European Conference on Optical Communication (ECOC) 2022 © Optica Publishing Group 2022

Deep Convolutional Recurrent Neural Network For Fiber Nonlinearity Compensation

Prasham Jain(1), Lutz Lampe(1), and Jeebak Mitra(2)

(1) Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC V6T 1Z4, Canada, prashamj@ece.ubc.ca, lampe@ece.ubc.ca
(2) Huawei Technologies Canada, Ottawa, ON K2K 3J1, Canada, Jeebak.Mitra@huawei.com

Abstract *An iterative deep convolutional recurrent neural network is proposed to mitigate fiber nonlinearity with distributed compensation of polarization mode dispersion, demonstrating 1.3 dB Q-factor gain over previous neural network based techniques for dual-polarized 960 km 32 Gbaud 64QAM transmission. ©2022 The Author(s)*

Introduction

Machine learning based methods have gained considerable interest for nonlinearity compensation (NLC) in optical fiber communication systems. It has been demonstrated that learned NLC methods are effective without apriori knowledge of system parameters[1]–[6]. The proposed neural network (NN) models can be broadly divided into two classes. Models such as multi-layer perceptrons (MLPs)[1], bidirectional recurrent neural networks (BiRNNs)[2] and convolutional BiRNNs (CBRNNs)[4], which characterize the nonlinear effects with few layers and mitigate them in a single module, can be classified as "lumped" NN models. Conversely, models which iteratively compensate distortions for short sections of the link in a piecemeal manner, such as learned digital backpropagation (LDBP)[5] and deep convolutional neural networks (DCNNs)[6] can be classified as "iterative" NN models. Despite notable progress in their development, comparison of various models has been restricted to their respective class[6],[7].

Against this backdrop, in this paper we propose a deep convolutional recurrent neural network (DCRNN), which advances the class of iterative learned NLC methods. In each step, the DCRNN compensates linear distortions using a complex valued convolutional layer followed by a bidirectional recurrent layer which captures the interaction between dispersion and nonlinearity to perform NLC. We also extend the piecemeal processing approach of DCRNN by incorporating polarization mode dispersion (PMD) compensation in each iterative step. We refer to this extension as the DCRNN-PMD model. We compare, for the first time, the effectiveness of the proposed models against both lumped and iterative neural networks in addition to deterministic methods such as digital backpropagation (DBP). For this, we account for the fact that different NN models have varying potential for complexity reduction, and we apply an iterative pruning and fine-tuning approach based on the lottery ticket hypothesis (LTH)[8] to operate the learned NLC methods at their optimal complexity. We demonstrate that iterative NN models outperform their lumped counterparts as well as deterministic methods including DBP. Moreover, we show that the proposed DCRNN-PMD model achieves the best performance among all schemes.

Deep Convolutional Recurrent Neural Network

Lumped models explored in the literature employ a combination of convolutional, recurrent, and fully connected layers to characterize the accumulated distortions from signal propagation along the entire length of the fiber. They draw from the ability of neural networks to learn complex functions efficiently from data and the proven potency of these topologies in applications of image and speech processing[4]. On the other hand, iterative models perform equalization in multiple repetitive steps, where each step can be considered equivalent to distortion compensation for a small section of the fiber. Among deterministic equalization methods, this notably applies to DBP, which performs channel inversion using split step Fourier method (SSFM). Häger *et al.* proposed a learned approximation of DBP (LDBP)[5] which iteratively compensates linear and nonlinear effects. However, LDBP does not consider the interaction between dispersion and nonlinearity which can degrade its performance when operated with fewer coarser steps needed to keep the complexity low. Recently, Sidelnikov *et al.* addressed this in the DCNN[6] model by introduc-

