Digital backpropagation accounting for polarization-mode dispersion

CRISTIAN B. CZEGLEDI,[1,*] GABRIELE LIGA,[2] DOMANIÇ LAVERY,[2] MAGNUS KARLSSON,[3] ERIK AGRELL,[1] SEB J. SAVORY,[4] AND POLINA BAYVEL[2]

[1]*Dept. of Signals and Systems, Chalmers University of Technology, Gothenburg, SE-412 96, Sweden*
[2]*Optical Networks Group, Dept. of Electronic and Electrical Engineering, University College London, London, WC1E 7JE, U.K.*
[3]*Dept. of Microtechnology and Nanoscience, Chalmers University of Technology, Gothenburg, SE-412 96, Sweden*
[4]*Dept. of Engineering, Electrical Engineering Division, University of Cambridge, Cambridge, CB3 0FA, U.K.*
[*]czegledi@chalmers.se

Abstract: Digital backpropagation (DBP) is a promising digital-domain technique to mitigate Kerr-induced nonlinear interference. While it successfully removes deterministic signal–signal interactions, the performance of ideal DBP is limited by stochastic effects, such as polarization-mode dispersion (PMD). In this paper, we consider an ideal full-field DBP implementation and modify it to additionally account for PMD; reversing the PMD effects in the backward propagation by passing the reverse propagated signal also through PMD sections, which concatenated equal the inverse of the PMD in the forward propagation. These PMD sections are calculated analytically at the receiver based on the total accumulated PMD of the link estimated from channel equalizers. Numerical simulations show that, accounting for nonlinear polarization-related interactions in the modified DBP algorithm, additional signal-to-noise ratio gains of 1.1 dB are obtained for transmission over 1000 km.

References and links

1. Y. Yamamoto, Y. Kawaguchi, and M. Hirano, "Low-loss and low-nonlinearity pure-silica-core fiber for C- and L-band broadband transmission," J. Lightw. Technol.**34**, 321–326 (2016).
2. X. Liu, A. R. Chraplyvy, P. J. Winzer, R. W. Tkach, and S. Chandrasekhar, "Phase-conjugated twin waves for communication beyond the Kerr nonlinearity limit," Nat. Photon. **7**, 560–568 (2013).
3. S. L. Jansen, D. van den Borne, P. M. Krummrich, S. Sälter, G.-D. Khoe, and H. de Waardt, "Long-haul DWDM transmission systems employing optical phase conjugation," IEEE J. Sel. Topics Quantum Electron. **12**, 505–520 (2006).
4. Y. Gao, J. C. Cartledge, A. S. Karar, S. S.-H. Yam, M. O'Sullivan, C. Laperle, A. Borowiec, and K. Roberts, "Reducing the complexity of perturbation based nonlinearity pre-compensation using symmetric EDC and pulse shaping," Opt. Express **22**, 1209–1219 (2014).
5. M. Secondini and E. Forestieri, "On XPM mitigation in WDM fiber-optic systems," IEEE Photon. Technol. Lett. **26**, 2252–2255 (2014).
6. R. J. Essiambre, P. J. Winzer, X. Q. Wang, W. Lee, C. A. White, and E. C. Burrows, "Electronic predistortion and fiber nonlinearity," IEEE Photon. Technol. Lett. **18**, 1804–1806 (2006).
7. E. Ip, "Nonlinear compensation using backpropagation for polarization-multiplexed transmission," J. Lightw. Technol. **28**, 939–951 (2010).
8. G. P. Agrawal, *Nonlinear Fiber Optics* (Academic, 2013), 5th ed.
9. E. Temprana, E. Myslivets, L. Liu, V. Ataie, A. Wiberg, B. P. P. Kuo, N. Alic, and S. Radic, "Two-fold transmission reach enhancement enabled by transmitter-side digital backpropagation and optical frequency comb-derived information carriers." Opt. Express **23**, 20774–83 (2015).
10. R. Maher, T. Xu, L. Galdino, M. Sato, A. Alvarado, K. Shi, S. J. Savory, B. C. Thomsen, R. I. Killey, and P. Bayvel, "Spectrally shaped DP-16QAM super-channel transmission with multi-channel digital back-propagation," Nat. Sci. Rep. **5**, 8214 (2015).
11. D. Lavery, D. Ives, G. Liga, A. Alvarado, S. J. Savory, and P. Bayvel, "The benefit of split nonlinearity compensation for optical fiber communications," IEEE Photon. Technol. Lett. **28**, 1803–1806 (2016).

12. D. Rafique and A. D. Ellis, "Impact of signal–ASE four-wave mixing on the effectiveness of digital back-propagation in 112 Gb/s PM-QPSK systems," Opt. Express **19**, 3449–3454 (2011).
13. F. Yaman and G. Li, "Nonlinear impairment compensation for polarization-division multiplexed WDM transmission using digital backward propagation," IEEE Photonics Journal **2**, 816–832 (2010).
14. G. Gao, X. Chen, and W. Shieh, "Influence of PMD on fiber nonlinearity compensation using digital back propagation," Opt. Express **20**, 14406–14418 (2012).
15. K. Goroshko, H. Louchet, and A. Richter, "Fundamental limitations of digital back propagation due to polarization mode dispersion," in "Proc. of Asia Communications and Photonics Conference (ACP)," (Hong Kong, China, 2015), p. ASu3F.5.
16. G. Liga, T. Xu, A. Alvarado, R. I. Killey, and P. Bayvel, "On the performance of multichannel digital backpropagation in high-capacity long-haul optical transmission," Opt. Express **22**, 30053–30062 (2014).
17. G. Liga, C. B. Czegledi, T. Xu, E. Agrell, R. I. Killey, and P. Bayvel, "Ultra-wideband nonlinearity compensation performance in the presence of PMD," in "Proc. of European Conference on Optical Communication (ECOC)," (Düsseldorf, Germany, 2016), pp. 794–796.
18. N. V. Irukulapati, H. Wymeersch, P. Johannisson, and E. Agrell, "Stochastic digital backpropagation," IEEE Trans. Commun. **62**, 3956–3968 (2014).
19. I. Fernandez de Jauregui Ruiz, A. Ghazisaeidi, P. Tran, and G. Charlet, "Impact of polarization mode dispersion on digital nonlinear compensation algorithms in dispersion unmanaged systems," in "Proc. of Optical Fiber Communication Conference (OFC)," (Anaheim, CA, 2016), p. Th3D.3.
20. I. Fernandez de Jauregui Ruiz, A. Ghazisaeidi, E. Awwad, P. Tran, and G. Charlet, "Polarization effects in nonlinearity compensated links," in "Proc. of European Conference on Optical Communication (ECOC)," (Düsseldorf, Germany, 2016), pp. 379–381.
21. K. Goroshko, H. Louchet, and A. Richter, "Overcoming performance limitations of digital back propagation due to polarization mode dispersion," in "Proc. of International Conference on Transparent Optical Networks (ICTON)," (Trento, Italy, 2016), p. Mo.B1.4.
22. C. B. Czegledi, G. Liga, D. Lavery, M. Karlsson, E. Agrell, S. J. Savory, and P. Bayvel, "Polarization-mode dispersion aware digital backpropagation," in "Proc. of European Conference on Optical Communication (ECOC)," (Düsseldorf, Germany, 2016), pp. 1091–1093.
23. C. B. Czegledi, G. Liga, D. Lavery, M. Karlsson, E. Agrell, S. J. Savory, and P. Bayvel, "Modified digital backpropagation accounting for polarization-mode dispersion," to appear in "Proc. of Optical Fiber Communication Conference (OFC)," (Los Angeles, CA, 2017), p. W1G.6.
24. S. J. Savory, "Digital coherent optical receivers: algorithms and subsystems," IEEE J. Sel. Topics Quantum Electron. **16**, 1164–1179 (2010).
25. H. Louchet, K. Kuzmin, and A. Richter, "Improved DSP algorithms for coherent 16-QAM transmission," in "Proc. of European Conference on Optical Communication (ECOC)," (Brussels, Belgium, 2008), p. Tu.1.E.6.
26. M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions With Formulas, Graphs, and Mathematical Tables* (Dover Publications, 1964).
27. R. Bellman, *Introduction to Matrix Analysis* (McGraw-Hill, 1960).
28. R. Haber and H. Unbehauen, "Structure identification of nonlinear dynamic systems—a survey on input/output approaches," Automatica **26**, 651–677 (1990).
29. D. Marcuse, C. R. Manyuk, and P. K. A. Wai, "Application of the Manakov-PMD equation to studies of signal propagation in optical fibers with randomly varying birefringence," J. Lightw. Technol. **15**, 1735–1746 (1997).
30. C. B. Czegledi, M. Karlsson, E. Agrell, and P. Johannisson, "Polarization drift channel model for coherent fibre-optic systems," Nat. Sci. Rep. **6**, 21217 (2016).
31. C. H. Prola, J. A. Pereira da Silva, A. O. Dal Forno, R. Passy, J. P. von der Weid, and N. Gisin, "PMD emulators and signal distortion in 2.48-Gb/s IM-DD lightwave systems," IEEE Photon. Technol. Lett. **9**, 842–844 (1997).
32. W. Kabsch, "A solution for the best rotation to relate two sets of vectors," Acta Crystallographica Section A **32**, 922–923 (1976).
33. A. J. Laub, *Matrix Analysis for Scientists & Engineers* (Society for Industrial and Applied Mathematics, 2005).

