Blind Polarization Demultiplexing of Shaped QAM Signals Assisted by Temporal Correlations

Bajaj, Vinod; Plas, Raf Van de; Wahls, Sander

DOI
10.1109/JLT.2023.3315370

Publication date
2024

Document Version
Final published version

Published in
Journal of Lightwave Technology

Blind Polarization Demultiplexing of Shaped QAM Signals Assisted by Temporal Correlations

Vinod Bajaj, Raf Van de Plas, and Sander Wahls, *Senior Member, IEEE*

***Abstract*—While probabilistic constellation shaping (PCS) enables rate and reach adaption with finer granularity [1] (Cho and Winzer, 2009), it imposes signal processing challenges at the receiver. Since the distribution of PCS-quadrature amplitude modulation (QAM) signals tends to be Gaussian, conventional blind polarization demultiplexing algorithms are not suitable for them [2] (Johnson et al., 1998). It is known that independently and identically distributed (iid) Gaussian signals, when mixed, cannot be recovered/separated from their mixture. For PCS-QAM signals, there are algorithms such as [3] and [4] Dris et al. (2019) and Athuraliya et al. (2004) which are designed by extending conventional blind algorithms used for uniform QAM signals. In these algorithms, an initialization point is obtained by processing only a part of the mixed signal, which have non-Gaussian statistics. In this article, we propose an alternative method wherein we add temporal correlations at the transmitter, which are subsequently exploited at the receiver in order to separate the polarizations. We will refer to the proposed method as frequency domain (FD) joint diagonalization (JD) probability aware-multi modulus algorithm (pr-MMA), and it is suited to channels with moderate polarization mode dispersion (PMD) effects. Furthermore, we extend our previously proposed JD-MMA [5] (Bajaj et al., 2022) by replacing the standard MMA with a pr-MMA, improving its performance. Both FDJD-pr-MMA and JD-pr-MMA are evaluated for a diverse range of PCS (entropy $\mathcal{H}$) of 64-QAM over a first-order PMD channel that is simulated in a proof-of-concept setup. A MMA initialized with a memoryless constant modulus algorithm (CMA) is used as a benchmark. We show that at a differential group delay (DGD) of 10% of symbol period T_{symb} and 18 dB SNR/pol., JD-pr-MMA successfully demultiplexes the PCS signals, while CMA-MMA fails drastically. Furthermore, we demonstrate that the newly proposed FDJD-pr-MMA is robust against moderate PMD effects by evaluating it over a DGD of up to 40% of T_{symb}. Our results show that the proposed FDJD-pr-MMA successfully equalizes PMD channels with a DGD up to 20% of T_{symb}.**

***Index Terms*—Digital signal processing, optical fiber communication, polarization demultiplexing, probabilistic constellation shaping.**

Manuscript received 24 February 2023; revised 21 August 2023; accepted 10 September 2023. Date of publication 14 September 2023; date of current version 16 January 2024. The work of Raf Van de Plas was supported by the National Institutes of Health's Common Fund, NIDDK, NEI, NIAID, NIA, and OD under Awards U54DK120058, U54DK134302, U54EY032442, R01AI138581, R01AI145992, and R01AG078803. *(Corresponding author: Vinod Bajaj.)*

Vinod Bajaj is with the Delft Center for Systems and Control, Delft University of Technology, 2628 CD Delft, The Netherlands (e-mail: v.bajaj-1@tudelft.nl).

Raf Van de Plas is with the Delft Center for Systems and Control, Delft University of Technology, 2628 CD Delft, The Netherlands, and also with the Mass Spectrometry Research Center and Department of Biochemistry, Vanderbilt University, Nashville, TN 37240 USA (e-mail: raf.vandeplas@tudelft.nl).

Sander Wahls is with the Karlsruhe Institute of Technology (KIT), Institute for Industrial Information Technology, Karlsruhe, TN, Germany (e-mail: s.wahls@tudelft.nl).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/JLT.2023.3315370.

Digital Object Identifier 10.1109/JLT.2023.3315370

I. Introduction

PROBABILISTIC constellation shaping (PCS) is a tool that allows for rate and reach adaption with finer granularity than a uniformly distributed quadrature amplitude modulation (QAM) signal [1]. It was shown recently that conventional blind digital signal processing (DSP) algorithms, which were originally developed for uniform QAM signals, are not suitable to equalize PCS-QAM signals from distortions such as frequency offset [6], phase noise [7], and polarization impairments [2], [8]. While it is possible to use pilot-aided algorithms, they require additional overhead in the data payload. There is thus a need for blind signal processing algorithms that can deal with PCS-QAM signals.

This article is concerned with blind algorithms for the compensation of polarization impairments. Already three decades ago, it was known that the commonly used constant modulus algorithm (CMA) is not suited for shaped sources [p. 136] [2]. PCS-QAM signals tend to have a Gaussian distribution. However, separating a mixture of two independently and identically distributed (iid) Gaussian signals into the original source signals is not possible. The main reason is that the contrast function that distinguishes the state of separation from the state of mixing, e.g. kurtosis or other higher order statistics (HOS), decreases or even vanishes completely when signal are Gaussian [9]. Therefore, classical methods of blind source separation (BSS), such as independent component analysis [9], CMA [2] which use HOS tend not to work well for PCS-QAM.

In the literature on blind polarization demultiplexing algorithms, this phenomenon is often referred to as the kurtosis problem [2], [8] or described as formation of a sphere instead of a lens-like structure in the Stokes-space [3], [10]. Recently, new algorithms based on the idea of magnitude-based separation of signal constellations (symbols) were proposed [3], [6]. In [3], 'inner QPSK' symbols were extracted and transformed into their three-dimensional representation in the Stokes-space, to find an initialization for the multi modulus algorithm (MMA). An extension of this algorithm was presented recently in which symbols beyond the 'inner QPSK' were used [10]. In [11], a variational auto-encoder based equalizer for the equalization of a first-order PMD channel was proposed and shown to work for PCS-QAM signals. We note the similarity between the idea of separating constellations in [3], [6], [10] and what is referred to as quantized CMA in the literature [4]. There, the received symbols are separated into two or more categories based on their magnitude. The symbols in separated categories have much

better contrast function or optimization landscape than the case where symbols are not categorized, i.e. all symbols are used. Therefore equalization is possible using classical algorithms like CMA. These algorithms extend methods that were originally designed for uniform QAM signals.

Here, we investigate an alternative idea. We exploit that if Gaussian signals exhibit diverse temporal correlations it is often possible to separate them using their second order statistics [9], [12]. An algorithm that is inherently suitable for Gaussian signals is expected to be suitable for PCS-QAM signals as well. Recently, we proposed a first method for blind polarization demultiplexing that exploits this idea [5]. The method is called joint diagonalization (JD) assisted MMA or JD-MMA, because it combines JD and a standard MMA (MMA) [13], [14]. The algorithm requires that temporal correlations are present in the transmitted signal, which were introduced by using additional filters at the transmitter. A major limitation of the JD-MMA is that it may not work when channel has memory such as PMD channel. Furthermore, it uses a MMA instead of probability aware pr-MMA which has been shown to perform better [3].

In this article, we present a new blind polarization demultiplexing algorithm based on second order statistics that can also be applied to channels with memory (i.e. with significant PMD). We furthermore improve our previous algorithm for the memoryless case [5] and investigate the performance of these methods in scenarios with different degrees of PMD. In addition, we discuss briefly the impact of pulse shaping, phase noise and frequency offset on the working of the proposed algorithms. Furthermore, the computational complexity and information loss due to the added temporal correlations are presented.

The article is organized as follows: in Section II, we briefly review the basics of PMD channel impairments and describe the first-order PMD model used in this article. Then, we discuss in Section III our previously proposed JD-MMA together with its improved version JD-pr-MMA, designed for separating signals from instantaneous mixtures. The newly proposed FDJD-pr-MMA algorithm, designed for PMD channels, is presented in Section IV. Some additional processing steps required by FDJD-pr-MMA are detailed in Section V. The simulation setup is explained in the Section VI, and numerical results are discussed in Section VII. The impact of pulse shaping, phase noise and frequency offset on the proposed algorithms, computational complexity and information loss due to filtering are briefly discussed in Section VIII. Finally, the article concludes in Section IX.

II. Brief Review of Polarization Mode Dispersion

An ideal optical fiber allows propagation of fundamental modes in two orthogonal polarizations without any interference. Real fiber, however, exhibits cylindrical asymmetry caused by factors such as manufacturing defects and mechanical stress, which leads to birefringence. That is, the two orthogonal polarization modes propagate with different group velocities. This phenomena is called polarization mode dispersion (PMD) [15], [16]. To understand PMD, let us consider a chunk of fiber with a constant birefringence along its length. In this chunk of fiber, there exist two axes known as 'slow' and 'fast' axes. For (linearly polarized) input light whose polarization aligns with either of these axes, the difference in the differential group velocities is maximum and the 'fast' and 'slow' light pulses thus reach the other end at different times. The difference of arrival times is called differential group delay (DGD). Although actual fiber does not have constant birefringence along its length, PMD can be modeled as a concatenation of several chunks with constant birefringence. We use such a first-order PMD model in our simulations [17], whose implementation is available online [18]. In Jones space, the polarization of light can be represented by a Jones vector. PMD effects that map input Jones vectors to output Jones vectors are given by a 2×2 unitary Jones matrix $\mathbf{U}(\omega)$, where ω is angular frequency. As explained, the PMD of the link can be described by a concatenation of Jones matrices, i.e. $\mathbf{U}(\omega) = \mathbf{U}_{n_c}(\omega)\mathbf{U}_{n_c-1}(\omega)\ldots\mathbf{U}_2(\omega)\mathbf{U}_1(\omega)$. Here, $\mathbf{U}_i(\omega)$ is the PMD induced by ith chunk and n_c is the number of chunks. The PMD of a single chunk is given by

