Optimal Polarization Demultiplexing for Coherent Optical Communications Systems

Ioannis Roudas, Athanasios Vgenis, Constantinos S. Petrou, Dimitris Toumpakaris, Jason Hurley, Michael Sauer, John Downie, Yihong Mauro, and Srikanth Raghavan

***Abstract*—Spectrally-efficient optical communications systems employ polarization division multiplexing (PDM) as a practical solution, in order to double the capacity of a fiber link. Polarization demultiplexing can be performed electronically, using polarization-diversity coherent optical receivers. The primary goal of this paper is the optimal design, using the maximum-likelihood criterion, of polarization-diversity coherent optical receivers for polarization-multiplexed optical signals, in the absence of polarization mode dispersion (PMD). It is shown that simultaneous joint estimation of the symbols, over the two received states of polarization, yields optimal performance, in the absence of phase noise and intermediate frequency offset. In contrast, the commonly used zero-forcing polarization demultiplexer, followed by individual demodulation of the polarization-multiplexed tributaries, exhibits inferior performance, and becomes optimal only if the channel transfer matrix is unitary, e.g., in the absence of polarization dependent loss (PDL), and if the noise components at the polarization diversity branches have equal variances. In this special case, the zero-forcing polarization demultiplexer can be implemented by a 2×2 lattice adaptive filter, which is controlled by only two independent real parameters. These parameters can be computed recursively using the constant modulus algorithm (CMA). We evaluate, by simulation, the performance of the aforementioned zero-forcing polarization demultiplexer in coherent optical communication systems using PDM quadrature phase shift keying (QPSK) signals. We show that it is, by far, superior, in terms of convergence accuracy and speed, compared to conventional CMA-based polarization demultiplexers. Finally, we experimentally test the robustness of the proposed constrained CMA polarization demultiplexer to realistic imperfections of polarization-diversity coherent optical receivers. The PMD and PDL tolerance of the proposed demultiplexer can be used as a benchmark in order to compare the performance of more sophisticated adaptive electronic PMD/PDL equalizers.**

***Index Terms*—Coherent communications, polarization demultiplexing, constant modulus algorithm.**

Manuscript received January 30, 2009; revised May 27, 2009, July 12, 2009. First published November 03, 2009; current version published March 12, 2010. This work was supported in part by the European Social Fund and in part by the Greek Ministry of Development-GSRT.

I. Roudas, A. Vgenis, C. S. Petrou, and D. Toumpakaris are with the Department of Electrical and Computer Engineering, University of Patras, Rio 26500, Greece (e-mail: roudas@ece.upatras.gr; vgenis@ece.upatras.gr; petrou@ece.upatras.gr; dtouba@ece.upatras.gr).

J. Hurley, M. Sauer, J. Downie, Y. Mauro, and S. Raghavan are with Corning Inc., Corning, NY 14831 USA (e-mail: RaghavanS@corning.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JLT.2009.2035526

I. Introduction

RECENT progress in fast data acquisition, in combination with the decreasing cost of high-speed digital electronics, is currently rendering the digital implementation of coherent optical receiver functionalities commercially viable at symbol rates equal to 10 GBd and beyond [1]. A growing number of research papers focuses on the evaluation of the performance of digital signal processing (DSP) algorithms, which can successfully counteract various transmission impairments that typically affect the performance of coherent optical receivers (see tutorials [2], [3], and references therein).

Among all transmission impairments, it would be difficult to overstate the importance of the impact of polarization effects on the performance of coherent optical receivers. For instance, random polarization rotations, caused by the birefringence of optical fibers, can be detrimental, since the states of polarization (SOPs) of the received optical signal and the local oscillator are not identical, as required. Polarization diversity [2], [3] is a practical means to detect all signal power, independent of the received SOP, by using a coherent optical receiver with two identical branches, one for the x- and one for the y-polarization component, respectively. The photocurrents at the output of the receiver branches must be appropriately combined using a two-input/one-output adaptive filter (electronic *polarization combiner*) [3], in order to retrieve all the information carried by the received signal.

The increased complexity and cost of polarization-diversity coherent optical receivers can be better justified when simultaneous polarization division multiplexing (PDM) is used at the transmitter, in order to double the spectral efficiency of the optical communications system. In this case, the electronic polarization combiner, used in the single-channel case, is replaced by a two-input/two-output adaptive filter, which separates the PDM channels into their respective outputs (electronic *polarization demultiplexer*) [4]–[12]. In addition, by increasing the number of adaptive filter coefficients, the polarization demultiplexer can also perform equalization of the intersymbol interference caused by polarization mode dispersion (PMD), polarization dependent loss (PDL), residual chromatic dispersion and other effects [13]–[17].

There is no unanimous agreement in the optical communications community regarding the merit of different DSP algorithms for electronic polarization demultiplexing and equalization. For example, depending on their operating mode, proposed algorithms can be distinguished into two categories: i) Data-aided (requiring a training sequence to achieve convergence, e.g., [8]); and ii) Blind, either decision-directed (employing estimates of the received symbols for adaptation, e.g., [13]) or based on other attributes of the received optical signal, which are affected by intersymbol interference, without attempting to recover the data. In the latter category, the constant modulus algorithm (CMA) [18]–[21] has been proposed for blind adaptive feed-forward polarization demultiplexers [10], [11] and equalizers [15]. The popularity of these CMA-based modules [22]–[24] is due to their low computational complexity and their robustness in the presence of intermediate frequency (IF) offsets and laser phase noise. The second feature allows for decoupling between polarization demultiplexing and carrier frequency/phase recovery, so the latter two impairments can be addressed by separate DSP modules. A disadvantage of CMA-based modules is their possible erroneous convergence to the same PDM channel [11], [12].

This article focuses on the issue of optimal polarization demultiplexing exclusively, in the absence of PMD. The purpose of the present study is the optimal design, using the maximum-likelihood criterion [25]–[27], of polarization-diversity coherent optical receivers, for the detection of polarization-multiplexed optical signals with orthogonal, albeit unknown, SOPs. In the absence of laser phase noise and IF offset, it is shown that simultaneous joint estimation of the symbols, over the two received orthogonal SOPs, yields optimal performance. In contrast, the commonly used zero-forcing polarization demultiplexer usually yields sub-optimal performance [26], [27]. The latter first employs an electronic polarization demultiplexer, in order to fully invert the Jones matrix of the optical fiber, followed by individual demodulation of the polarization-multiplexed tributaries. Jones matrix inversion can be achieved using a lattice adaptive filter with four complex taps [4]–[12]. Similar to [4], [11] we introduce constraints between the taps, in order to avoid convergence into the same PDM channel. The constraints take advantage of the fact that polarization rotations in optical fibers can be represented by unitary transformations. Then, the transfer matrix of the adaptive filter, in the absence of PMD and PDL, is expressed as a function of only two independent real parameters. These parameters can be estimated using either data-aided or blind channel estimation techniques. For their estimation, we use the CMA. We evaluate, by simulation, the performance of the proposed constrained CMA polarization demultiplexer in coherent optical communication systems using PDM quadrature phase shift keying (QPSK) signals. We show that it is, by far, superior, in terms of convergence accuracy and speed, compared to previously proposed, conventional CMA-based polarization demultiplexers [10]. A salient feature of the proposed constrained CMA polarization demultiplexer is that convergence is always guaranteed. Finally, we experimentally test the tolerance of the proposed constrained CMA polarization demultiplexer to realistic imperfections of polarization-diversity coherent optical receivers.

Despite its apparent simplicity, the proposed constrained CMA electronic demultiplexer, can be used as a benchmark in order to compare the performance of more sophisticated adaptive electronic equalizers in the presence of PMD and PDL. Since it does not possess any PMD/PDL compensation capabilities, it can be used as a reference for the PMD and PDL tolerance of uncompensated coherent optical systems.

It is worth mentioning at this point that our polarization demultiplexer is almost identical to the one proposed by Kikuchi in a recent paper [11]. The latter came to our attention only after the submission of our manuscript to the Journal of Lightwave Technology, by one of the reviewers. Despite their similarities, the two papers approach the issue of polarization demultiplexing from different angles. Kikuchi's main goal was to elucidate the physics behind the operation of the CMA-based polarization demultiplexers. In contrast, in our paper, we first derive the optimal polarization demultiplexer's structure, based on the maximum-likelihood criterion. Then, we prove that the performance of a zero-forcing polarization demultiplexer, in the absence of PMD and PDL, is optimal. Finally, we express the transfer matrix of the zero-forcing polarization demultiplexer, in the absence of PMD and PDL, as a function of only two real parameters (as opposed to two complex parameters in [11]), which are subsequently computed recursively using the CMA.

The rest of this paper is divided into three major sections, namely, theoretical model (Section II), simulation results and discussion (Section III), and experimental validation (Section IV). In Section II, we develop an equivalent, discrete-time model of a representative coherent optical communications system, using matrix formulation. Then, we apply the maximum-likelihood criterion, in order to derive an optimal decision metric for the joint estimation of the symbols, over both received SOPs. Based on the optimal decision metric, it is shown that, under certain ideal conditions, a zero-forcing linear receiver yields optimal performance. The final part of Section II is devoted to the proposed constrained polarization demultiplexer and the application of CMA for the blind adaptive estimation of its adjustable parameters. In Section III, we evaluate, by simulation, the performance of the proposed constrained CMA polarization demultiplexer, in terms of convergence properties and error probability. Finally, in Section IV, we experimentally test the capability of the proposed constrained CMA polarization demultiplexer in separating 2 GBd PDM QPSK optical signals. The details of the theoretical calculations are presented in the Appendices.

II. Theoretical Model

A. System Description

Fig. 1 shows the block diagram of a PDM QPSK optical communications system with a polarization- and phase-diversity coherent optical receiver. The modules of the optical transmitter are shown in detail in Fig. 1(a). The optical signal from a CW semiconductor laser diode (SLD) is equally split and fed into two parallel quadrature modulators (QM). Two independent pseudo-random bit sequences (PRBS), at a bit rate R_b each, are differentially encoded (DE) and transformed into pulse sequences, which, in turn, change the driving voltage of each QM. Two optical, differentially-encoded QPSK signals are generated, at a symbol rate $R_s = R_b$ each, at the output of the QMs. The two optical QPSK signals are superimposed with orthogonal SOPs, using two polarization controllers (PC) and

