hat{s}_k} \qquad (19)$$

the PEP in (17), 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

$$P(\mathbf{a} \to \hat{\mathbf{a}}) = P\left(\sum_k d_k y_k < \frac{1}{2} \sum_k (s_k - \hat{s}_k)\right). \qquad (20)$$

As given in Appendix A, the random variable (r.v.) y_k in (18) can be approximated as Gaussian with mean and variance, respectively, as

$$\eta_{y_k} = \alpha_k \sqrt{s_k} \qquad (21)$$
$$\sigma_{y_k}^2 = N_k/2 \qquad (22)$$

where

$$\alpha_k = \frac{s_k/N_k + \left(4\nu_k^2 - 1\right)/16}{s_k/N_k + (2\nu_k - 1)(2\nu_k - 3)/16} \qquad (23)$$

$$N_k = \frac{\sigma_{z_k}^2}{\eta_{z_k} + s_k} \qquad (24)$$

$$\nu_k = \frac{\eta_{z_k}^2 - s_k^2}{\sigma_{z_k}^2} \qquad (25)$$

η_{z_k} and $\sigma_{z_k}^2$ being the mean and the variance of z_k. Notice that for the oversampling case, as the postdetection filter is absent, the photodetected samples are chi-square distributed such that (24) and (25) 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.

$$x = \sum_k d_k y_k \qquad (26)$$

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

$$\eta_x = \sum_k d_k \eta_{y_k} \tag{27}$$

$$\sigma_x^2 = \sum_k d_k^2 \sigma_{y_k}^2 + 2 \sum_{\substack{k,\ell \\ k \neq \ell}} \rho_{k\ell} d_k d_\ell \sigma_{y_k} \sigma_{y_\ell} \tag{28}$$

where $\rho_{k\ell}$, which is 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

$$P(\mathbf{a} \to \hat{\mathbf{a}}) \simeq \frac{1}{2} \operatorname{erfc} \left(\frac{\eta_x - \frac{1}{2} \sum_k (s_k - \hat{s}_k)}{\sqrt{2} \sigma_x} \right) \tag{29}$$

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

$$P(\mathbf{a} \to \hat{\mathbf{a}}) \simeq \frac{1}{2} \operatorname{erfc} \left(\frac{\sum_k \left[(2\alpha_k - 1)\sqrt{s_k} - \sqrt{\hat{s}_k} \right] d_k}{2\sqrt{\sum_k N_k d_k^2}} \right). \tag{30}$$

Notice that a similar result was also found in [15], where, however, a different approximation to the χ^2 distribution was used in deriving the branch metrics. We would also like to point out that the results in [15] are only valid for low optical extinction ratios (10 dB or lower) and when the optical filter has a large bandwidth and the postdetection 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 next section.

VI. NUMERICAL RESULTS

Standard Monte Carlo simulations were carried out to evaluate 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,[6] so that all 5-bit interfering patterns were considered, the number of trellis states thus being equal to 16. We verified that these choices are redundant and that a four-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 a large number of states is only useful for large amounts of PMD and/or GVD,
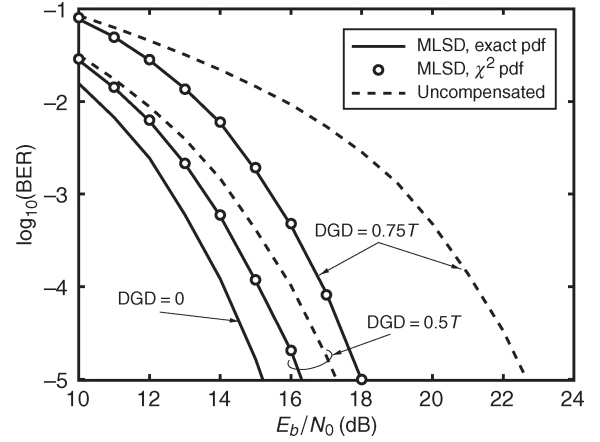

Fig. 5. Performance of the proposed MLSD receiver with first-order PMD and equal power splitting between PSPs.
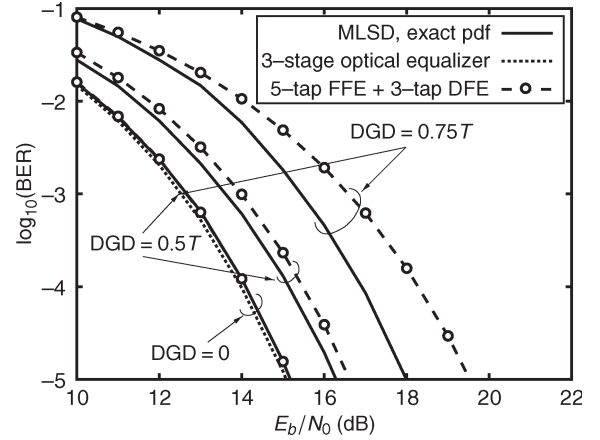

Fig. 6. Comparison between optical and electrical compensation schemes with first-order PMD and 50% power splitting.

when, however, the induced penalty is already several decibels and the achievable benefits are negligible).

A. 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. 5, 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 DGD increases. Moreover, we found no appreciable difference in the results obtained using the branch metrics in (10) or (12).

