



Optics Letters

Combining nonlinear Fourier transform and neural network-based processing in optical communications

OLEKSANDR KOTLYAR,*  MARYNA PANKRATOVA, MORTEZA KAMALIAN-KOPAE, 
ANASTASIIA VASYLCHENKOVA,  JAROSLAW E. PRILEPSKY, AND SERGEI K. TURITSYN 

Aston Institute of Photonic Technologies, Aston University, Birmingham, B4 7ET, UK

*Corresponding author: o.kotlyar1@aston.ac.uk

Received 31 March 2020; revised 14 May 2020; accepted 15 May 2020; posted 15 May 2020 (Doc. ID 394115); published 22 June 2020

We propose a method to improve the performance of the nonlinear Fourier transform (NFT)-based optical transmission system by applying the neural network post-processing of the nonlinear spectrum at the receiver. We demonstrate through numerical modeling about one order of magnitude bit error rate improvement and compare this method with machine learning processing based on the classification of the received symbols. The proposed approach also offers a way to improve numerical accuracy of the inverse NFT; therefore, it can find a range of applications beyond optical communications.

Published by The Optical Society under the terms of the [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/). Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

<https://doi.org/10.1364/OL.394115>

The nonlinear Fourier transform (NFT) [1] has recently attracted a great deal of interest as a possible way to combat nonlinear signal distortions in fiber-optic communication systems [2–5]. In NFT-based optical communications, either the parameters associated with the discrete eigenvalues (solitary modes) [2] or the continuous part of the nonlinear spectrum (dispersive modes) [3–6] can be utilized as data carriers. NFT-based methods effectively linearize the nonlinear fiber channel. However, there are many technical problems to be resolved before NFT will become a practical technology. The major challenge comes from the nonlinear interaction of signal and noise, which becomes even more complicated because of the dispersive spreading [7,8] of the signal and processing-related noise [9]. This nonlinear interaction of signal and noise leads to cross-talk between nonlinear spectrum components (nonlinear modes), degrading the quality of data transmission.

Among various solutions to the nonlinear interaction of signal and noise in optical communication, several machine learning (ML) techniques (both supervised and unsupervised) have already shown their high potential to improve the system performance (see, e.g., [10–13] and references therein). ML

methods designed for nonlinear system modeling, and inherently nonlinear approaches such as NFT, can be synergistically combined, enhancing each other. Perhaps, the first combination of NFT systems with the neural network (NN) equalization was shown in [14] for mapping of solitons in time domain directly to symbols, substituting the NFT procedure for the system employing only solitary modes. Previously, we demonstrated the successful application of various ML techniques such as k -nearest neighbors, support vector machine, and other algorithms in the processing of received symbols for both “traditional” NFT (with vanishing boundary conditions) schemes [15] and periodic NFT systems [16,17]. Those ML methods provide better decision boundaries at the detection stage when a received symbol is classified and attributed to a transmitted symbol. Therefore, these solutions can be especially useful when the optimum boundaries ensuing from the distribution of the received symbols over the complex plane are notably distorted. Although this is common in optical communication systems when nonlinearity plays a significant role, it might be different in the corresponding NFT-based communication schemes where the fiber is effectively linearized. This effective linearization of fiber is evident in continuous spectrum communication [18], and the recently introduced b -modulation transmission [19]. In the range of system parameters that we have studied, the receiver symbols, distorted by a complicated combination of noise and signal, form probabilistic clouds that typically retain a spherical shape—so much like in linear communication with additive noise. This makes the optimum detection scheme quite close to the minimum Euclidean distance and leaves little room to improve the system performance for ML methods. In such a case, making the cloud size smaller seems to be the only way to improve a constellation detection scheme. This translates into reducing the effective noise from the data-carrying signal (e.g., the continuous spectrum). This entails processing a continuous function such as the continuous spectrum, $r(\xi)$, where $\xi \in \mathbb{R}$ plays the role of frequency in the nonlinear Fourier (NF) domain, instead of processing the resulting constellation symbols. Denoising a signal is usually done through linear filtering with respect to the characteristics of noise. However, noise

