



HAL
open science

Attention-Based Neural Network Equalization in Fiber-Optic Communications

Abtin Shahkarami, Mansoor Yousefi, Yves Jaouën

► **To cite this version:**

Abtin Shahkarami, Mansoor Yousefi, Yves Jaouën. Attention-Based Neural Network Equalization in Fiber-Optic Communications. Asia Communications and Photonics Conference 2021, 2021, 10.1364/acpc.2021.m5h.3 . hal-03585118

HAL Id: hal-03585118

<https://telecom-paris.hal.science/hal-03585118>

Submitted on 22 Feb 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Attention-Based Neural Network Equalization in Fiber-Optic Communications

Abtin Shahkarami, Mansoor I. Yousefi, and Yves Jaouen

Department of Communications and Electronics, Polytechnic Institute of Paris, Telecom Paris, 91120 Palaiseau, France.

(email: abtin.shahkarami@telecom-paris.fr, yousefi@telecom-paris.fr, yves.jaoeun@telecom-paris.fr)

Abstract:

An attention mechanism is integrated into neural network-based equalizers to prune the fully-connected output layer. For a 100 GBd 16-QAM 20 × 100 km SMF transmission, this approach reduces the computational complexity by ~15% in a CNN+LSTM model.

© 2021 The Author(s)

1. Introduction

Nonlinear effects are a major limiting factor in fiber-optic communications. The complexity of receivers designed to equalize these impairments must be reduced to enable higher data rates. Non-model-driven deep learning-based equalization methods have shown promise in this area.

In this category of methods, a neural network that does not incorporate the channel model (in contrast to model-driven architectures [1, 2]) is trained using pairs of the sampled waveform at RX (or typically the waveform after chromatic dispersion compensation) and the vector of transmitted symbols. Several such models have been proposed, including multi-layer perceptrons (MLP) [3], convolutional neural network (CNN) [4], bidirectional long short-term memory (bi-LSTM) [5], gated recurrent unit (GRU) [6], and CNN+LSTM models [7]. It is shown that the CNN+LSTM model outperforms the other models in terms of bit-error-rate (BER). [7].

In these models, the output layer is often a fully-connected layer with one hidden unit per symbol in PAM systems, or two hidden units per symbol in QAM systems (one for the real and one for the imaginary part). This results in $2 \times n^{[L-1]}$ real multiplications per symbol in the output layer, where $n^{[L-1]}$ is the number of hidden units in the layer $L - 1$. Although dense layers could yield good BER performance, their complexity is high. Furthermore, the signal is already almost equalized in the previous layers, especially in the CNN+LSTM model; thus, paying the same attention to all the nodes in different temporal distances might not be optimal.

In this paper, we address this problem using an attention mechanism [8]. We consider a state-of-the-art CNN+LSTM model and reduce the number of floating-point operations (FLOPs) in it (for the inference mode) by learning and removing the low-impact connections in the output fully-connected layer. We also show that leveraging this technique facilitates replacing the LSTM units with simple RNN cells without performance deterioration. We present a cumulative FLOPs reduction of ~27% using this approach, broken down into around 15% for connections-dropping and around 12% for replacing LSTMs.

2. Fiber-optic system model

Nonlinear Schrödinger (NLS) equation models the evolution of signal with one polarization in the standard single-mode fiber (SMF)

$$\frac{\partial q(t, z)}{\partial z} = -\frac{\alpha}{2}q(z, t) - \frac{j\beta_2}{2}\frac{\partial^2 q}{\partial t^2} + j\gamma|q(t, z)|^2q(t, z), \quad 0 \leq z \leq \mathcal{L}, \quad (1)$$

where $q(t, z)$ is the complex envelope of the signal propagating in fiber as a function of time t and distance z . Here, γ is the nonlinearity parameter, β_2 and α are respectively chromatic dispersion and attenuation coefficients, and \mathcal{L} is the fiber length.

The optical link is split into several spans of equal length. Split-step Fourier method (SSFM) is applied to solve NLS equation in order to calculate the propagation of the signal in each span. To compensate for fiber loss, amplification is performed using Erbium-doped fiber amplifiers (EDFA) after each span, followed by ASE noise addition.

At the transmitter (TX) first the input bit-stream $\mathbf{m} = (m_1, m_2, \dots, m_{N_b})$, $m_i \in \{0, 1\}$, is translated to a sequence of symbols $\mathbf{S} = (s_1, s_2, \dots, s_{N_s})$, where the symbols are drawn from a QAM-constellation. Following this, \mathbf{S} is mapped to the waveform $q(t, 0) = \sum_{i=0}^{N_s} s_i p(t - i/R_s)$, where $p(t)$ is the pulse shape and R_s is the baud rate. The

CNN+LSTM model and the corresponding attention-integrated version were adopted to learn the equalization. The models were created and trained using Tensorflow 2.0 in Python. Logcosh was set as the loss function, and the optimization was conducted using Adam algorithm with the learning rate of 0.001, having the reduce on plateau property with the patience of 7 and the factor of 0.7, $\beta_1 = 0.85$, and $\beta_2 = 0.999$. The hyper-parameters, the same for both the models, were optimized using k-fold cross-validation in such a way as to achieve the maximum performance to the same level as digital back-propagation (DBP). The obtained attention-integrated model was then pruned with the drop-connection threshold of 0.1%.