It is interesting to compare the performance of the MLSD receiver with that of other electrical and optical equalization techniques. In Fig. 6, for the same DGD values, we also report the performance of combined (synchronous) FFEs and DFEs,[7] and that of a three-stage optical compensator, composed by

---

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

[7]The number of taps of both equalizers was optimized, finding no improvement with more than five taps for the FFE and three taps for the DFE.
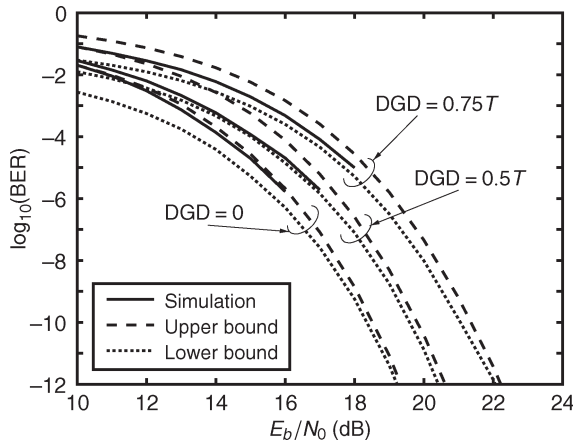
Fig. 7.   Lower and upper bounds of the receiver performance with first-order PMD and 50% power splitting.
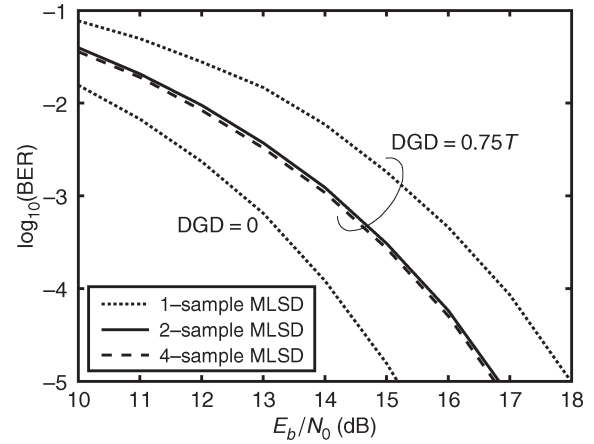

Fig. 8.   Simulation results for the histogram-based oversampling receiver for $n = 1, 2, 4$. First-order PMD and 50% power splitting.

a cascade of three polarization controllers and polarization maintaining fibers as in [2]. This optical compensator shows a negligible penalty for increasing DGD values (the curves are practically undistinguishable), whereas both electrical processing schemes suffer a larger penalty as DGD increases. We point out that this is because, 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 also have 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 $0.5\,T$, as expected.

The accuracy of the described lower and approximated upper bounds is shown in Fig. 7. In this case also, only the first-order PMD is considered, with DGD values of $0\,T$, $0.5\,T$, and $0.75\,T$. It is clear that it is not necessary to have recourse to time-consuming computer simulations since this tool predicts very well the receiver performance, where the higher the accuracy, the lower the BER, and can also be used to optimize the system parameters.

B. 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 reoptimized, finding that, in this case, its optimum value is $1.2/T$. Accounting for sample correlation, we considered up to four 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 can be seen from the simulation curves in Fig. 8, no significant improvement is attained for $n > 2$, so an oversampling factor of 2 is assumed in the following.

Since the optimal tradeoff between computability and efficiency is achieved by using just two samples per bit time, it
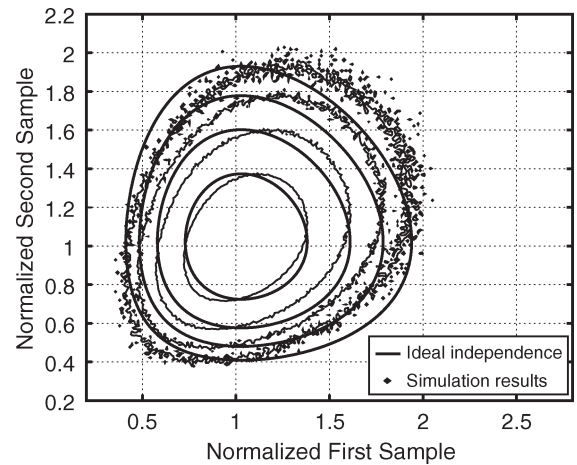

Fig. 9.   Joint pdf of two $T/2$-spaced samples at $E_b/N_0 = 12$ dB. Starting from the inner, contours are at 10^0, 10^{-1}, 10^{-2}, 10^{-3}. Ragged lines represent simulation results.

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 method for pdf computation, because the joint pdf could be approximated by the product of the marginal pdfs.[8] The joint pdf accounting for sample correlation and obtained by simulation was compared with the joint pdf evaluated under the hypothesis of sample independence and analytically derived as the product of the marginal pdfs. We found that the correlation between samples is almost negligible when in both $T/2$-spaced samples, the signal part is low (i.e., it corresponds to a logical "0"), whereas it is significant when the signal is high (i.e., it corresponds to a logical "1"). Fig. 9 compares the pdfs obtained as explained before when the signal is high in both samples. Indeed, the plot in Fig. 9 reveals a little amount of correlation. So Monte Carlo simulations were carried out with both joint and independent histograms to evaluate the impact of such a correlation on the performance of the receiver. Obviously, proper training sequences of different length were used to build

