On the Computational Complexity of Artificial Neural Networks for Short-Reach Optical Communication

Zhaopeng Xu
Peng Cheng Laboratory
Shenzhen, China
xuzhp@pcl.ac.cn

William Shieh
School of Engineering
Westlake University
Hangzhou, China
shiehw@westlake.edu.cn

Abstract—Artificial neural networks (ANNs) have been widely used for nonlinear equalization in short-reach optical communications due to the superior performance compared with traditional approaches. However, the computational complexity (CC) is a major challenge that hinder their applications. In this paper, we compare various types of ANNs for short-reach optical communication with CC constraint, as well as provide several means of CC reduction by employing transfer learning, pruning and multi-task networks.

Keywords—artificial neural network, nonlinear equalization, computational complexity

I. INTRODUCTION

Short-reach optical communication dominates data center supporting high-speed accompanied by the ever-increasing demand for data traffic [1]. Despite several self-coherent proposals [2-4], traditional intensity-modulated direct-detection (IMDD) links are mostcommonly used in short-reach applications due to the intrinsic simple structure and cost-effectiveness [5-7]. However, IMDD systems suffer from nonlinear impairments owing to the mixture of chromatic dispersion (CD) and the intensityonly square-law detection. Moreover, to meet the low-cost target, band-limited transceivers and cost-effective lasers such as directly modulated lasers (DMLs) are preferred for usage which possess non-ideal frequency response, the chirp effects, and the device nonlinear impairments [8], [9]. The overall mixed linear and nonlinear impairments could strongly degrade system performance and limit the achievable capacity. As such, efficient equalization techniques are of vital importance to guarantee a desired system performance.

With the fast development of machine learning technologies, various neural network (NN)-based equalizers have been proposed as the underlying digital signal processing (DSP) tools to effectively deal with the IMDD system impairments, including but not limited to feedforward NN (FNN) [10-22], radial basis function NN (RBFNN) [23], recurrent NN (RNN) [24-30], convolutional NN (CNN) [31-33], spiking NN (SNN) [34], [35], etc. The different types of NN-based equalizers have attracted a lot of attention since they usually outperform traditional digital signal processing (DSP) approaches such as feedforward equalization (FFE), decision feedback equalization (DFE) and the Volterra seriesbased equalization in system bit-error-rate (BER) performance, which enables higher data-rate signal transmission. However, the computational complexity (CC) of NNs is an important issue needed to be considered, where some of the NNs (E.g., CNNs, and part of the RNNs) are usually regarded too complex to be applied in real-time scenarios.

In this paper, we provide an overview on the CC issue of NN-based nonlinear equalization in short-reach communications, and provide several approaches to relax the CC requirements. The CC of NN-based equalization lies in both the NN training and the equalization (inference) processes. The following Sections will discuss the CC from the two aspects.

II. TRANING COMPLEXITY

The training process of NN-based equalizers can be roughly described as several rounds of forward-propagation (FP) and back-propagation (BP) steps. The FP step is performed to calculate the loss function, while at the same time determine whether the training should come to a stop according to the calculated results. The BP step is performed to calculate the partial derivatives of the weights and biases, allowing all the NN parameters to be fine-tuned by efficient weight updating algorithms. The FP and BP steps need to be repeated many times (iterations) in order to achieve a welltrained NN, which is rather computationally-intensive. In addition, the training process normally requires large amount of known data to prevent the underfitting problem and to dive further into system performance. As such, how to lower the number of iterations and training data is a key factor in reducing the training CC.

To speed up the training process, transfer learning approach [36-40] is introduced into short-reach optical communications. Apart from short-reach applications, it is also applied in optical performance monitoring [41-43] or nonlinear mitigation in medium to long-haul optical singlesideband systems [44] and coherent systems [45], [46]. Transfer learning can make use of a pre-trained NN model (from source system) and "transfer" it for fast training the NN model into a new scenario (target system). Instead of training a NN from scratch, the prior knowledge preserved by the source NN model can serve as a good starting point for NN training in a related target system. For short-reach optical communications, since NNs have the capability of equalizing the received distorted signals, they "learned" the mixed linear and nonlinear channel effects which influence the transmitted symbols. The information lies in the parameters of the NN models, to be specific, the weights and biases. As such, the NNs trained from source systems have already preserved a lot of channel information such as CD, frequency chirp, and band-limited effects, where the target system could benefit from by performing transfer learning. Transfer learning is well suitable when there is multiple links ready for NN training,

