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

On the Impact of the Optical Phase Conjugation on the Computational Complexity of Neural Network-Based Equalisers

D. Argüello Ron , K. Nurlybayeva, M. Kamalian-Kopae , A. A. Ali, E. Turitsyna , S. Turitsyn

Aston Institute of Photonic Technologies, d.arguelloron@aston.ac.uk

Abstract *We develop a low complexity complex-valued neural network to compensate the nonlinearity from the transmission of PDM 28 Gbaud 64QAM over 400km of SSMF, by combining midlink optical phase conjugation and pruning.*

Introduction

In recent years, machine learning has gained attraction as a promising technique for mitigating fibre nonlinearities in optical communication systems[1],[2]. The use of neural networks (NN) in particular allows for the enhancement of existing fibre-optic systems without any prior knowledge of their characteristics. Despite its outstanding performance, the computational cost inflicted by such equalisers is high when compared to existing approaches such as digital backpropagation (DBP)[3],[4]. Therefore, it is imperative to reduce the computational complexity (CC) of the neural-network equalisers, while maintaining adequate performance. In order to achieve this, we have previously suggested a neural network equaliser combined with optical solutions, such as Optical Phase Conjugation (OPC)[5]. We showed that combining two nonlinearity mitigation techniques improves the system performance, and argued that because OPC decreases the effective channel memory, it decreases the processing delay and has the potential to reduce the CC of the obtained NN equaliser[5]. In this paper, we go one step further and demonstrate that the required CC of an NN-based equaliser is indeed significantly lower when optical solutions such as OPC are implemented. We employ the NN compression technique known as pruning[6] to decrease the CC of the NN, and show that pruning is more successful when OPC is in place. This means that optical solutions like OPC (and maybe dispersion managed links) can help to simplify the NN-based equaliser's architecture. Furthermore, we demonstrate that the combination of the two techniques outperforms each technique on its own. The performance of OPC-aided optical communication is limited by a few factors of the system and is shown to deteriorate as the asymmetry of the link increases[7]. While there are techniques to alleviate the impact of this asymmetry (see[8] for example), we propose to use NN at the receiver to mitigate them. However, this NN equaliser needs to have acceptable CC.

In this work, we make use of complex-valued NNs (CVNNs) that are more effective in capturing the sequential relations between complex samples as the real and imaginary parts of complex-valued numbers are not separated. Our methodology is as the following: using experimental data from two systems with and without OPC, we use Bayesian optimisation algorithm (BO) to find the best NN architecture at each launch power and number of taps[5]. The resulting optimum configurations can be potentially different and smaller for the case with OPC. The possible reason for that is the simpler interconnection and underlying nonlinear relation between time samples of the signal due to the presence of the OPC. The evolution of signal in the fibre can be captured by a smaller NN as fewer features are required to explain it. This phenomenon is well known in the literature, particularly for inline dispersion-managed communication links[9]. However, machine learning solutions tend to be over-parameterized, i. e. the resulting NN may be unnecessarily large[6]. The technique known as *pruning* address this by removing weights (or nodes) with small contribution (i.e small magnitude or connectivity) at small or no cost to the overall accuracy[6]. After pruning the NN obtained using BO we show that the one with OPC has a greater potential to be pruned compared to the one without OPC.

Experimental Setup

Figure 2 depicts the experimental setup. A polarisation division multiplexing (PDM) transmission at 28GBaud using 64 QAM signal has been considered through the 4 spans of SSMF fibre SSMF ($\alpha = 0.2$ dB/km, $D = 17$ ps/nm/km, $\gamma = 1.3$ /W/km), and Mid-link OPC. In the case of the use of OPC, in the middle of the link, the signal was amplified by an EDFA with fixed output power (15 dBm), and its conjugate was transmitted in the