---

[8]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.
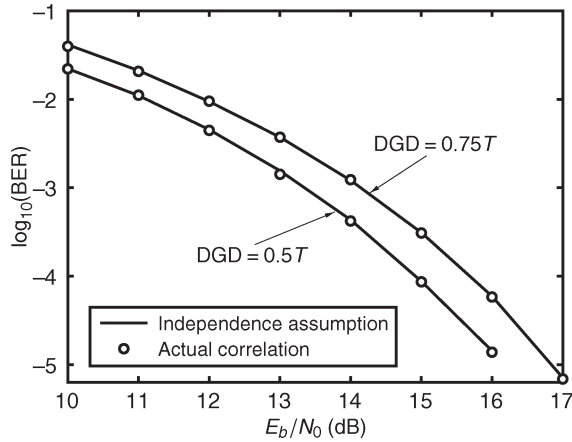
Fig. 10. Simulation results for independence assumption or actual correlation on noise statistics when taking two samples per bit (histogram-based receiver).
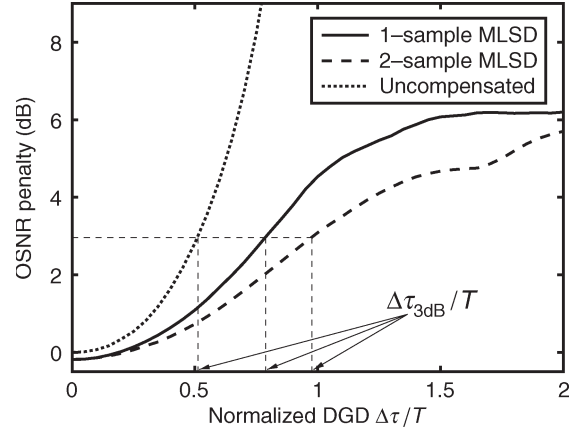

Fig. 11. Comparison between synchronous and oversampled receiver. First-order PMD and 50% power splitting.

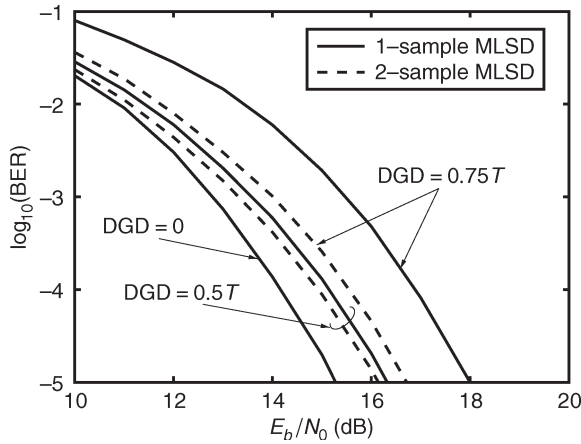Fig. 12. OSNR penalty at BER $= 10^{-12}$ versus normalized DGD. First-order PMD and 50% power splitting.

Fig. 13. Outage probability due to first-order PMD.

up the histograms so that the same degree of accuracy was assured. We report in Fig. 10 the results obtained by simulation under the two assumptions about sample correlation. As can be seen, assuming independence leads to a really negligible penalty (almost unnoticeable in the graphs), 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 reoptimization, its performance was evaluated using the reference DGD values 0, 0.5, and 0.75 T. The final results are shown in Fig. 11, evidencing that oversampling leads to a significant improvement for increasing values of DGD.

C. Outage Probability

The method that we developed to obtain lower and upper bounds allows for 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 [24]. In computing the
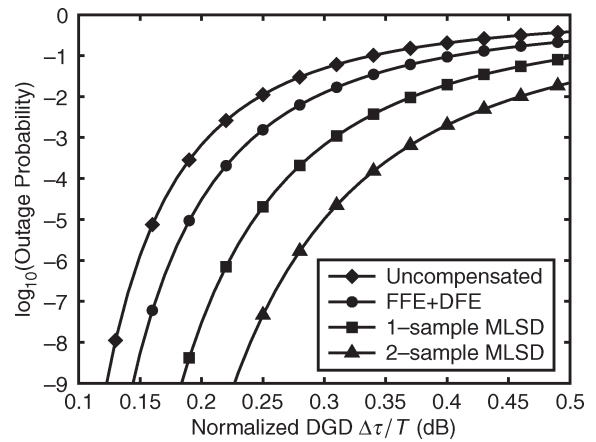
outage probability, we consider a reference BER of 10^{-12} and a 3-dB margin such that Monte Carlo simulations are infeasible. So, we evaluated the penalty by using the approximated upper bound previously discussed.

We initially considered first-order PMD only. In this case, once verified that the penalty contour at 3 dB versus DGD and power splitting is U shaped, the outage probability can be approximated as [24]

$$P_{\text{out}} \simeq \exp\left[-\frac{4}{\pi}\left(\frac{\Delta\tau_{\text{3dB}}}{\langle\Delta\tau\rangle}\right)^2\right] \qquad (31)$$

where $\langle\Delta\tau\rangle$ is the mean DGD, and $\Delta\tau_{\text{3dB}}$ is the instantaneous DGD giving a 3-dB penalty at 50% power splitting, as shown in Fig. 12, where the penalty versus DGD is reported. Note that, if the postdetection filter has a highly nonsymmetric impulse response, the isopenalty curves may not be U shaped and, in this case, (31) may not be accurate. We verified that this is not the case with the electrical fifth-order Bessel filter we used. The outage probability due to the first-order PMD is shown in Fig. 13. As can be seen, a synchronous MLSD receiver does fairly better than an FFE combined with a DFE, allowing for a mean DGD value of about 30% higher for $P_{\text{out}} < 10^{-6}$, whereas the oversampling MLSD receiver improves this value to about 60%.
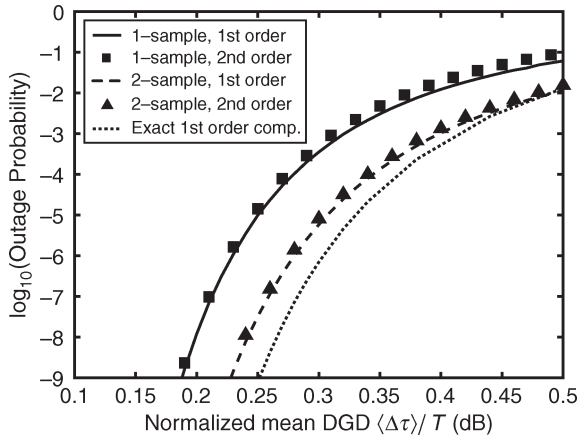
Fig. 14.   Outage probability due to first- and second-order PMD.

