Quantum Electronics **50** (12) 1105 – 1109 (2020)

https://doi.org/10.1070/QEL17463

Application of neural networks to determine the discrete spectrum of the direct Zakharov – Shabat problem

E.V. Sedov, I.S. Chekhovskoy, J.E. Prilepsky, M.P. Fedoruk

Abstract. **A neural network architecture is proposed to determine the number of solitons generated by random processes in optical wavelength-division multiplexed telecommunication systems with QPSK, 16-QAM, 64-QAM, and 1024-QAM modulation. The dependence of the prediction quality of a neural network with a special architecture on the number of soliton modes in the signal and the parameters of this signal is studied.**

Keywords: *nonlinear Schrödinger equation, inverse scattering problem method, Zakharov – Shabat problem, nonlinear Fourier transform, neural networks, machine learning, optical telecommunication systems, wavelength-division multiplexing.*

1. Introduction

Currently, there is an increased interest in optical telecommunications, since they are responsible for the transmission of more than 99% of global information traffic over transoceanic distances (1000 km or more). However, it is a matter of concern that a permanent increase in the volume of transmitted traffic may in the near future exceed the potential capabilities of information transmission lines based on modern technologies [1, 2]. In this regard, new promising ways to increase the capacity of communication lines are being actively explored [3]. Since the propagation of optical signals in optical fibres is described by the nonlinear Schrödinger equation (NSE), data transmission can be based on solitons, i.e. signals that do not change their shape during propagation [4]. To date, this idea has not been widely developed due to various limitations imposed on soliton communication lines, and major commercial technologies still use systems with wavelength-division multiplexing (WDM systems).

Although modelling optical channels has always been a difficult task, the use of coherent technologies to increase the information capacity of long-distance optical links has made the modelling of the behaviour of optical systems even more time-consuming. Many tunable parameters of fibre-optic communication lines (FOCLs), such as modulation formats, symbol rates, adaptive encoding rates, and adjustable channel separation, allow one to optimise data transmission systems, but require a large amount of calculations to determine the optimal values of the system parameters. In addition, the optimisation problem becomes even more challenging if the nonlinearity of the optical channel is taken into account.

Recently, there have been propositions to study modern signal modulation formats for the presence of solitons in them [5 – 8]. Such studies provide a better understanding of how telecommunication signals evolve over long distances and what causes nonlinear distortions that lead to loss of transmitted information. The method for studying signals in this problem is called the nonlinear Fourier transform (NFT). In fact, the name NFT is often found in the signal processing literature and serves to refer to the operations performed in the inverse scattering method used for integrating a special class of nonlinear equations. The inverse scattering method for NSE (that is, an explicit mathematical representation of NFT operations) was first described in the celebrated work by Zakharov and Shabat [9]. The direct NFT consists in attributing an optical signal to its nonlinear spectrum: the latter consists of (in the most general case) discrete and continuous parts [10]. The set of discrete eigenvalues corresponds to the soliton part of the signal. This representation is convenient, because the components (nonlinear modes), when applying the NFT to a signal with a finite norm, evolve in a trivial (linear) way as the signal propagates along a nonlinear optical fibre, while the discrete eigenvalues remain constant. For any value of the evolution variable, the signal can be completely restored using the inverse NFT if the propagation channel can be described well enough by the NSE. The NFT method is used for constructing analytical solutions of integrable equations [11, 12], as well as for analysing the contribution of solitons to signals obeying non-integrable equations that contain the NSE as one of the parts of a more general system [13 – 16].