and can drastically reduce the required number of training symbols and iterations (epochs). In [36] and [37], transfer learning-aided FNN and RNNs are proposed for nonlinear equalization in a 50-Gb/s 20-km pulse amplitude modulation (PAM)-4 link. s. A reduction of 90%/87.5% in epochs and 62.5%/53.8% in training symbols is achieved with FNNs/RNNs transferred from the most similar source system (60-Gb/s 15-km link). The transfer between different types of NNs is also investigated. FNNs can be transferred smoothly to their corresponding RNNs for equalization in the target system, while TL from RNNs to FNNs cannot work properly. Similar source and target systems can be found in [38], where CNN-based transfer learning is developed. The same conclusions also hold for CNN-based links. In [39] and [40], iterative pruning technique is proposed to further accelerate the transfer learning procedure of neural network equalizers. By adjusting the pruning threshold and span, a compromising performance can be achieved considering both performance stability and complexity. Transfer learning can be of great importance for future optical switched data center networks, where the parameters of the optical links need to be dynamically reconfigured. New optical interconnects can be established in an expeditious manner with the help of transferred NNs.

III. EQUALIZATION COMPLEXITY

Although the NN training process is complex, the good news is the optical links are quite stable. The statistics of the transmitted data and the channel environment change little over time. As such, for one certain optical links, the equalization complexity is what we care about most, since the well-trained NNs can be used as fixed equalizers without the need of retraining. The equalization CC of different types of NNs are thoroughly investigated and compared in [47] and [48], where auto-regressive RNN-based equalizers are recommended considering both BER performance and CC constraint. Moreover, for small-scale machine learning tasks such as nonlinear equalization in short-reach optical links, it is found that it is possible to control the number of multiplications of the NN-based receivers to the order of a few tens. This suggests NN-based equalizers have the potential to be applied in real time receivers such as field-programmable gate array (FPGAs) or specific integrated circuits (ASICs). Note that only the FNN and the FNN variants (such as autoregressive RNN and cascade FNN/RNN) are regarded as the DSP-lite NNs. Other types of RNNs such as long-short term memory (LSTM) and gate recurrent unit (GRU), as well as CNNs are deemed too complex for usage in short-reach optical communication systems.

To support higher data rate and achieve better system performance, NNs with additional structure and larger size are required. The consequent increase of equalization CC would lead to higher latency and larger power consumption of the receiver. Therefore, it is highly desirable to reduce the CC, and several means are proposed to address this issue. Inspired by multi-task learning, multi-symbol prediction is proposed for different types of NNs, which enables effective weight sharing and lower the number of multiplications required per received symbol [49-52]. Traditional NN-based equalization typically recovers the received symbols sequentially, which only takes one received symbol into consideration each time. However, this kind of architecture is not computationally efficient, since much information provided by the weights and biases trained to predict the current symbol may still be useful to predict the

following symbols. In other words, when performing equalization over different symbols, part of the weights and biases can be shared. Based on this critical fact, multi-symbol equalization scheme is proposed, where multiple symbols can be equalized by one NN-based equalizer at the same time. With the help of the multi-output architecture, a better usage of the weights and biases of NNs could be made. Although the entire network size for the new scheme needs to be larger due to the increased NN outputs, the averaged CC needed to recover the received symbol could be reduced, since the weights and biases are shared to simultaneously predict multiple symbols. Multi-symbol equalization schemes have been validated based on the FNN/cascade FNN/autoregressive RNN [49], [50] and the LSTM [51], [52] architectures, achieving more than 40% CC reduction ratio compared with traditional single-output counterparts. Besides reduction, another advantage of multi-symbol equalization scheme is parallel computing [53], [54]. Increasing the number of NN outputs could enable higher throughput at the same FPGA clock frequency. As such, less FPGAs are needed under a given symbol rate, facilitating the real-time FPGA implementation of NN-based equalizers.

