Modified Weighted Learned Digital Backpropagation with Pre-optimization in High-symbol-rate Coherent Systems

Du Tang[1], Zhen Wu[1], Xizi Tang[1], Jiating Luo[2], Ji Luo[3], Bofang Zheng[2], and Yaojun Qiao [1*]

[1] *The State Key Laboratory of Information Photonics and Optical Communications, School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China*
[2] *B&P Laboratory, Huawei Technologies Co. Ltd., Shenzhen, China*
[3] *Moscow Optic Algorithm Lab, Huawei Moscow Research Center, Moscow, Russia*
**qiao@bupt.edu.cn*

Abstract: A modified weighted learned digital backpropagation (M-W-LDBP) with pre-optimization is proposed for fiber nonlinearity compensation in high-symbol-rate coherent systems. Compared with LDBP, M-W-LDBP exhibits 1/0.7 dB signal-to-noise ratio gain in 90/128-GBaud systems, respectively.

1. Introduction

Digital backpropagation (DBP), as a standard algorithm for fiber nonlinearity compensation, has been widely studied for a long time. It's common knowledge that due to the nature of the dispersion and nonlinearity interaction, the step size of DBP should be much smaller than the dispersion length $L_D = 1/(|\beta_2|R_S^2)$ and nonlinear length $L_{NL} = 1/(\gamma P_L)$ with β_2, R_S, γ, P_L denoting the group velocity dispersion, symbol rate, nonlinear coefficient and launch power, respectively [1,2]. Apparently, when L_{NL} is fixed and much higher than L_D, the step size of DBP will be dominated by symbol rate. In other words, a higher symbol rate requires a higher-complexity DBP. However, due to the limited capability of digital signal processing (DSP) in practical systems, the application of DBP with high steps per span (StPS) may encounter challenges. If DBP is performed with a limited low StPS, the mismatch between the step size and L_D will become more serious with the increase of symbol rate. It is verified that the benefit of DBP degrades rapidly with the increase of symbol rate [3]. Hence, studying the effectiveness of low-complexity DBP in high-symbol-rate (HSR) systems is significant.

Recently, by leveraging the powerful neural network (NN), an improved version of DBP called learned DBP (LDBP) [4, 5] has been proposed. LDBP further improves the performance of DBP by globally optimizing all the parameters of DBP. Most importantly, LDBP can reduce complexity while improving performance, which has improved the nonlinear compensation algorithms to a new level. However, there is still little research regarding the application of LDBP in HSR systems.

In this paper, we focus on the optimization of low-complexity LDBP in HSR systems. First, we modify the base model of LDBP from conventional DBP to modified weighted DBP (M-W-DBP), which leverages and combines the features of weighted DBP (W-DBP) [6] and modified DBP (M-DBP) [7] together. Then, M-W-DBP is deep-unfolded into an NN called modified weighted LDBP (M-W-LDBP) and initialized with pre-optimized hyper-parameters. Compared with LDBP, M-W-LDBP provides 1 dB and 0.7 dB signal-to-noise ratio (SNR) improvement in 90-GBaud and 128-GBaud systems, respectively.

2. System setup and the limits of 1 StPS LDBP in HSR systems

The system setup is shown in Fig. 1 (a). One span contains a 100-km single-mode fiber with $\alpha = 0.2\ dB/km$, $\beta_2 = -21.7\ ps^2/km$ and $\gamma = 1.4\ /W/km$ as well as an EDFA with a 5-dB noise figure. A 10-span 1000-km transmission link is considered in our systems. In order to focus more on the fiber nonlinearity, no bandwidth limitation is considered in transceivers and the sampling rate of the receiver is set to twice the symbol rate. The symbol rate of DP-16QAM is set to 20 GBaud, 90 GBaud, and 128 GBaud to investigate the DBP performance with respect to the symbol rate. After coherent receiver, offline DSP is performed including chromatic dispersion compensation (CDC) or DBP, matched filtering, down sample, phase recovery, and data recovery. LDBP should be trained first according to [4] and then cascaded with data recovery. Mersenne twister random data is generated for training and testing separately according to [8] so that the NN will be immune to the data pattern.