Fig. 15.   OSNR penalty at BER $= 10^{-12}$ as a function of chromatic dispersion index. Both upper and lower bounds are reported for MLSD.

There is no closed-form approximation for the outage probability due to second-order PMD, and we analytically evaluated it as described in [24] using the model in [25] for the fiber Jones matrix. The results are shown in Fig. 14, 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 [24], which is also shown in Fig. 14. 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, and we are confident that our results also remain valid when taking into account higher PMD orders.

D. 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 is 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 in (1). As already mentioned, using the chromatic dispersion index γ instead of the residual dispersion D_r simply makes the results 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. 15, an oversampling MLSD receiver outperforms standard electrical equalization techniques, as well as synchronous one, except for values of the chromatic dispersion index lower than about 0.3. It is interesting to note that the synchronous receiver penalty shows a well-defined local minimum at around $\gamma = 0.4$, which can be explained by analyzing the Euclidean distances between correct and corresponding wrong patterns on the trellis diagram. Indeed, as shown in Fig. 16, the distance of the most likely error event, which determines the receiver performance, is smaller when $\gamma = 0.3$ 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 $t = T/2$
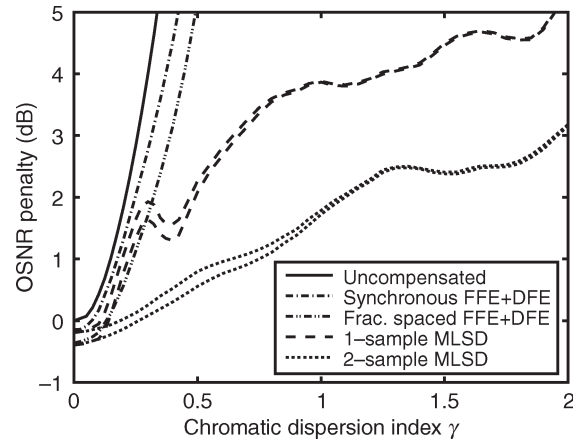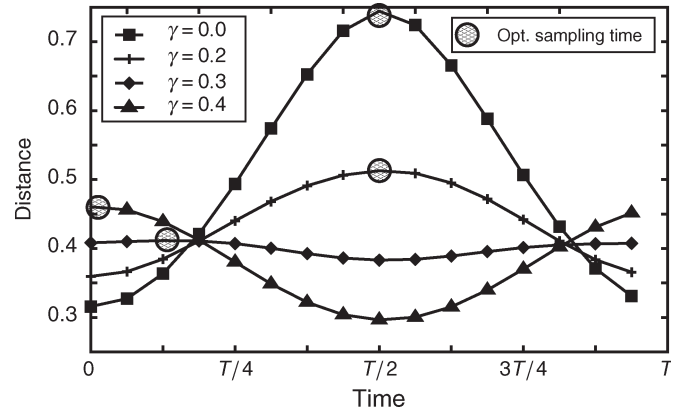

Fig. 16.   Euclidean distance between correct and corresponding wrong patterns on the trellis diagram versus sampling time.

to $t = 0$. Now, with reference to Fig. 16, 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 the MLSD that performs sequence rather than bit-by-bit detection such that a sequence of samples need not be taken at the maximum eye-opening times, and even when the eye is closed, it can still be correctly detected.

We next consider the combined effect of GVD and PMD. In Fig. 17, 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 of around $\gamma = 0.4$, which can be explained by the above considerations, while the maximum mean DGD that is tolerable by the oversampling receiver decreases monotonically.

VII. CONCLUSION

MLSD for OOK optical channels with PMD and GVD has been discussed, considering both the implementation aspects and the relevant performance. In particular, the implementation of the MLSD strategy making use of one sample per bit of the signal after photodetection has been considered first. In this case, starting from the exact statistics of the received
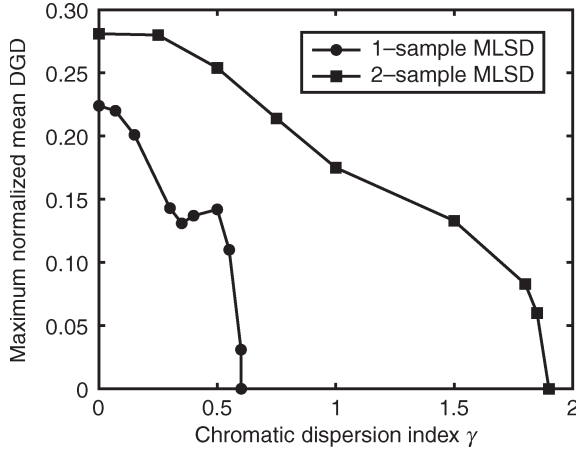
Fig. 17. Maximum mean DGD giving an outage probability of less than 10^{-6} versus chromatic dispersion index. First- and second-order PMD is considered.

signal that can be analytically computed, a simpler closed-form approximated expression of the branch metrics of the VA implementing the MLSD strategy has been obtained. Then, to improve the performance, a quest for sufficient statistics has been carried out, leading to the conclusion that the signal at the receiver should be sampled at twice the bit rate. Through comparison with the results obtainable by using the exact statistics, it has been shown that, sampling at twice the bit rate, the samples can still be regarded as independent for computing the metrics. Hence, a closed-form expression of the VA branch metrics is still available in this case. By means of simulations as well as accurate analytical performance evaluation, it has also been shown that this technique makes a significant penalty reduction possible. A performance comparison with electronic equalization techniques, such as FFE and DFE, and with optical equalization, has also been carried out, showing that the MLSD receiver based on two samples per bit time outperforms other electronic equalizers but is still far away from a three-tap optical equalizer based, for instance, on PLCs.

