Digital Pre-Distortion Coefficients Identification Using Gauss-Newton Based Direct Learning Architecture

Hexun Jiang[1], Mengfan Fu[1], Yixiao Zhu[1], Lilin Yi[1], Weisheng Hu[1, 2], and Qunbi Zhuge[1, 2*]
1. State Key Laboratory of Advanced Optical Communication Systems and Networks, Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai, 200240, China
2. Peng Cheng Laboratory, Shenzhen, 518000, China
*Corresponding author: qunbi.zhuge@sjtu.edu.cn

Abstract: We propose to identify the coefficients of a digital pre-distortion equalizer based on direct learning architecture (DLA) using the Gauss-Newton method. Experimental results show that DLA outperforms indirect learning architecture by 0.5dB.

1. Introduction

Due to the development of internet services, the Ethernet interface in datacenter optical interconnects is evolving from 400G to 800G and 1.6T, where coherent system with spectrally-efficient modulation formats will be a potential solution [1]. For coherent short-reach transmissions, transmitter nonlinear impairments limit system performance. To overcome this problem, digital pre-distortion (DPD) [2-4] has been applied to mitigate the transmitter nonlinear impairments. Nonlinear equalizer such as Volterra filter equalizer (VFE) is one of the most popular DPD techniques.

For the effectiveness of DPD, the key problem is to identify the coefficients of a DPD model, which is also called as learning architecture [5]. The two most popular learning architectures are indirect learning architecture (ILA) [2] and direct learning architecture (DLA) [3-4]. The DPDs based on ILA are simpler and investigated more commonly. Nonetheless, there exist two shortcomings in ILA: 1) The post-inverse equalizer is first trained and then used as a pre-inverse equalizer, which is not exactly mathematically equivalent due to the lack of commutative law; 2) the calculation of ILA is based on noisy signals, which will yield biased estimates. On the contrary, DLA can circumvent the two problems. In optical communication, DPDs based on DLA have recently been investigated, which achieve a better performance than ILA [3,4]. Nonetheless, the reported approaches all involve training an auxiliary neural network (NN), which makes the learning procedure more complicated. Moreover, since the structure and hyper parameters of NN may differ for different transmitters, an NN should be designed and trained for every transmitter, which is time-consuming and difficult in practice.

In this paper, we propose to identify the coefficients of digital pre-distortion using a Gauss-Newton (GN) based DLA method for optical communications. To the best of our knowledge, it is the first time to realize DLA without training an auxiliary NN in optical communications. An experiment is conducted, which demonstrates that Volterra filter equalizer (VFE) based DPDs using the GN-based DLA can mitigate the nonlinear distortion effectively and shows a 0.5 dB signal-to-noise ratio (SNR) improvement over ILA.

2. Principle

The third-order VFE is selected as the DPD model in this work, which can be written as

$$z(n)=\sum_{k_1=-L_1}^{L_1} h\left(k_1\right)x\left(n-k_1\right)+\sum_{k_1=-L_2}^{L_2}\sum_{k_2=k_1}^{L_2} h\left(k_1,k_2\right)\prod_{i=1}^{2}x(n-k_i)+\sum_{k_1=-L_3}^{L_3}\sum_{k_2=k_1}^{L_3}\sum_{k_3=k_2}^{L_3} h\left(k_1,k_2,k_3\right)\prod_{i=1}^{3}x(n-k_i) \tag{1}$$

where z, x and h denote the output signals, input signals, and the coefficients of the Volterra kernel, respectively. $\left(L_1,L_2,L_3\right)$ denotes the third-order single-side memories. $\boldsymbol{x}$, $\boldsymbol{y}$, and $\boldsymbol{z}$ denote column vectors of the original signals, the received signals, and the signals after pre-distortion, respectively. Eq. (1) can be transformed into a matrix form

$$\boldsymbol{z}=\boldsymbol{X}\boldsymbol{h} \tag{2}$$

where the matrix $\boldsymbol{X}$ includes all the product terms in Eq. (1):

$$\boldsymbol{X}=\left[\boldsymbol{x}\left(n+L_1\right),\ \boldsymbol{x}\left(n+L_1-1\right),\ldots,\ \boldsymbol{x}^3\left(n\text{-}L_3\right)\right] \tag{3}$$