The main difficulty in the wide application of the NFT method for the analysis of optical signals is the lack of fast and sufficiently accurate numerical methods for its implementation. Currently, there are a large number of methods for determining the nonlinear spectrum [17 – 20] and significant progress has been made in reducing the asymptotic complexity of algorithms (fast NFT) [21, 22] and improving their accuracy [23 – 25]. However, when applying the NFT to complex signals, problems may arise with the stability of

E.V. Sedov Aston Institute of Photonic Technologies, Aston University, Birmingham B4, 7ET, UK; Novosibirsk State University, ul. Pirogova 2, 630090 Novosibirsk, Russia;
I.S. Chekhovskoy Novosibirsk State University, ul. Pirogova 2, 630090 Novosibirsk, Russia; e-mail: i.s.chekhovskoy@nsu.ru;
J.E. Prilepsky Aston Institute of Photonic Technologies, Aston University, Birmingham B4, 7ET, UK;
M.P. Fedoruk Novosibirsk State University, ul. Pirogova 2, 630090 Novosibirsk, Russia; Institute of Computational Technologies, Siberian Branch, Russian Academy of Sciences, prosp. Akad. Lavrent'eva 6, 630090 Novosibirsk, Russia
Received 12 October 2020
Kvantovaya Elektronika **50** (12) 1105 – 1109 (2020)
Translated by M.A. Monastyrskiy

computational algorithms [26]. In addition, the NFT calculation for complex signal forms remains difficult, which limits the possibilities of implementing NFT at the hardware level for use in modern fibre-optic communication lines.

A promising direction in this case is the use of methods and systems for processing optical signals, based on the principles of machine learning, in particular using neural networks. In recent years, there has been a dramatic advance in the development of machine learning methods for solving algorithmically complex problems, such as image recognition and classification [27, 28]. The main steps in solving these problems are training the model using a set of certain data and applying it to obtain a prediction. The first stage may take a long time. However, the use of a trained model is usually much faster, which allows implementation of machine learning systems on various devices with low performance. We should also note that machine learning methods are used quite successfully to compensate for nonlinear effects that occur when a signal propagates in an optical fibre [29–32]. Recently, it was proposed to use machine learning in NFT-based data transmission systems at the post-processing stage [33]. In this work, we propose to implement a more radical approach and calculate the NFT itself using neural networks. Note that machine learning methods have already been used to process signals consisting only of soliton components when the number of soliton modes was small [34–36].

2. Nonlinear Fourier transform

The propagation of light in an optical fibre is well described using the NSE, which, with some restrictions, can be represented in a dimensionless form:

$$\mathrm{i}\frac{\partial q}{\partial z} + \frac{1}{2}\frac{\partial^2 q}{\partial t^2} + |q|^2 q = 0. \tag{1}$$

Here $q(z,t)$ is the slowly varying optical field in the fibre; z is the distance along the fibre axis; and t is time in the reference frame moving with the group velocity of the wave packet [37]. In this paper, we only consider the case of the focusing NSE for which there are soliton solutions, i.e. the sign of dispersion (anomalous dispersion) at the minimum-loss frequency corresponds to a standard single-mode fibre. To simplify the analysis, we do not take into account the gain and loss in the fibre, as well as the presence of noise components.

The NSE in form (1) belongs to the class of integrable equations that can be solved by the inverse scattering method [38]. The direct NFT allows one to determine the scattering data (characteristics of nonlinear modes) and consists in solving the Zakharov–Shabat spectral problem using a localised 'potential' $q(z,t)$, which is an optical signal:

$$\begin{aligned} -\partial_t\psi_1 + q(z,t)\psi_2 &= \mathrm{i}\lambda\psi_1, \\ \partial_t\psi_2 + q^*(z,t)\psi_1 &= \mathrm{i}\lambda\psi_2. \end{aligned} \tag{2}$$

Here ψ_i are auxiliary functions, and the complex parameter $\lambda = \xi + \mathrm{i}\eta$ is a nonlinear analogue of frequency. To determine the nonlinear spectrum associated with the profile $q = q(z,t)$, it is necessary to find a special solution $\Phi(t,\lambda) = \{\phi_1, \phi_2\}$ of system (2), satisfying the condition $\Phi \to \{\exp(-\mathrm{i}\xi t), 0\}$ at $t \to -\infty$. The main part of the direct NFT consists in calculating the scattering coefficients $a(\lambda)$ and $b(\lambda)$, defined using a special solution $\Phi(t,\lambda)$ as follows:

$$a(\xi) = \lim_{t\to+\infty}\phi_1(t,\xi)\exp(\mathrm{i}\xi t),$$

$$b(\xi) = \lim_{t\to+\infty}\phi_2(t,\xi)\exp(-\mathrm{i}\xi t).$$

We do not consider the continuous part of the nonlinear spectrum, and further analysis refers to the discrete part, i.e., the characteristics of solitons. The soliton modes correspond to discrete eigenvalues λ_n, which are set using the condition $a(\lambda_n) = 0$. The root of this equation is sought in the upper complex half-plane of the parameter λ, i.e. for $\lambda_n = \xi_n + \mathrm{i}\eta_n$ the condition $\eta_n > 0$ is satisfied. In fact, the presence of discrete eigenvalues that arise in the analysis of the Zakharov–Shabat problem with a potential in the form of our signal indicates the presence of soliton components in the signal.

3. Architecture of the neural network used for analysis of the soliton component of optical signals

Since linear, nonlinear, and noise effects manifest themselves simultaneously when we transmit data over fibre-optic communication lines, such systems are well fit to be dealt with using the latest advances in machine learning methods. Using these methods, it is possible to solve the problem of multidimensional optimisation (for example, in terms of data transmission quality and data throughput maximisation) without having to iterate through all possible parameter values.

Of particular relevance is the problem of identifying some internal features and patterns of the transmitted data, where neural networks can be used to simulate various effects that affect the signal when it propagates through a noisy nonlinear medium. In other words, neural networks can be used to simulate nonlinear transformations without the need for direct calculation of these transformations. The advantage lies in the speed and versatility of the transformation, as well as the flexibility and adaptability of operations based on a neural network: the network does not know what data it processes; it looks for the necessary features in the data that affect the final result, and then extracts them. This process is called feature extraction. Thus, if we want to calculate a certain value of a function, instead of (possibly) complex calculations, we can use a pre-trained network that, with a pre-known number of operations, will give the desired result. The difficulty is that the neural networks need to be trained up-front on the known data.

Another advantage of signal processing based on neural networks is that networks can reduce the noise component present in the analysed data [38]. In practice, we almost always encounter a situation where there is some noise in the data, for example, due to the finite accuracy of measurements, and its presence may be critical for accurate data processing methods. A neural network can effectively filter out unnecessary information within itself, leaving only the basic features needed for a specific task. Note that one of the disadvantages of using neural networks is the final accuracy of the result attainable with the use of a trained network. However, in practice, the accuracy of a neural network is sufficient for most tasks and sometimes even exceeds the accuracy of existing numerical methods if the necessary set of training data is available.

4. Discussion of results

In this work, we used a neural network to predict the number of discrete eigenvalues in a nonlinear spectrum of telecommunication signals. The discrete spectrum reflects the internal structure of the signal, and knowledge of this structure allows one to find out the signal properties and the features of its propagation over a nonlinear optical fibre.

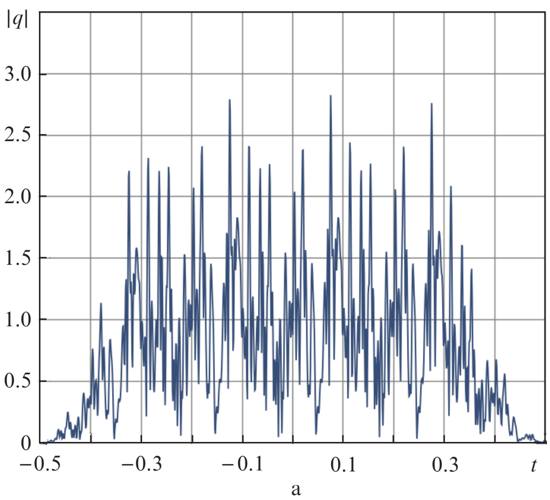
Optical signals encoded by the widely used WDM format were selected for the study. A single WDM symbol can be represented as the sum of independent optical carriers [8]:

$$s(t) = \sum_{k=1}^{M} C_k \exp(\mathrm{i}\omega_k t) f(t), \quad 0 \leqslant t < T, \tag{3}$$