Another effective complexity reduction approach for NNbased equalizers is weight pruning. Before the popularization of NN-based receivers, pruning techniques have been widely employed in Volterra series-based equalizers [55], [56] for nonlinear impairments mitigation in short-reach direct detection optical systems. Whether which types of equalizers are used, it has been demonstrated that a proportion of connections inherent in those equalizers are not necessary, and as such pruning techniques can be applied to cut off the insignificant weights and reduce the CC required for equalization. For NN-based nonlinear equalization, the pruning procedure is most commonly operated by cutting off the connections where the weight values are lower than a predefined threshold. The threshold can be carefully determined so that the BER performance of NNs can be maintained. When a NN-based equalizer is well-trained, the weights with small values are thought insignificant and negligible, and as such they can be removed to further reduce the CC. The NN can then be retrained based on the pruned structure to fine-tune the survived weights. Alternatively, the NN can be retrained based on the fully-connected structure iteratively where at the beginning of each iteration, the weight values determined to be pruned are set as zero. Either approach produces a sparsely connected NN which is computationally efficient. Pruning has been shown effective in reducing CC for different types of short-reach optical communications in [57-62]. Even though the minimum NN structure is used for equalization (further decrease the number of neurons will degrade BER), weight pruning is still applicable.

IV. CONCLUSIONS

This paper reviews different types of NN-based equalizers proposed for short-reach IMDD optical communication systems. Considering real-time implementation of NN-based receivers, the paper focuses on the CC of various NN-based equalizers, and discuss several CC reduction approaches. The methods include transfer learning for NN training acceleration, as well as multi-symbol prediction and pruning for computationally-efficient NN-based equalization. The proposed methods are well-suited for different NN structures, which speed up the DSP process for NN-based short-reach optical communications.