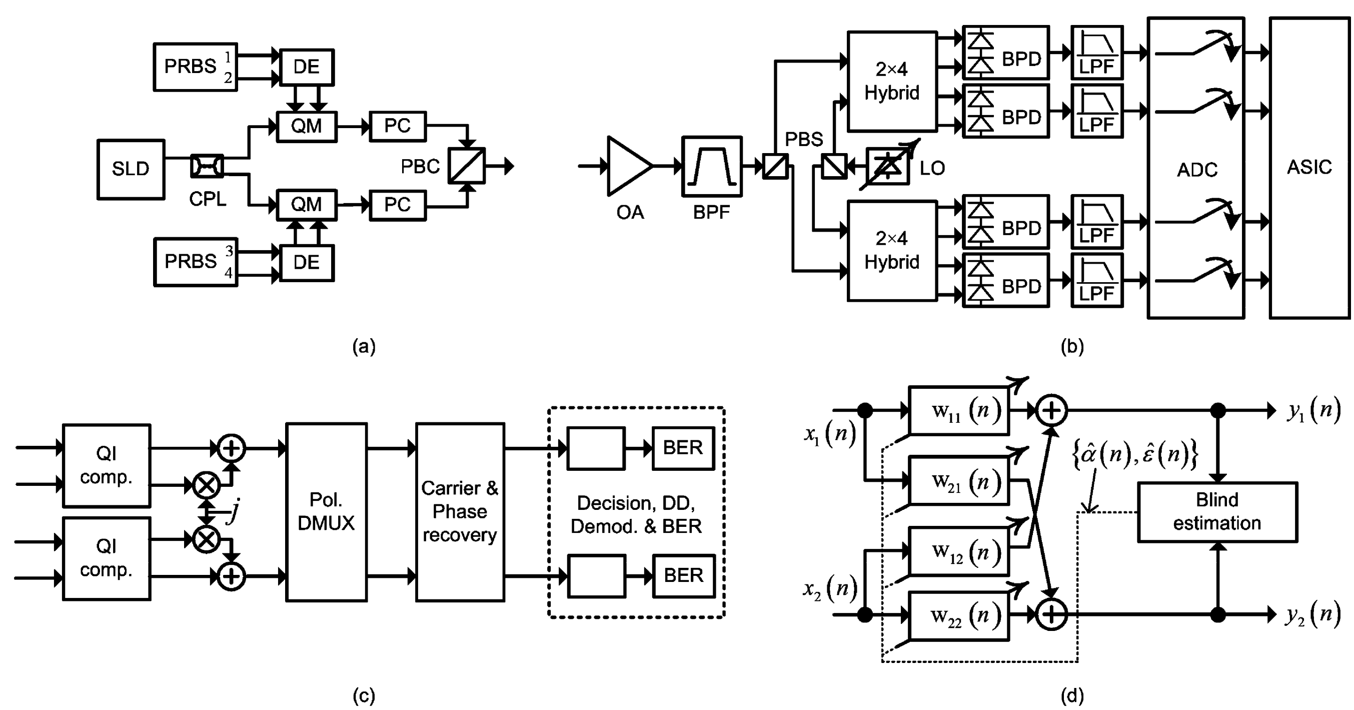

Fig. 1. Block diagram of a representative coherent optical system. (a) PDM QPSK transmitter (Symbols: PRBS: Pseudo-random bit sequence, DE: Differential encoding and pulse shaping, SLD: Semiconductor laser diode, CPL: 3-dB coupler, QM: Quadrature modulator, PC: Polarization controller, PBC: Polarization beam combiner.); (b) Polarization- and phase-diversity coherent optical receiver (Symbols: OA: Optical preamplifier, BPF: Optical bandpass filter, PBS: Polarization beam splitter, LO: Local oscillator, BPD: Balanced photodetectors, LPF: Lowpass filter, ADC: Analog-to-digital converter, ASIC: application specific integrated circuit.); (c) ASIC architecture (Symbols: QI comp.: Quadrature imbalance estimation and compensation, Pol. DMUX: Polarization demultiplexer, DD: Differential decoding, Demod.: Demodulation, BER: Bit error rate counter.); (d) Proposed constrained polarization demultiplexer (Symbols: $x_k(n)$: input photocurrents, $y_k(n)$: output photocurrents, $w_{kl}, k, l = 1, 2$: Complex taps, $\{\hat{\alpha}(n), \hat{\varepsilon}(n)\}$: Estimated azimuth and ellipticity).

a polarization beam combiner (PBC), to form a PDM QPSK signal, which is transmitted through an optical fiber.

The block diagram of an optical polarization- and phase-diversity digital coherent homodyne synchronous receiver is shown in Fig. 1(b). The optical receiver front-end is composed of an optical preamplifier (OA), an optical bandpass filter (BPF), a laser diode, acting as a local oscillator (LO), two polarization beam splitters (PBSs) with aligned principal axes, two 2×4 90° optical hybrids, and four balanced photodetectors (BPDs). The received optical signal is optically preamplified and filtered by the optical BPF, in order to reject the out-of-band amplified spontaneous emission (ASE) noise. The x- and y-polarization components of the received optical signal and the local oscillator are separately combined and detected by two identical phase-diversity receivers composed of a 2×4 90° optical hybrid and two BPDs each, at the upper and lower polarization branches, respectively. The photocurrents at the output of the four balanced detectors are low-pass filtered (LPF), sampled at integer multiples of the symbol period T_s, using an analog-to-digital converter (ADC), and fed to an application specific integrated circuit (ASIC) for DSP.

Fig. 1(c) shows the ASIC's architecture. The four sampled photocurrents are processed in pairs. Each pair corresponds to the in-phase and quadrature components of the coherent beating between the received signal and the signal of the LO. Initially, the quadrature imbalance (QI) occurring at each phase-diversity receiver is estimated and corrected [28]–[31]. The two quadratures are then combined, via complex addition, to form discrete-time, scaled replicas of the received complex electric field vectors at the x- and y-polarizations, respectively. Subsequently, polarization demultiplexing is performed [4]–[12], possibly combined with transmission impairments equalization [13]–[17]. The block diagram of the proposed constrained CMA polarization demultiplexer is shown in Fig. 1(d). The polarization demultiplexer attempts to counteract the channel effect by forming a linear superposition of the photocurrents. It has two inputs and two outputs and is composed of four complex multipliers $w_{kl}, k, l = 1, 2$, which are connected in a butterfly structure. The multipliers are iteratively adjusted, using the CMA. The rationale behind the structure of the proposed polarization demultiplexer is explained in detail in Section II-C.

Referring back to Fig. 1(c), after polarization demultiplexing, the complex envelopes of the electric fields of the PDM QPSK tributaries are recovered separately, at the upper and lower branches of the ASIC. The non-zero IF offset, due to the carrier frequency difference between the transmitter and the local oscillator lasers, is estimated and removed, using a feed-forward carrier recovery algorithm [32], [33]. A feed-forward phase noise removal circuit estimates and removes laser phase noise, e.g., [34]. Subsequently, the two waveforms are demodulated independently. Each symbol sequence is recovered using a decision circuit, is differentially decoded and transformed into two bit sequences, which are used for error counting.

B. Mathematical Formulation

In this subsection, an equivalent, discrete-time model of the coherent optical system of Fig. 1 is derived.

For mathematical convenience, an equivalent baseband representation [25] of the optical signals and components is used. In addition, the following notations are adopted: (i) To distinguish vectors from scalars, we identify vector quantities with boldface type; (ii) Matrices are also denoted by boldface type (the distinction should be clear from the context); and (iii) Dirac's bra and ket vectors denote normalized Jones vectors [35].

It is assumed that the electric fields of the two QPSK modulated waveforms, at the output of the QM, have orthogonal polarizations $|x\rangle$ and $|y\rangle$, respectively. It is also assumed that the optical fiber induces arbitrary, random, time-varying polarization rotations but maintains the orthogonality between the SOPs of the polarization multiplexed signals. After transmission through the optical fiber, at the output of the optical BPF, the electric field of the optical PDM QPSK signal can be written, in equivalent baseband notation, as

$$\mathbf{E}(t) = E_s(t)|e_s(t)\rangle + E_p(t)|e_p(t)\rangle \tag{1}$$

where $E_s(t), E_p(t)$ are the complex envelopes of the preamplified optical signals contaminated with ASE noise, and $|e_s(t)\rangle, |e_p(t)\rangle$ are the corresponding slowly-varying, normalized, Jones vectors along two arbitrary orthogonal SOPs, denoted by s, p. Based on the assumption of SOP orthogonality, the inner product of the two Jones vectors must vanish, i.e., $\langle e_k(t) e_l(t)\rangle = \delta_{kl}, k, l = s, p$, where δ_{kl} is Kronecker's delta.

The complex envelopes of the electric fields in (1) can be written as

$$\begin{aligned} E_s(t) &= [\sqrt{2P_s}e^{j\theta_s(t)}u_1(t) + n_s(t)]e^{j\omega_s t} \\ E_p(t) &= [\sqrt{2P_p}e^{j\theta_s(t)}u_2(t) + n_p(t)]e^{j\omega_s t} \end{aligned} \tag{2}$$

where P_s, P_p are the average optical powers at each SOP, $u_1(t), u_2(t)$ are the modulating signals, ω_s is the angular frequency offset from the channel's nominal frequency, and $\theta_s(t)$ is the phase noise of the received signal. The terms $n_s(t), n_p(t)$ represent independent, identically distributed, complex ASE noise components in the two orthogonal SOPs, which follow Gaussian distribution with zero mean and variance σ^2_{ASE}.

The normalized Jones vectors $|e_s(t)\rangle, |e_p(t)\rangle$ can be expressed in rectangular coordinates as [36]

$$|e_s(t)\rangle = \begin{bmatrix} \cos\alpha(t)\cos\varepsilon(t) - j\sin\alpha(t)\sin\varepsilon(t) \\ \sin\alpha(t)\cos\varepsilon(t) + j\cos\alpha(t)\sin\varepsilon(t) \end{bmatrix} \tag{3}$$

and

$$|e_p(t)\rangle = \begin{bmatrix} -\sin\alpha(t)\cos\varepsilon(t) + j\cos\alpha(t)\sin\varepsilon(t) \\ \cos\alpha(t)\cos\varepsilon(t) + j\sin\alpha(t)\sin\varepsilon(t) \end{bmatrix} \tag{4}$$

where $\alpha(t)$ and $\varepsilon(t)$ are the angles corresponding to the s-SOP's azimuth and ellipticity [36], respectively, and take values in the intervals $|\alpha(t)| \leq \pi/2, |\varepsilon(t)| \leq \pi/4$ [36].

In Appendix A, it is shown that an array of two complex photocurrents $\mathbf{X}(n)$ is generated at the output of the polarization- and phase-diversity coherent optical receiver, which, in the absence of phase noise, can be written as

$$\mathbf{X}(n) = \mathbf{H}(n)\mathbf{U}(n) + \mathbf{N}(n) \tag{5}$$

where $\mathbf{U}(n)$ is the array of the two sampled modulating signals scaled by a multiplication factor, $\mathbf{N}(n)$ is the array of the total photocurrent noises, and $\mathbf{H}(n)$ is the transfer function of the transmission channel

$$\mathbf{H}(n) = \begin{bmatrix} \langle x|e_s(n)\rangle & \langle x|e_p(n)\rangle \\ \langle y|e_s(n)\rangle & \langle y|e_p(n)\rangle \end{bmatrix}. \tag{6}$$

The mean and covariance matrices of the sampled modulating signals $\mathbf{U}(n)$ are

$$\boldsymbol{\mu}_{\mathbf{U}} = \mathrm{E}\{\mathbf{U}(n)\} = \mathbf{0} \tag{7}$$

and

$$\mathbf{K}_{\mathbf{U}} = \mathrm{E}\{\mathbf{U}(n)\mathbf{U}^\dagger(n)\} = \mathrm{diag}\left\{A_1^2, A_2^2\right\} \tag{8}$$

respectively, where $\mathrm{E}\{\cdot\}$ denotes expectation, $\mathrm{diag}\{\cdot\}$ denotes a diagonal matrix, dagger denotes the adjoint, i.e., conjugate transpose, matrix, and A_1, A_2 are the signal amplitudes at each receiver branch.

The mean and covariance matrices of the total photocurrent noises $\mathbf{N}(n)$ are given by

$$\boldsymbol{\mu} = \mathrm{E}\{\mathbf{N}(n)\} = \mathbf{0} \tag{9}$$

and

$$\mathbf{K} = \mathrm{E}\{\mathbf{N}(n)\mathbf{N}^\dagger(n)\} = \mathrm{diag}\left\{\sigma_1^2, \sigma_2^2\right\} \tag{10}$$

respectively, where σ_1^2, σ_2^2 are the variances of the photocurrent noise components at each receiver branch.

It is straightforward to verify that $\mathbf{H}(n)$ belongs to the special unitary group SU(2), i.e., $\mathbf{H}(n)\mathbf{H}^\dagger(n) = \mathbf{H}^\dagger(n)\mathbf{H}(n) = \mathbf{I}$, where $\mathbf{I}$ denotes the 2×2 unit matrix, and $\det[\mathbf{H}(n)] = 1$, where the operator $\det[\cdot]$ denotes the determinant of a matrix.

C. Optimal Receiver

In Appendix B, using the maximum-likelihood criterion [25]–[27], we derive the decision metric of the optimal receiver for joint detection of PDM QPSK signals transmitted over the memoryless, discrete-time, two-input two-output (TITO) linear channel described by (5).

It is shown that a sufficient statistic for estimating the transmitted symbols is the following (see (57) in Appendix B)

$$\hat{\mathbf{U}} = \arg\min_{\mathbf{U}\in\mathcal{U}}\{\mathbf{U}^\dagger\mathbf{H}^\dagger\mathbf{K}^{-1}\mathbf{H}\mathbf{U} - 2\Re\{\mathbf{U}^\dagger\mathbf{H}^\dagger\mathbf{K}^{-1}\mathbf{X}\}\} \tag{11}$$

where $\Re\{\cdot\}$ denotes real part and $\mathcal{U}$ is the joint, complex-symbol alphabet. We have dropped the time dependence of all matrices in order to avoid clutter.

