Complex-Valued Neural Network Design for Mitigation of Signal Distortions in Optical Links

Pedro J. Freire, Vladislav Neskornuik, Antonio Napoli, Bernhard Spinnler, Nelson Costa, Ginni Khanna, Emilio Riccardi, Jaroslaw E. Prilepsky, Sergei K. Turitsyn

Abstract—**Nonlinearity compensation is considered as a key enabler to increase channel transmission rates in the installed optical communication systems. Recently, data-driven approaches – motivated by modern machine learning techniques – have been proposed for optical communications in place of traditional model-based counterparts. In particular, the application of neural networks (NN) allows improving the performance of complex modern fiber-optic systems without relying on any a priori knowledge of their specific parameters. In this work, we introduce a novel design of complex-valued NN for optical systems and examine its performance in standard single mode fiber (SSMF) and large effective-area fiber (LEAF) links operating in relatively high nonlinear regime. First, we present a methodology to design a new type of NN based on the assumption that the channel model is more accurate in the nonlinear regime. Second, we implement a Bayesian optimizer to jointly adapt the size of the NN and its number of input taps depending on the different fiber properties and total length. Finally, the proposed NN is numerically and experimentally validated showing an improvement of 1.7 dB in the linear regime, 2.04 dB at the optimal optical power and 2.61 at the max available power on Q-factor when transmitting a WDM 30×200G DP-16QAM signal over a 612 km SSMF legacy link. The results highlight that the NN is able to mitigate not only part of the nonlinear impairments caused by optical fiber propagation but also imperfections resulting from using low-cost legacy transceiver components, such as digital-to-analog converter (DAC) and Mach-Zehnder modulator.**

Index Terms—**Neural network, nonlinear equalizer, channel model, metropolitan links, Bayesian optimizer, coherent detection.**

I. Introduction

MACHINE learning (ML) techniques have been recently proposed as a promising tool to address various challenges in optical communications. In particular, neural network (NN) based algorithms have demonstrated their potential to mitigate nonlinear transmission impairments in optical communication links [1–3]. The knowledge of physical effects, and underlying mathematical models, can be used to build

Pedro J. Freire, Vladislav Neskornuik, Jaroslaw E. Prilepsky and Sergei K. Turitsyn are with Aston Institute of Photonic Technologies, Aston University, United Kingdom, p.freiredecarvalhosouza@aston.ac.uk.

Antonio Napoli and Bernhard Spinnler are with Infinera R&D, Sankt-Martin-Str. 76, 81541, Munich, Germany, anapoli@infinera.com.

Nelson Costa is with Infinera Unipessoal, Lda, Rua da Garagem nº1, 2790-078 Carnaxide, Portugal, ncosta@infinera.com.

Ginni Khanna was with TUM, Munich, Germany.

Emilio Riccardi is with Telecom Italia Mobile, Torino, Italy, emilio.riccardi@telecomitalia.it

the respective NN's architecture for specific transmission systems to improve their performance. Importantly, those techniques showed performance comparable to that achieved by using conventional methods such as digital backpropagation (DBP) [4]. On the other hand, some works proposed NN to mitigate the fiber nonlinearity, taking advantage of the large amount of symbols passing through the line that can be used as a training data source. For the regression task, it is possible to include the delay taps into the NN architecture to account for the channel memory that reflects interaction with the neighboring received symbols [5, 6]. For the classification task, the ML-based techniques have been applied to determine the decision boundaries in QAM modulation formats [6–8].

In this paper, a new design of a complex-valued artificial NN – to be placed at the received (Rx) side – is proposed. Our strategy was numerically tested with transmission over 6×80 km and 12×80 km of standard single mode fiber (SSMF) and large effective area fiber (LEAF) links. The selected fiber types are among the most commonly deployed fibers in legacy optical networks [9]. Our numerical analysis considers a dual polarization (DP) single channel. We selected 64 quadrature amplitude modulation (QAM) format for the proof-of-principle analysis at a symbol rate of 32 GBd and roll-off = 0.06 for pulse shaping realized with a root-raised-cosine (RRC) filter. Here we report a Q-factor improvement, after 6×80 km, of ∼0.4 dB for SSMF and ∼2.1 dB for LEAF. In case of 12×80 km, the gain in Q-factor is ∼0.4 dB for SSMF and ∼1.25 dB for LEAF. The performance assessment of the proposed algorithm has been carried out against conventional electronic dispersion compensation and phase/amplitude normalization of the received signal. In addition, when compared to the NN proposed in [5, 10, 11] – with 2 dense layers and 192 neurons – an improvement of ∼0.4 dB (SSMF) and ∼1.7 dB (LEAF) after 6 spans; and ∼0.4 dB (SSMF) and ∼1 dB (LEAF) after 12 spans, was obtained. Finally, we experimentally validated the proposed technique by applying it to a 30×200G wavelength division multiplexing (WDM) system, with each channel employing DP-16QAM signal, transmitted over a 612 km of SSMF legacy link [12, 13]. In this scenario, the Q-factor improvement of 1.7 dB in the linear regime, 2.04 dB at the optimal optical power and 2.61 dB at the max available power are demonstrated compared to the use of electronic dispersion compensation combined with phase/amplitude normalization.

The remainder of the paper is organized as follows. In Sec. II, we review the theoretical background of our channel model and introduce the NN design. Sec. III presents
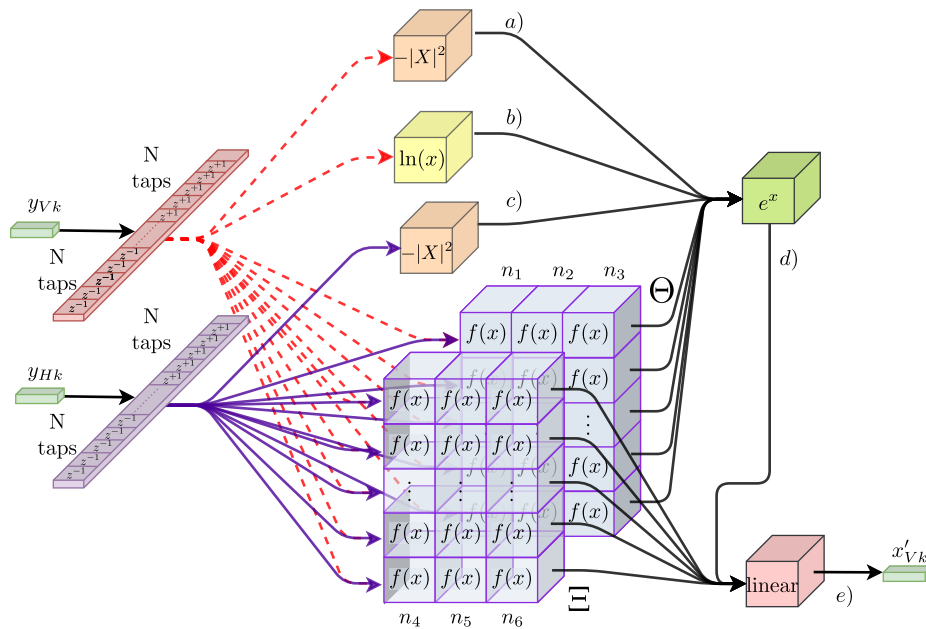
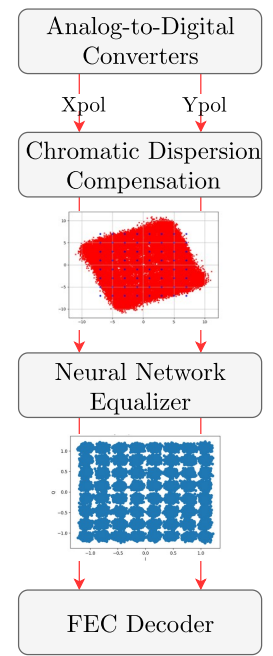Fig. 1: The architecture of the proposed neural network.

