



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Neural network-aided receivers for soliton communication impaired by solitonic interaction

Citation for published version:

Chen, Y, Baniasadi, A & Safari, M 2023, 'Neural network-aided receivers for soliton communication impaired by solitonic interaction', *Optics Express*, vol. 31, no. 26, pp. 43289-43306.
<https://doi.org/10.1364/OE.499296>

Digital Object Identifier (DOI):

[10.1364/OE.499296](https://doi.org/10.1364/OE.499296)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Publisher's PDF, also known as Version of record

Published In:

Optics Express

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.





Neural network-aided receivers for soliton communication impaired by solitonic interaction

YU CHEN,  MOHAMMADAMIN BANIASADI, AND MAJID SAFARI*

Institute for Imaging, Data and Communications, School of Engineering, University of Edinburgh, Edinburgh, UK

*majid.safari@ed.ac.uk

Abstract: In this paper, different neural network-based methods are proposed to improve the achievable information rate in amplitude-modulated soliton communication systems. The proposed methods use simulated data to learn effective soliton detection by suppressing nonlinear impairments beyond amplifier noise, including intrinsic inter-soliton interaction, Gordon-Haus effect-induced timing jitter, and their combined impact. We first present a comprehensive study of these nonlinear impairments based on numerical simulations. Then, two neural network designs are developed based on a regression network and a classifier. We estimate the achievable information rates of the proposed learning-based soliton detection schemes as well as two model-based benchmark schemes, including the nonlinear Fourier transform eigenvalue estimation and continuous spectrum-aided eigenvalue estimation schemes. Our results demonstrate that both learning-based designs lead to substantial performance gains when compared to the benchmark schemes. Importantly, we highlight that exploiting the channel memory, introduced by solitonic interactions, can yield additional gains in the achievable information rate. Through a comparative analysis of the two neural network designs, we establish that the classifier design exhibits superior adaptability to interaction impairment and is more suitable for symbol detection tasks in the context of the investigated scenarios.

© 2023 Optica Publishing Group under the terms of the [Optica Open Access Publishing Agreement](#)

1. Introduction

The spectral efficiency (SE) of conventional standard single-mode (SSM) fibers is often constrained by the Kerr nonlinearity, a signal-dependent effect that intensifies as the launch power increases [1]. Predicted by Zakharov and Shabat, the existence of the soliton pulse is first demonstrated via numerical simulation by Hasegawa [2]. The stability of the soliton evolution in nonlinear fiber shows its ability of balancing linear dispersion and Kerr nonlinearity, hence implying its potential application as information carrier [3]. However, it was pointed out that the amplifier noise will introduce Gordon-Haus timing jitter, which will limit the data rate of the proposed on-off keying soliton communication [4].

It was not until the new emergence of nonlinear Fourier transform (NFT) that the soliton communication started to regain some research interests. To potentially harness the nonlinearity effects in long-haul fibers, many recent works have employed the NFT to map a fiber-propagated optical signal from time domain to its degrees of freedom defined in the so-called nonlinear frequency spectrum: the discrete spectrum (DS) and the continuous spectrum (CS) [5–9]. Specifically, the DS signifies the solitonic component of the optical signal, while the CS corresponds to the non-solitonic radiation.

In a lossless and noise-free nonlinear fiber, the nonlinear channel effect governed by nonlinear Schrödinger equation (NLSE) becomes linearized in the nonlinear frequency domain. The eigenvalues of the DS remain invariant, while the evolution of both the norming constants of DS and CS can be described using a linear phasor [5]. Based on this property, nonlinear frequency division multiplexing (NFDm) has been proposed as a promising candidate to overcome the limitations imposed by nonlinearity [6]. The NFDm system encodes information in the nonlinear

frequency domain, similar to orthogonal frequency division multiplexing (OFDM). In particular, using invariant eigenvalues as information carriers has attracted interest within the community. However, determining the capacity of such an eigenvalue communication system, such as an amplitude-modulated (AM) single soliton transmission system, remains an open challenge. The complexity arises from nontrivial signal-dependent noise and other nonlinear impairments, including the Gordon-Haus effect and inter-soliton interaction.

In long-haul fiber transmission, the necessity of using optical amplifiers inevitably introduces amplifier spontaneous emission (ASE) noise. The ASE noise, which is added to the optical signal in the time domain, gives rise to a non-trivial signal-dependent noise in the received eigenvalue, i.e., in nonlinear frequency domain [10,11]. Further complexities arise when a sequence of solitons is transmitted. Intrinsic forces between solitons can result in inter-soliton interactions when they are temporally proximate [12], potentially causing interference between neighboring symbols and, consequently, errors in communication. The situation is further exacerbated by the ASE noise-induced Gordon-Haus timing jitter [4], which increases the chances of neighboring solitons becoming temporally close, thereby intensifying their interactions. Unfortunately, to the best of our knowledge, the existing literature offers inadequate investigation of these compounded impairments.

Based on the perturbation theory, the transmission of an amplitude modulated soliton pulse in isolation can be described by a signal-dependent noise model [10]. In addition, variance normalizing transform (VNT) can be used to effectively approximate such signal-dependent noise models with equivalent unit-variance additive white Gaussian channels (AWGN) [10,11,13]. However, it has been observed that as the operational power escalates, inter-soliton interaction starts to dominate over the signal-dependent noise, rendering the isolated soliton channel model ineffective [13]. Research has suggested that significant improvements in the detection of soliton sequences can be achieved by exploiting the correlation between the eigenvalue and the CS [14]. Furthermore, an intriguing concept of beneficial interaction has been introduced, proposing that a specially designed signaling scheme, that allows for precise post-interaction soliton position swaps, can be advantageous [15].

In addition to the aforementioned conventional detection schemes, machine learning has shown promising gains in detecting signals within highly nonlinear channels where precise channel models are unavailable. For example, the bidirectional long short-term memory (BLSTM) network has been found effective in CS-modulated NFDN systems where the exact channel model with memory remains unknown [16]. A two-stage neural network receiver scheme has been proposed for performing joint channel equalization in both the time and nonlinear frequency domains before and after the NFT, resulting in improved error performance [17]. In [18], an autoencoder design based on unsupervised neural networks was utilized to adaptively adjust the DS modulation scheme in response to data generated according to the NLSE model of the optical fiber channel. Classifier neural networks have been employed in DS NFDN systems to detect high-order solitons corresponding to two distinct eigenvalues [19]. In [20], a regression neural network was employed as an equalizer to compensate for channel perturbations in a b coefficient-modulated multi-eigenvalue DS NFDN systems. A comparison was made between classifier and regression networks in [21]. This comparison study highlights that, in high-accuracy systems, the use of a classifier network design with a non-smooth cross-entropy loss function can lead to learning problems such as gradient vanishing, thus limiting the achievable information rate (AIR) of the network. It was also suggested that implementing a regression network with a smooth mean square error (MSE) loss and early stopping criteria may overcome these learning problems, achieving superior error performance and AIR.

In this study, we develop different neural network based soliton detection schemes, analyze their performance in terms of achievable information rate, and demonstrate their superiority against best existing model-based techniques. The key contributions of the paper are listed below:

- We present a numerical study of the effect of intrinsic inter-soliton interaction using numerical simulations, followed by an examination of the Gordon-Haus effect and finally their combined impact.
- Noting that the inter-soliton effects either individually or combined with Gordon-Haus jitter are nonlinear, complex and non-tractable, we propose two neural network designs based on regression and classification BLSTM networks for soliton detection and optimize their design by adjusting the different network parameters.
- We estimate the achievable information rates of the two proposed neural network designs and compare them with two model-based soliton detection schemes, where our analysis suggests a distinct preference for classification neural network design.
- Our AIR results reveal that both classifier and regression network designs of the BLSTM network for soliton detection can outperform benchmark model-based schemes, potentially reaching the upper bound of the SE in low-to-mid power regimes. However, in higher power regimes, the expected gains over the benchmark methods are modest, indicating its struggle to cope with the increased interaction and noise.
- Beyond the standard hyperparameter analysis (e.g., different quantities of hidden units and BLSTM layers), we analyze the length of the training soliton sequences, which unveils the memory size of the underlying nonlinear channel, further contributing to our understanding of these systems.