From (11), we observe that, prior to decision, the optimal receiver must form a linear superposition $\mathbf{Y}$ of the complex photocurrents $\mathbf{X}$

$$\mathbf{Y} = \mathbf{W}\mathbf{X} \tag{12}$$

where $\mathbf{W}$ is the transfer matrix of a spatial electronic filter

$$\mathbf{W} = \mathbf{H}^{\dagger}\mathbf{K}^{-1}. \tag{13}$$

Substitution of (12) into (11) leads to the concise decision rule

$$\hat{\mathbf{U}} = \arg\min_{\mathbf{U}\in\mathcal{U}}\{\mathbf{U}^{\dagger}\mathbf{H}^{\dagger}\mathbf{K}^{-1}\mathbf{H}\mathbf{U} - 2\Re\{\mathbf{U}^{\dagger}\mathbf{Y}\}\}. \tag{14}$$

We conclude that, in the general case, when the total photocurrent noise is not spatially white, i.e., $\sigma_1 \neq \sigma_2$, the optimal receiver should perform the following steps: (i) spatial noise pre-whitening, i.e., multiplication by $\mathbf{K}^{-1/2}$; (ii) projection to $\mathbf{K}^{-1/2}\mathbf{H}$; and (iii) joint maximum-likelihood vector symbol estimation using the concise metric (14). This receiver is called the linear minimum mean squared error (MMSE) receiver [27].

D. Zero-Forcing Receiver

In the special case when the noises at the two branches of the polarization diversity receiver have the same variance (i.e., the photocurrent noise is spatially white), the optimal receiver may use the simplified decision rule (see (61) in Appendix B)

$$\hat{\mathbf{U}} = \arg\min_{\mathbf{U}\in\mathcal{U}}\|\mathbf{Y} - \mathbf{U}\|^2 \tag{15}$$

where the spatial electronic filter transfer matrix is now reduced to

$$\mathbf{W} = \mathbf{H}^{\dagger}. \tag{16}$$

The above relationship indicates that the optimal receiver must use an electronic polarization demultiplexer, in order to fully invert the fiber Jones matrix.

It would be instructive to gain some insight into why (16) is optimal. Assume that the receiver has perfect knowledge of the channel transfer matrix. According to (16), the receiver should set $\mathbf{W} = \mathbf{H}^{\dagger}$. Then, the output of the polarization demultiplexer is written as

$$\mathbf{Y}_0(n) = \mathbf{U}(n) + \mathbf{H}^{\dagger}(n)\mathbf{N}(n). \tag{17}$$

Since the multiplication with a matrix is a linear operation, the resulting noise $\mathbf{N}_0(n) = \mathbf{H}^{\dagger}(n)\mathbf{N}(n)$ is a complex Gaussian random vector. Using (9) and (10), for $\sigma_1^2 = \sigma_2^2 = \sigma^2$, we can calculate the mean $\boldsymbol{\mu}_0$ and covariance $\mathbf{K}_0$ matrices of $\mathbf{N}_0(n)$

$$\boldsymbol{\mu}_0 = \mathrm{E}\{\mathbf{N}_0(n)\} = \mathbf{H}^{\dagger}\mathrm{E}\{\mathbf{N}(n)\} = \mathbf{0} \tag{18}$$

and

$$\begin{aligned}\mathbf{K}_0 &= \mathrm{E}\left\{\mathbf{N}_0(n)\mathbf{N}_0^{\dagger}(n)\right\} \\ &= \mathbf{H}^{\dagger}\mathrm{E}\{\mathbf{N}(n)\mathbf{N}^{\dagger}(n)\}\mathbf{H} = \sigma^2\mathbf{I}.\end{aligned} \tag{19}$$

We observe that the photocurrent noise statistics are preserved after the proposed polarization demultiplexer. This indicates that polarization demultiplexing can be achieved without penalty. However, were it not for the equality of the total photocurrent noise variances and the unitarity of the channel transfer matrix, the performance of the linear zero-forcing receiver would be suboptimal, compared to joint maximum likelihood detection, since the noises at the output branches of the polarization demultiplexer would be correlated (19) [27].

As shown in Appendix B, the decision rule (15) can be further reduced so that each element $\hat{u}_k$ of $\hat{\mathbf{U}}$ can be individually estimated at each branch of the polarization-diversity receiver. Furthermore, the in-phase and quadrature components of $\hat{u}_k$ are independently retrieved, by comparison to a zero threshold.

Polarization demultiplexing, followed by separate detection of the two PDM channels, is a special case of a well-known receiver structure in the context of multiuser systems with space diversity, called *zero-forcing linear receiver, interference nuller*, or *decorrelator* [27]. It is worth noting that the proposed receiver is a direct extension to two dimensions of the maximal ratio combiner, used in single channel communications systems, with polarization-diversity coherent receivers [3]. In addition, the proposed polarization demultiplexer transfer matrix is the exact 2×2 equivalent of the matched filter transfer function [27], [11].

E. Estimation of Channel Transfer Matrix

The zero-forcing receiver must calculate an estimate $\hat{\mathbf{H}}(n)$ of the channel transfer matrix in (5) and set $\mathbf{W}(n) = \hat{\mathbf{H}}^{\dagger}(n)$. From (3)–(4) and (6), we observe that the elements of $\mathbf{H}(n)$ are functions of only two independent parameters $\alpha(n)$ and $\varepsilon(n)$. Therefore, one simply needs to calculate the estimates of the angles $\hat{\alpha}(n), \hat{\varepsilon}(n)$. Below, we show that this can be achieved blindly using the CMA. Its application on optical channel parameters estimation is a novel idea. Its adequacy, compared to other estimators [37], lies out of the scope of the present study. No claims about its optimality are made. Its adoption, as an appropriate scheme for the estimation of $\hat{\alpha}(n)$ and $\hat{\varepsilon}(n)$, can be justified by its tolerance to intermediate frequency offsets and laser phase noise, and its excellent bit error rate (BER) performance shown in Fig. 5(c).

The instantaneous error function is defined as

$$\mathcal{E}(n) = \mathbf{Y}^*(n) \bullet \mathbf{Y}(n) - \boldsymbol{\Lambda} \tag{20}$$

where the operator $\bullet$ denotes the Hadamard matrix product [39], defined as the component-wise multiplication of two matrices

$$\mathbf{C} = \mathbf{A} \bullet \mathbf{B} \Leftrightarrow [C]_{ij} = a_{ij}b_{ij} \tag{21}$$

and

$$\boldsymbol{\Lambda} = \begin{bmatrix} R_0^{(1)} \\ R_0^{(2)} \end{bmatrix} \tag{22}$$

where $R_0^{(k)}, k = 1, 2$ are the total signal and noise powers at each branch of the polarization-diversity receiver.

The cost function, which we seek to minimize, can be defined as the total mean-squared error

$$\xi = \mathrm{E}\{\mathcal{E}^T(n)\mathcal{E}(n)\}. \tag{23}$$

The instantaneous cost function can then be expressed in terms of $\mathcal{E}(n)$ as

$$\xi(n) = \mathcal{E}^T(n)\mathcal{E}(n). \tag{24}$$

We define the auxiliary column vector $\mathbf{Z}(n)$ with elements equal to the independent parameters

$$\mathbf{Z}(n) = [z_1(n) \quad z_2(n)]^{\mathrm{T}} = [\hat{\alpha}(n) \quad \hat{\varepsilon}(n)]^{\mathrm{T}}. \tag{25}$$

Taking the derivative of the instantaneous cost function with respect to z_k and using the chain rule, one finally obtains analytical expressions for $\partial\xi(n)/\partial z_k, k=1,2$

$$\begin{aligned}\partial\xi(n)/\partial z_k = 2\mathcal{E}^T(n)\{&\mathbf{Y}^*(n)\bullet[\partial\mathbf{W}(n)/\partial z_k\mathbf{X}(n)]\\ &+\mathbf{Y}(n)\bullet[\partial\mathbf{W}^*(n)/\partial z_k\mathbf{X}^*(n)]\}, k=1,2\end{aligned} \tag{26}$$

We define the gradient of the instantaneous cost function in the space of the independent variables

$$\nabla\xi(n) = \begin{bmatrix} \frac{\partial\xi(n)}{\partial z_1} & \frac{\partial\xi(n)}{\partial z_2} \end{bmatrix}^{\mathrm{T}}. \tag{27}$$

The stochastic gradient algorithm for updating the adaptive filter coefficients is written [21], [25]

$$\mathbf{Z}(n+1) = \mathbf{Z}(n) - \mu\nabla\xi(n) \tag{28}$$

where μ is a positive real constant (*step-size parameter*). Using periodic boundary conditions, the parameters $\hat{\alpha}(n)$ and $\hat{\varepsilon}(n)$ in (28) are confined within a unit cell, delimited by $|\hat{\alpha}| \leq \pi/2, |\hat{\varepsilon}| \leq \pi/4$.

It should be stressed that both the proposed constrained CMA polarization demultiplexer and its conventional counterpart suffer from output permutation and rotation ambiguities. The first type of ambiguity means that PDM channel ordering at the polarization demultiplexer outputs is unpredictable. The second type of ambiguity means that the recovered constellations might exhibit arbitrary rotations from their nominal position. In other words, both polarization demultiplexers cannot distinguish between the desired solution (in the absence of noise)

$$\mathbf{Y}_0(n) = \mathbf{U}(n) \tag{29}$$

and the undesired solution

$$\mathbf{Y}'(n) = [u_2(n)e^{j\phi_1} \quad u_1(n)e^{j\phi_2}]^{\mathrm{T}} \tag{30}$$

where ϕ_1, and ϕ_2 are arbitrary phase rotations.

The output permutation ambiguity can be addressed, for instance, by periodically transmitting channel identification training symbols (pilots) and by using a tracking scheme for their detection.

The rotation ambiguity can be readily unraveled by the feedforward laser phase noise estimation circuit [34], which is assisted by differential coding and decoding of the transmitted symbols [3].

It is worth mentioning that a combination of data-aided and blind estimation of $\hat{\alpha}, \hat{\varepsilon}$ can be performed, as well. Since polarization rotations, due to fiber birefringence, are slow, in comparison with the symbol rate, they can be considered a quasi-static effect. Estimation of quasi-static effects can be performed in two phases, i.e., training and tracking. During the training phase, a short training sequence, in conjunction with the least squares method [37], can be used to estimate $\hat{\mathbf{H}}$. This method asymptotically yields the best linear unbiased estimate [37]. During the time intervals between transmitting consecutive training sequences, a blind estimation algorithm, e.g., [38], can be used for tracking and continuously updating the values of the adjustable parameters. The values provided by the training phase can be used as initial guesses for the recursion of the tracking algorithm.

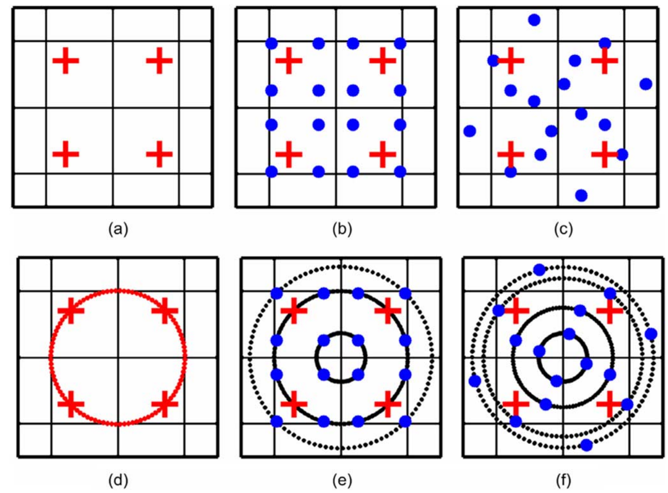

Fig. 2. Constellation diagrams for the x-polarization, at the output of the polarization diversity receiver, in the absence ((a), (b), (c)) and in the presence ((d), (e), (f)) of IF offset. (Conditions: Received SOP parameters: (a), (d) $\alpha = \varepsilon = 0$, (b), (e) $\alpha = \pi/3, \varepsilon = 0$, and (c), (f) $\alpha = \varepsilon = \pi/3$).

III. Simulation Results and Discussion

In order to theoretically evaluate the performance of the proposed constrained CMA polarization demultiplexer and compare it with its conventional counterpart [10], we perform computer simulations of the coherent optical communication system shown in Fig. 1. We use ideal non-return-to-zero (NRZ) QPSK signals. The optical fiber is modeled simply as a polarization rotator with transfer matrix $\mathbf{H}(n)$. The optical hybrids are considered ideal and all photodiodes are identical. All simulations are performed with initial guesses $\mathbf{Z}(0) = [\hat{\alpha}(0) \quad \hat{\varepsilon}(0)]^{\mathrm{T}} = [0 \quad 0]^{\mathrm{T}}$ for the proposed constrained CMA polarization demultiplexer and $\mathbf{W}(0) = \mathbf{I}$ for its conventional counterpart.

