Artificial Neural Network Symbol Estimator With Enhanced Robustness to Nonlinear Phase Noise

Artificial Neural Network Symbol Estimator with Enhanced Robustness to Nonlinear Phase Noise

João M. Santos, Andrea Carena, Paulo P. Monteiro and Fernando P. Guiomar

***Abstract*—This letter reports a novel approach for nonlinear phase noise mitigation, based on artificial neural networks (ANNs) tailored to classification applications and a pre-processing stage of feature engineering. Starting with a set of proof-of-concept simulations, we verify that the proposed system can achieve optimal performance for the additive white Gaussian noise (AWGN) channel. Then, considering a dispersion-less channel with strong nonlinear phase noise (NLPN) distortion, we demonstrate an increase in nonlinear tolerance by up to 2 dB, comparing with standard carrier-phase estimation (CPE) followed by minimum distance detection. Finally, simulating the propagation of 64 Gbaud PM-16QAM over standard single mode fiber (SSMF), we verify that the ANN-based solution is effective on wavelength-division multiplexing (WDM) transmission conditions, enabling to increase the maximum signal reach by approximately 1 fiber span over the legacy CPE-enabled NLPN compensation.**

***Index Terms*—Artificial neural network, coherent detection, nonlinear mitigation, optical communications.**

I. Introduction

THE impact of nonlinear interference noise (NLIN) on optical fiber links has been intensively studied during the last decade [1], [2]. It has been generally acknowledged that, in dispersion-unmanaged fiber links, NLIN is mostly generated in the form a Gaussian-like noise [2], whose mitigation requires complex channel inversion techniques [3]. Nevertheless, it has also been found that an important nonlinear phase noise (NLPN) component can likewise be generated, mainly when operating at high spectral-efficiency [1]. At the same time, it has been shown that this NLPN-induced distortion can be partially removed by a common carrier-phase estimation (CPE) algorithm [4]. However, the effectiveness of CPE-based NLPN mitigation is known to strongly depend on the extent of its temporal correlation, as well as on the signal-to-noise ratio (SNR) operating conditions [5].

Like many other fields, the optical communications community was not indifferent to the rise of machine learning (ML) and started to search for ways of exploiting its distinctive features [6]. One key advantage of ML techniques is their ability to extract knowledge from the data itself without requiring the derivation of analytical models. This characteristic makes ML algorithms excellent at identifying and compensating for nonlinear relationships.

Motivated by the challenges posed by NLIN, nonlinear equalization quickly became one of the most common applications of artificial neural networks (ANNs) in optical systems. In [7] it was shown that simple regenerative ANNs with just two inputs could increase both the laser linewidth tolerance and the resilience to IQ imbalances when compared to k-Means. In order to enable a joint treatment of the in-phase and quadrature signal components, complex-valued ANNs have also been addressed in the literature [8], [9]. In [10], an ANN-based compensation method has enabled a spectral efficiency-distance product record in a field-trial experiment.

In this letter, we devise a strategy for extracting more information from the complex signal representation, and then feed it to a modest-size ANN that replaces the symbol estimation subsystem, adjusting the decision boundaries for non-white noise introduced by the channel. This ANN-based nonlinear estimator proved to be effective in the presence of NLPN for both QPSK and 16QAM modulation schemes, surpassing the performance of legacy CPE algorithms both in dispersion-less and dispersion-unmanaged optical systems.

II. ANN-based Nonlinear Symbol Estimation

ANNs are a very powerful machine learning algorithm, being able to adapt to a wide range of scenarios that can usually be divided into regression or classification problems. While the concept of artificial neural networks covers a large set of variations such as convolution or recurrent neural networks, in this work we will restrict our approach to the use of feedforward neural networks, often known as multilayer perceptrons [11], as illustrated in Fig. 1. Topologically, these networks consist of a stack of layers composed by several parallel computing units, the *neurons*, only allowing the data to flow from the input to the output layer. These structural parameters are usually called *hyper-parameters*. ANNs owe their versatility to the hidden computing units, which when chained together can abstract the input data to a high-dimensional space thus enabling a theoretical approximation of any function. In its purest form, the learning process is an optimization problem where the algorithm learns to map a function $\hat{\mathbf{y}} = f(\mathbf{x}; \boldsymbol{\Theta})$ by inferring the data itself and adjusting its internal parameters, $\theta_{ij}^{[l]}$, based on the inference outcome.