1. Introduction

The ever-increasing global demand for digital-data traffic has resulted in an accelerated development of the optical networks. One of the key technologies supporting this demand is coherent optical detection, which has reached maturity and is widely deployed. Among the many benefits of this technology, it facilitates the use of polarization multiplexing and the recovery of all degrees of freedom of the optical field, thus improving receiver sensitivity and allowing the use of higher-order modulation formats. Furthermore, it naturally suits the use of digital signal processing (DSP) tools to compensate for channel impairments and hardware imperfections.

Despite the fact that the optical fiber channel is fundamentally nonlinear, historically, the

maximum achievable data rates through optical fibers were limited by linear impairments, such as chromatic dispersion (CD) and polarization-mode dispersion (PMD). However, in current systems, these impairments are effectively mitigated using DSP, and the intensity-dependent Kerr nonlinearity is suggested to be the ultimate obstacle to increase optical fiber transmission rates. Due to the nonlinear effects, there exists an optimal (effective) signal-to-noise ratio (SNR) point (where it is recommended to operate), which is a compromise between the additive noise and the nonlinear interference.

Various nonlinear mitigation techniques are currently under investigation, which can be categorized into two groups: optical and digital techniques. Optical solutions such as large effective area fibers [1] allow higher signal launch powers, thus increasing the available SNR, but transmission is still performed avoiding the nonlinear regime. Other optical solutions include optical phase conjugation using twin waves [2] or mid-span optical phase conjugation [3]. The drawback of the former is that it does not utilizes all degrees of freedom of the optical field efficiently, hence the spectral efficiency is reduced. The latter requires a symmetric map of both dispersion and power evolution along the link, thus making it fruitless in deployed systems. Various digital techniques are available in the literature to mitigate fiber nonlinearities, including perturbation-based precompensation [4], Kalman equalization [5], and digital backpropagation (DBP) [6, 7], which has proved significantly beneficial in experimental demonstrations.

DBP compensates for the deterministic nonlinear fiber impairments by solving the nonlinear propagation equation using the split-step Fourier (SSF) method [8, Sec. 2.4.1] and backpropagating the received optical field with inverted channel parameters. Several variations of DBP have been proposed and demonstrated including transmitter- [9], receiver-side [10], and both transmitter- and receiver-side [11] compensation. Having exact knowledge of the fiber parameters, it is believed that the deterministic nonlinear signal–signal interactions are completely removed using DBP and that the performance of a fiber-optical system is limited by the uncompensated stochastic effects, such as amplified spontaneous emission noise, which leads to signal–noise interactions [12], and PMD, leading to polarization-dependent nonlinear interactions [13–17], which considerably undermine the effectiveness of DBP. In order to account for the signal–noise interactions, a modified DBP algorithm [18] has been proposed that takes into account the additive noise, getting the performance of the optical fiber channel closer to the fundamental limits. On the other hand, PMD introduces a frequency-dependent delay that accumulates as a random-walk-like process along the fiber length and it is usually compensated for at the receiver after DBP. When applying DBP, the entire reverse propagation is performed with the accumulated PMD over the entire link; therefore the nonlinear compensation is mismatched and its accuracy degrades with the backpropagated distance. However, it has been shown that low-complexity DBP implementations (one or less steps per span) that backpropagate a single channel are marginally affected by polarization effects since the nonlinear signal–signal interactions are not entirely removed and dominate the achievable performance [19, 20]. On the other hand, considering an ideal DBP with sufficient number of steps where PMD is an issue, in order to remove its effects, PMD should be compensated for as it naturally occurs, i.e., in a distributed fashion along the link, rather than doing it once after DBP. It has been shown numerically that compensating for PMD on a per-span basis decreases its impact on DBP significantly [13]. However, this approach requires *a priori* PMD knowledge for every span, which is not known at the receiver.

Recently, a modified DBP algorithm that takes into account PMD was proposed in [21], where an appropriate amount of differential group delay (DGD) is introduced at the link principal states of polarization (PSP) after each span in the reverse propagation, such that the accumulation of DGD in the forward propagation is reversed. In [22], we recently proposed a modified DBP method that reverses the PMD effects in the backward propagation by passing the reverse propagated signal through PMD sections. Unlike [21], where the DGD is subtracted along the same PSP every time in the backward propagation, in [22] the PSPs of the backward PMD

sections are different and aligned using an optimization algorithm such that the Jones matrix modeling the total backward PMD equals the inverse of the Jones matrix modeling the PMD occurring in the forward propagation.

In [23], we proposed a DBP method accounting for PMD, where the backward PMD sections are computed analytically using a first-order linearization approach, which we extend in this work. Compared to [23], herein we detail the analysis of the proposed method and extend the algorithm to wavelength-division-multiplexed (WDM) systems.

2. Proposed algorithm

The conventional DBP algorithm is modified such that the signal is backpropagated also through PMD sections. These sections are chosen such that, when cascaded, they match the overall PMD accumulated over the entire link, as observed at the receiver but with opposite sign. This method gradually reverses the PMD effects, mimicking the accumulation of PMD in the forward propagation. The accumulated PMD at the receiver can be obtained from (blind) channel equalizers, such as the constant modulus algorithm (CMA) [24] or the multiple modulus algorithm (MMA) [25].