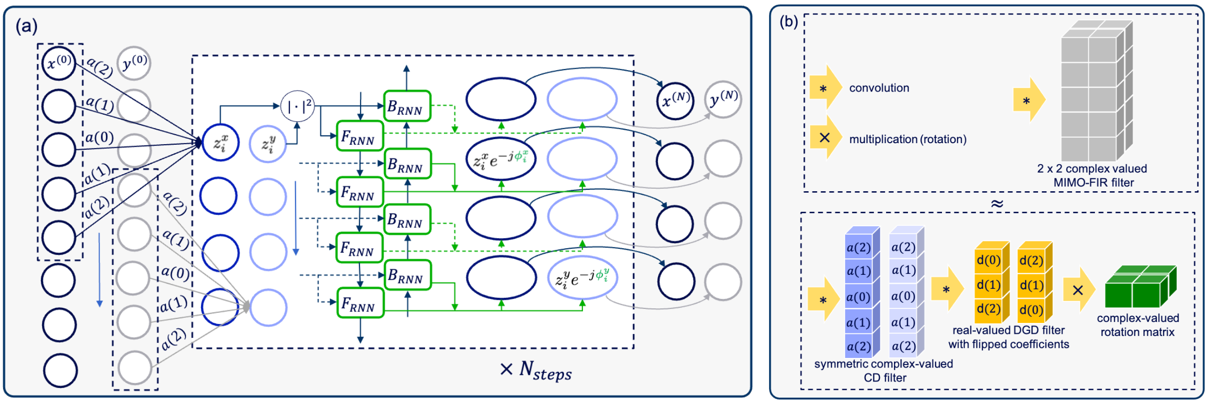

Fig. 1: (a) Architecture of the proposed DCRNN Model with N steps. (b) Decomposition of distributed linear filter.

ing an additional convolutional filter at the nonlinear compensation step. However, this approach is limited by the width of the filter and would incur considerable computational complexity if we try to cover the entire dispersion spread. In our proposed DCRNN model, depicted in Fig. 1a, we introduce a bidirectional recurrent neural network (BiRNN) layer, which can account for the interaction of dispersion with nonlinearity along the entire delay spread at a low cost by processing the symbols as a sequence and preserving information from past (and future) symbols in its memory. In each step of DCRNN, a complex-valued 1-D linear convolutional layer performs chromatic dispersion (CD) compensation for each polarization. The output of the i^{th} convolution step for the x-polarization can be written as

$$z_i^x(k) = \sum_{n=-l}^{l} a_i(n) x_i(k+n), \qquad (1)$$

where $(2l+1)$ is the width of the convolutional kernel, $x_i(n)$ are the x-pol. inputs and $a_i(n)$ are trainable complex valued weights at step i. A BiRNN layer then processes the energies of dispersion compensated symbols as a sequence:

$$h_{f,i}(k) = \tanh(W_{F,i}[h_{f,i}(k-1), |z_i^x(k)|^2, |z_i^y(k)|^2])$$
$$h_{b,i}(k) = \tanh(W_{B,i}[h_{b,i}(k+1), |z_i^x(k)|^2, |z_i^y(k)|^2])$$
$$\phi_i^{x/y}(k) = (f_i^{x/y} * h_{f,i})(k) + (b_i^{x/y} * h_{b,i})(k), \qquad (2)$$

where $W_{F,i}, W_{B,i}, f_i^{x/y}$ and $b_i^{x/y}$ are trainable real-valued weights. Each iteration ends with the application of the nonlinear compensation applied as $x_{i+1}(k) = z_i^x(k) e^{-j\phi_i^x(k)}$. For PMD compensation, we consider both a lumped and an iterative approach to highlight the impact of piecemeal compensation. For the base DCRNN model, we use a single lumped 2-D complex-valued convolutional layer, akin to a 2×2 MIMO-FIR filter, using output symbols of both polarizations from the last iterative block. In the extended DCRNN-PMD model, we modify the convolutional layer in each iterative block for distributed compensation of PMD. To reduce complexity, we simplify the linear compensation step based on the decomposition scheme proposed by Bütler *et al.*[9], as depicted in Fig. 1b, using a short differential group delay (DGD) filter followed by a rotation matrix.

Complexity Reduction Of Neural Networks

Fujisawa *et al.*[10] illustrated complexity reduction of NN-based NLC methods using weight pruning. Since the performance of deep composite networks is irregularly sensitive to the number of weights pruned from different layers, we use the automatic model compression (AMC) technique[11] to identify optimal sparsity factors for each of the relatively complex layers of the lumped models. For the iterative models with relatively simple and repeating layers, we choose the sparsity factors heuristically. Then, to effectively retrain the pruned model to compensate for the removed weights, we apply LTH[8], wherein we rewind the learning rate schedule before fine-tuning. Additionally, based on simple physical considerations, in the DCRNN model, we make the convolutional filters for CD compensation symmetrical. Also, to further reduce the number of trainable parameters, we use the same filter coefficients for the CD filters of each polarization.