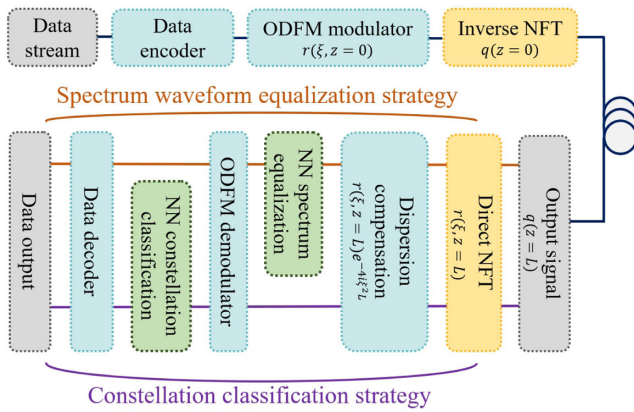


Fig. 1. Schematics of the NFT-based communication systems with two strategies of the receiver-side processing involving ML methods, marked with yellow and purple lines piercing the respective receiver processing blocks.

in the NF domain is a result of a complicated nonlinear transformation of a mixture of noise and signal. Although there are some ongoing studies of the noise properties in the NF domain [7–9], the stochastic characteristics of noise for the continuous spectrum are not straightforward to use for the design of the optimum receiver [7]. This calls for development of methods that can encompass the combined impact of various effects and distortions in the NF domain and mitigate its effects on the received signal.

We propose here to apply the supervised ML technique directly to the continuous nonlinear spectrum $r(\xi)$ of the NFT system (Fig. 1) rather than for processing the detected symbols in the received constellations. We employ the algorithm that utilizes a feedforward artificial NN (Fig. 2) for equalizing the received $r(\xi)$ waveform to effectively mitigate both the signal distortion and effective noise in the NFT domain. To show the efficiency of the proposed denoising NN regression-based approach, we compare it with an NN-based classifier for the received symbols. The performance of two approaches is compared in terms of achieved bit error rate (BER) for different signal powers at two data rates. The classification solution that we have used as a reference is similar to its successful applications in NFT communications that have been reported before [15–17]. We also anticipate that this method can be applicable to a wide range of similar NFT systems.

We consider the focusing nonlinear Schrödinger equation (NLSE) with additive white Gaussian noise η as the governing model for the propagation of signal q in a single-mode fiber [5,20]. The dimensionless NLSE reads

$$i \frac{\partial q}{\partial z} + \frac{1}{2} \frac{\partial^2 q}{\partial t^2} + |q|^2 q = \eta. \quad (1)$$

We assume ideally distributed amplification where gain fully compensates for losses. We simulate the propagation of optical signal down the fiber link for the propagation distance of 1000 km (see Fig. 1). For our simulations, we use an orthogonal frequency-division multiplexing (OFDM) modulation with 128 sub-carriers in the nonlinear continuous spectrum:

$$r(\xi, z=0) = \sum_{k=-63}^{64} c_k \text{sinc}(2\xi - k), \quad (2)$$

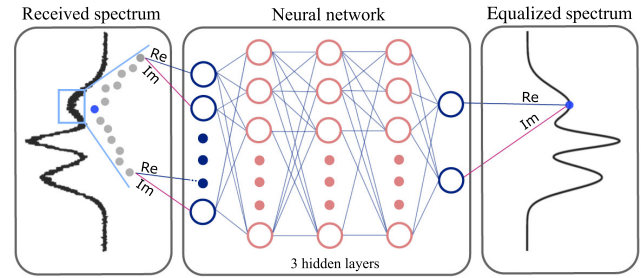


Fig. 2. Schematics of the spectra equalizer implemented as the feedforward NN with three fully connected hidden layers. As an input to the NN, we take a sample from received nonlinear spectra (blue circle) and its neighbors (gray circles) from each side. The output of the NN is an “equalized” sample.

where c_k are from the alphabet of 64- or 128-quadrature-amplitude modulation (QAM) symbols. The transmission is implemented in a burst mode, where the full temporal support of a single symbol (burst) is 9.8 ns. Details of the system parameters used in our work are in [8,9]. The achievable data rates are 63 Gbit/s and 74 Gbit/s for 64- and 128-QAM, correspondingly, accounting for the dispersion-broadening memory to the burst length.