First, we investigate the gain of 1 StPS DBP and 1 StPS LDBP compared to CDC in the mentioned three systems with different symbol rates. Note that all the parameters in DBP and hyper-parameters in LDBP have been fine-

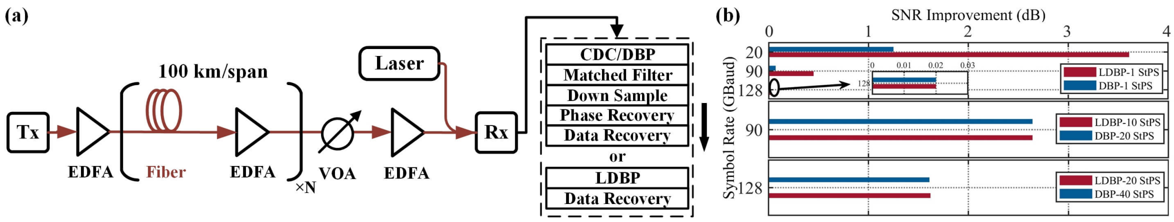

Fig. 1. (a) Block diagram of the simulation system. (b) The SNR improvement compared to CDC.

tuned. As is shown in Fig. 1 (b), at the optimal launch power, the improvement of DBP and LDBP decreases with the symbol rate increasing from 20 GBaud to 128 GBaud. For the 20-GBaud system, both 1 StPS DBP and LDBP provide considerable performance improvement compared to CDC, with 1.25 dB and 3.61 dB SNR improvement, respectively. However, for the 90-GBaud HSR system, 1 StPS DBP provides only 0.07 dB SNR improvement while 1 StPS LDBP seems also performance-limited by providing only 0.45 dB SNR improvement. Worse still, for the 128-GBaud HSR system, 1 StPS DBP provides only 0.02 dB SNR improvement while LDBP seems to have even lost most of the optimization ability and provides no obvious performance difference compared to 1 StPS DBP. The benefit of 1 StPS DBP and 1 StPS LDBP degrades rapidly with the increase of symbol rate.

Then, we re-perform DBP by selecting the L_D of the corresponding system (115.2 km for 20 GBaud, 5.7 km for 90 GBaud, and 2.8 km for 128 GBaud) as the step size, which means about 1 StPS, 20 StPS, and 40 StPS are required for 20-GBaud, 90-GBaud and 128-GBaud systems, respectively. As is shown in Fig. 1 (b), 20 StPS and 40 StPS DBP provide 2.64 dB and 1.61 dB SNR improvement for 90-GBaud and 128-GBaud systems, respectively. For LDBP, only 10 StPS and 20 StPS are required when providing 2.64 dB and 1.62 dB SNR improvement for 90-GBaud and 128-GBaud systems, respectively. These results imply that when the step size is near to L_D in HSR systems, LDBP once again demonstrates its powerful optimization capability, by providing similar SNR improvement compared with DBP but roughly saving about 50% steps. Hence, coarse step size (e.g. 1 StPS) leads to a coarse DBP model, which limits the performance of DBP and the optimization capability of LDBP.

3. Proposed M-W-LDBP with pre-optimization

To realize a low-complexity LDBP with only 1 StPS, first, we modify the base model of LDBP. For DBP with coarse step size, it is necessary to consider the correlation of neighboring symbols at the nonlinearity compensation (NLC) section of DBP, which can mitigate the model inaccuracy caused by the coarse step size and is also mentioned in [9]. Here, we adopt the weighted DBP (W-DBP) model with a Gaussian initialization (w_i in Fig.2 (a)) [6] where the correlation length is set to 21 and 45 symbols for 90-GBaud and 128-GBaud systems, respectively. Besides, we can also consider shifting the calculation position of nonlinear phase noise by introducing a shifting parameter r into DBP called modified DBP (M-DBP) [7]. Moreover, we combine both the M-DBP and W-DBP features together named M-W-DBP. LDBP with this new base model is called M-W-LDBP. All the mentioned models are depicted in Fig. 2 (a) in detail.