Numerical Results

We consider dual polarized 64-QAM transmission at 32 GBaud using root raised cosine (RRC) pulses at a roll-off factor of 0.06. The optical link consists of 12 spans of 80 km single mode fiber using EDFA amplification with a noise figure of 4.5 dB at the end of each span. The fiber has attenuation coefficient $\alpha = 0.21$ dB/km, dispersion coefficient $\beta_2 = -21.49$ ps^2/km, nonlinear coefficient $\gamma = 1.14$ (W·km)$^{-1}$ and DGD $D_{\text{PMD}} = 0.1$

ps/$\sqrt{\text{km}}$ at $\lambda = 1552.93$ nm. At the receiver, the signal is coherently detected, sampled at 2 samples/symbol and matched filtered with an RRC filter. For lumped NN models, a frequency domain CD equalizer is applied prior to the NN operation.

For our proposed DCRNN and DCRNN-PMD models, we operate at 1 step per span (StPS) with 29 taps in the symmetric convolutional CD filter, 2 hidden units in each BiRNN layer, 19 taps in the lumped PMD filter and 3 taps in each of the distributed DGD filters. We compare our results with previous works on NN based NLC from iterative and lumped paradigms. For iterative models, we consider the DCNN[6] model, also at 1-StPS, with a 29 tap convolutional CD filter and a 19 tap convolutional filter for nonlinearity. For lumped compensation, we consider a single hidden layer complex valued MLP[1] with 192 hidden units, a Bidirectional LSTM (BiLSTM)[2] with 100 hidden units and a Convolutional BiLSTM[4] with an additional 29 tap convolutional layer. For the BiLSTM, we apply the low complexity center-oriented LSTM (Co-LSTM)[3] implementation including the simplified mode and recycling mechanism. We train each model using 2^{18} symbols for 10^{3} epochs with a batch size of 10^{4} using the *Adam* optimizer and a cosine annealed learning rate schedule. Performance of the model is evaluated on a separate set of 2^{16} symbols. We also compare our results with conventional DBP[12] at 1, 2 and 3 StPS processed at 2 samples per symbol. Considering our significant efforts towards optimizing the NN training process, it is only fair that we include several key optimizations for DBP which have not been incorporated in several previous works on lumped NN based NLC[1] and hence may have caused some unintentional bias. In our DBP implementation, we numerically optimize the nonlinear parameter for each launch power and use effective step length accounting for signal attenuation. In addition, we compensate for the remaining phase rotation and PMD using an adaptive 2×2 LMS based MIMO-FIR filter with 19 taps.

In Fig. 2, we compare the Q-factor improvement achieved by each NLC method over linear equalization at various stages of complexity reduction. The rightmost point on each curve represents the performance and complexity of each model at the specifications described above. LTH-based weight pruning is then applied to reduce complexity of each model and obtain performance results at each complexity level. We find that pruning a fully parameterized network delivers better performance-complexity trade-off as compared to training a smaller network from scratch. This effect is analogous to pre-training as the pruned network serves as excellent initialization for the remaining weights. We find that the lumped models require considerably high complexity to keep up with DBP. Only convolutional BiLSTM is able to match the performance of DBP at 2-StPS but it does so at roughly 5 times the complexity even after pruning. Iterative models on the other hand outperform DBP at significantly lower complexity. The proposed DCRNN model outperforms linear compensation by 2.43 dB, 1-StPS DBP by 1.42 dB and next best learned method (DCNN) by 0.87 dB. This illustrates the improved capability of the DCRNN model to capture the interaction between dispersion and nonlinearity. Distributed PMD compensation using the DCRNN-PMD model provides another 0.43 dB Q-factor gain, clearly demonstrating the advantage of iterative compensation.

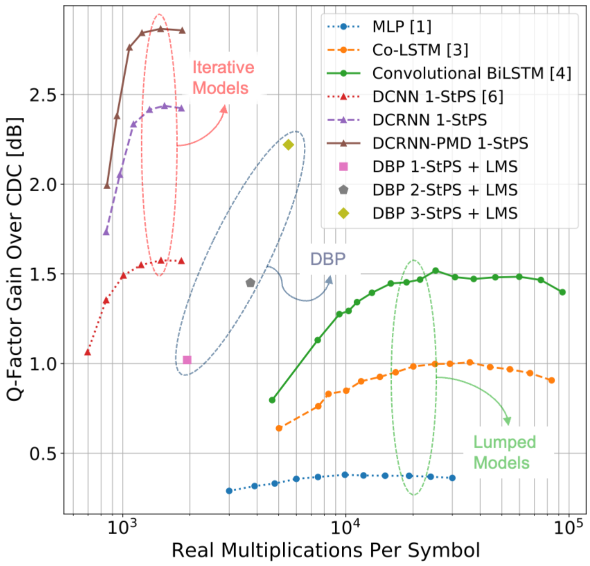

Fig. 2: Performance vs. complexity comparison of various NLC techniques with LTH-based NN complexity reduction.