The inverse of the accumulated PMD at the receiver can be modeled in baseband by a frequency-dependent 2×2 complex-valued Jones matrix $\mathbf{J}_k(f)$, where k denotes the time index. In order to reverse the PMD effects distributively, we divide $\mathbf{J}_k(f)$ into a product of $N_{\text{PMD}} + 1$ matrices as

$$\mathbf{J}_k(f) = \mathbf{J}_k(f_0) + (f - f_0)\frac{\mathbf{J}'_k(f_0)}{1!} + (f - f_0)^2\frac{\mathbf{J}''_k(f_0)}{2!} + \cdots \tag{1}$$

$$\approx \mathbf{J}_k(f_0) + (f - f_0)\mathbf{J}'_k(f_0) \tag{2}$$

$$= \mathbf{J}_k(f_0)\left(\mathbf{I} + (f - f_0)\mathbf{J}_k^{-1}(f_0)\mathbf{J}'_k(f_0)\right) \tag{3}$$

$$\approx \mathbf{J}_k(f_0)\exp\left((f - f_0)\mathbf{J}_k^{-1}(f_0)\mathbf{J}'_k(f_0)\right) \tag{4}$$

$$= \mathbf{J}_k(f_0)\mathbf{J}_k^{\text{s}}(f)^{N_{\text{PMD}}}, \tag{5}$$

where

$$\mathbf{J}_k^{\text{s}}(f) = \exp\left(\frac{(f - f_0)\mathbf{J}_k^{-1}(f_0)\mathbf{J}'_k(f_0)}{N_{\text{PMD}}}\right). \tag{6}$$

In Eq. (1), $\mathbf{J}_k(f)$ is expanded into its Taylor series [26, p. 880] evaluated at f_0 and then approximated to the first order in Eq. (2). The $(\cdot)'$, $(\cdot)''$, ... operators denote the first, second, ... derivatives with respect to frequency. In Eq. (4), $\mathbf{I} + (f - f_0)\mathbf{J}_k^{-1}(f_0)\mathbf{J}'_k(f_0)$ is approximated using the matrix exponential [27, p. 165] defined as

$$\exp \mathbf{A} = \mathbf{I} + \frac{\mathbf{A}}{1!} + \frac{\mathbf{A}^2}{2!} + \ldots, \tag{7}$$

which converges for any $\mathbf{A} \in \mathbb{C}^{n \times n}$, and $\mathbf{I}$ is the identity matrix, whereas $(\cdot)^{-1}$ denotes the inverse operator. Due to the unitary nature of $\mathbf{J}_k(f)$, the inverse can be obtained by the conjugate-transpose operation $\mathbf{J}_k^{-1}(f) = \mathbf{J}_k^{\text{H}}(f)$ for any f and k. Here we introduce the matrix exponential to facilitate the factorization of Eq. (4) in Eq. (5). Although higher order terms of $(f - f_0)$ are reintroduced in Eq. (4) with the wrong coefficient by the matrix exponential, thus introducing another approximation, counterintuitively, this is a better approximation of Eq. (1) than Eq. (2), where the higher order terms are absent. Unfortunately, we were not able to provide an analytic proof of this result, but observed it only numerically. Nevertheless, the approximation in Eq. (4) is accurate for small $(f - f_0)\mathbf{J}_k^{-1}(f_0)\mathbf{J}'_k(f_0)$; consequently it is inaccurate for large bandwidths, i.e., large $|f - f_0|$, or strong PMD, i.e., large $\mathbf{J}'_k(f_0)$, resulting in a performance penalty. In Eq. (5), the
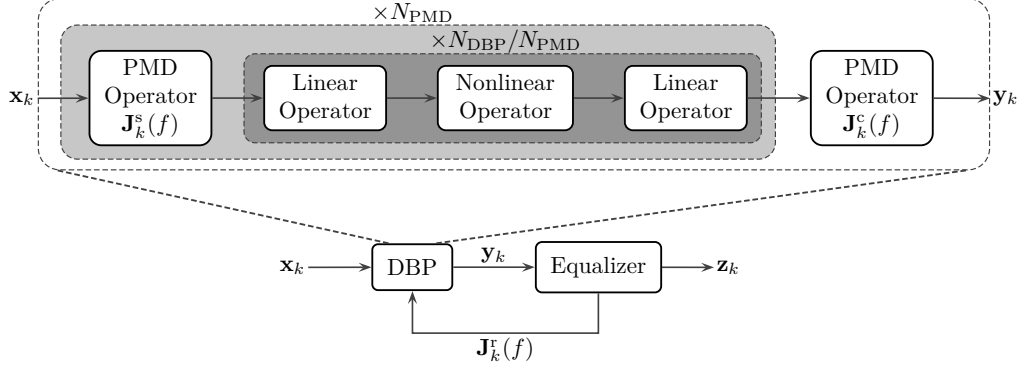
Fig. 1: Schematic of the proposed DBP method, where N_{DBP} is the number of steps used by the DBP algorithm over the entire link and the equalizer is a conventional adaptive channel equalizer such as the CMA or MMA. For brevity, the blocks modeling the amplifiers and attenuation are not shown.

exponential matrix is divided into N_{PMD} identical frequency-dependent matrices $\mathbf{J}_k^{\text{s}}(f)$. These matrices are inserted equidistantly in the DBP procedure, as illustrated in the top left of Fig. 1, to reverse the PMD effects in the backward propagation. As a last step, the result should be left-multiplied with $\mathbf{J}_k(f_0)$ to approximate $\mathbf{J}_k(f)$ as in Eq. (5), but as we will see below, we replace $\mathbf{J}_k(f_0)$ with $\mathbf{J}_k^{\text{c}}(f)$ to reduce the effects of approximations.

The approximations involved in deriving Eq. (5) hold tightly around f_0, which is chosen to be in the middle of the signal bandwidth, but the error diverges proportionally to the accumulated DGD as f deviates from f_0. Therefore, compensating for PMD by applying the approximation Eq. (5) instead of the true matrix $\mathbf{J}_k(f)$ leads to a penalty. However, this penalty can be mitigated by replacing $\mathbf{J}_k(f_0)$ in Eq. (5) with a frequency-dependent correction matrix

$$\mathbf{J}_k^{\text{c}}(f) = \mathbf{J}_k(f)\mathbf{J}_k^{\text{s}}(f)^{-N_{\text{PMD}}} \tag{8}$$

as illustrated in the top right of Fig. 1. Under ideal conditions (no nonlinear effects), this correction perfectly compensates for the approximations in Eqs. (2) and (4) and brings back $\mathbf{J}_k(f)$.