APPENDIX A

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

Although (11) proved to be adequate for deriving branch metrics for the VA, it is not adequate for performance evaluation. Indeed, as can be seen from Fig. 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 in (11) to fit the actual pdf of the sample z_k with very high accuracy as

$$p_{z_k}(z) \simeq \frac{c_k}{\sqrt{4\pi N_k z}} \left(\frac{z}{s_k}\right)^{\frac{2\nu_k - 1}{4}} \exp\left(-\frac{(\sqrt{z} - \sqrt{s_k})^2}{N_k}\right) \quad (32)$$

where N_k and ν_k are as in (24) and (25), and c_k is the normalization constant

$$c_k = \frac{2\sqrt{\pi}(s_k/N_k)^{(2\nu_k - 1)/4}e^{s_k/N_k}}{\Upsilon(1,1) + 2\sqrt{\frac{s_k}{N_k}}\Upsilon(3,3)} \quad (33)$$

having defined

$$\Upsilon(n,m) \triangleq \Gamma\left(\frac{2\nu_k + n}{4}\right) {}_1F_1\left(\frac{2\nu_k + n}{4}; \frac{m}{2}; \frac{s_k}{N_k}\right) \quad (34)$$

and where ${}_1F_1$ is the confluent hypergeometric function [31]. Using the approximation (less accurate for $x \approx 1$, but still valid)

$${}_1F_1(a;b;x) \simeq \frac{\Gamma(b)}{\Gamma(a)}x^{a-b}e^x\left(1 + \frac{(a-b)(a-1)}{x}\right), \quad x > 1 \quad (35)$$

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

$$c_k \simeq \frac{s_k/N_k}{s_k/N_k + (2\nu_k - 1)(2\nu_k - 3)/16} \quad (36)$$

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

From (32), it turns out that the pdf of $y_k = \sqrt{z_k}$ can be approximated as

$$p_{y_k}(y) = 2\sqrt{z}p_z(z)$$
$$\simeq \frac{c_k}{\sqrt{\pi N_k}}\left(\frac{y}{\sqrt{s_k}}\right)^{\nu_k - \frac{1}{2}}e^{-(y-\sqrt{s_k})^2/N_k}, \quad y \geq 0 \quad (37)$$

whose tails (only the right-hand one for small s_k/N_k values) are Gaussian. The mean value of y_k turns out to be

$$\eta_{y_k} \simeq \sqrt{N_k}\frac{\Upsilon(3,1) + 2\sqrt{\frac{s_k}{N_k}}\Upsilon(5,3)}{\Upsilon(1,1) + 2\sqrt{\frac{s_k}{N_k}}\Upsilon(3,3)}$$
$$\simeq \sqrt{s_k}\frac{s_k/N_k + \left(4\nu_k^2 - 1\right)/16}{s_k/N_k + (2\nu_k - 1)(2\nu_k - 3)/16} \quad (38)$$

such that (37) can be approximated as

$$p_{y_k}(y) \simeq \frac{1}{\sqrt{\pi N_k}}\exp\left(-\frac{(y - \eta_{y_k})^2}{N_k}\right). \quad (39)$$

We point out again that, for $s_k/N_k \ll 1$, (39) is a good approximation only for the right-hand tail of (37), but this is suitable in 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 photodetected signal.

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

$$s_i(t) = s_{i1}(t) + js_{i2}(t)$$
$$n_i(t) = n_{i1}(t) + jn_{i2}(t) \quad, \quad i = 1, 2 \quad (40)$$

and letting

$$x_{ij}(t) \triangleq s_{ij}(t) + n_{ij}(t), \qquad i,j = 1,2 \tag{41}$$

the photodetector output in (2) 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

$$p_{x_{ij}}(x,y) = \frac{\exp\left\{ -\frac{(x-\eta_{ij1})^2 - 2\rho(x-\eta_{ij1})(y-\eta_{ij2}) + (y-\eta_{ij2})^2}{2(1-\rho^2)\sigma^2} \right\}}{2\pi\sigma^2\sqrt{1-\rho^2}} \tag{42}$$

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

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

$$\rho_z = \frac{E\left\{\sqrt{z(t_1)z(t_2)}\right\} - E\left\{\sqrt{z(t_1)}\right\}E\left\{\sqrt{z(t_2)}\right\}}{\sigma_{\sqrt{z_1}}\sigma_{\sqrt{z_2}}} \tag{43}$$

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

$$\frac{1}{2}\sum_{k=1}^4 |a_k| \le \sqrt{\sum_{k=1}^4 a_k^2} \le \sum_{k=1}^4 |a_k| \tag{44}$$

and letting

$$y(t) \triangleq |x_{12}(t)| + |x_{12}(t)| + |x_{21}(t)| + |x_{22}(t)| \tag{45}$$

we have that

$$\sqrt{z(t_i)} = \kappa_i y(t_i) \tag{46}$$

where $1/2 \le \kappa_i \le 1$. 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 (43), we will evaluate

$$\rho_y = \frac{E\{y(t_1)y(t_2)\} - E\{y(t_1)\}E\{y(t_2)\}}{\sigma_{y_1}\sigma_{y_2}} \tag{47}$$

assuming that $\rho_z \simeq \rho_y$.