This work has been supported by the European Regional Development Fund (FEDER), through the Portugal 2020 framework (CENTRO 2020), and by national public funds (FCT, OE) projects ORCIP (CENTRO-01-0145-FEDER-022141) and FreeComm-B5G (UIDB/50008/2020). Fernando P. Guiomar acknowledges a fellowship from "la Caixa" Foundation (ID 100010434). The fellowship code is LCF/BQ/PR20/11770015.

João M. Santos, Paulo P. Monteiro and Fernando P. Guiomar are with Instituto de Telecomunicações and Department of Electronic, Telecommunications, and Informatics (DETI), University of Aveiro, 3810-193, Aveiro, Portugal. e-mail: joao.mp.santos@ua.pt

Andrea Carena is with Dipartimento di Elettronica e Telecomunicazioni, Politecnico di Torino, Corso Duca degli Abruzzi, 24, 10129 Torino, Italy (e-mail: andrea.carena@polito.it).

Manuscript received April XX, 2021; revised April XX, 2021.

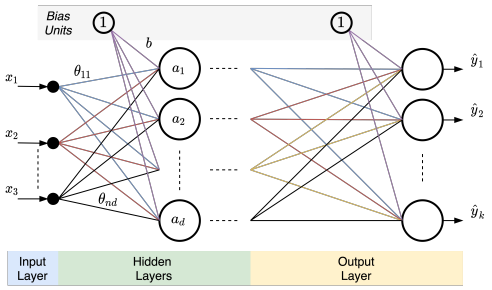

Fig. 1: Topology illustration of a standard feedforward ANN.

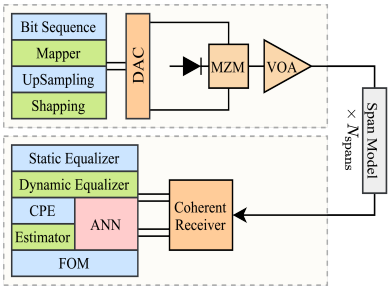

Fig. 2: Simulation setup utilized for coherent optical transmission.

A. System Architecture

Our proposal consists of a classification-tailored ANN that will replace the typical *minimum-distance* symbol estimation block, providing optimized nonlinear decision boundaries. To achieve this purpose, we idealized a three-layered ANN where the activation function of the hidden layer is an *hyperbolic-tangent* and the output is governed by the *softmax* function. The choice of input signals is crucial, as it defines how the ANN will only interact with the system. We mapped the input symbol, x_k, onto a set of mathematical characteristics, also denoted as *features*, that take advantage of the complex representation of the signal and project the estimation process to an hyper-dimensional space, enabling a decision to be made with more criteria than just the distance to the reference points. After some preliminary tests, a total of $11 + M$ features have been chosen, according to the following criteria:

- $I = \text{Re}\{x_k\}$; $Q = \text{Im}\{x_k\}$;
- $||\mathbf{v}||, \mathbf{v} = (I, Q)$, absolute value of the QAM symbol;
- $\angle\mathbf{v}$, phase of the QAM symbol;
- $I \times Q$, product of real and imaginary components;
- $||s_i - \mathbf{x_k}||$, distance to each constellation symbol;
- I^2; Q^2, square value of each component;
- $\sin I$; $\sin Q$;
- $\cos I$; $\cos Q$;

B. Development Pipeline

Having specified the number of layers and the respective activation functions, we shall now define the number of nodes in each layer. For the output layer, the use of the *softmax* function requires one node per object, thereby resulting in a number of outputs equals to the number of symbols in the QAM constellation. However, there is no clear rule for what should be the node count in the hidden layer. While, on the one hand, a larger hidden layer might potentially lead to an improved model extraction capability, on the other hand, its increased complexity might also hinder the learning process. To tackle this intricate challenge, we implement a grid-search strategy, which consists of training a network for each possible set of hyper-parameters, within a pre-determined interval, *e.g.* up to 2^{12} units, and select the best outcomes. For the supervised learning of the ANN, we use the resilient back-propagation algorithm, paired with early-stop regularization to prevent model overfitting, and the cross-entropy loss function to optimize the classification model. The dataset has been divided into training, validating and testing subsets with ratios of 70%, 15% and 15%, respectively. Note that the testing set allows to verify if the trained network is actually able to generalize from the training data.

