Perturbation-aided deep neural network for dual-polarization optical communication systems

Xiang Lin[1], Shenghang Luo[2], Octavia A. Dobre[1], Lutz Lampe[2], Deyuan Chang[3], and Chuandong Li[3]
[1] **Faculty of Engineering and Applied Science, Memorial University, St. John's, NL, Canada**
[2] **Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC,Canada**
[3] **Huawei Technologies Canada, Ottawa, ON, Canada**
Xiang.Lin@mun.ca

Abstract: We propose a perturbation-aided deep neural network for fiber nonlinearity compensation in polarization-multiplexed optical communication systems. The proposed technique achieves a fast convergence that is facilitated by the perturbation analysis and attains an enhanced performance.

OCIS codes: (060.2330) Fiber optic communications; (060.1660) Coherent communications.

1. Introduction

Fiber nonlinearity has been the major barrier against the performance improvement of high speed, long-haul optical fiber communication systems. A number of nonlinearity compensation (NLC) techniques has been proposed in the literature, such as the digital back-propagation (DBP), Volterra series-based equalizer, and those based on perturbation theory. These techniques attain a good performance, at the price of high complexity. Recently, machine learning techniques have paved a new pathway to design complexity and performance-balanced NLC techniques [1–3]. Among these, the learned DBP (LDBP) that unfolds the conventional DBP into a deep neural network (DNN) has attracted much interest [3–5]. The LDBP parametrizes the conventional DBP, and optimizes the parameters through supervised training. LDBP achieves a significantly improved balanced complexity-performance tradeoff than DBP, and its model-based nature facilitates the hyper-parameter selection in the DNN. However, it is noticed that the number of layers in LDBP needs to increase with the number of fiber spans for successful nonlinearity compensation, which also renders training of the LDBP DNN challenging [6].

In [7], we proposed a perturbation-aided (PA) LDBP to improve the nonlinearity compensation performance at each nonlinear step by incorporating the self phase modulation (SPM) and intra-channel cross phase modulation (IXPM) effects based on the first-order perturbation theory. In this paper, we extend this approach to polarization-multiplexed (PM) systems. To deal with the polarization effects, two more layers are introduced after the chromatic dispersion (CD) compensation layer at each step in the DNN. Specifically, the first layer accounts for the differential group delay (DGD), while the second one recovers the state of polarization. Furthermore, the cross-polarization modulation (CPM), in addition to the SPM and IXPM, is incorporated into the DNN. The resulting PA-LDBP for PM systems achieves a significantly-improved performance and enables a flexible structure with multiple spans per step. Finally, it is demonstrated that the initialization of DNN weights with perturbation coefficients reduces the training effort by helping the DNN converge to a favourable solution.

2. The Proposed PA-LDBP

Without loss of generality, we consider an arbitrary step in the PA-LDBP for PM systems. Taking the h polarization as an example, the first-order nonlinear distortion to the nth sample is given as $\Delta_{n,h} = \Gamma_{n,h} + \Phi_{n,h}$ with $\Gamma_{n,h}$ and $\Phi_{n,h}$ defined in (1) and (2), respectively. $\Gamma_{n,h}$ represents the intra-channel four-wave mixing effect, expressed as

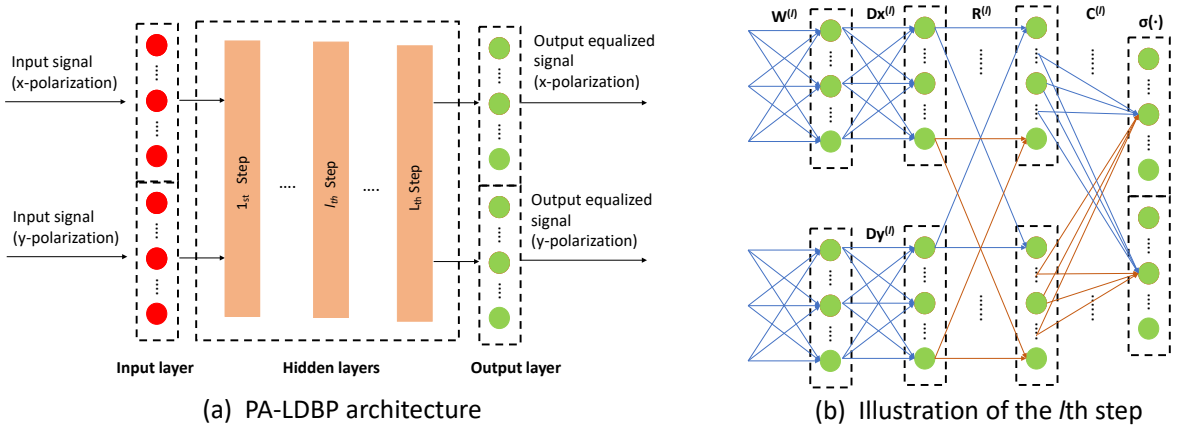
$$\Gamma_{n,h} = jP^{\frac{3}{2}}\left[\sum_{m\neq 0,k\neq 0} x^{\mathrm{CD}}_{n+k}\left(x^{\mathrm{CD}}_{n+m+k}\right)^{*} x^{\mathrm{CD}}_{n+m} C_{m,k} + \sum_{m\neq 0,k} y^{\mathrm{CD}}_{n+k}\left(y^{\mathrm{CD}}_{n+m+k}\right)^{*} x^{\mathrm{CD}}_{n+m} C_{m,k}\right], \tag{1}$$