Fig. 2: Receiver setup in the simulator.

the experimental setup and results, including the comparison between the proposed NN and DBP. Sec. IV analyses the complexity of the proposed NN compared to the standard DBP. The last section concludes the paper.

II. NEURAL NETWORK DESIGN

A. The Channel Model

The averaged evolution of the slowly varying complex-valued envelopes of the electric field in an optical fiber is described by the pass-averaged Manakov equation [14]:

$$\frac{\partial u_{H/V}(t,z)}{\partial z} + \frac{j\beta_2}{2}\frac{\partial^2 u_{H/V}(t,z)}{\partial t^2} =$$
$$= j\frac{8\tilde{\gamma}}{9}\left[|u_H(t,z)|^2 + |u_V(t,z)|^2\right]u_{H/V}(t,z), \quad (1)$$

where $u_{H/V}(t,z)$ are the normalized optical fields of horizontal (H) and vertical (V) polarization, respectively, β_2 is the group velocity dispersion, $\tilde{\gamma} = \gamma e^{-\alpha z}$ is the effective averaged nonlinearity coefficient, that includes the effective length scale $L_{eff} = (1 - e^{-\alpha L})/\alpha$ emerging due to averaging over periodic loss and gain, γ is the fiber nonlinear coefficient, L is the span length and α is the fiber loss coefficient. In case $\beta_2 = 0$, one can show that the analytical solution of Eq. (1), at the end of the transmission link, can be expressed for H-polarization in terms of the transmitted x_{H_k} and received y_{H_k} soft symbols as [15]:

$$x_{H_k} = y_{H_k} e^{-8j/9\gamma L_{eff} N_s \left[|y_{H_k}|^2 + |y_{V_k}|^2\right]}, \quad (2)$$

where N_s is the number of spans.

We suggest the application of Eq. (2) to recover digitally the transmitted symbols x_k out of the received signal when the chromatic dispersion (CD) was already compensated for.

For the Manakov equation (Eq. (1)), perturbative methods describing the field evolution in the highly nonlinear regime [16–19] have been developed under some limiting assumptions. Conversely, we suggest adding trainable general nonlinearity functions to the memory-less solution provided by Eq. (2): additions Θ_1, Θ_2 to the exponential power, responsible for phase distortion, and additions Ξ_1, Ξ_2, responsible for amplitude distortion. The additional degrees of freedom, introduced by these functions, will let the equalizer learn the features of the system behavior, which are not covered by the original Eq. (2). Particularly, we introduce the following model for the dual-polarization case:

$$x'_{H_k} = c_1 Y_{H_k} e^{\left[-c_2|Y_{H_k}|^2 - c_3|Y_{V_k}|^2 + \Theta_1\right]} + \Xi_1, \quad (3)$$

$$x'_{V_k} = c_4 Y_{V_k} e^{\left[-c_5|Y_{H_k}|^2 - c_6|Y_{V_k}|^2 + \Theta_2\right]} + \Xi_2, \quad (4)$$

where x'_{V_k/H_k} is the k-th recovered symbol in each polarization, $Y_{(V/H)_k}$ is the vector containing the sequence of received symbols $[y_{(V/H)_{k-N}}, ..., y_{(V/H)_k}, ..., y_{(V/H)_{k+N}}]$, $2N+1$, is the size of the memory in the model, c_k with $k = 1, ..., 5$ are complex vectors of size $2N+1$ each, and $\Theta_{1/2}(Y_{H_k}, Y_{V_k})$ and $\Xi_{1/2}(Y_{H_k}, Y_{V_k})$ are the two adaptive nonlinear functions mentioned above, which depend on the vectors Y_{H_k} and Y_{V_k}. We would like to stress that this model is used for the design of the NN, and not for solving the propagation equations.

The point-wise multiplication of the sequence of received symbols $Y_{(V/H)_k}$ and the weight coefficients c_k define the memory introduced into the nonlinear distortion Eq. (2) by CD and transceiver impairments. Furthermore, nonlinear functions $\Theta_{1/2}$ and $\Xi_{1/2}$ reflect the residual nonlinear distortion introduced by the optical fiber and the transmitter components. In other words, the proposed NN is designed to compensate at

the receiver the deterministic intra-channel nonlinear distortion originating in both optical fiber and transceiver equipment.

The model hyper-parameters depend on the total fiber length and on the relationship between β_2 and γ. Consequently, during the numerical study, we considered links based on two different types of optical fibers: large effective area fiber (LEAF) and standard single-mode fiber (SSMF). The considered link lengths of 6×80km (480 km) and 12×80km (960 km). The characteristics of the SSMF and LEAF are reported in Table I.

TABLE I: Considered fiber parameters.

Fiber	α [dB/km]	β_2 [ps/(nm km)]	γ [1/W km]
SSMF	0.21	16.8	1.14
LEAF	0.225	4.2	1.3

To describe the scope of the proposed ML algorithm, it is important to show qualitatively where strong nonlinear regime starts in both fibers. The definition of the different regimes can be carried out by taking into account the relationship between chromatic dispersion (L_D) and Kerr nonlinearity (L_{NLI}) effective lengths, described as [15]:

$$L_D = \frac{T_0^2}{|\beta_2|}, \quad L_{NLI} = \frac{1}{\gamma P}, \quad (5)$$

where T_0 is the pulse length and P is the launch optical power. Fig. 3 shows the dependency of those two lengths on the launch power considering a symbol rate equal to 32 GBd. In the "hot" region, the nonlinearity is the dominant signal distortion source whereas, when we are in the "cold" region, the linear distortions are dominant. As expected, LEAF has a stronger nonlinear dependence on the launch power than SSMF, resulting from its chromatic dispersion parameter which is four times smaller than the one of SSMF.
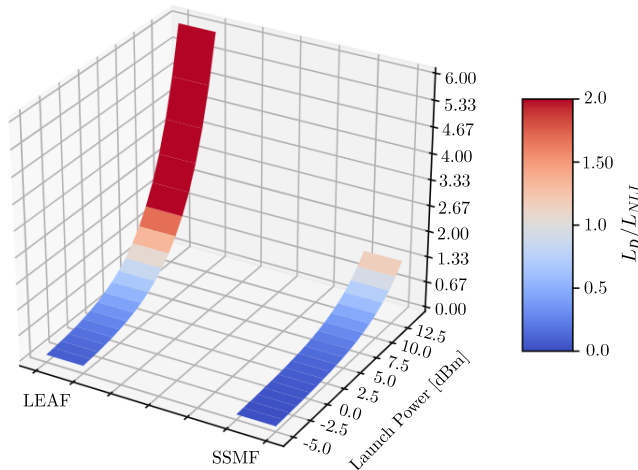

Fig. 3: Qualitative estimation of the relative strengths of chromatic dispersion and Kerr nonlinearity [Eq. (5)] for LEAF and SSMF at different power levels.

B. Neural Network Description

We propose to build the NN topology based on the loose signal evolution model Eqs. (3), (4). Contrary to the conven-

tional real-valued NN architectures [21], in our design we implement complex-valued weights, activation functions, and input symbols as suggested in [22, 23] to reflect the complex-valued laws describing the signal propagation. The proposed NN topology is schematically depicted in Fig. 1.