Fig. 3 demonstrates the effective SNR plot of the resulting models. As expected, both models provide the same level of performance. This is while the number of connections in the output layer of the attention-integrated model is dropped by 25%, resulting in around 15% FLOPs reduction over CNN+LSTM without attention. It is also noteworthy to elaborate that the output of the attention mechanism can be exploited more fruitfully in field-programmable gate array (FPGA) and circuit design frameworks by assigning lower bits to the connections with moderate impacts.

The performance of simple RNN cells in place of LSTMs was also investigated. It was observed the CNN+RNN achieves the same performance as CNN+LSTM in both cases. Considering this observation and the discussion in Section 3, it is highly motivated to use simple RNN cells in lieu of LSTMs. This replacement resulted in $\sim 12\%$ FLOPs reduction in our considered scenario.

5. Conclusion

An attention mechanism was introduced to detect and drop the unnecessary connections in the output layer of non-model-driven equalizers in optical fiber communication systems. For a recently proposed CNN+LSTM architecture, in a 100 GBd 16-QAM $20 \times 100\text{km}$ SMF optical transmission system, it was demonstrated that this technique drops 25% of connections in the output fully-connected layer without affecting the performance, resulting in 15% reduction in the total number of FLOPs. Further, it was shown that the LSTM units in the recurrent layer could be safely replaced with simple RNN cells, bringing an additional 12% FLOPs reduction in the considered system.

Acknowledgement

This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 766115.

References

1. R. M. Butler, C. Hager, H. D. Pfister, G. Liga, and A. Alvarado, "Model-based machine learning for joint digital backpropagation and pmd compensation," *J. Light. Technol.* pp. 1–1 (2020).
2. O. Sidelnikov, A. Redyuk, S. Sygletos, M. Fedoruk, and S. Turitsyn, "Advanced convolutional neural networks for nonlinearity mitigation in long-haul wdm transmission systems," *J. Light. Technol.* **39**, 2397–2406 (2021).
3. C. Catanese, R. Ayassi, E. Pincemin, and Y. Jaouën, "A fully connected neural network approach to mitigate fiber nonlinear effects in 200G DP-16-QAM transmission system," in *Proc. Int. Conf. TRN. Opt. Netw. (ICTON)*, (2020), pp. 1–4.
4. C. Chuang, L. Liu, C. Wei, J. Liu, L. Henrickson, W. Huang, C. Wang, Y. Chen, and J. Chen, "Convolutional neural network based nonlinear classifier for 112-Gbps high speed optical link," in *Proc. Opt. Fiber Commun. Conf. Expo. (OFC)*, (2018), pp. 1–3.
5. S. Deligiannidis, A. Bogris, C. Mesaritakis, and Y. Kopsinis, "Compensation of fiber nonlinearities in digital coherent systems leveraging long short-term memory neural networks," *J. Light. Technol.* **38**, 5991–5999 (2020).
6. X. Liu, Y. Wang, X. Wang, H. Xu, C. Li, and X. Xin, "Bi-directional gated recurrent unit neural network based nonlinear equalizer for coherent optical communication system," *Opt. Express* **29**, 5923–5933 (2021).
7. P. J. Freiredecarvalhosouza, Y. Osadchuk, B. Spinnler, A. Napoli, W. Schairer, N. M. S. da Costa, J. Prilepsy, and S. K. Turitsyn, "Performance versus complexity study of neural network equalizers in coherent optical systems," *J. Light. Technol.* (2021).
8. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, (2017), pp. 5998–6008.

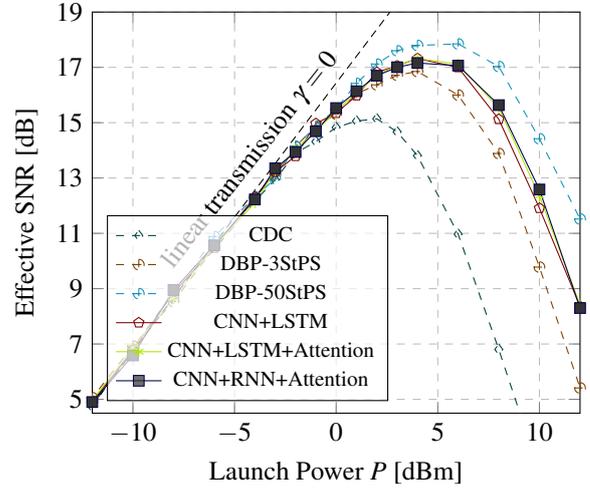


Fig. 3: The effective SNR of the models as a function of the launch power.