Polarization Tracking in the Presence of PDL and Fast Temporal Drift

(article starts on next page)

Polarization Tracking in the Presence of PDL and Fast Temporal Drift

Mohammad Farsi, *Student Member, IEEE,* Christian Häger, *Member, IEEE,* Magnus Karlsson, *Senior Member, IEEE*, and Erik Agrell, *Fellow, IEEE.*

Abstract—In this paper, we analyze the effectiveness of polarization tracking algorithms in optical transmission systems suffering from fast state of polarization (SOP) rotations and polarization-dependent loss (PDL). While most of the gradient descent (GD)-based algorithms in the literature may require step size adjustment when the channel condition changes, we propose tracking algorithms that can perform similarly or better without parameter tuning. Numerical simulation results show higher robustness of the proposed algorithms to SOP and PDL drift compared to GD-based algorithms, making them promising candidates to be used in aerial fiber links where the SOP can potentially drift rapidly, and therefore becomes challenging to track.

Index Terms—Constant modulus algorithm, hybrid algorithm, least square algorithm, polarization-dependent loss, polarization tracking, state of polarization.

I. INTRODUCTION

ADVANCED digital signal processing (DSP) enables the use of multilevel modulation formats and polarization-division multiplexed transmission to achieve high spectral efficiency [1]. However, spectrally-efficient systems have a lower tolerance to fiber impairments such as state of polarization (SOP) drift and polarization-dependent loss (PDL), which need to be adaptively estimated and tracked at the receiver [2].

In an optical link, the random SOP drift originates from the random variation of environmental conditions (mechanical/thermal stress, weather conditions, splices, etc.) and internal imperfections (asymmetry of the core, manufacturing process errors, etc.) of the fiber. Experimental measurements show that the SOP drift can be extremely slow (days and hours) in buried fibers [3] and very fast (microseconds) in aerial fibers [4]–[6]. For instance, field measurements of an aerial fiber link have revealed that the rotation rate of SOP might get up to 5.1 Mrad/s due to lightning strikes [7]. In the literature, the SOP fluctuation is often modeled as randomly chosen rotations without drift [8], [9] or cyclic/quasi-cyclic deterministic rotation [10]–[12]. Recently, an experimentally

M. Farsi, C. Häger, and E. Agrell are with the Department of Electrical Engineering, Chalmers University of Technology, SE-41296 Gothenburg, Sweden (e-mails: {farsim, christian.haeger, agrell}@chalmers.se).

M. Karlsson is with the Department of Microtechnology and Nanoscience, Chalmers University of Technology, SE-41296 Gothenburg, Sweden (e-mail: magnus.karlsson@chalmers.se).

verified model was proposed in [13], where SOP drift is described by a random walk on the Poincaré sphere.

In a long-haul fiber, several PDL-inducing components (e.g., isolators, amplifiers, multiplexers, and couplers) are in place, and the overall PDL might aggregate to several dBs. Historically, PDL has been modeled as a concatenation of many randomly oriented PDL elements along with the fiber, and its statistics have been extensively studied in [14]–[17], describing the aggregated PDL with a Maxwellian distribution. The impact of the average PDL on the performance of the optical link is studied in [18], [19]. The interplay of PDL with polarization-mode dispersion and Kerr nonlinearity has been studied in [20], and [21], respectively.

The SOP drift accompanied by PDL not only results in a time-varying power imbalance between the two polarizations, but also it causes optical signal-to-noise ratio (SNR) fluctuations and breaks the orthogonality between the two polarizations [22]. While it would be feasible to resolve the static SOP and PDL, the time-varying nature of SOP and the aggregated PDL makes the polarization tracking challenging at the receiver.

The gradient descent (GD)-based algorithms has been widely applied in both wireless and optical fields for adaptive tracking, and the most popular one is the constant modulus algorithm (CMA) [23] and its variants known as modified CMA (MCMA) [24] and multi-modulus algorithm (MMA) [25]. Although CMA is immune to phase noise (PN) and has low computational complexity, it is modulation dependent and suffers from phase ambiguity and the so-called singularity problem [26], where the equalizer converges only on one of two polarizations. The decision-directed least mean squares (DDLMS) algorithm [27], [28] removes the modulation format dependence of the blind algorithms [11], and the phase ambiguity can be resolved by differential encoding/decoding at the expense of a performance degradation [29, Sec. 2.6.1]. Although many have tried to overcome the singularity problem of CMA [30], [31], its solution remains an open research problem. The pilot-aided hybrid algorithms [32], [33] ensure reliable and fast convergence by proper tap initialization of the blind algorithms. For instance, in [33], a pilot-based filter tap initialization using the least mean squares (LMS) [34] algorithm has been proposed to address the singularity problem. For a linear time-invariant system with stationary noise, the LMS solution converges to the well-known Wiener filter solution [35], [36, Ch. 13.2].

The memoryless dual-polarization channel in the presence of SOP drift and a negligible amount of PDL can be described

2

Fig. 1: The DP-PDL model of a long fiber at time k as a concatenation of N, each containing a PDL component $\boldsymbol{\Gamma}_n$ coupled with an SOP drift element $\mathbf{J}_{k,n}$.

by a time-varying complex unitary matrix. Most of the tracking algorithms in the literature (e.g., CMA, MMA, DDLMS, etc.) are designed for a general complex channel matrix without constraining their channel estimation matrix to be unitary. The tracking performance can be improved by explicitly taking the unitary nature of the SOP rotations into account when designing tracking algorithms (even if the overall channel is not exactly unitary). For example, the block-wise decision-directed Kabsch (DD-Kabsch) algorithm was proposed in [37] to address the unitary constraints of the channel. In [38], a unitary GD-based decision-directed joint PN and SOP tracking algorithm was proposed, where the authors have shown a higher polarization drift tolerance than the Kabsch and CMA-based algorithms. We refer to the algorithm in [38] as decision-directed Czegledi (DD-Czegledi).

This paper, which extends the conference paper [39], proposes two polarization tracking algorithms called sliding window Kabsch (SW-Kabsch) and sliding window least squares (SW-LS), described in Sections III-A and III-B1, respectively. When the PDL is negligible, SW-Kabsch shows higher SOP drift tolerance than SW-LS and the GD-based algorithms (e.g., DD-Czegledi, CMA and its variants). However, considering PDL, the channel is no longer unitary, and SW-LS shows the highest SOP drift tolerance. Moreover, unlike GD-based algorithms that require step size adjustment, the proposed algorithms require no further parameter tuning, making them potential candidates for memoryless fast drifting optical systems.

Although the literature is replete with time-invariant PDL models [15], [16], the time-varying nature of PDL is not well-studied. Inspired by the previous work [13], we introduce a channel model called dual-polarization PDL (DP-PDL) (described in Section II) that accounts for memoryless time-varying PDL.

Notation: Column vectors are denoted by underlined letters $\underline{x}$ and matrices by uppercase roman letters X. We use bold-face letters $\mathbf{x}$ for random quantities and the corresponding nonbold letters x for their realizations. Sets are denoted by uppercase calligraphic letters $\mathcal{X}$. The Frobenius norm is denoted by $\|\cdot\|$ and the expectation over random variables is denoted by $\mathbb{E}[\cdot]$. The real zero-mean multivariate Gaussian distribution is denoted by $\mathbf{x} \sim \mathcal{N}(\underline{0}, \Lambda_{\underline{x}})$ and the complex zero-mean circularly symmetric Gaussian distribution of a vector is denoted by $\mathbf{x} \sim \mathcal{CN}(\underline{0}, \Lambda_{\underline{x}})$, where $\Lambda_{\underline{x}}$ is the covariance matrix.