The NN depicted in Fig. 1 predicts the k-th symbol transmitted in V polarization x_{V_k}. The proposed NN can obtain the transmitted symbols from the Horizontal (H) polarization x_{V_k} by swapping V, y_{V_k}, and H, y_{H_k}, polarization symbol sequences at the NN input. The architecture of the suggested NN is described by the following equations: each of the equations defines the exits of the neurons a), b), c), d) and e), as referred to in Fig. 1 (the corresponding exit labels are also marked in the figure with the sub-index "a", "b", etc.):

$$Neuron_{a,out} = -\left| \sum_{i=k-N}^{k+N} y_{V_i} a_i \right|^2, \quad (6)$$

$$Neuron_{b,out} = \ln\left(\sum_{i=k-N}^{k+N} y_{V_i} b_i \right), \quad (7)$$

$$Neuron_{c,out} = -\left| \sum_{i=k-N}^{k+N} y_{H_i} c_i \right|^2, \quad (8)$$

$$Neuron_{d,out} = \exp\left(Neuron_{a,out} d_1 + \right.$$
$$\left. + Neuron_{b,out} + Neuron_{c,out} d_2 + \sum_{i=0}^{n_3-1} \Theta_i d_{i+3} \right), \quad (9)$$

$$Neuron_{e,out} = Neuron_{d,out} + \sum_{i=0}^{n_6-1} \Xi_i e_{i+1}. \quad (10)$$

Eqs. (6)–(10) can be combined in a single pipeline connecting NN output x'_{V_k} to its inputs y_{V_i}, y_{H_i}. This form is convenient to show the similarities between the NN architecture and Eq. (4) (the similar expression can be written down for another polarization component):

$$x'_{V_k} = \left(\sum_{i=k-N}^{k+N} y_{V_i} b_i \right) \exp\left(-d_1 \left| \sum_{i=k-N}^{k+N} y_{V_i} a_i \right|^2 - \right.$$
$$\left. -d_2 \left| \sum_{i=k-N}^{k+N} y_{H_i} c_i \right|^2 + \sum_{i=0}^{n_3-1} \Theta_i d_{i+3} \right) + \sum_{i=0}^{n_6-1} \Xi_i e_{i+1}. \quad (11)$$

Here a_i, b_i, c_i, d_i and e_i are complex tensor weights constituting the parameters of the NN. After a random initialization, these weights are learned from propagated data by means of stochastic gradient descent optimization performed by Adam optimizer [24]. The optimizer minimizes the mean-squared error (MSE) loss function L, describing the difference between the predicted symbols x'_i and the actually transmitted ("desired") ones x_i:

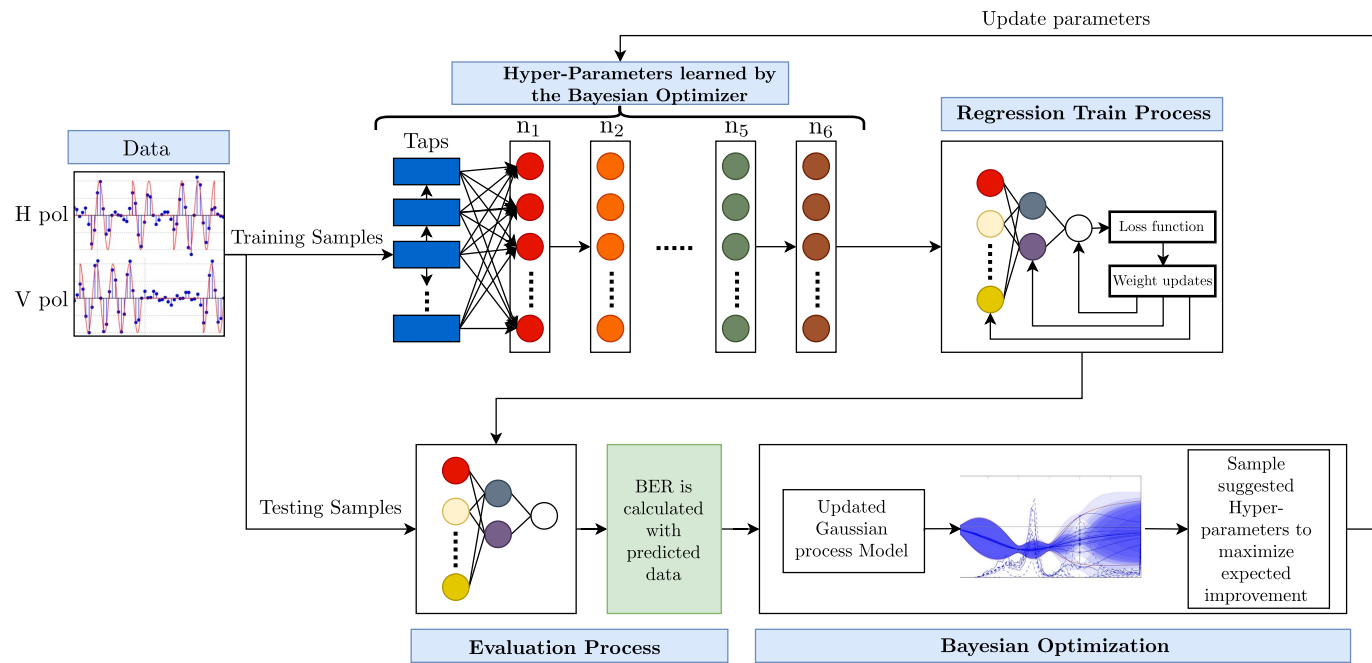$$L = \frac{1}{M} \sum_{i=0}^{M-1} |x'_i - x_i|^2, \quad (12)$$

Fig. 4: The principal scheme of the Bayesian optimizer [20] used to find hyper-parameters of NNs approximating nonlinear functions Θ and Ξ, Eqs. (3–4). The detailed explanation of the schematic is given in the main text.

where M is the batch size.

We used custom activation functions $-|x|^2$, $\ln(x)$ and e^x, to reflect in the architecture our estimation of the distortion form, given by Eqs. (3) (4). Furthermore, Θ and Ξ, as mentioned before, are the analytically unknown nonlinear functions taking into account the deviation of the considered link from the ideal zero-dispersion case provided by Eq. (2); these functions are modelled via the NNs. Particularly, a three layer perceptron was used for each function, given the fact that the multi-layer perceptron is a universal function approximator [21].

The NN that learns Θ has n_1 neurons in the first layer, n_2 neurons in the second, and n_3 neurons in the third one. In the same way, the NN that learns Ξ has n_4 neurons in the first layer, n_5 neurons in the second one, and, finally, n_6 neurons in the third one. For both NNs, all neurons have complex-valued activation function $f(x)$

$$f(x = x_r + jx_i) = \frac{e^{2x_r} - 1}{e^{2x_r} + 1} + j\frac{e^{2x_i} - 1}{e^{2x_i} + 1}. \qquad (13)$$

C. Hyper-parameters optimization

Since the functions approximated by Θ and Ξ may have different nonlinearity type and memory size at different power levels depending on the leading propagation effect: the Kerr nonlinearity or the CD (Fig. 3), a Bayesian optimization algorithm [20] was implemented to derive the optima values for the hyper-parameters of Θ and Ξ NNs for each studied communication system: the number of input taps N; neuron numbers n_i in each layer of each NN.

