Compressed Nonlinear Equalizers for Optical Interconnects: Efficiency and Stability

Ling Ge,Wenjia Zhang*, Yanci Zhang, Chenyu Liang, Jiangbing Du and Zuyuan He

State Key Laboratory of Advanced Optical Communication Systems and Networks
Shanghai Jiao Tong University, Shanghai 200240, China

**wenjia.zhang@sjtu.edu.cn*

Abstract: Efficiency and stability of pruned Volterra-Series and Neural-Network Equalizers are compared in the 112-Gbps optical interconnects. The results show NNE outperforms VE at equalization performance and complexity while VE is more stable with channel variation.

1. Introduction

Recent explosion of bandwidth-hungry applications like high definition video streaming and sharing in a social network has led to a dramatic increase of IP traffic in data centers. Bandwidth requirement for a single lane has reached 100-Gbps, which has been a basic building block for a 400 or 800-Gbps transceiver. For 100-Gbps transmission, optical and electrical components show grave nonlinear impairments especially for multilevel intensity modulation and direct detection (IM/DD). Nonlinear equalization, though having not yet been a standard component in a commercial optical transceiver, have showed great potentials to become an inevitable element to deal with complicated design and mass production issues. In order to realize the translation of nonlinear equalization from theory to practice, one important question should be addressed: how to design an advanced algorithm with great and stable equalization performance and low computational complexity.

In this paper, a comparison study of pruned Volterra-series based equalization (VE) and neural network based equalization (NNE) is presented for 112-Gbps vertical cavity surface emitting lasers (VCSELs) based optical interconnects, which has become a competitive candidate due to ultra-low cost and power consumption. However, with increasing bandwidth requirement, this solution suffers from bandwidth limitation and complex nonlinear effect of optical fiber channel and various optoelectronic devices, such as mode dispersion and relative intensity noise (RIN) [1]. VE [2] and NNE [3] are proposed to provide powerful nonlinear compensation capability. In order to overcome the high computational complexity of those complex nonlinear equalizers, coefficients of Volterra series out of the diagonal have been intentionally discarded to construct a memory polynomial Volterra equalizer in [4]. But such method can't fit various transmission scenarios and transmission performance may be out of reach due to that the process of simplification runs before the training. Moreover, in [2,5], the authors adopt weight-pruning to reach a sparse equalizer structure for VE or NNE. However, as equalization architectures for Volterra-series and neural network are different in the nonlinear construction and complexity increase, there lacks comparison study of the effectiveness and efficiency of VE and NNE and the corresponding pruning algorithms. More importantly, it is still unclear that whether pruned structures can adapt to changing channel characteristics.

In this work, a three-layer NNE and a three-order VE are both employed to deal with the 112-Gbps PAM-4 signals. The threshold based pruning algorithm is used to compare the pruning efficiency of the two nonlinear equalizers. In addition, we analyze the signaling performance under different bias voltages, and measure bit error rate by using pruned VE and pruned NNE respectively. Our experimental results show that NNE outperforms VE at equalization and computational complexity while pruned-VE has better stability than pruned-NNE.