II. SYSTEM MODEL

We consider dual-polarization transmission in the presence of PDL, SOP drift, and amplified spontaneous emission (ASE) noise at the receiver. We also assume that nonlinearities are negligible and chromatic dispersion is compensated. We particularly study the memoryless dual-polarization channel where polarization-mode dispersion is not considered. To model the channel, we take the experimentally verified SOP drift model in [13] and combine it with concatenated PDL elements. This model can be regarded as a spherical analogy of a two-dimensional Wiener process.

The transmitted signal at time k is a 2-dimensional random vector $\underline{\mathbf{s}}_k$, which takes on values from a set $\mathcal{S} = \{\underline{c}_1, \underline{c}_2, ..., \underline{c}_M\}$ of complex, zero-mean, equiprobable constellation points. The vector of received complex samples $\underline{\mathbf{x}}_k$ can be expressed as

$$\underline{\mathbf{x}}_k = \mathbf{H}_k \underline{\mathbf{s}}_k + \underline{\mathbf{z}}_k, \qquad (1)$$

where $\mathbf{H}_k$ is a 2×2 complex channel matrix, and $\underline{\mathbf{z}}_k \sim \mathcal{CN}(0, \sigma_z^2 \mathbf{I}_2)$. The physical model at time k is shown in Fig. 1, where a link with N segments, each consisting of a SOP element $\mathbf{J}_{k,n}$ and a PDL element $\boldsymbol{\Gamma}_n$, where n is the segment index. The DP-PDL channel matrix can be described by a 2×2 complex-valued matrix as

$$\mathbf{H}_k = \boldsymbol{\Gamma}_N \mathbf{J}_{k,N} \cdots \boldsymbol{\Gamma}_1 \mathbf{J}_{k,1} = \prod_{n=1}^{N} \boldsymbol{\Gamma}_n \mathbf{J}_{k,n}, \qquad (2)$$

where $\boldsymbol{\Gamma}_n$ is a 2×2 positive real-valued diagonal matrix modeling the power imbalance induced by PDL. For the sake of simplicity, we assume that a time-invariant deterministic matrix can describe each PDL component as [15]

$$\boldsymbol{\Gamma}_n = \begin{bmatrix} \sqrt{1 + \gamma_n} & 0 \\ 0 & \sqrt{1 - \gamma_n} \end{bmatrix} \qquad (3)$$

where $0 \le \gamma_n \le 1$ is each segment's PDL ratio indicating that in the extreme case, only one active polarization will remain ($\gamma_n = 1$). Moreover, $\mathbf{J}_{k,n}$ is a random 2×2 unitary matrix accounting for SOP drift defined in [13] as

$$\mathbf{J}_{k+1,n} = \exp(-j\underline{\boldsymbol{\alpha}}_{k,n} \cdot \overrightarrow{\sigma}) \mathbf{J}_{k,n}, \qquad (4)$$

where $\exp(\cdot)$ is the matrix exponential and

$$\underline{\boldsymbol{\alpha}}_{k,n} \sim \mathcal{N}(\underline{0}, \sigma_{\mathrm{p}}^2 \mathbf{I}_3), \qquad (5)$$

where $\sigma_{\mathrm{p}}^2 = 2\pi\Delta\mathrm{p}T$ and $\Delta\mathrm{p}$ is referred to as the *polarization linewidth* determining the speed of the SOP drift and T is the symbol duration. Finally, $\overrightarrow{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$ is a tensor of the Pauli spin matrices [40, eq. (2.5.19)]

$$\sigma_1 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \quad \sigma_2 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \sigma_3 = \begin{bmatrix} 0 & -j \\ j & 0 \end{bmatrix}. \quad (6)$$

The total polarization linewidth scales with N and can be defined as

$$\Delta\mathrm{p}_{\mathrm{tot}} = N \cdot \Delta\mathrm{p}. \qquad (7)$$

For the DP-PDL channel, we define the segment-wise PDL in dB as

$$\varphi_n = 10 \log_{10} \left(\frac{1 + \gamma_n}{1 - \gamma_n} \right), \qquad (8)$$
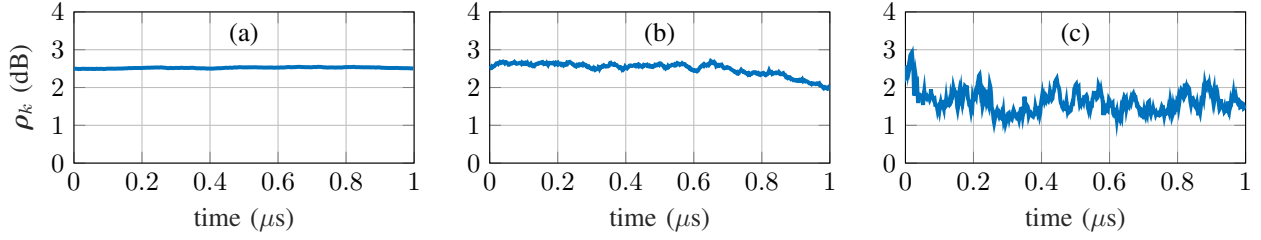
Fig. 2: The evolution of aggregated PDL ratio ρ_k in dB for different SOP drifts in a time period of 1 microsecond is plotted. The symbol rate is 28 Gbaud (i.e., the symbol duration $T = 3.57 \cdot 10^{-11}$) and $N = 20$. (a) $\Delta_{\text{Ptot}} \cdot T = 10^{-8}$, (b) $\Delta_{\text{Ptot}} \cdot T = 3.57 \cdot 10^{-6}$, and (c) $\Delta_{\text{Ptot}} \cdot T = 3.57 \cdot 10^{-5}$.

and the aggregated PDL ratio at time k as

$$\rho_k = \frac{\left\| \boldsymbol{\lambda}_k^{\max} \right\|^2}{\left\| \boldsymbol{\lambda}_k^{\min} \right\|^2}, \tag{9}$$

where $\boldsymbol{\lambda}_k^{\max}$ and $\boldsymbol{\lambda}_k^{\min}$ are the singular values of $\mathbf{H}_k$. The average aggregated PDL over K transmitted symbols in dB is defined as

$$\bar{\rho} = 10 \log_{10} \left(\mathbb{E}_{\mathbf{H}} \left[\frac{1}{K} \sum_{k=0}^{K-1} \boldsymbol{\rho}_k \right] \right). \tag{10}$$

Although each PDL component of the fiber is assumed to be constant, the aggregated PDL ρ_k of the channel will drift with time. Fig. 2 illustrates the dependence of the aggregated PDL dependence on the SOP drift. The PDL evolution is shown in a window of one microsecond with three different SOP drift speeds showing that ρ_k is strongly dependent on the speed of the SOP drift making the channel tracking challenging.

III. PROPOSED ALGORITHMS

Since the received data suffers from a time-varying SOP drift, adaptive channel tracking is needed to recover the transmitted data. This section describes the proposed channel tracking algorithms. First, an adaptive unitary tracking algorithm is proposed, which performs best when PDL is negligible (e.g., $\boldsymbol{\Gamma}_n = \mathrm{I}_2 \quad \forall n$); after that, a data-aided estimation algorithm is proposed that estimates and tracks both SOP and PDL of the DP-PDL channel.