Fig. 4 illustrates the hyper-parameters optimization process. During the optimization, we consider a signal optical launch power 3 dB higher than the CDC optimal level for every test-case in order to have a stronger response of the system performance to the hyper-parameters values. The learning rate and the batch size are fixed to 0.001 and 1000, respectively. In the beginning, the number of input taps N and the numbers of neurons in all layers n_1–n_6 are initialized as 20 and 100, respectively. The optimization cycle starts with training the NN via backpropagation with a fixed set of hyper-parameters for 5000 epochs, i.e. full passes through the whole dataset, over the training dataset containing 2^{18} restored symbols. After each training epoch, we calculate the BER obtained by the whole NN (Fig. 2) on the independently generated testing dataset containing 2^{16} symbols. The best BER obtained during the training is recorded. After training, the best BER is fed as the optimization target to the Bayesian optimizer [20]. The optimizer assumes that the conditional distribution of BERs – given the particular values of hyper-parameters – is a Gaussian process. Having received the new measured BER value, the optimizer updates the process model and generates a new set of hyper-parameters to be tested. Having carried out 20 Bayesian optimizer cycles, we selected the set of hyper-parameters with the lowest BER.

Fig. 5 shows an example of the benefits provided by the Bayesian optimizer method. This figure displays the optimization for the case of transmission of a DP-64QAM single channel over LEAF (6×80km) at a launch power of 3 dBm (that will be discussed in detail in the next section). As can be seen, this method reached the best set of hyper-parameters – and thus best Q-factor – on the 8th optimization cycle. The best configurations obtained for each considered scenario in this paper, are shown in Table II.

Noticeable, the optimizer suggested several sets with similar performance levels. Among these, one could select the set

with the best balance between the performance and numerical complexity. However, in this study, we decided to maximize optical performance and, hence, always chose the best performing point. Seeking the optimal balance between optical performance gain and cost is left beyond the scope of this work. Besides fine-tuning the Bayesian optimization, other advanced ML techniques such as pruning [25] and efficient scaling [26] can be used to achieve lower computational complexity levels without reducing the NN performance.
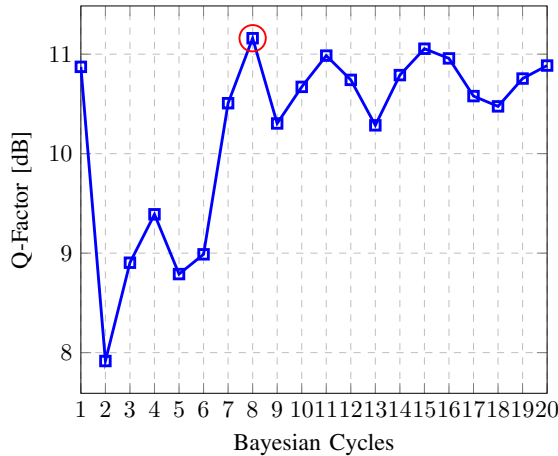

Fig. 5: Example of the iterative optimization for the Bayesian method in the case of DP-64 QAM single channel LEAF (6×80km) at 3 dBm.

To eliminate some possible dataset periodicity, which could cause overfitting and overestimation [27], the training dataset objects were randomly shuffled at the beginning of every epoch. The numerically generated training and testing datasets were measured to achieve a normalized cross correlation below 0.6%, to ensure their independence. The code developed to obtain the results depicted in this paper is provided online on the platforms Github and Zenodo [28].

TABLE II: Hyper-parameters of neural networks approximating Θ and Ξ functions from Eq. 11 learned by means of Bayesian optimization.

Scenario	N (taps)	n_1	n_2	n_3
		n_4	n_5	n_6
LEAF (Numerical – 6×80km)	11	54	46	40
		4	38	50
LEAF (Numerical – 12×80km)	18	62	34	56
		55	18	40
SSMF (Numerical – 6×80km)	20	51	54	53
		19	22	56
SSMF (Numerical – 12×80km)	26	54	47	50
		41	38	61
SSMF (Trial data – 8×76+4km)	20	90	53	88
		56	60	64

III. ML-Based Nonlinear Equalizer

A. Numerical study

1) Numerical setup: The system setup considered in the numerical study is illustrated in Fig. 6. First, a random bit stream is generated for each polarization. Afterwards, Gray coded 32

GBd 64QAM symbols are mapped from the bit sequences. The resulting signals are upsampled with a sampling rate equal to four times the symbol rate and shaped using an RRC with roll-off = 0.06. Finally, a polarization beam combiner (PBC) is used to combine both polarizations to the single-channel dual-polarized signal which is then fed into the optical fiber.
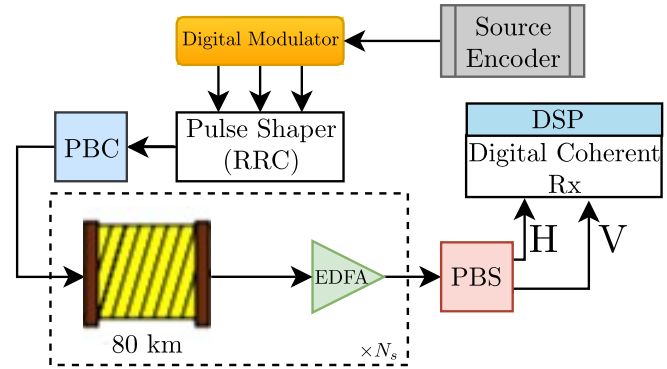

Fig. 6: The system setup of the transmission loop considered in our numerical simulations.

To simulate the (forward) propagation of the signal in the fiber, we used a symmetrized split-step Fourier method [15] solving the ordinary Manakov equation [29]. We considered 6×80 km and 12×80 km systems consisting of LEAF or SSMF. The considered fiber parameters are given in Table I. At the end of each span, the fiber losses are fully compensated by a lumped Erbium-doped fiber amplifier (EDFA). Additive White Gaussian Noise (AWGN) representing the amplified spontaneous emission (ASE) noise is also added. The EDFA noise figure is $NF = 4.5$ dB. At the receiver side, both H and V polarizations are first separated using a polarization beam splitter (PBS) and then fed to a digital coherent receiver where the signal passes through a matched filter and is downsampled to the symbol rate. The resulting signal is processed offline using the DSP outlined in Fig. 2 to recover the transmitted symbols. Since we considered no transceiver impairments in numerical simulations, the receiver DSP consisted of: an ideal CDC compensator followed by a single-tap adaptive filter performing amplitude and phase normalization of the received symbol stream to the desired one (CDC + Norm), and, finally, the tested nonlinearity mitigation algorithm.