$$\mathbf{U}_i(\omega) = \mathbf{Q}_i\left(-\theta_i, -\phi_i\right)\begin{pmatrix} e^{j\omega\triangle\tau_i/2} & 0 \\ 0 & e^{-j\omega\triangle\tau_i/2}\end{pmatrix}\mathbf{Q}_i\left(\theta_i, \phi_i\right) \tag{1}$$

where $\theta_i \sim \text{Uni}[-\pi/2, \pi/2]$, $\phi_i \sim \text{Uni}[-\pi/4, \pi/4]$, and $\mathbf{Q}_i$ is the rotation matrix of ith chunk,

$$\mathbf{Q}_i(\omega) = \begin{pmatrix} \cos(\theta_i) & \sin(\theta_i)e^{j\phi} \\ -\sin(\theta_i)e^{-j\phi} & \cos(\theta_i)\end{pmatrix}, \tag{2}$$

and $\triangle\tau_i$ is the DGD of an individual chunk. In the first-order model, the DGD τ_i is constant over the signal spectrum. The rotation of the state of polarization (RSOP) channel is a special case of (1) where the factor $\omega\triangle\tau_i$ is replaced with a phase shift that is constant over the signal bandwidth, i.e. $\omega\triangle\tau_i = \delta$. The RSOP channel can be implemented using a single chunk $n_c = 1$. Using a different notation, we can write the PMD channel propagation as

$$X_i(\omega) = \sum_{k=1}^{2} H_{ik}(\omega)S_k(\omega), \quad i,k \in \{1,2\}, \tag{3}$$

$$\text{or}\, \mathbf{X}(\omega) = \mathbf{H}(\omega)\mathbf{S}(\omega). \tag{4}$$

Here, $S_k(\omega)$ and $X_i(\omega)$ represent the discrete Fourier transform (DFT)[1] of the fiber input and output signals $s_k(t)$ and $x_i(t)$, respectively. In (4), bold capital letters represent matrices, i.e.

$$\mathbf{S}(\omega) = \begin{bmatrix} S_1(\omega) \\ S_2(\omega)\end{bmatrix}, \quad \mathbf{X}(\omega) = \begin{bmatrix} X_1(\omega) \\ X_2(\omega)\end{bmatrix}$$

and

$$\mathbf{H}(\omega) = \begin{bmatrix} H_{11}(\omega) & H_{12}(\omega) \\ H_{21}(\omega) & H_{22}(\omega)\end{bmatrix}.$$

[1]We follow the convention for DFT as $x(\omega) = \sum_k x(k)e^{(-j\frac{2\pi\omega k}{N})}$ and IDFT as $x(k) = \frac{1}{N}\sum_w x(w)e^{(j\frac{2\pi\omega k}{N})}$.

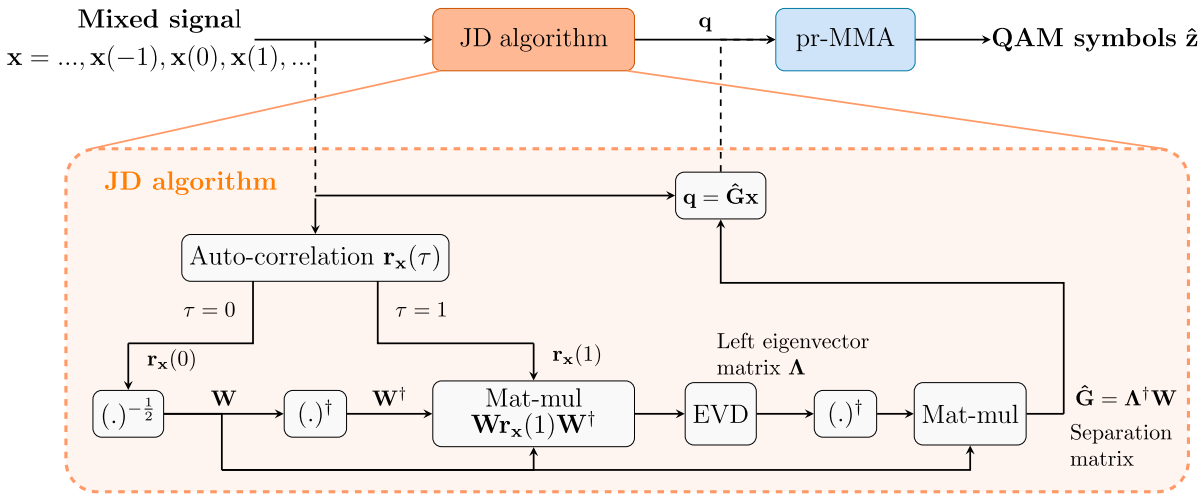

Fig. 1. Schematic of JD-pr-MMA. The JD algorithm estimates an instantaneous separation matrix $\hat{\mathbf{G}}$, which is applied to the mixed signal. Then, the added temporal correlations and residual distortions are removed by pr-MMA in order to recover PCS-QAM symbols. In the ideal case, the intermediate signal $\mathbf{q} = \mathbf{s}$ for the RSOP channel.

In the time-domain, (3) can be written as convolutive mixing

$$x_i(t) = \sum_{k=1}^{2} \sum_{\tau=-\ell_{ch}}^{+\ell_{ch}} h_{ik}(\tau) s_k(t-\tau), \quad i,k \in \{1,2\} \tag{5}$$

with ℓ_{ch} being the one-sided memory of the PMD channel. If $\ell_{ch} = 0$, (5) represents instantaneous mixing, i.e. the RSOP channel.

III. JD Assisted pr-MMA (JD-Pr-MMA)

The JD-pr-MMA method for blind demultiplexing of signals assumes that the mixing channel is memoryless [5]. Hence, it is ideally suited for RSOP channel effects. The method is composed of two algorithms working in a sequence. A schematic of the method is shown in Fig. 1. First, an approximate joint diagonalization (JD) algorithm estimates an inverse of the instantaneous mixing matrix and applies it to the mixed signal[9, Ch. 7.4.1]. Then, a pr-MMA removes any residual effects to recover the QAM symbols. The combined algorithm is referred to as JD assisted pr-MMA or JD-pr-MMA for short. We now explain this method in detail.

Let us denote a dual polarized transmit signal by $\mathbf{s} = [s_1,\ s_2]^T$, where $s_1(t)$ and $s_2(t)$ are base-band signals, and the mixed signal by $\mathbf{x}(t) = \mathbf{H}\mathbf{s}(t)$. Here, $\mathbf{H}$ is a instantaneous 2×2 unitary channel mixing matrix.

The method assumes that the transmit signals in the two orthogonal polarizations, i.e. s_1 and s_2 are

A1.1 zero-mean, second order ergodic, and wide sense stationary
A1.2 mutually uncorrelated, and
A1.3 have diverse temporal correlation, i.e. the temporal correlations of the sources are pairwise linearly independent.

We refer the reader to [9, Ch.7.3.1] for why the above assumptions are necessary. We meet the assumptions above by applying a filter $\mathbf{f} = [f_1,\ f_2]^T$ on the transmit iid PCS-QAM symbols on each polarization, $\mathbf{z} = [z_1,\ z_2]^T$, such that $s_i(t) = z_i(t) \star f_i(t)$ for $i \in 1, 2$. Here, $\star$ denotes the convolution operation. It is sufficient to filter only one of the signals at the transmitter to satisfy assumptions A1.

With the above assumption in place, it can be shown that the auto-correlation of the signals $\mathbf{s}$, i.e. $\mathbf{r}_\mathbf{s}(\tau) = \mathbb{E}\{\mathbf{s}(t)\mathbf{s}^\dagger(t+\tau)\} = \text{diag}(r_{s1}(\tau), r_{s2}(\tau))$, forms a sequence of diagonal matrices. Here, $r_{si}(\tau)$ is the correlation sequence of the ith polarization and $\dagger$ denotes the Hermitian transpose operation. It can also be shown that the auto-correlation of the mixed signal contains the instantaneous mixing (RSOP channel) matrix:

$$\mathbf{r}_\mathbf{x}(\tau) = \mathbb{E}\{\mathbf{x}(t)\mathbf{x}^\dagger(t+\tau)\} = \mathbf{H}\mathbf{r}_s(\tau)\mathbf{H}^\dagger. \tag{6}$$

Note that $\mathbf{H}\mathbf{r}_x(0)\mathbf{H}^\dagger = \mathbf{r}_\mathbf{s}(0)$ and $\mathbf{H}\mathbf{r}_x(1)\mathbf{H}^\dagger = \mathbf{r}_\mathbf{s}(1)$ are both diagonal because of the assumptions A1. We can thus recover $\mathbf{H}^\dagger$ up to rotations and ordering by finding a matrix that jointly diagonalizes $\mathbf{r}_\mathbf{x}(0)$ and $\mathbf{r}_\mathbf{x}(1)$.

A method for approximate joint diagonalization of correlation matrices $\mathbf{r}_\mathbf{x}(\tau)$ for $\tau \in \{\tau_1, \tau_2\}$ is as follows [9, Ch. 7.4.1]:

1) Compute matrix $\mathbf{W} = \mathbf{r}_\mathbf{x}(\tau_1)^{-1/2}/N$ by using a received block of length N;
2) Compute an eigendecomposition $\mathbf{\Lambda}\text{diag}(\lambda_1, \lambda_2)\mathbf{\Lambda}^{-1} = \mathbf{W}\mathbf{r}_\mathbf{x}(\tau_2)\mathbf{W}^\dagger$; and
3) Compute the inverse mixing or the separation matrix $\hat{\mathbf{G}} = \mathbf{\Lambda}^\dagger\mathbf{W}$.