III. Numerical Performance Assessment

Since the focus of this work is on the compensation of non-Gaussian nonlinear fiber impairments, we devised a simulation setup that allows assessing different scenarios and development conditions. The setup, depicted in Fig. 2, consists of a 64 GBaud M-QAM transmitter that employs root-raised cosine pulse shaping. In the preliminary validation stages, we start by considering simplified channel models, where single-polarization transmission is employed. After that, we consider a more realistic dual-polarization propagation over an actual optical fiber channel. At the receiver, a minimal stack of digital signal processing (DSP) subsystems is applied, as all hardware is deemed to be ideal. The benchmark for NLPN compensation was set by the well-known Viterbi & Viterbi CPE (VV-CPE) with an optimized memory length [4]. Finally, the system performance is evaluated through bit error ratio (BER) counting.

A. Preliminary Validation over an AWGN Channel

We start our numerical analysis by performing a series of tests over an additive white Gaussian noise (AWGN) channel. Apart from the methodology validation, these tests are also helpful to establish a performance baseline for the ANN estimator. Under this assumption, the channel model only accounts for the inline amplification noise,

$$\tilde{A}(t, z + L_{\text{span}}) = \tilde{A}(t, z) + n_{\text{ASE}}(t), \tag{1}$$

where t and z represent the temporal and spatial coordinates, respectively, and $\tilde{A}(t, z)$ is the complex envelope of the optical signal. $L_{\text{span}} = 80\,\text{km}$ is the fiber span length and $n_{\text{ASE}}(t) = h(G-1)Ff_0B$ is the amplified spontaneous emission (ASE) noise of the Erbium-dopped fiber amplifier (EDFA) with a noise figure of $F = 5\,\text{dB}$ and a gain, G, that exactly inverts the span loss. A typical fiber attenuation coefficient, $\alpha = 0.2\,\text{dB/km}$, is assumed, yielding $G = 16\,\text{dB}$. Finally, h is the Planck's constant, $f_0 = 193.4\,\text{THz}$ is the central frequency and $B = 64\,\text{GHz}$ is the noise reference bandwidth.

Using the simulation setup of Fig. 2, both QPSK and 16-QAM signals were generated, with a total of 91750 symbols. The BER evaluated on the test subset for a purely AWGN channel is depicted in Fig. 3. The ANN-based solution was able to acquire the Gaussian characteristic of the signal

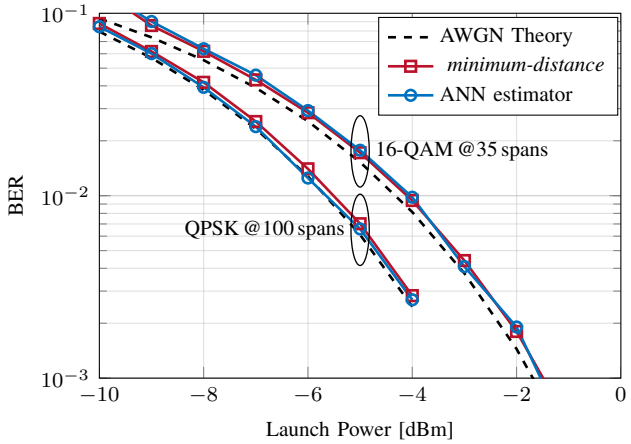

Fig. 3: BER performance of the ANN symbol estimator in a purely AWGN scenario.