2) Numerical Results: The proposed NN was benchmarked against the classic digital backpropagation (DBP) [30] with 2 samples/symbol and 2 and 3 steps per span (StPS), along with the NN-based equalizer proposed in [11]. This reference NN was implemented exactly as stated in [11]: it had two layers with 192 neurons and processed the same number of input taps as our proposed NN did. The sequences of symbols from both polarizations were simultaneously fed as input to the reference NN from Ref. [11]. In this case, however, real and imaginary parts are fed as different parts of the input vector [31], since this is a real-value neural network. All the algorithms used in this paper are static algorithms. As the performance metric,
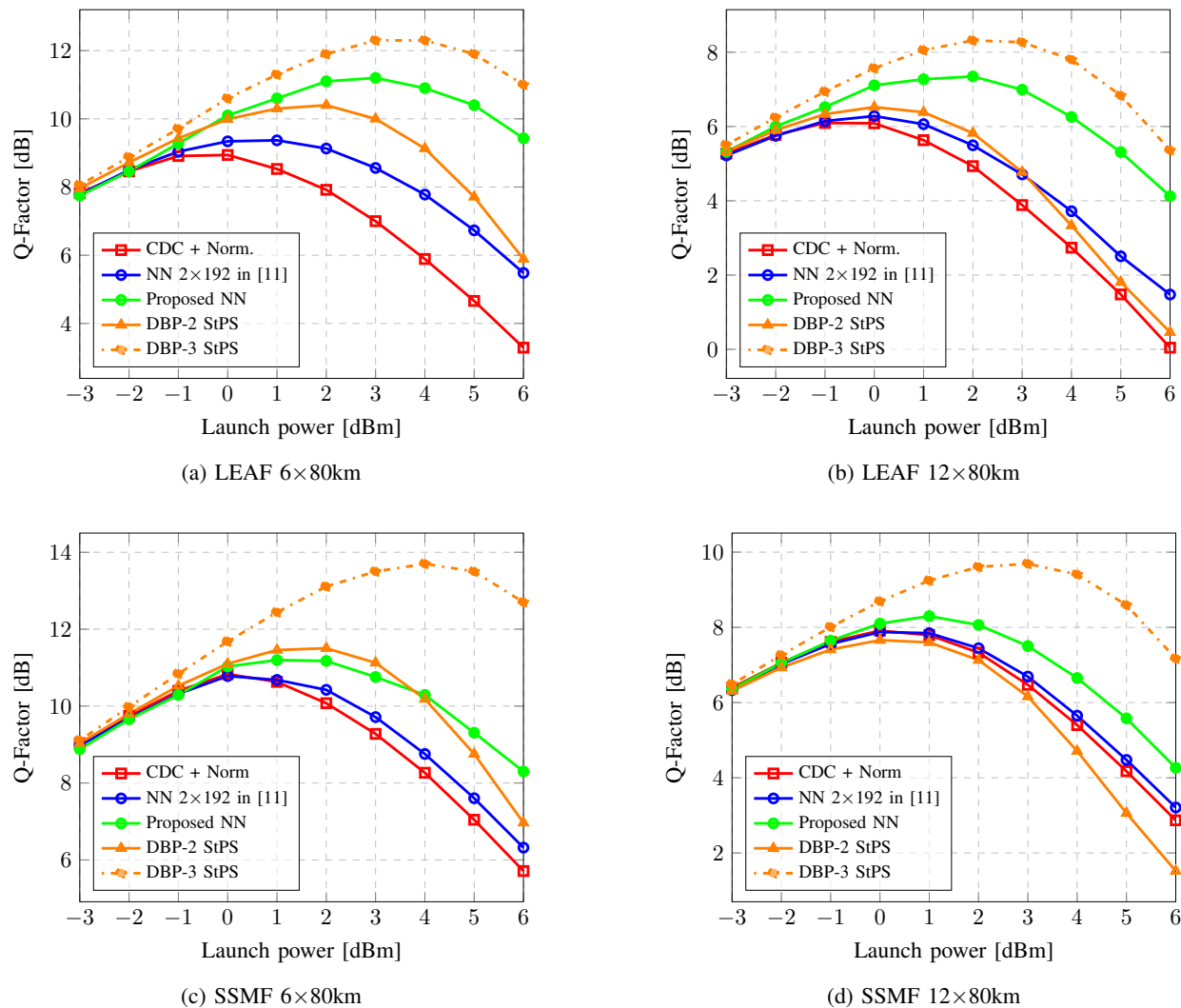
(a) LEAF 6×80km

(b) LEAF 12×80km

(c) SSMF 6×80km

(d) SSMF 12×80km

Fig. 7: Signal equalization performance of the proposed NN, benchmarked against digital backpropagation (DBP) and the reference neural network (2×192) [11] for various numerically studied fiber-optic links. In all testcases DP-64QAM 32 GBd single-channel optical signal was considered.

we used the Q-factor expressed through BER as follows:

$$Q = 20 \log_{10} \left[\sqrt{2} \ \text{erfc}^{-1}(2BER) \right], \quad (14)$$

where erfc^{-1} is the inverse complementary error function. All main results of this subsection are summarized on Table III.

Fig. 7 shows the comparison of Q-factors obtained by employing the considered algorithms for nonlinearity mitigation in the numerically studied testcases. For LEAF, Fig. 7a illustrates the 6×80 km case and Fig. 7b the 12×80 km case. In the first case, the proposed NN improved the Q-factor

by up to ∼2.1 dB when compared with CDC + Norm and up to ∼1.7 dB when compared with the reference NN from Ref. [11], at their respective optimal optical launch powers. The proposed NN also outperforms the DBP with 2 StPS performance by 0.7 dB. Evidently, it is expected that the DBP with higher complexity (with more steps or sampling points) outperforms our NN. Indeed, the DBP with 3 StPS offers a better performance. For the second scenario, a similar improvement trend was observed, see Fig. 7b. At the optimal launch powers, the proposed NN increased the Q-factor by up

TABLE III: Summary of all results obtained by CDC + Norm, NN 2×192 neurons[11], DBP and the proposed NN. The optimal launch power, in dBm, and the peak Q-factor, in dB, are indicated for all considered testcases.

Fiber	Q / Best P [CDC + Norm]	Q / Best P [NN in [11]]	Q / Best P [Proposed]	Q / Best P [DBP-2 StPS]	Q / Best P [DBP-3 StPS]
LEAF 6×80km	8.94 dB / 0 dBm	9.37 dB / 1 dBm	11.16 dB / 3 dBm	10.4 dB / 2 dBm	12.3 dB / 4 dBm
LEAF 12×80km	6.09 dB / 0 dBm	6.28 dB / 0 dBm	7.34 dB / 2 dBm	6.52 dB / 0 dBm	8.31 dB / 3 dBm
SSMF 6×80km	10.83 dB / 0 dBm	10.81 dB / 0 dBm	11.19 dB / 2 dBm	11.50 dB / 2 dBm	13.7 dB/ 4 dBm
SSMF 12×80km	7.91 dB / 0 dBm	7.98 dB / 0 dBm	8.29 dB / 1 dBm	7.66 dB/ 0 dBm	9.68 dB / 3 dBm
SSMF (Trial data)	6.81 dB / 3 dBm	7.58 dB / 3 dBm	8.85 dB / 4 dBm	-	-

to ~1.25 dB when compared to the CDC + Norm and by ~1 dB when compared with the reference NN from [11]. Also, the proposed NN outperforms the DBP 2 StPS performance by 0.8 dB.

For the SSMF system, Fig. 7c shows the result for a 6×80 km case, where we have observed an improvement of the Q-factor of up to ~0.4 dB when using the proposed NN with respect to both the CDC + Norm and the reference NN in [11]. In the 12×80 km case, (Fig. 7d) the proposed NN led to the Q-factor improvement of up to ~0.4 dB with respect to both the CDC + Norm and the NN from [11]. In this case, the DBP with 2 StPS was not able to outperform the simple CDC compensator followed by a single-tap adaptive filter performing the amplitude and phase normalization. As predicted in Sec. II-A, using a fiber with 4 times higher chromatic dispersion (β_2) impacts the performance of the proposed NN, since the proposed NN shows better performance when the nonlinearity produces the dominant contribution, as it is expected due to the chosen structure.