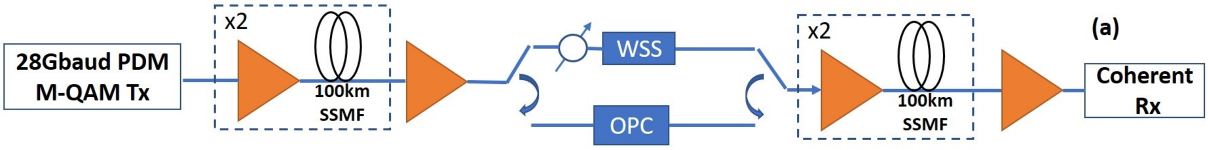

Fig. 1: Block diagram of the experimental setup

second half of the link. Otherwise, to simulate the insertion penalty of the OPC, a variable optical attenuator with a WSS was used, with the OSNR set to be the same in the second half of the connection for both situations (with and without the OPC). The coherent receiver is comprised of a local oscillator with a linewidth of 100 kHz, whose signal was mixed with the received signal in a 90° optical hybrid. Four balanced photodiodes were linked to the hybrid outputs, and an analogue-to-digital converter was used in the form of a real-time sampling scope (100 GS/s sampling rate, 33 GHz 3-dB bandwidth). On a desktop computer, offline data-assisted digital signal processing was performed, and machine learning was applied before symbol de-mapping. More about this setup can be found in[5].

Machine Learning-based Equalisation

The CVNN's configuration was inferred employing the BO algorithm for all combinations of launched power and input vector sizes. The values optimised were the number of neurons, number of layers, dropout rate and the L2 regularisation penalty These last two parameters have been used to limit the risk of over-fitting. The number of neurons tested was $n \in [50, 600]$, meanwhile that the possible values for the number of layers were $l \in [2, 6]$. The dropout rate was chosen between $p \in [0, 0.5]$ and the regularisation penalty between $L2 \in [10^{-4}, 10^{-1}]$. To investigate the memory reduction impact of the OPC, the input layer consists of a sequence of successive samples from the two polarisations with varying sizes. For the activation function, except for the final layer, we used the Cartesian Rectified Linear Unit(relu) described as in[10] (it applies relu to both real and imaginary parts), while for the final layer we use the linear activation function. For the training, we employed 2^{18} complex-valued samples from both polarisations and the corresponding target broadcast QAM symbols. The training, validation, and test sets each account for 80, 10, and 10 per cent of the total dataset, respectively. The numbers of epochs and the batch size are set to 300 and 2000, respectively. Following initialisation by He uniform[11], weights are learnt from propagating data using Adam optimiser with the learning rate of 0.001, which minimises the complex valued mean-squared error loss (MSE).

Pruning

As it was mentioned above, when designing a NN there is a trend toward using over-parameterized architectures, in order to guarantee good model performance and especially due to the benefits in the learning capabilities of the model[6],[12]. This last point is a consequence of the effect that a larger number of parameters has on the loss function, as it helps to make it smoother. Thus, it is easier for gradient descent techniques to converge[6]. Nevertheless, the parameter redundancy consequences of over-parameterization comes at the cost of larger computational and memory resources requirements[13],[14]. Therefore, noticeable efforts are being done in developing techniques that can help to simplify the NNs without drastically damaging their performance[13]–[15]. One of them is the already mentioned technique known as pruning, which removes redundant NN elements, reducing its size and computational complexity[6],[15],[16]. There are several types of pruning, depending on the element to be pruned and when this process takes place[6]. We apply pruning to the weights of the model, based on their magnitude. Thus, the smaller weights which contain less information are removed first, until a desired sparsity level (e.g. % of pruned weights) is reached. Here, the magnitude of the weights is defined as the norm of the complex value. One of the novelties of this work is the application of this technique to the general case of Deep CVNN in the context of optical channel equalisation. Although some efforts have been done to compress CVNN, they targeted other structural elements, techniques and applications[17] or very specific architectures[18].

Results

In this work, the CVNN were developed employing TensorFlow and based on a library[19], meanwhile that the pruning process was implemented using the TensorFlow Model Optimisation Toolkit – Pruning API. It is worth noticing that these frameworks are not ready to work with CVNN and therefore a custom implementation was carried out based on the code already available. The

Tab. 1: Optimised CVNN architectures

Optical Solutions	Optimal Parameters	Power
OPC	$n = [519, 505, 531, 151, 562, 140]$, $l = 6$, $p = 0.1$, $L2 = 6.5 \cdot 10^{-6}$	8 dBm
No OPC	$n = [366, 422, 600, 600, 242, 327]$, $l = 6$, $p = 0$, $L2 = 1 \cdot 10^{-6}$	8 dBm
OPC	$n = [430, 532, 406, 462, 315]$, $l = 5$, $p = 0.35$, $L2 = 6.5 \cdot 10^{-3}$	9 dBm
No OPC	$n = [600, 600, 600, 600, 568, 50]$, $l = 6$, $p = 0.1$, $L2 = 0.1$	9 dBm