2. System model

Adopting the normalization scheme as presented in [13], the unitless stochastic NLSE, that precisely models the propagation of optical signals over nonlinear long-haul fibers, is given by

$$jq_z(t, z) = q_{tt}(t, z) + 2|q(t, z)|^2q(t, z) + n(t, z), \quad (1)$$

where $q(t, z)$ (normalized by $\sqrt{|\beta_2|/\gamma T_0^2}$) is the normalized complex envelope of the optical field with the subscript z and tt denotes the first and second order partial differentiation with respect to z and t , respectively, while $n(t, z)$ represents the normalized ASE noise. The variables t and z denote unitless time (normalized by T_0) and propagation distance (normalized by $2T_0^2/|\beta_2|$), respectively. The autocorrelation of the normalized ASE noise is given by $E[n(t, z)n^*(t', z')] = \sigma^2\delta(t-t')\delta(z-z')$, where $\sigma^2 = \alpha h\nu_0 K_T \frac{2\gamma T_0^3}{|\beta_2|^2}$. The parameters α , β_2 and γ describe the fiber loss, group velocity dispersion and Kerr nonlinearity, respectively. Furthermore, $h\nu_0$ denotes the photon energy, while K_T is the phonon occupancy factor.

The NFT transforms the time domain signal into scattering data which evolves linearly in the nonlinear NLSE channel. The scattering data includes the CS, $\rho(\lambda, z)$, and the DS, $\{\lambda_m(z)_{m=1}^M, C_m(z)_{m=1}^M\}$. The DS contains eigenvalues λ_m and corresponding norming constants C_m pairs. If only the imaginary part of a single eigenvalue is employed and the soliton is centered at time zero, the solution of (1) is known as a first order soliton, given by

$$q(t, 0) = 2A\text{sech}(2At), \quad (2)$$

where $A = \Im(\lambda_1) > 0$ with the norming constant of $-2A$ at transmitter $z = 0$. Note that the soliton pulse (2) contains a sech hyperbolic shape in both time and frequency domains, dependent on its amplitude A . For practical implementation of the soliton, the pulse should be truncated within the symbol period timewidth of $[-\frac{t_w}{2}, \frac{t_w}{2}]$. Letting δ define the fraction of the soliton energy that

is outside $[-\frac{t_w}{2}, \frac{t_w}{2}]$, the relationship between t_w , δ , and the soliton amplitude, A , can be written as [13]

$$t_w(A, \delta) = \frac{1}{2A} \ln \left(\frac{2}{\delta} - 1 \right). \quad (3)$$

Similarly, the bandwidth of the soliton that preserves $(1 - \epsilon)$ of energy could be defined as

$$f_w(A, \epsilon) = \frac{2A}{\pi^2} \ln \left(\frac{2}{\epsilon} - 1 \right). \quad (4)$$

The $(1 - \delta)$ timewidth in (3) and $(1 - \epsilon)$ bandwidth in (4) effectively define the resources occupied by a single soliton pulse with amplitude A . The description of occupation of time and spectral resources for a single soliton amplitude can be extended to a constellation of soliton amplitudes, where the timewidth and the occupied bandwidth for the signaling are determined by the minimum and maximum amplitude of the constellation, A_{\min} and A_{\max} , respectively. Similar to [14], we consider a fixed time-bandwidth product to estimate the system performance with limited resources.

To fully utilize available resources and maximize the SE, constellation shaping can be applied to achieve an optimal constellation. The optimization problem corresponding to such shaping approach is formulated as

$$C = \sup_{P_A(a): A \in \{0\} \cup [A_{\min}, A_{\max}]} I(A; R), \quad (5)$$

where $I(A; R)$ indicates the mutual information (MI) between transmitted and received eigenvalues A and R . Based on the perturbation theory and the assumption of isolated soliton pulse transmission, the channel between A and R can be described by a noncentral chi-squared model [10]. However, applying the VNT transformation $T(u) = 2\sqrt{2u/\sigma^2}l$ as in [13], where l denotes the unitless propagation distance, allows us to simplify the estimation of the mutual information $I(A; R) = I(T(A), T(R)) = I(X; Y)$ assuming that the VNT transformed channel between X and Y is approximately Gaussian and independent of input. This assumption is validated in [13] by showing that the Kullback–Leibler (KL) divergence between the approximated AWGN channel and the VNT transformed noncentral chi-squared channel is close to zero in the regime of interest. Note that the range of soliton amplitudes is defined by the effective time-bandwidth product,

$$\text{TB}(A_{\min}, A_{\max}, \delta, \epsilon) = \frac{1}{\pi^2} \frac{A_{\max}}{A_{\min}} \ln \left(\frac{2}{\delta} - 1 \right) \ln \left(\frac{2}{\epsilon} - 1 \right), \quad (6)$$

where the energy truncation factors δ and ϵ could be independently selected. In this work, unless stated otherwise, we set δ and ϵ to be 10^{-2} , which means 99% of the soliton energy should be preserved in both time and frequency domains. Given the time-bandwidth product TB and available bandwidth f_w , (6) allows for the calculation of A_{\min} and A_{\max} , which define the range of symbol amplitudes comprising the constellation excluding the zero symbol. Note that the inclusion of zero symbol as part of the constellation is a reasonable design choice that expands the source entropy without affecting TB. In addition, the NFT detection of the zero symbol would be highly resilient against ASE noise, which does not commonly exceed the threshold to support the generation of an erroneous eigenvalue [5]. Finally, the zero symbol could also provide effective guarding between its neighbor solitons, potentially reducing the overall impact of inter-soliton interaction.

3. Impairments beyond additive noise

In much of the existing literature, solitons are typically assumed to be transmitted with ample separation between each other, resulting in the dominant perturbation being additive signal-dependent noise. However, in practical implementations of the AM soliton communication system,

solitons are transmitted sequentially. This results in the potential for attractive forces between solitons, even in the absence of ASE noise perturbation [12]. Such intrinsic interactions can soon become dominant over signal-dependent ASE noise [13] at higher soliton transmission rates. Given the non-tractable nature of the formulation of this problem, this section will first primarily focus on a numerical study of this impairment in the absence of ASE noise. In the numerical visualization of the impairments, the fiber with dispersion coefficient of $\beta_2 = -2.1 \times 10^{-26} \text{ s}^2/\text{m}$, nonlinearity factor of $\gamma = 1.27 \times 10^{-3} \text{ /W/m}$ and Raman amplifier compensated loss of $\alpha = 0.2 \text{ dB/km}$ is simulated with split-step Fourier method. When signaling the solitons, the 99% timewidth is chosen, and the number of samples per soliton is selected to be 64. Additionally, note that the phonon occupancy K_T of 1.13 and wavelength of $1.55 \text{ }\mu\text{m}$ are selected.

To isolate and visualize the intrinsic interaction and separating it from the ASE noise perturbation, we simulate the propagation of a sequence of three randomly selected solitons in a noiseless NLSE channel. Using a simple 4-pulse amplitude modulation (PAM) constellation of $[0, 0.75, 1, 1.25]$ and $t_w(0.75, 10^{-2})$ as one of the possible scenarios, we generate a series of example combinations of three-soliton sequences, as shown in Fig. 1(a) to 1(d). The constellation considered in this set of examples ensures well-separated amplitudes, allowing effective guarding intervals to be imposed for larger soliton amplitudes. It can be observed from Fig. 1(a) and 1(b) that when solitons with the smallest amplitude (i.e., widest pulse duration) are transmitted successively, there is effectively no guarding interval between the solitons due to truncation. In this case, the inter-solitic force causes the solitons to collide after propagation over the fiber, introducing a severe impact on the integrity of the soliton pulses. On the other hand, the combination $[1, 1, 1]$ sketched in Fig. 1(c) shows the neighboring soliton leaning towards the center soliton due to the effect of the inter-soliton force, the solitons are overall less impaired by the inter-solitic force owing to the effective guarding intervals.