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

$$E\{y(t_1)y(t_2)\} - E\{y(t_1)\}E\{y(t_2)\}$$

$$= \sum_{ij} \left[E\{|x_{ij}(t_1)x_{ij}(t_2)|\} - E\{|x_{ij}(t_1)|\}E\{|x_{ij}(t_2)|\} \right] \tag{48}$$

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

$$E\{|x_{ij}(t_k)|\} = \eta_{ijk}\,\mathrm{erf}\left(\frac{\eta_{ijk}}{\sqrt{2}\sigma}\right) + \sqrt{\frac{2}{\pi}}\sigma\,\exp\left(-\frac{\eta_{ijk}^2}{2\sigma^2}\right) \tag{49}$$

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

$$\eta_1 \triangleq \begin{cases} \eta_{ij1}, & \text{if } |\eta_{ij1}| < |\eta_{ij2}| \\ \eta_{ij2}, & \text{otherwise} \end{cases} \tag{50}$$

and

$$\eta_2 \triangleq \begin{cases} \eta_{ij2}, & \text{if } |\eta_{ij1}| < |\eta_{ij2}| \\ \eta_{ij1}, & \text{otherwise} \end{cases} \tag{51}$$

it can be shown that

$$E\{|x_{ij}(t_1)x_{ij}(t_2)|\} = \iint\limits_{-\infty}^{\infty} |xy| p_{x_{ij}}(x,y)\,dx\,dy$$

$$= \int_{-\infty}^{\infty} |x| f(x)\,dx \tag{52}$$

where

$$f(x) = \frac{\sqrt{1-\rho^2}}{\pi} \exp\left(-\frac{(x-\eta_1)^2}{2\sigma^2}\right)$$
$$\times \left\{ \exp\left[-g^2(x)\right] + \sqrt{\pi}g(x)\mathrm{erf}\left[g(x)\right] \right\} \tag{53}$$

and

$$g(x) = \frac{\rho(x-\eta_1) + \eta_2}{\sigma\sqrt{2(1-\rho^2)}}. \tag{54}$$

Notice that, due to (42), the result of the integral in (52) remains unchanged if we swap η_{ij1} and η_{ij2}, and we exploited this fact in (50) and (51) because the function $f(x)$ can be approximated as $f(x) \simeq |f_a(x)|$, where

$$f_a(x) = \frac{\rho(x-\eta_1) + \eta_2}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\eta_1)^2}{2\sigma^2}\right) \tag{55}$$

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

$$\xi \triangleq \eta_1 - \eta_2/\rho \tag{56}$$

we have

$$\int_{-\infty}^{\infty} |x| f(x)\,dx \simeq \int_{-\infty}^{\infty} |x f_a(x)|\,dx$$

$$= \int_{-\infty}^{\infty} x f_a(x)\,dx - 2\mathrm{sgn}(\xi)\int_0^{\xi} x f_a(x)\,dx. \tag{57}$$

The first integral in the last equation in (57) gives

$$\int_{-\infty}^{\infty} x f_a(x) dx = \rho \sigma^2 + \eta_1 \eta_2 \tag{58}$$

while the second one gives

$$\int_{0}^{\xi} x f_a(x) dx$$

$$= \frac{1}{2} \left(\rho \sigma^2 + \eta_1 \eta_2 \right) \left[\operatorname{erf}\left(\frac{\eta_1}{\sqrt{2}\sigma} \right) - \operatorname{erf}\left(\frac{\eta_2}{\sqrt{2}\rho\sigma} \right) \right]$$

$$+ \frac{\sigma}{\sqrt{2\pi}} \left[\eta_2 \exp\left(-\frac{\eta_1^2}{2\sigma^2} \right) - \rho \eta_1 \exp\left(-\frac{\eta_2^2}{2\rho^2\sigma^2} \right) \right]. \tag{59}$$

When both η_{ij1} and η_{ij2} are of the same order or smaller than σ, (57) loses accuracy, and we must add a correction term to it. Using the approximation $f(x) - |f_a(x)| \simeq f_e(x)$, where

$$f_e(x) = \frac{\sqrt{1-\rho^2}}{\pi} \exp\left(-\frac{(x-\eta_1)^2}{2\sigma^2} - \sqrt{\pi}|g(x)| - \frac{\pi-2}{2} g^2(x) \right) \tag{60}$$

the correction term to be added to (57) is approximated with great accuracy by

$$\int_{-\infty}^{\infty} |x| f_e(x)\, dx$$

$$= \varepsilon(\xi) \Big\{ 2e^{-v_1^2(\xi)} - e^{-v_2^2(\xi)} + \sqrt{\pi} v_1(\xi)$$

$$\times \left[2 \operatorname{erf} v_1(\xi) - \operatorname{erf} v_2(\xi) - \operatorname{sgn} \xi \right] \Big\}$$

$$+ \varepsilon(-\xi) \Big\{ e^{v_2^2(-\xi)} + \sqrt{\pi} v_1(-\xi) \left[\operatorname{erf} v_2(-\xi) + \operatorname{sgn} \xi \right] \Big\} \tag{61}$$

with ξ being as in (56), and

$$\varepsilon(\xi) = \frac{r^2 \sigma^2 \sqrt{1-\rho^2}}{\pi}$$

$$\times \exp\left(-\frac{\eta_2 \mu(\xi)}{\sigma^2} + \frac{(\pi-2)\eta_2^2 + \pi\rho^2\sigma^2}{2\sigma^2 \left[2 - (4-\pi)\rho^2 \right]} \right) \tag{62}$$