where M is the number of WDM optical channels; ω_k is the carrier frequency of the kth channel; C_k corresponds to digital data in the kth channel (the value of this coefficient is a random variable from a set corresponding to the selected modulation format); T is the symbol interval; and $f(t)$ is the signal envelope with zero values at the edges of the symbol interval, the normalised expression for which without loss of generality can be written as

$$f(t) = \begin{cases} \frac{1}{2}\left[1 - \cos\left(\frac{4\pi t}{T}\right)\right], & 0 \leqslant t \leqslant \frac{T}{4} \ \text{or} \ \frac{3T}{4} \leqslant t \leqslant T, \\ 1, & \frac{T}{4} < t < \frac{3T}{4}. \end{cases} \tag{4}$$

The signal was generated from a random dataset encoded in one of the modulation formats for C_k: QPSK, 16-QAM, 64-QAM, and 1024-QAM, corresponding to 4, 16, 64, and 1024 possible values of C_k, selected from the 'constellation' on a complex plane. The number M of spectral channels for each signal from the set was one of the following: {9, 11, 13, 15, 17, 31, 51}. This set of values was taken to cover the number of channels used in existing WDM systems. An example of the WDM signal amplitude is shown in Fig. 1a, while an example of a discrete spectrum for such a signal is shown in Fig. 1b. The neural network architecture was based on a simplified version of the VGG-16 network [39], which is used in image recognition problems (Fig. 2a). Such architectures, where convolutional layers with the same number of input channels are sequentially arranged, demonstrate high efficiency by reducing the number of trained parameters while maintaining the overall prediction accuracy. To further improve the accuracy, it is necessary to increase the number of convolutional layers. This increases the learning time of the model; however, it allows us to select more features in the input data, and therefore improve the operation accuracy. The neural network input receives a complex signal consisting of 1 024 points. This signal is converted to a vector with 2 048 elements, in which the real and imaginary parts of each point of the original complex signal are sequentially arrayed. The signal is then processed by several convolutional layers with activation functions and then passed through the fully connected layers. The network output shows the number of solitons in the signal. The number of trained parameters in the network was 3 834 145.



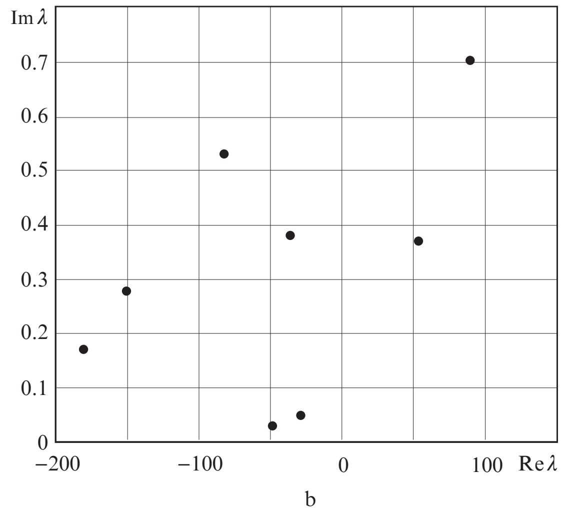


Figure 1. (a) Example of the dependence of the amplitude of the WDM signal under study on time (one of the possible implementations is given), and (b) example of the location of the discrete spectrum components in the complex half-plane of the spectral parameter λ for one of the signals under study.

In total, the training set consists of 174 847 generated signals, which contain from 0 to 20 solitons. The exact number of solitons in each signal was preliminarily calculated using a modification of the method of contour integrals [19], where the grid step in the spectral space was adaptively calculated depending on the number of solitons in the signal. To speed up the learning process on the training data set, for each signal we calculate only the number of discrete eigenvalues in the nonlinear spectrum, rather than the numerical value of each of the discrete spectral parameters corresponding to the soliton mode. A set of data with complete information on the nonlinear spectrum for each signal will be considered in subsequent works. The accuracy of network predictions was determined using a validation set of 19 427 signals (10 % of the total training set). The network