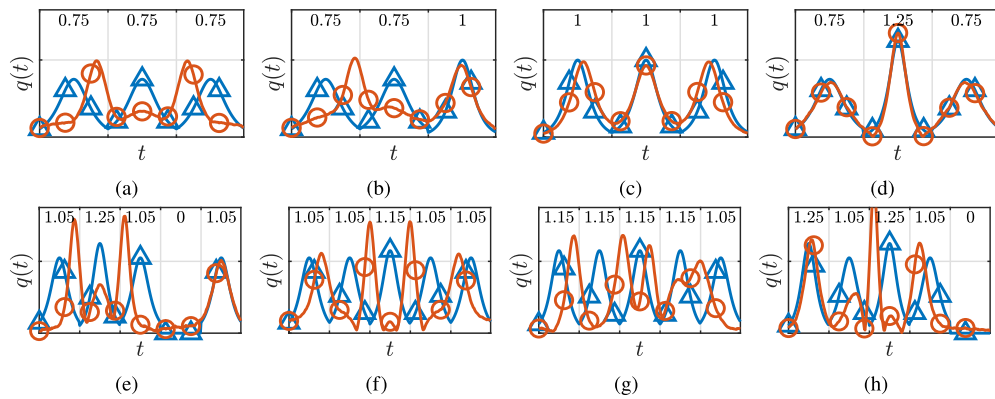


Fig. 1. Examples of intrinsic inter-soliton interaction occurs between three-soliton sequence ((a)-(d)) and five-soliton sequence ((e)-(h)). Blue lines indicate the transmitted pulses while red lines denote the received pulses. The three-soliton sequences are drawn from the constellation $[0, 0.75, 1, 1.25]$, while the five-soliton sequences are drawn from a more closely-spaced constellation $[0, 1.05, 1.15, 1.25]$.

By implementing a more closely-spaced constellation that takes advantage of temporal resources more efficiently and using a longer soliton sequence, we can observe more complex, non-tractable interactions, as shown in Fig. 1(e) to 1(h). In this set of simulations, five solitons are randomly drawn from a constellation $[0, 1.05, 1.15, 1.25]$ with smaller spacing between the non-zero constellation points. It becomes clear that the memory effect of the inter-soliton interaction can easily exceed a soliton sequence length of three. Longer solitons and denser constellations will result in more intricate interaction patterns. Note that although not shown

in the example figure, it is demonstrated with the same simulator that when transmitted with neighboring soliton amplitudes equal to zero, the information soliton is well preserved. This suggests that the distortion shown in Fig. 1 is primarily generated by the soliton interaction.

The presence of ASE noise induces the Gordon-Haus effect, which occurs due to random fluctuations in the group velocity of solitons, causing them to arrive at the receiver with random timing jitter [4]. Hence, it introduces a constraint on the maximum allowable soliton transmission rate for a given propagation distance [4]. To isolate the intrinsic inter-soliton interaction from coupling with the Gordon-Haus effect, three-soliton sequences with the first and last solitons set to zero are launched into an ASE noise-perturbed NLSE fiber. This experimental setup essentially simulates the propagation of a single soliton centered at time 0 with neighboring zero solitons as a guarding interval on both sides. By performing the NFT over the sampling window of $3t_w$, which covers the entire soliton sequence, a single eigenvalue and its corresponding norming constant can be extracted. The center of the received soliton is then estimated as [22,23]

$$t_{GH} = \frac{1}{2R} \log \frac{|C_1|}{2R}, \quad (7)$$

where C_1 denotes the detected norming constant for the single eigenvalue and R denotes the estimated amplitude of the received soliton (i.e. the imaginary parts of the eigenvalue). Since the soliton is initially centered at time 0 at the transmitter and there is no intrinsic interaction in the simulation, t_{GH} can be interpreted as the timing jitter introduced by the Gordon-Haus effect. The amount of the Gordon-Haus timing jitter is estimated and depicted in Fig. 2. It is observed that the amount of timing jitter introduced is dependent on the soliton amplitude, as larger variance of t_{GH} observed for larger solitons. The dependence on soliton amplitude has been shown to be linear using perturbation theory over the general solution of a propagated single soliton [4,24].

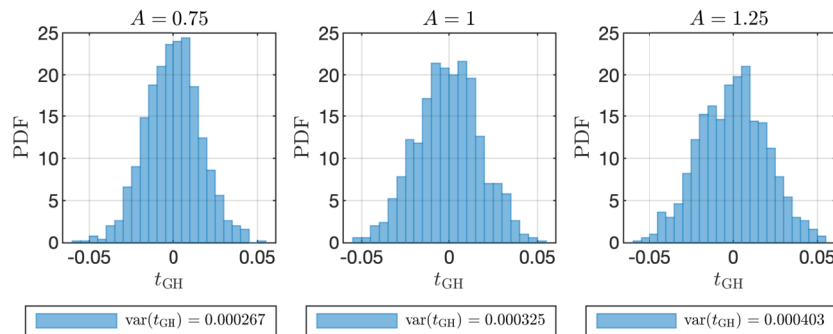


Fig. 2. Histograms of the Gordon Haus timing jitter t_{GH} of transmitting 0.75, 1 and 1.25 with zero solitons as guarding intervals on both sides. The NFT detection is performed on the whole sequence with a $3t_w$ sampling window.

The nature of the timing jitter is the added randomness of the position of the soliton. Such random position can increase the probability of error via two mechanisms, which are referred to as out-of-window error and enhanced interaction, respectively. In a practical implementation of soliton communication, NFT is expected to be applied to a limited sampling window, while the random timing jitter could shift the soliton away from such sampling window, resulting in excessive error beyond the original ASE noise effect. Hence, this type of error is defined as out-of-window error. To capture the isolated impact of the out-of-window errors here, NFT is applied on both the information soliton (sampling window being t_w) and the whole three-soliton sequence (sampling window being $3t_w$). The two neighboring solitons are set to zero, similar to the previous experiment, which ensures that the NFT over the whole sequence is free from out-of-window noise. Meanwhile, the NFT over the t_w sampling window provides a more realistic

representation of practical NFT detection. Therefore, the difference between the results of these two detection methods can be interpreted as the excessive out-of-window error introduced by the random jitter. Denoting this excessive out-of-window error as N_w , the histograms of N_w given different solitons are transmitted are illustrated in Fig. 3. In the fixed soliton interval system considered in this work, larger solitons correspond to larger effective guarding intervals compared to smaller solitons. Consequently, it is estimated that the out-of-window error power for the smallest soliton of 0.75 is almost 100 times that of the largest soliton of 1.25 from the constellation. Although not shown explicitly in this figure, the level of out-of-window error power is less significant compared to the ASE noise power in the regime of interest in this study. Recall that the capacity lower bounds estimated with isolated soliton transmission reported in [13] also suggest that when there is no solitonic interaction, the data rate of the system is primarily limited by the ASE noise rather than the out-of-window error introduced by the Gordon-Haus effect.

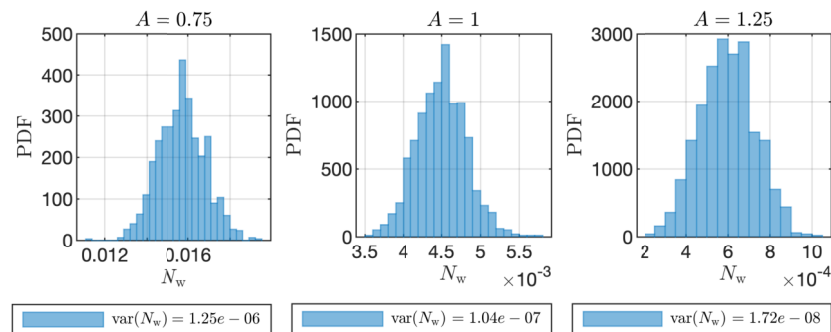


Fig. 3. Histograms of the out-of-window error N_w of transmitting 0.75, 1 and 1.25 with two zero solitons as guarding intervals on both sides. The NFT detection is performed on both the center soliton, with a t_w sampling window, and the whole sequence, with a $3t_w$ sampling window, to estimate the N_w .

As mentioned, the timing jitter can also increase the probability of error via a second mechanism, i.e., enhanced interaction, which occurs when the timing jitter drives the soliton towards its successive solitons during sequential transmission. When combined with the intrinsic interaction forces exerted between solitons, this leads to enhanced inter-soliton interaction. However, visualizing such an effect is challenging due to the coupling between the intrinsic interaction and Gordon-Haus induced interaction. Figure 4 illustrates 1000 transmissions of a three-soliton sequence over a noisy NLSE channel using the constellation $[0, 1.75, 1.9, 2.25]$, which has higher power compared to the previously considered constellations. We selected two sequences as examples: $[1.75, 1.75, 1.75]$ to demonstrate the performance of a sequence with high intrinsic inter-soliton interaction under both ASE noise and the Gordon-Haus effect, and $[1.9, 1.9, 1.9]$ as a representative of a sequence with less intrinsic inter-soliton interaction. For the sequence $[1.75, 1.75, 1.75]$, we observed that when coupled with the Gordon-Haus induced interaction, the received pulse is more distorted compared to the noiseless received pulse, indicating the effect of intensified interaction by Gordon-Haus effect. For the sequence $[1.9, 1.9, 1.9]$, because of the effective guarding intervals between solitons, less intrinsic interaction is observed by comparing the transmitted pulse and received pulse under a noiseless channel. The effective guarding intervals also allow the pulse to be less impaired by the Gordon-Haus enhanced interaction. As demonstrated in the examples, it is observed that the intrinsic forces between solitons changes the temporal position of each soliton deterministically, while the Gordon Haus effect would induce a random temporal position shift. Such random jitter can increase or decrease the distance among soliton pulses, which would in turn attenuate or amplify the intrinsic interaction effect. Due to the strong nonlinear coupling between these random and deterministic effects, we mainly

focus on the combined nonlinear behavior of the underlying system and employ a neural network design that can learn the complexity of optimal detection in the presence of both Gordon Haus and intrinsic interaction effects.