The modified DBP algorithm operates in a streaming block-wise fashion processing a block of samples $\mathbf{x}_k$ at a time and it compensates for PMD using the same Jones matrix $\mathbf{J}_k(f)$ over the entire block $\mathbf{x}_k$. Since PMD is a time-varying stochastic process and might not be constant over $\mathbf{x}_k$, a residual PMD after DBP might occur due to these temporal variations. This residue, denoted $\mathbf{J}_k^{\text{r}}(f)$, is mitigated by the channel equalizer and is periodically fed back to the DBP through a feedback loop to update $\mathbf{J}_k(f) = \mathbf{J}_k^{\text{r}}(f)\mathbf{J}_{k-1}(f)$. In this fashion, it is ensured that the processed samples are PMD-free and that the PMD Jones matrix $\mathbf{J}_k(f)$ implemented by the DBP algorithm is periodically updated, tracking the temporal variations of the PMD.

The schematic diagram of the proposed algorithm is shown in Fig. 1. A typical coherent receiver might include several different blocks between DBP and the equalizer, such as matched filtering, frequency offset compensation, or resampling, which are skipped here since they are not relevant to this work. The DBP algorithm receives as inputs a block of received samples $\mathbf{x}_k$ and the PMD residue $\mathbf{J}_k^{\text{r}}(f)$, and outputs the processed block of samples $\mathbf{y}_k$, which is then further processed by the equalizer and output as $\mathbf{z}_k$. Initially, at the first block of samples $k = 1$, the DBP algorithm operates as in the conventional approach, without compensation for PMD ($\mathbf{J}_1(f) = \mathbf{I}$). Thereafter, $\mathbf{J}_k^{\text{r}}(f)$ is estimated and fed back by the equalizer, which for $k = 1$ is the entire PMD of the link since the DBP did not compensate for it. Consequently, PMD will be compensated for in the DBP routine and $\mathbf{J}_k^{\text{r}}(f)$ will consist of only residual PMD.
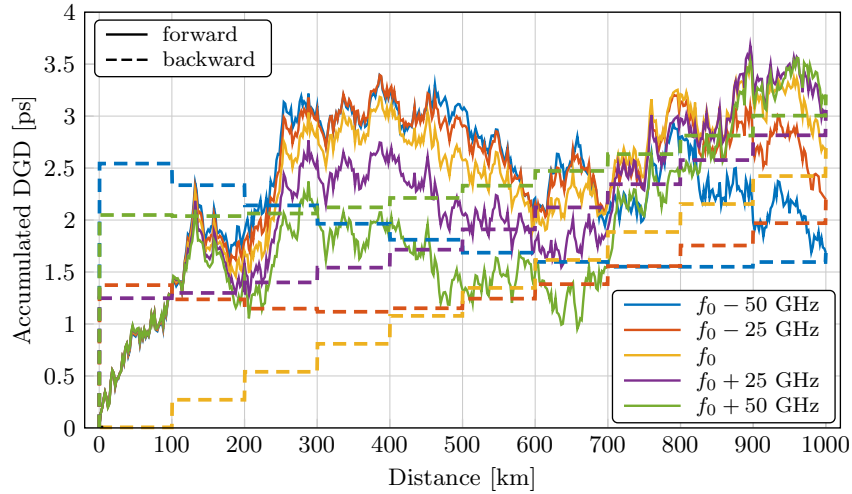
Fig. 2: The evolution of the accumulated DGD at different frequencies versus distance in the forward and backward propagation. The PMD parameter is 0.1 ps/$\sqrt{\text{km}}$ and $N_{\text{PMD}} = 10$.

The polarization-dependent nonlinear interactions, and thereby the algorithm's efficiency, are immune to temporal variations of the absolute state of polarization and are sensitive only to the frequency dependence of $\mathbf{J}_k(f)$. Therefore, fluctuations of the state of polarization, which drift much faster than the frequency dependence of $\mathbf{J}_k(f)$, do not have to be fed back by the equalizer, reducing the feedback frequency.

Figure 1 also illustrates the internal structure of the proposed DBP method. The conventional DBP algorithm is confined in the dark gray box, whereas two extra operators are added to account for PMD. The $\mathbf{J}_k^{\text{s}}(f)$ operator is nested in the DBP procedure and applied N_{PMD} times, while the $\mathbf{J}_k^{\text{c}}(f)$ operator is applied only once at the end. It should be noted that if N_{PMD} is chosen to be equal to the number of spans, the $\mathbf{J}_k^{\text{s}}(f)$ operator is applied at the beginning of each span in the backward propagation. In this paper, we chose the Wiener–Hammerstein model [28], but this is not a general requirement. The algorithm can be implemented using the Wiener or the Hammerstein models [28].

As mentioned earlier, $\mathbf{J}_k^{\text{r}}(f)$ is obtained from channel equalizers. However, in general, a coherent receiver estimates and cancels PMD using a time-domain complex-valued finite-impulse-response (FIR) filter bank arranged in a butterfly structure. Therefore, in order to calculate the PMD operators $\mathbf{J}_k^{\text{s}}(f)$ and $\mathbf{J}_k^{\text{c}}(f)$ used by the modified DBP, the time-domain filter coefficients of the equalizer must be converted to the frequency-domain Jones matrix $\mathbf{J}_k^{\text{r}}(f)$. Moreover, the derivative $\mathbf{J}_k'(f_0)$ must be calculated numerically in this case from noisy $\mathbf{J}_k(f)$. In the Appendix, we present a method to obtain the discrete-frequency unitary matrix $\mathbf{J}_k^{\text{r}}(f)$ from the filter coefficients of the equalizer, and an approach to calculate $\mathbf{J}_k'(f_0)$.

Figure 2 illustrates an example of the evolution of the accumulated DGD in the forward propagation and in the backward propagation applying the modified DBP scheme using Eqs. (6) and (8). As can be seen, different frequency components accumulate different amounts of DGD in the forward propagation in a random-walk-like fashion. In the backward propagation, we set $N_{\text{PMD}} = 10$, therefore the $\mathbf{J}_k^{\text{s}}(f)$ operator is applied at every multiple of 100 km starting from distance 1000 km, whereas $\mathbf{J}_k^{\text{c}}(f)$ is applied last, at distance 0 km. The DGD at f_0, where the first order approximation was performed, decreases after every $\mathbf{J}_k^{\text{s}}(f)$ operation with a constant step equal to $1/N_{\text{PMD}}$ of the total DGD at 1000 km. On the other hand, due to the approximations involved in deriving Eq. (5), the DGDs at the other four frequencies do not decrease monotonically

and can, in some cases, increase. Since the approximations are tighter around f_0, the DGDs at $f_0 \pm 25$ GHz have a lower residual DGD at 0 km compared to the ones at $f_0 \pm 50$ GHz. This residual DGD is corrected by applying $\mathbf{J}_k^c(f)$ at 0 km. As can be seen, this operation has no impact on the DGD at f_0, since there is no approximation in this case and the accumulated DGD has been successfully removed already at 100 km. The average DGD grows proportionally to the square root of the distance in the forward propagation. We replicated this in our study for the backward direction, i.e., instead of a linear decrease of the argument of the matrix exponential (which quantifies the DGD), we have chosen the decrease rate according to the inverse of the square root. But we observed minor gains compared to the linear decrease and decided not to include it as it involves extra complexity since the $\mathbf{J}_k^s(f)$ operators will be different w.r.t. distance.