We remark that we could in principle also jointly diagonalize $\mathbf{r}_\mathbf{x}(\tau_1)$ and $\mathbf{r}_\mathbf{x}(\tau_2)$ for other values than $\tau_1 = 0$ and $\tau_2 = 1$. Furthermore, note that the transmit filters ensure that $\mathbf{r}_\mathbf{x}(1)$ is invertible. The inverse mixing matrix $\hat{\mathbf{G}}$ is an inverse of the RSOP channel mixing matrix up to the ambiguity of complex rotation and order. Note that these ambiguities are also present in other blind algorithms such as CMA.

By applying $\hat{\mathbf{G}}$ to the mixed signal, we recover an intermediate signal $\mathbf{q} = \hat{\mathbf{G}}\mathbf{x}$. Then, a MMA equalizer can recover the transmit PCS-QAM symbols $\hat{\mathbf{z}}$ by removing residual channel distortions

as well as the transmit filters $\mathbf{f}$ from $\mathbf{q}$. It was shown in [3] that a variation of the MMA equalizer that accounts for the probabilistic distribution of the transmit symbols performs better than the standard MMA [13], [14]. From here onwards, we refer to the standard MMA as MMA in the text. For readers who are not familiar with the probability aware-MMA, we explain it briefly below for completeness.

Source probability aware MMA (pr-MMA)

Similar to MMA, the pr-MMA uses a 2×2 multi-input multi-output (MIMO) structure for adaptive equalization. In each iteration, blocks of input signals are processed to produce the two equalizer output symbols. Then, each of the output symbols is assigned a radius based on the closest QAM symbol. An error is computed as the difference in the absolute squares of the equalizer output and the assigned radius. Finally, the equalizer taps are updated using the computed errors.

In a multi-level constellation, symbols can be categorized based on their magnitudes (i.e., QAM moduli or radii). The pr-MMA in [3] uses the fact that the probabilities of a symbol having a given modulus are not uniformly distributed and should be accounted for in the update equation. This can be achieved by multiplying the error with the probability of the corresponding assigned modulus. Note that these probabilities will vary with the entropy of the PCS formats. Furthermore, assigning a modulus closest to the magnitude of the equalizer output is not optimal in presence of noise [3]. Therefore, the thresholds used to assign the radius can also be optimized. However, that requires knowledge of the signal to noise ratio. In our pr-MMA implementation, only the errors are weighted (i.e. multiplied) by the probability of the corresponding assigned modulus. The equalizer outputs are assigned to the closest modulus of the constellation.

A shortcoming of the JD method is that it is only suitable for a memoryless channel since if $\mathbf{H}$ is not memoryless then (6) does not hold true. As a result, for an RSOP channel the JD-pr-MMA works well, while in the presence of PMD effects the algorithm can fail. In [5], we found that a similar algorithm, JD-MMA, could tolerate weak PMD effects (DGD values up to 10% of the symbol duration). In the next section, we propose a method that can be used to separate PCS-QAM symbols in the presence of stronger PMD impairments.

IV. Proposed FDJD-Pr-MMA for Convolutive Mixing

The frequency-domain description (4) of the PMD effects can be interpreted as many parallel channels that all suffer from instantaneous mixing only. A method that uses this fact for blind-identification of a convolutive mixing channel was proposed in [19]. We refer to it here as a FDJD or frequency domain (FD) JD method. Although the original FDJD [19] is suitable only for a two-input-two-output channel that has only cross-mixing (i.e. $H_{11}(\omega) = H_{22}(\omega) = 1$), we can still apply this method to the PMD channel and combine it with other processing steps to propose a method that works for general PMD channels.

The underlying principle of the FDJD algorithm is described in [19]. Similar to JD, the FDJD algorithm makes several assumptions. The assumptions on the transmit signals are the same as for the JD-pr-MMA algorithm in Section III, with the only difference being that both of transmit signals need to exhibit temporal correlations. Like JD-pr-MMA, the assumption on transmit signal $\mathbf{s}$ is met in our setup by filtering the PCS-QAM symbols $\mathbf{z}$ before transmission, i.e $\mathbf{s} = \mathbf{f} \star \mathbf{x}$. Additionally, the FDJD algorithm requires assumptions about the channel:

A2.1 The channel is a linear time-invariant filter with a finite impulse response (FIR).
A2.2 The auto-channels are delta pulses, i.e. the diagonal entries satisfy $H_{ii}(\omega) = 1$ for $i \in \{1, 2\}$.
A2.3 The cross-channels have no common zeros.
A2.4 Each cross-channel has no zeros in conjugate reciprocal pairs or, equivalently, it does not contain zero-phase convolutional components.

The assumption A2.1 is not restrictive as any stable infinite impulse response (IIR) filter can be approximated by an FIR filter of appropriate length. The assumption A2.2 is not true for a PMD channel in general because the auto-channels are not impulses, i.e. $H_{ii}(\omega) \neq 1$ for $i \in 1, 2$. It was shown in [19] that the FDJD algorithm can still be applied when A2.2 is not satisfied. In that case, the estimated responses will be normalized cross-channels instead of true cross-channels. That is, instead of estimating the true channel responses $\hat{H}_{12}(\omega)$ and $\hat{H}_{21}(\omega)$, the FDJD algorithm estimates

$$\hat{H}'_{12}(\omega) = \frac{\hat{H}_{12}(\omega)}{\hat{H}_{22}(\omega)} \quad \text{and} \quad \hat{H}'_{21}(\omega) = \frac{\hat{H}_{21}(\omega)}{\hat{H}_{11}(\omega)}, \tag{7}$$

respectively.

We observed that the assumption A2.4 may not be true for the PMD channel in general. This assumption is required for solving the order ambiguity (explained later in Section V). Fortunately, it is not necessary when the temporal correlations of the transmit signals are known [19], as it is the case in our setup.

A. FDJD Algorithm

The FDJD algorithm takes a block of the mixed signals and uses the knowledge of the transmit signals' correlations (i.e. transmit filters $\mathbf{f}$) to provide the normalized cross channel response given in (7).

Similar to the JD algorithm in Section III, this algorithm is based on the insight that the channel matrix $\mathbf{H}(\omega)$ jointly diagonalizes the matrices $\mathbf{C_Y}(\omega)$ and $\mathbf{C_Y}(\omega, \omega + 1)$, where $\mathbf{C_Y}(\omega_1, \omega_2) = \mathbf{R_Y}(\omega_1, \omega_2)\mathbf{R_Y}(\omega_1, \omega_2)^\dagger$, $\mathbf{R_Y}(\omega_1, \omega_2) = \mathbb{E}\{\mathbf{Y}(\omega_1), \mathbf{Y}(\omega_2)\}$ [19, Eq. 21]. Here, $\mathbf{Y}(\omega) = \mathbf{V}(\omega)\mathbf{X}(\omega)$ denotes the whitened signal and $\mathbf{V}(\omega)$ the pre-whitening operator.

Before explaining the algorithm, let us define a $N_f \times N_f$ DFT matrix

$$\mathcal{F} \triangleq e^{-j\frac{2\pi}{N_f}\mathbf{u}\mathbf{u}^\dagger} \tag{8}$$

for $\mathbf{u} = [0, 1, \ldots, N_f - 1]^T$ and a $N_f \times N_f$ Kaiser window matrix

$$\mathbf{K} = K^\dagger K, \tag{9}$$

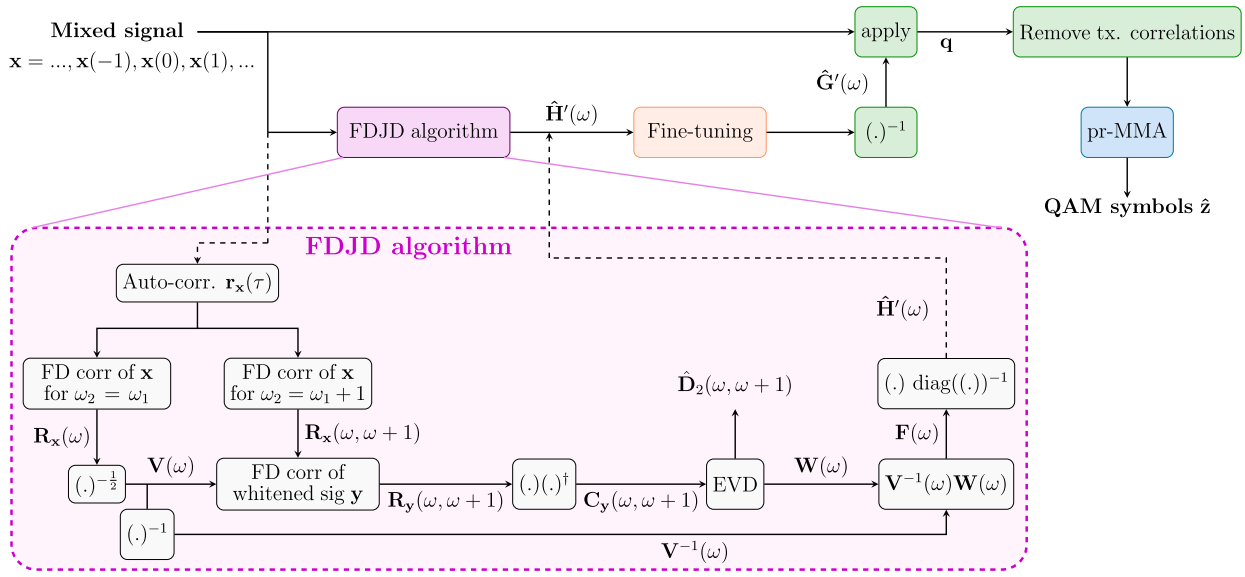

Fig. 2. Schematic of FDJD-pr-MMA. The FDJD algorithm blindly estimates the PMD channel's normalized response $\hat{\mathbf{H}}'(\omega)$. After ambiguity removal and reconstruction, an inverse of this response $\hat{\mathbf{G}}'(\omega)$ is applied, which removes the cross mixing components of the PMD channel. Finally, the temporal correlations are removed and PCS-QAM symbols are recovered using pr-MMA.

with K being a Kaiser window of dimension $1 \times N_f$.