In all the numerically considered testcases, the optimal launch power for the proposed NN was higher than for CDC + Norm and the reference NN in [11]. This shows that the performance gains of the proposed NN are obtained by improving nonlinearity mitigation.

B. Experimental Study

1) Experimental setup: In the experimental study, the data were obtained by using the setup of the field trial described in [12] and [13]. The transmission link consisted of 8×76 km G.652 SSMF spans along with two 2 km long SSMF connectors deployed between Torino and Chivasso in Italy, leading to a total length of 612 km. The transmitted spectrum consisted of 15×200G DP-16QAM neighbor channels on the right- and left-hand sides of the channel under test, resulting in a 31 channel WDM transmission system in the 37.5 GHz grid (33.01 GBd). At the receiver side, the WDM signal was first amplified and then converted into the electrical domain using a coherent front-end. An ADC with 18 GHz bandwidth operating at 80 GSample/s was used to capture sets of 5×10^5 samples per tributary. Digital signal processing is then applied at the receiver. Firstly, bulk accumulated chromatic dispersion was compensated using a frequency domain equalizer, followed by removal of carrier frequency offset. Constant-amplitude-zero-autocorrelation based training sequence was then located in the received frames, and the equalizer transfer function was estimated from it. After equalization, the two polarizations were de-multiplexed, and time corrected. Carrier phase estimation was then achieved through the aid of pilot symbols. Finally, two independent downsampled measurement shots, processed by DSP, were used as training and testing data for the offline testing of the considered nonlinearity mitigation algorithms.

2) Experimental result: On the contrary of the numerical studies, in the experiment we were not able to apply the DBP technique because the carrier had no appropriate information about the system parameters installed: they were laid in the field and are poorly accessible. So only the downsampled

data was available for our tests. Therefore, we will compare the proposed NN with the traditional CDC + norm and the reference NN from [11].

Fig. 8 shows the experimental results obtained in the field trial and processed by the CDC, the NN from [11], and, eventually, the proposed NN. First, the optimal launch power obtained using our new NN equalizer increased from 3 to 4 dBm, and the Q-factor improved by up to ~2 dB and ~1.2 dB when compared, at the best power level, with the CDC+Norm and the reference NN from [11], respectively. Moreover, we have highlighted the different gains on the linear and nonlinear regions which indicates that our NN is compensating for nonlinear effects since the performance improvement increases with the launch power. At the lowest measured launch power (0 dBm), we observed a gain of 1.7 dB when comparing with the traditional CDC + norm. We assume that this gain in the linear regime comes from the ability of the neural network, by using the specially designed nonlinear parts $\Xi_{1/2}$, to mitigate the impact of additional limitations such as, e.g., the impact of low-resolution digital-to-analog and analog-to-digital converters, the driver amplifier, and the dual polarization Mach-Zehnder modulator, that were not entirely tackled by the current DSP. The imperfections of the transmitter – e.g., the $S_{21}(f)$ – and of the DAC low-resolution – showed as ENOB versus frequency – are reported in Fig. 3 of [32]. On top of these linear regime gains, we found that, at the optimum power levels, (3 dBm for CDC+Norm and 4 dBm for the proposed algorithm) our NN provides additional 0.34 dB of Q-factor gain, with the total gain being 2.04 dB. For the proposed NN, the optimum launch power increases in the experiment by ~1 dB, which is in agreement with the numerical analysis. Notably, the reference NN from [11] did not improve the optical performance when increasing the optimal launch power.
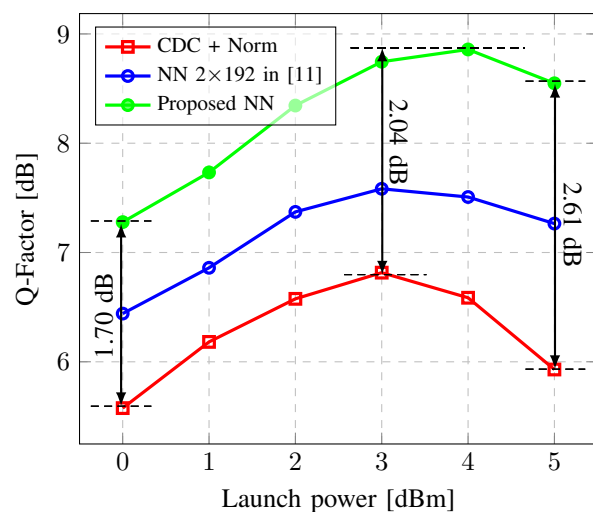

Fig. 8: Comparison of the equalisation performance obtained by the considered algorithms in the field trial of 612 km SSMF legacy link.

IV. Computational Complexity Analysis

The goal of this section is to discuss the computation complexity required by the proposed method and compare it with the traditional DBP method and the NN densely-connected architecture proposed in [11]. The comparison is performed in terms of the total number of real multiplications per transmitted bit required by each of the two nonlinear compensation schemes. To calculate the complexity of the DBP strategy, we followed the approaches given in [4, 33].

We start our analysis with the DBP-based receiver. As mentioned before, the basic implementation of the DBP algorithm is considered, where each propagation step comprises a linear part for dispersion compensation followed by a nonlinear phase cancellation stage. The linear part is achieved with a zero-forcing equalizer by transforming the signal in the frequency domain and multiplying with the inverse dispersion transfer function of the propagation section. The complexity of the method is [33]:

$$C_{\text{DBP}} = 4N_{span}N_{step}\left(\frac{n\,N_{\text{FFT}}[\log_2(N_{\text{FFT}}) + 1]}{(N_{\text{FFT}} - N_D + 1)\log_2(Mod)} + n\right), \tag{15}$$

where N_{step} is the number of steps per span used, N_{FFT} is the FFT size, Mod is the order of the modulation format, n is the oversampling ratio, and $N_D = \tau_D/T$, where τ_D corresponds to the dispersive channel impulse response and T is the symbol duration.

In the case of the NN, the complexity is evaluated for the most critical mode of operation when the trained NN is working as the part of the data communication chain, i.e., on the inference stage. In this case, the forward propagation step complexity is given by:

$$C_{\text{NN}} = \frac{4}{\log_2(Mod)}\Big[(2N + 1)(2n_1 + 2n_4 + 3) + \\ + (n_1n_2 + n_2n_3 + n_4n_5 + n_5n_6) + (n_3 + 3) + \\ + (n_6 + 1) + C_{\text{CDC}}\Big]. \tag{16}$$

In more details, we assume the neurons of first layer require $(2N + 1)(2n_1 + 2n_4 + 3)$ complex multiplications. The dense layers of Θ and Ξ contribute with $n_1n_2 + n_2n_3 + n_4n_5 + n_5n_6$ multiplications. Moreover, the neurons d) and e) add $n_3 + 3$ and $n_6 + 1$ complex multiplications, respectively. The found amount of complex multiplications was multiplied by 4 to convert from complex to real multiplications and divided by $\log_2(Mod)$ to convert from multiplication per symbol to those per bit. Finally, we have added the number of complex multiplications per symbol of the CDC step (C_{CDC}) that was taken equivalent to one linear step of the DBP algorithm, with the step covering the whole transmission link.