Although the evolution of the DGD in the backward propagation does not match its evolution in the forward propagation, as we will see in the Results section, this method does provide performance gains over the conventional approach. However, due to this mismatch between the two evolutions, the performance does not match the case without PMD. In the conventional approach, the backward propagation is performed with the DGD frozen to the values accumulated over the entire link, i.e., at 1000 km in Fig. 2. However, we would like to remind the reader that the DGD does not fully quantify PMD, therefore the match between the forward and backward evolution of the DGD cannot be fully regarded as a performance metric. Moreover, the example plotted in Fig. 2 is just one realization of the stochastic process PMD, and for other realizations the forward and backward DGD evolutions will match closely or loosely.

Judging by the step-wise evolution of the DGD in the backward direction, our approach seems to be similar to the one proposed in [21]. However, the two methods are different. Opposed to [21], where in the backward propagation $\mathbf{J}_k^s(f)$ is a diagonal matrix modeling a DGD element applied at the same PSP along the entire link, our approach changes the PSPs with every $\mathbf{J}_k^s(f)$ in the backward direction.

In WDM systems, the bandwidth of the signal is large and, as we saw earlier, $\mathbf{J}_k^s(f)$ becomes inaccurate for large bandwidths. This inaccuracy can be lessened by, instead of approximating $\mathbf{J}_k(f)$ around one central frequency f_0, approximating $\mathbf{J}_k(f)$ for each channel individually around their center frequency. The center frequency f_0 of each linearization Eq. (6) can be chosen in the middle of each channel's bandwidth. However, this approach introduces discontinuities between the $\mathbf{J}_k^s(f)$ operators of different channels. For example, if N_{ch} WDM channels with a frequency separation of w are considered, the frequency span (including the guard band) of the ith channel is $f \in [f_0^i - w/2, f_0^i + w/2]$ for $i = 1, \ldots, N_{\text{ch}}$, where f_0^i denotes the central frequency of the ith channel. The discontinuities in $\mathbf{J}_k^s(f)$ will emerge at $f = f_0^1 + w/2 + iw$ for $i = 0, \ldots, N_{\text{ch}} - 1$. These discontinuities affect the performance and should be removed. To do so, a continuous $\breve{\mathbf{J}}_k^s(f)$ operator can be calculated by first computing the $\mathbf{J}_k^s(f)$ operator according to Eq. (6) for $f \in [f_0^i - w/2, f_0^i + w/2]$ for each channel i. Here we chose to apply the linearization Eq. (6) once per channel. Nevertheless, this can be applied more than once, but we observed no benefits from doing so in our simulations. Thereafter, we set $\breve{\mathbf{J}}_k^s(f) = \mathbf{J}_k^s(f)$ for $f \in [f_0^{\lfloor N_{\text{ch}}/2 \rfloor} - w/2, f_0^{\lfloor N_{\text{ch}}/2 \rfloor} + w/2]$ at the central channel $\lfloor N_{\text{ch}}/2 \rfloor$, where $\lfloor \cdot \rfloor$ indicates the flooring function. The continuous operator $\breve{\mathbf{J}}_k^s(f)$ of the outer channels is calculated recursively as

$$\breve{\mathbf{J}}_k^s(f) = \breve{\mathbf{J}}_k^s(f_0^{i+1} - w/2)\mathbf{J}_k^s(f_0^i + w/2)^{-1}\mathbf{J}_k^s(f) \text{ for } i = \lfloor N_{\text{ch}}/2 \rfloor - 1, \ldots, 2, 1, \quad (9)$$

$$\breve{\mathbf{J}}_k^s(f) = \breve{\mathbf{J}}_k^s(f_0^{i-1} + w/2)\mathbf{J}_k^s(f_0^i - w/2)^{-1}\mathbf{J}_k^s(f) \text{ for } i = \lfloor N_{\text{ch}}/2 \rfloor + 1, \ldots, N_{\text{ch}}, \quad (10)$$

and $f \in [f_0^i - w/2, f_0^i + w/2]$, where in Eqs. (9) and (10), the recursion is backwards and forwards, respectively. Notice that the correction matrix has to be calculated accordingly in this case by using $\breve{\mathbf{J}}_k^s(f)$ in Eq. (8). It should be noted that the linearization of each WDM channel becomes less accurate by making $\breve{\mathbf{J}}_k^s(f)$ continuous in Eqs. (9) and (10). However, the continuity of $\breve{\mathbf{J}}_k^s(f)$ is more critical for the performance than having the linearization of each WDM channel intact.

Complexity-wise, the computational increase over the conventional DBP is marginal. The calculations of the $\mathbf{J}_k^s(f), \tilde{\mathbf{J}}_k^s(f), \mathbf{J}_k^c(f)$ operators can be performed in parallel to the data-processing stream. On the other hand, applying these operators to the data requires a single complex multiplication and can be performed after compensating for the linear effects in the frequency domain.

3. Simulation setup

We study through numerical simulations a point-to-point transmission link consisting of an ideal transmitter and coherent receiver, and N_s spans of 100 km standard single-mode fiber with: attenuation 0.2 dB/km, CD parameter 17 ps/(nm·km), and nonlinear coefficient 1.2 1/(W·km). Each span is followed by one erbium-doped fiber amplifier with a noise figure of 4.5 dB, compensating for the exact span loss. The transmitted signal consists of N_{ch} polarization-multiplexed 16-ary quadrature-amplitude modulated channels at 50 Gbaud shaped using a root-raised cosine (RRC) pulse with roll-off factor 0.01 and spaced at 50.1 GHz. The signal was oversampled by a factor of 2 and each channel consists of 2^{15} symbols. The signal propagation was simulated by solving the Manakov-PMD equation [29] using the SSF approach with steps of 0.1 km. Static PMD was emulated at every SSF step consisting of a polarization scrambler, which uniformly [30] scatters the state of polarization, and a retardation plate. The DGD introduced by each retardation plate was Gaussian distributed with mean $\Delta\tau_p$ and standard deviation $\Delta\tau_p/5$ [31]; thus the mean accumulated DGD is $\bar{\tau} = \sqrt{8N_{SSF}/(3\pi)}\Delta\tau_p$, where N_{SSF} is the total number of SSF steps. In order to capture the stochastic nature of PMD, 120 fiber realizations were simulated for each set of parameters, except Fig. 4(b) where we simulated 500 realizations to increase the accuracy of the scatter plot.

We consider two receiver DSP setups: i) conventional DBP followed by a linear PMD equalizer, and ii) modified DBP described in the previous section using Eqs. (6), (8–10) with $N_{PMD} = 10$, except in Fig. 4(a) where N_{PMD} is varied. For both setups, we consider the equalizer to operate under perfect PMD knowledge and then we compare this performance with results obtained using the blind MMA equalizer in Fig. 5(b). The length of equalizer's FIR filters was $N_{taps} = 31$ with two samples per symbol, and the numerical derivative $\mathbf{J}_k'(f_0)$ is averaged over 2/3 of the bandwidth of the signal, i.e., $L = 1/(3T\Delta f)$ (see appendix). DBP is performed with the same SSF step distribution and sampling rate as in the forward propagation for both setups. Full-field DBP is applied, i.e., all channels are backpropagated. DBP is followed by an ideal matched RRC filter applied to the signal, after which the linear-scale SNR is averaged over the two polarizations by comparing the transmitted and received symbols. Namely, the SNR of each polarization is evaluated as the ratio between the variance of the transmitted complex symbols ($\mathbb{E}[|U|^2]$) and variance of the noise ($\mathbb{E}[|U-V|^2]$), where V denotes the received symbols after DSP. Note that this definition of the SNR over one polarization is related to the error vector magnitude (EVM) according to SNR $\approx 1/\text{EVM}^2$.