Initially, the impact of polarization rotations on the constellation of received sampled complex photocurrents is investigated. We distinguish two cases, in the absence and in the presence of phase noise and IF offset, respectively. First, the ideal constellation is shown as a reference (Fig. 2(a), black (red) crosses). Due to cross-polarization interference, received constellations consist of 16 points, Fig. 2(b), (c). This occurs because the mixing matrix $\mathbf{H}(n)$ creates all possible combinations of two constellations of four points each. Depending on the specific values of α, ε, some of the constellation points may overlap. In the presence of IF offset, constellation points rotate either clockwise or counterclockwise, producing up to four concentric circles with unequal radii, as shown in Fig. 2(e)–(f).

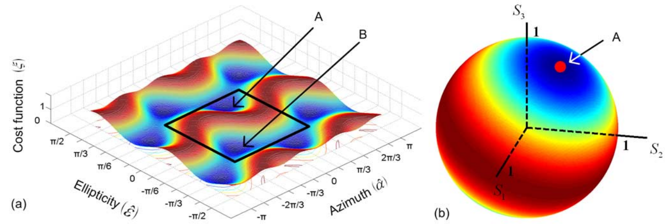

Fig. 3. (a) Three-dimensional plot of the cost function vs. estimated azimuth and ellipticity $\hat{\alpha}, \hat{\varepsilon}$. (Symbols: A, B: Global minima within the unit cell, Rectangle: Unit cell), (b) Corresponding contour plot of the cost function on the Poincaré sphere for the proposed constrained polarization demultiplexer. (Symbols: White (red) point (A): Minimum at $\hat{\alpha} = \hat{\varepsilon} = \pi/6$). (Conditions: Received SOP parameters: $\alpha = \varepsilon = \pi/6$). (Color coding: Black (blue) areas: Small values of the cost function, White (red) areas: Large values of the cost function).

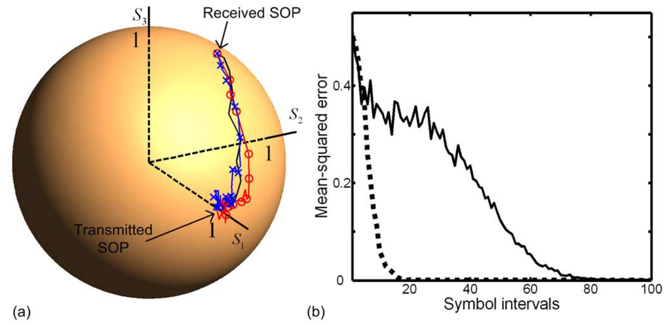

Fig. 4. (a) Trajectory of the estimated SOP at the output of the proposed constrained polarization demultiplexer. (Conditions: Received SOP parameters: $\alpha = \varepsilon = \pi/6$, BPF bandwidth= $32R_s$, LPF 3-dB bandwidth: $B_e = 0.8R_s$, LPF equivalent noise bandwidth: $B_e^{\mathrm{eq}} = 1.04B_e$, phase averaging block size $= 10$ symbols), (Symbols: Black (unmarked) line: Absence of noise, Red line (open circles): $\mathrm{OSNR} = 9$ dB and, Blue line (crosses): $\mathrm{OSNR} = 7$ dB) (b) Time evolution of the MSE for the x polarization for the proposed constrained (dotted line), and the conventional CMA-based (solid line), polarization demultiplexers, for the optimum step size $\mu = 0.1$. Averaging over 500 experiments is performed. (Conditions: Received SOP parameters: $\alpha \approx \pi/4$ and $\varepsilon = 0$).

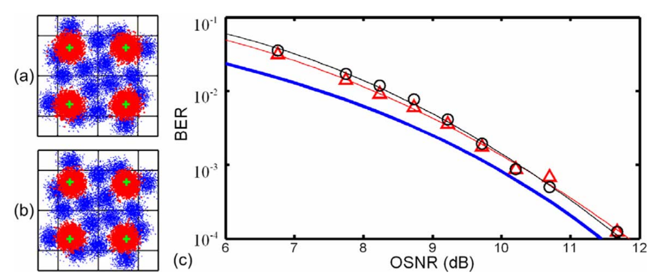

Fig. 5. Representative constellations of the received (blue (gray) points), equalized (red (black) points) and ideal ((green) crosses) signals for the x polarization, using (a) the proposed constrained polarization demultiplexer, and (b) the conventional CMA-based polarization demultiplexer; (c) BER vs. OSNR for the ideal (i.e., distortionless) case (blue (thick black) curve), and using the proposed constrained ((red) triangles), and the conventional CMA-based ((black) circles), polarization demultiplexer. Quadratic polynomial fitting is used, drawn as a dotted line (in red) for the proposed constrained and as a thin solid line (in black) for the conventional CMA-based polarization demultiplexer. (Conditions: Received SOP parameters: $\alpha = \pi/6$ and $\varepsilon = \pi/12$, BPF $\mathrm{bandwidth} = 32R_s$, LPF 3-dB bandwidth: $B_e = 0.8R_s$, 4th-order Bessel LPF equivalent noise bandwidth: $B_e^{\mathrm{eq}} = 1.04B_e$, phase averaging block $\mathrm{size} = 10$ symbols).

We proceed with exploring the performance surface $\xi = \mathrm{E}\{\mathcal{E}^T(n)\mathcal{E}(n)\}$ as a function of the proposed demultiplexer adjustable parameters. Fig. 3(a), (b) show three-dimensional and Poincaré sphere-contour plots, respectively, of the instantaneous cost function $\xi(n)$ vs. $\hat{\alpha}, \hat{\varepsilon}$, in the absence of noise, for $\alpha = \varepsilon = \pi/6$. Time averaging of the cost function $\xi(n)$, over 500 consecutive symbols, is performed instead of ensemble averaging. In Fig. 3(a), we observe that, within the limits of the unit cell, denoted by a rectangle, there are two global minima A, B. In Fig. 3(b), these minima correspond to two antipodal points on the Poincaré sphere. The minimum at the point $A = [\hat{\alpha} \quad \varepsilon]^{\mathrm{T}} = [\alpha \quad \varepsilon]^{\mathrm{T}}$ (also denoted by a (white) red point on the Poincaré sphere), corresponds to the correct ordering of the output signals, whereas the other minimum at the point $B = [\hat{\alpha} \quad \hat{\varepsilon}]^{\mathrm{T}} = [\alpha \pm \pi/2 \quad -\varepsilon]^{\mathrm{T}}$ (not shown on the Poincaré sphere), results in a permutation of the output channels. Bisecting the line connecting the two minima on the Poincaré sphere with a perpendicular plane, divides the sphere into two hemispheres. The intersection of the plane, with the surface of the sphere, creates a rotated equator line, which corresponds to the ridge within the unit cell of Fig. 3(a). The hemisphere of each minimum in Fig. 3(b) corresponds to a valley within the unit cell of Fig. 3(a). The constrained CMA converges to the minimum lying in the same hemisphere as the initialization point $\hat{\alpha}(0), \hat{\varepsilon}(0)$. If the initial point lies exactly on the equator, in the presence of noise, the algorithm may converge to either minimum.

Subsequently, we study the impact of ASE noise on the convergence behavior of the proposed constrained CMA polarization demultiplexer. We assume that the transmitted orthogonal SOPs are $|x\rangle, |y\rangle$, corresponding to the Stokes vectors $[1 \quad 0 \quad 0]^{\mathrm{T}}$ and $[-1 \quad 0 \quad 0]^{\mathrm{T}}$, respectively. We also assume that the received SOPs have been rotated due to fiber birefrigence in relation to the transmitted ones, such that the s-SOP angles are $\alpha = \varepsilon = \pi/6$. The polarization demultiplexer iteratively estimates the angles $\hat{\alpha}, \hat{\varepsilon}$ and restores the SOPs of the PDM QPSK signals back to their initial values. Fig. 4(a) shows the trajectories followed on the Poincaré sphere during restoration from $|e_s\rangle$ to $|x\rangle$, both in the absence (black (unmarked) line) and in the presence of ASE noise (for two different values of the optical signal-to-noise ration (OSNR)). We observe that, in the absence of ASE noise, the restored SOP eventually coincides with the transmitted one, whereas, as the OSNR is reduced, the restored SOP fluctuates more significantly around the initially transmitted SOP.

In Fig. 4(b), we compare the decay time of the error magnitude for the proposed constrained CMA polarization demultiplexer (dotted line), and the conventional one (solid line) [10]. More specifically, for both demultiplexer types, we plot the time evolution of the mean-squared error $\mathrm{E}\{e_1^2(n)\}$, for the x-polarization, for the optimal value of the step-size parameter μ. We choose a maximal initial perturbation; that is to say, the initial point is selected adjacent to the rotated equator and within the appropriate hemisphere, in order to prevent output reversal. Ensemble averaging over 500 simulation runs is performed. The optimum step-size parameter is $\mu = 0.1$ for both polarization and demultiplexers, providing fast convergence and negligible residual mean squared error (MSE). The constrained CMA polarization demultiplexer clearly exhibits

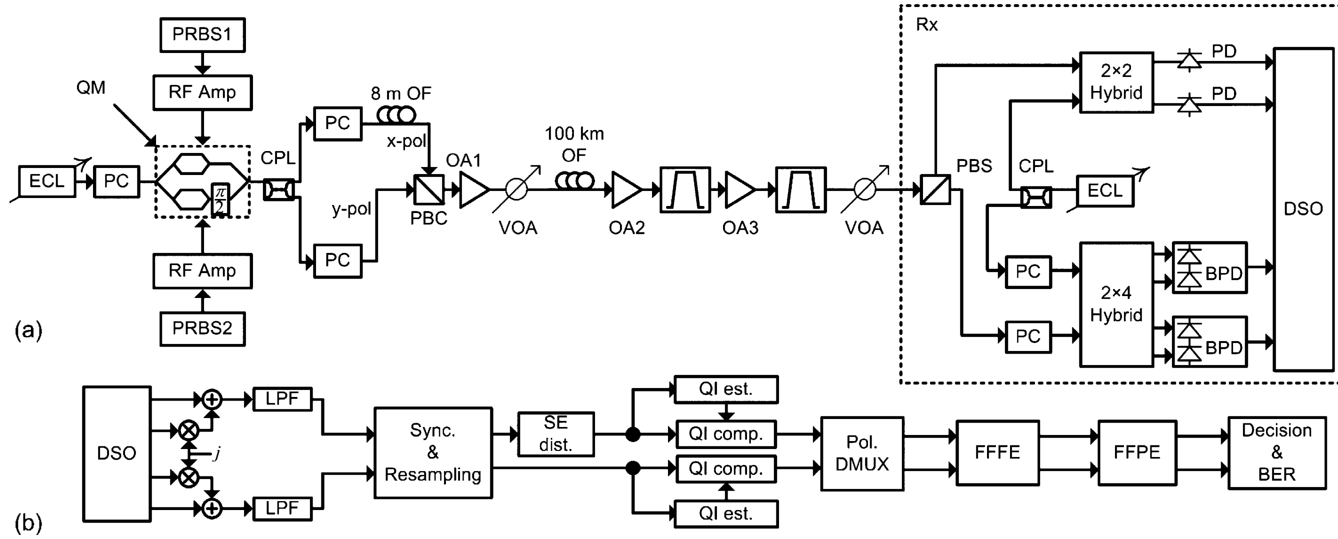

Fig. 6. (a) Block diagram of the experimental setup (Symbols: ECL: External cavity laser, QM: Quadrature modulator, CPL: 3-dB coupler, PC: Polarization controller, PRBS: Pseudo-random binary sequence, RF Amp: Radio frequency amplifier, OF: Optical fiber, PBC: Polarization beam combiner, VOA: Variable optical attenuator, OA: Optical amplifier, Rx: Polarization- and phase-diversity coherent optical receiver, BPD: Balanced photodetector, DSO: Digital storage oscilloscope.); (b) Block diagram of the DSP modules used to analyze the experimental results. (Symbols: Sync. & Resampling: Quadrature and polarization time synchronization and resampling, SE dist: Single-ended distortion mitigation, FFFE: Feed-forward frequency estimation and removal, FFPE: Feed-forward phase estimation and removal, BER: Bit error rate tester.).

superior performance, in terms of convergence speed. For example, for $\mu = 0.1$, the constrained CMA polarization demultiplexer requires less than 20 symbol intervals to minimize the MSE, whereas its conventional counterpart [10] requires more than 60 symbol intervals, i.e., it is more than three times slower.