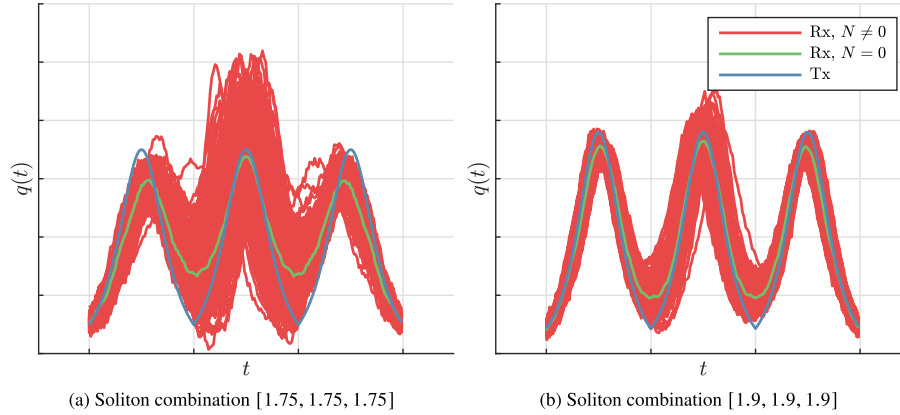


Fig. 4. Examples of transmitted and received three-soliton sequences through a noisy, lossless nonlinear fiber with 4-PAM of constellation of $[0, 1.75, 1.9, 2.25]$. The received pulse over a noiseless fiber is also included for reference, and the edges of each soliton segment are highlighted with grid lines.

4. Soliton detection schemes

In this paper, we propose two learning based schemes for soliton detection and estimate their AIR using suitable auxiliary mismatch decoding schemes [25]. Before introducing our NN-based detection schemes, we describe two recently proposed soliton detection techniques that are used as benchmarks in our AIR performance analysis in the next section. Note that soliton communication using the energy detection scheme was a popular design choice before the development of NFT, because of its simplicity in implementation. However, as the energy detection scheme is not as efficient as the presented benchmark schemes, it will not be included in the discussion. The block diagram of the soliton communication system employing the two proposed NN-based detection schemes, and the two benchmark schemes is sketched in Fig. 5. The figure also includes an example of a five-soliton sequence evolution along the fiber.

4.1. Direct NFT detection

The first benchmark detection scheme is direct NFT detection after applying VNT, which is used to mitigate the signal dependency of noise [13]. In this detection scheme, the soliton sequence is divided into segments of equal length based on the specified timewidth t_w . The NFT is performed on each segment to extract the eigenvalue R . Subsequently, VNT is applied as $Y = T(R)$ to reduce the signal dependency of noise [13]. The AIR for this scheme can be estimated as

$$I_{\text{iso}} = E_{X,Y} \left[\log \frac{Q_1(y|x)}{Q_1(y)} \right], \quad (8)$$

where a unit-variance AWGN channel $Q_1(y|x)$ is used as the auxiliary mismatch channel and the numerical expectation is taken over a large number of channel realizations. It has been shown that the unit-variance AWGN channel provides a tight lower bound when solitonic interaction is not taken into account [13].

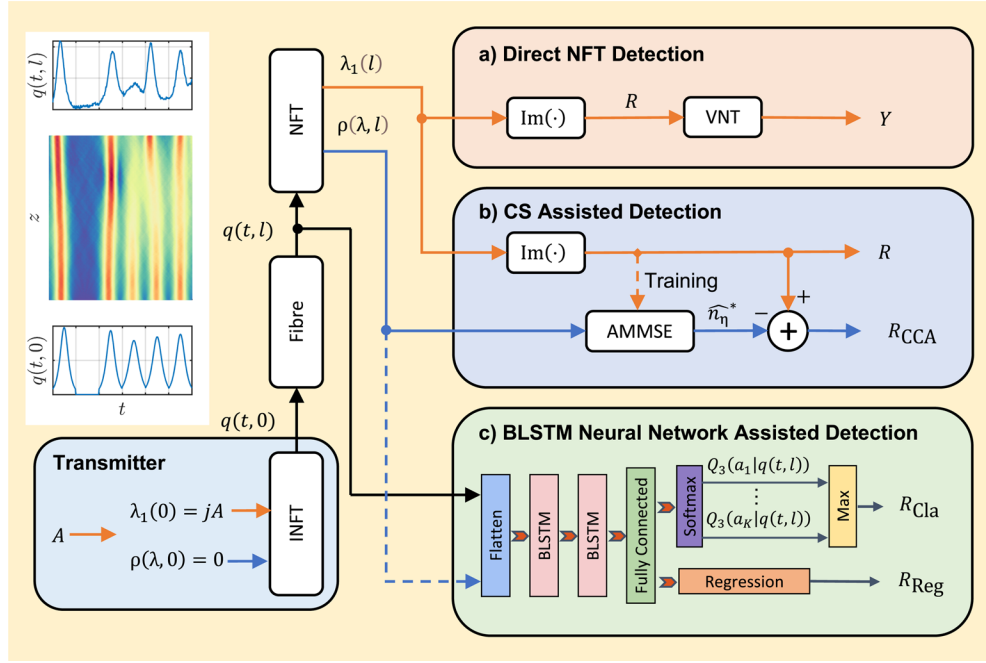


Fig. 5. The system block diagram for the three categories of detection schemes discussed. Examples of transmitted and received soliton sequence pulse shape are presented along with the corresponding pulse evolution.

4.2. CS assisted detection

The second benchmark detection scheme aims at exploiting the potential correlation between the eigenvalue distortion (i.e., soliton amplitude noise) and the excessive CS generated at the end of fiber [14]. The estimated correlation is then used to compute an affine minimum mean squared error (AMMSE) estimation of the noise [14]. The AMMSE estimator consists of the weight \mathbf{W}^* and bias B^* that minimize the mean squared error between the affine transformed CS and soliton amplitude noise as

$$[\mathbf{W}^*, B^*] = \arg \min_{[\mathbf{W}, B]} E_{A, R, \rho'} |(R - A) - (\mathbf{W}\rho'(\lambda, l) + B)|^2, \quad (9)$$

where the received CS spectrum $\rho(\lambda, l)$ is rewritten as $\rho'(\lambda, l) = [\Re(\rho(\lambda, l)), \Im(\rho(\lambda, l))]$ with the dimension of $2N_{CS} \times 1$ and N_{CS} denotes the number of samples taken in CS domain. Using a large number of channel realizations for more reliable correlation estimation, canonical correlation analysis (CCA) is employed to identify the AMMSE estimator parameters [14]. If the AMMSE estimated noise is removed from the received soliton amplitude R , the resultant received symbol $R_{CCA} = R - (\mathbf{W}^*\rho'(\lambda, l) + B^*)$ will have lower noise power, notwithstanding the effect of the solitonic interaction [14]. To formulate a comparable performance metric for this scheme, the AIR for this scheme could be estimated as

$$I_{CCA} = E_{A, R_{CCA}} \left[\log \frac{Q_2(r_{CCA}|a)}{Q_2(r_{CCA})} \right], \quad (10)$$

where a Gaussian channel with statistics estimated from the training data is selected to be the auxiliary channel model $Q_2(r_{CCA}|a)$, and the expectation should be taken on a different set of channel realizations to show transferability. Note that the interaction of the soliton is not taken

into account in I_{iso} since the auxiliary channel only characterizes the channel model for isolated soliton communication systems. However, it was demonstrated that the AMMSE method could take advantage of CS generated by interaction to some extent [14]. It should be noted that the outputs of the benchmark schemes are interpreted as the combination of transmitted symbol and noise, which requires further decision making scheme, such as maximum likelihood based on the auxiliary channel to obtain the estimated transmitted symbols.