The algorithm is as follows (also see Fig. 2)

1) Compute the auto-correlation $\mathbf{r}_\mathbf{x}(\tau)$ for $\tau \in [-N_f, N_f]$ of a block of the mixed signal $\mathbf{x} = [\mathbf{x}(0), \mathbf{x}(1), \ldots, \mathbf{x}(N-1), \mathbf{x}(N)]$ using (6).
2) Compute two frequency domain (FD) correlation matrices $\mathbf{R}_\mathbf{X}(\omega, \omega) = \mathbf{R}_\mathbf{X}(\omega)$ and $\mathbf{R}_\mathbf{X}(\omega, \omega+1)$ with $\omega \in \{0, 1, 2, \ldots, N_f - 1\}$, using
$$R_X^{ij}(\omega_1, \omega_2) = \mathcal{F}(\omega_1, :)\, \mathbf{K}\, \mathbf{r}^{x_i, x_j}\, \mathcal{F}(\omega_2, :)^\dagger, \tag{10}$$
where $\mathbf{r}^{x_i, x_j}$ denotes the Toeplitz cross-correlation matrix of size $N_f \times N_f$ derived from $\mathbf{r}_\mathbf{x}(\tau)$ and $\mathcal{F}(\omega_1, :)$ denotes the ω_1th row of the matrix $\mathcal{F}$.
3) Compute the pre-whitening operator
$$\mathbf{V}(\omega) = \mathbf{R}_\mathbf{X}^{-1/2}(\omega). \tag{11}$$
4) Next, compute the FD correlation matrix of the whitened signal $\mathbf{y}(t)$ given by $\mathbf{Y}(\omega) = \mathbf{V}(\omega)\mathbf{X}(\omega)$ as
$$\mathbf{R}_\mathbf{Y}(\omega, \omega+1) = \mathbf{V}(\omega)\, \mathbf{R}_\mathbf{X}(\omega, \omega+1)\, \mathbf{V}^\dagger(\omega+1). \tag{12}$$
5) Compute the matrix
$$\mathbf{C}_\mathbf{Y}(\omega, \omega+1) \triangleq \mathbf{R}_\mathbf{Y}(\omega, \omega+1)\mathbf{R}_\mathbf{Y}{}^\dagger(\omega, \omega+1). \tag{13}$$
6) The eigenvalue decomposition $\mathbf{C}_\mathbf{Y}(\omega, \omega+1)$
$$\mathbf{C}_\mathbf{Y}(\omega, \omega+1) = \mathbf{W}(\omega)\hat{\mathbf{D}}_\mathbf{2}(\omega, \omega+1)\mathbf{W}^\dagger(\omega+1) \tag{14}$$
provides the column eigenvector matrix $\mathbf{W}(\omega)$.
7) The normalized MIMO channel response is given by
$$\hat{\mathbf{H}}'(\omega) = \mathbf{F}(\omega)\, \mathrm{diag}\left(\mathbf{F}(\omega)\right)^{-1}, \tag{15}$$
where
$$\mathbf{F}(\omega) \triangleq \mathbf{V}(\omega)^{-1}\mathbf{W}(\omega). \tag{16}$$

By applying the inverse of the normalized MIMO filters $\hat{\mathbf{H}}'(\omega)$ to the mixed signal $\mathbf{X}(\omega)$, the cross mixing components will be removed. However, a filter within each transmit signals remains (see [19, Sec. IV]), i.e.
$$\mathbf{Q}(\omega) = \hat{\mathbf{H}}'(\omega)^{-1}\mathbf{X}(\omega) = \begin{bmatrix} \hat{H}_{11}(\omega)s_1(\omega) \\ \hat{H}_{22}(\omega)s_2(\omega) \end{bmatrix}. \tag{17}$$
In order to recover the PCS-QAM symbols $\hat{\mathbf{z}}$, the filters $\hat{H}_{11}(\omega)$ and $\hat{H}_{22}(\omega)$ and the transmit filters $\mathbf{f}(t)$ need to be removed. As we assume that the transmit filters are known, their inverse can be applied to the signal right after removing the PMD channel's cross-component using (17). Finally, a MMA equalizer can be used to remove filters remaining from the FDJD algorithm as well as any distortions.

The FDJD-pr-MMA requires three additional processing blocks to resolve the order ambiguity and to improve FDJD estimates and pr-MMA convergence. We explain them in the next section.

V. Additional Processing Blocks

The three blocks of processing pertain to ambiguity removal, reconstruction of the normalized PMD channel response, and candidate phase response search (CPRS). The first block is needed since the estimated normalized response $\hat{\mathbf{H}}'(\omega)$ from the FDJD algorithm is not necessarily equal to the true normalized channel, and can contain ambiguities similar to polarization switching in CMA. This block is described in subsection V-A. After ambiguity removal, the next block performs reconstruction of the normalized PMD channel as described in subsection V-B. This block discards frequency bins that contain relatively large estimation errors and reconstructs the FDJD estimated response. These two blocks are applied post-FDJD algorithm, and the last block, CPRS, is applied right before pr-MMA. The CPRS

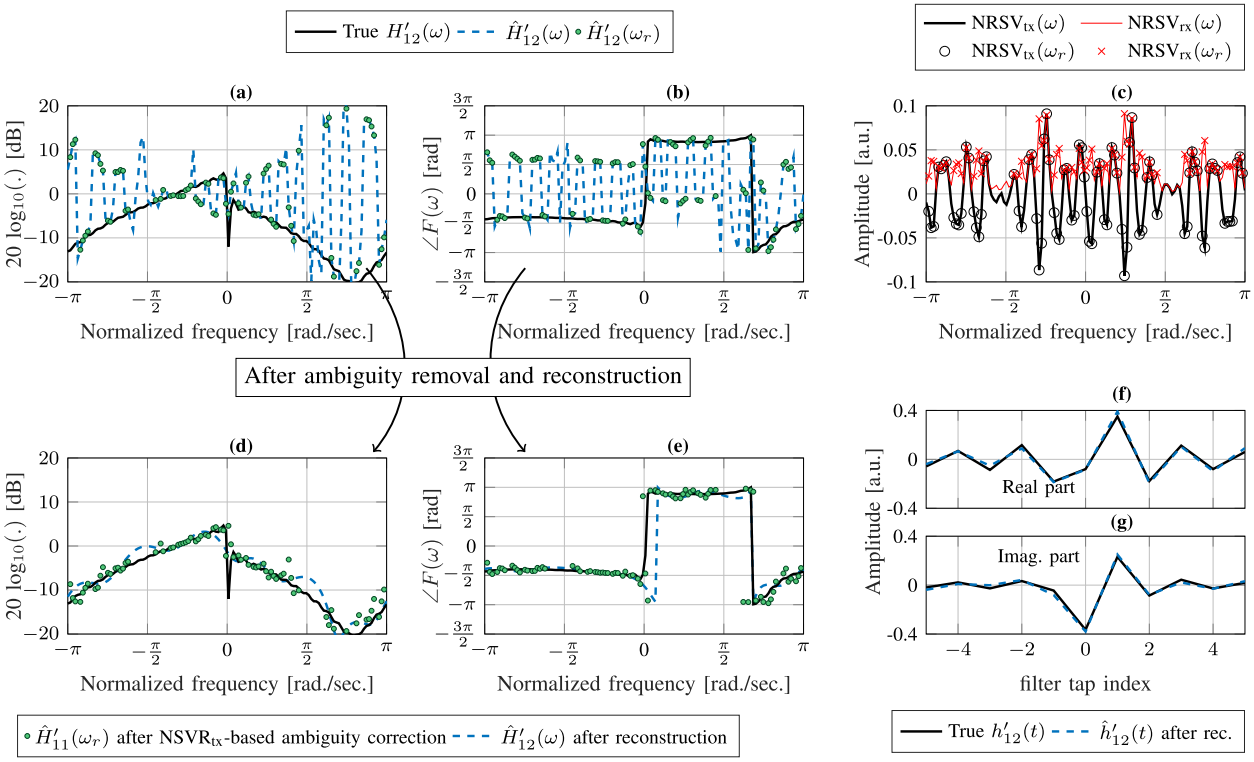

Fig. 3. Example of ambiguity removal and reconstruction of the normalized PMD channel's response estimates obtained through the FDJD algorithm: (a), (b) raw estimates, (c) NRSV based selection and decision on reliable frequencies ω_r, (d) and (e) reconstructed channel responses after removing ambiguity, and (f) true and estimated impulse responses.

block was added because, during evaluation, we found that in some cases pr-MMA would fail in equalizing for the PMD auto-channels, i.e. $H_{11}(\omega)$ and $H_{22}(\omega)$ as given in (17), despite FDJD estimates being quite good. The CPRS block improved convergence of the pr-MMA, which will be elaborated on in the Results section. The working of the CPRS block is provided in Section V-C.

A. Ambiguity Removal

The ambiguity in FDJD estimated responses tends to be more severe than that present in algorithms such as CMA or JD-(pr-)MMA. As the FDJD algorithm is applied in the frequency domain, the order ambiguity is present in every frequency component. Because the PMD channel is a two-input-two-output system, at any given ω the estimated normalized response will be either [19]

$$\hat{\mathbf{H}}'(\omega) = \begin{bmatrix} 1 & \hat{H}'_{12}(\omega) \\ \hat{H}'_{21}(\omega) & 1 \end{bmatrix} \tag{18}$$

or

$$\hat{\mathbf{H}}'(\omega) = \begin{bmatrix} 1 & 1/\hat{H}'_{21}(\omega) \\ 1/\hat{H}'_{12}(\omega) & 1 \end{bmatrix}. \tag{19}$$

This frequency component switching problem is visualized in Fig. 3, where we plot 128 point DFTs of normalized PMD channel response. In Fig. 3(a) and (b), the curve in black shows the underlying true response $H'_{12}(\omega)$. The blue dashed line as well as the green circles show the FDJD estimated response $\hat{H}'_{12}(\omega)$. It is clear that at some frequency bins the estimated response is switched/flipped, which is just a change of sign on the log-scale of vertical axis in Fig. 3(a).