Fig. 5(a), (b) show representative input/output constellation diagrams, with ASE noise and zero IF offset, obtained by using the constrained CMA polarization demultiplexer and the conventional one, respectively. We assume that the received SOPs correspond to $\alpha = \pi/6, \varepsilon = \pi/12$. We see that both polarization demultiplexers are able to transform the received spiral constellation (blue (gray) points) into four approximately circular points (in red (black)) that approach the transmitted constellation (green crosses).

Fig. 5(c) shows BER curves, as a function of the OSNR, measured in a resolution bandwidth $B_o \cong 1.25R_s$, (e.g., OSNR measured in 0.1 nm resolution bandwidth for $R_s = 10$ GBd) for both polarization demultiplexers. The BER is calculated using Monte Carlo simulation. The red triangles correspond to the proposed constrained CMA polarization demultiplexer, the black circles correspond to the conventional CMA polarization demultiplexer, and the blue (thick black) curve corresponds to the ideal case, with no polarization rotations [2], [3]. Obviously, both polarization demultiplexers exhibit almost identical performance, with a negligible penalty relative to the ideal case at high OSNRs.

IV. Experimental Validation

In order to test the validity of the simplifying assumptions of the theoretical model presented in Section II, we performed a series of PDM QPSK experiments. The experimental set-up is shown in Fig. 6(a).

Light from an external cavity laser (ECL), acting as a transmitter, is QPSK modulated using a QM, driven by two 2 Gb/s PRBSs. The optical signal at the output of the QM is split into two equal amplitude components, using a 3-dB coupler. One of the two components is delayed using approximately 8 m of optical fiber. Their SOPs are adjusted, using two PCs, so that they become aligned with the principal axes of the PBC. The PDM QPSK signal, at the output of the PBC, is amplified using a booster optical amplifier (OA1) and is subsequently transmitted through 100 km of LEAF® optical fiber. The latter, at 2 GBd, simply acts as a polarization rotator and attenuator. The received optical signal is preamplified and filtered in two stages, using two tunable fiber Bragg grating (FBG) filters. The first FBG filter has 0.6 nm bandwidth, in order to emulate a WDM DMUX. The second FBG filter has 0.25 nm bandwidth, in order to emulate the ASE noise-limiting filter, typically used after the optical preamplifier. As the optical field reaches the polarization- and phase-diversity coherent optical receiver, it is split using a PBS. The two polarization components are combined with the light of an ECL, acting as a LO. Local-oscillator-to-signal power ratio (LOSPR) was kept small due to limitations of the lasers used at the experiment. Two different optical hybrid technologies were used, namely, a bulk-component, 2×2 90° optical hybrid [40], and a commercially-available, integrated 2×4 90° optical hybrid [6]. At the output of the 2×2 90° optical hybrid, two, almost identical, 10-GHz bandwidth PDs are used. The integrated 2×4 90° optical hybrid is followed by two pairs of 40-GHz bandwidth BPDs. Finally, an 8-GHz electrical bandwidth, 40 GSa/s, real-time, sampling oscilloscope samples the photocurrents and stores the signals for off-line processing. An electrical spectrum analyzer, not shown in Fig. 6(a), is used for the manual adjustment of the transmitter and LO frequencies within $\sim$400 MHz from each other. The duration of a single measurement is equal to 51.25 μs.

Fig. 6(b) shows the block diagram of the DSP modules used to analyze the experimental results. First, we filter the translated spectrum using an LPF, in order to remove out-of-band noise.

Timing recovery is manually performed, in order to remove any differential delays between the signals, which are caused by optical and electrical path differences. Subsequently, signals are resampled to one sample per symbol.

Distortion due to small LOSPR and single-ended detection at the 2×2 90° optical hybrid is first partly removed [41]. Inaccuracies in the bias voltages of the 2×4 90° optical hybrid, as well as non-optimal setting of the four PCs within the 2×2 90° optical hybrid, in conjunction with differences in the responsivity of the PDs, cause QI [29]. QI is a slowly varying impairment, essentially constant over the duration of a single measurement. Several methods have been proposed for QI estimation and compensation, both in optical communications [28]–[31], and in digital communications, e.g., [42]. Here we use the algorithm described in [29], [31]. After QI compensation, the PDM QPSK signals are fed into the proposed electronic polarization demultiplexer. After polarization demultiplexing, any residual IF is estimated and removed using a feed-forward frequency estimation algorithm [32], [33]. A feed-forward phase noise removal circuit estimates and removes laser phase noise [34]. Finally, the signal corresponding to each quadrature passes through a decision circuit. The symbol sequence for each PDM QPSK signal is recovered and is differentially decoded. Then, it is transformed into two bit sequences, which are compared to the transmitted ones in order to perform error counting.

It is important to note that all DSP algorithms based on the assumption of the envelope constancy of QPSK signals, (i.e., [32]–[34]), cannot be applied prior to polarization demultiplexing. Otherwise, symbol errors occur due to the presence of multiple signal levels, caused by cross-polarization interference, as shown in Fig. 2. Therefore, residual IF estimation and phase noise estimation should be performed only *after* the proposed constrained CMA polarization demultiplexer. The latter is insensitive to IF offset and laser phase noise, so their presence does not affect the correct estimation of the fiber transfer matrix parameters $\hat{\alpha}, \hat{\varepsilon}$.

Fig. 7 shows typical constellations for the x-polarization (upper row) and y-polarization (lower row), respectively, immediately after synchronization and downsampling (Fig. 7(a), (b)), at the input of the proposed polarization demultiplexer (Fig. 7(c), (d)), at the output of the proposed polarization demultiplexer (Fig. 7(e), (f)), and after the IF and phase noise removing circuits (Fig. 7(g), (h)). The scale of constellations (a)–(d) is different from the scale of constellations (e)–(g) because the samples are normalized to unit magnitude at the input of the polarization demultiplexer. The initial constellations shown in Fig. 7(a), (b) contain QI, as witnessed by their elliptical shape. The constellation of Fig. 7(a), corresponding to the 2×2 90° optical hybrid, has the form of eccentric ellipses, a shape due to the distortion introduced by single-ended detection, combined with small LOSPR. The QI compensated constellations of Fig. 7(c), (d), resemble the ones plotted in Fig. 7(e), (f). Concentric circles with unequal radii are a tell-tale sign of cross-polarization interference between the two PDM QPSK signals. Due to the presence of ASE noise, circles are transformed into thick rings, whose circumferences may overlap. The constellations at the output of the polarization demultiplexer, seen in Fig. 7(e), (f), are single circles, indicating that fiber transfer matrix inversion was successfully performed. The difference in sizes between the

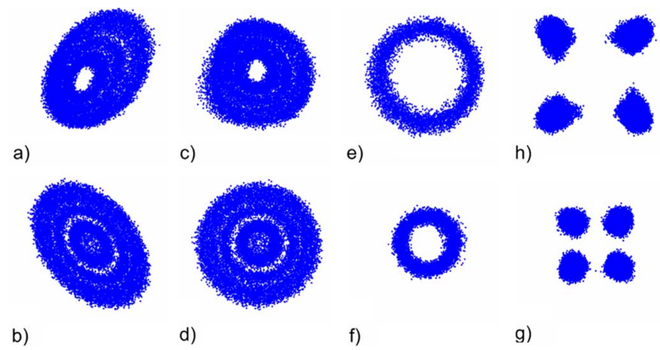

Fig. 7. Typical constellation diagrams in the x polarization (upper row) and the y polarization (lower row). (a), (b) after synchronization and downsampling; (c), (d) after QI compensation; (e), (f) at the output of the proposed polarization demultiplexer; (g), (h) after the IF and phase noise removing circuits. (Conditions: $R_s = 2$ GBd).

final x- and y-polarization constellations is primarily attributed to *polarization imbalance* (not to be confused with QI) due to gain and phase differences between the two branches of the polarization-diversity receiver (see analysis in Appendix C). In addition, the two polarization tributaries that are combined at the PBC, may have slightly unequal average powers due to maladjustment of the PCs or the non-ideal power splitting ratio of the 3-dB coupler. Fig. 7(g), (h) show the final constellations. The impact of polarization imbalance is obvious since the recovered constellations have unequal radii. No errors occur during a single measurement (i.e., 100 000 symbols) in both branches of the polarization diversity receiver.

Fig. 8(a) illustrates the time evolution of the estimates of the azimuth $\hat{\alpha}$ and ellipticity $\hat{\varepsilon}$. A variable step-size μ is used. In order to bring the operating point near the optimum quickly we start with a relatively large value of μ. The value of μ is halved after 100 symbols and again after another 200 symbols, to avoid large baseline wander. We can see that while the ellipticity angle $\hat{\varepsilon}$ approaches its final value after fewer than 100 symbols, the azimuth $\hat{\alpha}$ requires around 200 symbols to do the same. The azimuth and ellipticity remain stable over the rest of the measurement, confirming that polarization rotations are a slowly varying effect. Fig. 8(b) shows the time evolution of the instantaneous squared error function for the x-polarization $e_1^2(n)$, for both the proposed constrained CMA polarization demultiplexer and the conventional one. Curves have been smoothed by moving averaging for illustration purposes. The theoretically observed three-fold increase in convergence speed is hereby qualitatively confirmed, although the absolute time scales are different compared to Fig. 4(b).

V. Summary

This article addressed the optimal design, using the maximum-likelihood criterion, of polarization-diversity coherent optical receivers, for the detection of orthogonal polarization-multiplexed optical signals in the absence of PMD and PDL. It was shown that a zero-forcing linear receiver, performing polarization demultiplexing and individual demodulation of the demultiplexed tributaries, yields, under certain conditions, optimal performance. We showed that polarization demultiplexing can be performed using a lattice adaptive filter with four complex, mutually-dependent taps in the absence of PMD and PDL. The taps can be expressed as a function of only two, independently-controlled real parameters. For their estimation, we used the CMA. We studied, by simulation, the performance of the proposed polarization demultiplexer in coherent optical communication systems, using PDM QPSK signals. We showed that it was, by far, superior, in terms of convergence speed, compared to a conventional, CMA-based polarization demultiplexer [10]. Apart from this difference, both polarization demultiplexers exhibited almost identical performance. Nevertheless, a salient feature of the proposed constrained CMA polarization demultiplexer is that it always achieves convergence, unlike its conventional counterpart, which occasionally gets caught in singularities. Both the proposed constrained CMA polarization demultiplexer and the conventional one suffer from output permutation and rotation ambiguities, but these problems can be remedied by other DSP techniques. Finally, we experimentally tested the tolerance of the proposed constrained CMA polarization demultiplexer to realistic imperfections of polarization-diversity coherent optical receivers.

Appendix A

In this Appendix, we model the polarization- and phase-diversity coherent receiver and derive the matrix equation (5). Our analysis is similar to the one by [2], [3] and is reported here for completeness. Relationship (1) is used as a starting point.

The polarization-diversity coherent optical receiver splits the received complex electric field vector $\mathbf{E}(t)$ into its x- and y-polarization components $\mathbf{E}_x(t)$ and $\mathbf{E}_y(t)$ using a PBS with principal axes $|x\rangle, |y\rangle$

$$\begin{aligned}\mathbf{E}_x(t) &= [E_s(t)\langle x|e_s(t)\rangle + E_p(t)\langle x|e_p(t)\rangle]|x\rangle \\ \mathbf{E}_y(t) &= [E_s(t)\langle y|e_s(t)\rangle + E_p(t)\langle y|e_p(t)\rangle]|y\rangle. \end{aligned} \tag{31}$$

The SOP at the output of the LO is assumed to be linear 45°. After a PBS with principal axes $|x\rangle, |y\rangle$, at the input of each hybrid, the electric field of the LO can be written, in equivalent baseband notation, as

$$\mathbf{E}_{\mathrm{lo},k}(t) = \frac{1}{\sqrt{2}} E_{\mathrm{lo}}(t)|k\rangle \tag{32}$$

where $k = x, y$ and $E_{\mathrm{lo}}(t)$ is the complex envelope of the local oscillator given by

$$E_{\mathrm{lo}}(t) = \sqrt{2P_{\mathrm{lo}}} e^{j\omega_{\mathrm{lo}}t + j\theta_{\mathrm{lo}}(t)}. \tag{33}$$

In (33), P_{lo} is the average optical power of the local oscillator, ω_{lo} is the local oscillator's angular frequency offset from the channel's nominal frequency, and $\theta_{\mathrm{lo}}(t)$ is the phase noise of the local oscillator.

