Research Article Vol. 33, No. 19 / 22 Sep 2025 / Optics Express 40851
Optics EXPRESS

Enhancing generalization in neural network-based waveform-level channel modeling for optical fiber transmission through parameter encoding structures

MINGHUI SHI,† HANG YANG,† CHUYAN ZENG,† ZEKUN NIU, JUNZHE XIAO, MINGZHE CHEN, WEISHENG HU, AND LILIN YI*

State Key Laboratory of Photonics and Communications, Publications Department, School of Integrated Circuit Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China
†These authors contributed equally
*lilinyi@sjtu.edu.cn

Abstract: Fast and accurate optical fiber channel waveform modeling is essential for the evaluation of optical communication systems. Traditional modeling methods, such as the split-step Fourier method (SSFM), suffer from significant computational complexity. In contrast, neural network (NN)-based approaches not only achieve comparable accuracy but also significantly reduce computational burden. Enhancing the generalization capability of NNs is crucial to facilitate flexible system design and optimization across varying system parameters. In this work, we introduce a novel parameter encoding structure, which significantly improves NN generalization by pre-encoding system parameters. In the generalized scenario, with a wide range of launch power from -2 to 7 dBm and arbitrary transmission distances, the parameter encoding structure improves waveform modeling accuracy by 49.9% and 69.7% compared to the non-encoded scheme. Notably, for the first time, we develop a single NN that generalizes across multiple system parameters–modulation format, symbol rate, WDM channel space, phase noise, and frequency offset of lasers, launch power, accumulated chromatic dispersion, span length, and total transmission distance–simultaneously. This enhanced generalized NN holds significant potential for the design and optimization of optical transmission systems.

1. Introduction

Optical fiber communication systems form the backbone of modern information transmission infrastructure. The optical fiber channel model characterizes the propagation of optical signals through the optical fiber, playing a crucial role in predicting system performance, developing advanced digital signal processing (DSP) algorithms [1–5], and performing end-to-end (E2E) optimization [6–9]. The propagation of signals within optical fibers is governed by the nonlinear Schrödinger equation (NLSE) [10]. Except for a few special cases, NLSE lacks an analytical solution. Over the past decades, numerous modeling schemes have been proposed to effectively model the optical fiber channel, typically categorized into power-level and waveform-level approaches. Power-level modeling approaches, represented by Gaussian noise (GN) models [11,12] and their variants—including enhanced GN (EGN) models [13] and inter-channel stimulated Raman scattering GN (ISRSGN) models [14]—treat nonlinearity as Gaussian noise, providing fast and accurate generalized signal-to-noise ratio (GSNR) estimations. However, these GN-like models fail to provide detailed signal waveform information, limiting their applicability in DSP algorithm development, particularly for nonlinearity compensation. The traditional waveform-level modeling method is the split-step Fourier method (SSFM), which resolves NLSE by iterative calculation. SSFM has high accuracy but low running efficiency, especially for long-haul wavelength multiplexing (WDM) transmission modeling [15].

#573803 https://doi.org/10.1364/OE.573803
Journal © 2025 Received 18 Jul 2025; revised 4 Sep 2025; accepted 9 Sep 2025; published 17 Sep 2025

Optics EXPRESS

Recently, neural network (NN)-based techniques have been explored to achieve fast and accurate waveform-level modeling. These methods learn signal and channel characteristics from labeled data [16] combined with prior knowledge [17]. Compared to SSFM, NN-based approaches offer similar accuracy while significantly reducing computational complexity, demonstrating successful applications in various scenarios, including Intensity Modulation Direct Detection (IM-DD) systems and coherent transmission across both single- and multi-channel configuration [16–25]. For effective NN-based transmission performance prediction and system design across varying configurations, it is crucial for a single NN to possess strong generalization capabilities across a broad range of system parameters. Although training the NN with a rich dataset [20–24] or fine-tuning the already trained model [25]can achieve certain generalization capabilities, generalization across a wide range of launch powers—particularly under conditions outside the training distribution—and over arbitrary transmission distances, including varying span lengths, remains a challenging problem that warrants further investigation. Additionally, the generalization for symbol rates, channel space, channel accumulated chromatic dispersion (aCD), or phase noise (PN) and frequency offset (FO) of lasers in WDM scenarios require further exploration. In summary, addressing the generalization ability of NN-based methods for large-scale, multi-parameter scenarios is an urgent and necessary task.