This ambiguity on the frequency component can be resolved by using assumption A2.4 or by the knowledge on the transmit signals correlations $\mathbf{r_s}(\tau)$ [19]. We utilize the latter. We know that PCS-QAM symbols $\mathbf{z}$ are iid. By assuming $\mathbf{z}$ is normalized, it can be shown that the transmit signal correlations $\mathbf{r_s}(\tau) = \mathbf{r_f}(\tau)$, are known at the receiver since the transmit filters $\mathbf{f}$ are known. From that, the FD correlations of $\mathbf{f}$, i.e. $\mathbf{R_f}(\omega)$ and $\mathbf{R_f}(\omega, \omega+1)$, can be obtained. Finally, the matrix $\mathbf{D_2}(\omega, \omega+1)$ can be computed using $\mathbf{R_y} = \mathbf{R_f}$ in (13) and (14). The diagonal entries of the matrix $\hat{\mathbf{D}}_\mathbf{2}(\omega, \omega+1)$ are the real and positive eigenvalues of $\mathbf{C_f}(\omega, \omega+1)$. By comparing the true order of eigenvalues in $\mathbf{D_2}$ against the estimated order in $\hat{\mathbf{D}}_\mathbf{2}$, we can rearrange the estimated $\mathbf{W}(\omega)$ and thus determine if $\hat{\mathbf{H}}'(\omega)$ at the given ω is from relation (18) or (19).

We observed that it was not straightforward to resolve the frequency components' switching in the presence of estimation errors. Therefore, we select a set of reliable frequencies ω_r and correct the frequency components' switching of them. The selection of the reliable frequencies ω_r is based on a quantity called a normalized ratio of singular values (NRSV) [19], defined as

$$\text{NRSV}_{\text{tx}}(\omega) = \frac{|\mathbf{D_2}^{11}(\omega) - \mathbf{D_2}^{22}(\omega)|}{\max\left(\mathbf{D_2}^{11}(\omega), \mathbf{D_2}^{22}(\omega)\right)}, \tag{20}$$

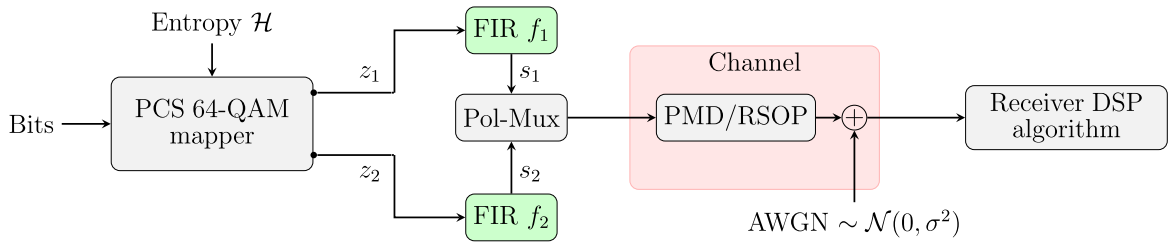

Fig. 4. Simulation setup. Receiver DSP algorithm candidates are JD-(pr-)MMA, the proposed FDJD-pr-MMA, and CMA-(pr-)MMA to benchmark against. The FIR filters f_1 and f_2 are only used for JD-(pr-)MMA and FDJD-pr-MMA. In other words, for CMA-(pr-)MMA, $f_1(t)$ and $f_2(t)$ are Dirac impulses $\delta(t)$.

$$\mathrm{NRSV}_{\mathrm{rx}}(\omega) = \frac{|\hat{\mathbf{D}}_{\mathbf{2}}^{11}(\omega) - \hat{\mathbf{D}}_{\mathbf{2}}^{22}(\omega)|}{\max\left(\hat{\mathbf{D}}_{\mathbf{2}}^{11}(\omega), \hat{\mathbf{D}}_{\mathbf{2}}^{22}(\omega)\right)}. \tag{21}$$

For a given threshold $\mathrm{NRSV}_{\mathrm{th}}$, frequencies for which $\mathrm{NRSV}_{\mathrm{rx}}(\omega) > \mathrm{NRSV}_{\mathrm{th}}$ are added to the set of reliable frequencies. Choosing a large threshold $\mathrm{NRSV}_{\mathrm{th}}$ seems to require the transmit filters $\mathbf{f}(t)$ to be more frequency selective, while choosing a small threshold requires an accurate estimation of the statistics of the mixed signal. In our simulations, we chose not to make the transmit filter more frequency selective and used a small value of 0.02 for $\mathrm{NRSV}_{\mathrm{th}}$. In addition to $\mathrm{NRSV}_{\mathrm{th}}$, more restrictions specific to the PMD channel are imposed for selecting reliable frequencies. From the PMD model described in Section II, it can be inferred that $|H_{11}(\omega)| = |H_{22}(\omega)|$ and $|H_{12}(\omega)| = |H_{21}(\omega)|$ and thus, $|\hat{H'}_{21}(\omega)| = |\hat{H'}_{12}(\omega)|$. Their ratio $|I_1(\omega)|$, defined as

$$\hat{I}_1(\omega) = \frac{\hat{H}'_{21}(\omega)}{\hat{H}'_{12}(\omega)}, \tag{22}$$

should be 1 at all ω. The frequencies for which $|I_1(\omega) - 1|$ is larger than a set threshold $I_{1,tol.}$ (chosen as 0.2 in our case) are also discarded. The $\mathrm{NRSV}_{\mathrm{tx}}(\omega)$ and $\mathrm{NRSV}_{\mathrm{rx}}(\omega)$ are shown in Fig. 3(c). There values at reliable frequencies ω_r are indicated by markers. At frequency bins where $\mathrm{NRSV}_{\mathrm{tx}}(\omega_r)$ and $\mathrm{NRSV}_{\mathrm{rx}}(\omega_r)$ are not in agreement (i.e. positive/negative), the frequency components are switched. In other words, at those ω values, FDJD estimated the solution from (19), which can be corrected to (18) by a reciprocal operation followed by a swap operation on $\hat{H}'_{12}(\omega)$ and $\hat{H}'_{21}(\omega)$.

An illustration of the corrected spectrum is shown in Fig. 3(d) and (e). The frequency switching problem shown in Fig. 3(a) and (b) is corrected at the reliable ω_r bins.

B. Reconstruction of the Normalized PMD Channel Response

After removing ambiguity at reliable frequencies, the channel response is reconstructed using the following iterative procedure.

Let us assume that the memory-length of the normalized cross-channel impulse responses $\hat{h'}_{12}(n)$ and $\hat{h'}_{21}(n)$ is roughly known and referred to as estimator memory. We carry out multiple iterations of the following four steps: First, an IDFT operation is applied on $\hat{H'}_{12}(\omega)$ and $\hat{H'}_{21}(\omega)$. Then, memory of the obtained time domain impulse responses $\hat{h'}_{12}(n)$ and $\hat{h'}_{21}(n)$ is limited to the estimator memory length, i.e. taps beyond the estimator memory are forced to zero. In the third step, a DFT operation is applied to get frequency responses. Finally, the magnitude responses are constrained to be equal at ω_r, i.e. $|\hat{H'}_{21}(\omega_r)| = |\hat{H'}_{12}(\omega_r)|$. After a few iterations, an improved estimate of the normalized channel responses $\hat{H'}_{12}(\omega)$ and $\hat{H'}_{21}(\omega)$ is obtained. An example of the reconstructed spectrum is shown in Fig. 3(d) and (e) along with their corresponding impulse response in Fig. 3(f) and (g). The MIMO inverse of the normalized channel response is finally given by $\hat{\mathbf{G}}(\omega) = \hat{\mathbf{H}}'^{-1}(\omega)$.

C. Candidate Phase Response Search

Ideally, the task of the pr-MMA in the FDJD-pr-MMA is reduced to finding the phase response of $G_{11}(\omega) = H_{11}^{-1}(\omega)$ and $G_{22}(\omega) = H_{22}^{-1}(\omega)$ because the magnitude responses $|G_{11}(\omega)| = |H_{22}^{-1}(\omega)|$ can be calculated from the normalized cross channel responses $\hat{H}_{21}(\omega)$ and $\hat{H}_{12}(\omega)$ estimated by the FDJD algorithm. The PMD channel matrix $\mathbf{H}(\omega)$ is a 2×2 unitary matrix for every ω. Then, algebraic manipulation shows that $|G_{11}(\omega)| = |H_{11}^{-1}(\omega)| = \sqrt{1 - H_{12}(\omega)H_{21}(\omega)}$.

As pr-MMA failed often for larger DGD, we added an additional processing block before the pr-MMA, called the candidate phase response search (CPPS). CPRS takes a phase response $\angle G_{11}(\omega)$ from 27 pre-chosen phase responses, combines it with $|G_{11}(\omega)|$ to make a candidate $G_{11}(\omega)$ and applies this to the signal. Then, the cost function of the pr-MMA cost is measured. The candidate that achieves the minimum cost function is applied on the signal prior to pr-MMA. This filter was applied in the time domain, on the first polarization, by doing an IDFT operation and limiting memory to (estimator memory)/2. Similarly, another filter was derived and applied to the signal in the second polarization. We observed that after applying these inverse filters to the signal, the PCS-QAM constellation starts appearing. Feeding the filtered signal to the pr-MMA recovers the PCS-QAM symbols in most cases where pr-MMA failed previously.