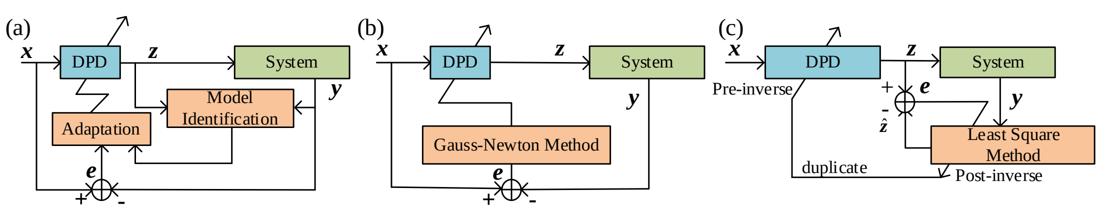

Fig. 1. Schematic diagram of (a) DLA with model identification, (b) GN-based DLA and (c) ILA.

The target of DLA is identifying the coefficients $\boldsymbol{h}$. Normally, for DLA, a model of the system should be identified firstly, and then the coefficients are calculated based on the extracted model and error signals as shown in Fig.1(a). Fig. 1(b) depicts that GN-based DLA, in which Euclidean norm $J(\boldsymbol{h})=\|\boldsymbol{x}-\boldsymbol{y}\|_2$ is minimized, can be converted into the following equation:

$$J'(\boldsymbol{h})=0 \tag{4}$$

We use the GN method to obtain the iterative solution of $\boldsymbol{h}$:

$$\boldsymbol{h}_{k+1}=\boldsymbol{h}_k-J'(\boldsymbol{h}_k)/J''(\boldsymbol{h}_k),\quad k=0,1,2... \tag{5}$$

where $\boldsymbol{h}_k$ denotes the coefficients of the *k-th* iteration. The specific derivation of $J'(\boldsymbol{h})$ and $J''(\boldsymbol{h})$ can be found in [5]. Finally, the analytical formula of the GN-based DLA is

$$\boldsymbol{h}_{k+1}=\boldsymbol{h}_k-\mu\left(\boldsymbol{X}^H\boldsymbol{X}\right)^{-1}\boldsymbol{X}^H(\boldsymbol{y}-\boldsymbol{x}) \tag{6}$$

where μ denotes the learning rate.

As a comparison, we also show the structure of ILA in Fig. 1(c). The post-inverse model is first identified, where the errors $\boldsymbol{e}$ between the output signals of post-inverse equalizer $\hat{\boldsymbol{z}}$ and the input signals of the system $\boldsymbol{z}$ are minimized. This process can be obtained by the least square (LS) method

$$\boldsymbol{h}_{k+1}=\boldsymbol{h}_k-\mu\left(\boldsymbol{Y}^H\boldsymbol{Y}\right)^{-1}\boldsymbol{Y}^H\boldsymbol{z} \tag{7}$$

where the matrix $\boldsymbol{Y}$ is similar to $\boldsymbol{X}$, which can be expressed as

$$\boldsymbol{Y}=\left[\boldsymbol{y}\left(n+L_1\right),\boldsymbol{y}\left(n+L_1-1\right),...,\boldsymbol{y}^3\left(n\text{-}L_3\right)\right] \tag{8}$$

Then, the post-inverse model is duplicated to the pre-inverse model. ILA has two fundamental drawbacks: 1) The matrix $\boldsymbol{Y}$ is built from the received noisy signal $\boldsymbol{y}$, which results in biased estimates and degrades the DPD performance; 2) ILA is based on the open-loop structure and assumes that the post-inverse of a nonlinear system can be duplicated to the pre-inverse of a nonlinear system, which is not strictly satisfied in mathematics. On the contrary, DLA is based on the close-loop and directly calculates the pre-inverse model. Besides, the matrix $\boldsymbol{X}$ is built from the original signals $\boldsymbol{x}$ according to Eq. (6), which is not affected by the noise. From the above analysis, the performance of DLA is considered to be superior to ILA.