The last complexity calculation to be presented here is from the NN given in [11] that is a real-value multilayer perceptron with two layers and 192 neurons in each of them. Eq. 17 displays how the complexity of this NN was calculated. Firstly, the input layer feeds separately the real and imaginary parts of the complex symbol for both polarizations. This step adds $2(4N + 2)n_{1a}$ operations, considering that N is the number of taps and n_{1a} is the number of neurons of the first layers. After

that, we accounted for the multiplication of the hidden layer, $n_{1a}n_{2a}$, where n_{2a} is the number of neurons in the second layer, and the number of multiplications for the output layer is $2n_{2a}$, which is multiplied by 2 because we have two outputs (real and imaginary) to recover the symbol. Finally, similarly to the addition of the dispersion compensation complexity to the overall value of our proposed NN, Eq. (16), we add this part C_{CDC} to the complexity of the NN from [11]. We have divided the final result by $\log_2(Mod)$ to convert from multiplication per symbol to per bit. The result reads as

$$C_{\text{ref. [11]}} = \frac{1}{\log_2(Mod)}\Big[2(4N + 2)n_{1a} + n_{1a}n_{2a} + \\ 2n_{2a} + 4C_{\text{CDC}}\Big]. \tag{17}$$
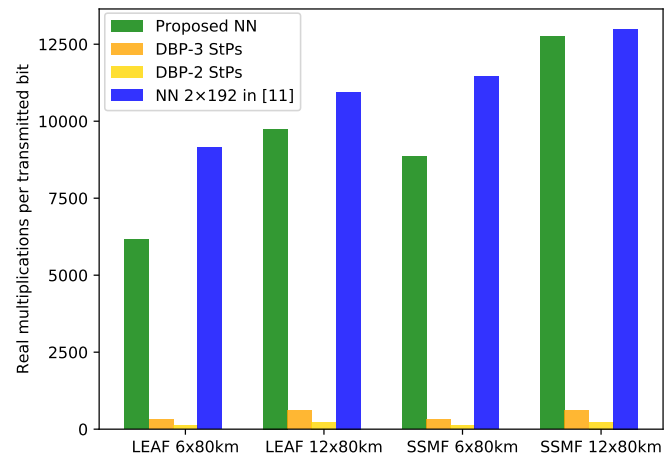

Fig. 9: Comparison of computational complexity in terms of real multiplications per transmitted bit among the proposed NN, NN in [11] and DBP with 2 and 3 steps-per-span (StPS) and 2 samples-per-symbol.

Fig. 9 presents the complexity for all the suggested NNs shown in Table II, the reference NN from [11], and for the DBP with 2 and 3 StPS. First of all, we can see that the method [11] that uses the NN in a form of a black box not only gives somewhat worse results, but also requires more real multiplications than our proposed NN; this notion refers to all scenarios. Secondly, one can see that the proposed NN leads to much higher complexity than the DBP. Nonetheless, the proposed method has an important advantage over DBP. In fact, NN is by nature channel-agnostic, and therefore it does not require a priori exact knowledge of the channel parameters, like in DBP, since it learns them from training dataset. Also, as presented in [34], DBP performance might be considerably degraded in case of imperfect link knowledge[1]. Therefore, the implementation of proposed NN – for the systems where no channel parameters are known in advance – represent a valuable option.

To conclude this section, we investigated the relationship between the performance and computational complexity which

---

[1]In [35] the performance of a DBP with wrong values for the nonlinear coefficient γ was compared against the alternative proposed by the authors to automatically derive the correct value of this parameter.

TABLE IV: Topologies of the proposed neural network learned by means of Bayesian optimization and used in the Fig. 10.

Topologies	N (taps)	n1	n2	n3
		n4	n5	n6
T.1	10	3	4	3
		3	2	2
T.2	10	4	3	2
		3	4	4
T.3	11	6	8	7
		4	2	8
T.4	11	16	16	16
		16	16	16
T.5	12	15	25	12
		23	31	20
T.6	14	31	12	34
		2	41	38
T.7	11	32	32	32
		32	32	32
T.8	11	54	46	40
		4	38	50

can be achieved with the proposed NN. We have considered here the case of DP-64QAM single channel LEAF (6×80km) at 3 dBm, the one with the best gains in numerical study. Fig. 10 relates the Q-factor gains and the complexity for eight different topologies of the proposed NN. The topologies differ in the number of neurons and the input sizes of $\Theta_{1/2}$ and $\Xi_{1/2}$ – the parameters optimized by the Bayesian optimizer. The different numerical complexity – for the 8 topologies – were obtained by starting the optimization from its own initial set of hyper-parameters. The found hyper-parameter sets – corresponding to every topology – are described in Table IV. One can observe on Fig. 10 that the Q-factor gains by the NN are proportional to the complexity, expressed via a number of real multiplications. Noteworthy, the proposed NN requires considerable complexity (topologies T.6, T.7 and T.8) to outperform the Q-factor achieved by DBP with 2 StPS. Nonetheless, even the simplest topologies (T.1 and T.2) had provided considerable performance gains in the order of ∼ 1 dB over CDC+norm. Further optimization of the performance/complexity ratio is possible, notably, by using advanced ML techniques that can optimize the number of connections between neurons without reducing their amount. However, this last consideration is beyond the scope of the current study.

V. Conclusions

We present a novel design of a complex-valued neural network for the signal equalisation on the receiver side. This approach is based on the assumption of Kerr nonlinearity being the leading distortion. We investigated the performance of the suggested neural network in several fiber systems where we compared it with the multi-layer neural network proposed in [11] and standard digital back-propagation. We implemented Bayesian optimizer to fine-tune the parameters of the proposed architecture, i.e. its hyper-parameters, to every considered testcase. In our study, the focus was on the metro networks for which we considered both standard single mode fiber and large effective-area fiber base. The presented numerical and experimental results demonstrate that the proposed neural network leads to significant system performance improvement. Moreover, our approach was also shown to be able to mitigate not only nonlinear fiber transmission distortions, but also impairments arising in the components of both transmitter and receiver.
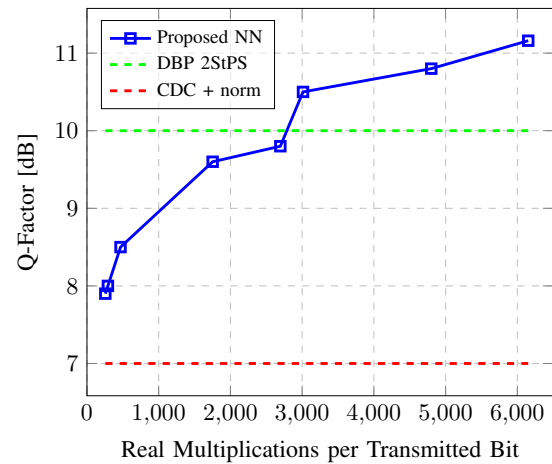

Fig. 10: Relationship between computational complexity and Q factor performance for some possible topologies (Table IV) of the proposed NN in the case of DP-64 QAM single channel LEAF (6×80km) at 3 dBm and its comparison with the performance of DBP with 2 steps-per-span (StPS) and 2 samples-per-symbol and CDC + norm techniques.

VI. Acknowledgment

References

[1] C. Häger and H. D. Pfister, "Nonlinear interference mitigation via deep neural networks," in *2018 Optical Fiber Communications Conference and Exposition (OFC)*. IEEE, 2018, pp. 1–3.

[2] S. Zhang *et al.*, "Field and lab experimental demonstration of nonlinear impairment compensation using neural networks," *Nature communications*, vol. 10, no. 1, pp. 1–8, 2019.