A. Decision-directed Sliding Window Kabsch Algorithm

In this section, we introduce a modification to the DD-Kabsch algorithm [37] and make it suitable for polarization tracking in fast drifting channels. The key idea is to add a sliding window called *equalization window* with a size of L and update the channel estimate for each window. We define

$$\begin{aligned} \mathbf{X}_k &= [\underline{\mathbf{x}}_k, \underline{\mathbf{x}}_{k+1}, \ldots, \underline{\mathbf{x}}_{k+L-1}], \\ \hat{\mathbf{S}}_k &= [\hat{\underline{\mathbf{s}}}_k, \hat{\underline{\mathbf{s}}}_{k+1}, \ldots, \hat{\underline{\mathbf{s}}}_{k+L-1}], \end{aligned} \tag{11}$$

where $\mathbf{X}_k$ is a $2 \times L$ matrix of received symbols and $\hat{\mathbf{S}}_k$ is a $2 \times L$ matrix of estimated received symbols. Ignoring the PDL, the channel matrix can be expressed as

$$\mathbf{H}_{k+1} = \mathbf{R}_k \mathbf{H}_k, \tag{12}$$

where

$$\mathbf{R}_k = \mathbf{J}_{k,N} \mathbf{J}_{k,N-1} \cdots \mathbf{J}_{k,1}, \tag{13}$$

is obtained from (2) by substituting $\boldsymbol{\Gamma}_n = \mathrm{I}_2$.

Although the channel $\mathbf{H}_k$ changes for every k, a constant estimated channel $\hat{\mathbf{H}}_k$ is used over an equalization window of length L to obtain the required decision-directed symbols for the tracking algorithm. This can be justified when the speed of the drift is not extremely high, and L is small. Then the transmitted symbols in an equalization window can be estimated by the previous estimate of the channel $\hat{\mathbf{H}}_k$ and using the minimum Euclidean distance criterion

$$\hat{\underline{\mathbf{s}}}_k = \arg \min_{\underline{c} \in \mathcal{S}} \left\| \hat{\mathbf{H}}_k^{-1} \underline{\mathbf{x}}_k - \underline{c} \right\|^2. \tag{14}$$

Then, we define the channel estimation problem as

$$\arg \min_{\hat{\mathbf{R}}_k} \left\| \mathbf{X}_k - \hat{\mathbf{R}}_k \hat{\mathbf{H}}_k \hat{\mathbf{S}}_k \right\|^2 \quad \text{subject to} \quad \hat{\mathbf{R}}_k \hat{\mathbf{R}}_k^{\dagger} = \mathrm{I}_2, \tag{15}$$

which is known as the orthogonal procrustes problem. The optimal solution is given in [41]–[43] as

$$\hat{\mathbf{R}}_k = \mathbf{U}_k \mathbf{V}_k^{\dagger}, \tag{16}$$

where

$$\mathbf{U}_k \Sigma_k \mathbf{V}_k^{\dagger} = \mathrm{svd}(\mathbf{X}_k \hat{\mathbf{S}}_k^{\dagger} \hat{\mathbf{H}}_k^{\dagger}), \tag{17}$$

where $\mathrm{svd}(\cdot)$ stands for singular value decomposition and Σ_k is a positive definite diagonal matrix (not used in the algorithm) containing the singular values. Note that $\hat{\mathbf{R}}_k$ is not an estimate of $\mathbf{R}_k$. Instead it is an averaged estimation of all the rotations in one equalization window. Finally, initializing the estimated channel matrix as $\hat{\mathbf{H}}_0 = \mathrm{I}_2$, the next estimated channel will be updated by

$$\hat{\mathbf{H}}_{k+\nu} = \hat{\mathbf{R}}_k \hat{\mathbf{H}}_k, \tag{18}$$

where ν is the sliding stride of the equalization window, meaning that the equalization window slides ν symbols at a time. The complexity of the algorithm is roughly L/ν times higher than the DD-Kabsch algorithm in [37]. However, we will show that, unlike the GD-based algorithms, once L and ν are set properly, no further active parameter (e.g., step-size) adjustment is needed.

Algorithm 1 describes the proposed SW-Kabsch algorithm. It takes as inputs the previous instance of the estimated channel $\hat{\mathbf{H}}_k$, the $2 \times L$ received symbols matrix $\mathbf{X}_k$, time instance k, and sliding stride ν. Then, it returns the updated channel matrix $\hat{\mathbf{H}}_{k+\nu}$, the $2 \times \nu$ matrix of decided symbols $\tilde{\mathbf{S}}_k$, and the updated time instance k.
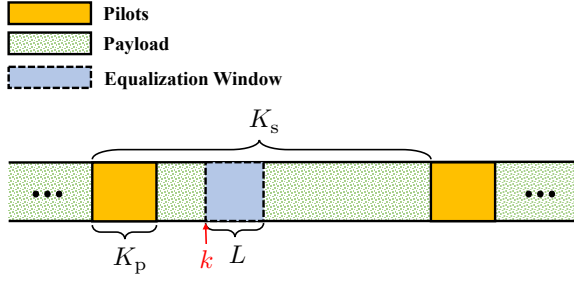
Fig. 3: Data frame with inserted pilot sequence and pilot symbols. The transmission block length is K_{s} and the length of the pilot sequence is K_{p}.

Algorithm 1 SW-Kabsch

Input: $\mathbf{X}_k, \hat{\mathbf{H}}_k, k, \nu$
Output: $\hat{\mathbf{H}}_{k+\nu}, \tilde{\mathbf{S}}_k, k$

1: **for** $l = 0, \ldots, L-1$ **do**

    2: $\hat{\underline{\mathbf{s}}}_{k+l} = \arg\min_{\underline{c} \in \mathcal{S}} \left\| \hat{\mathbf{H}}_k^{-1} \underline{\mathbf{x}}_{k+l} - \underline{c} \right\|^2$ // Eq. (14)

    end

3: Compute $\mathbf{U}_k$ and $\mathbf{V}_k$ using Eq. (17)
4: $\hat{\mathbf{R}}_k = \mathbf{U}_k \mathbf{V}_k^{\dagger}$
5: $\hat{\mathbf{H}}_{k+\nu} = \hat{\mathbf{R}}_k \hat{\mathbf{H}}_k$
6: $\tilde{\mathbf{S}}_k = [\hat{\underline{\mathbf{s}}}_k, \hat{\underline{\mathbf{s}}}_{k+1}, \ldots, \hat{\underline{\mathbf{s}}}_{k+\nu-1}]$
7: $k = k + \nu$

B. Pilot-Aided Algorithms

In this part, we propose a pilot-aided adaptive algorithm to estimate and track both SOP drift and accumulated PDL. The inserted pilots are chosen from the quadrature phase-shift keying (QPSK) constellation.

Fig. 3 shows the data frame construction where a pilot sequence of length K_{p} is inserted at the beginning of each transmission block of length K_{s} and as previously defined, L is the length of the equalization window. We also define the $2 \times K_{\mathrm{p}}$ matrix of pilots as

$$\mathrm{S}^{\mathrm{P}} = [\underline{s}_0^{\mathrm{p}}, \underline{s}_1^{\mathrm{p}}, \ldots, \underline{s}_{K_{\mathrm{p}}-1}^{\mathrm{p}}], \tag{19}$$

where the pilots are orthogonal with respect to each polarization (i.e., $\mathrm{S}^{\mathrm{P}}\mathrm{S}^{\mathrm{p}\dagger} = \delta \mathrm{I}_2$ where $\delta \in \mathbb{R}$ is a constant). The pilot sequence length is set to be larger than 2, which coincides with the optimal pilot selection suggested in [44], [45]. By tuning K_{p} and K_{s}, the performance of the proposed data-aided algorithm can be optimized for different channels; however, there is a trade-off between the overhead and the performance.