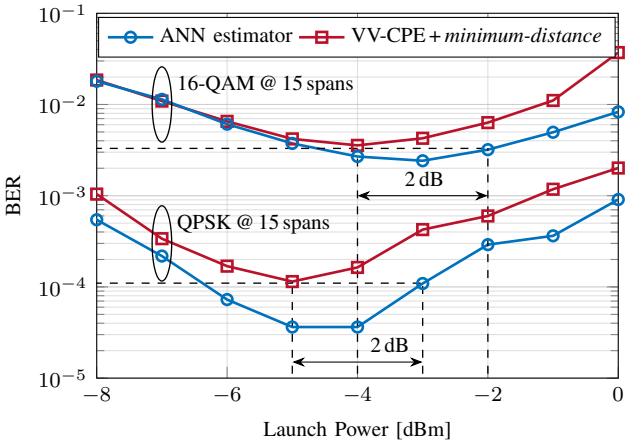

Fig. 5: BER versus launched optical power after propagation over 15 spans of a dispersion-less fiber.

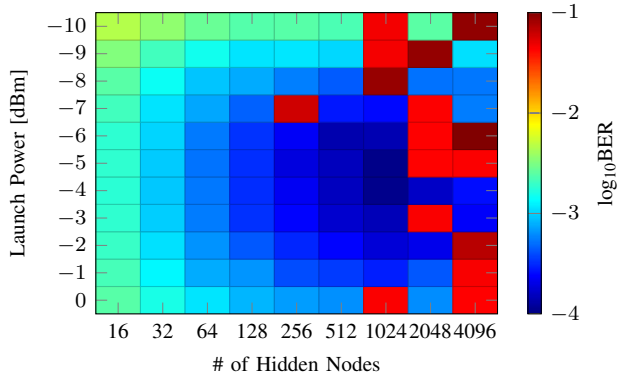

Fig. 4: Grid-search optimization of the number of neurons on the hidden layer of the ANN estimator when applied to an NLPN-only channel and considering QPSK modulation.

as it converged to the theoretical performance curve. The overlap with the *minimum-distance* symbol detection curve in an AWGN channel validates the ANN-based solution as a reliable symbol estimator and sets the baseline for more advanced scenarios. In addition, this initial AWGN validation also ensures that the ANN-based estimator can safely be employed in optical systems that are temporarily operating at low launched powers when the SNR budget is high enough (e.g. for higher energy efficiency), without incurring into performance penalties.

B. Validation over a Static Nonlinear Fiber Channel

After successfully validating over an AWGN channel, we now proceed to assess the potential of the ANN estimator over a non-linearly impaired signal. We start by considering a simplified multi-span dispersion-less fiber channel,

$$\tilde{A}(t, z + L_{\text{span}}) = \tilde{A}(t, z) \exp\left(-j\Phi_{\text{NL}}(t, z)\right) + n_{\text{ASE}}(t), \quad (2)$$

where $\Phi_{\text{NL}}(t, z) = \gamma P(t, z) L_{\text{eff}}$ is the nonlinear phase-shift, with $\gamma = 1.3\,\text{W}^{-1}\text{km}^{-1}$ as the nonlinear coefficient, $P(t, z) = |A(t, z)|^2$ the signal power and $L_{\text{eff}} = (1 - \exp(-\alpha L_{\text{span}}))/\alpha$ the effective span length.

In order to optimize the ANN architecture, we have performed a grid-search to obtain the optimal number of hidden units. An example of the application of this procedure is depicted in Fig. 4, for the QPSK transmitted signal. From the observation of this figure, we can clearly identify a 2D valley of optimized BERs, both in terms of launched power and number of hidden units. It can be seen that the best performance around the optimal launched power tends to be achieved with 512–1024 hidden units. Nevertheless, it is also clear that the probability of divergence during the ANN training also increases substantially when using a larger number of hidden units, which is a natural consequence of the increased ANN complexity. This behavior clearly outlines the engineering tradeoff between complexity and tractability, which should be carefully taken into account during the design of the ANN architecture. After determining the optimum number of hidden units for each launch power, the obtained BER results are presented in Fig. 5. Even though the ANN-based solution cannot fully compensate the nonlinear rotation, it was able to outperform the benchmark VV-CPE with an optimized number of taps, achieving a BER reduction up to 68% for the QPSK signal and 22% for the 16QAM signal, together with an increase of 2 dB on the nonlinear tolerance.

