Training-Aided Frequency-Domain Channel Estimation and Equalization for Single-Carrier Coherent Optical Transmission Systems

Fabio Pittalà, *Student Member, IEEE, Student Member, OSA*, Israa Slim, *Student Member, IEEE*,
Amine Mezghani, *Student Member, IEEE*, and Josef A. Nossek, *Life Fellow, IEEE*

Abstract—Frequency-domain equalization supported by training-aided 2×2 multi-input multi-output channel estimation (CE) is proposed for coherent optical transmission systems. In particular, novel training sequence (TS) schemes based on a constant amplitude zero auto-correlation code, efficient CE algorithms and equalizer-tap updating solutions are derived from the theoretical linear fiber channel model and investigated through simulations. The robustness of the proposed methods is demonstrated with respect to the main degrading optical propagation effects, such as amplified spontaneous emission noise, chromatic dispersion, polarization-mode dispersion, polarization-dependent loss, time misalignment between received and reference TS, frequency offset between transmitter and receiver lasers, as well as dynamic state of polarization rotation.

Index Terms—Channel estimation, coherent communications, digital signal processing, equalization, fiber-optics communications, training sequences.

I. INTRODUCTION

COHERENT optical demodulation in combination with digital signal processing (DSP) have revolutionized the design and operation of optical fiber communication systems [1]. Coherent detection provides the entire optical field within the electronic bandwidth of the receiver and consequently DSP allows for efficient mitigation of the main degrading optical propagation effects, namely chromatic dispersion (CD) and all-order polarization-mode dispersion (PMD) of an uncompensated linear or weakly nonlinear fiber channel [2].

The impact of the linear optical channel on a transmitted signal can be described as a multiplication of the signal spectra with the channel transfer function or by a convolution of the transmitted time-domain (TD) signal with the channel impulse response (CIR). In the digital coherent receiver, the according inverse operation could be performed relating to a de-convolution using a TD filter or a multiplication with an according frequency-domain (FD) filter. For TD filters with M taps, each filtered

symbol requires at least M complex multiplications such that the implementation complexity grows linearly with the filter length. In contrast, the according filter transfer function of same length in FD applies M parallel multiplications for M symbols such that each symbol requires only one complex multiplication. This is independent from the filter length. However, it must be considered that the received TD signal has to be transferred into the FD, which requires serial-parallel conversion with overlapping blocks and block-wise fast Fourier transform (FFT) and inverse FFT (IFFT) en-framing the filtering operation. Nonetheless, the additional complexity by the overlap, FFT and IFFT only grows logarithmically, such that FD filter implementation results in much lower implementation complexity compared to TD filtering. This is valid in particular for large CIR lengths that require long filters, (i.e., the CD compensating filter) [3].

Current digital coherent receivers widely use a hybrid FD/TD equalization approach. CD is a time-invariant effect, unless the optical path is switched, and does not create interference between polarizations. Therefore, the bulk of CD can be estimated once by blind algorithms [4]–[9] and compensated by a zero-forcing (ZF) FD equalizer (FDE) [9]–[11]. PMD effects are mitigated by a 2×2 multi-input multi-output (MIMO) TD equalizer (TDE) adapted by blind channel acquisition using gradient algorithms like constant-modulus algorithm (CMA), multi-modulus algorithm or decision-directed (DD) least mean square (LMS) [12] which converge to the minimum mean square error (MMSE) solution. Unfortunately, such filter updates are strongly dependent on the modulation format, which require complex implementation with individual cost functions for each modulation [13], [14] and suffer from a relatively slow convergence with potential sub-optimum acquisition and even failures [15], [16]. In contrast, FDE combined with training-aided (TA) channel estimation (CE) fulfills these requirements in many ways. This approach requires periodical transmission of a known training sequence (TS) slightly widening the spectrum of the transmitted signal to track time-varying polarization effects [17]. From each TS, a CE with instantaneous filter acquisition can be obtained [18], [19].

In this paper, a dual-stage FD/FD equalization approach is considered with main focus on TA adaptive 2×2 MIMO FDE employing periodically transmitted short TS for CE.

The paper is structured as following. In Section II the theoretical model of the 2×2 MIMO transmission channel is given. In Section III equalizer-tap updating methods based on the ZF and MMSE solutions are derived for two-fold oversampled

systems with and without considering that the received signal needs to be downsampled after equalization. In Section IV novel TS schemes based on a constant amplitude zero auto-correlation (CAZAC) code are illustrated in combination with efficient TA FD CE algorithms. Finally, Section V covers the performance investigation based on a 28-GBaud polarization-division multiplexed (PDM) quadrature phase-shift keying (QPSK) and 16-ary quadrature amplitude modulation (16-QAM) coherent optical systems transmitting return-to-zero (RZ), non-RZ (NRZ) or Nyquist shaped pulses. The robustness of the proposed methods is demonstrated with respect to amplified spontaneous emission (ASE) noise, CD, PMD, polarization-dependent loss (PDL), time misalignment between received and reference TS, frequency offset (FO) between transmitter and receiver lasers and dynamic state of polarization (SOP) rotation.

Notation: Bold letters indicates matrices, non-bold letters refer to vectors and non-bold italic letters express scalars. The complex-conjugate operation is denoted by $(\cdot)^*$, the transpose operation by $(\cdot)^T$, the complex-conjugate transpose operation by $(\cdot)^H$ and the expectation operation by $E[\cdot]$. The floor function denoting the largest integer less than or equal to x is expressed by $\lfloor x \rfloor$, the nearest integer by $\lfloor \cdot \rceil$, the convolution operation by $\otimes$ and the FFT operation by $\mathcal{F}\{\cdot\}$. The $i \times i$ identity matrix is expressed by $\mathbf{I}_{i \times i}$ while the zero matrix with i rows and j columns is defined by $\mathbf{0}_{i \times j}$.

II. OPTICAL TRANSMISSION SYSTEM MODEL

The optical transmission system considered in this paper is modeled as shown in Fig. 1. The complex-valued input signal $\mathbf{x}[n] = [x_x[n], x_y[n]]^T$ has a sampling rate of $R_s = 1/T$ where T is the symbol period in s. The spectrum $\mathbf{X}(\omega) = [X_x(\omega), X_y(\omega)]^T$ of $\mathbf{x}[n]$ is defined over the non-normalized continuous angular frequency axis $-\frac{\pi}{T} \leq \omega \leq \frac{\pi}{T}$ where ω is the angular frequency defined by $\omega = 2\pi f$ in rad/s and f is the fundamental frequency measured in Hz = s^{-1}. In the TD, the two-fold upsampling adds a zero between each of the samples of $\mathbf{x}[n]$. After upsampling, the sampling rate of $\mathbf{x}[n]$ is increased by a factor of 2 which means that the temporal spacing between its subsequent samples becomes $T/2$. The new obtained sequence is denoted as $\mathbf{y}[m]$ given as

$$\mathbf{y}[m] = \begin{cases} \mathbf{x}[n], & m = 2n \\ 0, & \text{elsewhere.} \end{cases} \qquad (1)$$

The spectrum of the upsampled signal $\mathbf{Y}(\omega)$ is identical to the spectrum of the original signal $\mathbf{X}(\omega)$

$$\mathbf{Y}(\omega) = \mathbf{X}(\omega) \qquad (2)$$

since $\mathbf{X}(\omega)$ is defined over a non-normalized continuous angular frequency axis ω. However, what differs after upsampling is the Nyquist bandwidth (bandwidth of uniqueness) of both signals $\mathbf{X}(\omega)$ and $\mathbf{Y}(\omega)$ since the sampling rate has changed.

The signal $\mathbf{X}(\omega)$ is then sent through a channel which, under the linear or weakly nonlinear assumption, can be described by concatenation of all deterministic channel impairments [20]

$$\mathbf{H}(\omega) = \mathbf{H}_{Rx}(\omega)\mathbf{H}_{PMD}(\omega)\mathbf{H}_{PDL}(\omega)\mathbf{H}_{CD}(\omega)\mathbf{H}_{Tx}(\omega). \qquad (3)$$

The transfer function accounting for any pulse shaper or low-pass filtering at the transmitter is defined by $H_{Tx}(\omega)$.

The CD transfer function $H_{CD}(\omega)$ is expressed by

$$H_{CD}(\omega) = \exp\left\{j\frac{D\lambda^2\omega^2}{4\pi c}\right\} \qquad (4)$$

where $D = \beta_2 d$ in ps/nm is responsible for the accumulated CD of a signal after propagating through an optical fiber of length d km and with group velocity dispersion β_2 in ps^{2}nm, λ in nm is the central wavelength of the signal, c in nm/ps is the speed of light in vacuum and $j = \sqrt{-1}$.

The Hermitian PDL matrix $\mathbf{H}_{PDL}(\omega)$ is defined by

$$\mathbf{H}_{PDL}(\omega) = \begin{bmatrix} \cos(\alpha') & -\sin(\alpha') \\ \sin(\alpha') & \cos(\alpha') \end{bmatrix} \times \begin{bmatrix} 1 & 0 \\ 0 & \kappa \end{bmatrix}$$
$$\times \begin{bmatrix} \cos(\alpha'') & -\sin(\alpha'') \\ \sin(\alpha'') & \cos(\alpha'') \end{bmatrix} \qquad (5)$$

where α' and α'' are the polarization rotation angles at the input and output of the PDL element, respectively [21]. The PDL coefficient κ $(0 < \kappa < 1)$ defined by the attenuation between the orthogonal lossy axis and the lossless axis can be expressed in dB as: $\kappa_{dB} = -20\log(\kappa)$.

PMD can be modeled by concatenating S independently orientated fiber sections with independent uniform birefringence:

$$\mathbf{H}_{PMD}(\omega) = \prod_{s=1}^{S} \mathbf{H}_{DGD_s}(\omega)$$
$$= \prod_{s=1}^{S} \begin{bmatrix} \cos(\alpha'_s) & -\sin(\alpha'_s) \\ \sin(\alpha'_s) & \cos(\alpha'_s) \end{bmatrix}$$
$$\times \begin{bmatrix} e^{j(\phi_s + \omega\tau_s)/2} & 0 \\ 0 & e^{-j(\phi_s + \omega\tau_s)/2} \end{bmatrix}$$
$$\times \begin{bmatrix} \cos(\alpha''_s) & -\sin(\alpha''_s) \\ \sin(\alpha''_s) & \cos(\alpha''_s) \end{bmatrix}. \qquad (6)$$

Analogously to the PDL case, the input and output field is rotated according to the principal SOP of the fiber section s. For each differential group delay (DGD) element, the input and output angles between the slow axis and the polarization state of the fast axis are α'_s and α''_s, respectively. Each Hermitian birefringence matrix is defined by ϕ_s which refers to the polarization phase and τ_s characterizing the DGD between the two orthogonal fast and slow axes. The all-order PMD approximation is obtained for $S = 20$, while the first-order PMD (referred also as DGD) is obtained for $S = 1$.

The transfer function $H_{Rx}(\omega)$ contains the contribution of the optical and electrical low-pass filtering at the receiver [22].

The cross coupling between the 2×2 channel transfer function $\mathbf{H}(\omega)$ and the x- and the y-polarization component of the transmitted signal $\mathbf{X}(\omega)$, perturbed with additive white Gaussian noise (AWGN) $\boldsymbol{\eta}(\omega)$ with variance σ_η^2, provides the
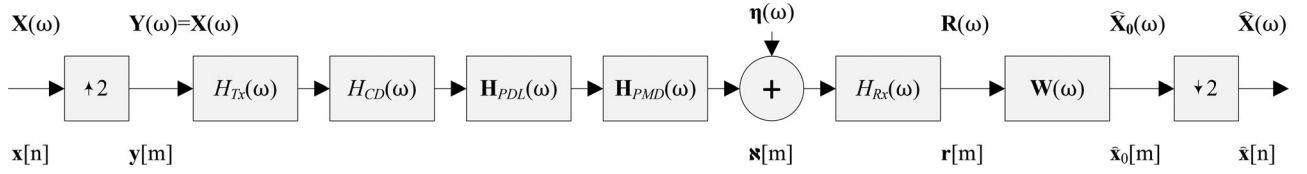
Fig. 1. Optical transmission system model.

received signal $\mathbf{R}(\omega) = [R_x(\omega), R_y(\omega)]^T$:

$$\mathbf{R}(\omega) = \mathbf{H}(\omega)\mathbf{X}(\omega) + H_{Rx}(\omega)\boldsymbol{\eta}(\omega)$$

$$= \begin{bmatrix} H_{xx}(\omega) & H_{xy}(\omega) \\ H_{yx}(\omega) & H_{yy}(\omega) \end{bmatrix} \begin{bmatrix} \mathbf{X}_x(\omega) \\ \mathbf{X}_y(\omega) \end{bmatrix}$$

$$+ H_{Rx}(\omega) \begin{bmatrix} \eta_x(\omega) \\ \eta_y(\omega) \end{bmatrix}. \qquad (7)$$

The equalized signal $\widehat{\mathbf{X}}_0(\omega)$ is obtained from the product between $\mathbf{R}(\omega)$ and the 2×2 MIMO equalizer transfer function $\mathbf{W}(\omega)$ designed to compensate for any channel distortion

$$\widehat{\mathbf{X}}_0(\omega) = \mathbf{W}(\omega)\mathbf{R}(\omega) = \begin{bmatrix} W_{xx}(\omega) & W_{xy}(\omega) \\ W_{yx}(\omega) & W_{yy}(\omega) \end{bmatrix} \begin{bmatrix} R_x(\omega) \\ R_y(\omega) \end{bmatrix}. \qquad (8)$$

Finally, to allow decision on a symbol basis $\widehat{\mathbf{X}}_0(\omega)$ is two-fold decimated such that, in the TD, only each second sample of $\widehat{\mathbf{x}}_0[n]$ is retained. After downsampling, the sampling rate of $\widehat{\mathbf{x}}_0[n]$ is decreased by a factor of 2 which means that the temporal spacing between its subsequent samples becomes T. The spectrum $\widehat{\mathbf{X}}(\omega)$ of the downsampled sequence $\widehat{\mathbf{x}}[n] = \widehat{\mathbf{x}}_0[2m]$ can be expressed as follows:

$$\widehat{\mathbf{X}}(\omega) = \frac{1}{2} \left\{ \widehat{\mathbf{X}}_0(\omega) + \widehat{\mathbf{X}}_0 \left(\omega + \frac{\pi}{T} \right) \right\} \qquad (9)$$

where $\widehat{\mathbf{X}}_0(\omega + \frac{\pi}{T})$ expresses the aliasing terms that appear due to downsampling. Thus, the spectrum of $\widehat{\mathbf{X}}(\omega)$ and $\widehat{\mathbf{X}}_0(\omega)$ are different since in the bandwidth of uniqueness of $\widehat{\mathbf{X}}_0(\omega)$ aliasing terms appear. Still, both spectra are defined over the same non-normalized continuous angular frequency axis ω.