REFERENCES

- [1] A. Ghiasi, "Large data centers interconnect bottlenecks," *Opt. Express*, vol. 23, no. 3, pp. 2085-2090, Jan. 2015.
- [2] C. Sun et al., "Experimental demonstration of complex-valued DSB Signal field recovery via direct detection," *IEEE Photon. Technol. Lett.*, vol. 32, no. 10, pp. 585-588, May 2020.
- [3] T. Ji *et al.*, "Theoretical and experimental investigations of interleaved carrier-assisted differential detection," *J. Lightw. Technol.*, vol. 39, no. 1, pp. 122-128, Jan. 2021.
- [4] H. Ji et al., "Carrier assisted differential detection with generalized and simplified receiver structure," J. Lightw. Technol., vol. 39, no. 22, pp. 7159-7167, Nov. 2021.
- [5] X. Chen et al., "High-capacity direct-detection systems," in Opt. Fiber Telecommunication VII, Orlando, FL, USA: Academic, pp. 419-441, 2020.
- [6] G. N. Liu et al., "IM/DD transmission techniques for emerging 5G fronthaul, DCI, and metro applications," J. Lightw. Technol., vol. 36, no. 2, pp. 560-567, Jan. 2018.
- [7] X. Pang et al., "200 Gbps/lane IM/DD technologies for short reach optical interconnects", J. Lightw. Technol., vol. 38, no. 2, pp. 492-503, Dec. 2019.
- [8] K. Zhong et al., "Digital signal processing for short-reach optical communications: A review of current technologies and future trends," J. Lightw. Technol., vol. 36, no. 2, pp. 377-400, Jan. 2018.
- [9] C. Yang et al., "IM/DD-based 112-Gb/s/lambda PAM-4 transmission using 18-Gbps DML," *IEEE Photon. J.*, vol. 8, no.3, pp. 1-7, Jun. 2016.
- [10] J. Estaran et al., "Artificial neural networks for linear and non-linear impairment mitigation in high-baudrate IMDD systems," in Proc. Eur. Conf. Opt. Commun. (ECOC), Dusseldorf, Germany, Sep. 2016, paper M.2.B.2.
- [11] A. G. Reza and J-K. K. Rhee, "Nonlinear equalizer based on neural networks for PAM-4 signal transmission using DML," *IEEE Photon. Technol. Lett.*, vol. 30, no. 15, pp. 1416-1419, Aug. 2018.
- [12] L. Yi et al., "Neural network-based equalization in high-speed PONs," in *Proc. Opt. Fiber Commun. Conf. (OFC)*, San Diego, CA, USA, Mar. 2020, paper T4D.3.
- [13] C. Y. Chuang *et al.*, "Employing deep neural network for high speed 4-PAM optical interconnect," in *Proc. Eur. Conf. Opt. Commun. (ECOC)*, Gothenburg, Sweden, Sep. 2017, paper W.2.D.2.
- [14] B. Karanov et al., "End-to-end deep learning of optical fiber communications," J. Lightw. Technol., vol. 36, no. 20, pp. 4843-4855, Oct. 2018.
- [15] Z. Xu et al., "Joint equalization of linear and nonlinear impairments for PAM4 short-reach direct detection systems", IEEE Photon. Technol. Lett., vol. 33, no. 9, pp. 425-428, May 2021.
- [16] L. Zhang et al., "160-Gb/s Nyquist PAM-4 transmission with GeSi-EAM using artificial neural network based nonlinear equalization," in Proc. Opt. Fiber Commun. Conf. (OFC), San Diego, CA, USA, Mar. 2020, paper M4F.5.
- [17] M. Luo *et al.*, "Transmission of 4×50-Gb/s PAM-4 signal over 80-km single mode fiber using neural network," in *Proc. Opt. Fiber Commun. Conf. (OFC)*, San Diego, CA, USA, Mar. 2018, paper M2F.2.
- [18] F. Da Ros et al., "Reservoir-computing and neural-network-based equalization for short reach communication," in Proc. Optica Adv. Photon. Cong. (APC), Maastricht, Limburg, Netherlands, Jul. 2022, paper SpTu1J.1.
- [19] Z. Xu et al., "Investigation on the impact of additional connections to feedforward neural networks for equalization in PAM4 short-reach direct detection links", in *Proc. Asia Commun. Photon. Conf. (ACP)*, Beijing, China, Nov. 2020, paper S4I.1.
- [20] C. Ye *et al.*, "Demonstration of 50Gbps IM/DD PAM4 PON over 10GHz class optics using neural network based nonlinear equalization," in *Proc. Eur. Conf. Opt. Commun. (ECOC)*, Gothenburg, Sweden, Sep. 2017, paper W.2.B.4.
- [21] N. Kaneda et al., "FPGA implementation of deep neural network based equalizers for high-speed PON," in Proc. Opt. Fiber Commun. Conf. (OFC), San Diego, CA, USA, Mar. 2020, paper T4D.2.
- [22] C. Bluemm et al., "800Gb/s PAM4 transmission over 10km SSMF enabled by low-complex duobinary neural network equalization," in Proc. Eur. Conf. Opt. Commun. (ECOC), Basel, Switzerland, Sep. 2022, paper Tu3C.4.