VI. Simulation Setup

The block diagram of the simulation setup is shown in the Fig. 4. At the transmitter, a PCS-64QAM mapper is used to generate two 32 GBaud sequences z_1 and z_2 of length N_{symb} according the Maxwell-Boltzmann distribution for the given entropy $\mathcal{H}$. Next, the sequences of symbols are passed through transmit filters that add time correlations to the symbols. The filtered signals s_1 and s_2 are multiplexed and transmitted through

the channel. The channel was simulated as a cascade of the PMD/RSOP model (explained in Section II), followed by addition of noise from an additive white Gaussian noise (AWGN) source to the mixed signal. The DGD parameter is provided as a percentage of the symbol duration. At the receiver, candidate DSP algorithms are applied to recover the transmit PCS-64QAM symbols. Then, the performance is computed in terms of 2-D mutual information (MI) bits.

The receiver DSP algorithm candidates are our previously proposed JD-MMA, the newly proposed JD-pr-MMA and FDJD-pr-MMA, and CMA-(pr-)MMA, which serves as a reference for our benchmark. The transmit filters are introduced to satisfy the assumptions involved in JD-(pr-)MMA and FDJD-pr-MMA algorithms. For CMA-(pr-)MMA, these filters are not needed and therefore omitted from the signal path or rather treated as Dirac impulses $f_1(t) = f_2(t) = \delta(t)$.

It should be noted that simulation setup as well as algorithms under investigation are implemented at 1 sample per symbol (sps).

VII. Results

We carried out a comparative study on the performance of the reference CMA-(pr-)MMA, the improved JD-(pr-)MMA for memoryless channels described in Section III, and the new FDJD-MMA from Section IV. We first present results for the JD-(pr-)MMA, and then show the performance of the new FDJD-pr-MMA.

A. JD-Pr-MMA (RSOP Channel and Low PMD)

Here, we compare the gains obtained by replacing MMA by pr-MMA. We use the same configuration that was used in our previous paper [5]. That is, we chose $N_{\text{symb}} = 2^{16}$ and $f_1(t) = [1, 0.05 + 0.05j]$, $f_2(t) = \delta(t)$ for the JD-(pr-)MMA.

The number of taps for (pr-)MMA is fixed to 11. The learning rate of CMA and MMA was 5×10^{-3} and 5×10^{-4}, respectively. The learning rate of pr-MMA was scaled up according to the probability distribution of the transmit signal constellation such that on average the step size of pr-MMA matches that of MMA. The length of the training sequence is fixed at 2^{14}, and 101 different random realizations of the channel were studied. In the figures, the average of the MI values calculated for each realization and each polarization is reported, unless stated otherwise.

First, we compare the CMA-(pr-)MMA and JD-(pr-)MMA on the RSOP channel. In Fig. 5(a), the average MI for CMA-(pr-)MMA is plotted against the SNR per polarization. For reference, the AWGN channel spectral efficiency (SE) is plotted as well. We observe that CMA-MMA performs close to the AWGN bound for SNRs up to 18 and then saturates at 6 bits for entropy $\mathcal{H} = 6$, i.e. uniform 64-QAM signal, which is expected. Next, we look at the performance for shaped QAMs of various entropy ranges. We observe that the average MI performance for entropies $\mathcal{H} \neq 6$ is worse compared to $\mathcal{H} = 6$. For example, at higher SNR > 20 dB, the average MIs are not reaching the corresponding entropy value, but instead stay below 3 bits. By looking into the MI for each realization, we found that for only around half of the

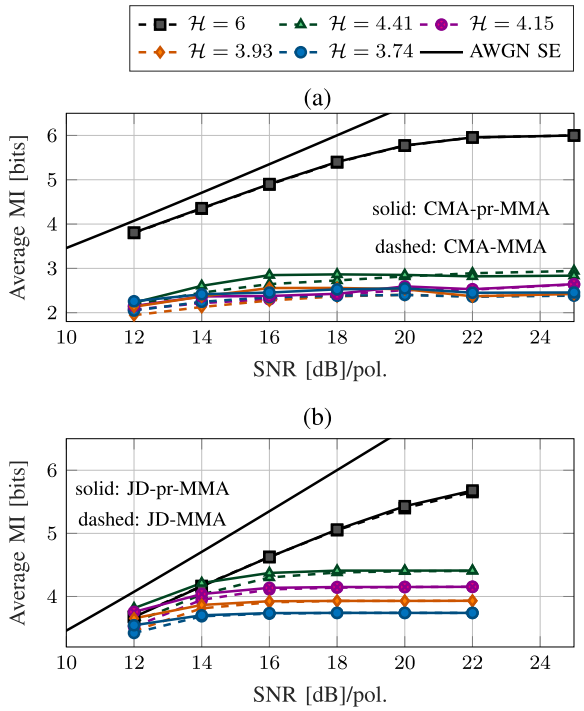

Fig. 5. Average MI of (a) CMA-(pr-)MMA and (b) JD-(pr-)MMA for the RSOP channel. $\mathcal{H}$ denotes the entropy of the PCS-64QAM format. Average over 101 realizations and two polarizations.

realizations of shaped 64-QAM reach closer to their entropy value at SNR 20 dB. The rest of the realizations produced MI < 3 bits with some producing MI of only around 1 bit. These realizations were not successfully equalized by CMA-(pr-)MMA.

Now, we look at the performance of JD-MMA, which is shown in Fig. 5(b). An average MI > 3 bits is observed for all the entropies in the considered SNR range. We also see that average MI values reach their corresponding entropy values at higher SNRs(> 18 dB), indicating successful equalizations in all the cases. A noticeable difference in the average MI for uniform 64-QAM is observed (for e.g., at SNR $= 20$ dB) when compared with CMA-MMA. This difference can be reduced by running more training iterations of MMA. We also observe that using JD-pr-MMA instead of JD-MMA increases the average MI in the smaller SNR range (< 15 dB). These results shows that using pr-MMA helped in improving convergence of both CMA-MMA and JD-MMA. Therefore, we use the pr-MMA instead of MMA for the results in the following part of the article.

The RSOP channel is an instantaneous mixing case for which JD-(pr-)MMA should always work in the absence of noise. Next, we test these algorithms for a PMD channel. We set a DGD value of 0.1 T_{symb} and evaluate the performance. The results are shown in Fig. 6. A similar trend is observed in average MI performance of CMA-pr-MMA and JD-pr-MMA. Though when compared with the RSOP channel, both CMA-pr-MMA and JD-pr-MMA achieves sligthly lower MI because the channel memory is increased with DGD.

When we further increase the DGD parameter of the PMD channel, we found that a memory of 11 taps is no longer sufficient. Therefore, we increased the MMA memory to 31

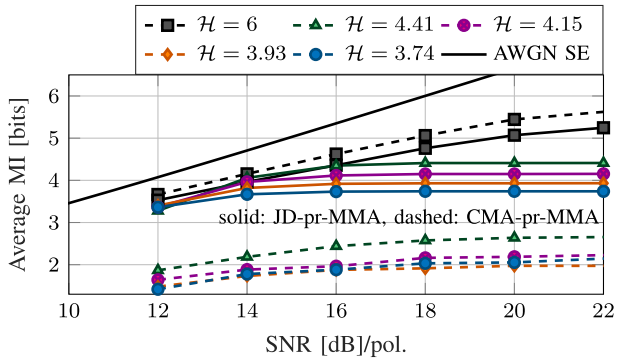

Fig. 6. Average MI of CMA-pr-MMA and JD-pr-MMA for DGD = 10% of symbol duration T_{symb}, averaged over 101 realizations of two polarizations.

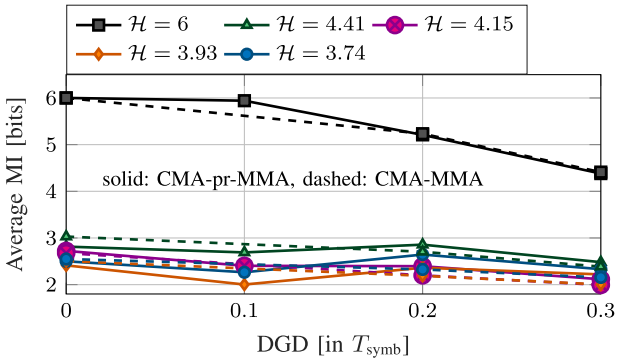

Fig. 7. Average MI of CMA-(pr-)MMA for various DGD values for SNR = 30 dB.

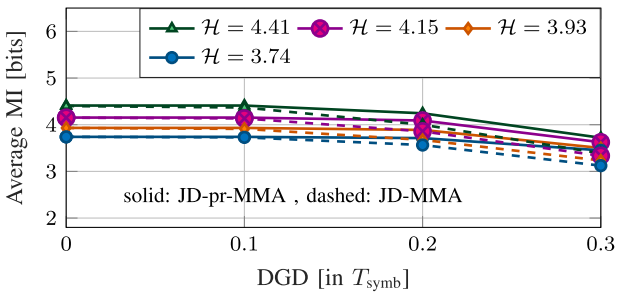

Fig. 8. Average MI of JD-(pr-)MMA for various DGD values for SNR = 22 dB.

taps. The average MI for CMA-(pr-)MMA for a fixed SNR of 30 dB and varying DGD values are shown in Fig. 7. We see that CMA-MMA only achieves average MI >3 bits for the case of uniform QAM signals. For shaped QAMs, average MI is <3 bits. On observing MI for each PMD channel realization, we found more cases of unsuccessful equalization relative to the RSOP channel.