III. FDE DESIGN

Since the DSP handles with discrete signals, the equalizer will be evaluated at discrete angular frequencies $\omega_k = 2\pi f_k, k = 0, \dots, M - 1$, where M is the length of the filter. The equalizer is designed to have one tap per frequency component k considering the following approaches: *A.* the equalization is done over the whole channel's bandwidth, i.e., without considering the downsampling in the system model illustrated in Fig. 1; *B.* the equalization is done over half the channel's bandwidth and the noise is minimized in the other half, i.e., considering the downsampling operation.

A. Without Taking Into Account Downsampling

Without considering the downsampling operation, the equalizer design is based on the detected signal $\widehat{\mathbf{X}}_0(\omega_k)$ given by

$$\widehat{\mathbf{X}}_0(\omega_k) = \mathbf{W}(\omega_k)\mathbf{H}(\omega_k)\mathbf{X}(\omega_k) + \mathbf{W}(\omega_k)H_{Rx}(\omega_k)\boldsymbol{\eta}(\omega_k). \qquad (10)$$

The equalizer matrix $\mathbf{W}(\omega_k)$ in (10) can be obtained by using: 1) the MMSE, 2) the ZF or 3) the matched filter (MF) design. Here, all M degrees of freedom are exploited for the design of the equalizer.

1) MMSE Solution: The MMSE aims to minimize the expectation of the squared Euclidean distance between the equalized signal $\widehat{\mathbf{X}}_0(\omega_k)$ and the corresponding transmitted signal $\mathbf{X}(\omega_k)$

$$\mathbf{W}(\omega_k) = \underset{\{\mathbf{W}(\omega_k)\}}{\mathrm{argmin}} \mathbf{J}(\omega_k)$$

$$= \underset{\{\mathbf{W}(\omega_k)\}}{\mathrm{argmin}} \mathrm{E}\left[\left\| \widehat{\mathbf{X}}_0(\omega_k) - \mathbf{X}(\omega_k) \right\|_2^2 \right] \qquad (11)$$

leading to

$$\mathbf{J}(\omega_k) = \mathrm{E}\left[\left\| \widehat{\mathbf{X}}_0(\omega_k) - \mathbf{X}(\omega_k) \right\|_2^2 \right]$$

$$= \mathrm{E}\left[\left\| (\mathbf{W}(\omega_k)\mathbf{H}(\omega_k) - \mathbf{I}_{2\times2}) \mathbf{X}(\omega_k) \right\|_2^2 \right]$$

$$+ \mathrm{E}\left[\left\| \mathbf{W}(\omega_k)H_{Rx}(\omega_k)\boldsymbol{\eta}(\omega_k) \right\|_2^2 \right]$$

$$= \sigma_X^2 \mathbf{W}(\omega_k)\mathbf{H}(\omega_k)\mathbf{H}^H(\omega_k)\mathbf{W}^H(\omega_k)$$

$$- \sigma_X^2 \mathbf{W}(\omega_k)\mathbf{H}(\omega_k)$$

$$- \sigma_X^2 \mathbf{W}^H(\omega_k)\mathbf{H}^H(\omega_k) + \sigma_X^2 \mathbf{I}_{2\times2}$$

$$+ \sigma_\eta^2 \mathbf{W}(\omega_k)\mathbf{W}^H(\omega_k)H_{Rx}(\omega_k)H_{Rx}^*(\omega_k) \qquad (12)$$

where the expectation is taken over the signal and noise which are considered uncorrelated

$$\mathrm{E}\left[\|\mathbf{X}(\omega_k)\|_2^2 \right] = \mathrm{E}\left[\mathbf{X}(\omega_k)\mathbf{X}^H(\omega_k) \right] = \sigma_X^2 \mathbf{I}_{2\times2} \qquad (13)$$

$$\mathrm{E}\left[\|\boldsymbol{\eta}(\omega_k)\|_2^2 \right] = \mathrm{E}\left[\boldsymbol{\eta}(\omega_k)\boldsymbol{\eta}^H(\omega_k) \right] = \sigma_\eta^2 \mathbf{I}_{2\times2}. \qquad (14)$$

To find the matrix $\mathbf{W}(\omega_k)$ that minimizes (11), the partial derivative of (12) with respect to $\mathbf{W}^H(\omega_k)$ has to be calculated

$$\frac{\partial \mathbf{J}(\omega_k)}{\partial \mathbf{W}^H(\omega_k)} = \mathbf{W}(\omega_k)\left[\mathbf{H}(\omega_k)\mathbf{H}^H(\omega_k) + \boldsymbol{\Xi}(\omega_k) \right] - \mathbf{H}^H(\omega_k) \qquad (15)$$

where

$$\boldsymbol{\Xi}(\omega_k) = \frac{\sigma_\eta^2}{\sigma_X^2}\Theta(\omega_k)\mathbf{I}_{2\times2} \qquad (16)$$

with $\Theta(\omega_k)$ defined as

$$\Theta(\omega_k) = H_{Rx}(\omega_k)H_{Rx}^*(\omega_k) = |H_{Rx}(\omega_k)|^2. \quad (17)$$

Setting (15) to zero and solving it for $\mathbf{W}(\omega_k)$ the MMSE equalizer matrix is obtained

$$\mathbf{W}(\omega_k) = \mathbf{H}^H(\omega_k)\big[\mathbf{H}(\omega_k)\mathbf{H}^H(\omega_k) + \mathbf{\Xi}(\omega_k)\big]^{-1}. \quad (18)$$

2) ZF Solution: The ZF approach brings the inter-symbol interference (ISI) down to zero in a noiseless scenario. Therefore, by setting $\sigma_\eta^2 = 0$ in (18), $\mathbf{W}(\omega_k)$ becomes

$$\mathbf{W}(\omega_k) = \mathbf{H}^{-1}(\omega_k). \quad (19)$$

To avoid noise enhancement due to non-constant power distribution of the channel, $\mathbf{H}(\omega_k)$ can be normalized, prior to matrix inversion, as

$$\mathbf{H}_N(\omega_k) = \mathbf{H}(\omega_k)\Theta^{-1}(\omega_k) \quad (20)$$

where the normalization factor $\Theta(\omega_k)$ can be defined as in (17), or alternatively, to avoid system calibration, as a p-norm (with $p \in \{1,2\}$) of the channel matrix $\mathbf{H}(\omega_k)$

$$\Theta(\omega_k) = \|\mathbf{H}(\omega_k)\|_p = \left(\sum_{i=1}^{I}\sum_{j=1}^{J}|H_{i,j}(\omega_k)|^p\right)^{\frac{1}{p}} \quad (21)$$

with $I = J = 2$ for a 2×2 MIMO channel.

The function $\Theta(\omega_k)$ is real, therefore, the normalized ZF preserves the inversion of the phase of the channel but not of the amplitude.

3) MF Solution: The equalizer can also be designed with the MF approach. The MF is a linear filter that maximizes the signal-to-noise ratio (SNR) of the received signal. It is obtained by assuming that the noise term σ_η^2 dominates in the denominator of (18) and reads as

$$\mathbf{W}(\omega_k) = \mathbf{H}^H(\omega_k)\Theta^{-1}(\omega_k) \quad (22)$$

where $\Theta(\omega_k)$ serves as a noise-whitening filter (NWF).

B. Taking Into Account Downsampling

Considering the two-fold downsampling operation, the equalizer design is based on the detected signal $\widehat{\mathbf{X}}(\omega_k)$ given by

$$\widehat{\mathbf{X}}(\omega_k) = \widehat{\mathbf{X}}_X(\omega_k) + \widehat{\mathbf{X}}_\eta(\omega_k) = \frac{1}{2}\big(\mathbf{E}^T\mathbf{G}\mathbf{X}(\omega_k) + \mathbf{E}^T\mathbf{\Gamma}\mathbf{N}\big) \quad (23)$$

where

$$\mathbf{E}^T = \begin{bmatrix} \mathbf{W}(\omega_k) & \mathbf{W}(\omega_k + \frac{\pi}{T}) \end{bmatrix} \quad (24)$$

$$\mathbf{G}^T = \begin{bmatrix} \mathbf{H}(\omega_k) & \mathbf{H}(\omega_k + \frac{\pi}{T}) \end{bmatrix} \quad (25)$$

$$\mathbf{\Gamma} = \begin{bmatrix} H_{Rx}(\omega_k) & 0 & 0 & 0 \\ 0 & H_{Rx}(\omega_k) & 0 & 0 \\ 0 & 0 & H_{Rx}(\omega_k + \frac{\pi}{T}) & 0 \\ 0 & 0 & 0 & H_{Rx}(\omega_k + \frac{\pi}{T}) \end{bmatrix} \quad (26)$$

$$\mathbf{N}^T = \begin{bmatrix} \boldsymbol{\eta}(\omega_k) & \boldsymbol{\eta}(\omega_k + \frac{\pi}{T}) \end{bmatrix} \quad (27)$$

and considering that $\mathbf{X}(\omega_k) = \mathbf{X}(\omega_k + \frac{\pi}{T})$.

The equalizer matrix $\mathbf{W}(\omega_k)$ can be obtained from (23) by using: 1) the MMSE, 2) the ZF or 3) the MF approach. To avoid ambiguity with the previous section, the ZF and MMSE are renamed as DZF and DMMSE, respectively. The MF approach leads to the same solution presented in the previous section, therefore, the name will not be changed. The design of the equalizer takes into account the aliasing due to the downsampling operation, therefore, $M/2$ degrees of freedom are exploited to design the equalizer over half of the total bandwidth and the remaining $M/2$ degrees of freedom are exploited to minimize the noise over the other half of the bandwidth.

1) DMMSE Solution: Contrary to the MMSE derived in the previous section, the target of the DMMSE is to minimize the square error between $\widehat{\mathbf{X}}(\omega_k)$ and $\mathbf{X}(\omega_k)$

$$\left\{\mathbf{W}(\omega_k), \mathbf{W}\left(\omega_k + \frac{\pi}{T}\right)\right\} = \mathrm{argmin}_{\{\mathbf{W}(\omega_k), \mathbf{W}(\omega_k + \frac{\pi}{T})\}}\mathbf{J}$$

$$= \mathrm{argmin}_{\{\mathbf{W}(\omega_k), \mathbf{W}(\omega_k + \frac{\pi}{T})\}}\mathrm{E}\left[\left\|\widehat{\mathbf{X}}(\omega_k) - \mathbf{X}(\omega_k)\right\|_2^2\right] \quad (28)$$

leading to

$$\mathbf{J} = \mathrm{E}\left[\left\|\widehat{\mathbf{X}}(\omega_k) - \mathbf{X}(\omega_k)\right\|_2^2\right]$$

$$= \mathrm{E}\left[\left\|\frac{1}{2}\big((\mathbf{E}^T\mathbf{G} - 2\mathbf{I}_{2\times2})\mathbf{X}(\omega_k) + \mathbf{E}^T\mathbf{\Gamma}\mathbf{N}\big)\right\|_2^2\right]$$

$$= \mathrm{E}\left[\frac{1}{4}\big((\mathbf{E}^T\mathbf{G} - 2\mathbf{I}_{2\times2})\mathbf{X}(\omega_k) + \mathbf{E}^T\mathbf{\Gamma}\mathbf{N}\big)\right.$$

$$\left.\times \big((\mathbf{E}^T\mathbf{G} - 2\mathbf{I}_{2\times2})\mathbf{X}(\omega_k) + \mathbf{E}^T\mathbf{\Gamma}\mathbf{N}\big)^H\right]$$

$$= \frac{1}{4}\big(\sigma_X^2\big(\mathbf{E}^T\mathbf{G} - 2\mathbf{I}_{2\times2}\big)\big(\mathbf{G}^H\mathbf{E}^* - 2\mathbf{I}_{2\times2}\big)\big)$$

$$+ \frac{1}{4}\big(\mathbf{E}^T\mathbf{\Gamma}\mathbf{K}_\eta\mathbf{\Gamma}^*\mathbf{E}^*\big) \quad (29)$$

where $\mathbf{K}_\eta$ is the noise correlation matrix defined by

$$\mathbf{K}_\eta = \mathrm{E}\big[NN^H\big] = \sigma_\eta^2\mathbf{I}_{4\times4}. \quad (30)$$

In addition to (13) and (14), the following assumptions are considered:

$$\mathrm{E}\left[\left\|\boldsymbol{\eta}\left(\omega_k + \frac{\pi}{T}\right)\right\|_2^2\right] = \mathrm{E}\left[\boldsymbol{\eta}\left(\omega_k + \frac{\pi}{T}\right)\boldsymbol{\eta}^H\left(\omega_k + \frac{\pi}{T}\right)\right] = \sigma_\eta^2\mathbf{I}_{2\times2} \quad (31)$$

$$\mathrm{E}\left[\boldsymbol{\eta}(\omega_k)\boldsymbol{\eta}^H\left(\omega_k + \frac{\pi}{T}\right)\right] = \mathbf{0}_{2\times2}. \quad (32)$$

The partial derivative of (29) with respect to $\mathbf{E}^*$ leads to

$$\frac{\partial\mathbf{J}}{\partial\mathbf{E}^*} = \frac{1}{4}\big(\sigma_X^2\mathbf{G}^*(\mathbf{G}^T\mathbf{E} - 2\mathbf{I}_{2\times2}) + \mathbf{\Gamma}^*\mathbf{K}_\eta\mathbf{\Gamma}\mathbf{E}\big). \quad (33)$$

Setting (33) to zero, the equalizer $\mathbf{E}$ is found to be

$$\mathbf{E} = 2\sigma_X^2\big(\sigma_X^2\mathbf{G}^*\mathbf{G}^T + \mathbf{\Gamma}^*\mathbf{K}_\eta\mathbf{\Gamma}\big)^{-1}\mathbf{G}^*. \quad (34)$$

The subband equalizer transfer function $\mathbf{W}(\omega_k)$ is obtained applying the matrix inversion lemma in (34), leading to

$$\mathbf{W}(\omega_k) = 2\mathbf{A}(\omega_k)\mathbf{H}^H(\omega_k)\Theta^{-1}(\omega_k) \qquad (35)$$

where

$$\begin{aligned}
\mathbf{A}(\omega_k) = \ & \Big(\mathbf{H}(\omega_k)\mathbf{H}^H(\omega_k)\Theta^{-1}(\omega_k) \\
& + \mathbf{H}\left(\omega_k + \frac{\pi}{T}\right)\mathbf{H}^H\left(\omega_k + \frac{\pi}{T}\right) \\
& \times \Theta^{-1}\left(\omega_k + \frac{\pi}{T}\right) + \frac{\sigma_\eta^2}{\sigma_X^2}\mathbf{I}_{2\times 2} \Big)^{-1}
\end{aligned}$$

with $\Theta(\omega_k)$ defined as in (17).