3. Experimental setup and results

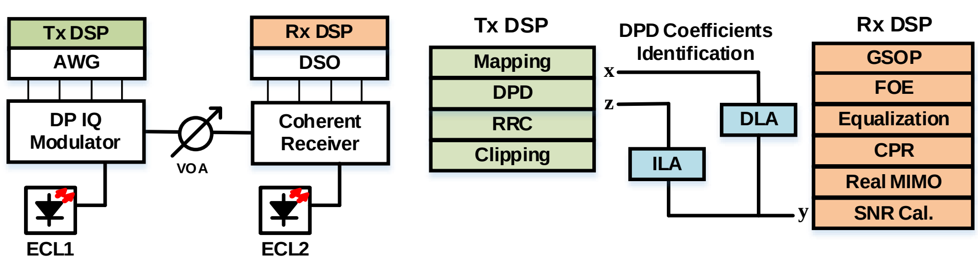

Fig. 2 Experimental setup and DSP flow.

The experimental setup and DSP flow are shown in Fig. 2. In the Tx DSP, 40Gbaud 64QAM symbols are generated and pre-distorted with the trained VFE based DPD with tap parameters (5,3,2). In particular, 4 independent VFEs are implemented on the in-phase or quadrature (IQ) parts of symbols for two polarizations XI, XQ, YI, YQ, respectively. Then, the distorted symbols are up-sampled to 2 samples per symbol (SPSs), shaped by a root raised cosine (RRC) filter and clipped with a clipping ratio of 10dB. The generated sample sequence is resampled from 2SPSs to 3SPSs to match the 120GSa/s sampling rate of an arbitrary waveform generator (AWG, Keysight M8194) with 45GHz bandwidth. An external cavity laser (ECL) is used as the optical source, with 12.7dBm launch power and linewidth less than 50kHz. Then, the electrical signals are fed into a 40GHz dual-polarization IQ (DP-IQ) transmitter (Neophotonics HB-CDM class 40), which includes four drivers and a DP-IQ modulator. In such an amplifier-less short-reach system, the peak-to-peak voltage of AWG is fixed at 800 mV to boost the optical power, where nonlinear distortions caused by the drivers and the modulator are severe and the comparison between ILA and DLA is more notable. The optical signal is attenuated by a fixed variable optical attenuator (VOA) and then received by a coherent receiver (Neophotonics ICR class 40) using another similar ECL with 10.7dBm power. The electrical waveforms are sampled and stored by a real-time digital storage oscilloscope (DSO, Tektronix DPO75902SX) with a sampling rate of 100GSa/s. In the Rx DSP, IQ imbalance in the Rx is first compensated. After 4th-power frequency offset estimation (FOE), the equalization is performed, and then the phase-locked loop (PLL) is used to recover the laser phase noise. Finally, a 4×4 real-value MIMO equalizer is used to compensate for the Tx IQ crosstalk. The IQ parts of the transmitted symbols **x**, pre-distorted symbols **z** and the output symbols of real-value equalizers **y** are used to identify the coefficients of the 4 independent VFEs based on DLA or ILA.

Fig. 3 depicts the SNR performance versus the iteration for three different received optical powers (ROPs). The first iteration represents the initial stage without DPD. Both DLA and ILA can mitigate the nonlinear distortions effectively. The performance improves with the increase of iteration and saturates at the fourth iteration both for DLA and ILA. At three different ROPs, the performances of DLA are all superior to ILA, which demonstrates that coefficients can be identified more accurately based on DLA. Moreover, when the ROP is reduced from -14dBm to -18dBm, the SNR gap between the DLA and ILA increases from 0.31dB to 0.50dB. This is because ILA is more sensitive to noise than DLA.