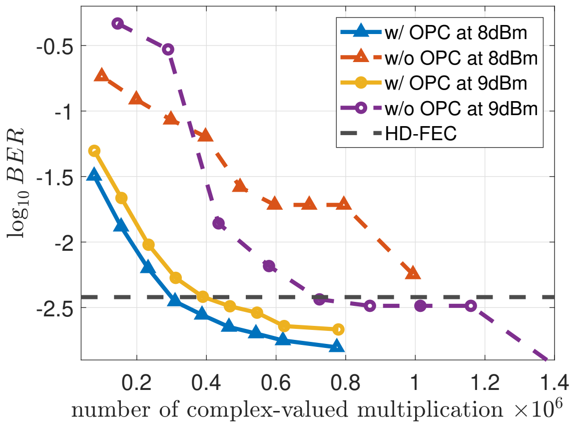

Fig. 2: BER achieved for different CC values for 2 datasets where optical solutions -OPC- are used and 2 where not. 2 launch powers are considered : 8 dBm and 9 dBm

optimal configuration values for the CVNNs produced by the BO can be found in Table 1. From these results, it can be concluded that regarding the architecture, using OPC permits to have less dense layers and even a lower number of them. As it was mentioned above, in addition to performance, the CC is another largely relevant metric for the deployment of equalisers developed using NNs in real optical communication systems. In this work, it is defined as the number of complex-valued multiplications[3],[20]:

$$CC = (1-s) \cdot (n_s n_i n_1 + \sum_{l=1}^{L-1} n_l n_{l+1} + n_o n_L) \quad (1)$$

where s is the level of sparsity achieved by pruning the NN, n_s is the memory size (e.g $2N+1$, being N the number of taps), n_i is the number of input features (e.g 4 as the real and imaginary parts of two polarisation components are used), n_o is the number of outputs (e.g 2, as the real and imaginary parts of each symbol are recovered) and n_l corresponds to the number of neurons in each layer with $l \in [1, L]$ and can be found for the different architectures in Table 1. In Fig. 2 it is possible to appreciate the performance vs the complexity trade-off achieved thanks to pruning and taking into account the use or not of OPC for 8 dBm and 9 dBm launch powers, and 9 taps. Thus, in the y-axis, the value of the performance, defined as BER, is represented. The x-axis contains the CC values, calculated following Eq. (1), for the different NN architectures described in Table 1 before and after having been pruned to different degrees. In this case, the sparsity level s ranged from 20 % to 90 %, with a 10 % increment, for the pruned models. $s = 0$ % corresponds to the original models, previous to pruning. This way we demonstrate that using OPC not only helps to have simpler NNs previous to pruning, but it also allows to reach larger pruning levels without reaching the HD-FEC threshold (e.g. $BER = 3.8 \cdot 10^{-3}$). In fact, for a launch power of 8 dBm, it is possible to prune up to 60 % of the weights and still be below the HD-FEC threshold. For a power of 9 dBm, sparsity levels up to 40% can be achieved and still stay below the threshold of interest. When OPC is not used, the optimised CVNN is not able to be lower than the HD-FEC threshold for any CC value for a launch power of 8 dBm. Meanwhile, for 9 dBm, only the performance of the original model is clearly below the threshold and those pruned up to $s = 60$ % are slightly above or below it. This demonstrates the powerful impact of OPC on simplifying the architecture of the NN-based equalisers. OPC not only improves the performance of the optical communication system (see[5]) but it also helps the pruning technique when trying to reduce the CC of the equaliser.

Conclusion

We have implemented an optimised CVNN-based equaliser using BO. Moreover, we have demonstrated that it is able to reduce nonlinear impairments for a PDM 28 Gbaud OPC-aided fibre-optic communication in a 400 km transmission. Furthermore, we showed evidence that OPC helps to reduce the effective channel memory allowing to use smaller and less complex CVNN-based equaliser. Finally, we implement the compression technique known as pruning for the case of CVNN and show that better performance vs complexity ratios can be achieved if combined with OPC.

Acknowledgements

This work has received funding from: EU H2020 MSCA project No. 860360 and EPSRC project TRANSNET.

References

[1] E. Giacoumidis, Y. Lin, J. Wei, I. Aldaya, A. Tsokanos, and L. Barry, "Harnessing machine learning for fiber-induced nonlinearity mitigation in long-haul coherent optical ofdm", *Future Internet*, vol. 11, no. 1, p. 2, 2018. DOI: 10.3390/fi11010002.