2) DZF Solution: The DZF solution is obtained by setting $\sigma_\eta^2 = 0$ in (35). The subband equalizer transfer function $\mathbf{W}(\omega_k)$ is given by

$$\mathbf{W}(\omega_k) = 2\mathbf{B}(\omega_k)\mathbf{H}^H(\omega_k)\Theta^{-1}(\omega_k) \qquad (36)$$

where

$$\begin{aligned}
\mathbf{B}(\omega_k) = \ & \Big(\mathbf{H}(\omega_k)\mathbf{H}^H(\omega_k)\Theta^{-1}(\omega_k) + \mathbf{H}\left(\omega_k + \frac{\pi}{T}\right) \\
& \times \mathbf{H}^H\left(\omega_k + \frac{\pi}{T}\right)\Theta^{-1}\left(\omega_k + \frac{\pi}{T}\right) \Big)^{-1}
\end{aligned}$$

with $\Theta(\omega_k)$ defined as in (17).

3) MF Solution: The MF is obtained by assuming that the noise term σ_η^2 dominates in the denominator of (35) and read as in (22), with $\Theta(\omega_k)$ defined as in (17).

IV. TRAINING-AIDED 2×2 MIMO FREQUENCY DOMAIN CHANNEL ESTIMATION

In the previous section the tap of the equalizer matrix have been calculated knowing the channel $\mathbf{H}(\omega_k)$. However, in a real system the exact channel matrix is not known and the receiver needs to retrieve such information from the received data. In this section, TA 2×2 MIMO FD CE based on a special type of CAZAC sequences is described.
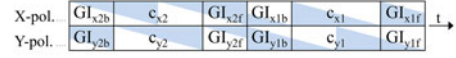
A. TS Design

An important class of complex polyphase sequences with optimal properties for CE and synchronization is the perfect-square minimum-phase (PS-MP) CAZAC [23] or more in general Frank–Zadoff sequence [24], defined as

$$\begin{aligned}
c[n] = \ & \exp\Big\{ j\frac{2\pi}{\sqrt{N}} \Big[\mathrm{mod}\left(n-1, \sqrt{N}\right) + 1 \Big] \\
& \times \Big[\Big\lfloor \frac{n-1}{\sqrt{N}} \Big\rfloor + 1 \Big] \Big\}
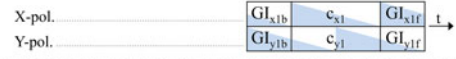\end{aligned} \qquad (37)$$

with $n = 1, 2, \ldots, N$, where N is the sequence length in symbols. These sequences have flat spectrum, and their auto-correlation

$$\mathrm{ACF}_i = \sum_{n=1}^{N} = c[n+i]c^*[n] \qquad (38)$$

Fig. 2. TS schemes for 2×2 MIMO systems.

where $c[n+N] = c[n]$, has only one contribution different from zero

$$\mathrm{ACF}_i = \begin{cases} \sum_{n=1}^{N} |c[n]|^2, & \text{for } i = 0 \\ 0, & \text{otherwise.} \end{cases} \qquad (39)$$

The minimum number of distinct phases of the polyphase code (37) given by $\sqrt{N}$ is obtained for sequences length $N = p^2$ symbols with $p \in \{1, 2, 3, \ldots\}$. However, for lengths $N = 2^{2p}$ symbols with $p \in \{1, 2, 3, \ldots\}$ the constellation plot of a PS-MP CAZAC sequence refers to a $\log(N)$ PSK modulated signal (i.e., $p = 1$ refers to binary PSK, $p = 2$ to QPSK, $p = 3$ to 8PSK and so on). In principle, the modulation of the TS and of the payload data are independent assuming that the modulator allows generating all relevant constellation points of both, the TS and the payload data. This provides significant benefits for rate-adaptive transponders where data can be switched between different higher-order modulation formats depending on the required optical SNR (OSNR) and the demanded channel capacity. Therefore, PS-MP CAZAC sequences of length $N = 2^{2p}$ symbols have a twofold benefit on the system: relax the requirement of the modulator at the transmitter and facilitate the use of FFT operations at receiver keeping complexity significantly lower than that of conventional single carrier systems with TDE structure [3]. In addition, the computation of the reciprocal of a CAZAC sequence does not require divisions: $c^{-1}[n] = c^*$.

When PS-MP CAZAC codes are used as TS for CE, their length has to be chosen in accordance to the longest CIR expected at the point of estimation of the channel, taking into account that the quality of a single CE is proportional to the length of the training word and that to track time-varying polarization effects it is required that the known symbols are periodically transmitted between the payload data.

For a 2×2 MIMO, the TS based on PS-MP CAZAC codes can be classified in two groups: double-block TS (DB-TS) scheme and single-block TS (SB-TS) scheme.

1) Double-Block Training Sequence: With this approach the TS consists of four blocks $c_{x1}[n]$, $c_{x2}[n]$, $c_{y1}[n]$, $c_{y2}[n]$, transmitted in couple in two consecutive time slots (see Fig. 2-top).

A simple DB-TS scheme requires that for each time slot only one training block $c[n]$ is transmitted in one polarization whereas no signal is present in the other polarization. This approach for

the estimation of the transmission parameters is suboptimal for the modulator and the gain control since the power envelope of the signal varies for certain intervals of time. In addition, such power fluctuation would lead to nonlinear crosstalk in the fiber reducing the performance of the whole optical system. Therefore, this approach will not be further considered in the following sections of this paper.

A more sophisticated TS scheme which uses the diversity principle of the Alamouti code [18], [25] is defined as

$$
\begin{aligned}
c_{x1}[n] &= c[n] \\
c_{y1}[n] &= c[n+N/2] \\
c_{x2}[n] &= -c^*[n+N/2] \\
c_{y2}[n] &= c^*[n].
\end{aligned}
\tag{40}
$$

These sequences are time-space uncorrelated. Orthogonality is assured by sending at the same time two PS-MP CAZAC sequences shifted with respect to each other by a number of symbols greater or equal to the CIR. Since the CIR length is not known in advance, the simplest implementation would be to choose a fixed orthogonality interval equal to $N/2$ symbols which is also the optimum solution for a 2×2 MIMO system. In addition, the TS blocks can be (optionally) framed by adding a guard interval (GI) of length N_{GI} symbols at the beginning and end of each TS block. The GIs continuously pursue the adjacent PS-MP CAZAC sequence as described in Fig. 2. The total length of the DB-TS is $N_{TS} = 2N + 4N_{GI}$ symbols.

2) Single-Block Training Sequence: Alternatively, as illustrated in Fig. 2-bottom, the TS can be composed of only two independent blocks $c_{x1}[n]$ and $c_{y1}[n]$ sent simultaneously one per polarization [19]

$$
\begin{aligned}
c_{x1}[n] &= c[n] \\
c_{y1}[n] &= c[n+N/2].
\end{aligned}
\tag{41}
$$

The total length of the SB-TS is $N_{TS} = N + 2N_{GI}$ symbols. Respect to the DB-TS scheme the training overhead (TOH) is reduced by 50%.

B. Channel Estimation

In what follows, the CE algorithms using the DB-TS and SB-TS are presented. The estimated channel matrix $\widehat{\mathbf{H}}(\omega_k)$ is used to update the 2×2 MIMO FDE described in Section III.

The CE is performed at discrete angular frequencies $\omega_k = 2\pi f_k, k = 0,\ldots,M-1$, where $M = 2N$ is the length of the two-fold oversampled TS.

1) Double-Block Training Sequence: With this approach, the transmission system in a noiseless scenario can be modeled by

$$
\begin{aligned}
&\begin{bmatrix} R_{x1}(\omega_k) & R_{x2}(\omega_k) \\ R_{y1}(\omega_k) & R_{y2}(\omega_k) \end{bmatrix} \\
&= \begin{bmatrix} H_{xx}(\omega_k) & H_{xy}(\omega_k) \\ H_{yx}(\omega_k) & H_{yy}(\omega_k) \end{bmatrix} \times \begin{bmatrix} C_{x1}(\omega_k) & C_{x2}(\omega_k) \\ C_{y1}(\omega_k) & C_{y2}(\omega_k) \end{bmatrix}.
\end{aligned}
\tag{42}
$$

Multiplying left and right side of (44) by

$$
\begin{bmatrix} C_{x1}(\omega_k) & C_{x2}(\omega_k) \\ C_{y1}(\omega_k) & C_{y2}(\omega_k) \end{bmatrix}^{-1} = Q(\omega_k)^{-1} \begin{bmatrix} C_{y2}(\omega_k) & -C_{x2}(\omega_k) \\ -C_{y1}(\omega_k) & C_{x1}(\omega_k) \end{bmatrix}
\tag{43}
$$

with

$$
Q(\omega_k) = C_{x1}(\omega_k)C_{y2}(\omega_k) - C_{y1}(\omega_k)C_{x2}(\omega_k)
\tag{44}
$$

leads to the channel components $\widehat{H}_{xx}(\omega_k)$, $\widehat{H}_{xy}(\omega_k)$, $\widehat{H}_{yx}(\omega_k)$, $\widehat{H}_{yy}(\omega_k)$

$$
\widehat{H}_{xx}(\omega_k) = \big(R_{x1}(\omega_k)C_{y2}(\omega_k) - R_{x2}(\omega_k)C_{y1}(\omega_k) \big) Q(\omega_k)^{-1}
\tag{45}
$$

$$
\widehat{H}_{xy}(\omega_k) = \big(R_{x2}(\omega_k)C_{x1}(\omega_k) - R_{x1}(\omega_k)C_{x2}(\omega_k) \big) Q(\omega_k)^{-1}
\tag{46}
$$

$$
\widehat{H}_{yx}(\omega_k) = \big(R_{y1}(\omega_k)C_{y2}(\omega_k) - R_{y2}(\omega_k)C_{y1}(\omega_k) \big) Q(\omega_k)^{-1}
\tag{47}
$$

$$
\widehat{H}_{yy}(\omega_k) = \big(R_{y2}(\omega_k)C_{x1}(\omega_k) - R_{y1}(\omega_k)C_{x2}(\omega_k) \big) Q(\omega_k)^{-1}.
\tag{48}
$$

2) Single-Block Training Sequence: For SB-TS, the noiseless transmission system can be modeled by

$$
\begin{bmatrix} R_{x1}(\omega_k) \\ R_{y1}(\omega_k) \end{bmatrix} = \begin{bmatrix} H_{xx}(\omega_k) & H_{xy}(\omega_k) \\ H_{yx}(\omega_k) & H_{yy}(\omega_k) \end{bmatrix} \times \begin{bmatrix} C_{x1}(\omega_k) \\ C_{y1}(\omega_k) \end{bmatrix}.
\tag{49}
$$

In this case, to perform the CE, left and right side of (49) need to be multiplied by

$$
\begin{bmatrix} C_{x1}^*(\omega_k) & 0 \\ 0 & C_{y1}^*(\omega_k) \end{bmatrix}
\tag{50}
$$

to obtain $\widehat{H}_A(\omega_k)$ and $\widehat{H}_B(\omega_k)$:

$$
\begin{aligned}
\widehat{H}_A(\omega_k) &= C_{x1}(\omega_k)C_{x1}^*(\omega_k) = H_{xx}(\omega_k)C_{x1}(\omega_k)C_{x1}^*(\omega_k) \\
&\quad + H_{xy}(\omega_k)C_{x1}(\omega_k)C_{x1}^*(\omega_k)e^{j\frac{M}{2}\omega_k}
\end{aligned}
\tag{51}
$$

$$
\begin{aligned}
\widehat{H}_B(\omega_k) &= C_{y1}(\omega_k)C_{y1}^*(\omega_k) = H_{yy}(\omega_k)C_{y1}(\omega_k)C_{y1}^*(\omega_k) \\
&\quad + H_{yx}(\omega_k)C_{y1}(\omega_k)C_{y1}^*(\omega_k)e^{j\frac{M}{2}\omega_k}
\end{aligned}
\tag{52}
$$

which in TD read as

$$
\begin{aligned}
\widehat{h}_A[m] &= c_{x1}[m] \otimes c_{x1}^*[-m] \\
&= h_{xx}[m] \otimes \delta[m] + h_{xy}[m] \otimes \delta\left(m - \frac{M}{2}\right)
\end{aligned}
\tag{53}
$$

$$
\begin{aligned}
\widehat{h}_B[m] &= c_{y1}[m] \otimes c_{y1}^*[-m] \\
&= h_{yx}[m] \otimes \delta\left(m - \frac{M}{2}\right) + h_{yy}[m] \otimes \delta[m]
\end{aligned}
\tag{54}
$$

where $\delta[m]$ is the Dirac function defined as

$$
\delta[m] = \begin{cases} 1, & m = 1 \\ 0, & \text{otherwise.} \end{cases}
\tag{55}
$$