1) Sliding Window Least Square Algorithm (SW-LS): The key idea is similar to SW-Kabsch. Note that the standard least squares (LS) algorithm is designed to estimate deterministic channels; however, we combine LS with a decision-directed sliding window enabling us to track the DP-PDL channel adaptively. Considering the PDL and SOP drift, we can write the next instance of the channel matrix as

$$\mathbf{H}_{k+1} = \mathbf{G}_k \mathbf{H}_k, \tag{20}$$

where $\mathbf{G}_k$ is 2×2 a complex matrix accounting for the overall rotations and aggregated PDL since time instance k. The channel estimation problem can be defined as

$$\arg\min_{\hat{\mathbf{G}}_k} \left\| \mathbf{X}_k - \hat{\mathbf{G}}_k \hat{\mathbf{H}}_k \hat{\mathbf{S}}_k \right\|^2, \tag{21}$$

where the decision-directed transmitted symbols can be estimated by (14). Hence, $\hat{\mathbf{S}}_k$ and $\mathbf{X}_k$ can be formed in the same way as (11). The optimal solution of (21) is given by the LS algorithm as

$$\hat{\mathbf{G}}_k = \mathbf{X}_k \hat{\mathbf{S}}_k^{\dagger} \hat{\mathbf{H}}_k^{\dagger} \left(\hat{\mathbf{H}}_k \hat{\mathbf{S}}_k \hat{\mathbf{S}}_k^{\dagger} \hat{\mathbf{H}}_k^{\dagger} \right)^{-1}. \tag{22}$$

Initializing the estimated channel matrix as $\hat{\mathbf{H}}_0 = \mathrm{I}_2$, the estimated channel $\hat{\mathbf{H}}_k$ then is updated by

$$\hat{\mathbf{H}}_{k+\nu} = \hat{\mathbf{G}}_k \hat{\mathbf{H}}_k. \tag{23}$$

Since $\hat{\mathbf{S}}_k$ is not hand-picked and comes from the received data symbols, the matrix inverse at the right-hand side of the (22) might become singular. A way to deal with this problem is to use the minimum mean square error (MMSE) criterion, which regularize the inverse matrix using he covariance matrix of the ASE noise. However, we assume that the receiver has no knowledge of the ASE noise power, and hence we do not consider the MMSE estimator. Besides, we observed that for high order modulations ($M \geq 16$) and $L \geq 16$, the chance of running to singularity is low.

The complexity of the proposed algorithm is approximately L/ν times higher than LS and can be controlled by proper selection of ν and L.

Algorithm 2 details the pilot-aided SW-LS algorithm. It takes as inputs the received symbols $\mathbf{X}_k$ defined in (11), the previous estimate of the channel $\hat{\mathbf{H}}_k$, time index k, and sliding stride ν as inputs, and returns the updated channel matrix $\hat{\mathbf{H}}_{k+\nu}$, the $2 \times \nu$ matrix of decided symbols $\tilde{\mathbf{S}}_k$, and the updated time instance k.

Algorithm 2 Pilot-Aided SW-LS

Input: $\mathbf{X}_k, \hat{\mathbf{H}}_k, k, \nu$
Output: $\hat{\mathbf{H}}_{k+\nu}, \tilde{\mathbf{S}}_k, k$

1: **for** $l = 0, \ldots, L-1$ **do**

    2: $i = (k + l \mod K_{\mathrm{s}})$

    3: **if** $i \leq K_{\mathrm{p}} - 1$ **then** // Check if the symbol is pilot

    4:     $\hat{\underline{\mathbf{s}}}_{k+l} = \underline{s}_i^{\mathrm{p}}$

    5: **else**

    6:     $\hat{\underline{\mathbf{s}}}_{k+l} = \arg\min_{\underline{c} \in \mathcal{S}} \left\| \hat{\mathbf{H}}_k^{-1} \underline{\mathbf{x}}_{k+l} - \underline{c} \right\|^2$ // Eq. (14)

    end

7: $\hat{\mathbf{G}}_k = \mathbf{X}_k \hat{\mathbf{S}}_k^{\dagger} \hat{\mathbf{H}}_k^{\dagger} \left(\hat{\mathbf{H}}_k \hat{\mathbf{S}}_k \hat{\mathbf{S}}_k^{\dagger} \hat{\mathbf{H}}_k^{\dagger} \right)^{-1}$ // Eq. (22)

8: $\hat{\mathbf{H}}_{k+\nu} = \hat{\mathbf{G}}_k \hat{\mathbf{H}}_k$
9: $\tilde{\mathbf{S}}_k = [\hat{\underline{\mathbf{s}}}_k, \hat{\underline{\mathbf{s}}}_{k+1}, \ldots, \hat{\underline{\mathbf{s}}}_{k+\nu-1}]$
10: $k = k + \nu$

---

2) Pilot-Aided Hybrid Algorithms: We consider a DP-PDL channel with constant PDL components Γ_n. We assume that for a short enough transmission block length K_{s}, the position of the maximum singular value of the channel matrix $\mathbf{H}_k$

remains the same, and that the aggregated PDL ratio ρ_k is almost constant. Thus, one can argue that after compensating for the channel using the pilots, the residual channel matrix is almost unitary, which can be tracked using an arbitrary unitary channel tracking algorithm. Therefore, considering the frame structure in Fig. 3, we propose two hybrid algorithms, which are described in the following.

Since the frame structure is periodic, the algorithms are described for the first frame, where the same procedure is repeated for the subsequent frames. The first K_{p} transmitted symbols are pilots as defined in (19) and the corresponding $2 \times K_{\mathrm{p}}$ matrix of received symbols is $\mathbf{X}^{\mathrm{P}} = [\underline{\mathbf{x}}_0, \ldots, \underline{\mathbf{x}}_{K_{\mathrm{p}}-1}]$. We define the set of payload indices as $\mathcal{P} = \{K_{\mathrm{p}}+1, \ldots, K_{\mathrm{s}}\}$. In the first stage of the hybrid algorithms, a coarse estimation of the channel is obtained by applying the LS estimator as

$$\tilde{\mathbf{H}}_0 = \frac{1}{\delta} \mathbf{X}^{\mathrm{P}} \mathbf{S}^{\mathrm{P}}. \qquad (24)$$

The compensated payload symbols can be obtained according to

$$\mathbf{X}_{\mathrm{c}} = \tilde{\mathbf{H}}_0^{-1} \mathbf{X}^{\mathrm{d}}, \qquad (25)$$

where $\mathbf{X}^{\mathrm{d}} = [\underline{\mathbf{x}}_{K_{\mathrm{p}}}, \ldots, \underline{\mathbf{x}}_{K_{\mathrm{p}}+K_{\mathrm{s}}-1}]$ is the matrix of payload symbols. The residual channel matrix at time $k \in \mathcal{P}$ can be written as

$$\mathbf{H}_k^{\mathrm{r}} = \tilde{\mathbf{H}}_0^{-1} \mathbf{H}_k. \qquad (26)$$

For a short K_{s}, the aggregated PDL ratio ρ_k for $k \in \mathcal{P}$ is assumed to be constant. Thus, assuming that the aggregated PDL is compensated by $\tilde{\mathbf{H}}_0^{-1}$, the residual channel matrix $\mathbf{H}_k^{\mathrm{r}}$ can be regarded as an almost unitary matrix. Therefore, in the second stage, given $\mathbf{X}_{\mathrm{c}}$, a unitary adaptive algorithm is used to track the residual channel matrix $\mathbf{H}_k^{\mathrm{r}}$ for $k \in \mathcal{P}$.

In this paper, using LS for the first stage, we propose the LS-SW-Kabsch and LS-DD-Czegledi hybrid algorithms where SW-Kabsch and DD-Czegledi [38] are used in the second stage, respectively.