In this work we use the fast NFT and inverse NFT (INFT) algorithms from [21] for the NF signal processing: these methods have signal processing complexity of order $O(n \log^2 n)$, where n is the number simulation samples, and second-order accuracy with regard to sample size. The algorithms demonstrate the highest ratio between performance and accuracy characteristics. It is worthwhile mentioning that as shown in [9], the performance of the NFT-based communication system is determined by the interplay between in-line noise and the “processing” noise arising from the imperfectness of finite-precision NFT procedures. The impact of processing noise is especially pronounced for high signal powers. At the same time, for low signal powers, the main corruption source is amplified spontaneous emission (ASE) noise. So our equalization methods deal with both the processing and in-line noise-induced corruptions simultaneously.

We start with estimating the performance of our NFT-based communication system without the NN equalizer for 64- and 128-QAM. As seen in Fig. 3 (purple curves), using the hard decision (HD), we obtain BER values that are much higher than the HD forward error correction (FEC) threshold for signal powers around the optimal value and for both data rates.

One solution is to improve the classification stage by using ML techniques. In our previous work [15], we improved the performance of a similar NFT-based optical transmission system by using several ML methods for the classification of received symbols. It was demonstrated that the employment of ML techniques makes it possible to improve BER of the system by 30% for 16-QAM symbols. Therefore, first, we apply the ML technique for the classification of received symbols to benchmark the achievable improvements. We utilize the feed-forward NN to create the optimum boundaries ensuing from the distribution of the received symbols. Figure 1 shows the “constellation classification strategy” for the exact point where we apply ML in our scheme. We train the NN by providing the received symbols as input and the transmitted ones as labels. Then, the trained NN classifies newly received symbols

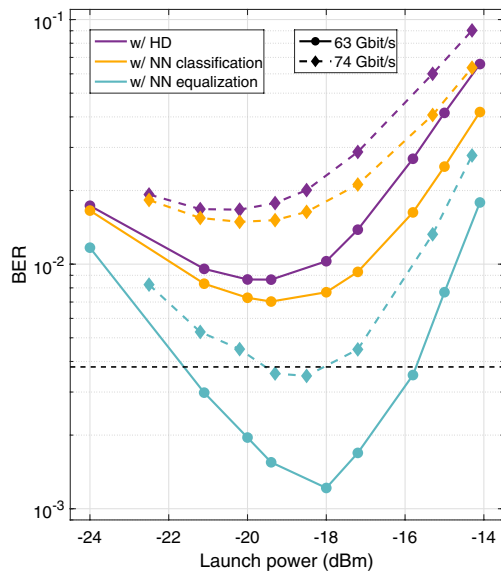


Fig. 3. BER versus power dependence for propagation distance 1000 km, using hard decision (purple curves), classification of received symbols by NN (yellow curves), or applying NN for equalization of the received spectrum (blue curves) for two data rates of 63 Gbit/s (solid lines) and 74 Gbit/s (dashed lines). Horizontal black dashed line represents HD FEC threshold.

in order to increase performance of the system, reducing BER by more accurate decision boundaries. For classification, we use a feed-forward fully connected NN network that consists of three hidden layers with 64 nodes in each layer, with a leaky rectified linear unit (ReLU) activation function. For the training and validation set, we use 7.7×10^5 and 6.4×10^4 symbols, respectively. Results presented in Fig. 3 are obtained by classifying a test set of 3.4×10^6 that was not used during the training process. We would like to emphasize that in this approach, NN is used to provide more accurate decision boundaries but without shifting the positions of received symbols itself.