By TD windowing (TDW) $\widehat{h}_A[m]$ the CIR corresponding to $\widehat{h}_{xx}[m]$ and $\widehat{h}_{xy}[m]$ can be extracted, while $\widehat{h}_{yx}[m]$ and $\widehat{h}_{yy}[m]$ can be retrieved by TDW $\widehat{h}_B[m]$

$$\widehat{H}_{xx}(\omega_k) = \mathcal{F}\left\{\widehat{h}_A[m]\text{rect}[m]\right\} \tag{56}$$

$$\widehat{H}_{xy}(\omega_k) = \mathcal{F}\left\{\widehat{h}_A\left[m+\frac{M}{2}\right]\text{rect}[m]\right\} \tag{57}$$

$$\widehat{H}_{yx}(\omega_k) = \mathcal{F}\left\{\widehat{h}_B\left[m+\frac{M}{2}\right]\text{rect}[m]\right\} \tag{58}$$

$$\widehat{H}_{yy}(\omega_k) = \mathcal{F}\left\{\widehat{h}_B[m]\text{rect}[m]\right\} \tag{59}$$

where $\text{rect}[m]$ is a rectangular windowing function. For $M_{\text{TDW}} = M/2$ taps windowing, it is defined as

$$\text{rect}[m] = \begin{cases} 0, & \frac{M}{4}+1 \leq \text{mod}(m-1,M) \leq \frac{3M}{4} \\ 1, & \text{elsewhere.} \end{cases} \tag{60}$$

However, the length of the TDW can be adjusted depending on the function to be performed.

3) Averaging Over CEs: For noisy scenarios, linear averaging over L consecutive CEs is performed to reduce the impact of the ASE noise [16]

$$\widehat{\mathbf{H}}_{\text{LIN}}(\omega_k) = \frac{1}{L+1}\sum_{l=0}^{L}\widehat{\mathbf{H}}_l(\omega_k)\cdot\exp\left\{-jl\angle\psi\right\}. \tag{61}$$

The term ψ makes phase correction eliminating the phase jump between consecutive CEs induced by the carrier frequency-offset (CFO). It is defined by

$$\psi = \begin{cases} 0, & L = 0 \\ \frac{1}{M-1}\sum_{k=0}^{M-1}\Psi(\omega_k), & L \geq 1 \end{cases} \tag{62}$$

where

$$\Psi(\omega_k) = \frac{1}{4}\sum\text{diag}\left\{\frac{1}{L}\sum_{l=0}^{L-1}\widehat{\mathbf{H}}_{l+1}(\omega_k)\cdot\widehat{\mathbf{H}}_l^H(\omega_k)\right\}. \tag{63}$$

V. PERFORMANCE EVALUATION

The performance evaluation is based on simulated 28-GBaud PDM-QPSK and PDM-16-QAM coherent optical systems transmitting RZ, NRZ or Nyquist (0.1 roll-off factor) shaped pulses. Simulations of the linear channel include CD, PMD, polarization rotation angle α, polarization phase ϕ, PDL and CFO. AWGN is loaded onto the signal before an optical Gaussian band-pass filter (second-order, double-sided 35-GHz) and an electrical Bessel filter (fifth-order, 19-GHz) which defines the opto-electronic front-end of the coherent receiver, modeled by $H_{Rx}(\omega)$ in Fig. 1. An analog-to-digital converter digitalizes the received signal at two samples per symbol. The two-fold over-sampled signal is then processed by the TA-DSP architecture illustrated in Fig. 3. Equalization is performed in two stages: a couple of identical CD-FDEs (one per polarization) are dedicated to perform coarse compensation of the static CD, while a shorter adaptive TA 2×2 MIMO FDE is dedicated to PDM

signals de-multiplexing, residual CD (RCD)/PMD compensation and tracking of time-varying effects [3], [11], [15], [16], [26]. Equalization in FD is performed by transferring 50% over-lapping blocks of the serial data into the FD by a discrete FFT, applying the compensating function to each block, transferring the signal back into the TD by discrete IFFT and cutting off the overlap to restore the serial data stream [overlap-save (OS) method]. Typical single-tap per sub-band FDEs are used for CD compensation. The length M of each CD-FDE is chosen accordingly with the maximum CD expected at the point of estimation. FD CD compensation based on a non-maximally decimated discrete FFT filter bank with a trivial prototype filter and a delayed single-tap equalizer per sub-band could be employed to extend the CD tolerance without increasing the FFT size [10]. CD estimation can be performed either in TD or in FD as proposed in [4]–[9]. However, in this paper, CD estimation is not considered. The taps of the CD-FDEs are calculated using the desired CD value. RCD considered to evaluate the TA 2×2 MIMO FDE refers to the CD which in a real system is not compensated at CD-FDE due to imperfect CD estimation. After CD compensation a block of the data stream is taped to a TS finder and TA CFO estimator [27]. CFO compensation is performed on the TS extracted from the received data prior to CE. Residual CFO (RCFO) refers to the TA CFO estimation error. The buffer is an optional module and can be used to compensate for the CE and filter-update processing-delay. The TA 2×2 MIMO FDE has a single-tap per sub-band. The number of sub-band M' depends on the TS scheme and the length N of the PS-MP CAZAC chosen for CE. When $M' > 2N$, the interpolation of the channel spectra is performed by zero-padding the estimated CIRs. Note that a tap of the CD-FDE refers to one complex coefficient while a tap of the 2×2 MIMO FDE consists of four complex coefficients. FD decimation allows to use IFFT of $M'/2$ points in front of the 2×2 MIMO FDE. Viterbi and Viterbi carrier recovery is performed [28] before bit decision and bit-error counting. In contrast to the architecture proposed for equalization which is modulation format transparent, the carrier phase recovery is modulation format dependent. To adapt the Viterbi and Viterbi algorithm to 16-QAM, here, the QPSK partitioning approach is used [29].

The performance investigation considers time-invariant and time-varying channels. In the former case the TA 2×2 MIMO FDE is evaluated with respect to 1) convergence speed (number of averaging over CEs), 2) RCD, 3) DGD, 4) PDL, 5) misalignment between received and reference TS, 6) RCFO and 7) tap-update solutions for both CD-FDE and TA 2×2 MIMO FDE. In the latter case the TA 2×2 MIMO FDE is evaluated with respect to SOP-rotation speed.

In time-invariant scenarios the taps of 2×2 MIMO FDE are update only at the initialization stage so that the frame structure and the TS overhead are not considered in the discussion. However, such concepts acquire fundamental importance in time-varying channels where continuous tap-update is required.

In what follows, (if not otherwise stated) the RZ50 pulse shaping is employed. The $M = 1024$ coefficients of the CD-FDE are calculated by ZF as in (19) where $H(\omega_k) = H_{\text{CD}}(\omega_k)$. Following the 2×2 matrix notation used in
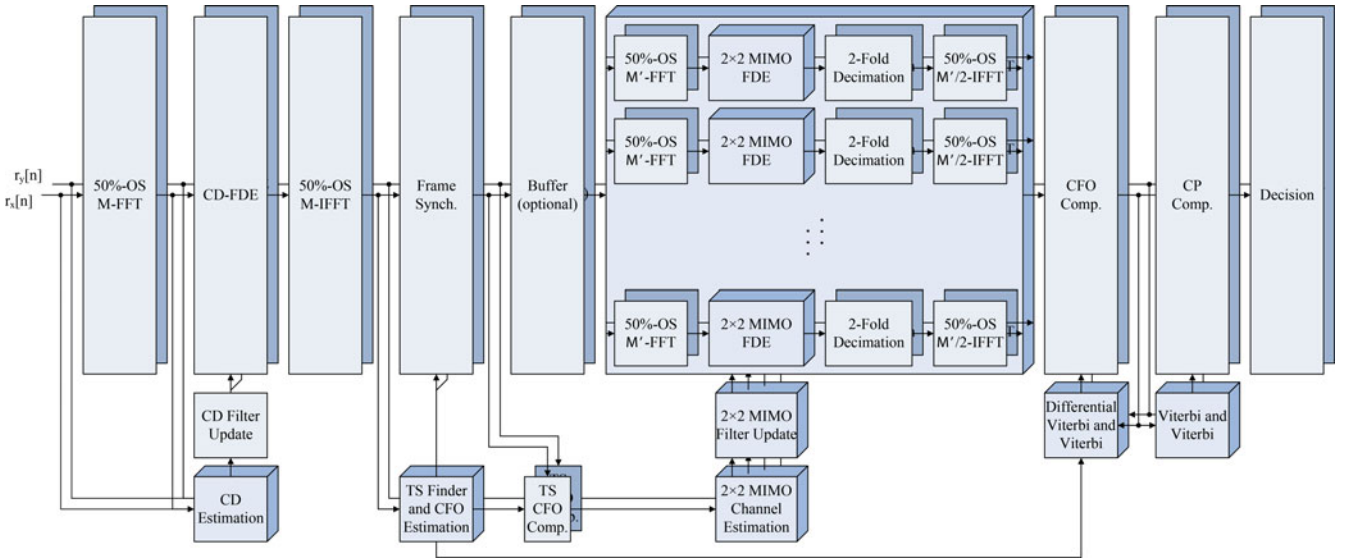
Fig. 3.   Receiver TA DSP. OS: overlap-save, FFT: fast Fourier-transform, CD: chromatic dispersion, FDE: frequency-domain equalization, IFFT: inverse FFT, MIMO: multi-input multi-output, CFO: carrier-frequency offset, CP: carrier-phase, TS: training sequence, Synch.: synchronization, Comp.: compensation.

Section III, $H(\omega_k)$ refers to the elements in the diagonal of $\mathbf{H}(\omega_k)$, so that $H(\omega_k) = H_{xx}(\omega_k) = H_{yy}(\omega_k)$. While, the $4M'$ coefficients of the TA 2×2 MIMO FDE are calculated by DZF as expressed in (36) with $\Theta(\omega_k) = 1$. The extracted TS is perfectly aligned with the reference TS and the RCFO is equal to zero. The required OSNR (ROSNR) (0.1 nm) at bit-error-rate (BER) equal to 10^{-3} is considered as figure of merit for performance evaluation.

A. TA 2×2 MIMO FDE in Time-Invariant Channels

1) Convergence Speed: Fig. 4 shows the evolution of the ROSNR at BER $= 10^{-3}$ as function of the number of linear averaging over CEs required to update the equalizer. Fig. 4(a) and (b) refer to the PDM-QPSK system, instead Fig. 4(c) refers to the PDM-16-QAM transmission system. The optical channel consists of only ASE noise. The CIR is estimated by using SB-TS and DB-TS of length $N \in \{16, 64, 256\}$ symbols with GI length $N_{\mathrm{GI}} = N/4$ symbols. For 16-QAM, the TSs with $N = 16$ symbols are mapped either in the inner-, middle- or outer-ring of the constellation diagram [30]. While the TSs with $N > 16$ symbols are mapped crossing the middle-ring. Since for both PDM-QPSK and PDM-16-QAM the TA 2×2 MIMO FDE exhibits same convergence trend, in Fig. 4(c) only the effect of the TS mapping is reported. The TA 2×2 MIMO FDE has length $M' = 2N$ taps. CE based on SB-TS requires TDW with, in general, $N_{\mathrm{TDW}} = N/2$ symbols. Although CE based on DB-TS does not need TDW, it can be optionally applied to smooth the transfer function of the estimated channel. For both, SB-TS and DB-TS the case of $N_{\mathrm{TDW}} = 8$ symbols is investigated. Results show that by keeping fix N_{TDW} and increasing N, a lower number of averaging over CEs is needed for a given ROSNR. More specifically, averaging over ξ CEs based on TS with $N = 256$ symbols and $N_{\mathrm{TDW}} = 8$ symbols leads to similar performance than aver-

aging over 4ξ CEs and 16ξ CEs based on TS with $N = 64$ symbols and $N = 16$ symbols with $N_{\mathrm{TDW}} = 8$ symbols, respectively. Thus, the higher difference between the CAZAC block and the TDW ($N - N_{\mathrm{TDW}}$) is the higher is the effectiveness of the smoothing filter. It should be noted that the length of the TDW needs to be chosen such to cover the longest expected CIR $\max\{m_{\mathrm{Ch}}\}$ at the point of CE.

The number of averaging over CEs L based on a given TS G required to reach the same performance of L_R CEs averaging based on a reference TS R is given by

$$L = \xi_R \frac{N_R}{N_{\mathrm{TDW}_R}} L_R \times \frac{1}{\xi_G} \frac{N_{\mathrm{TDW}_G}}{N_G} \qquad (64)$$

where ξ refers to the TS scheme: $\xi = 1$ for SB-TS and $\xi = 2$ for DB-TS. To be (64) valid: the same power needs to be used to map the single symbols of both TSs and $\{N_{\mathrm{TDW}}, 2N_{\mathrm{GI}}\} \geq m_{\mathrm{Ch}}$, with m_{Ch} the channel memory defined in terms of adjacent symbols over which a transmitted pulse spreads due to the influence of the channel.