After changing the base model of LDBP, we perform pre-optimization for hyper-parameters. This is significant since the performance of 1 StPS M-W-LDBP extremely depends on the initial hyper-parameters because of the coarse step size. In our trials, the empirical value of γ is not proper for the NN initialization anymore. We perform coarse and fine scanning for γ and r to make 1 StPS M-W-DBP performs the best when w_i keeps the Gaussian shape. These fine-tuned values, e.g., for 128 GBaud, $\gamma = 0.17\ /W/km$ (severely deviated from the empirical value $1.4\ /W/km$ because of the coarse step size) and $r = 0.26$, are necessary for the initialization of M-W-LDBP. Additionally, some key parameters of the NN should also be fine-tuned, including learning rate (LR), batch size, epoch, and the scaling factor (SF) of the trained parameters. Note that the scaling factor is a key parameter. We find the trained parameters β_2 and γ should be in the same order of magnitude since they share the same $LR = 0.001$. In our trials, γ is multiplied with a SF equal to the step size, and β_2 is multiplied with a SF equal to 1e27. After pre-optimization, M-W-LDBP is trained according to the architecture in Fig. 2 (b). Mean squared error (MSE) and Adam are set as the loss function and optimizer, respectively. All the trainable parameters include β_2, γ, r and w_i.

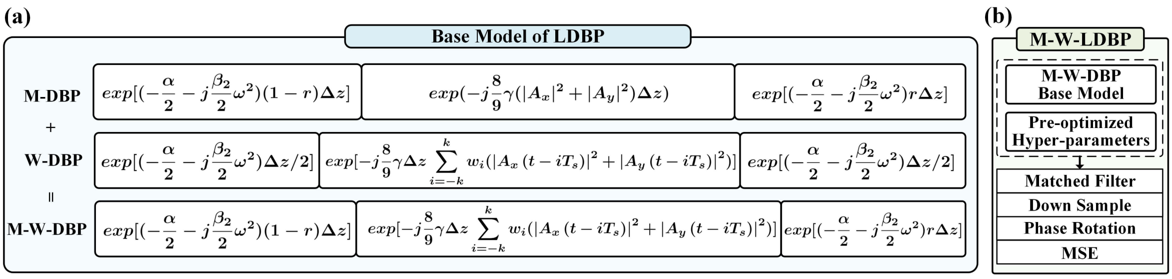

Fig. 2. (a) The base model of LDBP. (b) The architecture of M-W-LDBP.

Table 1. SNR improvement (dB) compared with CDC.

	DBP/LDBP	W-DBP/W-LDBP	M-DBP/M-LDBP	M-W-DBP/M-W-LDBP
90 GBaud	0.07/0.45	0.28/0.68	0.29/0.54	0.85/1.42
128 GBaud	0.02/0.02	0.17/0.46	0.05/0.17	0.33/0.72