At the four outputs of an ideal, lossless, polarization-independent, 90° optical hybrid, we obtain (omitting the time dependence, for brevity) [43]

$$\begin{aligned}\mathbf{E}_{k1} &= \tfrac{1}{2}(\mathbf{E}_k - \mathbf{E}_{\mathrm{lo},k}) \quad & \mathbf{E}_{k3} &= \tfrac{j}{2}(\mathbf{E}_k + j\mathbf{E}_{\mathrm{lo},k}) \\ \mathbf{E}_{k2} &= \tfrac{j}{2}(\mathbf{E}_k + \mathbf{E}_{\mathrm{lo},k}) \quad & \mathbf{E}_{k4} &= \tfrac{j}{2}(j\mathbf{E}_k + \mathbf{E}_{\mathrm{lo},k})\end{aligned} \tag{34}$$

where $k = x, y$.

Neglecting, for the moment, the contribution of shot and thermal noises, which will be taken into account later on, the photocurrents, at the output of the photodiodes, are given by

$$i_{kl} = \frac{R_{kl}}{2} \mathbf{E}_{kl}^\dagger \mathbf{E}_{kl} \tag{35}$$

where $k = x, y, l = 1, \ldots, 4$. In (35), R_{kl} are the responsivities of the photodiodes and the dagger denotes the adjoint, i.e., conjugate transpose, matrix.

By substitution of (34) into (35), we obtain

$$\begin{aligned}i_{k1} &= \frac{R_{k1}}{8}\left\{\|\mathbf{E}_k\|^2 + \|\mathbf{E}_{\mathrm{lo},k}\|^2 - 2\Re\left[\mathbf{E}_{\mathrm{lo},k}^\dagger \mathbf{E}_k\right]\right\} \\ i_{k2} &= \frac{R_{k2}}{8}\left\{\|\mathbf{E}_k\|^2 + \|\mathbf{E}_{\mathrm{lo},k}\|^2 + 2\Re\left[\mathbf{E}_{\mathrm{lo},k}^\dagger \mathbf{E}_k\right]\right\} \\ i_{k3} &= \frac{R_{k3}}{8}\left\{\|\mathbf{E}_k\|^2 + \|\mathbf{E}_{\mathrm{lo},k}\|^2 + 2\Im\left[\mathbf{E}_{\mathrm{lo},k}^\dagger \mathbf{E}_k\right]\right\} \\ i_{k4} &= \frac{R_{k4}}{8}\left\{\|\mathbf{E}_k\|^2 + \|\mathbf{E}_{\mathrm{lo},k}\|^2 - 2\Im\left[\mathbf{E}_{\mathrm{lo},k}^\dagger \mathbf{E}_k\right]\right\}\end{aligned} \tag{36}$$

where $k = x, y$, and $\Re\{\cdot\}, \Im\{\cdot\}$ denote the real and imaginary parts, respectively.

In the case of identical photodiodes with responsivity equal to R, at the outputs of the balanced receivers, we obtain

$$\begin{aligned}i_{\mathrm{tot},k1} &= i_{k2} - i_{k1} = \frac{R}{2}\Re\left[\mathbf{E}_{\mathrm{lo},k}^\dagger \mathbf{E}_k\right] \\ i_{\mathrm{tot},k2} &= i_{k3} - i_{k4} = \frac{R}{2}\Im\left[\mathbf{E}_{\mathrm{lo},k}^\dagger \mathbf{E}_k\right]\end{aligned} \tag{37}$$

where $k = x, y$.

After sampling at integer multiples of the symbol period T_s, we can form the discrete-time complex photocurrents via complex addition

$$\begin{aligned}i_{\mathrm{tot},k}(n) &= i_{\mathrm{tot},k1}(n) + j i_{\mathrm{tot},k2}(n) \\ &= \frac{R}{2}\mathbf{E}_{\mathrm{lo},k}^\dagger(n)\mathbf{E}_k(n)\end{aligned} \tag{38}$$

where $k = x, y$.

By substitution of (31) and (32) into (38), we obtain

$$i_{\mathrm{tot},k}(n) = \frac{RE_{\mathrm{lo}}^*(n)}{2\sqrt{2}}[E_s(n)\langle k|e_s(n)\rangle + E_p(n)\langle k|e_p(n)\rangle] \tag{39}$$

where $k = x, y$ and the superscript $*$ denotes complex conjugation.

We can define the column-vectors of the photocurrents $\mathbf{X}(n)$, the modulating signals $\mathbf{U}(n)$, and the total photocurrent noise $\mathbf{N}(n)$ as

$$\begin{aligned}\mathbf{X}(n) &= [i_{\mathrm{tot},x}(n) \quad i_{\mathrm{tot},y}(n)]^{\mathrm{T}} \\ \mathbf{U}(n) &= [u_1(n) \quad u_2(n)]^{\mathrm{T}} \\ \mathbf{N}(n) &= [n_1(n) \quad n_2(n)]^{\mathrm{T}}\end{aligned} \tag{40}$$

where $n_1(n), n_2(n)$ include the contribution of ASE, shot, and thermal noises, and the superscript T denotes transposition.

From (2), (39), we observe that the fiber-induced polarization rotation and the optoelectronic conversion, at the polarization- and phase-diversity coherent optical receiver, can be described as a matrix equation

$$\mathbf{X}(n) = A\mathbf{H}(n)\mathbf{U}(n) + \mathbf{N}(n) \tag{41}$$

where A is a multiplicative factor

$$A = \frac{R}{\sqrt{2}}\sqrt{P_s P_{\mathrm{lo}}} e^{j\Delta\omega_{\mathrm{IF}} n T_s + j\Delta\theta(n)}. \tag{42}$$

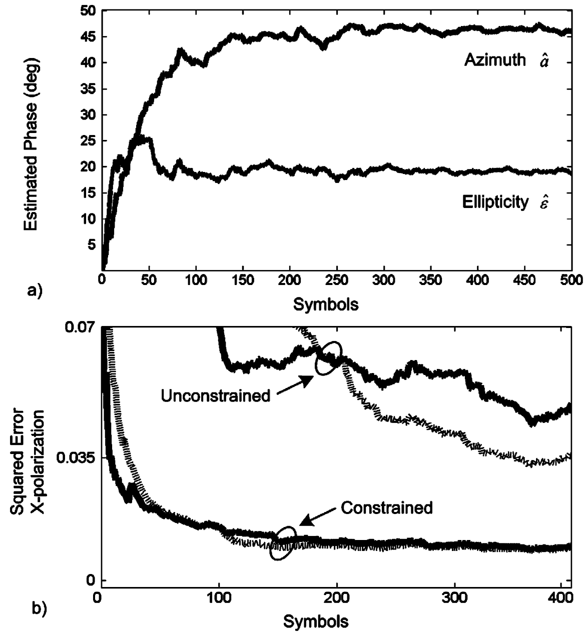

Fig. 8. (a) Time evolution of the estimates of the s-SOP angles $\hat{\alpha}$ and $\hat{\varepsilon}$ computed by the proposed constrained polarization demultiplexer. For visualization purposes, we have removed the unit cell angle restrictions. (b) Time evolution of the squared error function for the x-polarization $e_1^2(n)$, for the proposed polarization demultiplexer and the conventional CMA-based one. Curves are smoothed by two hundred-points moving averaging. (Symbols: Dotted lines: Initial convergence step $\mu = 0.05$, Solid lines: Initial convergence step $\mu = 0.5$).

In the above, $\Delta\omega_{\rm IF} = \omega_s - \omega_{\rm lo}$ is the IF offset and $\Delta\theta(n) = \theta_s(n) - \theta_{\rm lo}(n)$ is the total laser phase noise.

In (41), we also defined the transfer function of the transmission channel as a 2×2 matrix

$$\mathbf{H}(n) = \begin{bmatrix} \langle x|e_s(n)\rangle & \langle x|e_p(n)\rangle \\ \langle y|e_s(n)\rangle & \langle y|e_p(n)\rangle \end{bmatrix}. \tag{43}$$

It is straightforward to verify that $\mathbf{H}(n)$ belongs to the special unitary group SU(2), i.e., $\mathbf{H}(n)\mathbf{H}^\dagger(n) = \mathbf{H}^\dagger(n)\mathbf{H}(n) = \mathbf{I}$ and $\det[\mathbf{H}(n)] = 1$, where $\mathbf{I}$ denotes the 2×2 unit matrix and the operator $\det[\cdot]$ denotes the determinant of a matrix.

For ideal NRZ QPSK modulation, the sampled modulating signals $u_k(n), k = 1, 2$ take equiprobable discrete complex values $u_k(n) = e^{j\varphi_k(n)}, k = 1, 2$, where $\varphi_k(n) \in \{\pm\pi/4, \pm 3\pi/4\}$.

Consequently, the mean and covariance matrices of the modulating signals $\mathbf{U}(n)$ are

$$\boldsymbol{\mu}_{\mathbf{U}} = \mathrm{E}\{\mathbf{U}(n)\} = \mathbf{0} \tag{44}$$

and

$$\mathbf{K}_{\mathbf{U}} = \mathrm{E}\{\mathbf{U}(n)\mathbf{U}^\dagger(n)\} = \mathbf{I} \tag{45}$$

respectively, where $\mathrm{E}\{\cdot\}$ denotes expectation.

From the properties of ASE, shot, and thermal noises, the mean and covariance matrices of the total photocurrent noise $\mathbf{N}(n)$ are given by

$$\boldsymbol{\mu} = \mathrm{E}\{\mathbf{N}(n)\} = \mathbf{0} \tag{46}$$

and

$$\mathbf{K} = \mathrm{E}\{\mathbf{N}(n)\mathbf{N}^\dagger(n)\} = \sigma^2\mathbf{I} \tag{47}$$

respectively, where σ^2 is the variance of the total photocurrent noise components at the two receiver branches.

In the general case, when the modulating signal amplitudes and the noise variances at the two receiver branches are not identical, (41) can be rewritten, with some abuse of notation, as

$$\mathbf{X}(n) = \mathbf{H}(n)\mathbf{U}(n) + \mathbf{N}(n) \tag{48}$$

where we scale $\mathbf{U}(n), \mathbf{N}(n)$, so that

$$\mathbf{K}_{\mathbf{U}} = \mathrm{diag}\{|A_1|^2, |A_2|^2\} \tag{49}$$

and

$$\mathbf{K} = \mathrm{diag}\left\{\sigma_1^2, \sigma_2^2\right\} \tag{50}$$

where $\mathrm{diag}\{\cdot\}$ denotes diagonal matrix, A_1, A_2 are the modulating signal amplitudes, and σ_1^2, σ_2^2 are the total photocurrent noise variances at the two branches of the polarization diversity receiver, respectively.

Appendix B

This Appendix provides a detailed derivation of the optimal receiver for joint detection of PDM QPSK signals transmitted over a memoryless, discrete-time, two-input two-output (TITO) linear channel, based on the maximum-likelihood criterion [25]–[27].

As a starting point for the derivation, we use the matrix equation (48)

$$\mathbf{X}(n) = \mathbf{H}(n)\mathbf{U}(n) + \mathbf{N}(n). \tag{51}$$

The optimal maximum-likelihood receiver estimates the vector $\hat{\mathbf{U}}(n)$ by maximizing the metric [26],

$$\hat{\mathbf{U}} = \arg\max_{\mathbf{U}\in\mathcal{U}} P(\mathbf{X}|\mathbf{U}) \tag{52}$$

where $P(\mathbf{X}|\mathbf{U})$ is the conditional probability of the observed vector $\mathbf{X}$ given that the transmitted vector was $\mathbf{U}$ and $\mathcal{U}$ is the joint, complex-symbol alphabet. We have dropped the time dependence of all matrices in order to avoid clutter.