IV. RESULTS

A. Simulation Setup

Polarization-multiplexed 16 quadrature amplitude modulation (PM-16-QAM) at a symbol rate of $R_s = 28$ Gbaud (i.e., $T = 1/Rs_s$) is considered. A random sequence of $K = 10^5$ symbols is transmitted on each polarization where the initial matrices $\mathbf{J}_{0,1}, \ldots, \mathbf{J}_{0,N}$ are drawn from the set of all 2×2 unitary matrices. Thereby the SOP gets a uniform distribution over the Poincaré sphere. The presented results are evaluated by averaging over 10^5 such sequences. For the DP-PDL channel, it is assumed that the link has $N = 20$ segments where all the segments have identical polarization linewidth $\Delta\mathrm{p}$ and segment-wise PDL ratio γ_n. The launch power in each polarization is $P = \mathbb{E}[\underline{\mathbf{s}}_k^\dagger \underline{\mathbf{s}}_k]$ and the SNR per polarization is defined as $\mathrm{SNR} = P/\sigma_z^2$. The performance is assessed by estimating the symbol error rate (SER) for different setups in the presence of SOP drift, PDL, and ASE noise.

1) Blind / Decision-Directed Algorithms Setup: For comparison, the results obtained from the MCMA [24], DD-Kabsch [37], DD-Czegledi [38] algorithms are presented as benchmarks. The DD-Kabsch algorithm operates in a decision-directed block-wise fashion where it uses a block size of $L_{\mathrm{Kabsch}} = 16$, which gives the best polarization drift tolerance for this algorithm. To ensure the convergence of the DD-Czegledi and MCMA algorithms, they are implemented with two stages of convergence where the first stage uses a larger tracking step size μ than the one is used in the second stage; for detailed implementation, refer to [38, Algorithm 1]. To compare the SW-Kabsch algorithm with the best version of the benchmarks, the tracking step size μ of the DD-Czegledi and MCMA algorithms is optimized as a function of SOP drift speed and SNR.

Moreover, coherent differential coding is used to resolve the four-fold phase ambiguity of the PM-16-QAM constellation for blind and decision-directed algorithms. Deploying differential coding induces an SNR penalty meaning that an extra SNR is needed to attain a given SER compared to nondifferential schemes [46]. For comparison, differential coding is used even when the channel is perfectly known (i.e., $\hat{\mathbf{H}}_k = \mathbf{H}_k$). Both decision-directed and blind algorithms may swap the equalized channels yielding polarization ambiguity. The polarization ambiguity may be resolved by inserting a few pilots in the data load. In most communication systems, it is common to use pilots for different purposes such as timing synchronization, carrier frequency estimation, etc. Therefore, in this paper, a genie-aided ambiguity resolution is used for both benchmarks and the proposed algorithm in our simulation, whereas we expect real deployed receivers to use pilots.

The SW-Kabsch operates with a sliding window of size $L = 24$. The sliding stride is $\nu = 6$, which results in 4 times higher computational complexity than the DD-Kabsch algorithm. The parameters L and ν are chosen such that SW-Kabsch outperform the benchmarks with relatively low complexity. Note that optimizing ν and L as a function of SOP drift might alter the performance of the algorithm for a fixed computational complexity.

2) Pilot-aided Algorithms Setup: The pilot length is chosen to be $K_{\mathrm{p}} = 16$, and the block length is assumed to be $K_{\mathrm{s}} = 1016$ symbols yielding 1.6% overhead. Note that optimizing K_{p} and K_{s} as a function of the polarization linewidth, $\bar{\rho}$, and SNR might improve the performance of the pilot-aided tracking algorithms; however, this requires a feedback channel between the receiver and the transmitter, which might not be available in many applications. Thus, we used fixed K_{p} and K_{s}, which performs reasonably well for a large range of channel parameters. The SW-LS algorithm is set to use an equalization window of $L = 24$ with a $\nu = 6$ to get a fair performance while keeping the complexity low.

As for the benchmarks, a hybrid algorithm called LS-DDLMS is used, where the DDLMS algorithm is initialized with a coarse estimate of the channel obtained by applying LS on the pilot sequence. The maximum likelihood (ML) detection for a known channel at the receiver is also used to serve as a benchmark.
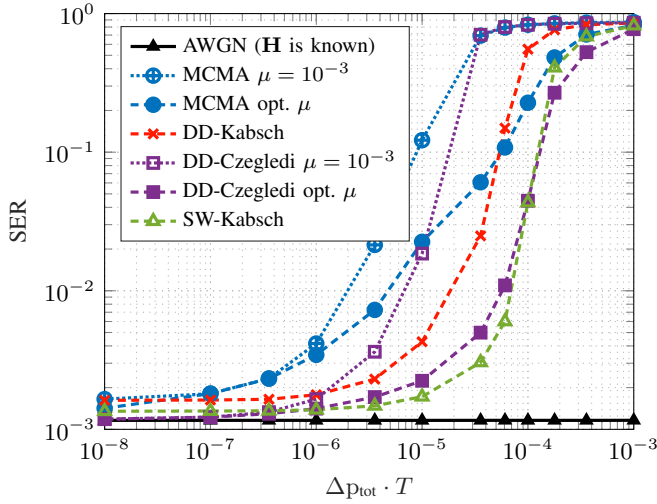
Fig. 4: The polarization drift tolerance at SNR = 18 dB is shown for four tracking algorithms using PM-16-QAM constellations where the vertical axis shows the symbol error rate (SER).

B. Polarization Drift Tolerance

This part evaluates the channel tracking ability of the proposed and conventional algorithms for the DP-PDL channel. The polarization drift sensitivity is measured by sweeping Δp while the SNR is fixed. The polarization movement between two consecutive symbols on the Poincaré sphere can be quantitatively characterized by the angle of the polarization movement θ_{SOP} [7, Eq. (2)]. For a fixed symbol rate of $R_s = 28$ Gbaud (i.e., $T = 35.7$ ps), we sweep $\Delta p_{tot} \cdot T$ from 10^{-8} to 10^{-3} which gives an average θ_{SOP} from $5 \cdot 10^{-4}$ to $1.57 \cdot 10^{-1}$ rad/symbol, respectively. Note that θ_{SOP} does not correspond to the angular velocity of the SOP movement in rad/s. The angular velocity is quantified as the time derivative of θ_{SOP}, which is undefined for our model (i.e., the time derivative of a Wiener process is undefined).

1) Channel with Negligible PDL: The SER versus $\Delta p_{tot} \cdot T$ is plotted in Fig. 4. The SNR is set to achieve SER = 10^{-3} for a known channel at the receiver. As can be seen, the proposed SW-Kabsch algorithm offers a better polarization drift tolerance than the benchmarks at the expense of approximately 4 times higher complexity than DD-Kabsch. More specifically, when the channel drifts slowly, i.e., $\Delta p_{tot} \cdot T < 10^{-6}$, SW-Kabsch and DD-Czegledi show almost the same performance; SW-Kabsch gradually outperforms DD-Czegledi when the channel gets faster. Finally, DD-Czegledi takes over at $\Delta p_{tot} \cdot T > 10^{-4}$, but in this regime, the SER is out of practical interest.

To investigate the step size sensitivity of the GD-based algorithms, the SER of MCMA and DD-Czegledi for fixed and optimal μ is also plotted (see blue and purple dotted curves). The step size is set to $\mu = 10^{-3}$ such that MCMA and DD-Czegledi perform their best for slowly varying channels. Evidently, MCMA and DD-Czegledi show a higher step size sensitivity and hence a lower polarization drift tolerance for a fixed μ. While the tracking capability of the GD-based algorithms is highly dependent on the proper adjustment of the tracking step size, SW-Kabsch requires no parameter tuning.

This could be advantageous in bursty channels where the SOP drift speed does not remain constant during the whole transmission. Thus, an algorithm that is tailor-made for a specific SOP drift speed may fail to track sudden changes in the channel.