SB-TS with $N_{\mathrm{TDW}} = N/2$ symbols and DB-TS without TDW show the same and the slowest convergence. However, for a given N and N_{TDW}, DB-TS exhibits two-fold faster convergence than SB-TS. Nevertheless, despite the fact that the number of averaging over CEs is halved, the total number of training symbols ($L \times N_{\mathrm{TS}}$) remains constant. The fastest convergence is obtained by averaging over $L = 5$ CEs based on DB-TS with $N = 256$ symbols and $N_{\mathrm{TDW}} = 8$ symbols, reaching the same performance of FDE based on ideal channel. In addition, it is important to note that even in the worst case (no CE averaging) and with a short TS ($N = 16$ symbols) in both PDM-QPSK and PDM-16-QAM systems acceptable performance is achieved and the OSNR penalty is around 3 dB. Such fast and robust convergence makes the TA 2×2 MIMO FDE an attractive solution for burst-mode transmissions.
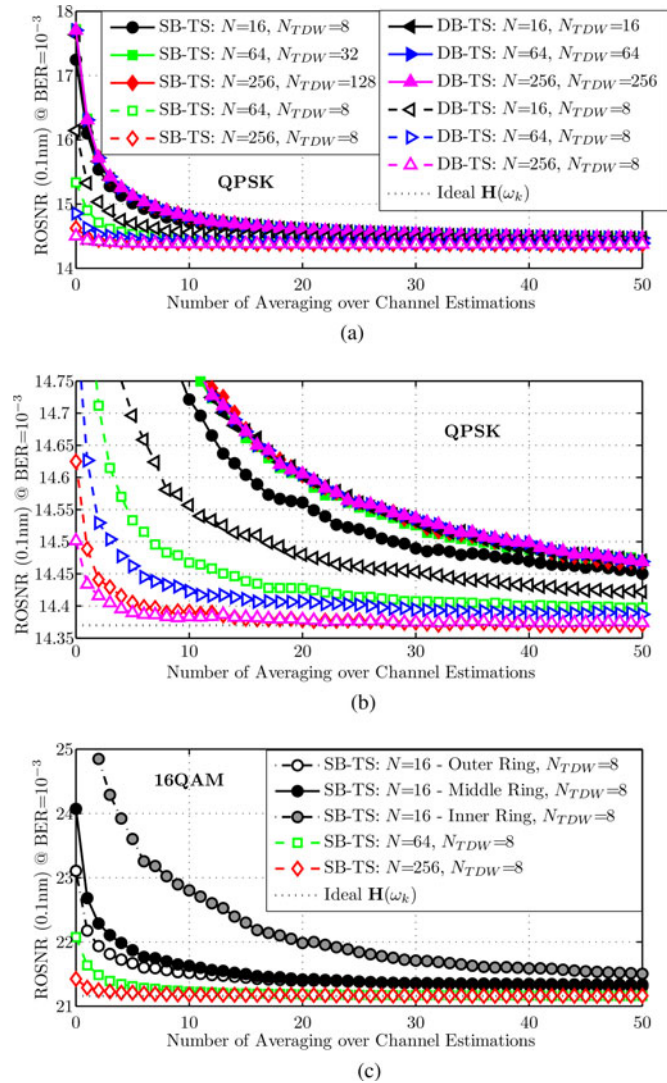
(a)

(b)

(c)

Fig. 4. TA 2×2 MIMO FDE convergence speed. CE based on SB-TS and DB-TS employing PS-MP CAZAC codes of length $N \in \{16, 64, 256\}$ symbols with GIs of length $N_{GI} = N/4$ symbols. TDW of length $N_{TDW} = N/2$ symbols and $N_{TDW} = 8$ symbols considered. Equalization performed using the 50% OS method by a $M' = 2N$ taps FDE updated with the DZF defined in (36) and $\Theta(\omega_k) = 1$. (a) PDM-QPSK. (b) Zoom in of Fig. 4(a). (c) PDM-16-QAM.

(a)

(b)

Fig. 5. TA 2×2 MIMO FDE: RCD tolerance. CE based on (a) SB-TS and (b) DB-TS employing PS-MP CAZAC codes of length $N = 16$ symbols with GIs of length $N_{GI} \in \{0, 4, 8\}$ symbols. Equalization performed using the 50% OS method by a $M' \in \{32, 64\}$ taps FDE updated with the DZF defined in (36) and $\Theta(\omega_k) = 1$.

(a)

(b)

Fig. 6. TA 2×2 MIMO FDE: DGD tolerance. CE based on (a) SB-TS and (b) DB-TS employing PS-MP CAZAC codes of length $N = 16$ symbols with GIs of length $N_{GI} \in \{0, 4, 8\}$ symbols. Equalization performed using the 50% OS method by a $M' \in \{32, 64\}$ taps FDE updated with the DZF defined in (36) and $\Theta(\omega_k) = 1$.

The results shown in the following sections are mainly based on TS with $N = 16$ symbols. SB-TS has $N_{TDW} = 8$ symbols, while for DB-TS no TDW is performed. The number of averaging over CEs is set to $L = 30$.

2) RCD and DGD Tolerance: Figs. 5 and 6 present the TA 2×2 MIMO FDE tolerance with regard to RCD and DGD, respectively. The ROSNR at BER $= 10^{-3}$ is plotted as function of the impairment under consideration and as function of m_{Ch}. Assuming that CD and DGD are the major players for ISI in an optical transmission system: $m_{Ch} = m_{CD} + m_{DGD}$, where $m_{CD} = D \cdot \Delta_\omega \cdot B_s \cdot \frac{\lambda_c^2}{c}$ and $m_{DGD} = \tau \cdot B_s$, with Δ_ω the signal bandwidth, B_s the signal baudrate and $\lambda_c = 1550$ nm the carrier wavelength. The channel is estimated by using SB-TS and DB-TS with $N = 16$ symbols and $N_{GI} \in \{0, 4, 8\}$ symbols. The 2×2 MIMO FDE has $M' \in \{32, 64\}$ taps.
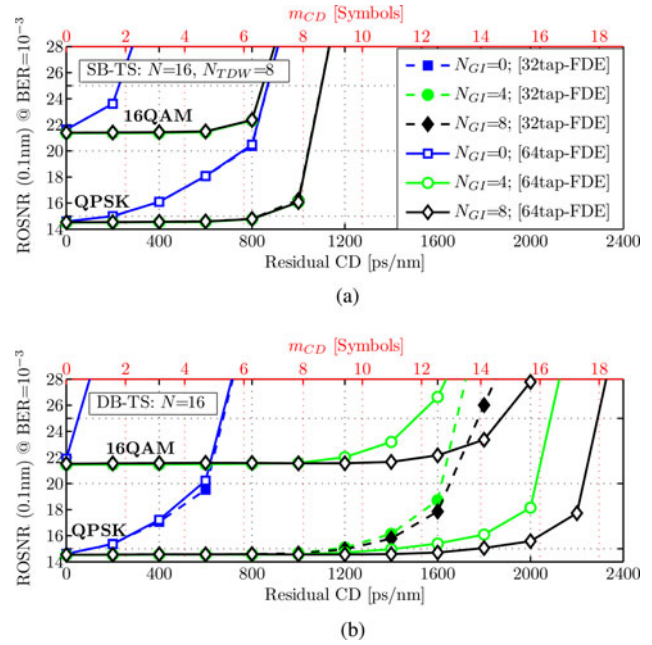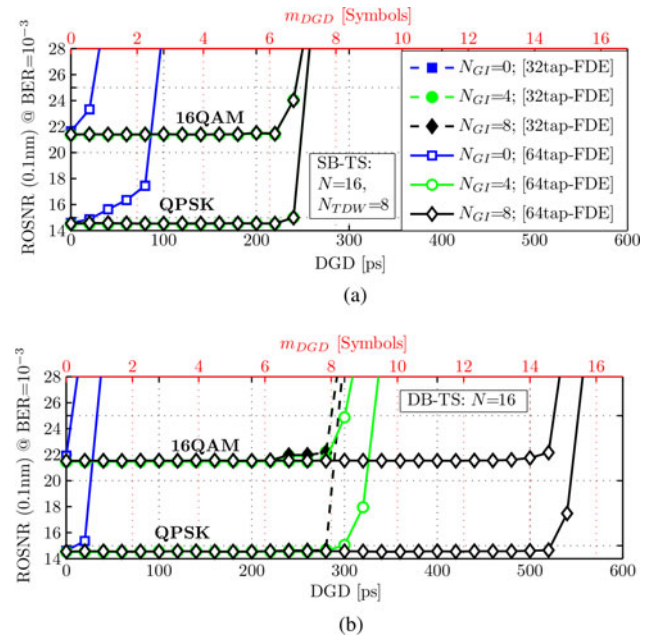
Results show that if no GI is used to absorb the ISI between the adjacent TS blocks (DB-TS) or the TS block and user data (DB-TS and SB-TS), the accuracy of the estimated CIR rapidly degrades as far as m_{Ch} increases leading to poor equalization performances. The robustness of the CE can be
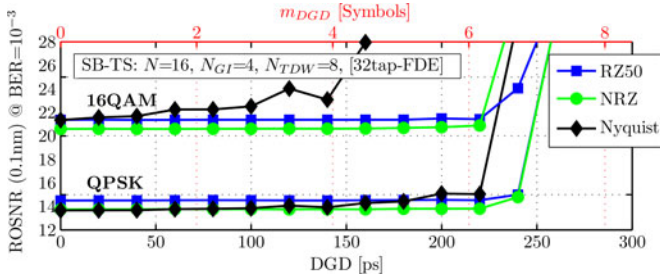
Fig. 7. TA 2×2 MIMO FDE: DGD tolerance for different pulse shaping. CE based on SB-TS employing PS-MP CAZAC codes of length $N = 16$ symbols with GIs of length $N_{GI} = 4$ symbols. Equalization performed using the 50% OS method by a $M' = 32$ taps FDE updated with the DZF defined in (36) and $\Theta(\omega_k) = 1$.
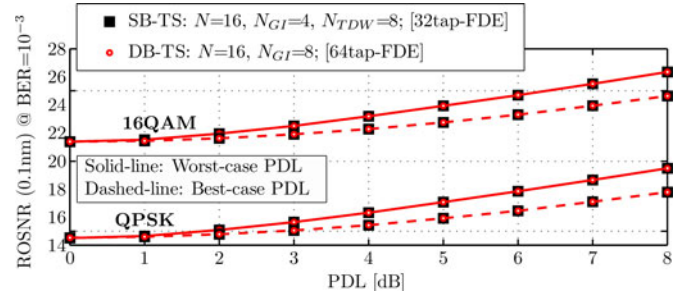

Fig. 8. TA 2×2 MIMO FDE: PDL tolerance. CE based on SB-TS and DB-TS employing PS-MP CAZAC codes of length $N = 16$ symbols with GIs of length $N_{GI} = N/4$ symbols. Equalization performed using the 50% OS method by a FDE updated with the DZF defined in (36) and $\Theta(\omega_k) = 1$: $M' = 2N$ taps for SB-TS; $M' = 4N$ taps for DB-TS.

improved by adding a GI at the beginning and end of each TS block as described in Fig. 2. For both PDM-QPSK and PDM-16-QAM transmission systems, GIs of four symbols allow penalty-free CE estimation up to 600 and 800 ps/nm RCD for SB-TS and DB-TS, respectively. With the same GI length, penalty-free CE estimation up to 220 and 280 ps DGD is obtained for SB-TS and DB-TS, respectively. However, respect to PDM-QPSK, in PDM-16-QAM a sharper performance degradation is observed. Because the CIR of DGD is much shorter than that of CD, RCD tolerance leads to a smooth degradation while high values of DGD experience a steep degradation. In addition, DB-TS exhibits larger tolerance with respect to RCD and DGD due to the fact that TDW is not required. SB-TS of length $N = 16$ allows estimation of CIRs with $\max\{m_{Ch}\} = 8$ symbols. Therefore, a couple of GIs of four symbols totally cover the ISI between the CAZAC block and the user data. In contrast, DB-TS uses the full length of the TS to estimate the CIR, that is, 16 symbols. Thus, to fully cover $m_{Ch} = 16$ symbols, two GIs per TS block of eight symbols each are required. The dashed-black lines with black-diamond markers in both Figs. 5(b) and 6(b) show that even tough the CE can cover a CIR of $\max\{m_{Ch}\} = 16$ symbols the tolerance with respect of RCD and DGD of the 32-tap FDE is limited to $\max\{m_{Ch}\} = 8$. This is due to the fact that the FDE is implemented by using a 50%-OS method which means that an equalizer of M' taps effectively covers CIRs of $\max\{m_{Ch}\} = M'/2$ symbols. Considering that the data are two-fold oversampled, the equalizer finally covers CIRs of $\max\{m_{Ch}\} = M'/4$ symbols, that is, $\max\{m_{Ch}\} = 8$ symbols for $M' = 32$ taps. Hence, to maximize the system efficiency for CEs based on DB-TS with $N = 16$ symbols and $N_{GI} = 8$ symbols the FDE length needs to be extended to 64 taps. Thus, for both PDM-QPSK and PDM-16-QAM transmission systems, penalty-free CE estimation up to 1400 ps/nm RCD and 520 ps DGD is then obtained. For CE based on SB-TS the required FDE length is $M' = 32$ taps.

Fig. 7 shows the effect of the pulse shaping for TA 2×2 MIMO CE. It should be noted, that the PS-MP CAZAC used for CE at the receiver consists of RZ50 pulses independently of the pulse shaping used at the transmitter. This choice has been made to preserve the constant amplitude and zero-circular auto-correlation properties of the oversampled TS used as reference at the receiver. Results show that, apart from a different

ROSNR due to the bandwidth chosen in the simulations for the receiver optical and electrical filters, between RZ50 and NRZ no significant difference is observed in the DGD tolerance. In contrast, for Nyquist pulses a reduction in DGD tolerance is observed. This effect is due to the increased CIR length obtained by correlating a RZ50 and a Nyquist shaped TS. Such concepts applies also for RCD tolerance.

Although the performance evaluation in this section was based only on TS with $N = 16$ symbols, similar conclusions can be drawn for TS with longer PS-MP CAZAC blocks. In general, such large DGD tolerance is typically not required, making it clear that the TS is mainly designed to meet the CD requirement, which follows the worst-case RCD after the first CD compensation stage. Therefore, longer TS employing TDW could be required only to reduce the effect of ASE noise from the CE without performing averaging over CEs.