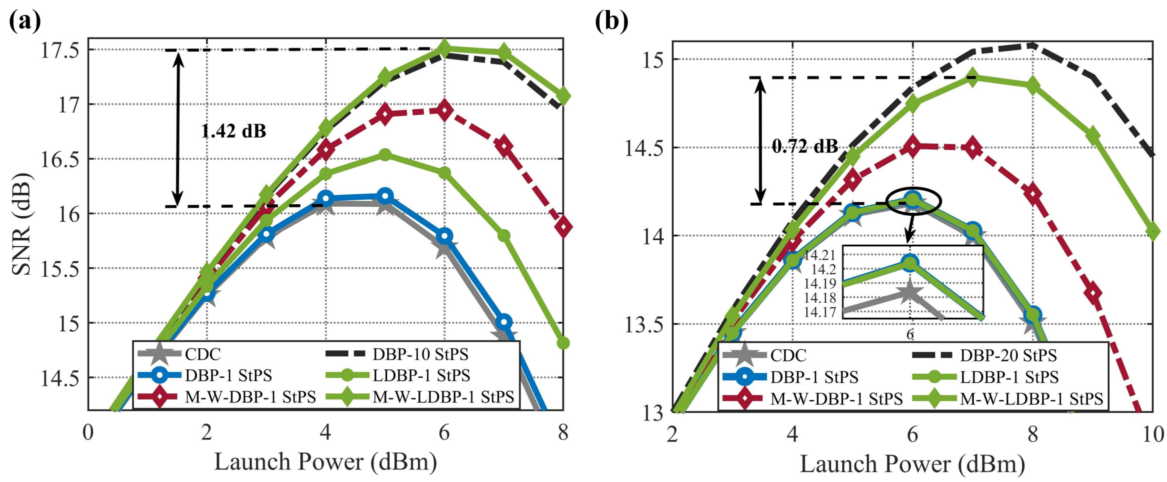

Fig. 3. The results of different algorithms in (a) the 90-GBaud system and (b) the 128-GBaud system.

4. Results and discussion

We reuse the system mentioned in section 2 and investigate the performance of the proposed M-W-LDBP with pre-optimization. First, Fig. 3 (a) and (b) depict that, 1 StPS LDBP encounters the performance limitation either in the 90-GBaud or 128-GBaud systems as mentioned in section 2. Then, as is shown in Table. 1, by considering the symbol correlation at the NLC section or shifting the position of NLC section of DBP, either 1 StPS W-DBP or M-DBP provides an improved performance than 1 StPS DBP. Then, by introducing the feature of W-DBP into LDBP and using the pre-optimized hyper-parameters, the corresponding 1 StPS weighted LDBP (W-LDBP) outperforms LDBP by providing 0.23 dB and 0.44 dB SNR improvement in 90-GBaud and 128-GBaud systems, respectively. Similarly, by introducing the feature of M-DBP into LDBP and using the pre-optimized hyper-parameters, the corresponding 1 StPS modified LDBP (M-LDBP) also outperforms LDBP by providing 0.09 dB and 0.15 dB SNR improvement in 90-GBaud and 128-GBaud systems, respectively.

When we consider M-W-DBP, the performance achieves a new level with 0.85 dB and 0.33 dB SNR improvement compared with CDC in 90-GBaud and 128-GBaud systems, respectively. As expected, 1 StPS M-W-LDBP provides the best performance, with another 0.57 dB (90 GBaud) and 0.39 dB (128 GBaud) SNR improvement compared with 1 StPS M-W-DBP as well as 1.42 dB (90 GBaud) and 0.72 dB (128 GBaud) SNR improvement compared with CDC. The performance of 1 StPS M-W-LDBP in the 90-GBaud system is similar to 10 StPS DBP and 1 StPS M-W-LDBP provides a similar performance of 20 StPS DBP in the 128-GBaud system. As a result, M-W-LDBP outperforms conventional LDBP by providing 1 dB and 0.7 dB SNR improvement in 90-GBaud and 128-GBaud systems, respectively.

5. Conclusion

For the nonlinearity compensation task in HSR systems, the realization of 1 StPS DBP has encountered a considerable challenge and the optimization capability of 1 StPS LDBP is not as powerful as shown in low-symbol-rate systems. The proposed 1 StPS M-W-LDBP with pre-optimization has outperformed LDBP by providing 1 dB and 0.7 dB SNR improvement in 90-GBaud and 128-GBaud systems, respectively. With these improvements, M-W-LDBP may be considered as a promising NLC algorithm in HSR systems.

References

1. L. Zhu et al., Electron. Lett.46, 1140–1141 (2010).
2. E. F. et al., Opt. Express 19, 570–583 (2011).
3. H. Nguyen Tan and S. T. Le, Opt. Commun. 433, 36–43 (2019).
4. C. Häger and H. D. Pfister, Proc. OFC, W3A.4 (2018).
5. Q. Fan et al., Nat. Commun. 11, 1–11 (2020).
6. D. Rafique et al., Opt. Express 19, 9453–9460 (2011).
7. C.-Y. Lin et al., Proc. ECOC, P3.16 (2010).
8. T. Liao et al., Opt. Lett. 45, 5113–5116 (2020).
9. C. Häger and H. D. Pfister, IEEE J. on Sel. Areas Commun. 39, 280–294 (2021).