The ANN-optimized symbol decision boundaries for the 16QAM signal at a launch power of -3 dBm are depicted in Fig. 6. It can be observed that the ANN is effectively able to follow the rotation of the signal, which verifies the concept of this system as a nonlinear symbol estimator. Moreover, we also see that the training algorithm is capable of identifying the specific broadening of each cluster instead of capturing the average rotation, such as done by *m-th* power CPE algorithms.

C. Performance Assessment over WDM Transmission

Building on the promising results delivered by the ANN estimator over a static NLPN-only nonlinear channel, in this section we evolve to a more realistic wavelength-division multiplexing (WDM) transmission system in which the NLPN component is known to be dominated by the nonlinear interference from the co-propagating channels. We generate 11 independent channels with identical bandwidth and modulation (64 Gbaud 16QAM) and 75 GHz channel spacing. Dual-polarization transmission is now considered and

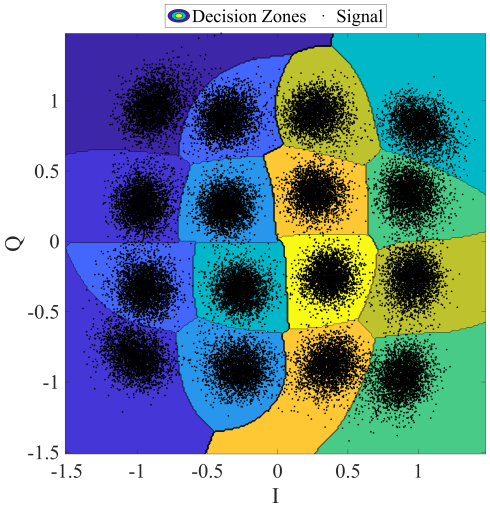

Fig. 6: Decision zones used by the ANN-based estimator with 512 hidden units for a 16QAM signal with −3 dBm launched power after propagation over 15 dispersion-less spans.

the fiber propagation is simulated with the split-step Fourier implementation of the Manakov equation. Standard single mode fiber (SSMF) parameters are assumed, i.e. an attenuation coefficient $\alpha = 0.2$ dB/km, a group velocity dispersion coefficient $\beta_2 = -20.4\,\text{ps}^2/\text{km}$ and $\gamma = 1.3\,\text{W}^{-1}\text{km}^{-1}$. Due to the increased computational complexity, the simulation length has been reduced to 60314 symbols per channel. The central channel is the designated channel under test, over which the ANN estimator is applied. The same grid-search procedure was performed, and the corresponding optimized number of neurons on the hidden layer has been chosen for each input power, denoting a tendency for requiring a higher number of neurons with the increase of nonlinear interference. The ANN symbol estimator was then trained, validated and tested over a wide range of launch optical powers and number of fiber spans. From the BER results obtained for each launch power and propagation distance, and considering a threshold pre-FEC BER of 2.4×10^{-2}, we evaluate the maximum reach of the transmitted signal as depicted in Fig. 7. As a performance benchmark for the ANN estimator, Fig. 7 also includes the maximum reach results obtained with and without VV-CPE. It can be observed that, while the optimized VV-CPE allows to partially compensate the NLIN distortion and thus increase the transmission reach by ∼30 km, the ANN-based symbol estimator is able to roughly double this maximum reach gain, achieving an extended reach of approximately 75 km.

IV. Conclusion

We have developed and numerically assessed an ANN classifier that is able to optimize the decision boundaries of an M-QAM symbol demapper to minimize the number of erroneous decisions in scenarios with additive nonlinear noise. The ANN-based demapper was shown to be particularly effective in scenarios with strong NLPN distortion, enabling a 2 dB increase in nonlinear tolerance, when compared to the standalone use of CPE-enabled NLPN mitigation. Relevant gains in terms of maximum signal reach were also identified for WDM transmission over dispersion-unmanaged