Using (51), the above relationship can be rewritten as

$$\hat{\mathbf{U}} = \arg\max_{\mathbf{U}\in\mathcal{U}} P_{\mathbf{N}}(\mathbf{X} - \mathbf{H}\mathbf{U}) \tag{53}$$

where $P_{\mathbf{N}}(\mathbf{N})$ is the joint probability density function (pdf) of the complex Gaussian random variables n_1, n_2, with zero mean given by (46) and covariance matrix $\mathbf{K}$ given by (50) [25], [26]

$$P_{\mathbf{N}}(\mathbf{N}) = \frac{1}{2\pi\sqrt{\det[\mathbf{K}]}} \exp\left[-\frac{1}{2}\|\mathbf{K}^{-1/2}\mathbf{N}\|^2\right]. \tag{54}$$

Since $P_{\mathbf{N}}(\mathbf{N})$ is a decreasing function of the argument of the exponential, the maximization in (53) is equivalent to minimizing the Euclidean distance

$$D = \|\mathbf{K}^{-1/2}(\mathbf{X} - \mathbf{H}\mathbf{U})\|^2. \tag{55}$$

Expanding the distance metric yields

$$D = \mathbf{X}^\dagger\mathbf{K}^{-1}\mathbf{X} + \mathbf{U}^\dagger\mathbf{H}^\dagger\mathbf{K}^{-1}\mathbf{H}\mathbf{U} - 2\Re\{\mathbf{U}^\dagger\mathbf{H}^\dagger\mathbf{K}^{-1}\mathbf{X}\}. \tag{56}$$

The first term is independent of $\mathbf{U}$ and, can be ignored. Thus, a sufficient statistic for estimating the transmitted symbols is the following

$$\hat{\mathbf{U}} = \arg\min_{\mathbf{U}\in\mathcal{U}}\{\mathbf{U}^\dagger\mathbf{H}^\dagger\mathbf{K}^{-1}\mathbf{H}\mathbf{U} - 2\Re\{\mathbf{U}^\dagger\mathbf{H}^\dagger\mathbf{K}^{-1}\mathbf{X}\}\}. \tag{57}$$

In the special case when the noises at the two branches of the polarization diversity receiver have the same variance σ^2 (i.e., the photocurrent noise is spatially white), the covariance matrix of the total photocurrent noise is given by (47). By substitution of (47) into (56), and taking into account that $\mathbf{H}$ is unitary, the distance metric is simplified

$$D = \sigma^{-2}\{\|\mathbf{X}\|^2 + \|\mathbf{U}\|^2 - 2\Re\{\mathbf{U}^\dagger\mathbf{H}^\dagger\mathbf{X}\}\} \tag{58}$$

or, equivalently,

$$D = \sigma^{-2}\{\|\mathbf{Y}-\mathbf{U}\|^2 - \|\mathbf{Y}\|^2 + \|\mathbf{X}\|^2\} \tag{59}$$

where we have defined

$$\mathbf{Y} = \mathbf{H}^\dagger\mathbf{X}. \tag{60}$$

Since only the first term of (59) depends on the candidate symbol vector $\mathbf{U}$, the optimal receiver simply needs to minimize the metric

$$\hat{\mathbf{U}} = \arg\min_{\mathbf{U}\in\mathcal{U}}\|\mathbf{Y}-\mathbf{U}\|^2. \tag{61}$$

We observe that it is sufficient to estimate each element $\hat{u}_k$ of $\hat{\mathbf{U}}$ individually, at each branch of the polarization-diversity receiver, using the metric

$$\hat{u}_k = \arg\min_{u_k\in\mathcal{A}}|y_k - u_k|^2, k = x, y \tag{62}$$

where $\mathcal{A}$ is the complex symbol alphabet of each polarization-multiplexed tributary. Furthermore, expanding the above relationship into its quadrature components, it follows that it is sufficient to choose the real and imaginary parts of each symbol independently at each branch of the phase-diversity receiver, in order to minimize the metric

$$\hat{u}_{k,l} = \arg\min_{u_{k,l}\in\Re\{\mathcal{A}\}}(y_{k,l} - u_{k,l})^2, k = x, y, l = r, i. \tag{63}$$

To summarize, the above analysis indicates that the optimal receiver is reduced to a zero-forcing linear receiver [26], [27]. The latter first uses an electronic polarization demultiplexer with transfer matrix $\mathbf{W}(n) = \mathbf{H}^\dagger(n)$. Subsequently, the in-phase and quadrature components of the symbols, at the two outputs of the polarization demultiplexer, can be independently detected, by comparing each individual quadrature component to a zero threshold.

In conclusion, the proposed constrained CMA polarization demultiplexer is optimal, when the photocurrent noise is spatially white. In this case, the joint maximum-likelihood receiver and the zero-forcing linear receiver are equivalent.

APPENDIX C

In this Appendix, we examine how the equivalent channel formalism can be modified, in order to accommodate PDL and polarization imbalance at the polarization-diversity receiver. It is shown that both effects result in a perturbation transfer matrix that creates additional cross-polarization interference.

Consider a partial polarizer with eigenaxes in Jones space denoted by the vectors $|p_+\rangle, |p_-\rangle$. The corresponding eigenaxes in Stokes space are denoted by $\pm\hat{p}$. The transmittances associated with these eigenaxes are $T_{\max}, T_{\min}$, respectively. Both the eigenaxes and the transmittances are independent of frequency. PDL is defined in dB units as $\rho_{\mathrm{dB}} \triangleq 10\log(T_{\max}/T_{\min})$.

The Jones matrix of the partial polarizer, in the absence of birefringence, is written as [44], [45]

$$\mathbf{M} = \sqrt{T_{\max}}|p_+\rangle\langle p_+| + \sqrt{T_{\min}}|p_-\rangle\langle p_-|. \tag{64}$$

Using the expansion of the projection coefficient in Pauli spin matrices (relation [3.9] of [35]), the Jones matrix of the partial polarizer can be expressed in the alternative form

$$\mathbf{M} = a(\mathbf{I} + \Delta a\hat{p}\vec{\boldsymbol{\sigma}}) \tag{65}$$

where $\mathbf{I}$ is the 2×2 identity matrix, $\vec{\boldsymbol{\sigma}}$ is the Pauli spin vector [35], and we defined the average amplitude attenuation coefficient a and the differential amplitude attenuation coefficient Δa as

$$a = (\sqrt{T_{\max}} + \sqrt{T_{\min}})/2 \tag{66}$$
$$\Delta a = (\sqrt{T_{\max}} - \sqrt{T_{\min}})/(\sqrt{T_{\max}} + \sqrt{T_{\min}}). \tag{67}$$

The electric field of the optical PDM QPSK signal at the input of the PDL is given by relationship (1). The electric field of the optical PDM QPSK signal at the output of the partial polarizer can be written, in the absence of noise, as

$$\mathbf{E}_{\mathrm{out}}(t) = \mathbf{M}\mathbf{E}(t). \tag{68}$$

Then (41) can be rewritten, in the absence of noise, as

$$\mathbf{X}(n) = a[\mathbf{H}(n) + \Delta a\boldsymbol{\Delta}\mathbf{H}(n)]\mathbf{U}(n) \tag{69}$$

where we defined the perturbation transfer matrix

$$\boldsymbol{\Delta}\mathbf{H}(n) = \begin{bmatrix} \hat{p}\langle x|\vec{\boldsymbol{\sigma}}|e_s(n)\rangle & \hat{p}\langle x|\vec{\boldsymbol{\sigma}}|e_p(n)\rangle \\ \hat{p}\langle y|\vec{\boldsymbol{\sigma}}|e_s(n)\rangle & \hat{p}\langle y|\vec{\boldsymbol{\sigma}}|e_p(n)\rangle \end{bmatrix}. \tag{70}$$

As a sanity check, we observe that the second term of (69) becomes negligible for $\Delta a \to 0$, i.e., when $T_{\max} = T_{\min}$.

The zero-forcing polarization demultiplexer must calculate an estimate $\hat{\mathbf{H}}(n)$ of the total channel transfer matrix $a[\mathbf{H}(n) + \Delta a\boldsymbol{\Delta}\mathbf{H}(n)]$ and set $\mathbf{W}(n) = \hat{\mathbf{H}}^\dagger(n)$. The independent parameters of $\hat{\mathbf{H}}(n)$ are the azimuth and the ellipticity of the input SOP $|e_s(n)\rangle$, the azimuth and the ellipticity of the PDL eigenaxis $\hat{p}$, and the PDL parameters $a, \Delta a$.

In summary, the description of the perturbation matrix would require four additional control parameters. The total channel matrix requires six independent control parameters. Still, the current approach is advantageous compared to the CMA, which requires control of eight independent real parameters for polarization demultiplexing. Increasing the dimensionality of the independent parameter space will obviously slow down the search for the global optimum of the transfer function. On the other hand, conventional CMA-based demultiplexers are not only slow but also suffer from singularities. Their large number of independent parameters increases the number of degrees of freedom and the effects that can be accommodated, at the expense of execution speed and perhaps convergence altogether.

It is instructive to estimate the impact of PDL and of polarization imbalance on the performance of the proposed constrained polarization demultiplexer, which possesses a unitary transfer matrix. Since the addition of two unitary matrices is not a unitary matrix, the total channel transfer matrix $a[\mathbf{H}(n)+\Delta a\boldsymbol{\Delta}\mathbf{H}(n)]$ is not unitary. As a result, the product of the total channel transfer matrix with the unitary transfer matrix of the proposed constrained polarization demultiplexer is not a unit matrix. Consequently, the transmitted constellations cannot be not fully detangled.

The output of the polarization demultiplexer can be written, in the absence of noise, as

$$\mathbf{Y}(n) = a[\mathbf{H}_{\mathrm{eq}}(n) + \Delta a\boldsymbol{\Delta}\mathbf{H}_{\mathrm{eq}}(n)]\mathbf{U}(n) \tag{71}$$

where $\mathbf{H}_{\mathrm{eq}}(n), \boldsymbol{\Delta}\mathbf{H}_{\mathrm{eq}}(n)$ are the transfer matrices of the channel and the perturbation after polarization demultiplexing, which can be written

$$\begin{aligned}\mathbf{H}_{\mathrm{eq}}(n) &= \mathbf{W}\mathbf{H}(n)\\ \boldsymbol{\Delta}\mathbf{H}_{\mathrm{eq}}(n) &= \mathbf{W}\Delta\mathbf{H}(n)\end{aligned} \tag{72}$$

The action of the polarization demultiplexer is to rotate the principal axes of the receiver PBS $|x\rangle, |y\rangle$ in order to match the Jones vectors of the received polarization tributaries. It is straightforward to show that

$$\mathbf{H}_{\mathrm{eq}}(n) = \begin{bmatrix} \langle \hat{e}_s | e_s(n)\rangle & \langle \hat{e}_s | e_p(n)\rangle \\ \langle \hat{e}_p | e_s(n)\rangle & \langle \hat{e}_p | e_p(n)\rangle \end{bmatrix} \tag{73}$$

and

$$\boldsymbol{\Delta}\mathbf{H}_{\mathrm{eq}}(n) = \begin{bmatrix} \hat{p}\langle \hat{e}_s(n)|\vec{\boldsymbol{\sigma}}|e_s(n)\rangle & \hat{p}\langle \hat{e}_s(n)|\vec{\boldsymbol{\sigma}}|e_p(n)\rangle \\ \hat{p}\langle \hat{e}_p(n)|\vec{\boldsymbol{\sigma}}|e_s(n)\rangle & \hat{p}\langle \hat{e}_p(n)|\vec{\boldsymbol{\sigma}}|e_p(n)\rangle \end{bmatrix} \tag{74}$$

where $|\hat{e}_s(t)\rangle, |\hat{e}_p(t)\rangle$ are estimates of the Jones vectors of the received polarization tributaries.

Assuming that the presence of the perturbation transfer matrix $\boldsymbol{\Delta}\mathbf{H}(n)$ does not drastically change the estimate of CMA, we can postulate that after convergence, $|\hat{e}_s(t)\rangle \cong |e_s(t)\rangle, |\hat{e}_p(t)\rangle \cong |e_p(t)\rangle$. Since $\langle e_s(n)|\vec{\boldsymbol{\sigma}}|e_s(n)\rangle = \hat{s}$ and $\langle e_p(n)|\vec{\boldsymbol{\sigma}}|e_s(n)\rangle = \hat{s}_2 + j\hat{s}_3$ [35], where $\hat{s}$ is the Stokes vector of the s-SOP and $\hat{s}, \hat{s}_2, \hat{s}_3$ are a right handed orthogonal set in Stokes space,

$$\mathbf{Y}(n) = \begin{bmatrix} a+\Delta a(\hat{p}\hat{s}) & 0 \\ 0 & a-\Delta a(\hat{p}\hat{s}) \end{bmatrix}\mathbf{U}(n) + \Delta\mathbf{Y}(\Delta a) \tag{75}$$

where

$$\Delta\mathbf{Y}(\Delta a) = \Delta a \begin{bmatrix} 0 & \hat{p}(\hat{s}_2 - j\hat{s}_3) \\ \hat{p}(\hat{s}_2 + j\hat{s}_3) & 0 \end{bmatrix}\mathbf{U}(n) \tag{76}$$

The first term in (75) indicates that the polarization tributaries can be essentially recovered but they are distorted (i.e., the constellations have different size, differing by $2\Delta a(\hat{p}\hat{s})$ in radius). This distortion is an artifact induced by the assumption of the unitarity of the polarization demultiplexer's transfer matrix. The second term in (75) indicates that there is a residual cross-polarization interference due to the anti-diagonal transfer matrix in (76). This interference term is relatively small since its magnitude depends on the differential amplitude attenuation coefficient Δa.