4. Results and discussion

In order to study the effectiveness of the proposed algorithm, we consider different PMD scenarios by studying two transmission distances: 1000 (10×100) km and 4000 (40×100) km; two different simulated bandwidths: 50 GHz ($N_{ch} = 1$ channel) and 350 GHz ($N_{ch} = 7$ channels); and two PMD parameters: 0.04 and 0.1 ps/$\sqrt{\text{km}}$, corresponding to typical modern optical fibers. These PMD parameters yield an average accumulated DGD of $\bar{\tau} = 1.26$ ps and $\bar{\tau} = 3.16$ ps, respectively, for 1000 km, and $\bar{\tau} = 2.52$ ps and $\bar{\tau} = 6.32$ ps, respectively, for 4000 km.

Figure 3 shows the achieved performance obtained for the studied scenarios. In Fig. 3(a), the single-channel scenario is considered and, as can be seen, DBP obtains gains of 9.2 and 6.5 dB compared to CDC for the two transmission distances of 1000 and 4000 km, respectively, when PMD is not considered. This performance degrades in the presence of PMD with parameter 0.04
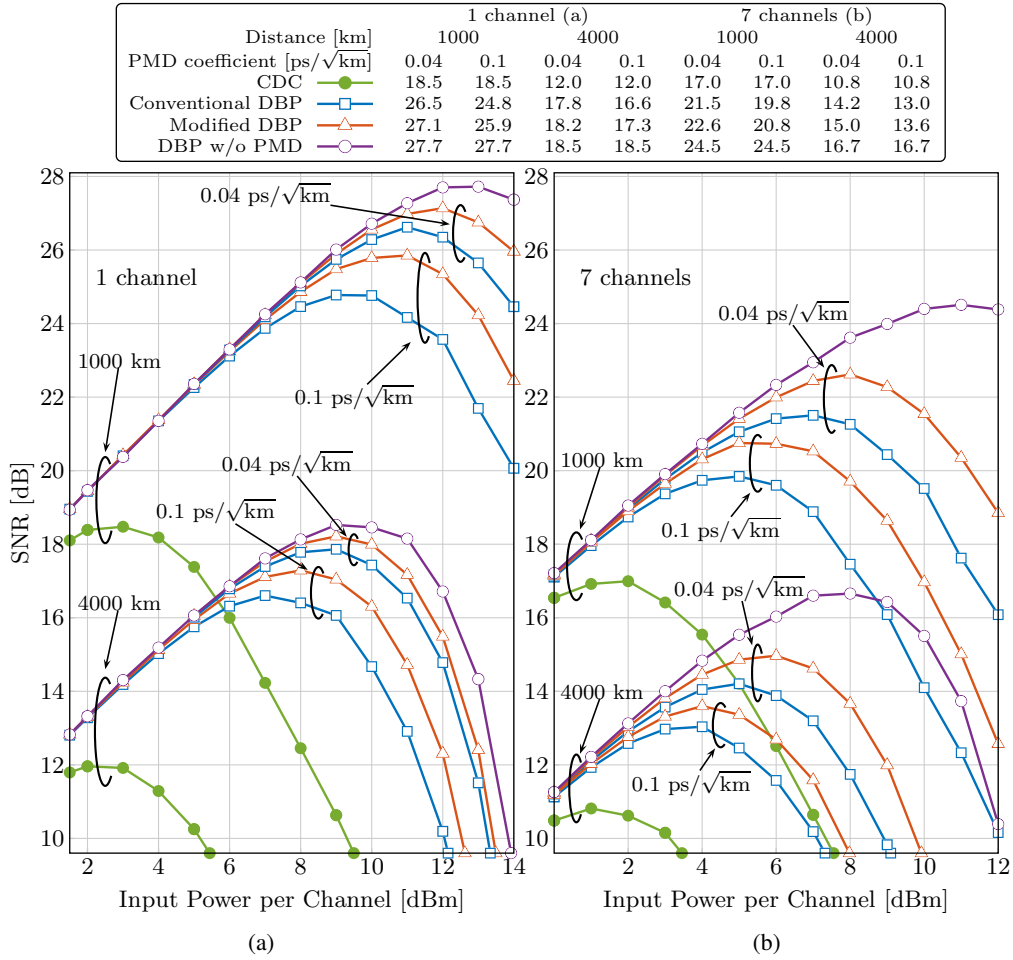
	1 channel (a)				7 channels (b)			
Distance [km]	1000		4000		1000		4000	
PMD coefficient [ps/$\sqrt{km}$]	0.04	0.1	0.04	0.1	0.04	0.1	0.04	0.1
CDC	18.5	18.5	12.0	12.0	17.0	17.0	10.8	10.8
Conventional DBP	26.5	24.8	17.8	16.6	21.5	19.8	14.2	13.0
Modified DBP	27.1	25.9	18.2	17.3	22.6	20.8	15.0	13.6
DBP w/o PMD	27.7	27.7	18.5	18.5	24.5	24.5	16.7	16.7

Fig. 3: Average SNR versus input power per channel for various setups: ((a)) $N_{\mathrm{ch}} = 1$ and $N_{\mathrm{s}} = \{10, 40\}$; ((b)) $N_{\mathrm{ch}} = 7$ and $N_{\mathrm{s}} = \{10, 40\}$. For comparison, the performance of CD compensation (CDC) only and DBP, both without PMD, are shown. The maximum (in dB) of each curve from the plots is summarized in the legend above.