The results of this symbol classification are presented in Fig. 3 by yellow lines for both data rates. It can be seen in the figure that performance improvement is insignificant and can be linked in this case to slight constellation rotation and scaling, which can be easily removed with simple methods. As the classification of symbols cannot provide desirable improvement that would decrease the BER below the HD FEC threshold in the considered case, we turn to the main point of our Letter: the application of NN-based equalization to the received nonlinear spectrum itself. The method is illustrated schematically in Fig. 2. For details on how we apply the ML in our new receiver scheme, see Fig. 1, where the nonlinear spectrum waveform equalization strategy is marked with yellow lines. So we use the received continuous nonlinear spectrum as an input for the NN in order to equalize the nonlinear spectrum. For such a regression task, we employ a fully connected NN that has the same number of hidden layers and nodes as the classification task described earlier: three hidden layers with 64 nodes in each layer. We have chosen the leaky ReLU because it provides better performance than the ordinary ReLU and hyperbolic tangent activation functions. Real and imaginary parts of each spectrum point at the input (output) are considered as separate nodes [see the left (right) part of Fig. 2]. The NN was trained using the Nesterov-accelerated

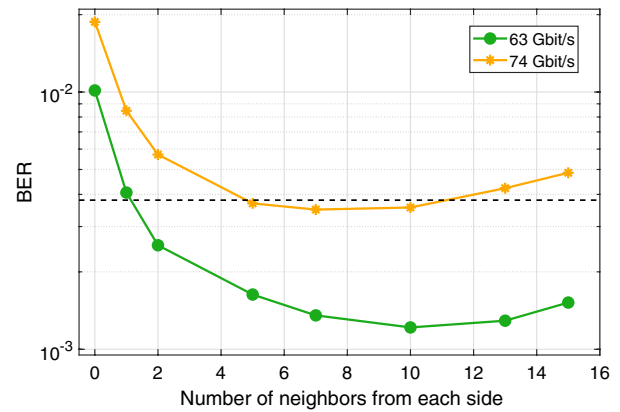


Fig. 4. BER as a function of the number of neighbors (taps) used in the NN equalization for propagation distance of 1000 km, where the power was taken in the vicinity of optimal value in -18 dBm.

adaptive moment estimation (NADAM) [22] optimization algorithm. Training was performed on 1.2×10^4 realizations of in-line noise; another 1×10^3 noise realizations were used as the validation set. To present results in the Letter, we use a distinct test set that consists of 5.2×10^4 in-line noise realizations not used in the training step. The training is performed for a data set consisting of received and initial spectra samples at each signal power separately.

As we mentioned above, the physical noise and numerical processing noise both contribute to effective crosstalk and lead to emerging correlations between the nonlinear spectrum components. An important part of our approach is that this correlation between the nonlinear spectrum components is taken into account, such that we perform a multi-tap equalization (see the left side of Fig. 2).

In particular, on the preprocessing stage, we create samples for NN training that consist of an initial element (blue circle in blue processing window in Fig. 2) from a nonlinear spectrum and its nearest neighbors (gray circles in processing window). Then, these new training samples are treated as independent. Before training the NN, all training samples from different noise realizations were interleaved and then shuffled. We consider here 10 nearest neighbors of $r(\xi_m)$ from both sides (i.e., 21 taps in total) of the received r -spectrum to equalize the central sample. To identify the optimal number of neighbors, we studied the dependence of our equalizer performance on the number of neighbors. We trained the NN and computed the respective BER values for the number of neighbors varying from zero to 30 in total, i.e., up to 15 from each side of the sample of interest. As shown in [9], the correlation between the nonlinear spectrum components rapidly decays; therefore, one might expect to see an improvement in the equalizer performance as the considered neighbors increase up to the actual effective correlation length produced by all channel distortions. Beyond this point, processing a larger window only adds more noise into the system, hence, degrading the performance. In Fig. 4, the BER first decreases with the number of neighbors; however, there is a slight growth of BER after seven and 10 neighbors for 74 and 63 Gbit/s transmission rates, respectively. We fixed the number of neighbors to 10 because such a configuration provides optimal performance for 63 Gbit/s. For 74 Gbit/s, our simulations show almost equal