Conclusions

The proposed DCRNN-PMD model is able to outperform conventional DBP and previous NN based equalizers in terms of both performance and complexity. LTH based pruning is able to further reduce DCRNN-PMD complexity by almost 50% with negligible performance loss. Lumped NN models require considerably more parameters to adequately characterise and mitigate nonlinear impairments. Iterative models take advantage of simpler characterisation of nonlinear effects on a per step basis and perform well using very few learnable parameters.

References

[1] P. J. Freire, V. Neskornuik, A. Napoli, *et al.*, "Complex-valued neural network design for mitigation of signal distortions in optical links", *Journal of Lightwave Technology*, vol. 39, no. 6, pp. 1696–1705, 2021. DOI: 10.1109/JLT.2020.3042414.

[2] S. Deligiannidis, C. Mesaritakis, and A. Bogris, "Performance and complexity evaluation of recurrent neural network models for fibre nonlinear equalization in digital coherent systems", in *European Conference on Optical Communications (ECOC)*, 2020, pp. 1–4. DOI: 10.1109/ECOC48923.2020.9333288.

[3] H. Ming, X. Chen, X. Fang, L. Zhang, C. Li, and F. Zhang, "Ultralow complexity long short-term memory network for fiber nonlinearity mitigation in coherent optical communication systems", *Journal of Lightwave Technology*, vol. 40, no. 8, pp. 2427–2434, 2022. DOI: 10.1109/JLT.2022.3141404.

[4] P. J. Freire, Y. Osadchuk, B. Spinnler, *et al.*, "Experimental study of deep neural network equalizers performance in optical links", in *Optical Fiber Communications Conference and Exhibition (OFC)*, 2021, pp. 1–3.

[5] C. Häger and H. D. Pfister, "Physics-based deep learning for fiber-optic communication systems", *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 1, pp. 280–294, 2021. DOI: 10.1109/JSAC.2020.3036950.

[6] O. Sidelnikov, A. Redyuk, S. Sygletos, M. Fedoruk, and S. Turitsyn, "Advanced convolutional neural networks for nonlinearity mitigation in long-haul wdm transmission systems", *Journal of Lightwave Technology*, vol. 39, no. 8, pp. 2397–2406, 2021. DOI: 10.1109/JLT.2021.3051609.

[7] P. J. Freire, Y. Osadchuk, B. Spinnler, *et al.*, "Performance versus complexity study of neural network equalizers in coherent optical systems", *Journal of Lightwave Technology*, vol. 39, no. 19, pp. 6085–6096, 2021. DOI: 10.1109/JLT.2021.3096286.

[8] A. Renda, J. Frankle, and M. Carbin, "Comparing rewinding and fine-tuning in neural network pruning", in *International Conference on Learning Representations*, 2020. [Online]. Available: https://openreview.net/forum?id=S1gSj0NKvB.

[9] R. M. Bütler, C. Häger, H. D. Pfister, G. Liga, and A. Alvarado, "Model-based machine learning for joint digital backpropagation and pmd compensation", vol. 39, no. 4, pp. 949–959, 2021. DOI: 10.1109/JLT.2020.3034047.

[10] S. Fujisawa, F. Yaman, H. G. Batshon, *et al.*, "Nonlinear impairment compensation using neural networks", in *2021 Optical Fiber Communications Conference and Exhibition (OFC)*, 2021, pp. 1–3.

[11] Y. He, J. Lin, Z. Liu, H. Wang, L.-J. Li, and S. Han, "AMC: AutoML for model compression and acceleration on mobile devices", in *Computer Vision – ECCV 2018*, V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, Eds., Cham: Springer International Publishing, 2018, pp. 815–832. DOI: 10.1007/978-3-030-01234-2_48.

[12] E. Ip and J. M. Kahn, "Compensation of dispersion and nonlinear impairments using digital backpropagation", *Journal of Lightwave Technology*, vol. 26, no. 20, pp. 3416–3425, 2008. DOI: 10.1109/JLT.2008.927791.