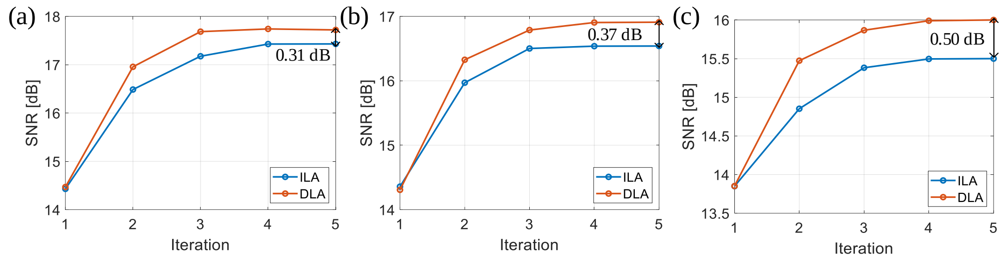

Fig. 3 SNR performance versus iteration at the ROP of (a) -14dBm, (b) -16dBm and (c) -18dBm.

In Fig. 4(a), the SNR versus ROP is shown with 4 various DPD schemes. The coefficients of the DLA or ILA are trained at -13dBm or -18dBm, respectively. In general, the SNR of DLA is higher than ILA for all the ROP values. Moreover, the performance of DLA trained at different ROP is almost the same, which indicates that the performance of DLA is insensitive to noise. On the contrary, the performance of ILA trained at -18dBm is worse than -13dBm since ILA is easily influenced by the noise. Finally, when the ROP is -13dBm, constellations of three different conditions are shown in Fig. 4(b-d), which indicates that the nonlinear impairments can be mitigated effectively by the DPD based on DLA or ILA. It is confirmed that the performance of DLA is better since the outer constellations of DLA are closer to the standard constellations than ILA.

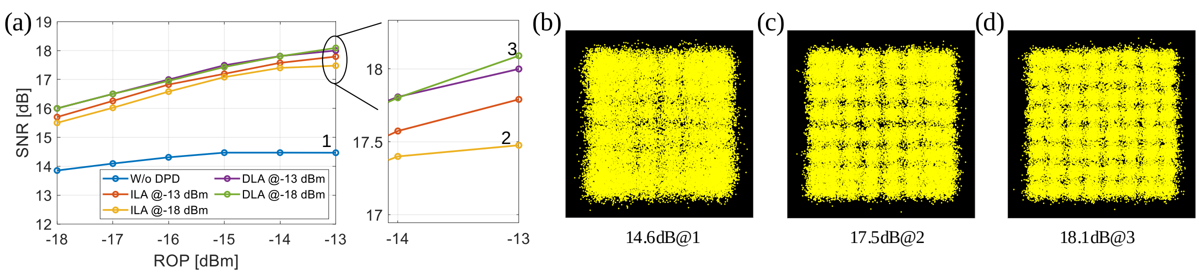

Fig. 4 (a) SNR performance versus ROP with different DPD coefficients. Constellations of received signals at -13dBm ROP (b) without DPD, with DPD based on (c) ILA @ -18dBm ROP, (d) DLA @ -18dBm ROP.

It is worth mentioning that SNR is lower in the systems with longer transmission distance, so the performance of DPD based on ILA will degrade further. Under this condition, DLA will outperform ILA with larger gains than the short-reach system.

4. Conclusion

We propose to identify the coefficients of DPD using GN-based DLA. To the best of our knowledge, it is the first time to train DPD using DLA without training an auxiliary NN in optical communications. We have experimentally verified that the performance of DLA outperforms ILA by 0.5dB.

This work was supported by National Key R&D Program of China (2022YFB2903500) and NSFC (62175145).

5. Reference

[1] X. Zhou et al., *J. Light. Technol*., vol. 38, no. 2, pp. 475-484 (2020).
[2] P. W. Berenguer et al., *J. Light. Technol.*, vol. 34, no. 8, pp. 1739-1745 (2015).
[3] V. Bajaj et al., *J. Light. Technol.*, vol. 40, no. 3, pp. 597-606 (2022).
[4] G. Paryanti, et al., *J. Light. Technol.*, vol. 38, no. 15, pp. 3883-3896 (2020).
[5] Z. Yu et al., *Proc. Asia-Pacific Microwave Conference* (2015).