3) PDL Tolerance: Despite the channel impairments with memory, PDL is one of the limiting memoryless effects. It results in a polarization-dependent OSNR degradation which can be estimated but not compensated by filtering [20].

Fig. 8 shows the TA 2×2 MIMO FDE tolerance with respect to PDL. The channel is estimated by using SB-TS and DB-TS with $N = 16$ symbols. The GIs have length $N_{GI} = 4$ symbols and $N_{GI} = 8$ symbols for SB-TS and DB-TS, respectively. The 2×2 MIMO FDE has length $M' = 32$ taps for SB-TS while for DB-TS $M' = 64$ taps. However, the length of the TS or the length of the FDE has no impact on the PDL tolerance of the TA 2×2 MIMO FDE. Worst-case PDL attenuating the signal of only one polarization, which leads to a worst-case BER performance and best-case PDL equally attenuating the signal in both polarizations, which leads to a best-case BER performance are considered. Results show no difference between SB-TS and DB-TS. The PDL tolerance curves of the TA 2×2 MIMO FDE follow the same evolution as those of a 2×2 MIMO TDE with CMA/DD-LMS update, as reported in [21]. For both PDM-QPSK and PDM-16-QAM transmission systems, between 0.5 and 1 dB OSNR penalty and between 2 and 3 dB OSNR penalty are observed at 3 dB PDL and at 6 dB PDL, respectively.

4) Misalignment Between Received and Reference TS: Up to now, it was assumed that the TS finder was ideal, that is, the received TS was perfectly aligned with the TS used as reference
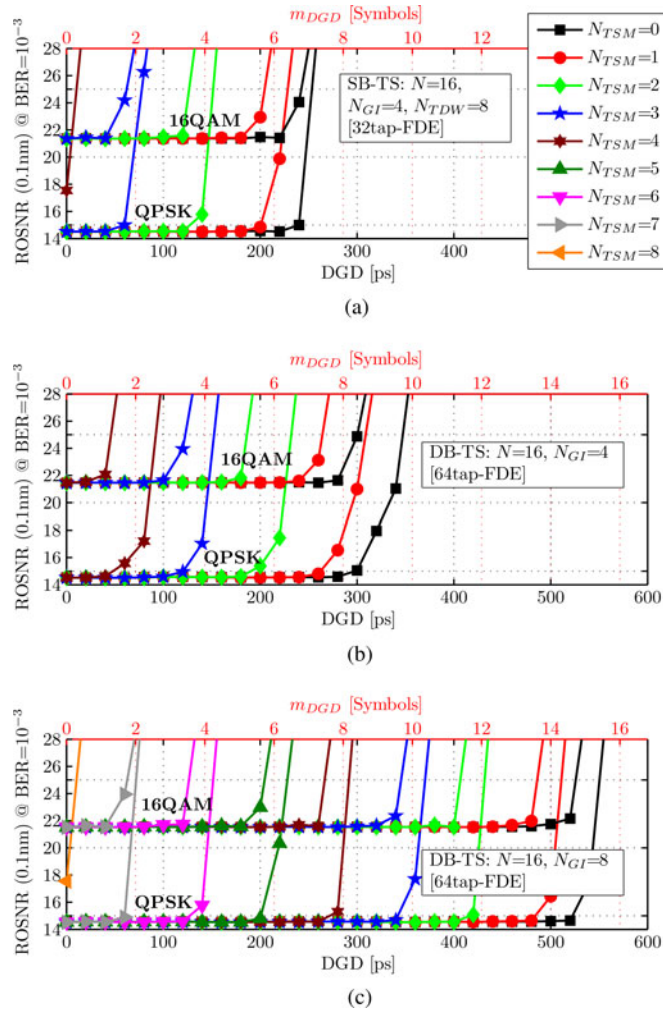
Fig. 9. DGD tolerance with respect to misalignment between received and reference TS. CE based on SB-TS: (a) $N = 16$ with $N_{GI} = 4$ or 8 symbols, DB-TS: $N = 16$ with (b) $N_{GI} = 4$ symbols and (c) $N_{GI} = 8$ symbols. Equalization performed using the 50% OS method by a FDE updated with the DZF defined in (36) and $\Theta(\omega_k) = 1$: $M = 2N$ taps for SB-TS; $M = 4N$ taps for DB-TS.

at the receiver. However, in a real system TS misalignment, specially in the initialization stage of the DSP, could be expected. The effect of such TS misalignment is reported only with respect to DGD in Fig. 9 because the RCD investigation would lead to same conclusions. The channel is estimated by using SB-TS and DB-TS with $N = 16$ symbols and $N_{GI} \in \{4, 8\}$ symbols. The 2×2 MIMO FDE has length $M' = 32$ taps and $M' = 64$ taps for SB-TS and DB-TS, respectively. The TS misalignment is $N_{TSM} \in \{0, 8\}$ symbols.

Fig. 9(a) shows that SB-TS tolerates $\max\{N_{TSM}\} = 3$ symbols. The limiting factor is the length of the TDW $N_{TDW} = 8$ symbols, therefore, GIs of length $N_{GI} \geq 4$ symbols lead to same results. Fig. 9(b) and (c) shows the tolerance for DB-TS with $N_{GI} = 4$ symbols and $N_{GI} = 8$ symbols, respectively. The limiting factor now is the length of the GIs. If the GIs have length $N_{GI} = 4$ symbols, then the TS tolerates $\max\{N_{TSM}\} = 3$ symbols. Instead, if $N_{GI} = 8$ symbols, the TS can tolerate $\max\{N_{TSM}\} = 7$ symbols.
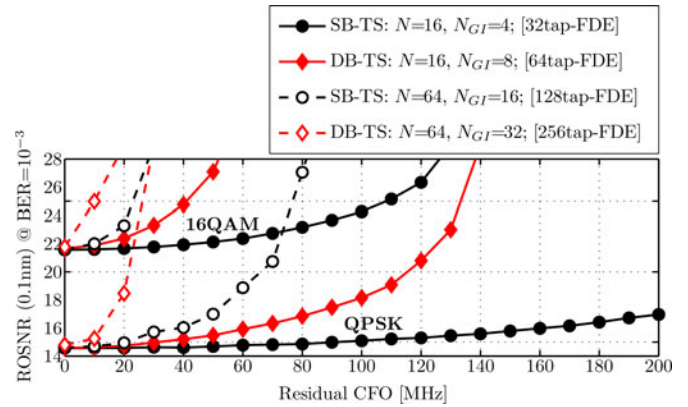

Fig. 10. TA 2×2 MIMO FDE: RCFO tolerance. CE based on SB-TS and DB-TS: $N \in \{16, 64\}$, $N_{GI} = N/4$ symbols for SB-TS and $N_{GI} = N/2$ symbols for DB-TS. Equalization performed using the 50% OS method by a FDE updated with the DZF defined in (36) with $\Theta(\omega_k) = 1$: $M' = 2N$ taps for SB-TS; $M' = 4N$ taps for DB-TS.

In general, the TS misalignment reduces the $\max\{m_{Ch}\}$ of the CIR that a given TS can estimate. Therefore, apart from increasing the tolerance with respect to RCD and DGD, the GIs have an important role in relaxing the requirement for the TS finder.

5) RCFO Tolerance: CFO is added to the signal emulating intradyne coherent detection. Fig. 10 shows the TA 2×2 MIMO FDE tolerance with respect to RCFO. The TS CFO compensation module is switched off. The channel is estimated by using SB-TS and DB-TS with $N \in \{16, 64\}$ symbols. SB-TS has $N_{GI} = N/4$ symbols with $N_{TDW} = N/2$ symbols, while DB-TS has $N_{GI} = N/2$. The 2×2 MIMO FDE has length $M' = 2N$ taps for SB-TS and $M' = 4N$ taps for DB-TS. SB-TS exhibits larger RCFO tolerance compared to DB-TS: in the PDM-QPSK system, TS with $N = 16$ symbols experience 1 dB OSNR penalty at 140 MHz for SB-TS and 55 MHz for DB-TS. Longer TS are more affected by RCFO such that TSs with $N = 64$ symbols show 1 dB OSNR penalty at 30 MHz for SB-TS and at 15 MHz for DB-TS. In the PDM-16-QAM system, TS with $N = 16$ symbols exhibit 1 dB OSNR penalty at 65 MHz for SB-TS and 25 MHz for DB-TS. TS with $N = 64$ symbols show 1 dB OSNR penalty at 15 MHz for SB-TS and at 5 MHz for DB-TS.

DB-TS shows lower RCFO tolerance compared to SB-TS because its longer length N_{TS}. In fact, the phase correction in (61) eliminates the phase jump induced by CFO between consecutive CEs. However, phase correction within a single CE is not performed leading to performances degradation for long TS. Nevertheless, it has to be notated that CFO estimation is jointly performed with the TS finder before TA 2×2 MIMO CE [27]. The robustness and accuracy of such algorithms improves for longer TS [31]. Thus, longer TS allow low RCFO tolerant CE but at the same time permit accurate CFO estimation, while shorter TS allow high RCFO tolerant CE but less accurate CFO estimation.

6) Tap-Update Solutions for Both CD-FDE and TA 2×2 MIMO FDE: The performances of the different FDE designs presented in Section III are reported in the following. Filter
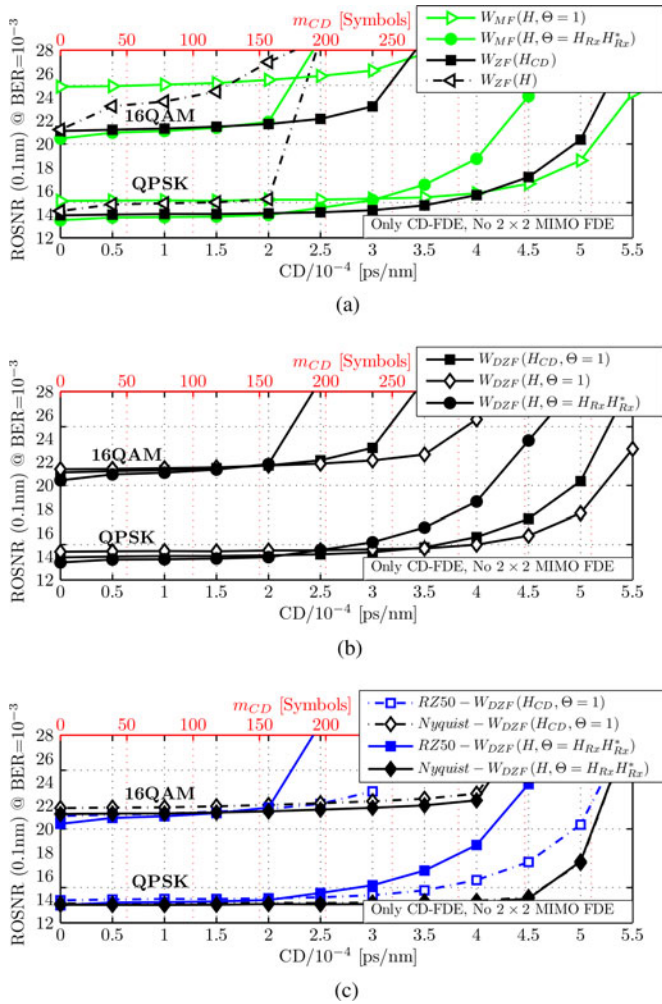
Fig. 11.   CD-FDE without TA 2×2 MIMO FDE: CD tolerance. 1024-tap CD-FDE with 50% OS implementation, ideal CD estimation. (a) Equalizer design without taking into account downsampling. (b) Equalizer design taking into account downsampling. (c) NWF effect for RZ50 and Nyquist pulses.

designs based on the MMSE criterion require higher implementation complexity due to the necessary estimation of the signal and noise power. In addition, for the systems considered, MMSE approach brings marginal BER improvement respect to ZF filter solutions, therefore, only the performances based on the simpler MF or ZF approach are discussed.

Fig. 11 shows the BER performance of the CD-FDE (no 2×2 MIMO equalizer) versus CD: in Fig. 11(a) the equalizer design does not take into account downsampling, while in Fig. 11(b) the downsampling is taken into account. In Fig. 11(c), the effect of the NWF is investigated for RZ50 and Nyquist pulse shaping. Depending on the channel information available at the receiver, the tap-calculation solutions of the CD-FDE can be divided into three groups: 1) only the CD transfer function $H_{\mathrm{CD}}(\omega_k)$ is known, therefore $H(\omega_k) = H_{\mathrm{CD}}(\omega_k)$; 2) the full transfer function of the channel is known $H(\omega_k) = H_{Rx}(\omega_k) H_{\mathrm{CD}}(\omega_k) H_{Tx}(\omega_k)$ but no NWF is applied; 3) the channel is as in 2) and a NWF as defined in (17) is applied.

The first group of filter designs uses the entire equalizer memory for CD compensation leading to best tradeoff between back-to-back (B2B) performance and large CD tolerance.

The second group of filter designs shows the worst BER performance. The $W_{\mathrm{ZF}}(\omega_k)$ inverts the amplitude of the whole channel bandwidth leading to noise enhancement occurring on the outer frequencies of the FDE where the power of the estimated channel is low. In contrast, the performance degradation for the $W_{\mathrm{MF}}(\omega_k)$ solution is due the additional low-pass filtering that the equalizer applies on the signal. The $W_{\mathrm{DZF}}(\omega_k)$ shows interesting behavior for high values of CD but unacceptable OSNR penalty at low CD values.

The last group of filter designs exhibits the best B2B performance due to the use of the NWF. However, the memory of the equalizer is not anymore fully dedicated for CD compensation leading to a reduced CD tolerance.