4.3. NN-based soliton detection

As discussed in the last section, the inter-soliton interaction effect either individually or when combined with Gordon-Haus effect is non-tractable and highly nonlinear. Therefore, we propose the use of NN-based soliton detection schemes using a BLSTM network classifier to learn the nonlinear patterns of eigenvalue distortions when sequences of closely spaced solitons are transmitted. In this scheme, the sampled time-domain soliton pulse is first fed into the sequence input layer, which matches the number of extracted features. The input layer is then connected to two BLSTM layers to further explore the relevant correlation between samples. Finally, the outputs of the BLSTM layers are fully connected to a softmax layer. The output of the softmax layer can be interpreted as the a posteriori probability $Q_3(A|q(t, l))$, which implies the probability of transmitting the symbol A given $q(t, l)$ is received [21]. The detected symbol R_{Cla} can be obtained after a hard decision by taking the maximum a posteriori probability among all. Note that the other detection schemes will require additional mapping to individual symbols to obtain the estimation of transmitted symbol. The a posteriori outputs of the softmax layer could also be interpreted as auxiliary mismatch decoding rules, which are then used to estimate AIR [21]. The AIR of BLSTM classifier detector I_{Cla} can be calculated as

$$I_{\text{Cla}} = H(A) - E_{A, R_{\text{Cla}}} \left[\log \frac{1}{Q_3(a|R_{\text{Cla}})} \right], \quad (11)$$

where $H(A)$ denotes the source entropy and the expectation is also taken on a separate test set that was not used in either the training or validation of the network.

The BLSTM network could also be designed to perform a regression-type task for estimating the possible transmitted symbol from the received soliton pulse. It is highlighted that the training of network could benefit from the better gradient behavior of the smooth loss function in the regression network [21]. It has been shown that the regression network, with a Gaussian auxiliary channel, could result in higher AIR over the classifier network, which adapts to the learning data's channel statistics, in some scenarios [21]. Therefore, the second proposed NN-based detection scheme considered in this work is a BLSTM regression network detector. Compared to the NN-based detection scheme with the BLSTM classifier, it shares the sequence input layer, two BLSTM layers, and a fully connected layer. Information from the fully connected layer is then fed into the regression layer to generate the estimated transmitted soliton amplitude R_{reg} . Unlike the classifier output, where the detected symbol decision is determined by taking the maximum softmax layer output, the regression network output R_{reg} requires additional mapping following the mismatch decoding rules similar to the benchmark schemes. The AIR of this detection scheme, I_{Reg} , is estimated similarly to the previous AMMSE scheme with a Gaussian auxiliary channel [21]. The expression of the AIR is given as

$$I_{\text{Reg}} = E_{A, R_{\text{Reg}}} \left[\log \frac{Q_4(r_{\text{Reg}}|a)}{Q_4(r_{\text{Reg}})} \right], \quad (12)$$

where the $Q_4(r_{\text{Reg}}|a)$ is an auxiliary Gaussian channel whose statistics are generated from the network behavior on the training data. It is important to note that the Gaussian channel here is not necessarily the optimal auxiliary channel, and a better auxiliary channel could yield a

higher AIR. In the next sections, we will present more detailed information about the training, validation, testing, parameter optimization of the proposed NN-based techniques. Note that the R_{Cla} is a discrete random variable as opposed to the continuous random variables R_{iso} , R_{CCA} and R_{Reg} . This nature does not affect the effectiveness of the AIR estimated as the only the a posteriori probability outputs of the softmax layer are used for AIR estimation rather than the R_{Cla} itself [21].

5. Achievable information rate analysis

This section presents a numerical AIR analysis for the four detection schemes introduced in the last section. We employ a series of 8 and 16 PAM constellations in this work. These constellations are optimally shaped by solving the optimization problem (5) under different peak amplitude constraints, using the interior-point algorithm in MATLAB. Recall that the optimization problem in (5) is simplified by performing VNT over the noncentral chi-squared channel to obtain an equivalent AWGN channel. The transmitted symbols are randomly drawn from the probabilistic and geometrically shaped constellation and encoded on a purely imaginary eigenvalue. Examples of the optimally shaped constellation are shown in Fig. 6. It can be observed that the optimized constellations make up quasi-uniform distributions due to the similarity between the optimization problem (5) and the peak amplitude constrained problem in [26] and [13]. Furthermore, note that the relationship between the A_{max} and the average power of the signal depends on the distribution of the input and the choice of the signaling parameter δ . For the $\delta = 10^{-2}$ employed in this work, the dynamic range of launch power for the shaped PAM constellation is estimated to be within -4 to 10 dBm for $A_{\text{max}} = 0.5$ to 2.40 .

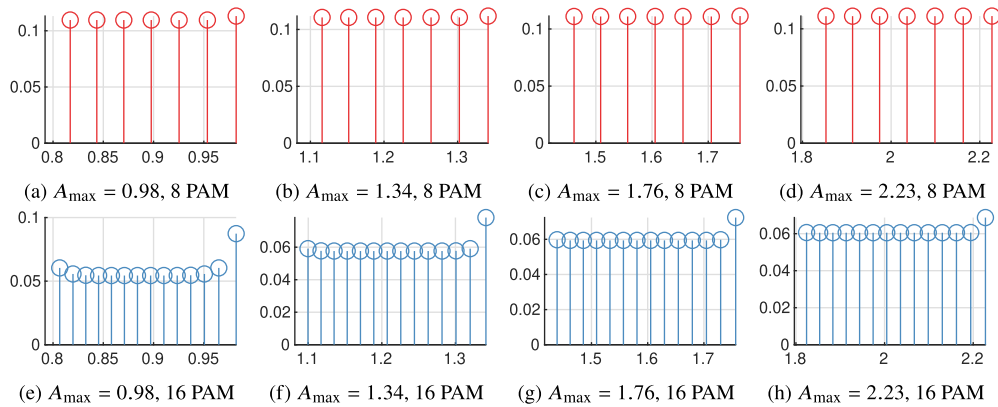


Fig. 6. Examples of input constellation for $A_{\text{max}} = 0.98, 1.34, 1.76$ and 2.23 for the geometric and probabilistic hybrid shaped 8 PAM and 16 PAM. Note that the mass point at 0 is omitted for simplicity.

Performing inverse NFT on the eigenvalue, the corresponding time domain pulse, i.e. a soliton, are then obtained. The pulse evolution over a long haul fiber is simulated with the split step Fourier method (SSFM), where an ideal distributed Raman amplified 2000 km fiber is assumed, with parameters $\alpha = 0.2$ dB/km, $\beta_2 = -2.1 \times 10^{-26}$ s²/m, and $\gamma = 1.27 \times 10^{-3}$ /W/m. Moreover, phonon occupancy K_T of 1.13 and wavelength of $1.55 \mu\text{m}$ are used. As for the normalization parameters, the normalization time T_0 is selected to be 0.1 ns, leading to the normalization length of $2T_0^2/|\beta_2| = 910$ km and normalization power of $|\beta_2|/\gamma T_0^2 = 2.39$ dBm. Subsequently, the four detection schemes are employed. In the direct NFT detection and also the AMMSE detection scheme, a fast NFT algorithm developed for MATLAB is used to compute the nonlinear frequency spectrum efficiently [27]. According to insights from observing numerical simulation

results and literature [14], only a fraction of the sampled CS is required for correlation estimation. In this work, 1/4 of the CS samples are taken for AMMSE estimator if not specified otherwise.

For the BLSTM network detectors, the classifier network employed here consists of a sequence input layer with an input size of 2 (real parts and imaginary parts of the pulse), and two BLSTM layers, each with 100 hidden units. The BLSTM layers are followed by a fully connected layer and a softmax output layer with 8 or 16 outputs, depending on the modulation format. In the regression network, a regression layer with a single real output is connected after the fully connected layer. Each of the training and validation channel realizations includes the transmission of a 16-soliton sequence through the SSFM fiber with a step size of 200 m. In contrast, each of the testing channel realizations includes transmission of a 256-soliton sequence, which is much longer and more practical. The number of samples per soliton is rounded up to the next closest power of 2, taking into account the 10 times oversampling of the 99% bandwidth. For instance, 99% and 99.9% timewidths would both require 64 samples per soliton. The BLSTM networks are trained with 80000 channel realizations as the training set, and a different set of 80000 channel realizations as the validation set. The validation set is used to avoid and monitor the possibility of overfitting, and the criteria of stopping the training is when the validation loss ceases to improve. The AIRs for both the classifier and regression detectors, I_{Cla} and I_{Reg} , are estimated with 1000 test realizations. The neural network is trained with adaptive moment estimation (ADAM) optimizer. The same training data are also employed to compute the AMMSE estimator, and I_{CCA} is also estimated with the same test data.