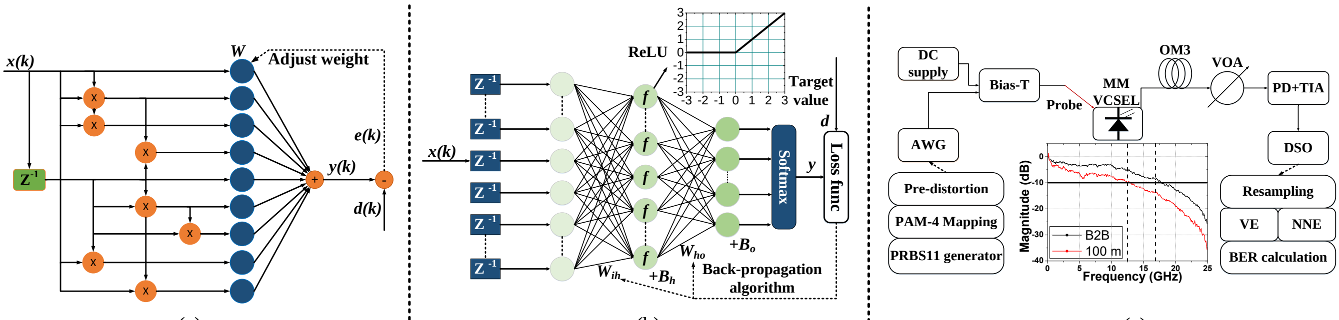
2. Principle

2.1. Volterra series based equalizer and neural network based equalizer

Volterra series is considered to describe nonlinear dynamic systems, which makes it a focus in the field of nonlinear equalization. The classic VE with P-order and memory length M_r can be expressed as Eq. (1),

$$y(k) = W_{dc} + \sum_{r=0}^{P} \sum_{k_1=0}^{M_r-1} \cdots \sum_{k_r=k_{r-1}}^{M_r-1} W_r(k_1,k_2,...,k_r) \cdot x(k-k_1) \cdots x(k-k_r). \tag{1}$$

where $x(k)$ is the k^{th} sample of the received signal and $y(k)$ is the output after equalization, W_r is the r^{th}-order Volterra kernel. The coefficient W_{dc} is responsible for the dc component, which is not included in the final model in the AC-coupled circuits. With the increasing of P and memory length M_r, the number of coefficients will suffer from rapid growth. Especially the higher Volterra order will lead to more number of coefficients. According to [6] and our tests, the three-order VE seems to be sufficient for an IM/DD system to compensate the linear and nonlinear impairments. Therefore, in this work, a three-order VE (P = 3) is applied. The basic $VE(2,2,2)$ structure with 9 coefficients is illustrated in Fig. 1(a) in order to facilitate the understanding of VE process. Note that we use



(a) (b) (c)

Fig. 1: The structure of (a) three-order VE and (b) three-layer NNE and (c) the experimental setup.

$VE(M_1, M_2, M_3)$ here to represent a three-order VE, where M_1, M_2, M_3 represent the memory length of the first, second and third order of VE.

Neural networks have shown much better performance than traditional algorithms in fields like image classification and natural language processing. Recently, NNEs have been widely studied and applied in the field of communications as nonlinear equalizer due to their powerful nonlinear representation capability. The mathematical formula of the conventional NNE process can be expressed as Eq. (2),

$$y = argmax[softmax(ReLU(x^T(k) \times W_{ih} + B_h) \times W_{ho} + B_o)]. \tag{2}$$

where *x(k)* is the resampled signal sequence, W_{ih} and W_{ho} are the weight matrices of input layer to hidden layer and hidden layer to output layer respectively, B_h and B_o are the bias vectors of hidden layer and output layer, *ReLU* (Rectified Linear Unit), as shown in the inset of Fig. 1(b), is chosen as the activation function of hidden layer. Function $softmax(\cdot)$ converts the results of output layer to probability distribution, which represents the probabilities of each class. Finally, $argmax[\cdot]$ is equivalent to decision function, which returns an index of the maximum value of the output probability vector which indicates the specific class of the equalization results. The structure of the 3-layer NNE is shown in Fig. 1(b). Same as VE, we use $NNE(N_1, N_2, N_3)$ to represent a 3-layer NNE, where N_1, N_2, N_3 are the number of neurons in input layer, hidden layer and output layer.

2.2. *Pruning algorithm*

It has been proved that most of the computations in the nonlinear equalizers are redundant [2], which means that there is possibility to realize the lightest equalizer without sacrificing the transmission performance. Pruning algorithm has shown significant ability to reduce the computational complexity while maintaining equalization performance. The coefficients of the nonlinear equalizer after initial training are discarded through a threshold and then recovered the performance damage caused by pruning through a retraining phase. The pruning