2) Channel with considerable PDL: Fig. 5 shows the SER versus $\Delta p_{tot} \cdot T$ for three different average aggregated PDL $\bar{\rho}$ levels where the SNR is set such that SER = 10^{-3} is achieved with the ML detection. Note that to change the average aggregated PDL $\bar{\rho}$ in Figs. 5(a), (b), and (c), the segment-wise PDL φ_n (defined in (8)) of all the segments is set to 0.25, 0.70, and 1.10 dB, respectively. It can be seen that SW-LS shows the best tolerance to SOP drift in all considered average aggregated PDL $\bar{\rho}$ levels. All pilot-aided algorithms behave roughly the same at low SOP drifts ($\Delta p_{tot} \cdot T < 10^{-6}$), at higher drift speeds ($\Delta p_{tot} \cdot T \geq 10^{-6}$), SW-LS shows the best polarization drift tolerance.

Interestingly, for $\bar{\rho} = 1.0$ dB, the hybrid unitary tracking algorithms (LS-DD-Czegledi and LS-SW-Kabsch) have a better polarization drift tolerance than the LS-DDLMS algorithm. A possible justification might be that for small $\bar{\rho}$, the residual channel $\mathbf{H}_k^r$ (26) is almost unitary, and therefore a unitary estimation of the channel might result in a better estimate. This implies that even for $\bar{\rho} = 1.0$ dB, nonunitary tracking algorithms could be replaced by unitary tracking algorithms to obtain higher polarization drift tolerance. The performance of the proposed hybrid algorithms might alter for larger $\bar{\rho}$ by decreasing the transmission block length K_s at the expense of higher overhead.

C. Additive Noise Tolerance

Fig. 6 compares the SER of SW-Kabsch with the benchmarks for a DP-PDL channel for three different SOP drift speeds. The PDL is assumed to be negligible (i.e., $\gamma_n = 0$). All the studied algorithms have roughly similar performances for a very slowly drifting channel, see Fig. 6(a). However, when the channel drifts quickly, as shown in Fig. 6(b), DD-Czegledi and SW-Kabsch perform almost similarly, and MCMA is no longer an option. Finally, for an even faster channel presented in Fig. 6(c), SW-Kabsch outperforms all the other algorithms.

Fig. 7 shows the SER of various algorithms for a DP-PDL channel with $\bar{\rho} = 3.0$ dB. From Fig. 7(a), it can be concluded that all the considered algorithms except MCMA perform similarly for a slowly drifting channel. However, for fast drifting channels, Figs. 7(b) and 7(c) show that the proposed SW-LS algorithm outperforms the other algorithms.

D. Computational Complexity vs. Performance

In [38, Table I], it has been shown that DD-Czegledi is almost twice as complex as DD-Kabsch for a dual-polarization channel. Therefore, taking DD-Kabsch as a reference, a rough complexity comparison of the proposed algorithm is presented. Compared to DD-Kabsch, the SW-Kabsch algorithm has an additional sliding window, resulting in approximately L/ν times higher complexity. The SW-LS algorithm has a matrix inverse instead of the singular value decomposition in SW-Kabsch, which essentially has the same complexity. Therefore,
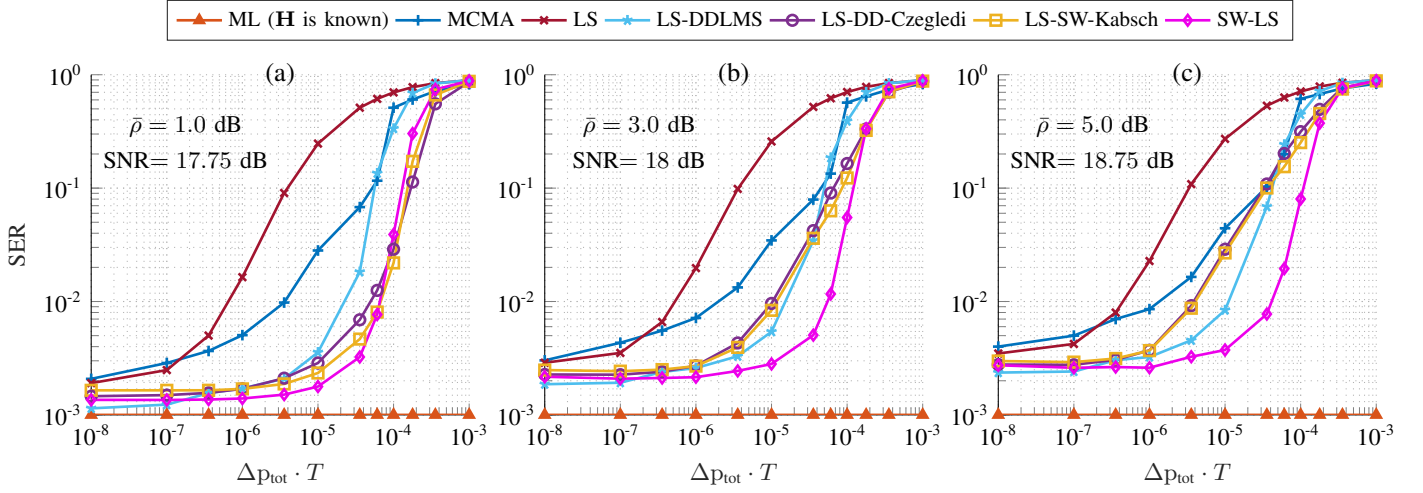
Fig. 5: The achievable SER performance of pilot-aided algorithms for a fiber with $N = 20$ segments is plotted. In (a), (b), and (c), to change the average aggregated PDL $\boldsymbol{\rho}$, the segment-wise PDL φ_n of all the segments is set to 0.25, 0.70, and 1.10 dB, respectively.
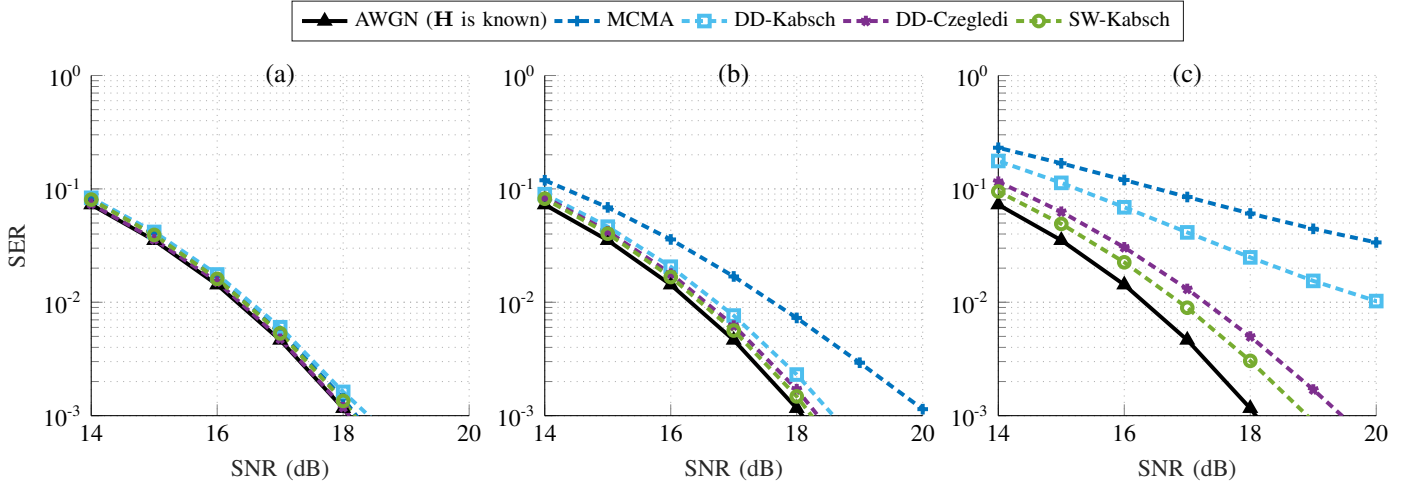