As shown in Fig. 11(c), the NWF used in the CD-FDE improves the B2B BER performance independently of the pulse shaping used at the transmitter and, in contrast to the RZ50, does not reduce the CD tolerance. For Nyquist systems a larger CD tolerance is obtained due to the narrower spectrum of the signal. In addition, for complexity saving the outer taps of the CD-FDE can be set to zero. Here, 448 taps over 1024 taps are set to zero resulting in about 44% complexity reduction.

Fig. 12 shows the equalization performances of the CD-FDE in combination with the TA 2×2 MIMO 32-tap FDE. Two filter solutions are considered for the CD-FDE: the one which offers the best tradeoff between B2B performance and large CD tolerance, $W_{\mathrm{ZF}}(\omega_k)$, $W_{\mathrm{MF}}(\omega_k)$ or $W_{\mathrm{DZF}}(\omega_k)$ based on $H(\omega_k) = H_{\mathrm{CD}}(\omega_k)$ with $\Theta(\omega_k) = 1$ [see Fig. 12(a)], and the one which exhibits the best B2B BER performance, $W_{\mathrm{DZF}}(\omega_k)$ or $W_{\mathrm{MF}}(\omega_k)$, both, based on $H(\omega_k) = H_{Rx}(\omega_k) H_{\mathrm{CD}}(\omega_k) H_{Tx}(\omega_k)$ with $\Theta(\omega_k)$ defined as in (17) [see Fig. 12(b)]. The 2×2 MIMO CE is based on SB-TS with $N = 16$ symbols, $N_{\mathrm{GI}} = 4$ symbols and $N_{\mathrm{TDW}} = 8$ symbols. The filter design for the TA 2×2 MIMO FDE can be divided into two groups: 1) no NWF is used or 2) a NWF defined as in (21) [or channel normalization for the solution (20)] is used.

Results show that the NWF is a nice tool to trade B2B performance with CD tolerance. The best B2B performance is obtained when both, CD-FDE and TA 2×2 MIMO FDE use a NWF. Instead, at the cost of degraded B2B BER performance, the largest CD tolerance is obtained when none of the equalizers uses a NWF. It should be notated that the NWF for the CD-FDE requires calibration of the system. If such calibration could not be performed, the NWF or the channel normalization expressed as in (21), applied to the TA 2×2 MIMO FDE, can still be beneficial for both B2B performance and CD tolerance.

In Fig. 12(c) the effect of NWF for the TA 2×2 MIMO FDE designed with the DZF solution is investigated for Nyquist channels. In such scenario, the NWF exhibits some limitations mainly observed in PDM-16-QAM systems. Indeed, due to the null parts of the signal spectrum, the NWF cannot be applied to the full bandwidth. Therefore, a certain number of taps of the TA 2×2 MIMO FDE needs to be set to zero. Such operation leads to performance improvement and DSP power saving. Here, 448
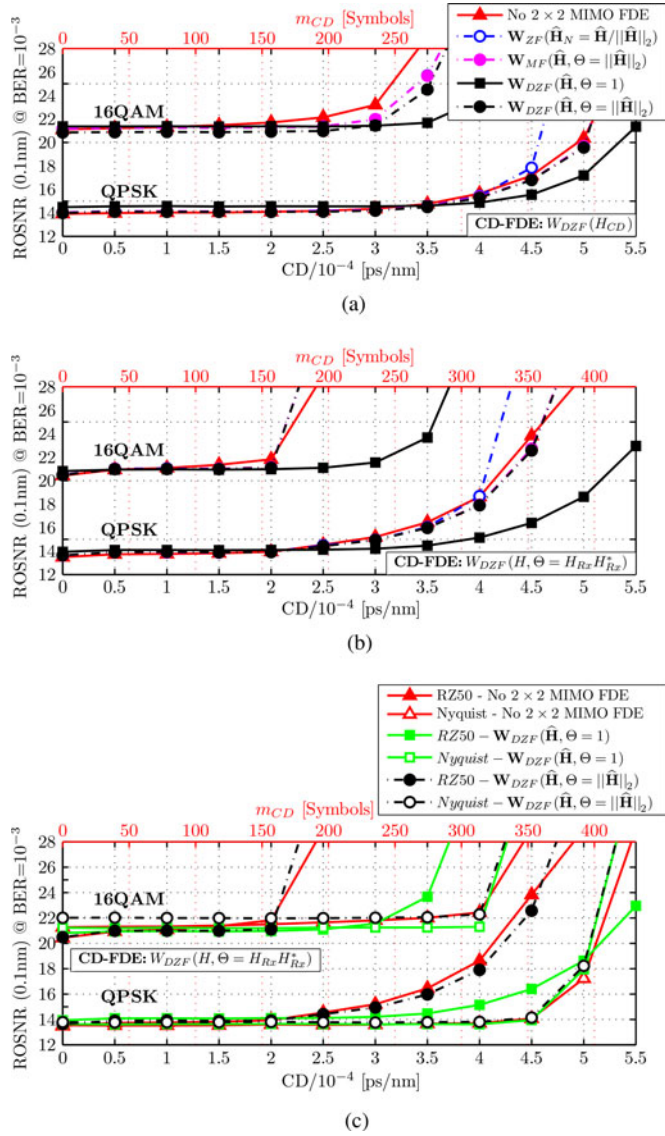
(a)

(b)

(c)

Fig. 12. CD-FDE with TA 2×2 MIMO FDE: CD tolerance. CD-FDE: 1024-tap equalizer using 50% OS method, ideal CD estimation. TA 2×2 MIMO FDE: 32-tap equalizer using 50% OS method, CE based on SB-TS with $N = 16$ symbols and $N_{\mathrm{GI}} = 4$ symbols. (a) CD-FDE updated by (36) with $\Theta(\omega_k) = 1$. (b) CD-FDE updated by (36) with $\Theta(\omega_k) = H_{Rx}(\omega_k)H_{Rx}^H(\omega_k)$. (c) NWF for RZ50 and Nyquist pulses. CD-FDE updated by (36) with $\Theta(\omega_k) = H_{Rx}(\omega_k)H_{Rx}^H(\omega_k)$.

taps over 1024 taps of the CD-FDE and 14 taps over 32 taps of the TA 2×2 MIMO FDE are set to zero. The B2B OSNR penalty caused by the NWF, in both PDM-QPSK and PDM-16-QAM systems, could be further reduced by increasing the number of taps set to zero in the TA 2×2 MIMO FDE.

B. TA 2×2 MIMO FDE in Time-Varying Channels

In addition to RCD and PMD compensation, the second stage equalizer should be able to operate in time-varying channels. For this investigation, a time-varying channel is generated by a Jones matrix with endless polarization rotation, i.e., by taking the PDL element defined in (5) and setting $\alpha' = 0$, $\kappa = 1$, and $\alpha'' = \vartheta t$,
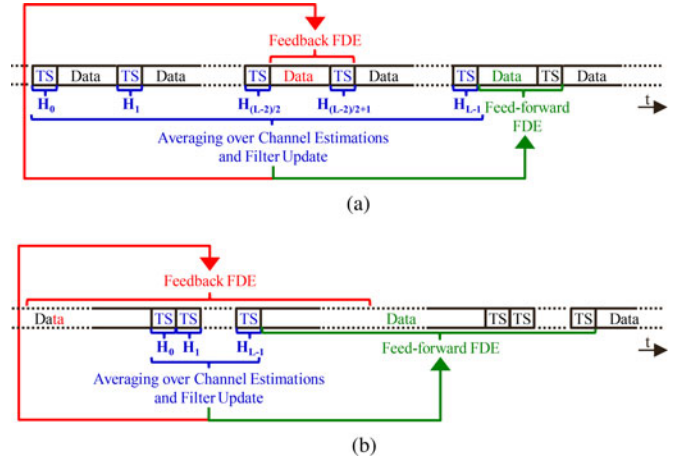

Fig. 13. Feed-forward (green) and feedback (red) FD equalization with filter-tap update after averaging over L consecutive CEs in time-varying channels. (a) $\chi = \gamma = 1$, $L \geq \chi$. (b) $\chi > 1, \gamma > 1, L = \chi$.

where ϑ is the angular frequency in rad/s governing the SOP rotation speed and t is the time in s. In such scenario, the TS needs to be framed with the payload data such that the frequency repetition of the TS is higher than that of the dynamic effects to be tracked (i.e., SOP rotation). In general, a frame has length N_F symbols including χ SB-TS of total length $N_{\mathrm{TS}} = 2N_{\mathrm{GI}} + \chi N$ symbols or χ DB-TS of total length $N_{\mathrm{TS}} = \chi(4N_{\mathrm{GI}} + 2N)$ symbols and $N_D = N_F - N_{\mathrm{TS}}$ symbols of user data. It should be noted that for $\chi > 1$, if SB-TS are employed then no GI is required between the repetition of the PS-MP CAZAC sequences. In fact, for each repetition the adjacent sequences continuously pursue the PS-MP CAZAC sequence, therefore, acting as GIs. Such concept cannot be applied to DB-TS since the sequences in the two time slots differs: $c_{x1}[n] \neq c_{x2}[n]$ and $c_{y1}[n] \neq c_{y2}[n]$. The TOH required for CE is given by $\zeta = \frac{N_{\mathrm{TS}}}{N_F}$. To keep the TOH constant when χ varies the new frame length should be defined by $N_F' = \gamma N_F$, where $\gamma = \chi$ for DB-TS while $\gamma = \frac{\chi N + 2N_{\mathrm{GI}}}{N + 2N_{\mathrm{GI}}}$ for SB-TS. In the following investigation, to keep the frame length N_F' as multiple of N_F for SB-TS the definition of γ is changed as

$$\gamma = \left\lfloor \frac{\chi N + 2N_{\mathrm{GI}}}{N + 2N_{\mathrm{GI}}} \right\rfloor. \tag{65}$$

Fig. 13 illustrates CEs averaging and equalization employing the feed-forward and feedback equalization schemes for data framed with $\chi = \gamma = 1$, where $L \geq \chi$ [see Fig. 13(a)] and with $\chi > 1, \gamma > 1$, where $L = \chi$ [see Fig. 13(b)]. The reference frame considered here has length $N_F = 1024$ symbols including a SB-TS of length $N_{\mathrm{TS}} = 24$ symbols ($N = 16$ symbols plus two GIs of length $N_{\mathrm{GI}} = 4$ symbols), which leads to 2.34% TOH required for CE. The taps of the 32-tap FDE are updated after every frame. For simplicity the feed-forward equalization scheme does not take into account processing delay. The feedback equalization approach requires buffering of data (see Fig. 3). The results of both PDM-QPSK and PDM-16-QAM system exhibit same trend so that only performances of PDM-QPSK are a shown.
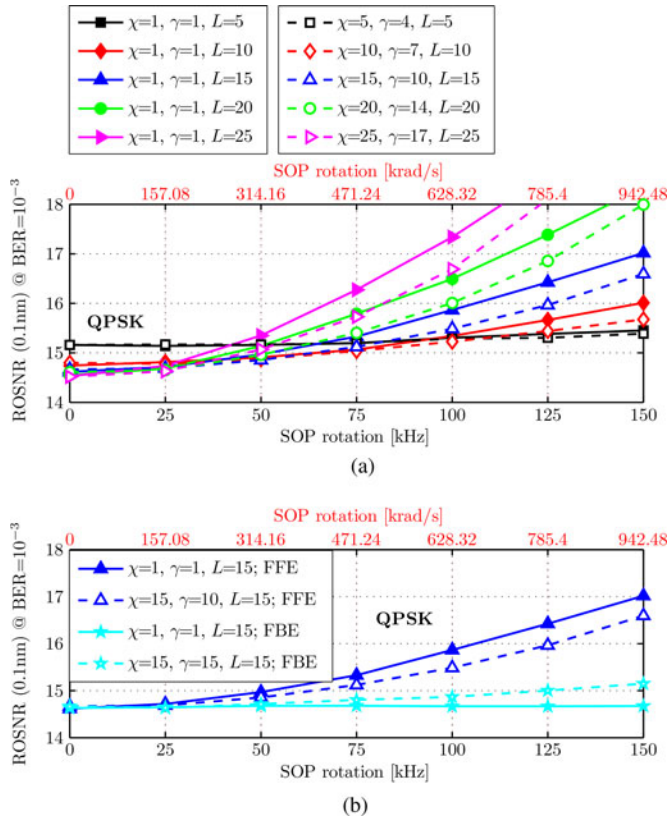
Fig. 14. TA 2×2 MIMO FDE: SOP-tracking speed. CE based on SB-TS employing PS-MP CAZAC codes of length $N = 16$ symbols and GIs of length $N_{\mathrm{GI}} = 4$ symbols. The 32-tap FDE is updated using the DZF defined in (36) and $\Theta(\omega_k) = 1$. (a) Frame comparison (feed-forward equalization): $N_F (\chi = 1, \gamma = 1)$ versus $N_{F'}(\chi > 1, \gamma > 1)$. (b) Comparison between feed-forward equalization (FFE) and feedback equalization (FBE) approach.

To reduce the impact of ASE noise linear averaging over L consecutive CEs is performed. A comparison between different averaging methods can be found in [17]. Ideally in static and noisy scenarios an infinite number of linear averages over CEs would eliminate the impact of ASE noise from the acquired channels, as discussed in Section V.A.1. Nevertheless, if the channel is varying in time, a linear averaging over many past CEs would lead to sub-optimum actual channel state. Even making the update non-causal by buffering the data (feedback equalization) a long averaging method would reduce the dynamic of the channel. Therefore, it becomes clear that the number L of averages over CEs has to be chosen in accordance with the speed of the SOP rotation.

Fig. 14(a) illustrates the TA 2×2 MIMO FDE tracking capability with respect to the dynamic change of SOP. The two frames schemes presented in Fig. 13 are compared with equalization performed by using the feed-forward approach. If only the last five CEs ($L = 5$) are considered, the two methods show similar performances. A B2B OSNR penalty is experienced respect to the case where $L = 25$ due to the non-averaged noise present on the channel matrix. However, the equalizer is able to track fast SOP rotation (in the range of hundreds kHz) with a negligible tracking penalty. By increasing L the B2B performance improves at the cost of slower SOP tracking capability.