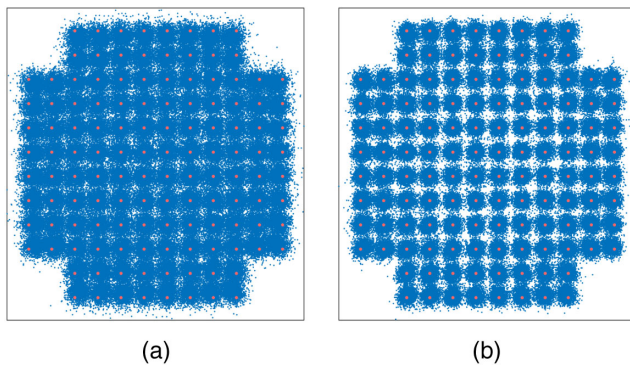


Fig. 5. Received constellations without (a) and with (b) ML-based spectrum equalization processing for the launch power of -18.5 dBm of 74 Gbit/s transmission.

BER for seven and 10 neighbors; therefore, we picked the number of neighbors to be the same for both data rates. The NN with the defined number of neighbors was trained using the received nonlinear spectrum $r(\xi)$ and initial spectrum as a target, and the procedure was done for every signal power and data rate separately. Then, the trained NN was applied to the test dataset (not used in the training) and the performance after the NN estimated in terms of BER, using an HD scheme. The results are presented in Fig. 3 by blue lines for both data rates. It can be seen in the figure that the improvements of BER for optimum powers are 8.6 and 5.75 times for 63 Gbit/s and 74 Gbit/s, respectively.

To analyze the impact of the size of the NN on performance, we varied the number of hidden layers and nodes in them. For optimal power of 74 Gbit/s transmission, performance of the NN consisting of two hidden layers with 64 nodes in each reduced by about 25% in comparison to the one described earlier, with a BER value lower than the HD FEC threshold and equal to 4.3×10^{-3} . On the other hand, by increasing the number of nodes in the hidden layers of the three-layer NN up to 96, we got only 6% BER improvement with a significant rise in computational complexity.

Additionally, to demonstrate the effect of the NN, in Fig. 5, we present the received 128-QAM constellation in (a) the absence and (b) presence of the NN equalizer with optimal power of -18.5 dBm. The improvement in system performance is visible in this figure. The BER value for the system without the NN equalizer is about 2×10^{-2} , whereas by employing the NN equalizer, it can be pushed below the HD FEC threshold and equals 3.5×10^{-3} . We note again that the NN-based nonlinear spectrum equalization can be useful for both types of noises in NFT systems: processing and ASE noise. This is the important benefit of the NN equalizer in comparison to more traditional ones, as the NN can target several degradation factors simultaneously.

We applied the NN-based equalizer at the receiver to improve the performance of the NFT-based optical transmission system. It is shown that by using this method, we can achieve almost six times BER improvement at 74 Gbit/s for the propagation distance of 1000 km at the optimal signal power. We believe that the proposed equalization method can be applied to other similar NFT-based optical communication systems, targeting different types of in-line and NFT computation-induced distortions, and also can be useful in increasing the accuracy of NFT procedures in general.

Funding. H2020 Marie Skłodowska-Curie Actions (GA-2015-713694, 751561); Engineering and Physical Sciences Research Council (Project TRANSNET EP/R035342/1); Leverhulme Trust (Grant RP-2018-063).