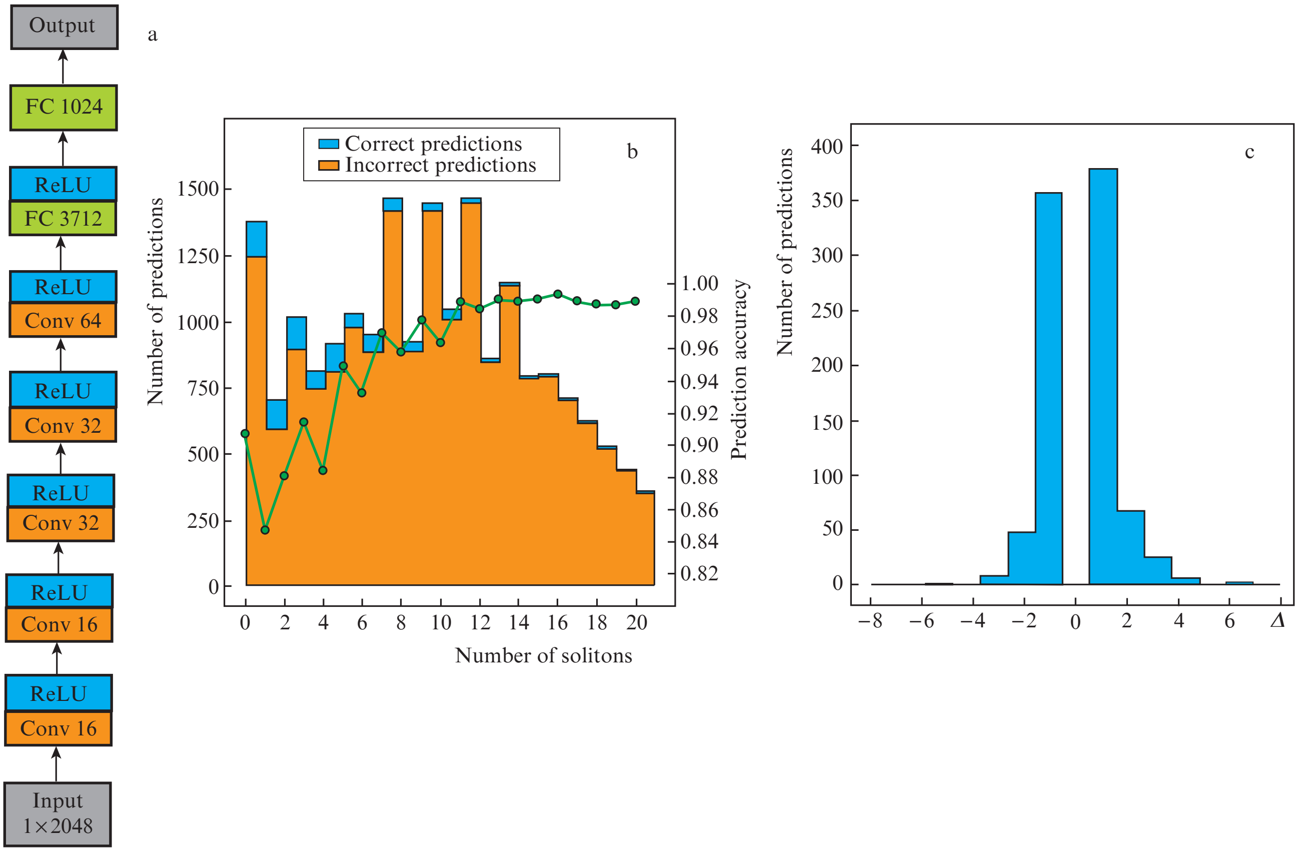


Figure 2. (Colour online) (a) Neural network architecture for predicting the number of discrete eigenvalues (solitons) in the nonlinear Fourier spectrum for a WDM signal [39], (b) the distribution of correct and incorrect predictions of the neural network as a function of the number of solitons in signals from the validation sample (the green curve shows the network prediction accuracy for each set of signals with the same number of solitons), and (c) the distribution of the deviation Δ of the predicted number of solitons in the signals from the validation sample versus their actual number.

was trained for 300 epochs, and the final prediction accuracy on the validation set was 95.39%. In the training process, we use for optimisation the Adam algorithm (adaptive moment estimation). The learning rate was varied during the training process from 10^{-3} at the beginning to 10^{-5} at the end. Its further reduction did not lead to an increase in the prediction accuracy.