where P is the launch power, x^{CD}_k and y^{CD}_k are the samples in the h and v polarization, respectively, and $C_{m,k}$ denotes the perturbation coefficient. $\Phi_{n,h}$ includes the IXPM, CPM, and SPM effects:

$$\Phi_{n,h} = jP^{\frac{3}{2}} x^{\mathrm{CD}}_n\left[2\sum_{k\neq 0}|x^{\mathrm{CD}}_{n+k}|^2 C_{0,k} + \sum_{k}|y^{\mathrm{CD}}_{n+k}|^2 C_{0,k} + |x^{\mathrm{CD}}_n|^2 C_{0,0}\right]. \tag{2}$$

To compensate for the IXPM, CPM, and SPM effects, the nonlinear operation, $\sigma(x^{\mathrm{CD}}_n)$, in PA-LDBP is given by

$$x^{\mathrm{CD}}_n - \Phi_{n,h} = x^{\mathrm{CD}}_n(1 - j\phi_{n,h}) \approx x^{\mathrm{CD}}_n \exp(-j\phi_{n,h}) = x^{\mathrm{CD}}_n \exp\left(-jP^{\frac{3}{2}}\left((2\mathbf{x}^{\mathrm{p}}_k + \mathbf{y}^{\mathrm{p}}_k)^T \mathbf{c}_0 - |x^{\mathrm{CD}}_n|^2 C_{0,0}\right)\right), \tag{3}$$



(a) PA-LDBP architecture (b) Illustration of the *l*th step

Fig. 1: Diagram of the proposed PA-LDBP.

where $\phi_{n,h} = P^{\frac{3}{2}}\left[2\sum_{k\neq 0}|x_{n+k}^{\mathrm{CD}}|^2 C_{0,k} + \sum_k |y_{n+k}^{\mathrm{CD}}|^2 C_{0,k} + |x_n^{\mathrm{CD}}|^2 C_{0,0}\right]$, $\mathbf{x}_k^{\mathrm{p}}$ and $\mathbf{y}_k^{\mathrm{p}}$ denote the vectors $[|x_{n-k}^{\mathrm{CD}}|^2, ..., |x_n^{\mathrm{CD}}|^2, ..., |x_{n+k}^{\mathrm{CD}}|^2]^T$ and $[|y_{n-k}^{\mathrm{CD}}|^2, ..., |y_n^{\mathrm{CD}}|^2, ..., |y_{n+k}^{\mathrm{CD}}|^2]^T$, respectively. Here $\mathbf{c}_0$ represents the perturbation coefficient vector, i.e., $[C_{0,k}, ..., C_{0,0}, ..., C_{0,k}]^T$. Note that (3) holds when $\phi_{n,h} \ll 1$. A careful examining of (3) indicates that PA-LDBP involves one more operation when compared to LDBP at each nonlinear step. This operation aims to learn the nonlinearity interaction between a number of adjacent samples. The proposed PA-LDBP is illustrated in Fig. 1, where the superscript (l) represents the lth step, $l = 1, ..., L$; $\mathbf{W}^{(l)}$ is a matrix that accounts for CD compensation at the lth step; $\mathbf{D}_{\mathbf{x}}^{(l)}$ and $\mathbf{D}_{\mathbf{y}}^{(l)}$ aim to handle the DGD; $\mathbf{R}^{(l)}$ copes with the polarization rotation; and $\mathbf{C}^{(l)}$ represents the perturbation coefficients. For $\mathbf{W}^{(l)}$, $\mathbf{D}_{\mathbf{x},\mathbf{y}}^{(l)}$, and $\mathbf{R}^{(l)}$, We follow the design and initialization strategies developed in [3] and [4]. Note that $\mathbf{c}_0$ depends on the transmission distance: a longer distance results in a larger size of $\mathbf{c}_0$.

