Performance and Complexity Evaluation of Recurrent Neural Network Models for Fibre Nonlinear Equalization in Digital Coherent Systems

Stavros Deligiannidis(1), Charis Mesaritakis(2), Adonis Bogris(1),
(1) Department of Informatics and Computer Engineering, University of West Attica, Aghiou Spiridonos, Egaleo, 12243, Athens, Greece, sdeligiannid@uniwa.gr
(2) University of the Aegean, Dept. of Information and Communication Systems Engineering, Palama 2, Karlovassi, 83200-Samos-Greece. cmesar@aegean.gr

Abstract *We investigate the complexity and performance of recurrent neural network models as post-processing units for the compensation of fibre nonlinearities in digital coherent systems carrying polarization multiplexed 16-QAM signals. We show that RNNs are promising nonlinearity compensators especially in dispersion unmanaged systems at reasonable complexity.*

Introduction

Fibre-optic communication industry is struggling to cope with the exponentially increasing capacity demands coming from next generation mobile networks and high bandwidth applications. Advanced modulation formats, nonlinear modulation techniques, space division multiplexing and/or bandwidth extension towards other bands such as O-band are the dominant methods being currently considered for capacity enhancement. The major constraint factor of capacity seems to be the nonlinear Shannon limit attributed to Kerr-induced fibre nonlinearities in their intra-channel and inter-channel form and their interaction with amplified spontaneous emission noise. Lately, there exists an upward trend in the investigation of machine learning techniques either for the mitigation of transmission impairments[1] or for the estimation of quality of transmission (QoT) of modern optical communication systems [2]. Different paradigms based on artificial neural networks (ANNs) [3], convolutional neural networks (CNNs)[4], recurrent neural networks (RNNs)[5] are among the techniques that have been successfully applied mostly in intensity modulation/direct detection systems (IM/DD) and in orthogonal frequency division multiplexing (OFDM) [1]. Very recently, we proposed for the first time the utilization of Long Short-Term Memory (LSTM) network, which is a well-known RNN model [6] for the compensation of fibre nonlinearities in digital coherent systems for multi-channel polarization multiplexed 16-QAM systems. A detailed analysis regarding the effect of LSTM model parameters and channel memory was conducted in order to reveal the limits of LSTM based receiver with respect to performance and complexity in comparison to Digital Back Propagation (DBP). In the present work, we extend the analysis by considering two more RNN models that are in principle less complex

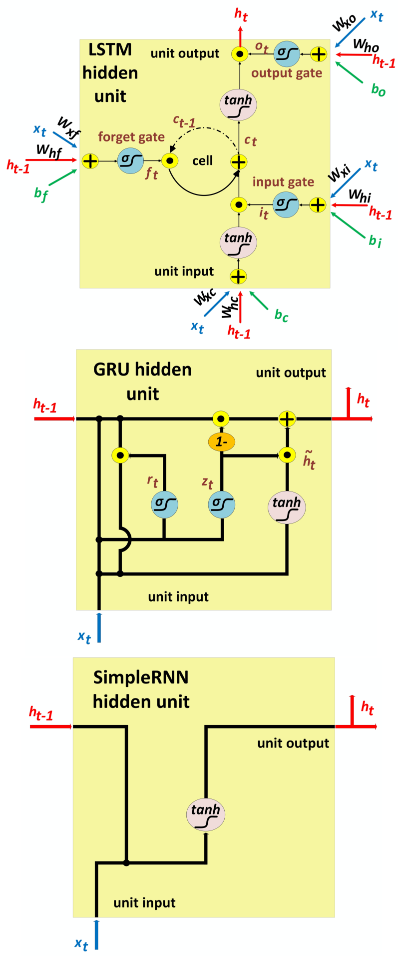

Fig. 1: Conceptual illustration of the LSTM, GRU and Simple RNN units

than LSTM in order to investigate the potential of adopting bidirectional RNN models in next generation digital coherent optical communication systems at low complexity.