Figure 2b shows the distribution of correct and incorrect neural network predictions as a function of the number of solitons in the signal for the validation sample. The green curve shows the neural network accuracy as a function of the number of solitons. The network worked best for signals where the number of solitons was greater than 10. For such cases, the accuracy exceeded 98%. Signals with a single soliton component were processed the worst, with an accuracy of 84%. In this case, the maximum 'error' of the network operation (the difference between the real number of solitons in the signal and the soliton number predicted by our neural network) was 8 (Fig. 2c). Negative error values correspond to the case when the network predicts a smaller number of solitons in the signal than the true number, and positive values occur when the neural network predicts a larger number of solitons than is present in the signal. Figure 2b shows the distribution of the deviation of the predicted number of solitons in the signals generated by the validation sample from their actual number: most of the incorrect results are in the range [–2; 2]. Thus, the neural network's prediction often deviates by a small value. This means that it captures general features indicating the number of solitons but cannot fully identify particular features. Obviously, with an increase in the number of convolutional layers, the neural network will be able to determine the internal features of signals more accurately, which means that its accuracy can be improved. The results show that neural networks have a great potential for implementing various stages of NFT with their help.

5. Conclusions

Thus, machine learning and neural network are modern technologies that are actively explored in nonlinear signal processing and for optical communication. The proposed neural network architecture demonstrates the fundamental possibility of its application for the analysis of complex optical signals. This opens up prospects for improving existing systems without the need for a deep understanding of the internal nonlinear processes that affect the signal transmission quality. We have found that modern neural networks can determine the internal structure of optical signals, and therefore can be used as a practical tool for their analysis. This stage is undoubtedly only the beginning of research on the possibility of using neural networks for optical communication. A promising direction is the development of autoencoders that will not only generate optical signals with the necessary parameters, but also select the optimal modulation and encoding formats. We stress that the method proposed in this work is only the first step in the development of machine learning methods for studying optical signals. The obtained results demonstrate that even a network with a small number of trainable parameters can identify complex nonlinear structures of optical signals with high accuracy.

Acknowledgements. The work was supported by the Fund of the President of the Russian Federation for State Support of Young Russian Scientists (Grant No. MK-677.2020.9). The work of I.S. Chekhovskoy was supported by the state assignment for fundamental research (FSUS-2020-0034), and the work of J.E. Prilepsky was supported by the Leverhulme Trust (Project RPG-2018-063).