The average MI of JD-(pr-)MMA for a fixed SNR of 22 dB and varying DGD values is shown in Fig. 8. We observe that the average MI obtained with JD-pr-MMA is reducing with increasing DGD values but it is more than 3 bits in all cases. We found that only around 2% of channel realizations gave MI < 3 bits at DGD $= 0.2$ T_{symb} and $\mathcal{H} = 4.15$ bits. On the other hand, this percentage was around 50% for CMA-MMA even for DGD ≤ 0.2 T_{symb}. We note that at higher DGD, JD-pr-MMA performance is not acceptable. But, for low DGD of $= 0.1$ T_{symb}, JD-pr-MMA obtained MI >3 bits for all the channel realizations and entropies. It shows that JD-pr-MMA is quite effective in compensating polarization impairments in the weak PMD distortions regime. For the strong PMD channel, we propose FDJD-pr-MMA, which is investigated next.

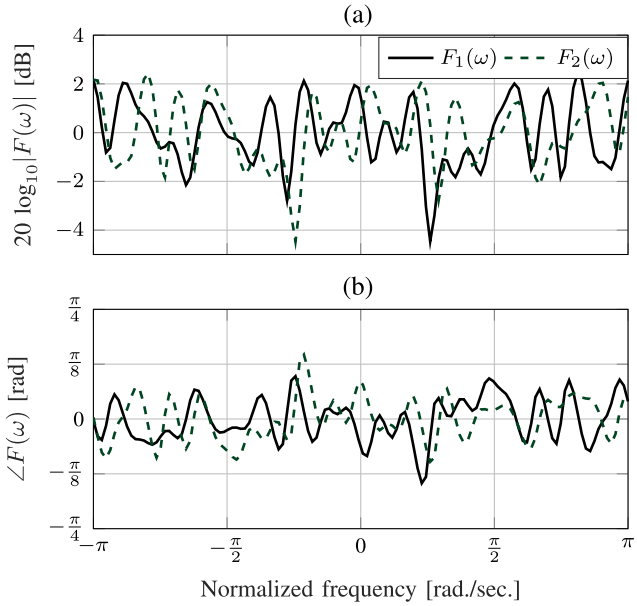

Fig. 9. Magnitude and phase responses of the transmit filters $\mathbf{f}$ used for FDJD-pr-MMA results. (a) Magnitude and (b) Phase responses of the transmit filters.

B. FDJD-Pr-MMA (Medium PMD)

For evaluation of the FDJD-pr-MMA, the first task was to design transmit filters f_1 and f_2. A transmit filter is considered well-fit if it produces reliable frequencies (explained in Section V), well-distributed over the spectrum, while not being too frequency selective. We found that if both filters are conjugate of each other, better channel estimations are obtained. We chose a 21 taps long transmit filter f_1 and accordingly $f_2 = f_1^{\dagger}$. The inverse of filter $\mathbf{f}$ was truncated to 51 taps. We remark that the chosen filters and memory length parameters are hand-tuned.

The frequency responses of the transmit filters are plotted in Fig. 9. It is visible that the magnitude responses have many fluctuations but most of them are within ± 2 dB and nearly all of them are within -4.5 dB and +3 dB. Furthermore, the phase responses are also well within $\pm\frac{\pi}{4}$.

We also found that an accurate estimation of the normalized cross-channel impulse responses is required by the FDJD algorithm or the subsequent steps may fail. A straightforward way is to increase the data length for the auto-correlation measurement. Therefore, we use a data length of 2^{18} symbols for FDJD, while the following MMA uses only 2^{14} symbols. During the evaluation of the FDJD-pr-MMA, we fix the FDJD estimator memory to 11 taps, $\text{NSVR}_{\text{th}} = 0.02$, $I_{1,tol.} = 0.2$, $N_f = 128$, and the Kaiser window parameter $\beta = 24$.

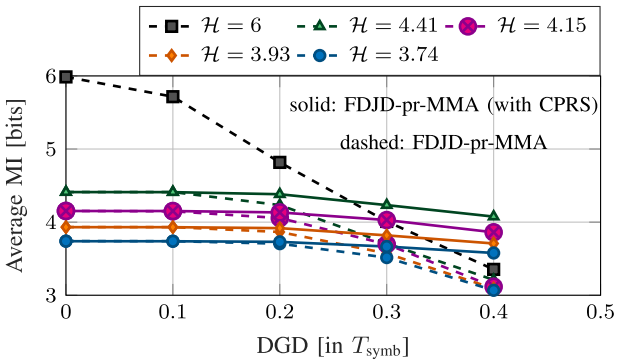

Fig. 10. Average MI of FDJD-pr-MMA for various DGD values and SNR = 30 dB. Adding a CPRS stage before pr-MMA helps in equalizing PCS-QAMs.

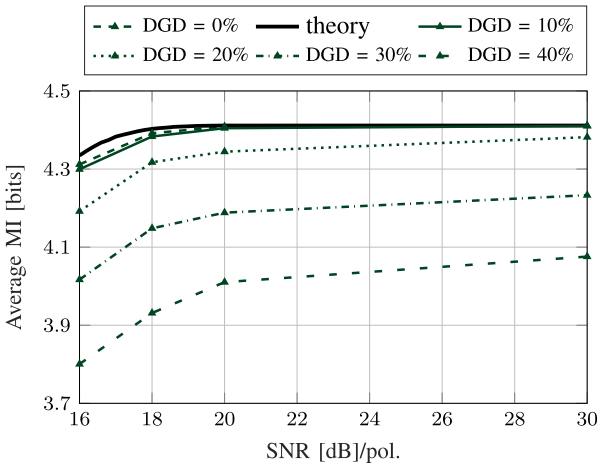

Fig. 11. FDJD-pr-MMA (with CPRS) performance: Average MI against SNR over various DGD values in % of symbol duration.

The average MI of FDJD-pr-MMA are shown in Fig. 10 for 30 dB SNR and varying DGD values. The dashed curves show the aveage MI for the FDJD-pr-MMA. We see that for DGD values $< 0.2T_{\text{symb}}$, an average MI close to the corresponding entropy values is obtained. The average MI decreases with increasing DGD values. This is because for more number of realizations, FDJD-pr-MMA failed to equalize. On investigating why the FDJD-pr-MMA algorithm failed for some random realizations, we found that despite the FDJD estimation of modified channel responses being quite good in those cases, the subsequent pr-MMA did not perform well. Therefore, we add the CPRS block, whose working is explained in V-C, right before pr-MMA.

In Fig. 10, we also plot the performance of FDJD-pr-MMA after including CPRS block. We observe that average MI is quite steady for DGD values up to 40% of symbol duration and a comparatively much slower decrease in MI is observed when the CPRS block is used. Furthermore, for DGD values $\leq 0.2T_{\text{symb}}$, none of the channel realization gave MI < 3 bits for $\mathcal{H} \neq 6$ bits. We also plot the average MI obtained by the FDJD-pr-MMA (with CPRS) in Fig. 11 for $\mathcal{H} = 4.41$ bits. A curve with MI obtained with an ideal demultiplexing is also plotted for reference. The average MI obtained for DGD $\leq 10\%$ of T_{symb} is

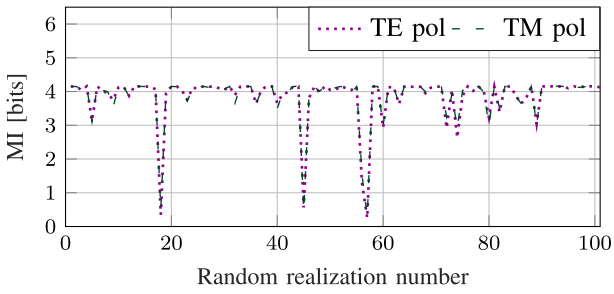

Fig. 12. FDJD-pr-MMA (with CPRS) performance: MI obtained for 101 random realizations (DGD = 0.4 T_{symb}, SNR = 30 dB, $\mathcal{H} = 4.15$ bits).

quite close to the theoretically expected value. We also observed that for all of the tested entropies, not plotted here for brevity, the performance is quite close to theory for low DGD ($\leq 0.1\text{T}_{\text{symb}}$). An increase in the DGD causes decrease in the average MI. At a DGD of $0.4\text{T}_{\text{symb}}$, a few ($\sim 7$) out of 101 random realizations achieved MI < 3 bits as shown in Fig. 12. We observe that MI values around 4 bits are obtained in most of the realizations; 3 out of 101 realizations gave MI less than 1 bit showing those realizations failed to equalize completely.

VIII. Discussion

In this study, we proposed two algorithms and demonstrated them on a simplified setup. The algorithms were evaluated at 1 sample per symbol and impairments such as phase noise and frequency offset were not included. In contrast, in a typical coherent receiver DSP, polarization demultiplexing algorithms usually operate at 2 sample per symbol in the presence of phase noise and frequency offset. Furthermore, the algorithms proposed in this work require filters at the transmitter, whose impact on the information rate is not yet clear. In this section, we discuss on these aspects together with the computational complexity of the algorithms.

A. Pulse Shaping

The proposed algorithms (and setup) require a few modifications for oversampling scenarios. There are two possibilities for adding temporal correlations at the transmitter: before or after the pulse shaping stage. The former approach seems preferable because it is easier to implement as the filter is applied to an iid sequence. The added correlations can be extracted at the receiver by feeding a down-sampled sequence (at 1 sps) to the JD and FDJD algorithm.