$$v_1(\xi) = \frac{\eta_1 - \rho\mu(\xi)}{\sqrt{2}r\sigma} \tag{63}$$

$$v_2(\xi) = \frac{\eta_2 - \rho^2 \mu(\xi)}{\sqrt{2}\rho\, r\, \sigma} \tag{64}$$

$$r = \sqrt{\frac{2(1-\rho^2)}{2 - (4-\pi)\rho^2}} \tag{65}$$

$$\mu(\xi) = \frac{(\pi-2)\eta_2 - \operatorname{sgn}(\xi)\sigma\sqrt{2\pi(1-\rho^2)}}{2 - (4-\pi)\rho^2}. \tag{66}$$
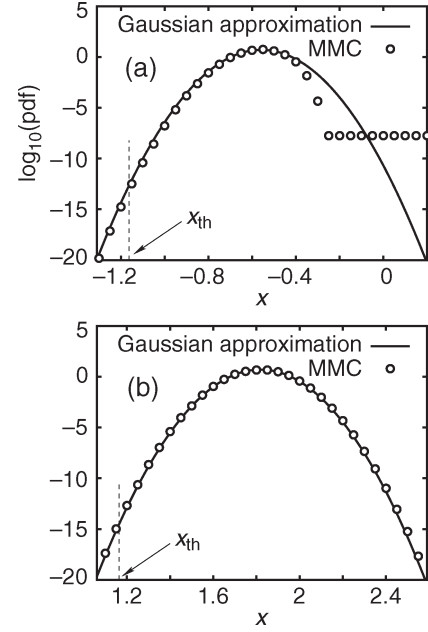

Fig. 18. PDF of the r.v. x in (26) when (a) $\mathbf{a} = \{1, 1, 1, 0, 1, 1, 1, 0\}$ and $\hat{\mathbf{a}} = \{1, 1, 1, 1, 1, 1, 1, 0\}$, and (b) sequences $\mathbf{a}$ and $\hat{\mathbf{a}}$ are exchanged. Samples are spaced by $T/2$ and $E_b/N_0 = 20$ dB.

Notice that, when η_1 and η_2 are chosen as in (50) and (51), the correction term (61) can always be safely added to (57); 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 (49), we have

$$\sigma_{y_k}^2 = 4\sigma^2 + \sum_{ij} \eta_{ijk}^2$$

$$- \sum_{ij} \left[\eta_{ijk} \operatorname{erf}\left(\frac{\eta_{ijk}}{\sqrt{2}\sigma} \right) + \sqrt{\frac{2}{\pi}}\sigma \exp\left(-\frac{\eta_{ijk}^2}{2\sigma^2} \right) \right]^2. \tag{67}$$

To check our hypotheses, we evaluated by multicanonical Monte Carlo (MMC) technique [32] the pdf of the r.v. x in (26) and compared it with a Gaussian pdf whose mean and variance are as in (27) and (28), respectively. In Fig. 18(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. 18(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.

REFERENCES

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

[2] E. Forestieri, G. Colavolpe, and G. Prati, "Novel MSE adaptive control of optical PMD compensator," *J. Lightw. Technol.*, vol. 20, no. 12, pp. 1997–2003, Dec. 2002.

[3] 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.

[4] F. Buchali and H. Bulow, "Adaptive PMD compensation by electrical and optical techniques," *J. Lightw. Technol.*, vol. 22, no. 4, pp. 1116–1126, Apr. 2004.

[5] J. H. Winters and R. D. Gitlin, "Electrical signal processing techniques in long-haul fiber-optic systems," *IEEE Trans. Commun.*, vol. 38, no. 9, pp. 1439–1453, Sep. 1990.

[6] J. H. Winters and S. Kasturia, "Adaptive nonlinear cancellation for high-speed fiber-optic systems," *J. Lightw. Technol.*, vol. 10, no. 7, pp. 971–977, Jul. 1992.

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

[8] H. F. Haunstein, W. Sauer-Greff, A. Dittrich, K. Sticht, and R. Urbansky, "Principles for electronic equalization of polarization-mode dispersion," *J. Lightw. Technol.*, vol. 22, no. 4, pp. 1169–1182, Apr. 2004.

[9] G. D. Forney, Jr., "The Viterbi algorithm," *Proc. IEEE*, vol. 61, no. 3, pp. 268–278, Mar. 1973.

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

[11] P. S. Henry, "Error-rate performance of optical amplifiers," presented at the Optical Fiber Communication (OFC), Houston, TX, Feb. 1989, Paper THK3.

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

[13] P. A. Humblet and M. Azizoğlu, "On the bit error rate of lightwavex systems with optical amplifiers," *J. Lightw. Technol.*, vol. 9, no. 11, pp. 1576–1582, Nov. 1991.

[14] 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," *J. Lightw. Technol.*, vol. 21, no. 3, pp. 727–734, Mar. 2003.

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

[16] A. J. Weiss, "On the performance of electrical equalization in optical fiber transmission systems," *IEEE Photon. Technol. Lett.*, vol. 15, no. 9, pp. 1225–1227, Sep. 2003.

[17] H. Meyr, M. Oerder, and A. Polydoros, "On sampling rate, analog prefiltering, and sufficient statistics for digital receivers," *IEEE Trans. Commun.*, vol. 42, no. 12, pp. 3208–3214, Dec. 1994.

[18] A. Farbert *et al.*, "Performance of a 10.7 Gb/s receiver with digital equaliser using maximum likelihood sequence estimation," presented at the Eur. Conf. Optical Communication (ECOC), Stockholm, Sweden, 2004, Paper Th4.1.5.