A fixed maximum and minimum soliton amplitude ratio of $A_{\text{max}}/A_{\text{min}} = 1.24$ is considered, implying a fixed effective time bandwidth product of $\text{TB} = 3.52$ for $\delta = \epsilon = 10^{-2}$. The AIRs are estimated for varying maximum soliton amplitudes A_{max} , (corresponding to varying bandwidth), as depicted in Fig. 7. Note that the SE could be calculated by dividing the AIR with the effective time-bandwidth product TB. The results show that the inter-soliton interaction becomes dominant compared to the signal-dependent noise of the isolated soliton channel. This dominance makes the perturbative theory noncentral chi-squared channel model effectively invalid. Consequently, the equivalent AWGN channel after VNT is no longer valid as well. This results in significant degradation of the mismatch capacity lowerbound, I_{iso} , tending it to 0 and 0.041 bits/symbol respectively, for 16-PAM and 8-PAM under lowest interaction at $A_{\text{max}} = 0.5$. Such degradation of mismatch capacity is expected even at $\delta = 10^{-3}$ due to excessive interaction at higher launch power as shown in [13]. Additionally, the AMMSE estimator can compensate for interaction to some extent, achieving 2.93 bits/symbol at $A_{\text{max}} = 0.51$, which is very close to the limit of 8 PAM. However, I_{CCA} decreases when the soliton amplitude increases. This decrease is attributed to the AMMSE estimator's inability to cope with the increasing interaction in the higher soliton amplitude regime.

Figure 7 also shows that the BLSTM classifier network detectors can extend the AIR-increasing power regime further. It is worth noticing that I_{Cla} saturates at 3 bits/symbol for $A_{\text{max}} < 0.98$ for the 8 PAM constellation, due to the limit of the constellation size. This suggests that the interaction can be fully compensated by the network within this operation regime. However, the BLSTM regression network is not able to completely adapt to the channel model and is limited by the misalignment between the auxiliary Gaussian channel model and the network equalized channel model. It is observed that I_{Reg} reaches 2.99 bits/symbol at $A_{\text{max}} = 0.68$, which is slightly lower than the 3 bits/symbol achieved by the classifier network. For 16 PAM where the constellation points are distributed closer, the task becomes more challenging for the classifier network. A clear peak at around $A_{\text{max}} = 0.98$ could be identified, and 3.87 bits/symbol could be achieved. When A_{max} is less than 0.98, I_{Cla} could be improved by increasing the power, suggesting that the system is more perturbed by noise in this regime. If the power is further increased, I_{Cla} could no longer be improved, because the system becomes more interaction dominant in this regime. For the regression network, similar trend could be observed with the peak appearing at $A_{\text{max}} = 0.68$,

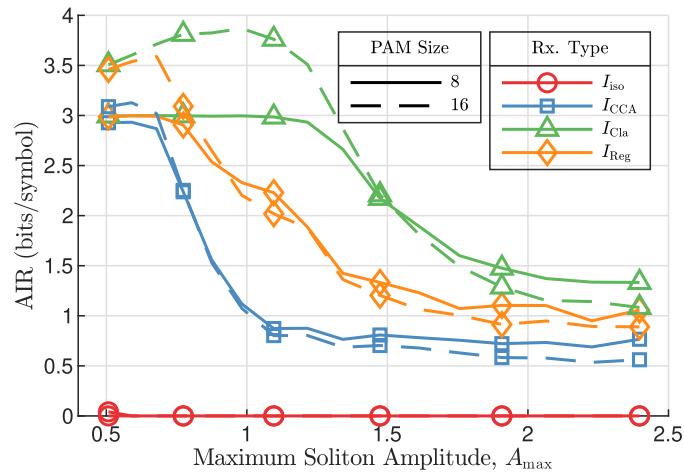


Fig. 7. AIRs of AM soliton communication system with probabilistically and geometrically shaped 8 and 16 PAM.

achieving a lower AIR of 3.59 bits/symbol compared to the I_{Cla} . It is also seen that the AIRs for both networks for 16 PAM degrade to 1.08 and 0.89 bits/symbol at $A_{max} = 2.40$ which are lower than the 1.33 and 1.06 bits/symbol provided by 8 PAM. This is because of the dominant soliton interaction having higher impact on the closer spaced constellation. Moreover, it is also worth highlighting that the classifier could achieve the highest AIR of 3.87 bits/symbol for the 16 PAM at $A_{max} = 0.98$ showing significant gain over the other counterparts, where AIRs of 2.20, 1.07, and 0 bits/symbol are achieved by regression network, AMMSE estimator, and direct NFT, respectively. In general, it can be concluded that the BLSTM classifier suits the system design and power regime of interest in this work better than the regression network counterpart.

Previously, we discussed the AIRs in bits/symbol under certain time-bandwidth constraints, and it was clear that the classifier network outperforms the regression network as well as the other benchmark schemes. For a more practical insight, data rate in bits/second should also be investigated for the BLSTM classifier detector. The data rate could be determined by the SE and the timewidth of the pulse. In the AM soliton system considered in this work, the timewidth of the pulse could be altered by selecting different time-domain energy truncation factors δ . The sech pulse has a trailing edge that degrades to effectively zero, hence this is equivalent to introducing guarding intervals between soliton pulses. Consequently, a gain in SE should be expected due to lower interaction. This reveals a trade-off between SE and timewidth. Since the optimal SE in the interested regime for classifier network occurs at $A_{max} = 0.98$, and the AIR for 8 PAM and 16 PAM approximately coincide at $A_{max} = 1.34$, the data rates at the two A_{max} s are investigated by selecting a smaller δ and shown in Table 1. Note that the two A_{max} s correspond to bandwidths around 10GHz and 14GHz respectively. As expected, reducing δ results in lower interaction, allowing a higher AIR to be achieved. For example, an AIR of 2.96 bits/symbol and a data rate of 6.17 Gbps could be achieved by the AMMSE estimator for 8 PAM at $A_{max} = 0.98$ and $\delta = 10^{-3}$. However, for the BLSTM classifier detector, the negative impact on the data rate from extending the timewidth cannot be offset by the SE gain in the operational setup considered in this work. For example, although a higher AIR of 3.83 bits/symbol could be achieved with 16 PAM and $\delta = 10^{-3}$ at $A_{max} = 0.98$, a lower data rate of 10.89 Gbps could be supported compared to 11.72 Gbps with $\delta = 10^{-2}$. It's worth noting that a higher time bandwidth product TB will be consumed when a smaller δ is selected, making the employment of the BLSTM classifier network detector with a more compact truncation $\delta = 10^{-2}$ a reasonable choice in the considered system.

Table 1. Data rates in Gbps for BLSTM classifier and AMMSE detectors

PAM, δ	$A_{\max} = 0.98$, BW = 10GHz		$A_{\max} = 1.34$, BW = 14GHz	
	\mathbb{R}_{Cla} (Gbps)	\mathbb{R}_{CCA} (Gbps)	\mathbb{R}_{Cla} (Gbps)	\mathbb{R}_{CCA} (Gbps)
$8, 10^{-2}$	8.95	3.34	10.87	3.12
$8, 10^{-3}$	6.24	6.17	8.52	7.76
$16, 10^{-2}$	11.56	3.21	11.72	2.80
$16, 10^{-3}$	7.94	6.73	10.89	7.82

6. Hyperparameter analysis

The performance of the BLSTM network detectors is influenced by various hyperparameters. Before examining the hyperparameters that determine the complexity of the neural network, the channel memory should be investigated. Given the constraints of computational power, only a limited number of realizations of the soliton propagation can be generated. On one hand, it is essential to generate as many realizations of the practical NLSE channel as possible to ensure the sample statistics accurately characterize the physical channel. On the other hand, as many solitons as possible should be generated within each realization of transmitting a soliton sequence over the NLSE channel to capture diverse interaction patterns.

To identify the shortest soliton sequence necessary to capture both the interaction patterns and the randomness of the NLSE channel, the training and validation data should be prepared with soliton sequences of different lengths. The network trained with each sequence length is then tested with the previously used 1000 realizations of 256-soliton sequences. If the performance of the network on the test data aligns with that on the validation data, it can be concluded that the training and validation data sufficiently characterize the nonlinear stochastic channel.