The subsequent stages of FDJD algorithm such as applying MIMO inverse $\hat{\mathbf{G}}(\omega)$ and inverse of the transmit filters and candidate phase response search (CPRS) need to account for the change in the sample rate. The inverse of modified channel response $\mathbf{G}^{-1}(\omega)$ and filters that remove temporal correlations $\mathbf{f}^{-1}$ are designed for the critically sampled sequence. Therefore, these filters need to be upsampled before applying them to the original oversampled sequence. After removing temporal correlations, a downsampled sequence can be fed to the CPRS stage. Then, the candidate response filter can be upsampled and applied to the original oversampled sequence. The MMA

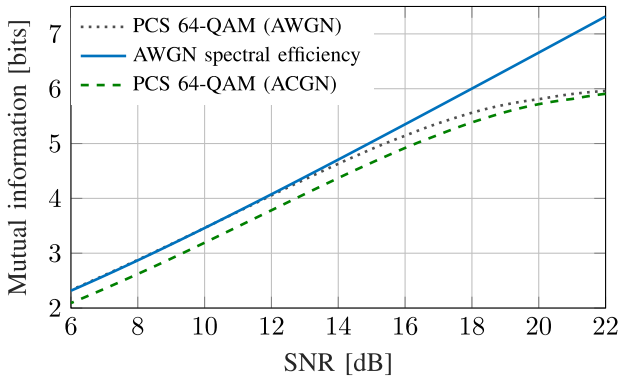

Fig. 13. Estimated MI over AWGN channel and ACGN channel (i.e. with FDJD filters).

equalizer can still be implemented with more than 1 sps and perform matched filtering.

B. Effect of Phase Noise and Frequency Offset

The proposed algorithms require the computation of cross-correlations of the mixed signal. When phase noise and frequency offset are present, the correlation measurement can be erroneous. For lasers of 100 kHz linewidth, which is often used in current coherent transceivers, We find that phase noise varies slowly over the considered correlation time-lags. Therefore, its impact is expected to be negligible on estimation. This was verified by quick simulations not presented here. The effect of a frequency offsets on the JD algorithm is also observed to be minimal. On the other hand, the effect of a frequency offset as small as 1 GHz is detrimental to FDJD algorithm. Therefore, a frequency offset correction is necessary for the FDJD algorithm. frequency offset correction is necessary before applying the FDJD algorithm.

C. Effect of Transmit Filters

Due to the presence of inverse filters $\mathbf{f}^{-1}$ at the receiver, the white Gaussian noise present in the received signal becomes colored. To quantify the impact of the transmit filter, we considered a simpler setup without any polarization impairments. In such a case, the end-to-end communication channel is a cascade of the transmit filters, an additive white Gaussian noise (AWGN) channel and the inverse transmit filters. This channel is equivalent to a additive colored Gaussian noise (ACGN) channel, where the colored noise arises from the applying the inverse filter at the receiver to the AWGN. In order to quantify the change in MI compared to AWGN channel, we run single polarization transmission simulations for two cases: one with and other without a transmit filter. The transmit filter used in this test is same as the one used in horizontal polarization of FDJD setup. The maximum MI achieved at given SNR is plotted in Fig 13. We see that maximum MI is reduced due to the transmit filters. In Fig. 14, we plot the difference in maximum MI achieved over ACGN channels and AWGN channels. We see that on average 0.21 bits of information are lost because of the filters. The maximum MI loss is 0.27 bits at 12 dB SNR per polarization.

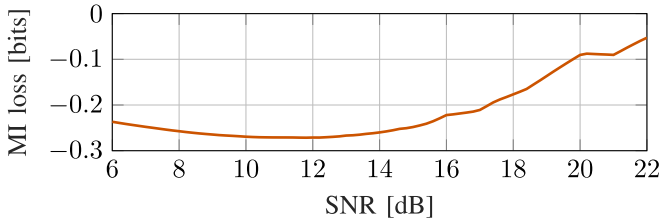

Fig. 14. Estimated MI loss: difference in MI for ACGN and AWGN channel for PCS-64QAM.

We note that these results are only for one polarization. The other polarization has a similar magnitude response and expected to have similar reduction in maximum MI. Moreover, we believe a detection scheme that could exploit the correlation present in the noise samples could improve MI performance.

D. Computational Complexity

The JD algorithm processes $N_t = 2^{14}$ (training length) received mixed symbols to determine a 2×2 instantaneous demixing matrix G. The dominant computation in this algorithm is the estimation of the correlations $r_x[\tau]$ ($\tau \in \{0, 1\}$), which requires $2 \times 4 \times N_t$ complex multiplications. Therefore, the time complexity of JD algorithm as $\mathcal{O}(8N_t)$.

The FDJD algorithm has many computational stages. Here, the major computational cost is due to step 1, 2 and CPRS stage. The overall complexity of these three steps scales with $\mathcal{O}(N_t \log_2 N_t + N_f^3 + N_{mma})$, where N_{mma} is the training length of the MMA.

IX. Conclusion

We investigated algorithms for blind polarization demultiplexing of PCS-QAM signals by inserting temporal correlations at the transmitter and exploiting them at the receiver. We proposed a novel algorithm called frequency domain joint diagonalization (FDJD) pr-MMA (FDJD-pr-MMA), which combines two algorithms, namely FDJD and pr-MMA. The first algorithm, FDJD, removes the cross mixing components of PMD channel and subsequently, the pr-MMA removes residual effects. A proof-of-concept simulation study is presented for a first-order PMD channel. We show that for moderate PMD effects, the FDJD-pr-MMA is well-suited, evaluating it on PMD channels with DGD up to 40% of the symbol duration (T_{symb}). The newly proposed FDJD-pr-MMA algorithm successfully equalizes PMD channels for DGD up to 20% of T_{symb}. Furthermore, we extended our previously proposed JD-MMA [5], which works ideally for RSOP channels, to improve its robustness to PMD effects.

We believe that FDJD estimates can be further improved by correcting the frequency-switching on non-reliable frequencies and by including them in the reconstruction of normalized channel responses. These algorithms' merit should be assessed in a more sophisticated setup using a higher-order PMD model and including other impairments that are often present in coherent transmission system. Finally, the effects of added temporal correlations lend themselves to further investigation as well.

ACKNOWLEDGMENT

The content is the sole responsibility of the authors and does not necessarily represent official views of the National Institutes of Health.

REFERENCES

[1] J. Cho and P. J. Winzer, "Probabilistic constellation shaping for optical fiber communications," *J. Lightw. Technol.*, vol. 37, no. 6, pp. 1590–1607, Mar. 2019.

[2] R. Johnson, P. Schniter, T. J. Endres, 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.

[3] S. Dris, S. Alreesh, and A. Richter, "Blind polarization demultiplexing and equalization of probabilistically shaped QAM," in *Proc. Opt. Fiber Commun. Conf.*, 2019, Paper W1D–2.

[4] S. A. Athuraliya and L. M. Garth, "Quantized CMA equalization for shaped signal constellations," *IEEE Signal Process. Lett.*, vol. 11, no. 2, pp. 67–70, Feb. 2004.

[5] V. Bajaj, R. V. d. Plas, V. Aref, and S. Wahls, "Blind polarization demultiplexing of probabilistically shaped signals," in *Proc. IEEE Photon. Conf.*, 2022, pp. 1–2.

[6] Q. Yan, L. Liu, and X. Hong, "Blind carrier frequency offset estimation in coherent optical communication systems with probabilistically shaped M-QAM," *J. Lightw. Technol.*, vol. 37, no. 23, pp. 5856–5866, Dec. 2019.

[7] F. A. Barbosa, S. M. Rossi, and D. A. A. Mello, "Phase and frequency recovery algorithms for probabilistically shaped transmission," *J. Lightw. Technol.*, vol. 38, no. 7, pp. 1827–1835, Apr. 2020.

[8] Q. Yan, C. Guo, and X. Hong, "The impact of probabilistic constellation shaping on channel equalization with constant modulus algorithm," in *Proc. Opto-Electron. Commun. Conf.*, 2021, pp. 1–3.

[9] P. Comon and C. Jutten, *Handbook of Blind Source Separation: Independent Component Analysis and Applications*. Cambridge, MA, USA: Academic Press, 2010.

[10] P. Zhang, Q. Yan, and X. Hong, "Probability-aware stokes space blind polarization demultiplexing for probabilistically shaped signals," *J. Lightw. Technol.*, vol. 39, no. 19, pp. 6120–6129, Oct. 2021.

[11] V. Lauinger, F. Buchali, and L. Schmalen, "Blind equalization and channel estimation in coherent optical communications using variational autoencoders," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 9, pp. 2529–2539, Sep. 2022.

[12] A. Belouchrani, K. Abed-Meraim, J. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *IEEE Trans. Signal Process.*, vol. 45, no. 2, pp. 434–444, Feb. 1997.

[13] M. J. Ready and R. P. Gooch, "Blind equalization based on radius directed adaptation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 1990, pp. 1699–1702.

[14] F. P. Guiomar et al., "Fully blind linear and nonlinear equalization for 100G PM-64QAM optical systems," *J. Lightw. Technol.*, vol. 33, no. 7, pp. 1265–1274, Apr. 2015. [Online]. Available: https://opg.optica.org/jlt/abstract.cfm?URI=jlt-33-7-1265

[15] A. Galtarossa and C. R. Menyuk, *Polarization Mode Dispersion*, vol. 296. Berlin, Germany: Springer, 2005.

[16] G. P. Agrawal, *Fiber Optic Communication Systems*. Hoboken, NJ, USA: Wiley, 2012.

[17] C. Xie and L. Möller, "The accuracy assessment of different polarization mode dispersion models," *Opt. Fiber Technol.*, vol. 12, no. 2, pp. 101–109, 2006.

[18] P. Serena, M. Bertolini, and A. Vannucci, "Optilux: An open source of light," 2009. [Online]. Available: http://www.optilux.sourceforge.net

[19] K. I. Diamantaras, A. P. Petropulu, and B. Chen, "Blind two-input-two-output FIR channel identification based on frequency domain second-order statistics," *IEEE Trans. Signal Process.*, vol. 48, no. 2, pp. 534–542, Feb. 2000.