3. Simulation Setup and Training

A dual polarization single carrier system is considered with a 32 Gbaud 64-quadrature amplitude modulation signal, which is pulse-shaped by a root raised cosine filter with a roll-off factor of 0.1. The link includes 16 spans of single mode fiber (SMF). For each span, the SMF is 80 km, followed by an Erbium-doped fiber amplifier with a 5 dB noise figure and 16 dB gain. The attenuation coefficient is 0.2 dB/km, the dispersion parameter is 17 ps/nm/km, the nonlinear coefficient is 1.3 /W/km, and the maximum individual PMD is 0.1 ps/$\sqrt{\text{km}}$. The propagation is simulated by the split-step Fourier method with 400 steps per span. The PMD effect is emulated with the PMD transfer function in frequency domain, $\prod_{i=1}^{I=6400} \mathbf{A}^i \mathbf{B}^i(\omega)$, where $\mathbf{A}^i$ is the unitary rotation matrix, $\mathbf{A}^i = [\cos\theta, \sin\theta; -\sin\theta, \cos\theta]$, $\theta \in [0, 2\pi)$, and $\mathbf{B}^i(\omega) = \mathrm{diag}(e^{-j\frac{\omega\tau_i}{2}}, e^{j\frac{\omega\tau_i}{2}})$ denotes the first-order PMD with τ_i as the DGD at the ith step. The Q factor, defined as $20\log_{10}(\sqrt{2}\mathrm{erfc}^{-1}(2\mathrm{BER}))$, is used as performance metric, where $\mathrm{erfc}(\cdot)$ is the complementary error function and BER is the system bit error rate.

The PA-LDBP is implemented in TensorFlow, where the training data set includes 256 frames and the testing data set includes 64 frames. Each frame has 2048 samples. Therefore, the Q factor is evaluated from 393,216 bits. The optimizer is Adam with a learning rate of 0.001, and the batch size is 32. The cost function is the mean-squared error defined as $\mathscr{L}(\mathbf{s}, \hat{\mathbf{s}}) = \sum_{n=1}^{N} |s_n - \hat{s}_n|^2 / N$, where s_n and $\hat{s}_n$ are the nth transmitted symbol and the output symbol of DNN after downsampling and phase de-rotation, respectively. To investigate the flexibility of the proposed scheme, we consider the three cases of PA-LDBP: with 1, 2, and 4 spans per step, respectively.

The PA-LDBP for PM systems is trained in two stages. Firstly, we obtain a PA-LDBP model without considering the polarization effects, i.e., the training data does not include polarization rotation or DGD effects and the corresponding compensation modules are not included in the DNN. In the second step, data with polarization effects are used to train the PA-LDBP for PM transmission. The $\mathbf{W}^{(l)}$ are transferred from the previously trained PA-LDBP model, and considered as non-trainable parameters during the second training stage. The rotations $\mathbf{R}^{(l)}$ are randomly initialized and unitary constrained. The initialization of the DGD filters and perturbation coefficient vector depends on the number of spans per step. The DGD filters in both polarizations (the asymmetric DGD filter in the v polarization is a flipped version of that in the h polarization) are initialized symmetrically as [0,...,0,1,0,...,0]; the length of the filter is 5, 9, 19 for the three cases with 1, 2, and 4 spans per step, respectively. The perturbation coefficients are calculated with fiber lengths of 80 km, 160 km, and 320 km, and we found that the optimized length of $\mathbf{c_0}$ for the three cases is 11, 25, and

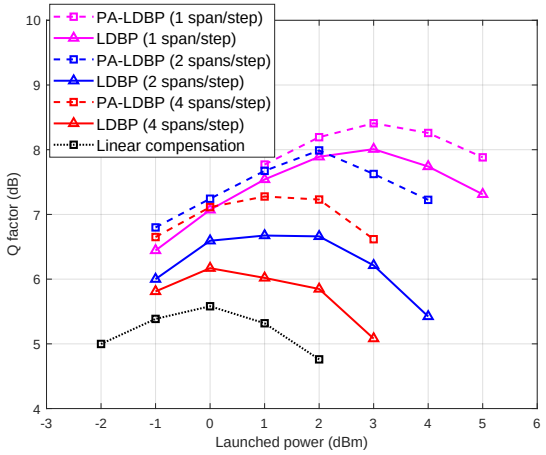


Fig. 2: Performance of PA-LDBP (dashed line) and LDBP (solid line).