Fig. 6: The achievable SER performance of the blind tracking algorithms when PDL is negligible, for three drift speeds is depicted where (a) $\Delta_{\mathrm{p_{tot}}} \cdot T = 10^{-8}$, (b) $\Delta_{\mathrm{p_{tot}}} \cdot T = 3.57 \cdot 10^{-6}$, and (c) $\Delta_{\mathrm{p_{tot}}} \cdot T = 3.57 \cdot 10^{-5}$.
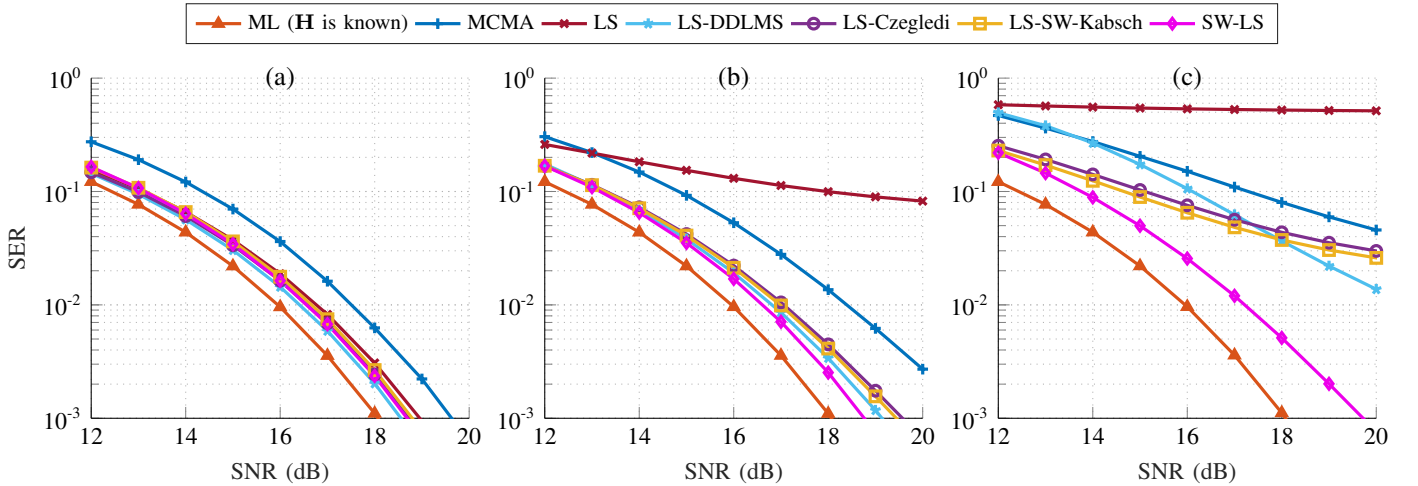

Fig. 7: The achievable SER performance of various algorithms for a DP-PDL channel with an average aggregated PDL of $\bar{\rho} = 3.0$ dB is plotted where (**a**) $\Delta_{\mathrm{p_{tot}}} \cdot T = 10^{-8}$, (**b**) $\Delta_{\mathrm{p_{tot}}} \cdot T = 3.57 \cdot 10^{-6}$, and (**c**) $\Delta_{\mathrm{p_{tot}}} \cdot T = 3.57 \cdot 10^{-5}$.

SW-LS is also at least L/ν times more complex than DD-Kabsch. While the complexity of the proposed algorithms is tractable for low-dimensional channels (e.g., DP-PDL channel, few-mode fiber, etc.), it becomes challenging for high-dimensional channels since the complexity of singular value decomposition and matrix inversion operations increases cubically with the number of dimensions.

The complexity of the algorithms can be decreased at the expense of performance. For instance, reducing the L/ν ratio decreases the complexity of the proposed algorithms but also degrades the performance. Besides, adjusting the L/ν ratio as a function of channel parameters might reduce the complexity of the algorithms for certain channel conditions (e.g., slowly drifting channels).

V. CONCLUSION

We have proposed SW-Kabsch and SW-LS algorithms to track memoryless DP-PDL channels. Both proposed algorithms use a sliding window in a decision-directed fashion, processing multiple symbols at a time. Numerical simulations are used to evaluate and compare the performance of the proposed algorithms with the conventional tracking algorithms, including MCMA which is the most popular tracking algorithm in the literature. Results show that the proposed algorithms are more robust to polarization drift than the benchmarks. While parameter adjustment is required for GD-based benchmarks (e.g., MCMA, DDLMS, DD-Czegledi, etc.), the proposed algorithms need no such adjustments and still outperform the benchmarks. Besides, unlike the DD-Czegledi algorithm [38], which cannot be scaled for higher dimensional channels, the proposed algorithm can be applied to any number of dimensions at the expense of increasing the computational complexity.

The proposed algorithms are analyzed assuming negligible polarization-mode dispersion, which is not the case in more realistic fiber channels. Although it is not possible to compensate for polarization-mode dispersion and SOP separately, the proposed algorithms can be used in a hybrid fashion as in [47], where polarization-mode dispersion is compensated in the frequency domain and SOP is tracked in the time domain. The effect of polarization-mode dispersion is left for future work.

REFERENCES

[1] K. Kikuchi, "Fundamentals of coherent optical fiber communications," *J. Lightwave Technol.*, vol. 34, no. 1, pp. 157–179, 2016.

[2] K. Roberts, M. O'Sullivan, K.-T. Wu, H. Sun, A. Awadalla, D. J. Krause, and C. Laperle, "Performance of dual-polarization QPSK for optical transport systems," *J. Lightwave Technol.*, vol. 27, no. 16, pp. 3546–3559, Aug. 2009.

[3] M. Karlsson, J. Brentel, and P. A. Andrekson, "Long-term measurement of PMD and polarization drift in installed fibers," *J. Lightwave Technol.*, vol. 18, no. 7, pp. 941–951, Jul. 2000.

[4] D. Waddy, P. Lu, L. Chen, and X. Bao, "Fast state of polarization changes in aerial fiber under different climatic conditions," *IEEE Photon. Technol. Lett.*, vol. 13, no. 9, pp. 1035–1037, Sep. 2001.

[5] J. Wuttke, P. Krummrich, and J. Rosch, "Polarization oscillations in aerial fiber caused by wind and power-line current," *IEEE Photon. Technol. Lett.*, vol. 15, no. 6, pp. 882–884, June 2003.

[6] P. M. Krummrich, D. Ronnenberg, W. Schairer, D. Wienold, F. Jenau, and M. Herrmann, "Demanding response time requirements on coherent receivers due to fast polarization rotations caused by lightning events," *Opt. Express*, vol. 24, no. 11, pp. 12 442–12 457, May 2016.

[7] D. Charlton, S. Clarke, D. Doucet, M. O'Sullivan, D. L. Peterson, D. Wilson, G. Wellbrock, and M. Bélanger, "Field measurements of SOP transients in OPGW, with time and location correlation to lightning strikes," *Opt. Express*, vol. 25, no. 9, pp. 9689–9696, May 2017.

[8] K. Kikuchi, "Polarization-demultiplexing algorithm in the digital coherent receiver," in *Proc. IEEE/LEOS Summer Topical Meetings*, 2008, p. MC2.2.

[9] I. Roudas, A. Vgenis, C. S. Petrou, D. Toumpakaris, J. Hurley, M. Sauer, J. Downie, Y. Mauro, and S. Raghavan, "Optimal polarization demultiplexing for coherent optical communications systems," *J. Lightwave Technol.*, vol. 28, no. 7, pp. 1121–1134, 2009.

[10] F. Heismann and K. Tokuda, "Polarization-independent electro-optic depolarizer," *Opt. Lett.*, vol. 20, no. 9, pp. 1008–1010, 1995.

[11] S. J. Savory, "Digital filters for coherent optical receivers," *Opt. Express*, vol. 16, no. 2, pp. 804–817, Jan. 2008.