[3] A. Redyuk, E. Averyanov, O. Sidelnikov, M. Fedoruk, and S. Turitsyn, "Compensation of nonlinear impairments using inverse perturbation theory with reduced complexity," *J. Lightwave Technol.*, vol. 38, no. 6, pp. 1250–1257, Mar 2020. [Online]. Available: http://jlt.osa.org/abstract.cfm?URI=jlt-38-6-1250

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

[5] 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, pp. 32 765–32 776, 2018.

[6] S. Owaki and M. Nakamura, "Equalization of optical nonlinear waveform distortion using neural-network based digital signal processing," in *2016 21st OptoElectronics and Communications Conference (OECC) held jointly with 2016 International Conference on Photonics in Switching (PS)*. IEEE, 2016, pp. 1–3.

[7] A. Amari, X. Lin, O. A. Dobre, R. Venkatesan, and A. Alvarado, "A machine learning-based detection technique for optical fiber nonlinearity mitigation," *IEEE Photonics Technology Letters*, vol. 31, no. 8, pp. 627–630, 2019.

[8] T. Koike-Akino, D. S. Millar, K. Parsons, and K. Kojima, "Fiber nonlinearity equalization with multi-label deep learning scalable to high-order DP-QAM," in *Signal Processing in Photonic Communications*. Optical Society of America, 2018, pp. SpM4G–1.

[9] M. Filer, J. Gaudette, M. Ghobadi, R. Mahajan, T. Issenhuth, B. Klinkers, and J. Cox, "Elastic optical networking in the Microsoft cloud [invited]," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 8, no. 7, pp. A45–A54, 2016.

[10] O. Sidelnikov, A. Redyuk, and S. Sygletos, "Nonlinear equalization in long haul transmission systems using dynamic multi-layer perceptron networks," in *2018 European Conference on Optical Communication (ECOC)*. IEEE, 2018, pp. 1–3.

[11] O. S. Sidelnikov, A. A. Redyuk, S. Sygletos, and M. P. Fedoruk, "Methods for compensation of nonlinear effects in multichannel data transfer systems based on dynamic neural networks," *Quantum Electronics*, vol. 49, no. 12, p. 1154, 2019.

[12] G. Khanna, T. Rahman, E. De Man, E. Riccardi, A. Pagano, A. C. Piat, B. Spinnler, S. Calabro, D. Rafique, U. Feiste *et al.*, "Comparison of single carrier 200G 4QAM, 8QAM and 16QAM in a WDM field trial demonstration over 612 km SSMF," in *ECOC 2016; 42nd European Conference on Optical Communication*. VDE, 2016, pp. 1–3.

[13] G. Khanna, T. Rahman, E. De Man, E. Riccardi, A. Pagano, A. C. Piat, S. Calabro, B. Spinnler, D. Rafique, U. Feiste *et al.*, "Single-carrier 400G 64QAM and 128QAM DWDM field trial transmission over metro legacy links," *IEEE Photonics Technology Letters*, vol. 29, no. 2, pp. 189–192, 2016.

[14] S. Gaiarin, A. M. Perego, E. P. da Silva, F. Da Ros, and D. Zibar, "Dual-polarization nonlinear Fourier transform-based optical communication system," *Optica*, vol. 5, no. 3, pp. 263–270, 2018.

[15] G. P. Agrawal, "Nonlinear science at the dawn of the 21st century," *Lecture Notes in Physics*, vol. 542, pp. 195–211, 2000.

[16] I. Terekhov, A. Reznichenko, and S. Turitsyn, "Calculation of mutual information for nonlinear communication channel at large signal-to-noise ratio," *Physical Review E*, vol. 94, no. 4, p. 042203, 2016.

[17] A. Reznichenko and I. Terekhov, "Investigation of non-linear communication channel with small dispersion via stochastic correlator approach," in *Journal of Physics: Conference Series*, vol. 1206, no. 1. IOP Publishing, 2019, p. 012013.

[18] A. V. Reznichenko and I. S. Terekhov, "Path integral approach to nondispersive optical fiber communication channel," *Entropy*, vol. 22, no. 6, p. 607, May 2020. [Online]. Available: http://dx.doi.org/10.3390/e22060607

[19] V. Oliari, E. Agrell, and A. Alvarado, "Regular perturbation on the group-velocity dispersion parameter for nonlinear fibre-optical communications," *Nature communications*, vol. 11, no. 1, pp. 1–11, 2020.

[20] M. Pelikan, D. E. Goldberg, E. Cantú-Paz *et al.*, "Boa: The bayesian optimization algorithm," in *Proceedings of the genetic and evolutionary computation conference GECCO-99*, vol. 1, 1999, pp. 525–532.

[21] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.

[22] M. Nakamura, Y. Fukumoto, S. Owaki, T. Sakamoto, and N. Yamamoto, "Experimental demonstration of SPM compensation using a complex-valued neural network for 40-Gbit/s optical 16QAM signals," *IEICE Communications Express*, vol. 8, no. 8, pp. 281–286, 2019.

[23] T. Kim and T. Adali, "Fully complex multi-layer perceptron network for nonlinear signal processing," *Journal of VLSI signal processing systems for signal, image and video technology*, vol. 32, no. 1-2, pp. 29–43, 2002.

[24] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[25] D. Blalock, J. J. G. Ortiz, J. Frankle, and J. Guttag, "What is the state of neural network pruning?" *arXiv preprint arXiv:2003.03033*, 2020.

[26] M. Tan and Q. V. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," *arXiv preprint arXiv:1905.11946*, 2019.

[27] T. A. Eriksson, H. Bülow, and A. Leven, "Applying neural networks in optical communication systems: possible pitfalls," *IEEE Photonics Technology Letters*, vol. 29, no. 23, pp. 2091–2094, 2017.

[28] P. J. Freire and V. Neskornuik, "Complex-Valued Neural Network Design for Mitigation of Signal Distortions in Optical Links in Highly Nonlinear Regimes - CODE," Oct. 2020. [Online]. Available: https://doi.org/10.5281/zenodo.4065404 and https://github.com/pedrofreiree/Complex-NN-based-on-NLSE

[29] R.-J. Essiambre, G. Kramer, P. J. Winzer, G. J. Foschini, and B. Goebel, "Capacity limits of optical fiber networks," *Journal of Lightwave Technology*, vol. 28, no. 4, pp. 662–701, 2010.

[30] 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.

[31] M. Schaedler, C. Bluemm, M. Kuschnerov, F. Pittalà, S. Calabrò, and S. Pachnicke, "Deep neural network equalization for optical short reach communication," *Applied Sciences*, vol. 9, no. 21, p. 4675, 2019.

[32] A. Napoli, P. W. Berenguer, T. Rahman, G. Khanna, M. M. Mezghanni, L. Gardian, E. Riccardi, A. C. Piat, S. Calabrò, S. Dris *et al.*, "Digital pre-compensation techniques enabling high-capacity bandwidth variable transponders," *Optics Communications*, vol. 409, pp. 52–65, 2018.

[33] B. Spinnler, "Equalizer design and complexity for digital coherent receivers," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 16, no. 5, pp. 1180–1192, 2010.

[34] R. Rath and W. Rosenkranz, "On the performance of digital back-propagation for imperfect knowledge of link design," in *2015 Tyrrhenian International Workshop on Digital Communications (TIWDC)*. IEEE, 2015, pp. 39–42.

[35] C.-Y. Lin, A. Napoli, B. Spinnler, V. Sleiffer, D. Rafique, M. Kuschnerov, M. Bohn, and B. Schmauss, "Adaptive digital back-propagation for optical communication systems," in *Optical Fiber Communication Conference*. Optical Society of America, 2014, pp. M3C–4.