System modeling

In this paper we numerically investigate the efficiency of three types of bidirectional Recurrent

Neural Networks (bi-RNN) in compensating fibre nonlinearity in digital coherent optical communication systems: LSTM, GRU, and a Simple RNN. Fig 1 illustrates the RNN units that we use whilst Eq.(1)-(3) calculates the output h_t.

$$
\begin{aligned}
i_t &= \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i)\\
f_t &= \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f)\\
o_t &= \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o)\\
c_t &= f_t * c_{t-1} + i_t * tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c)\\
h_t &= o_t * tanh(c_t) \qquad \text{(1, LSTM)}
\end{aligned}
$$

$$
\begin{aligned}
z_t &= \sigma\,(W_z\,[h_{t-1}, x_t])\\
r_t &= \sigma\,(W_r\,[h_{t-1}, x_t])\\
\tilde{h} &= tanh\,(W\,[r_t * h_{t-1}, x_t])\\
h_t &= (1 - z_t) * h_{t-1} + z_t * \tilde{h} \qquad \text{(2, GRU)}
\end{aligned}
$$

$$
h_t = tanh\,(W\,[h_{t-1}, x_t]) \qquad \text{(3, SimpleRNN)}
$$

The sequential neural model is demonstrated in fig. 2. The input x_t is the distorted symbol sequence which has the following form $x_{t,m}=[x_{t-k},\dots,x_{t-1}, x_t, x_{t+1},\dots,x_{t+k}]$, where m stands for the overall length of the word which is equal to m=2k+1. That is for the symbol at time t we also launch k preceding and k succeeding symbols so as to track intersymbol dependencies. The length of m depends on the foreseen channel memory strictly related to accumulated chromatic dispersion. Each symbol in each window contains four values/features (I and Q for both polarizations) as the input Xx-pol and Xy-pol feeding the Bidirectional RNN layer of L hidden units. In order to calculate bit error rate (BER) we drive the RNN network output to a Fully Connected Layer of 16 units and then to a Softmax layer that carries out the classification among 16 classes/QAM symbols of both polarizations for all the symbols at the output [6]. We train the model using the many to many approach which is beneficial as it takes into account the nonlinear interplay among adjacent bits caused by chromatic dispersion [6]. The RNN models are built, trained and evaluated in Keras with Tensorflow 2.3 GPU backend. We consider

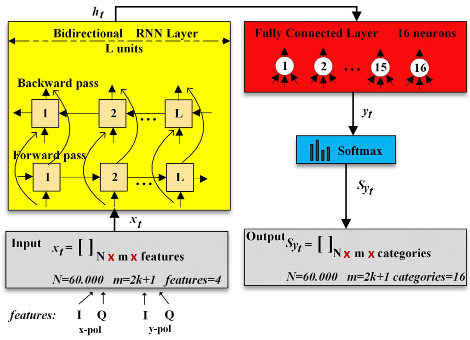

Fig. 2: Bi-RNN model architecture

40.000 symbols for training, 20.000 for validation and 60.000 for testing with unknown data. We consider fibre parameters similar to that of single mode fibres @ 1550 nm. The symbol rate is 25 Gbaud per 16-QAM polarization, channel spacing equals to 50 GHz, amplifier spacing equals to 50 km with noise figure being equal to 5 dB as in [6]. At the receiver we consider ideal carrier synchronization, polarization demultiplexing and dispersion compensation with the use of a frequency domain equalizer (FDE). Fibre propagation was modelled based on Manakov's equations using split-step Fourier method. The RNN processor was applied the central WDM channel that is in principle the most heavily impaired and requires only one sample per symbol.