Due to the signal dependence of the interaction, the largest $A_{\max} = 2.40$ is selected in this set of experiments to numerically estimate the longest memory of the channel in the sparse-spaced interacted 8-PAM soliton transmission. To maintain a similar total number of soliton transmissions, $2^6 \times 10^4$, $2^5 \times 10^4$, $2^4 \times 10^4$, $2^2 \times 10^4$ realizations of 3, 8, 16, 64-soliton sequences are generated as training and validation data respectively. Focusing on the better-performing classifier network, the trained network is then tested with the same 256-soliton sequence testing data used previously. The performance of the network on both the validation set and test set is illustrated in Fig. 8(a). The memory of the interacted channel can be estimated to be around 16, as the AIRs on the validation set and test set start to converge. Note that the size of the channel memory is dependent on channel parameters, including A_{\max} , δ , and the order of constellation. For example, for the closer-spaced 16-PAM constellation, we expect the channel memory to be at least 16-soliton long.

The previous intrinsic interaction and soliton sequence length experiments clearly demonstrate that the interaction soliton patterns are dependent on the neighboring solitons in a nonlinear and dynamic manner. This suggests the potential of exploiting such a memory effect in detecting interacted solitons. Without introducing extra communication overhead, a certain portion of the previous soliton is included in the neural network's input, along with the samples corresponding to the target soliton. When the first soliton of the sequence is received, the memory buffer at the receiver is filled with zeros, matching the number of samples of the specified portion of the previous soliton. Let D denote the number of samples for each soliton and f_{mem} denotes the memory factor, which indicates the portion of previous soliton taken by the receiver to assist detection. Then the size of the input vector is determined as $(D + \lceil f_{\text{mem}} D \rceil) \times 2$, where $\lceil \cdot \rceil$ denotes the rounding operation and the factor 2 indicates that the real and imaginary parts are fed into the network as two features of the input pulses. Using the same training, validation and test data, both classifier and regression networks are trained for 16 PAM constellation with the additional

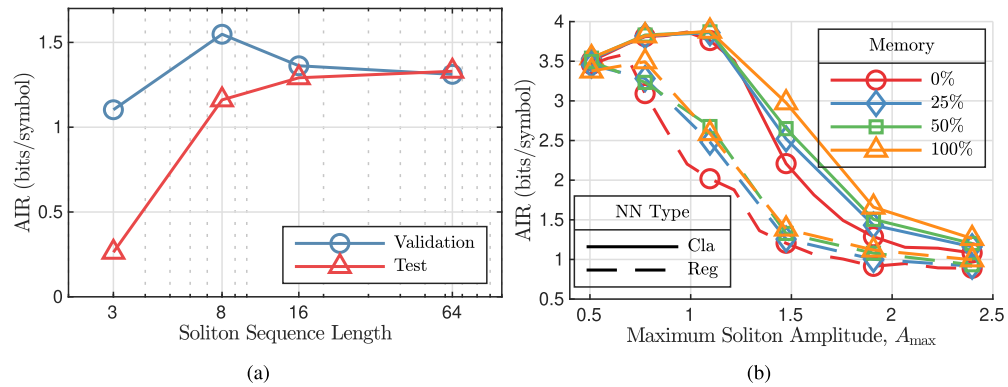


Fig. 8. The AIRs achieved by (a) 8 PAM at $A_{max} = 2.40$ and BLSTM classifier network trained with transmission of different length soliton sequence. (b) 16 PAM and BLSTM classifier and regression networks with specified portion of the soliton received previously as memory to assist the detection.

25%, 50% and 100% of the previous soliton. The estimated AIRs are then compared with the results of the network without such memory, as illustrated in Fig. 8(b). A similar conclusion can be drawn that even the best regression network with 100% of the previous soliton cannot outperform the classifier network without memory in the regime of interest, indicating that the regression is not as suitable for this particular application. As for the classifier network, it can be concluded that the memory of the previous soliton is beneficial when the system is in the interaction dominant regime. When $A_{max} < 0.98$, the system is primarily limited by the ASE noise as highlighted before, and increasing f_{mem} to 1 only bring marginal improvement of 0.01 bits/symbol at $A_{max} = 0.77$. However, when the system becomes interaction dominant, larger improvement can be expected, for example, improvement of 0.78 bits/symbol is achieved by completely including the previous soliton at $A_{max} = 1.47$. Additionally, it is observed that the AIR gain from including the previous soliton as memory also reduces due to the signal-dependent interaction as observed at $A_{max} = 1.91$, where the AIR gain is reduced to 0.37 bits/symbol for $f_{mem} = 1$. When stronger interaction is encountered, more previous solitons might be required to improve the accuracy of detection.

The aforementioned network requires two features to be extracted: the real and imaginary part of the time domain pulses. It has been highlighted that implementing an AMMSE estimator using the CS of the time-domain pulse could potentially improve performance. Therefore, the real and imaginary parts of the CS are also included as two additional features for the network, as indicated by the dash line connection in Fig. 5. This inclusion aims to identify a superior nonlinear estimator that could potentially outperform the AMMSE estimator [14]. In Fig. 9(a), the AIRs of both the BLSTM classifier and regression networks, using the real and imaginary parts of both the time-domain pulse and CS spectrum, are illustrated and compared with the AIRs achieved by networks that only leverage the time-domain pulse. It can be observed that the additional CS feature does not provide any significant gain. This suggests that the network's structure might be sufficiently complex to learn the useful information provided by the CS of the corresponding pulse.

The size of the training data can significantly impact the performance of the network. If the training data set is small, it may not adequately characterize the target problem, leading to poor performance when the trained network encounters test data. It is also worth investigating whether providing more training data, given the availability of computational power, could enhance the performance of the trained network. By selecting the 8 PAM constellation at $A_{max} = 2.40$, we

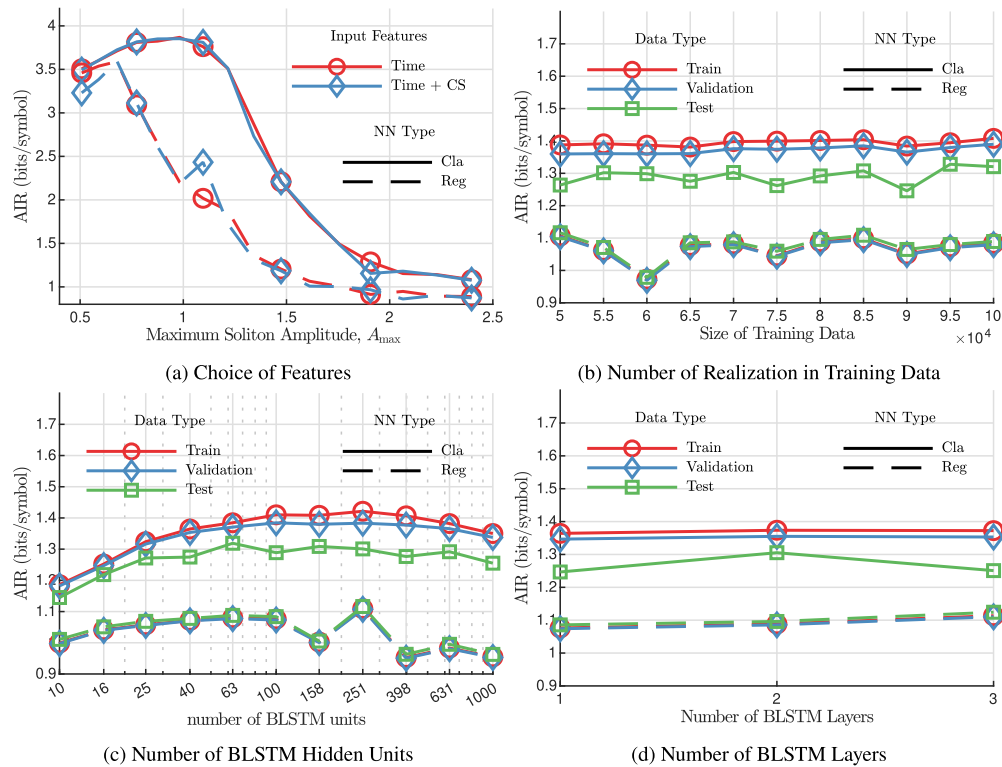


Fig. 9. The AIRs achieved by (a) 16 PAM and BLSTM network detector with different features of the input; (b) 8 PAM at $A_{\max} = 2.40$ and BLSTM network trained with transmission of different length soliton sequence; (c) 8 PAM at $A_{\max} = 2.40$ and BLSTM network with different number of hidden units within each BLSTM layer; (d) 8 PAM at $A_{\max} = 2.40$ and BLSTM network with different number of BLSTM layers.

intend to observe the potential improvements under the strongest interaction for a well-spaced constellation. The estimated AIRs for networks trained with varying amounts of realizations in the training data are sketched in Fig. 9(b). The results display a marginal AIR improvement of 0.01 bits/symbol for classifier network on the testing data when trained with 9.5×10^4 realizations of sequence transmissions, compared to training with 8×10^4 realizations.