[19] E. Forestieri, "Evaluating the error probability in lightwave systems with chromatic dispersion, arbitrary pulse shape and pre- and postdetection filtering," *J. Lightw. Technol.*, vol. 18, no. 11, pp. 1493–1503, Nov. 2000.

[20] 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-Verlag, 2004, pp. 499–506.

[21] G. P. Agrawal, *Nonlinear Fiber Optics*, 3rd ed. San Diego, CA: Academic, 2001.

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

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

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

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

[26] C. W. Helstrom, "Approximate evaluation of detection probabilities in radar and optical communications," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-14, no. 7, pp. 630–640, Jul. 1978.

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

[28] M. V. Eyuboğlu and S. U. Qureshi, "Reduced-state sequence estimation with set partitioning and decision feedback," *IEEE Trans. Commun.*, vol. 36, no. 1, pp. 13–20, Jan. 1988.

[29] G. Ferrari, G. Colavolpe, and R. Raheli, *Detection Algorithms for Wireless Communications*. Hoboken, NJ: Wiley, 2004.

[30] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inf. Theory*, vol. IT-20, no. 2, pp. 284–287, Mar. 1974.

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

[32] 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, no. 20, pp. 1894–1896, Oct. 2003.

Tommaso Foggi was born in Parma, Italy, on May 7, 1978. He received the Dr. Ing. degree in telecommunications engineering from the University of Parma in April 2003. He is currently working toward the Ph.D. degree at the Dipartimento di Ingegneria dell'Informazione (DII), University of Parma.

From July 2003 to July 2004, he was with the Photonic Networks National Laboratory, Pisa, Italy, under a Italian National Inter-University Consortium for Telecommunication (CNIT) scholarship. From July 2005 to 2006, he was with DII, University of Parma, under another CNIT scholarship. His main research interests include electrical compensation of impaired optical transmission systems and satellite navigation and positioning.

Enrico Forestieri (S'91–M'91) was born in Milazzo, Italy, in 1960. He received the Dr. Ing. degree in electronics engineering from the University of Pisa, Pisa, Italy, in 1988.

He was a Postdoctoral Scholar at the University of Parma, Parma, Italy, working on optical communication systems from 1989 to 1991 and was a Research Scientist and Faculty Member from 1991 to 2000. Since 2001, he has been with Scuola Superiore Sant'Anna di Studi Universitari e di Perfezionamento, Pisa, where he is currently a Professor of telecommunications. He is the leader of the "Optical Transmission Theory" area at the Integrated Research Center, Photonic Networks National Laboratory, Pisa. His research interests are in the general area of digital communication theory and optical communication systems, with special attention to adaptive optical and electronic equalization, channel coding, and advanced modulation formats for optical systems. His research activity has led to numerous scientific publications in leading international journals and conference proceedings, as well as a few patents.

Prof. Forestieri was the General Chairman of the Tyrrhenian International Workshop on Digital Communications in 2004.

Giulio Colavolpe was born in Cosenza, Italy, in 1969. He received the Dr. Ing. degree (*cum laude*) in telecommunications engineering from the University of Pisa, Pisa, Italy, in 1994 and the Ph.D. degree in information technologies from the University of Parma, Parma, Italy, in 1998.

Since 1997, he has been with the University of Parma, where he is currently an Associate Professor of telecommunications. In 2000, he was a Visiting Scientist at the Institut Eurécom, Valbonne, France. He is also the co-author of the book *Detection Algorithms for Wireless Communications, with Applications to Wired and Storage Systems* (New York: Wiley, 2004). His main research interests include digital transmission theory, adaptive signal processing, channel coding, and information theory. His research activity has led to several scientific publications in leading international journals and conference proceedings and a few industrial patents.

Prof. Colavolpe is the Principal Investigator of several research projects funded by the European Space Agency (ESA-ESTEC) and important telecommunications companies. In 2005, he received the Best Paper Award for the paper *VLSI Design of a Fully Parallel Decoder for LDPC-Like Codes* at the 13th International Conference on Software, Telecommunications, and Computer Networks.

Giancarlo Prati (M'80–SM'02–F'04) was born in Rome, Italy, on November 13, 1946. He received the Dr. Ing. degree (*cum laude*) in electronics engineering from the University of Pisa, Pisa, Italy, in 1972.

From 1975 to 1978, he was an Associate Professor of electrical engineering at the University of Pisa. From 1978 to 1979, he was on a NATO-supported Fellowship Leave in the Department of Electrical Engineering, University of Southern California, Los Angeles, working in optical communications. In 1982, he was a Visiting Associate Professor in the Department of Electrical and Computer Engineering, University of Massachusetts, Amherst. From 1976 to 1986, he was a Research Scientist of the Italian National Research Council (CNR) at the Centro di Studio per Metodi e Dispositivi di Radiotrasmissione, Pisa. From 1986 to 1988, he was a Professor of electrical engineering at the University of Genoa, Genoa, Italy. From 1988 to 2000, he was a Professor of telecommunications engineering at the University of Parma, Parma, Italy, where he served as Dean of the Engineering Faculty from 1992 to 1998. He is currently a Professor of telecommunications at the Scuola Superiore Sant'Anna di Studi Universitari e Perfezionamento, Pisa. He is also the President of Italian National Inter-University Consortium for Telecommunication (CNIT), incorporating 35 universities. His professional and academic interests are in telecommunication systems and digital signal processing in communications. The activity has focused on optical communications and radio frequency communications, with application to satellite communications, high-capacity terrestrial digital radio links, mobile radio, modems for switched telephone lines, and fiber communications.

Prof. Prati has been a member of the Technical Program Committee of the European Conference on Optical Communications since 1997.