- [23] Z. Yang *et al.*, "Radial basis function neural network enabled C-band 4 × 50 Gbs PAM-4 transmission over 80 km SSMF," *Opt. Lett.*, vol. 43, no. 15, pp. 3542-3545, Aug. 2018.
- [24] Z. Xu et al., "Cascade recurrent neural network enabled 100-Gb/s PAM4 short-reach optical link based on DML," in Proc. Opt. Fiber Commun. Conf. (OFC), San Diego, CA, USA, Mar. 2020, paper W2A.45.
- [25] C. Ye et al., "Recurrent neural network (RNN) based end-to-end nonlinear management for symmetrical 50Gbps NRZ PON with 29dB+ loss budget," in Proc. Eur. Conf. Opt. Commun. (ECOC), Rome, Italy, Sep. 2018, paper Mo4B.3.
- [26] Q. Zhou et al., "Low computationally complex recurrent neural network for high speed optical fiber transmission," Opt. Commun., vol. 441, pp. 121–126, Jun. 2019.
- [27] X. Dai et al., "LSTM networks enabled nonlinear equalization in 50-Gb/s PAM-4 transmission links," Appl. Opt., vol. 58, no. 22, pp. 6079-6084, Jul. 2019.
- [28] Z. Xu et al., "Cascade recurrent neural network-assisted nonlinear equalization for a 100 Gb/s PAM4 short-reach direct detection system", Opt. Lett., vol. 45, no. 15, pp. 4216–4219, Aug. 2020.
- [29] B. Karanov et al., "End-to-end optimized transmission over dispersive intensity-modulated channels using bidirectional recurrent neural networks," Opt. Express, vol. 27, no. 14, pp. 19650-19663, Jul. 2019.
- [30] C. W. Peng et al., "Long short-term memory neural network for mitigating transmission impairments of 160 Gbit/s PAM4 microring modulation," in *Proc. Opt. Fiber Commun. Conf. (OFC)*, San Francisco, CA, USA, Jun. 2021, paper Tu5D.3.
- [31] C. Y. Chuang et al., "Convolutional neural network based nonlinear classifier for 112-Gbps high speed optical link," in Proc. Opt. Fiber Commun. Conf. (OFC), San Diego, CA, USA, Mar. 2018, paper W2A.43.
- [32] J. Zhang et al., "56-Gbit/s PAM-4 optical signal transmission over 100-km SMF enabled by TCNN regression model," *IEEE Photon. J.*, vol. 13, no.4, pp. 1-6, Aug. 2021.
- [33] P. Li et al., "100Gbps IM/DD transmission over 25km SSMF using 20G-class DML and PIN enabled by machine learning," in Proc. Opt. Fiber Commun. Conf. (OFC), San Diego, CA, USA, Mar. 2018, paper W2A.46.
- [34] E. Arnold et al., "Spiking neural network equalization on neuromorphic hardware for IM/DD optical communication," in Proc. Eur. Conf. Opt. Commun. (ECOC), Basel, Switzerland, Sep. 2022, paper Th1C.5.
- [35] G. Böcherer *et al.*, "Spiking neural network linear equalization: Experimental demonstration of 2km 100Gb/s IM/DD PAM4 optical transmission," in *Proc. Opt. Fiber Commun. Conf. (OFC)*, San Diego, CA, USA, Mar. 2023, Paper W4E.1.
- [36] Z. Xu et al., "Transfer learning aided neural networks for nonlinear equalization in short-reach direct detection systems," in Proc. Opt. Fiber Commun. Conf. (OFC), San Diego, CA, USA, Mar. 2020, paper TAD 4
- [37] Z. Xu et al., "Feedforward and recurrent neural network-based transfer learning for nonlinear equalization in short-reach optical links", J. Lightw. Technol., vol. 39, no. 2, pp. 475-480, Jan. 2021.
- [38] X. Liu et al., "Transfer-learning based convolutional neural network for short-reach optical interconnects," in Proc. Asia Commun. Photon. Conf. (ACP), Shanghai, China, Nov. 2021, paper T4A.79.
- [39] J. Xiao et al., "Fastly converged transfer learning using neuron-pruning nonlinear equalizer for intra data center networking," in Proc. Opto Electron. Commun. Conf. (OECC), Toyama, Japan, Jul. 2022, paper TuP-A-8.
- [40] J. Xiao et al., "Optimizations and investigations for transfer learning of iteratively pruned neural network equalizers for data center networking," Opt. Express, vol. 30, no. 20, pp. 36358–36367, Sep. 2022.
- [41] L. Xia et al., "Transfer learning assisted deep neural network for OSNR estimation," Opt. Express, vol. 27, no. 14, pp. 19398–19406, Jun. 2019.
- [42] D. Wang *et al.*, "Comprehensive eye diagram analysis: A transfer learning approach," *IEEE Photon. J.*, vol. 11, no. 6, Dec. 2019.
- [43] D. Azzimonti et al., "Active vs transfer learning approaches for QoT estimation with small training datasets," in Proc. Opt. Fiber Commun. Conf. (OFC), San Diego, CA, USA, Mar. 2020, paper M4E.1.