[12] N. J. Muga and A. N. Pinto, "Adaptive 3-D Stokes space-based polarization demultiplexing algorithm," *J. Lightwave Technol.*, vol. 32, no. 19, pp. 3290–3298, 2014.

[13] C. B. Czegledi, M. Karlsson, E. Agrell, and P. Johannisson, "Polarization drift channel model for coherent fibre-optic systems," *Scientific Reports*, vol. 6, no. 1, pp. 1–11, 2016.

[14] N. Gisin, "Statistics of polarization dependent losses," *Opt. Commun.*, vol. 114, no. 5-6, pp. 399–405, 1995.

[15] A. Mecozzi and M. Shtaif, "The statistics of polarization-dependent loss in optical communication systems," *IEEE Photon. Technol. Lett.*, vol. 14, no. 3, pp. 313–315, 2002.

[16] A. Galtarossa and L. Palmieri, "The exact statistics of polarization-dependent loss in fiber-optic links," *IEEE Photon. Technol. Lett.*, vol. 15, no. 1, pp. 57–59, 2003.

[17] C. Vinegoni, M. Karlsson, M. Petersson, and H. Sunnerud, "The statistics of polarization-dependent loss in a recirculating loop," *J. Lightwave Technol.*, vol. 22, no. 4, pp. 968–976, 2004.

[18] M. Shtaif, "Performance degradation in coherent polarization multiplexed systems as a result of polarization dependent loss," *Opt. Express*, vol. 16, no. 18, pp. 13 918–13 932, 2008.

[19] T. Duthel, C. Fludger, J. Geyer, and C. Schulien, "Impact of polarisation dependent loss on coherent POLMUX-NRZ-DQPSK," in *Proc. Opt. Fiber Commun. (OFC)*, 2008, p. OThU5.

[20] M. Shtaif and O. Rosenberg, "Polarization-dependent loss as a waveform-distorting mechanism and its effect on fiber-optic systems," *J. Lightwave Technol.*, vol. 23, no. 2, p. 923, 2005.

[21] N. Rossi, P. Serena, and A. Bononi, "Polarization-dependent loss impact on coherent optical systems in presence of fiber nonlinearity," *IEEE Photon. Technol. Lett.*, vol. 26, no. 4, pp. 334–337, 2014.

[22] C. Xie and S. Chandrasekhar, "Two-stage constant modulus algorithm equalizer for singularity free operation and optical performance monitoring in optical coherent receiver," in *Proc. Opt. Fiber Commun. Conf. (OFC)*, Mar. 2010, p. OMK3.

[23] D. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems," *IEEE Trans. Commun.*, vol. 28, no. 11, pp. 1867–1875, Nov. 1980.

[24] K. N. Oh and Y. O. Chin, "Modified constant modulus algorithm: blind equalization and carrier phase recovery algorithm," in *Proc. IEEE Int. Conf. Commun. (ICC)*, vol. 1, 1995, pp. 498–502.

[25] J. Yang, J.-J. Werner, and G. Dumont, "The multimodulus blind equalization and its generalized algorithms," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 5, pp. 997–1015, Jun. 2002.

[26] K. Kikuchi, "Performance analyses of polarization demultiplexing based on constant-modulus algorithm in digital coherent optical receivers," *Opt. Express*, vol. 19, no. 10, pp. 9868–9880, May 2011.

[27] R. W. Lucky, "Techniques for adaptive equalization of digital communication systems," *The Bell System Technical Journal*, vol. 45, no. 2, pp. 255–286, 1966.

[28] S. J. Savory, "Digital coherent optical receivers: Algorithms and subsystems," *IEEE J. Sel. Topics Quantum Electron.*, vol. 16, no. 5, pp. 1164–1179, Sep.–Oct. 2010.

[29] M. Seimetz, *High-Order Modulation for Optical Fiber Transmission*. Heidelberg, Germany: Springer, 2009.

[30] L. Liu, Z. Tao, W. Yan, S. Oda, T. Hoshida, and J. C. Rasmussen, "Initial tap setup of constant modulus algorithm for polarization demultiplexing in optical coherent receivers," in *Proc. Opt. Fiber Commun. Conf. (OFC)*, 2009, p. OMT2.

This article has been accepted for publication in IEEE/OSA Journal of Lightwave Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JLT.2022.3194734

9

[31] A. Vgenis, C. S. Petrou, C. B. Papadias, I. Roudas, and L. Raptis, "Nonsingular constant modulus equalizer for PDM-QPSK coherent optical receivers," *IEEE Photon. Technol. Lett.*, vol. 22, no. 1, pp. 45–47, 2009.

[32] M. Morsy-Osman, M. Chagnon, Q. Zhuge, X. Xu, M. E. Mousa-Pasandi, Z. A. El-Sahn, and D. V. Plant, "Training symbol based channel estimation for ultrafast polarization demultiplexing in coherent single-carrier transmission systems with M-QAM constellations," in *Proc. Eur. Conf. Opt. Commun. (ECOC)*, 2012, p. Mo.1.A.4.

[33] M. S. Faruk, Y. Mori, C. Zhang, and K. Kikuchi, "Proper polarization demultiplexing in coherent optical receiver using constant modulus algorithm with training mode," in *Proc. OptoElectron. Commun. Conf. (OECC)*, 2010, pp. 768–769.

[34] B. Widrow, "Adaptive filters," in *Aspects of Network and System Theory*, R. E. Kalman and N. DeClaris, Eds.   New York, NY: Holt, Rinehart and Winston, Apr. 1971, pp. 563–587.

[35] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series: with Engineering Applications*.   MIT Press Cambridge, MA, 1949.

[36] J. G. Proakis and D. G. Manolakis, *Digital Signal Processing: Principles Algorithms and Applications*, 4th ed.   New York: Macmillan, 2007.

[37] H. Louchet, K. Kuzmin, and A. Richter, "Joint carrier-phase and polarization rotation recovery for arbitrary signal constellations," *IEEE Photon. Technol. Lett.*, vol. 26, no. 9, pp. 922–924, May 2014.

[38] C. B. Czegledi, E. Agrell, M. Karlsson, and P. Johannisson, "Modulation format independent joint polarization and phase tracking for coherent receivers," *J. Lightwave Technol.*, vol. 34, no. 14, pp. 3354–3364, Jul. 2016.

[39] M. Farsi, C. Häger, M. Karlsson, and E. Agrell, "Improved polarization tracking in the presence of PDL," in *Proc. Eur. Conf. Opt. Commun. (ECOC)*, 2022.

[40] J. N. Damask, *Polarization Optics in Telecommunications*.   New York, NY, USA: Springer, 2005.

[41] W. Gibson, "On the least-squares orthogonalization of an oblique transformation," *Psychometrika*, vol. 27, no. 2, pp. 193–195, 1962.

[42] P. H. Schönemann, "A generalized solution of the orthogonal procrustes problem," *Psychometrika*, vol. 31, no. 1, pp. 1–10, Mar. 1966.

[43] W. Kabsch, "A solution for the best rotation to relate two sets of vectors," *Acta Crystallographica Section A*, vol. 32, no. 5, pp. 922–923, Sep. 1976.

[44] T. L. Marzetta, "Blast training: Estimating channel characteristics for high capacity space-time wireless," in *Proc. Annual Allerton Conf. Commun. Control and Computing*, 1999, pp. 958–966.

[45] B. Hassibi and B. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inform. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.

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

[47] Z. Zheng, N. Cui, H. Xu, X. Zhang, W. Zhang, L. Xi, Y. Fang, and L. Li, "Window-split structured frequency domain Kalman equalization scheme for large PMD and ultra-fast RSOP in an optical coherent PDM-QPSK system," *Optics Express*, vol. 26, no. 6, pp. 7211–7226, 2018.