For a given L, in general, better performances are observed for frames containing multiple TSs ($\chi > 1, \gamma > 1$).

Fig. 14(b) compares the TA 2×2 MIMO FDE performance for equalization performed by using the feed-forward and the feedback approach. The feedback approach shows superior performance than the feed-forward approach. In fact, the processing delay needed for CE and filter-update can totally compensated by buffering the received data. Note that, due to the shorter frame length and, therefore, the more uniform time distribution of the TSs, for the feedback approach better dynamic performance is obtained for frames with $\chi = \gamma = 1$.

VI. Conclusion

FD equalization in combination with TA 2×2 multi-input multi-output MIMO CE has been proposed and investigated in a 28-GBaud PDM quadrature phase-shift keying and 16-ary QAM single-carrier optical transmission system.

TS schemes based on the PS-MP CAZAC code, referred as DB-TS and SB-TS, have been reported. For each TS scheme an efficient FD 2×2 MIMO CE method has been introduced. In addition, FDE tap-solutions which maximize the CD and PMD tolerances and minimize the effect of the ASE noise present in the estimated channel have been presented. The FDE has been designed to work with two-fold oversampled signals.

Results have shown that the design and optimization of the 2×2 MIMO FDE need to jointly cope with the pulse shaping desired at the transmitter, the TS scheme to be employed, its length, its GI length, the TOH desired for CE, the equalizer length and the equalizer tap-solution which have to be set taking into account effects such as ASE noise, CD, PMD, time misalignment between received and reference TS, FO between transmitter and receiver lasers and dynamic SOP rotation changing rate.

In conclusion, the energy efficient architecture and the feed-forward filter update make TA 2×2 MIMO FDE highly suitable for implementation in high-speed optical DSP-based coherent receivers. Furthermore, its transparency with respect to the modulation format of the payload and its fast convergence allow for software-defined flexible transponders.

References

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

[2] F. N. Hauske, M. Kuschnerov, B. Spinnler, and B. Lankl, "Optical performance monitoring in digital coherent receivers," *J. Lightw. Technol.*, vol. 27, no. 16, pp. 3623–3631, Aug. 2009.

[3] B. Spinnler, "Equalizer design and complexity for digital coherent receivers," *IEEE J. Sel. Topics Quantum Electron.*, vol. 16, no. 5, pp. 1180–1192, Sep./Oct. 2010.

[4] R. A. Soriano, F. N. Hauske, N. G. Gonzalez, Z. Zhuhong, Y. Ye, and I. T. Monroy, "Chromatic dispersion estimation in digital coherent receivers," *J. Lightw. Technol.*, vol. 29, no. 11, pp. 1627–1637, Jun. 2011.

[5] C. Malouin, P. Thomas, B. Zhang, J. O'Neil, and T. Schmidt, "Natural expression of the best-match search godard clock-tone algorithm for blind chromatic dispersion estimation in digital coherent receivers," presented at the Adv. Photon. Congr., Signal Process. Photon. Commun. Top. Meet., Colorado Springs, CO, USA, Jun. 2012, Paper SpTh2B.4.

[6] C. Malouin, M. Arabaci, P. Thomas, B. Zhang, T. Schmidt, and R. Marcoccia, "Efficient, non-data-aided chromatic dispersion estimation via generalized, FFT-based sweep," presented at the Opt. Fiber Commun./Nat. Fiber Opt. Eng. Conf., Anaheim, CA, USA, Mar. 2013, Paper JW2A.45.

[7] Q. Sui, A. P. Tao Lau, and C. Lu, "Fast and robust blind chromatic dispersion estimation using auto-correlation of signal power waveform for digital coherent systems," *J. Lightw. Technol.*, vol. 31, no. 2, pp. 306–312, Jan. 2013.

[8] N. Stojanovic, B. Mao, and F. Karinou, "Efficient and low-complexity chromatic dispersion estimation in coherent optical systems," in *Proc. 21st Telecommun. Forum*, Belgrade, Serbia, Nov. 2013, pp. 153–156.

[9] R. Corsini, A. Peracchi, E. Matarazzo, T. Foggi, J. Nijhof, G. Meloni, L. Poti, R. Magri, and E. Ciaramella, "Blind adaptive chromatic dispersion compensation and estimation for DSP-based coherent optical systems," *J. Lightw. Technol.*, vol. 31, no. 13, pp. 2131–2139, Jul. 2013.

[10] I. Slim, A. Mezghani, L. G. Baltar, J. Qi, F. N. Hauske, and J. A. Nossek, "Delayed single-tap frequency-domain chromatic-dispersion compensation," *IEEE Photon. Technol. Lett.*, vol. 25, no. 2, pp. 167–170, Jan. 2013.

[11] K. Ishihara, R. Kudo, T. Kobayashi, A. Sano, Y. Takatori, T. Nakagawa, and Y. Miyamoto, "Frequency-domain equalization for coherent optical transmission systems," presented at the Opt. Fiber Commun./Nat. Fiber Opt. Eng. Conf., Los Angeles, CA, USA, Mar. 2011, Paper OWW4.

[12] C. R. S. Fludger, T. Duthel, D. Van den Borne, C. Schulien, E. D. Schmidt, J. G. T Wuth, E. De Man, G. D. Khoe, and H. De Waardt, "Coherent equalization and POLMUX-RZ-DQPSK for robust 100-GE transmission," *J. Lightw. Technol.*, vol. 26, no. 1, pp. 64–72, Jan. 2008.

[13] P. J. Winzer, A. H. Gnauck, C. R. Doerr, M. Magarini, and L. L. Buhl, "Spectrally efficient long-haul optical networking using 112-Gb/s polarization-multiplexed 16-QAM," *J. Lightw. Technol.*, vol. 28, no. 4, pp. 547–556, Feb. 2010.

[14] J. Zhang, J. Yu, N. Chi, Z. Dong, J. Yu, X. Li, L. Tao, and Y. Shao, "Multi-modulus blind equalizations for coherent quadrature duobinary spectrum shaped PM-QPSK digital signal processing," *J. Lightw. Technol.*, vol. 31, no. 7, pp. 1073–1078, Apr. 2013.

[15] M. Kuschnerov, K. Piyawanno1, B. Spinnler, P. Kainzmaier, M. S. Alfiad, and A. N. and B. Lankl, "Data-aided versus blind single-carrier coherent receivers," *IEEE Photon. J.*, vol. 2, no. 3, pp. 386–403, Jun. 2010.

[16] F. Pittalà, F. N. Hauske, Y. Ye, N. G. Gonzalez, and I. T. Monroy, "Data-aided frequency-domain 22 MIMO equalizer for 112 gbit/s PDM-QPSK coherent transmission systems," presented at the Opt. Fiber Commun./Nat. Fiber Opt. Eng. Conf., Los Angeles, CA, USA, Mar. 2012, Paper OM2H.4.

[17] F. Pittalà, M. Msallem, F. N. Hauske, Y. Ye, I. T. Monroy, and J. A. Nossek, "Frequency domain training-aided channel estimation and equalization in time-varying optical transmission systems," presented at the IEEE Photon. Conf., San Francisco, CA, USA, Sep. 2012, Paper WE4.

[18] F. Pittalà, F. N. Hauske, Y. Ye, N. G. Gonzalez, and I. T. Monroy, "Fast and robust CD and DGD estimation based on data-aided channel estimation," presented at the 13th Int. Conf. Transparent Opt. Netw., Stockholm, Sweden, Jun. 2011, Paper We.D1.5.

[19] F. Pittalà, A. Mezghani, F. N. Hauske, Y. Ye, I. T. Monroy, and J. A. Nossek, "Efficient training-based channel estimation for coherent optical communication systems," presented at the Adv. Photon. Congr., Signal Process. Photon. Commun Top. Meet., Colorado Springs, CO, USA, Jun. 2012, Paper SpTu3A.4.

[20] C. C. Chan, *Optical Performance Monitoring: Advanced Techniques for Next-Generation Photonic Networks*. New York, NY, USA: Elsevier, 2010.

[21] T. Duthel, C. R. S. Fludger, J. C. Geyer, and C. Schulien, "Impact of polarisation dependent loss on coherent POLMUX-NRZ-DQPSK," presented at the Opt. Fiber Commun./Nat. Fiber Opt. Eng. Conf., San Diego, CA, USA, Feb. 2008, Paper OThU5.

[22] G. P. Agrawal, *Fiber-Optic Communication Systems*. New York, NY, USA: Wiley, 2002.

[23] U. H. Rohrs and L. P. Linde, "Some unique properties and applications of perfect squares minimum phase CAZAC sequences," in *Proc. South Afr. Symp. Commun. Signal Process*, Cape Down, South Africa, Sep. 1992, pp. 155–160.

[24] R. L. Frank and S. A. Zadoff, "Phase shift pulse codes with good periodic correlation properties (correspondence)," *IRE Trans. Inf. Theory*, vol. 8, pp. 381–382, Oct. 1962.

[25] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1451–1458, Oct. 1998.

[26] M. Kuschnerov, F. N. Hauske, K. Piyawanno, B. Spinnler, M. S. Alfiad, A. Napoli, and B. Lankl, "DSP for coherent single-carrier receivers," *J. Lightw. Technol.*, vol. 27, no. 16, pp. 3614–3622, Aug. 2009.

[27] F. Pittalà, J. Qi, and J. A. Nossek, "Joint frame synchronization and frequency offset estimation in coherent optical transmission systems," presented at the Eur. Conf. Opt. Commun, London, U.K., Sep. 2013, Paper Mo.4.D.4.

[28] A. J. Viterbi and A. M. Viterbi, "Nonlinear estimation of PSK-modulated carrier phase with application to burst digital transmission," *IEEE Trans. Inf. Theory*, vol. 29, no. 4, pp. 543–551, Jul. 1983.

[29] I. Fatadin, D. Ives, and S. J. Savory, "Carrier phase recovery for 16-QAM using QPSK partitioning and sliding window averaging," *IEEE Photon. Technol. Lett.*, vol. 26, no. 9, pp. 854–857, May 2014.

[30] F. Pittalà, F. N. Hauske, Y. Ye, I. T. Monroy, and J. A. Nossek, "Training-based channel estimation for signal equalization and OPM in 16-QAM optical transmission systems," presented at the Eur. Conf. Opt. Commun., Amsterdam, The Netherlands, Sep. 2012, Paper P3.16.

[31] F. Pittalà and J. A. Nossek, "Training-aided frequency offset estimation in 16-QAM Nyquist transmission systems," presented at the IEEE Photon. Conf., Bellevue, WA, USA, Sep. 2013, Paper MG2.3.

Fabio Pittalà (S'12) received the M.Sc. degree in telecommunication engineering from the Technical University of Denmark, Kongens Lyngby, Denmark, in 2011. He is currently working toward the Ph.D. degree at Technische Universität München, Munich, Germany. He is also working as a Researcher at the European Research Center, Huawei Technologies, Munich.

His research interests include high-speed digital signal processing for optical transmission systems with emphasis on synchronization, channel estimation, channel equalization and optical performance monitoring.

Israa Slim (S'10) received the M.Sc. degree in communications engineering from the Technische Universität München (TUM), Munich, Germany, in 2009. Since then, she has been working toward the Ph.D. degree at the Institute of Circuit Theory and Signal Processing, TUM, where she is also a Researcher.

Her research interests include high-speed digital signal processing for optical transmission systems with emphasis on channel equalization, filter design, and performance monitoring. She is also working on multirate systems with emphasis on system models and equalization.

Amine Mezghani (S'08) received the "Diplome d'Ingénieur" degree from École Centrale de Paris, Paris, France, and the Dipl.-Ing. degree in electrical engineering from the Technische Universität München (TUM), Munich, Germany, both in 2006. Since then, he has been working toward the Ph.D. degree at the Institute of Circuit Theory and Signal Processing, TUM, where he is also a Researcher.

His research interests include the study of wireless communications, information theory, and signal processing under low-precision analog-to-digital converters.

Josef A. Nossek (S'72–M'74–SM'81–F'93–LF'12) received the Dipl.-Ing. and Dr. techn. degrees in electrical engineering from the Vienna University of Technology, Vienna, Austria, in 1974 and 1980, respectively. He joined SIEMENS AG, Munich, Germany, in 1978, where he was involved in the design of both passive and active filters for communication systems. In 1978, he became the Supervisor, and in 1980, the Head of a Group of Laboratories involve in designing monolithic filters (analog and digital). Since 1982, he has been the Head of a Group of Laboratories designing digital radio systems with the Transmission Systems Department, SIEMENS AG. In 1984, he was a Visiting Professor at the University of Capetown, Capetown, Africa. From 1987 to 1989, he was the Head of the Radio Systems Design Department, where he was instrumental in introducing high-speed VLSI signal processing into digital microwave radio. Since April 1989, he has been a Professor of circuit theory and design with the Technische Universität München (TUM), Munich, where he teaches undergraduate and graduate courses in the field of circuit and system theory and leads research on signal processing algorithms in communications, particularly multiantenna communication systems.

Dr. Nossek was the President Elect, President, and Past President of the IEEE Circuits and Systems Society in 2001, 2002, and 2003, respectively. He was the Vice President of Verband der Elektrotechnik, Elektronik und Informationstechnik e.V. (VDE) 2005 and 2006, and was the President of VDE in 2007 and 2008. He received the ITG Best Paper Award in 1988, the Mannesmann Mobilfunk (currently Vodafone) Innovations Award in 1998, and the Award for Excellence in Teaching from the Bavarian Ministry for Science, Research and Art in 1998. From the IEEE Circuits and Systems Society, he received the Golden Jubilee Medal for Outstanding Contributions to the Society in 1999 and the Education Award in 2008. He received the Bundesverdienstkreuz am Bande in 2008. In 2009, he became an elected Member of Acatech. In 2010, he received the Guillemin-Cauer Best Paper Award of the IEEE Circuits and Systems Society.