Results and Discussion

Our analysis considers transmission along 1000km for 10 WDM channels. RNN processors are always placed after the FDE which undertakes dispersion compensation. Fig. 3 shows BER as a function of the number of hidden units (h.u.) that the network uses. Dispersion is -21 ps^2/km (dispersion unmanaged SMF transmission) and we train the different RNNs with a symbol word of 201 symbols that exceeds channel memory. It can be seen that all models perform adequately and equivalently as they improve BER by almost an order of magnitude

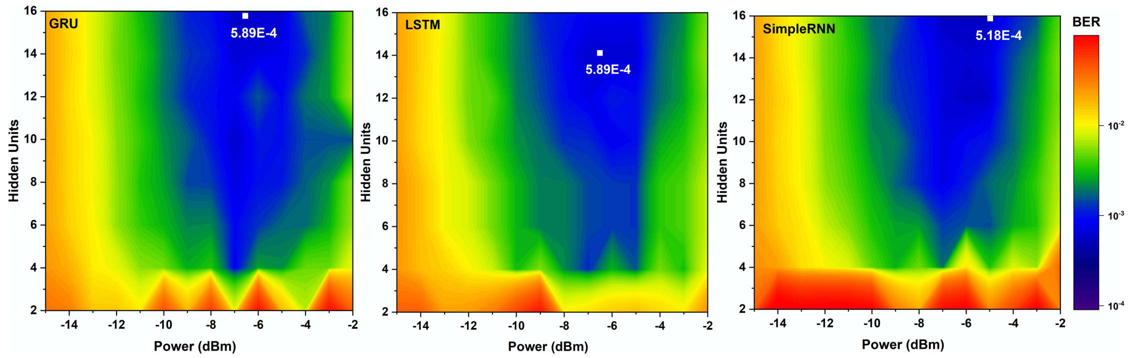

Fig. 3: BER as a function of optical power and number of hidden units for the Bidirectional GRU (a), LSTM (b) and SimpleRNN (c) in 1000km optical transmission with a dispersion of -21 ps^2/km. The system compensated only with the use of FDE exhibited best BER equal to 5x10^{-3}

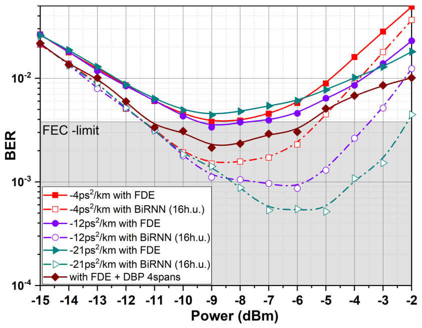

Fig. 4: BER as a function of optical power for dispersion of -4, -12 and -21 ps2/km, with linear equalization (FDE) and with a Simple Bidirectional RNN of 16 hidden units

compared to a transmission system that uses only linear equalization (best BER=4x10^{-3} for FDE not illustrated in the figure). It can be seen that LSTM offers best performance for a minimum of 14 units, whilst GRU and simpleRNN for 16 units. As already stated, memory size increases with the amount of end-to-end accumulated dispersion. In order to study the efficiency of Bi-RNN models for different channel memory scenarios we carried out numerical simulations for -4, -12 and -21 ps^2/km second order dispersion values. Although in real systems a dispersion value different than that of a typical SMF (~-21 ps^2/km) is translated to the use of dispersion compensating modules or other types of fibres that differentiate other critical parameters of the link, here we assume that all the other parameters of the system are not affected in order to identify how the RNN-based equalizer behaves at different channel memories assuming that signal to noise ratio and nonlinearity are kept constant. Apart from FDE equalization, we also conducted numerical simulations for digital back-propagation (DBP), only for -21 ps^2/km dispersion considering single-channel processing as in [6]. In fig. 4, one can see that FDE compensated systems exhibit almost identical behavior with minimum BER close to 4x10^{-3} despite differences in accumulated dispersion. On the contrary, simpleRNN compensation exhibits better BER at larger accumulated dispersion. This behavior was verified for GRU and LSTM as well and is related to the coherence time of the channel which is much longer than the symbol period as dispersion increases. Hence inter-channel effects become very slow and easily tracked by the bi-RNN equalizer [6], [7]. Bi-RNN performs better than DBP as well as the latter equalizes solely intra-channel effects and ignores inter-channel ones [6]. Finally, we investigated the receiver complexity in all cases. In general, the overall model complexity depends on the number of parameters (weights) that each network needs to calculate (see Eq. (1)-(3) and fig. 1), on the number of hidden units and the length of the input