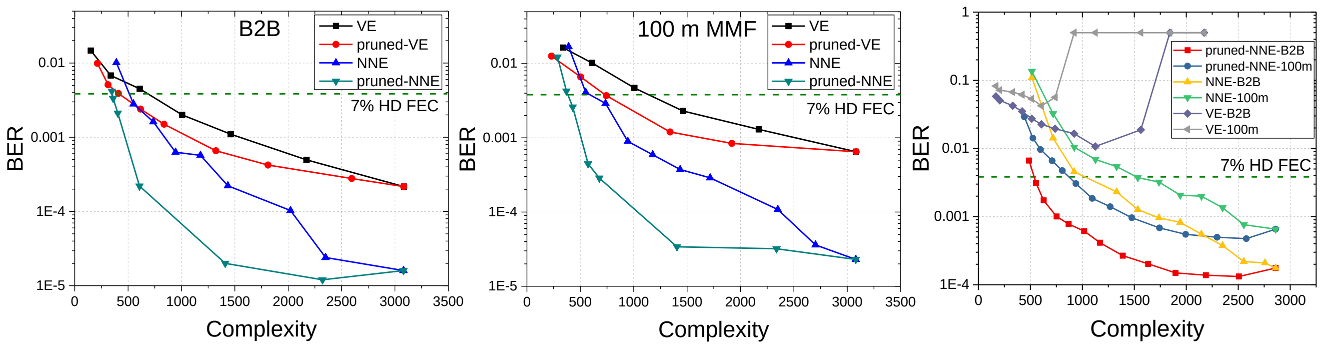
$$S(W) = 0, \quad when \quad S(\cdot) < T. \tag{3}$$

process can be easily expressed as Eq. (3), where $S(\cdot)$ represents the weight set and T is the threshold. Part of connections are pruned when their absolute weights are below the threshold. For VE, we just prune the second and third order coefficients because the linear term play an important role for equalization performance while occupies only a small fraction of computational complexity. For NNE, we cut off all the coefficients from each layer once they are below the threshold. In this work, we define the computational complexity as the number of multiplication operation [2] as calculated in Eq. (4).

$$VE: M_1 + M_2 \cdot (M_2+1) + M_3 \cdot (M_3+1) \cdot (M_3+2)/2. \qquad NNE: N_1 \cdot N_2 + N_2 \cdot N_3. \tag{4}$$

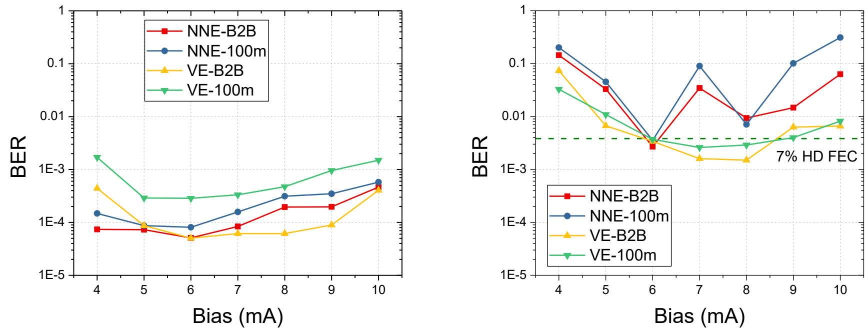
3. Experiment setup and results

Fig. 1(c) shows the experimental setup. In the Tx side, PAM-4 electrical signals with a period of $2^{11}-1$ are generated by an arbitrary waveform generator (AWG, Keysight M8195A) with 64-GS/s sampling rate. Both of the



(a) (b) (c)

Fig. 2: The efficiency comparison for (a) B2B case and (b) 100-m MMF case at 100-Gbps PAM-4 with pre-distortion. (c) BER as a function of complexity at 112-Gbps PAM-4 without pre-distortion.



(a) (b)

Fig. 3: (a) The BER as a function of VCSEL bias, (b) Pruned equalizers for different bias cases.

DC bias and high speed PAM-4 signals are applied to the 850-nm multi-mode VCSEL through a probe followed by OM3 MMFs and a multi-mode variable optical attenuator (VOA) to adjust the received optical power (ROP). In the Rx side, the signals are detected by a photo-detector (PD, New Focus 1484-A-50). The optical signals are converted into electrical signals by the PD and then captured by a real time digital storage oscilloscope (DSO, Keysight DSOZ592A) with 160-GS/s sampling rate for offline processing. The captured discrete signals are firstly resampled to one sample per symbol, followed by nonlinear processing and bit error rate (BER) calculation.