and 0.1 ps/$\sqrt{km}$ by 1.2 and 2.9 dB, respectively, for 1000 km, and by 0.7 and 1.9 dB, respectively, for 4000 km. In Fig. 3(b), the seven-channel scenario is studied, where DBP obtains gains of 7.5 and 5.9 dB compared to CDC for the two transmission distances (1000, 4000 km) when PMD is not considered. Same behaviour as above, the performance of DBP degrades in the presence of PMD with parameter 0.04 and 0.1 ps/$\sqrt{km}$ by 3 and 4.7 dB, respectively, for 1000 km, and by 2.5 and 3.7 dB, respectively, for 4000 km. Without PMD, DBP achieves the highest gains when both the distance and bandwidth are the smallest due to the weaker signal–noise nonlinear interactions. PMD has a stronger impact on the performance as the bandwidth and/or the PMD parameter increases, a phenomenon also observed in [17], where PMD incurred a penalty of 16 dB for a transmission of approximately 1 THz bandwidth. However, it is interesting to notice that the degradation of the DBP performance due to PMD relative to the gain obtained without PMD is constant with respect to the distance. In the single-channel scenario (Fig. 3(a)), PMD reduces the SNR gains of DBP for both distances by approximately 10% and 30% when the PMD parameter is 0.04 and 0.1 ps/$\sqrt{km}$, respectively. On the other hand, in the seven-channel scenario
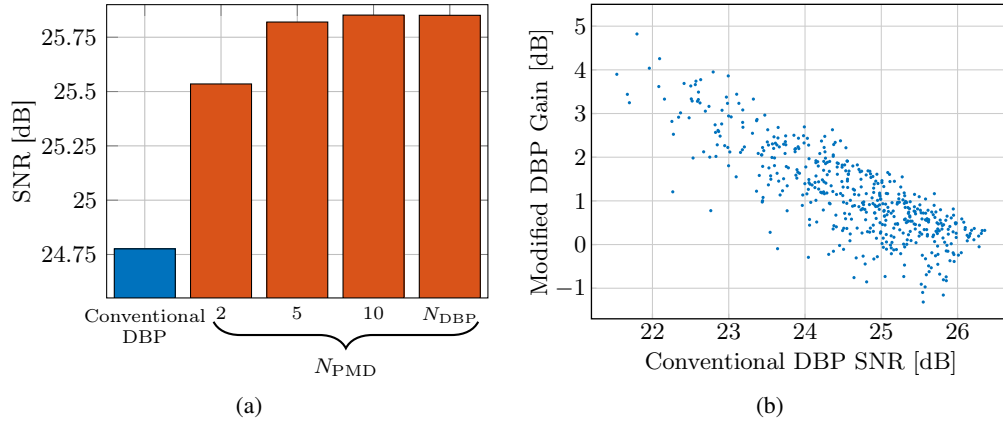
Fig. 4: (a): Average SNR obtained at the optimal input power by the conventional DBP and by the modified DBP scheme with varying number of PMD sections N_{PMD}.
(b): Scatter plot of the achieved SNR gain by the proposed DBP modification versus the SNR attained by the conventional DBP obtained over 500 fiber realizations at the optimum input power 11 and 10 dBm, respectively.

(Fig. 3(b)), the gains are reduced by approximately 40% and 60% when the PMD parameter is 0.04 and 0.1 ps/$\sqrt{km}$, respectively, irrespective of the distance.

The proposed modified DBP algorithm improves the optimum SNR in the single-channel scenario (Fig. 3(a)) by 0.6 and 1.1 dB for 1000 km, and by 0.4 and 0.7 dB for 4000 km compared to the conventional DBP approach when the PMD parameter is 0.04 and 0.1 ps/$\sqrt{km}$, respectively. These gains represent 50% and 37% of the SNR reduction due to PMD, regardless of the distance. In the seven-channel scenario (Fig. 3(b)), the modified DBP improves the optimum SNR by 1.1 and 1 dB for 1000 km, and by 0.8 and 0.6 dB for 4000 km, with PMD parameter 0.04 and 0.1 ps/$\sqrt{km}$, respectively. Here the gains represent 14% and 12% of the SNR reduction due to PMD. It is likely that the performance of the proposed algorithm decreases for larger bandwidths and higher PMD parameters due to the approximations used to derive Eq. (5), which becomes less accurate.

In [22], an average gain of only 0.7 dB (cf. 1.1 dB) was obtained in the single-channel configuration for a PMD coefficient of 0.1 ps/$\sqrt{km}$ and distance 1000 km shown in Fig. 3(a). There the backward PMD sections are chosen quasi-randomly and obtained from an optimization algorithm such that the Jones matrix modeling the total backward PMD equals the inverse of the Jones matrix modeling the PMD occurring in the forward propagation. This leads to large performance fluctuations since the optimization algorithm outputs both sequences of PMD sections that match (almost) exactly the one in the forward propagation, and sequences that are the opposite and perform poorly, lowering the average gain. On the other hand, the performance of the method proposed here does not exhibit these two extremes of poor and ideal, which is an advantage and disadvantage, respectively. It would be desirable to use [22] if one could distinguish between good and bad solutions of the optimization algorithm to maximize performance.

From now on, the presented results are based on the single-channel configuration over 1000 km and PMD parameter 0.1 ps/$\sqrt{km}$, except in Fig. 5(a) where the PMD parameter is varied.

The achieved SNR at the optimal input power is shown in Fig. 4(a) as a function of N_{PMD}. As can be seen, the performance increases with N_{PMD}, providing a gain of 1.1 dB for $N_{PMD} = 10$ (same as the number of spans), after which it saturates, providing negligible extra gain.

Figure 4(b) illustrates the scatter plot of the achieved SNR gain by the proposed DBP scheme
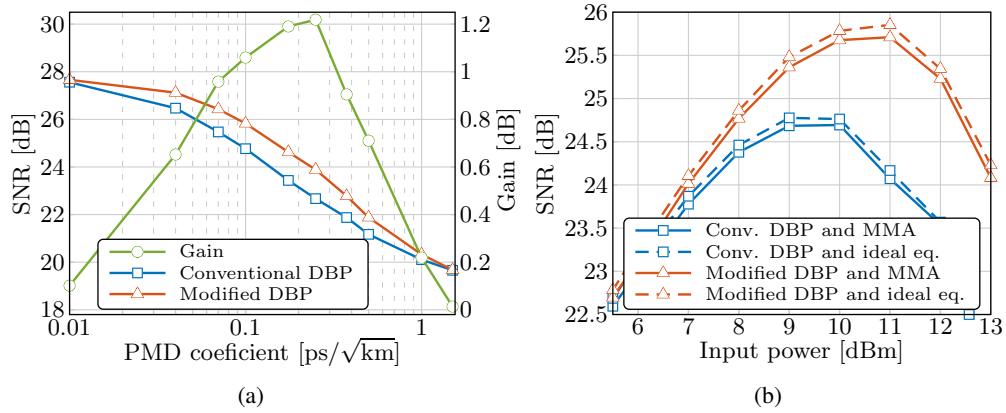
Fig. 5: (a): Average SNR and SNR gain versus fiber PMD parameter obtained at the optimum input power for each scenario. Note the y-axis on the right for the gain.
(b): Results obtained using the blind MMA equalizer compared to an ideal equalizer for both DBP schemes.

versus the SNR of the conventional DBP at the optimum input power of 11 and 10 dBm, respectively, for different PMD realizations. As can be seen, the gain is positive in the majority of the cases (>90%), and negative gain mainly occurs when the achieved SNR by the conventional DBP is relatively high. From the same figure, it can also be inferred that the worst/best SNR achieved by the proposed method is 2/0.5 dB better than the conventional DBP, thus improving the required SNR margin and lowering the outage probability.

Figure 5(a) shows the average performance of the two schemes as a function of the PMD parameter at the optimal input power for each case. As the PMD parameter increases, the performance of both schemes degrades significantly from 27.6 dB down to 19.6 dB when the PMD coefficient is 1.5 ps/√km. The proposed DBP method provides a minor gain (0.1 dB) at 0.01 ps/√km, which increases up to a maximum of 1.2 dB at 0.25 ps/√km. After this point, the gain declines and becomes 0.1 dB at 1.5 ps/√km. The performance of DBP in the weak PMD regime is lightly affected by PMD; therefore the proposed algorithm provides small gains. On the other hand, it is likely that the gain deteriorates at high PMD parameters due to the approximation used to derive Eq. (5), which becomes less accurate. However, the proposed DBP method provides gains greater than 0.8 dB over the range of PMD coefficient 0.04–0.4 ps/√km, which covers most of the modern optical fibers.