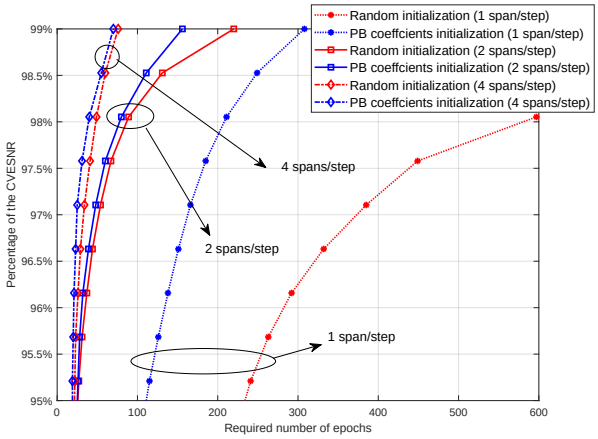


Fig. 3: PA-LDBP convergence behaviour with different initializations.

41, respectively.

4. Results and Discussion

The Q factors of the proposed PA-LDBP and the LDBP basedline are depicted in Fig. 2. The performance gains of PA-LDBP over linear compensation are approximately 2.7 dB, 2.4 dB, and 1.6 dB, when the numbers of spans per step are 1, 2, and 4, respectively. Further, PA-LDBP outperforms LDBP by about 0.3 dB, 1.3 dB, and 1.0 dB in these three cases. This performance improvement comes from the fact that PA-LDBP compensates for IXPM and CPM along with SPM at each nonlinear operation. Similar to LDBP, the performance gains over linear compensation of PA-LDBP for PM systems are less than those for single-polarization shown in [7], due to polarization effects [4]. Notably, PA-LDBP with 2 spans per step attains a similar gain as LDBP with 1 span per step. This is important as it has been shown that the overall complexity of PA-LDBP with 2 spans per step is lower than that of LDBP with 1 span per step [7]. In the cases of multiple spans per step, CD compensation in frequency domain can significantly reduce the complexity.

To investigate the impact of the initialization of the nonlinear step in PA-LDBP on the convergence of the DNN, two initialization cases are studied: the initialization with perturbation coefficients and random initialization. For the former, the perturbation coefficients are calculated based on the link configuration per step [7], and are used to initialize the weights, $\mathbf{C}^{(l)}$. For the latter, the real and imaginary parts are drawn from a normal distribution. The number of epochs required to converge to certain levels of the effective SNR (CVESNR) is presented in Fig. 3. Note that the effective SNR is calculated by $10\log_{10}\left(\mathscr{L}(\mathbf{s},\hat{\mathbf{s}})^{-1}\right)$. From the figure, we notice that a fast and good convergence of the DNN is achieved for the initialization with the perturbation coefficients. This could be an advantage for the elastic optical network where training is required more frequently to cope with the adaptive change of lightpath.

5. Conclusion

In this paper, we have proposed PA-LDBP for dual-polarization coherent optical fiber communication systems. Based on the first-order perturbation theory, the proposed technique improves the nonlinearity compensation at each nonlinear step by incorporating the SPM, CPM, and IXPM effects. When compared to LDBP, the proposed PA-LDBP provides a flexible DNN structure and achieves an enhanced performance.

References

1. C. Häger *et al.*, *Proc. Opt. Fiber Commun. Conf.*, paper W3A.4 (2018).
2. A. Amari *et al.*, IEEE Photon. Technol. Lett. **31**, 627–630 (2019).
3. C. Häger *et al.*, IEEE J. Sel. Areas Commun. **39**, 280–294 (2021).
4. C. Häger *et al.*, *Proc. Opt. Fiber Commun. Conf.*,paper W3D.3 (2020).
5. Q. Fan *et al.*, Nat. Commun. **11**, 1–11 (2020).
6. I. Goodfellow *et al.*, *Deep Learning* (MIT Press Cambridge, 2016).
7. X. Lin *et al.*, arXiv: 2110.05563 (2021).
8. X. Liang *et al.*, Opt. Express **22**, 29,733–29,745 (2014).