We capture 100-Gbps signals at 1-dBm ROP to compare the efficiency of VE and NNE, and the results are shown in Fig. 2(a, b). The initial equalizer configurations are set to VE(51, 31, 15) (3083 complexity) and NNE(51, 56, 4) (3080 complexity) respectively to balance the initial complexity of VE and NNE for fair comparison. As we can see, NNE performs more than one order of magnitude BER advantage over VE for both B2B and 100-m MMF cases under the same computational complexity, which means NNE has a higher performance limit than VE. In addition, the introduction of pruning algorithms leads to lower computational complexity and BER degradation. For example, the computational complexities that meet 7% HD FEC limit are 1144, 743, 603 and 386 for general VE, pruned-VE, general NNE and pruned-NNE respectively, in which pruned-NNE shows 48% complexity performance improvement than pruned-VE. To further investigate the efficiency difference between VE and NNE, we increase the PAM4 signal rate to 112-Gbps, skip the pre-distortion phase, and reduce the ROP to 0-dBm to create a worse transmission condition. The results are illustrated in Fig. 2(c). In this case, VE shows serious numerical instability due to its intricate structure. With the increasing of VE memory length, BER performance shows a slow changing and even starts to increase when the complexity increases to a certain degree till misconvergence ($BER \approx 0.5$). While NNE still achieves a remarkable equalization performance and far outperforms VE. In addition, the pruned-NNE reaches around 50% computational complexity reduction then conventional NNE at the FEC limit of 3.8e-3. However, in low complexity condition, VE performs better than NNE because VE can always maintain at least the linear equalizer, even after pruning. While the performance of NNE will decline rapidly after large-scale pruning, as NNE doesn't have the property of separate order like VE.

In order to verify the stability of pruning algorithm with changing link conditions, DC bias of VCSEL is tuned to emulate the change of output power and channel bandwidth, and measure BER under different bias as shown in fig. 3. In this part, we used 100-Gbps pre-distorted PAM-4 signals for transmission experiments in order to make the performance of VE and NNE comparable. VE(51, 23, 11) (1461 complexity) and NNE(31, 41, 4) (1435 complexity) are set respectively. Bias level of 6-mA, which is obvious the optimal operating point of VCSEL, is chosen as the benchmark point. We perform pruned-VE and pruned-NNE training on data with 6-mA bias and then equalize data of other biases with the pruned equalizers. The pruned-VE optimized for 6-mA bias also shows excellent tolerance for bias of 7-mA and 8-mA even near 9-mA, which means that the sparse VE structure can be adaptive within a certain range (+ 2.5-mA in this work). However, pruned-NNE shows poor stability for different bias cases, as the pruned-NNE optimized by data of 6-mA bias seems less effective for cases of other biases.

4. Conclusion

We conduct a comparison study of 100-Gbps and 112-Gbps optical interconnect with pruned nonlinear equalizers: VE and NNE. The results show more than one order of magnitude BER improvement for NNE than VE and around 50% complexity superiority for pruned-NNE than pruned-VE to reach the same FEC limit. While the pruned-VE has better stability (+ 2.5-mA) then pruned-NNE to adapt the variation of physical channel.

5. Acknowledgment

This work is supported by the National Natural Science Foundation of China under Grant 61605111.

6. Reference

[1] S.E. Ralph, et al, SPIE (2017) paper 101220E.
[2] L. Ge, et al, J. Lightw. Technol. 37(13), 0733-8724 (2019).
[3] Q. Zhou, et al, Opt. Commun. 441, 121–126 (2019).
[4] Diamantopoulos, et al, ECOC (2017) paper SC5.5.
[5] Z. Wan, et al, Opt. express. 26(8), 10631–10642 (2018).
[6] Stojanovic, et al, J. Lightw. Technol. 35(21), 4583–4594 (2017).