- [44] J. Zang et al., "Fast remodeling for nonlinear distortion mitigation based on transfer learning," Opt. Lett., vol. 44, no. 17, pp. 4243–4246, Sep. 2019.
- [45] P. J. Freire *et al.*, "Transfer learning for neural networks-based equalizers in coherent optical systems," *J. Lightw. Technol.*, vol. 39, no. 21, pp. 6733-6745, Nov. 2021.
- [46] P. J. Freire et al., "Domain adaptation: the key enabler of neural network equalizers in coherent optical systems," in Proc. Opt. Fiber Commun. Conf. (OFC), San Diego, CA, USA, Mar. 2022, paper Th2A.35.
- [47] Z. Xu et al., "Computational complexity analysis of neural network-based nonlinear equalization for short reach direct detection systems", in Proc. Asia Commun. Photon. Conf. (ACP), Chengdu, China, Nov. 2019, paper T2G.4
- [48] Z. Xu et al., "Computational complexity comparison of feedforward/radial basis function/recurrent neural network-based equalizer for a 50-Gb/s PAM4 direct-detection optical link", Opt. Express, vol. 27, no. 25, pp. 36953-36964, 2019.
- [49] Z. Xu et al., "Towards low computational complexity for neural network-based equalization in PAM4 short-reach direct detection systems by multi-symbol prediction," in Proc. Opt. Fiber Commun. Conf. (OFC), San Francisco, CA, USA, Jun. 2021, paper Th1A.47.
- [50] Z. Xu et al., "Low-complexity multi-task learning aided neural networks for equalization in short-reach optical interconnects", J. Lightw. Technol., vol. 40, no. 1, pp. 45-54, Jan. 2022.
- [51] B. Sang et al., "Multi-symbol output long short-term memory neural network equalizer for 200+ Gbps IM/DD system," in Proc. Eur. Conf. Opt. Commun. (ECOC), Bordeaux, France, Sep. 2021, paper We3C2.4.
- [52] B. Sang et al., "Low complexity neural network equalization based on multi-symbol output technique for 200+ Gbps IM/DD short reach optical system," J. Lightw. Technol., vol. 40, no. 9, pp. 2890-2900, May 2022
- [53] X. Huang et al., "Recurrent neural network based equalizer with embedded parallelization for 100Gbps/λ PON," in Proc. Opt. Fiber

- Commun. Conf. (OFC), San Francisco, CA, USA, Jun. 2021, paper M3G.2.
- [54] X. Huang et al., "Low-complexity recurrent neural network based equalizer with embedded parallelization for 100-Gbit/s/λ PON," J. Lightw. Technol., vol. 40, no. 5, pp. 1353-1359, Mar. 2022.
- [55] L. Ge et al., "Threshold-based pruned retraining Volterra equalization for 100 Gbps/lane and 100-m optical interconnects based on VCSEL and MMF," J. Lightw. Technol., vol. 37, no. 13, pp. 3222-3228, Jul. 2019.
- [56] C. Y. Chuang et al., "Sparse Volterra nonlinear equalizer by employing pruning algorithm for high-speed PAM-4 850-nm VCSEL optical interconnect," in Proc. Opt. Fiber Commun. Conf. (OFC), San Diego, CA, USA, Mar. 2019, paper M1F.2.
- [57] Z. Xu et al., "Sparsely-connected cascade recurrent neural network-based nonlinear equalizer for a 100-Gb/s PAM4 optical interconnect", in Proc. Asia Commun. Photon. Conf. (ACP), Shanghai, China, Nov. 2021, paper M5H.4.
- [58] L. Ge et al., "Compressed neural network equalization based on iterative pruning algorithm for 112-Gbps VCSEL-enabled optical interconnects," J. Lightw. Technol., vol. 38, no. 6, pp. 1323-1329, Mar. 2020.
- [59] M. Li et al., "High-throughput hardware deployment of pruned neural network based nonlinear equalization for 100-Gbps short-reach optical interconnect," Opt. Lett., vol. 46, no. 19, pp. 4980–4983, Oct. 2021.
- [60] G. S. Yadav et al., "Computation efficient sparse DNN nonlinear equalization for IM/DD 112 Gbps PAM4 inter-data center optical interconnects," Opt. Lett., vol. 46, no. 9, pp. 1999-2002, May 2021.
- [61] G. S. Yadav et al., "Sparsity learning deep neural network nonlinear equalization method for 112Gbps PAM4 short-reach transmission links," J. Lightw. Technol., vol. 41, no. 8, pp. 2333-2342, Apr. 2023.
- [62] Z. Xu et al., "Computationally-efficient sparsely-connected multioutput neural networks for IM/DD system equalization", in Proc. Opt. Fiber Commun. Conf. (OFC), San Diego, CA, USA, Mar. 2022, paper W2A.26.