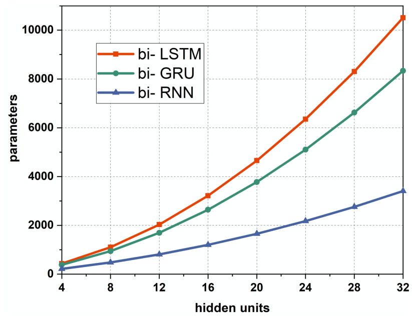

Fig. 5: Number of parameters as a function of the hidden units used in bi-LSTM, GRU and RNN models

word. According to fig. 3, LSTM needs at least 14 units whilst GRU and SimpleRNN need 16 units to achieve optimal BER performance. So, in terms of hidden units the three models are more or less equivalent. Regarding word length, this does not depend on the model architecture, it is strictly related to channel memory. The only parameter differentiating the three models is the complexity of each unit. Based on fig. 1 one can easily calculate the number of parameters for each model based on the expression $biRNN_{param} = 2B[L(L+f)+L]+(2L+1)M$ where *B*=4,3,1 for LSTM, GRU and SimpeRNN respectively, *L* the number of hidden units, *f*=4 the number of input features (see fig. 2) and *M* the number of categories (M=16 in the case of 16-QAM). The results are depicted in fig. 5 as a function of hidden units. It is evident that simpleRNN is by far less complex than LSTM and GRU. Taking into account that LSTM complexity is comparable to DBP's at long distances[6] it can be safely assumed that a simple bi-RNN is even more promising as a nonlinearity compensation scheme performing better than DBP both in terms of BER and complexity.

Conclusions

In this paper we studied three RNN models, (LSTM, GRU and simple RNN), as potential fibre nonlinearity compensators in high capacity digital coherent systems. At distances of 1000 km all models exhibited BER improvement of an order of magnitude compared to systems utilizing exclusively linear equalization. Their efficacy becomes stronger for dispersion unmanaged systems. Finally, among the three models, the simple RNN exhibits the lower complexity without lagging behind in BER performance. This model is a promising candidate for mid-term deployment in nonlinearity impaired coherent transmission systems.

Acknowledgements

This work has been partially funded by University of West Attica.

References

[1] Giacoumidis, et al. “Harnessing machine learning for fiber-induced nonlinearity mitigation in long-haul coherent optical OFDM”, *Future Internet*,11(1), pp. 2-11, 2019, 10.3390/fi11010002.

[2] D. Wang, et al., “Intelligent constellation diagram analyzer using convolutional neural network-based deep learning,” *Opt. Express*, 25, pp. 17150–17166, 2017, 10.1364/OE.25.017150.

[3] C. Hager and H. Pfister, “Nonlinear interference mitigation via deep “neural networks,” *in Proc. OFC*, 2018, paper W3A.4.

[4] P. Li, L. Yi, L. Xue, and W. Hu, “56 Gbps IM/DD PON based on 10G-class optical devices with 29 dB loss budget enabled by machine learning,” *in Proc. Opt. Fiber Commun. Conf.,* 2018, Paper M2B.2

[5] 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. Europ. Conf. on Opt. Comm*., 2018

[6] Deligiannidis et al, “Compensation of Fiber Nonlinearities in Digital Coherent Systems Leveraging Long Short-Term Memory Neural Networks“, *Journal of Lightwave Technology*, Early Access Article, 2020

[7] M. Secondini and E. Forestieri, “On XPM mitigation in WDM fiber-optic systems,” IEEE Photon. Technol. Lett., vol. 26, no. 22, pp. 2252–2255, Nov. 15, 2014