The complexity of the neural network reflects its potential to replicate the target nonlinear mapping relationship. Commonly, the number of hidden units within each BLSTM layer and the number of BLSTM layers (also known as the depth of the network) are adjusted to alter the network's complexity. However, increasing complexity also raises the number of hidden parameters needed to describe the network, thereby necessitating more training data. As illustrated in Fig. 9(c), both networks demonstrate a gain in AIR with an increase in the number of hidden units in each BSLTM layer, up to 63 hidden units. Beyond this point, the AIRs tend to saturate and even decrease after 251 units. Regarding the network's depth, a trade-off between complexity and the required size of training data can be observed for the classifier network, while the regression network maintains marginal AIR gain even with 3 BLSTM layers. Nevertheless, from all of the above analysis, it is observed that the regression network with varied hyperparameters cannot outperform its classifier counterparts for the constellation and operational regime of interest, further demonstrating the suitability of the classifier within the discussed system.

7. Conclusion

In conclusion, the application of data-driven learning techniques, such as regression and classifier network detectors, in AM soliton communication systems with closely-spaced soliton transmission can lead to significant gains compared to conventional DSP techniques, like the direct NFT and AMMSE estimator. However, we also found that even the designed NN-based systems can enter an interaction-dominant regime where the signal-dependent interaction surpasses the capability of both networks. In the ASE dominant regime, the AIR of the system can be improved by increasing the operational power, in contrast to the interaction dominant regime, where the system shows performance degradation as power increases. Considering the signal-dependent signaling timewidth, a trade-off between interaction and time efficiency is observed. From both the AIR and hyperparameter analyses, we see the evidence that the classifier design of the network is more resilient to interaction impairments than its regression counterpart. Furthermore, we explored the use of channel memory introduced by the inter-soliton interaction to aid symbol detection, demonstrating that incorporating the previous soliton into the detection process can enhance AIR when the system is in the interaction-dominant regime. The dependence of the interaction on the propagation distance remain less explored here; hence, one of the possible future research directions is to investigate the effective design and training of NN techniques and their potential achievable gains for different propagation distances.

Funding. Leverhulme Trust Research Project Grant; China Scholarship Council.

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

References

1. R.-J. Essiambre, G. Kramer, P. J. Winzer, *et al.*, "Capacity limits of optical fiber networks," *J. Lightwave Technol.* **28**(4), 662–701 (2010).
2. A. Hasegawa and F. Tappert, "Transmission of stationary nonlinear optical pulses in dispersive dielectric fibers. i. anomalous dispersion," *Appl. Phys. Lett.* **23**(3), 142–144 (1973).
3. A. Hasegawa and T. Nyu, "Eigenvalue communication," *J. Lightwave Technol.* **11**(3), 395–399 (1993).
4. J. P. Gordon and H. A. Haus, "Random walk of coherently amplified solitons in optical fiber transmission," *Opt. Lett.* **11**(10), 665–667 (1986).
5. M. I. Yousefi and F. R. Kschischang, "Information transmission using the nonlinear fourier transform, part i: Mathematical tools," *IEEE Trans. Inf. Theory* **60**(7), 4312–4328 (2014).
6. S. K. Turitsyn, J. E. Prilepsky, S. T. Le, *et al.*, "Nonlinear fourier transform for optical data processing and transmission: advances and perspectives," *Optica* **4**(3), 307–322 (2017).
7. I. Tavakkolnia and M. Safari, "Capacity analysis of signaling on the continuous spectrum of nonlinear optical fibers," *J. Lightwave Technol.* **35**(11), 2086–2097 (2017).
8. I. Tavakkolnia and M. Safari, "Signaling on the continuous spectrum of nonlinear optical fiber," *Opt. Express* **25**(16), 18685–18702 (2017).
9. Y. Chen, M. Baniasadi, and M. Safari, "On optimally shaped signals for nonlinear frequency division multiplexed fiber systems," *IEEE Trans. Commun.* **71**(9), 5379–5391 (2023).
10. S. A. Derevyanko, S. Turitsyn, and D. Yakushev, "Non-gaussian statistics of an optical soliton in the presence of amplified spontaneous emission," *Opt. Lett.* **28**(21), 2097–2099 (2003).
11. S. A. Derevyanko, J. E. Prilepsky, and D. A. Yakushev, "Statistics of a noise-driven manakov soliton," *J. Phys. A: Math. Gen.* **39**(6), 1297–1309 (2006).
12. J. Gordon, "Interaction forces among solitons in optical fibers," *Opt. Lett.* **8**(11), 596–598 (1983).
13. Y. Chen, I. Tavakkolnia, A. Alvarado, *et al.*, "On the capacity of amplitude modulated soliton communication over long haul fibers," *Entropy* **22**(8), 899 (2020).
14. Q. Zhang and F. R. Kschischang, "Improved soliton amplitude estimation via the continuous spectrum," *J. Lightwave Technol.* **37**(13), 3087–3099 (2019).
15. G. Zhou, T. Gui, C. Lu, *et al.*, "Improving soliton transmission systems through soliton interactions," *J. Lightwave Technol.* **38**(14), 3563–3572 (2019).
16. O. Kotlyar, M. Kamalian-Kopae, M. Pankratova, *et al.*, "Convolutional long short-term memory neural network equalizer for nonlinear fourier transform-based optical transmission systems," *Opt. Express* **29**(7), 11254–11267 (2021).
17. X. Chen, H. Ming, C. Li, *et al.*, "Two-stage artificial neural network-based burst-subcarrier joint equalization in nonlinear frequency division multiplexing systems," *Opt. Lett.* **46**(7), 1700–1703 (2021).

18. S. Gaiarin, F. Da Ros, R. T. Jones, *et al.*, “End-to-end optimization of coherent optical communications over the split-step fourier method guided by the nonlinear fourier transform theory,” *J. Lightwave Technol.* **39**(2), 418–428 (2020).
19. R. T. Jones, S. Gaiarin, M. P. Yankov, *et al.*, “Time-domain neural network receiver for nonlinear frequency division multiplexed systems,” *IEEE Photonics Technol. Lett.* **30**(12), 1079–1082 (2018).
20. Y. Wu, L. Xi, X. Zhang, *et al.*, “Robust neural network receiver for multiple-eigenvalue modulated nonlinear frequency division multiplexing system,” *Opt. Express* **28**(12), 18304–18316 (2020).
21. P. J. Freire, J. E. Prilepsky, Y. Osadchuk, *et al.*, “Deep neural network-aided soft-demapping in coherent optical systems: Regression versus classification,” *IEEE Trans. Commun.* **70**(12), 7973–7988 (2022).
22. M. J. Ablowitz and H. Segur, *Solitons and the inverse scattering transform* (SIAM, 1981).
23. M. Borghese, R. Jenkins, and K. D.-R. McLaughlin, “Long time asymptotic behavior of the focusing nonlinear schrödinger equation,” in *Annales de l’Institut Henri Poincaré C, Analyse non linéaire*, vol. 35 (Elsevier, 2018), pp. 887–920.
24. E. Meron, M. Feder, and M. Shtaif, “On the achievable communication rates of generalized soliton transmission systems,” *arXiv*, arXiv:1207.0297 (2012).
25. P. Sadeghi, P. O. Vontobel, and R. Shams, “Optimization of information rate upper and lower bounds for channels with memory,” *IEEE Trans. Inf. Theory* **55**(2), 663–688 (2009).
26. J. G. Smith, “The information capacity of amplitude-and variance-constrained scalar gaussian channels,” *Inf. Control.* **18**(3), 203–219 (1971).
27. S. Wahls, S. Chimmalgi, and P. Prins, “FNFT: A Software Library for Computing Nonlinear Fourier Transforms,” *The J. Open Source Softw.* **3**(23), 597 (2018).