Finally, it is worth noting that the polarization imbalance at the polarization-diversity receiver is the electronic domain equivalent of using a partial polarizer in the optical domain. Its impact can be accounted for by substituting the polarizer's transmission parameters $\sqrt{T_{\max}}, \sqrt{T_{\min}}$ with the photodiode responsivities and the polarizer eigenaxis in Stokes space by $\hat{p} = \hat{x}$.

ACKNOWLEDGMENT

The authors would like to thank Prof. G. Moustakides and Mr. V. U. Prabhu, Department of Electrical and Computer Engineering, University of Patras, Greece, for stimulating discussions on estimation theory, and the anonymous reviewers for their helpful comments and suggestions.

REFERENCES

[1] H. Sun, K.-T. Wu, and K. Roberts, "Real-time measurements of a 40 Gb/s coherent system," *Opt. Exp.*, vol. 16, no. 2, pp. 873–879, Jan. 2008.

[2] E. Ip, A. P. T. Lau, D. J. F. Barros, and J. M. Kahn, "Coherent detection in optical fiber systems," *Opt. Exp.*, vol. 16, no. 2, pp. 753–791, Jan. 2008.

[3] K. Kikuchi, "Coherent optical communication systems," in *Optical Fiber Telecommunications*, I. P. Kaminow, T. Li, and A. E. Willner, Eds. San Diego, CA: Academic, 2008, pp. 95–129, vol. VB, ch. 3.

[4] M. Tseytlin, O. Ritterbush, and A. Salamon, "Digital, endless polarization control for polarization multiplexed fiber-optic communications," in *Proc. OFC 2003*, Atlanta, GA, Mar. 2003, paper MF83.

[5] S. Calabro, T. Dullweber, E. Gottwald, N. Hecker-Denschlag, E. Mullner, B. Opitz, G. Sebald, E. Schmidt, B. Spinnler, C. Weiske, and H. Zech, "An electrical polarization-state controller and demultiplexer for polarization multiplexed optical signals," in *Proc. 29th Eur. Conf. Optical Communication (ECOC)*, Rimini, Italy, Sept. 2003, vol. 4, pp. 950–951.

[6] R. Noé, "Phase noise tolerant synchronous QPSK/BPSK baseband-type intradyne receiver concept with feedforward carrier recovery," *J. Lightw. Technol.*, vol. 23, no. 2, pp. 802–808, Feb. 2005.

[7] R. Noé, "PLL-free synchronous QPSK polarization multiplex/diversity receiver concept with digital I & Q baseband processing," *IEEE Photon. Technol. Lett.*, vol. 17, no. 4, pp. 887–889, Apr. 2005.

[8] Y. Han and G. Li, "Coherent optical communication using polarization multiple-input-multiple-output," *Opt. Exp.*, vol. 13, no. 19, pp. 7527–7534, Sept. 2005.

[9] M. T. Core, "Cross polarization interference cancellation for fiber optic systems," *J. Lightw. Technol.*, vol. 24, no. 1, pp. 305–312, Jan. 2006.

[10] A. Leven, N. Kaneda, and Y.-K. Chen, "A real-time CMA-based 10 Gb/s polarization demultiplexing coherent receiver implemented in an FPGA," in *Proc. OFC/NFOEC 2008*, Anaheim, CA, Feb. 2008, paper OTuO2.

[11] K. Kikuchi, "Polarization-demultiplexing algorithm in the digital coherent receiver," in *Proc. IEEE LEOS Summer Topicals*, Acapulco, Mexico, Jul. 2008, paper MC2.2.

[12] H. Zhang, Z. Tao, L. Liu, S. Oda, T. Hoshida, and J. C. Rasmussen, "Polarization demultiplexing based on independent component analysis in optical coherent receivers," in *Proc. 34th Eur. Conf. Optical Communication (ECOC)*, Brussels, Belgium, Sept. 2008, paper Mo.3.D.5.

[13] D. E. Crivelli, H. S. Carter, and M. R. Hueda, "Adaptive digital equalization in the presence of chromatic dispersion, PMD, and phase noise in coherent fiber optic systems," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, Dallas, TX, Nov. 2004, vol. 4, pp. 2545–2551.
[14] E. Ip and J. M. Kahn, "Digital equalization of chromatic dispersion and polarization mode dispersion," *J. Lightw. Technol.*, vol. 25, no. 8, pp. 2033–2043, Aug. 2007.
[15] S. J. Savory, "Digital filters for coherent optical receivers," *Opt. Exp.*, vol. 16, no. 2, pp. 804–817, Jan. 2008.
[16] C. Laperle, B. Villeneuve, Z. Zhan, D. McGhan, H. Sun, and M. O'Sullivan, "WDM performance and PMD tolerance of a coherent 40 Gbit/s dual-polarization QPSK transceiver," *J. Lightw. Technol.*, vol. 26, no. 1, pp. 168–175, Jan. 2008.
[17] C. R. S. Fludger, T. Duthel, D. van den Borne, C. Schulien, E. D. Schmidt, T. Wuth, J. Geyer, E. De Man, G. D. Khoe, and H. de Waardt, "Coherent equalization and POLMUX-RZ-DQPSK for robust 100-GE transmission," *J. Lightw. Technol.*, vol. 26, no. 1, pp. 64–72, Jan. 2008.
[18] D. Godard, "Self-recovering equalization and carrier tracking in two dimensional data communication systems," *IEEE Trans. Comm.*, vol. 28, no. 11, pp. 1867–1875, Nov. 1980.
[19] J. R. Treichler and M. G. Larimore, "New processing techniques based on the constant modulus adaptive algorithm," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-33, no. 2, pp. 420–431, Apr. 1985.
[20] C. R. Johnson Jr., P. Schniter, T. J. Edres, J. D. Behm, D. R. Brown, and R. A. Casas, "Blind equalization using the constant modulus criterion: A review," *Proc. IEEE*, vol. 86, no. 10, pp. 1927–1950, Oct. 1998.
[21] S. Haykin, *Adaptive Filter Theory*, 3rd ed. Englewood Cliffs, NJ: Prentice-Hall, 1996.
[22] J. Renaudier, G. Charlet, M. Salsi, O. B. Pardo, H. Mardoyan, P. Tran, and S. Bigo, "Linear fiber impairments mitigation of 40-Gbit/s polarization-multiplexed QPSK by digital processing in a coherent receiver," *J. Lightw. Technol.*, vol. 26, no. 1, pp. 36–42, Jan. 2008.
[23] X. Zhou, J. Yu, D. Qian, T. Wang, G. Zhang, and P. Magill, "8×114 Gb/s, 25-GHz-spaced, polmux-RZ-8PSK transmission over 640 km of SSMF employing digital coherent detection and EDFA-only amplification," in *Proc. OFC 2008*, San Diego, CA, Feb. 2008, paper PDP6.
[24] P. J. Winzer and A. J. Gnauck, "112-Gb/s polarization-multiplexed 16-QAM on a 25-GHz WDM grid," in *Proc. 34th Eur. Conf. Optical Communication (ECOC)*, Brussels, Belgium, Sept. 2008, paper Th.3.E.5.
[25] J. Proakis, *Digital Communications*, 4th ed. New York: McGraw-Hill, 2000, pp. 177–178.
[26] J. R. Barry, E. A. Lee, and D. G. Messerschmitt, *Digital Communication*, 3rd ed. Norwell, MA: Kluwer, 2003.
[27] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005, Section 3.3.3.
[28] H. Sun and K.-T. Wu, "Method for Quadrature Phase Angle Correction in a Coherent Receiver of a Dual-Polarization Optical Transport System," U.S. Patent 6 917 031, Jul. 12, 2005.
[29] I. Roudas, M. Sauer, J. Hurley, Y. Mauro, and S. Raghavan, "Compensation of coherent DQPSK receiver imperfections," in *Proc. IEEE LEOS Summer Topicals*, Portland, OR, July 2007, paper MA3.4.
[30] I. Fatadin, S. Savory, and D. Ives, "Compensation of quadrature imbalance in an optical QPSK coherent receiver," *IEEE Photon. Technol. Lett.*, vol. 20, no. 20, pp. 1733–1735, Oct. 2008.
[31] C. S. Petrou, A. Vgenis, A. Kiourti, I. Roudas, J. Hurley, M. Sauer, J. Downie, Y. Mauro, and S. Raghavan, "Impact of transmitter and receiver imperfections on the performance of coherent optical QPSK communication systems," in *Proc. LEOS 2008*, Newport Beach, CA, Nov. 2008, paper TuFF3.
[32] H. Meyr, M. Moeneclaey, and S. Fechtel, *Digital Communication Receivers: Channel Estimation and Signal Processing*. New York: Wiley, 1998, ch. 8.
[33] M. Morelli and U. Mengali, "Feedforward frequency estimation for PSK: A tutorial review," *Eur. Trans. Telecommun.*, vol. 9, no. 2, pp. 103–116, Mar./Apr. 1998.
[34] A. Viterbi and A. Viterbi, "Nonlinear estimation of PSK-modulated carrier phase with application to burst digital transmission," *IEEE Trans. Inf. Theory*, vol. IT-29, no. 4, pp. 543–551, Jul. 1983.
[35] J. P. Gordon and H. Kogelnik, "PMD fundamentals: Polarization mode dispersion in optical fibers," *PNAS*, vol. 97, no. 9, pp. 4541–4550, Apr. 2000.
[36] S. Huard, *Polarization of Light*. New York, NY: Wiley, 1997.
[37] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice Hall, 1993.
[38] E. Karami, "Tracking performance of least squares MIMO channel estimation algorithm," *IEEE Trans. Commun.*, vol. 55, no. 11, pp. 2201–2209, Nov. 2007.
[39] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1991, ch. 5.
[40] L. G. Kazovsky, L. Curtis, W. C. Young, and N. K. Cheung, "All-fiber 90° optical hybrid for coherent communications," *Appl. Opt.*, vol. 26, no. 3, pp. 437–439, Feb. 1987.
[41] X. Zhou, J. Yu, D. Qian, T. Wang, G. Zhang, and P. D. Magill, "High-spectral-efficiency 114-Gb/s transmission using PolMux-RZ-8PSK modulation format and single-ended digital coherent detection technique," *J. Lightw. Technol.*, vol. 27, no. 3, pp. 146–152, Feb. 2009.
[42] M. Valkama, M. Renfors, and V. Koivunen, "Advanced methods for I/Q imbalance compensation in communication receivers," *IEEE Trans. Signal Proc.*, vol. 49, no. 10, pp. 2335–2344, Oct. 1998.
[43] D. Hoffmann, H. Heidrich, G. Wenke, R. Langenhorst, and N. K. Cheung, "Integrated opticseight-port 90° hybrid on $LiNbO_3$," *J. Lightw. Technol.*, vol. 7, no. 5, pp. 794–798, May 1989.
[44] N. Gisin, "Statistics of polarization dependent losses," *Opt. Commun.*, vol. 114, no. 5–6, pp. 399–405, 1995.
[45] I. Roudas and N. Antoniades, "Performance outages in CWDM optical networks due to the polarization-dependent gain of semiconductor optical amplifiers," *IEEE Photon. Technol. Lett.*, vol. 18, no. 1, pp. 48–50, Jan. 2007.

Ioannis Roudas, photograph and biography not available at the time of publication.

Athanasios Vgenis, photograph and biography not available at the time of publication.

Constantinos S. Petrou, photograph and biography not available at the time of publication.

Dimitris Toumpakaris, photograph and biography not available at the time of publication.

Jason Hurley, photograph and biography not available at the time of publication.

Michael Sauer, photograph and biography not available at the time of publication.

John Downie, photograph and biography not available at the time of publication.

Yihong Mauro, photograph and biography not available at the time of publication.

Srikanth Raghavan, photograph and biography not available at the time of publication.