Acknowledgment. The research leading to these results has received funding from the EU Horizon 2020 Research and Innovation Program under the Marie Skłodowska-Curie grant agreements GA-2015-713694 (O.K.) and 751561 (M.P.), and the EPSRC Project TRANSNET (EP/R035342/1 [O.K., M.K.K., S.K.T.]). J.E.P. and S. K. T. acknowledge Leverhulme Trust Grant RP-2018-063.

Disclosures. The authors declare no conflicts of interest.

REFERENCES

- V. E. Zakharov and A. B. Shabat, *Soviet Phys. JETP* **34**, 62 (1972).
- A. Hasegawa and T. Nyu, *J. Lightwave Technol.* **11**, 395 (1993).
- J. E. Prilepsky, S. A. Derevyanko, K. J. Blow, I. Gabitov, and S. K. Turitsyn, *Phys. Rev. Lett.* **113**, 013901 (2014).
- M. Yousefi and F. Kschischang, *IEEE Trans. Inf. Theory* **60**, 4312 (2014).
- S. K. Turitsyn, J. E. Prilepsky, S. T. Le, S. Wahls, L. L. Frumin, M. Kamalian, and S. A. Derevyanko, *Optica* **4**, 307 (2017).
- E. G. Turitsyna and S. K. Turitsyn, *Opt. Lett.* **38**, 4186 (2013).
- S. Civelli, E. Forestieri, and M. Secondini, *IEEE Photon. Technol. Lett.* **29**, 1332 (2017).
- S. A. Derevyanko, J. E. Prilepsky, and S. K. Turitsyn, *Nat. Commun.* **7**, 307 (2016).
- M. Pankratova, A. Vasylichenkova, S. A. Derevyanko, N. B. Chichkov, and J. E. Prilepsky, *Phys. Rev. Appl.* **13**, 054021 (2020).
- F. Musumeci, C. Rottondi, A. Nag, I. Macaluso, D. Zibar, M. Ruffini, and M. Tornatore, *IEEE Commun. Surv. Tutorials* **21**, 1383 (2019).
- F. N. Khan, Q. Fan, C. Lu, and A. P. T. Lau, *J. Lightwave Technol.* **37**, 493 (2019).
- J. Mata, I. de Miguel, R. J. Durán, N. Merayo, S. K. Singh, A. Jukan, and M. Chamania, *Opt. Switch. Netw.* **28**, 43 (2018).
- D. Zibar, F. Da Ros, G. Brajato, and U. C. de Moura, *Opt. Photon. News* **31**, 34 (2020).
- R. T. Jones, S. Gaiarin, M. P. Yankov, and D. Zibar, *IEEE Photon. Technol. Lett.* **30**, 1079 (2018).
- O. Kotlyar, M. Pankratova, M. Kamalian, A. Vasylichenkova, J. E. Prilepsky, and S. K. Turitsyn, in *2018 IEEE British and Irish Conference on Optics and Photonics (BICOP)* (2018).
- O. Kotlyar, M. Kamalian, M. Pankratova, J. E. Prilepsky, and S. K. Turitsyn, in *European Conference on Optical Communication (ECOC)* (2019).
- M. Kamalian-Kopae, A. Vasylichenkova, O. Kotlyar, M. Pankratova, J. Prilepsky, and S. Turitsyn, in *2019 Conference on Lasers and Electro-Optics Europe and European Quantum Electronics Conference (CLEO/Europe-EQEC)* (IEEE, 2019).
- S. T. Le, V. Aref, and H. Buelow, *J. Lightwave Technol.* **36**, 1296 (2018).
- S. Wahls, in *2017 European Conference on Optical Communication (ECOC)* (2017).
- R.-J. Essiambre, G. Kramer, P. J. Winzer, G. J. Foschini, and B. Goebel, *J. Lightwave Technol.* **28**, 662 (2010).
- S. Wahls, S. Chimmalgji, and P. J. Prins, *J. Open Source Software* **3**, 597 (2018).
- T. Dozat, in *International Conference on Learning Representations (ICLR)* (2016).