So far, the presented results were obtained assuming that the equalizer has ideal knowledge of the accumulated PMD. In Fig. 5(b), results obtained using ideal equalization are compared with results based on the blind MMA equalizer and the processing procedure presented in the Appendix. It can be seen that the MMA induces a penalty of about 0.1 dB, which is approximately constant over the entire range of input power and is the same for both schemes. This penalty is likely due to the inability of the equalizer to perfectly estimate the PMD. In general, the equalizer does not always return the true Jones matrix $\mathbf{J}_k(f)$, but a matrix that minimizes the equalizer's cost function. Due to the symmetries of the polarization-multiplexed constellations, the returned matrix by the equalizer might be affected by phase and polarization ambiguities, therefore the returned matrix is approximately $\mathbf{J}_k(f)$ multiplied with a constant matrix modeling these ambiguities. However, the proposed algorithm is not affected by these ambiguities since the performance relies only on the dependence of $\mathbf{J}_k(f)$ with frequency.

5. Conclusions

We studied the achievable performance by ideal full-field DBP in systems with PMD. We saw that the performance of DBP degrades with bandwidth and the strength of PMD. However, interestingly, we observed that the ratio between the achieved gains by DBP considering PMD and the gains obtained without PMD does not change with respect to the transmission distance for all studied PMD parameters and bandwidth. Moreover, in order to improve the performance of DBP in the presence of PMD, we proposed a simple modification of the DBP algorithm to take into account polarization effects by blindly reversing the PMD effects in the backward propagation. The algorithm has been proved to work with PMD information obtained from the blind MMA equalizer and provides SNR gains over an ample range of different transmission setups. Interesting directions for future research include performance characterization in the presence of a time-varying PMD, as well as a comprehensive study including [21,22] to investigate which of the three methods offers the best performance for various simulation setups and channel conditions.

Appendix

In this Appendix is presented the methodology of obtaining the PMD residue $\mathbf{J}_k^{\mathrm{r}}(f)$ from the FIR filters of the equalizer and a numerical method to calculate the derivative $\mathbf{J}_k'(f_0)$. In the discussion above, the matrices modeling PMD were defined in the continuous-frequency domain. Since DSP operates on finite-length discrete-time or discrete-frequency signals, this analysis is carried in the discrete domain. We distinguish between continuous and discrete functions adopting the notation $f(\cdot)$ and $f[\cdot]$, respectively.

The filter impulse response of the equalizer can be represented by a discrete-time 2×2 complex-valued matrix

$$\mathbf{T}_k^{\mathrm{eq}}[n] = \begin{pmatrix} T_k^{\mathrm{xx}}[n] & T_k^{\mathrm{yx}}[n] \\ T_k^{\mathrm{xy}}[n] & T_k^{\mathrm{yy}}[n] \end{pmatrix} \tag{11}$$

sampled at time instants nT_s, where x and y denote the two degenerate polarizations, T_s is the sampling period, and $n = 0, \ldots, N_{\mathrm{taps}} - 1$, whereas N_{taps} is the length of the FIR filter. Typically, an oversampling of two samples per symbol is employed, i.e., $T = 2T_s$, where T is the symbol duration. The impulse response of the filter is usually two-sided centered at $\lceil N_{\mathrm{taps}}/2 \rceil$, where $\lceil \cdot \rceil$ indicates the ceiling function.

The discrete-frequency representation of $\mathbf{T}_k^{\mathrm{eq}}[n]$ can be obtained by taking the element-wise discrete Fourier transform, denoted $\mathscr{F}\{\cdot\}$, with respect to n

$$\mathbf{J}_k^{\mathrm{eq}}[m] = \begin{pmatrix} \mathscr{F}\left\{T_k^{\mathrm{xx}}[n]\right\} & \mathscr{F}\left\{T_k^{\mathrm{yx}}[n]\right\} \\ \mathscr{F}\left\{T_k^{\mathrm{xy}}[n]\right\} & \mathscr{F}\left\{T_k^{\mathrm{yy}}[n]\right\} \end{pmatrix}, \tag{12}$$

which is the discrete representation of $\mathbf{J}_k^{\mathrm{eq}}(f)$ sampled with resolution $\Delta f = 1/(T_s N_{\mathrm{taps}})$ at frequency instants $m\Delta f$ for $m = -\lfloor N_{\mathrm{taps}}/2 \rfloor, \ldots, \lceil N_{\mathrm{taps}}/2 \rceil - 1$, i.e., $\mathbf{J}_k^{\mathrm{eq}}(m\Delta f)$. Typically, N_{taps} is chosen to be less than 100, resulting in a matrix $\mathbf{J}_k^{\mathrm{eq}}[m]$ with a coarse frequency-domain resolution, i.e., large Δf. In order to refine this resolution, $\mathbf{T}_k^{\mathrm{eq}}[n]$ can be symmetrically padded with zeros, which increases N_{taps}, before applying Eq. (12) such that the desired resolution is obtained. The finer resolution of $\mathbf{J}_k^{\mathrm{eq}}[m]$ also improves the accuracy of the numerical differentiation, which is discussed below.

However, $\mathbf{T}_k^{\mathrm{eq}}[n]$ may compensate for polarization-dependent losses (PDL), which will be reflected in $\mathbf{J}_k^{\mathrm{eq}}[m]$ through Eq. (12). Considering the unitary nature of PMD, PDL can be removed

from $\mathbf{J}_k^{\text{eq}}[m]$ by finding its closest unitary matrix [32]

$$\mathbf{J}_k^{\text{r}}[m] = \mathbf{U}_k[m]\mathbf{V}_k^{\text{H}}[m], \tag{13}$$

which is the discrete-frequency representation of $\mathbf{J}_k^{\text{r}}(f)$ and can be used in the DBP routine as described in Sec. 2. In Eq. (13), the columns of $\mathbf{U}_k[m]$, $\mathbf{V}_k[m]$ are the left-, right-singular vectors of $\mathbf{J}_k^{\text{eq}}[m]$, i.e., $\mathbf{J}_k^{\text{eq}}[m] = \mathbf{U}_k[m]\Sigma_k[m]\mathbf{V}_k^{\text{H}}[m]$ is the singular-value decomposition [33, p. 35] of $\mathbf{J}_k^{\text{eq}}[m]$.

When calculating $\mathbf{J}_k^{\text{s}}(f)$ and $\mathbf{J}_k^{\text{c}}(f)$ in Eqs. (6–8), the derivative of $\mathbf{J}_k(f_0)$ is needed. Usually numerical differentiation is done using the straightforward two-point derivative $\mathbf{J}_k'[m] = (\mathbf{J}_k[m + 1] - \mathbf{J}_k[m])/\Delta f$. However, this derivative is very sensitive to noisy signals, but it can be improved by using multipoint differentiation. In the most general case, the multipoint differentiation can be approximated as

$$\mathbf{J}_k'[m] \approx \frac{1}{\Delta f} \sum_{l=1}^{L} c_l (\mathbf{J}_k[m + l] - \mathbf{J}_k[m - l]), \tag{14}$$

where c_l are the filters coefficients. In this paper, we chose c_l as

$$c_l = \frac{3(L + 1 - l)}{L(L + 1)(L + 2)}, \tag{15}$$

so that the biggest weight is given to the samples in the vicinity of $\mathbf{J}_k'[m]$ and then the weight decreases linearly with l.

Funding

Acknowledgments