References

1. Essiambre R.J., Foschini G.J., Kramer G., Winzer P.J. *Phys. Rev. Lett.*, **101** (16), 163901 (2008).
2. *Cisco Annual Internet Report (2018–2023). White Paper* (2020).
3. Winzer P.J., Neilson D.T., Chraplyvy A.R. *Opt. Express*, **26** (18), 24190 (2018).
4. Doran N.J. et al. *IEEE J. Quantum Electron.*, **19** (12), 1883 (1983).
5. Yousefi M.I. et al. *IEEE Trans. Inf. Theory*, **60** (7), 4312 (2014).
6. Le S.T., Prilepsky J.E., Turitsyn S.K. *Opt. Express*, **23**, 8317 (2015).
7. Sedov E.V., Redyuk A.A., Fedoruk M.P., Gelash A.A., Frumin L.L., Turitsyn S.K. *Opt. Lett.*, **43** (24), 5985 (2018).
8. Turitsyn S.K. et al. *J. Lightwave Technol.*, **38** (2), 352 (2020).
9. Zakharov V.E., Shabat A.B. *Zh. Eksp. Teor. Fiz.*, **61** (1), 118 (1972).
10. Turitsyn S.K., Prilepsky J.E., Le S.T., Wahls S., Frumin L.L., Kamalian M., Derevyanko S.A. *Optica*, **4** (3), 307 (2017).
11. Satsuma J., Yajima N. *Prog. Theor. Phys. Supp.*, **55**, 284 (1974).
12. Randoux S., Suret P., El G. *Sci. Rep.*, **6**, 29238 (2016).
13. Prilepsky J.E. et al. *J. Opt. Soc. Am. B*, **24** (6), 1254 (2007).
14. Chekhovskoy I.S., Shtyrina O.V., Fedoruk M.P., Medvedev S.B., Turitsyn S.K. *Phys. Rev. Lett.*, **122** (15), 153901 (2019).
15. Ryczkowski P., Närhi M., Genty G., Billet C., Merolla J.-M., Dudley J.M. *Nat. Photonics*, **12** (4), 221 (2018).
16. Sugavanam S., Kopae M.K., Prilepsky J.E., Turitsyn S.K., Peng J. *Nat. Commun.*, **10** (1), 5663 (2019).
17. Delves L.M., Lyness J.N. *Math. Comput.*, **21**, 543 (1967).
18. Boffetta G., Osborne A.R. *J. Comput. Phys.*, **102** (2), 252 (1992).
19. Burtsev S., Camassa R., Timofeyev I. *J. Comput. Phys.*, **147** (1), 166 (1998).
20. Vasylchenkova A., Prilepsky J.E., Turitsyn S.K. *Opt. Lett.*, **43** (15), 3690 (2018).
21. Wahls S., Poor H.V. *Proc. Int. Conf. on Acoustics, Speech and Signal Processing* (Vancouver, IEEE, 2013) pp 5780–5784.
22. Wahls S., Poor H.V. *IEEE Trans. Inf. Theory*, **61** (12), 6957 (2015).
23. Mullyadzhanov R., Gelash A. *Opt. Lett.*, **44** (21), 5298 (2019).
24. Medvedev S.B., Chekhovskoy I.S., Vaseva I.A., Fedoruk M.P. *Opt. Lett.*, **45** (7), 2082 (2020).
25. Medvedev S.B., Vaseva I.A., Chekhovskoy I.S., Fedoruk M.P. *Opt. Express*, **28** (1), 20 (2020).
26. Gelash A., Mullyadzhanov R. *Phys. Rev. E*, **101** (5), 052206 (2020).
27. Bishop Ch.M. *Pattern Recognition and Machine Learning* (New York: Springer, 2006).
28. Duda R.O. et al. *Pattern Classification* (New York: John Wiley & Sons, 2012).
29. Zibar D. et al. *Opt. Express*, **20** (26), B181 (2012).
30. Zibar D. et al. *J. Lightwave Technol.*, **34** (6), 1442 (2015).
31. Sidelnikov O.V. et al. *Quantum Electron.* **49** (12), 1154 (2019) [*Kvantovaya Elektron.*, **49** (12), 1154 (2019)].
32. Sidelnikov O.S., Redyuk A.A., Sygletos S. *Opt. Express*, **26** (25), 32765 (2018).
33. Kotlyar O. et al. *Opt. Lett.*, **45** (13), 3462 (2020).
34. Jones R.T., Gaiarin S., Yankov M.P., Zibar D. *IEEE Photonics Technol. Lett.*, **30** (12), 1079 (2018).
35. Wu Y., Xi L., Zhang X., Zheng Z., Wei J., Du S., Zhang W. *Opt. Express*, **28** (12), 18304 (2020).
36. Yamamoto S. et al. *IEICE Commun. Express*, **8** (12), 507 (2019).
37. Agrawal G.P. *The Nonlinear Fiber Optics* (Oxford, UK: Academic Press, 2007).
38. Tamura S. *Proc. Int. Conf. on Acoustics, Speech, and Signal Processing* (Glasgow, UK, 1989) pp 2001–2004.
39. Simonyan K., Zisserman A. arXiv preprint arXiv:1409.1556 (2014).