[2] F. Musumeci, C. Rottondi, A. Nag, *et al.*, "An overview on application of machine learning techniques in optical networks", *IEEE Communications Surveys Tutorials*, vol. 21, no. 2, pp. 1383–1408, 2019. DOI: 10.1109/comst.2018.2880039.

[3] 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.

[4] O. Sidelnikov, A. Redyuk, and S. Sygletos, "Equalization performance and complexity analysis of dynamic deep neural networks in long haul transmission systems", *Optics Express*, vol. 26, no. 25, p. 32 765, 2018. DOI: 10.1364/oe.26.032765.

[5] M. Kamalian-Kopae, A. A. Ali, K. Nurlybayeva, A. Ellis, and S. Turitsyn, "Neural network-enhanced optical phase conjugation for nonlinearity mitigation", in *2022 Optical Fiber Communications Conference and Exhibition (OFC)*, IEEE, 2022, pp. 1–3.

[6] T. Hoefler, D. Alistarh, T. Ben-Nun, N. Dryden, and A. Peste, *Sparsity in deep learning: Pruning and growth for efficient inference and training in neural networks*, 2021. arXiv: 2102.00554 [cs.LG].

[7] M. A. Al-Khateeb, M. E. McCarthy, C. Sánchez, and A. D. Ellis, "Nonlinearity compensation using optical phase conjugation deployed in discretely amplified transmission systems", *Optics express*, vol. 26, no. 18, pp. 23 945–23 959, 2018.

[8] M. Al-Khateeb, M. Tan, T. Zhang, and A. D. Ellis, "Combating fiber nonlinearity using dual-order raman amplification and opc", *IEEE Photonics Technology Letters*, vol. 31, no. 11, pp. 877–880, 2019.

[9] A. Napoli, Z. Maalej, V. A. Sleiffer, *et al.*, "Reduced complexity digital back-propagation methods for optical communication systems", *Journal of lightwave technology*, vol. 32, no. 7, pp. 1351–1362, 2014.

[10] Y. Kuroe, M. Yoshid, and T. Mori, "On activation functions for complex-valued neural networksâexistence of energy functionsâ", in *Artificial Neural Networks and Neural Information ProcessingâICANN/ICONIP 2003*, Springer, 2003, pp. 985–992.

[11] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification", in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1026–1034.

[12] Z. Allen-Zhu, Y. Li, and Z. Song, "A convergence theory for deep learning via over-parameterization", in *International Conference on Machine Learning*, PMLR, 2019, pp. 242–252.

[13] S. Han, H. Mao, and W. J. Dally, *Deep compression: Compressing deep neural networks with pruning, trained quantization and huffman coding*, 2016. arXiv: 1510.00149 [cs.CV].

[14] V. Sze, Y.-H. Chen, T.-J. Yang, and J. S. Emer, "Efficient processing of deep neural networks: A tutorial and survey", *Proceedings of the IEEE*, vol. 105, no. 12, pp. 2295–2329, 2017. DOI: 10.1109/JPROC.2017.2761740.

[15] T. Liang, J. Glossner, L. Wang, S. Shi, and X. Zhang, *Pruning and quantization for deep neural network acceleration: A survey*, 2021. arXiv: 2101.09671 [cs.CV].

[16] X. Dong and L. Zhou, *Understanding over-parameterized deep networks by geometrization*, 2019. arXiv: 1902.03793 [cs.LG].

[17] Z. Dong and H. Huang, "A structural optimization algorithm for complex-valued neural networks", in *2019 IEEE Symposium Series on Computational Intelligence (SSCI)*, 2019, pp. 1530–1535. DOI: 10.1109/SSCI44817.2019.9003040.

[18] D. Jianping, N. Sundararajan, and P. Saratchandran, "Communication channel equalization using complex-valued minimal radial basis function neural networks", *IEEE Transactions on Neural Networks*, vol. 13, no. 3, pp. 687–696, 2002. DOI: 10.1109/TNN.2002.1000133.

[19] J. A. Barrachina, *Complex-valued neural networks (cvnn)*, version v1.0.3, Jan. 2021. DOI: 10.5281/zenodo.4452131. [Online]. Available: https://doi.org/10.5281/zenodo.4452131.

[20] D. A. Ron, P. J. F. D. C. Sourza, J. E. Prilepsky, M. Kamalian-Kopae, A. Napoli, and S. K. Turitsyn, "Experimental implementation of a neural network optical channel equalizer in restricted hardware using pruning and quantization", 2022.