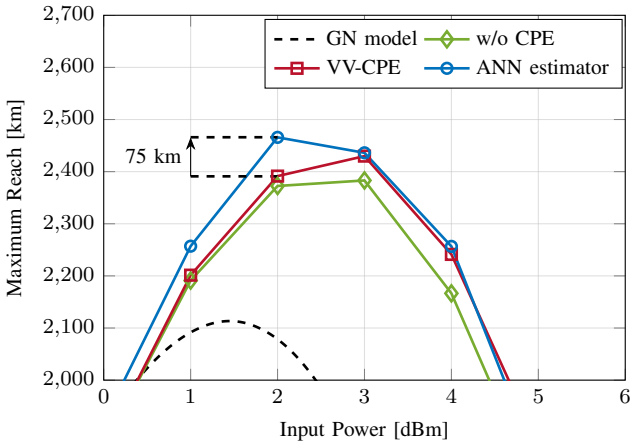

Fig. 7: Maximum system reach of a 64 Gbaud 16-QAM signal after transmission over an 11-channel WDM link.

fiber links, thereby demonstrating the potential applicability of the ANN-based symbol estimator in practical coherent optical communication systems.

References

[1] A. Carena, G. Bosco, V. Curri, Y. Jiang, P. Poggiolini, and F. Forghieri, "EGN model of non-linear fiber propagation," *Opt. Express*, vol. 22, no. 13, pp. 16 335–16 362, 2014.

[2] R. Dar, M. Feder, A. Mecozzi, and M. Shtaif, "Pulse collision picture of inter-channel nonlinear interference in fiber-optic communications," *Journal of Lightwave Technology*, vol. 34, no. 2, pp. 593–607, Jan 2016.

[3] ——, "Inter-channel nonlinear interference noise in WDM systems: Modeling and mitigation," vol. 33, no. 5, pp. 1044–1053, Mar 2015.

[4] F. P. Guiomar, A. Carena, G. Bosco, L. Bertignono, A. Nespola, and P. Poggiolini, "Nonlinear mitigation on subcarrier-multiplexed PM-16QAM optical systems," *Opt. Express*, vol. 25, no. 4, pp. 4298–4311, Feb 2017.

[5] O. Golani, D. Pilori, F. P. P. Guiomar, G. Bosco, A. Carena, and M. Shtaif, "Correlated nonlinear phase-noise in multi-subcarrier systems: modeling and mitigation," *Journal of Lightwave Technology*, vol. 38, no. 6, pp. 1148–1156, Mar 2020.

[6] D. Zibar, M. Piels, R. Jones, and C. G. Schäeffer, "Machine learning techniques in optical communication," *Journal of Lightwave Technology*, vol. 34, no. 6, pp. 1442–1452, Mar 2016.

[7] D. Wang, M. Zhang, Z. Li, C. Song, M. Fu, J. Li, and X. Chen, "System impairment compensation in coherent optical communications by using a bio-inspired detector based on artificial neural network and genetic algorithm," *Optics Communications*, vol. 399, pp. 1–12, Sep. 2017.

[8] S. Liu, Y. M. Alfadhli, S. Shen, H. Tian, and G.-K. Chang, "Mitigation of Multi-user Access Impairments in 5G A-RoF-based Mobile Fronthaul utilizing Machine Learning for an Artificial Neural Network Nonlinear Equalizer," in *Optical Fiber Communication Conference*. San Diego, California: OSA, 2018, p. W4B.3.

[9] Y. Fukumoto, S. Owaki, T. Sakamoto, N. Yamamoto, and M. Nakamura, "Experimental Demonstration of SPM Compensation Based on Digital Signal Processing Using a Complex-Valued Neural Network for 40-Gbit/s Optical 16QAM Signals," in *2018 23rd Opto-Electronics and Communications Conference (OECC)*. Jeju Island, Korea (South): IEEE, Jul. 2018, pp. 1–2.

[10] V. Kamalov, L. Jovanovski, V. Vusirikala, S. Zhang, F. Yaman, K. Nakamura, T. Inoue, E. Mateo, and Y. Inada, "Evolution from 8QAM live traffic to PS 64-QAM with Neural-Network Based Nonlinearity Compensation on 11000 km Open Subsea Cable," in *Optical Fiber Communication Conference Postdeadline Papers*. San Diego, California: OSA, 2018, p. Th4D.5.

[11] C. M. Bishop, *Pattern Recognition and Machine Learning*, corrected at 8th printing 2009 ed., paper Information Science and Statistics. New York, NY: Springer, 2009.