In this paper, we extend our conference work in [26] and introduce a novel parameter encoding structure integrated with our previously proposed feature decoupling distributed (FDD) scheme [20], significantly enhancing the generalization capability of NN-based waveform-level channel modeling. The parameter encoding structure pre-encodes system parameters through a fully connected layer, capturing their impact on channel characteristics from a higher-dimensional perspective. In the generalized scenario, with a wide range of launch power from -2 to 7 dBm and arbitrary transmission distances, the parameter encoding structure improves waveform modeling accuracy by 49.9% and 69.7% compared to the non-encoded scheme. Importantly, the introduction of the parameter encoding structure does not introduce excessive additional complexity. The NN with the parameter encoding scheme still reduces computational time by 99.3% relative to SSFM. Notably, for the first time, we develop a single NN that generalizes across multiple system parameters–including modulation format, symbol rate, WDM channel space, PN and FO of lasers, launch power, aCD, span length and total transmission distance–simultaneously. This enhanced generalized NN holds significant potential for the design and optimization of optical transmission systems. It can accurately predict transmission performance across various system parameter configurations, with a Q-factor error of less than 0.12 dB compared to SSFM. Additionally, the NN demonstrates comparable performance to the SSFM implemented in digital backpropagation (DBP) algorithms with varying step size configurations, highlighting its potential in the design and optimization of nonlinear compensation algorithms.

This study extends our preliminary conference work through the following key contributions:

1. A comprehensive analysis of the superior performance of the parameter encoding scheme over non-encoded approaches in terms of symbol rate and WDM channel space generalization.

2. An investigation into the capability of a single NN integrated with the parameter encoding scheme, to simultaneously generalize multiple system parameters. The enhanced generalized neural network is validated for accurate transmission performance prediction across varying system configurations.

3. An assessment of the efficacy of the enhanced generalized NN in designing nonlinear compensation algorithm.

The rest of this paper is organized as follows. Section 2 introduces the coherent simulation system setup and principle of parameter encoding structure integrated with FDD. Section 3

Optics EXPRESS

discusses the performance gain of parameter encoding structure. Section 4 shows the enhanced generalized NN-based channel modeling for optical transmission systems. Finally, Section 5 provides conclusions.

2. Principles

2.1. Coherent WDM simulation system setup

To collect the training data and analyze the signal modeling performance, we construct a coherent WDM optical fiber transmission simulation system, including the transmitter, the optical fiber channel, and the receiver. The complete system block diagram is shown in Fig. 1(a). The main simulation parameters are listed in Table 1. At the transmitter, signals undergo modulation, pulse shaping via a root raised-cosine (RRC) filter and resampling. Subsequently, the signals from different channels are modulated on separate frequency carriers to create full-field WDM signals for transmission within the optical fiber channel. The FO and PN induced by lasers are considered. The impact of PN and FO on the k-th symbol can be modeled as (1), where $s(k)$ and $l(k)$ represent the k-th symbol before and after passing through the laser model, respectively, and φ_k is the total phase shift caused by PN and FO, $\Delta\omega$ represents the FO, T is the sampling period, and $\theta_L(k)$ is PN modeled as a Wiener process [27]. The FO and PN are introduced at the oversampled level.

$$l(k) = s(k)\exp(\mathrm{j}\varphi_k),\ \varphi_k = k\Delta\omega T + \theta_L(k). \tag{1}$$

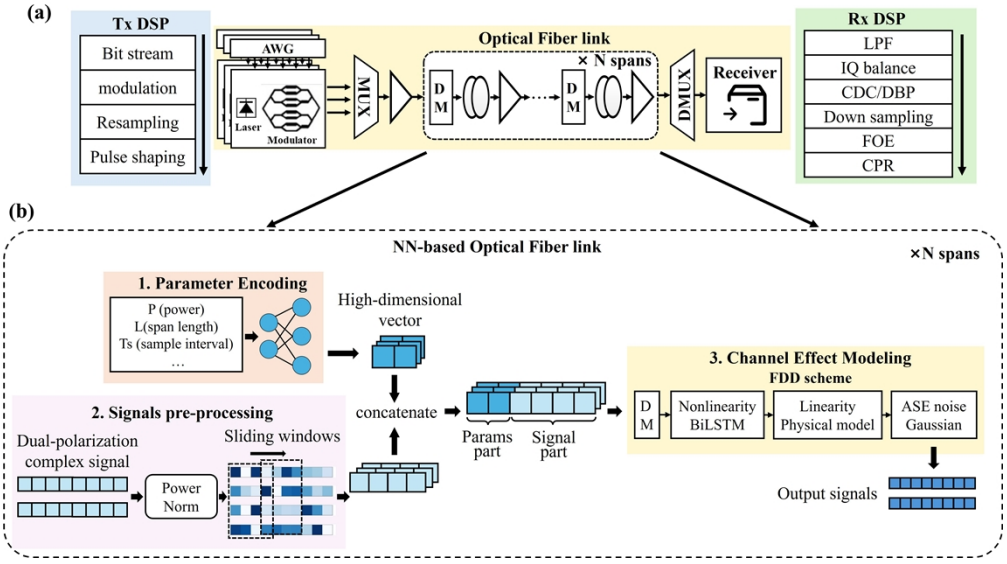

Fig. 1. Coherent WDM simulation diagram and the NN-based channel modeling scheme details. (a) Schematic of the coherent WDM optical fiber transmission simulation system. (b) NN-based single-span optical fiber channel model scheme combined with parameter encoding structure and FDD.

The evolution of the signal within the optical fiber channel is governed by NLSE. In this work, the dual-polarization signal is considered, and its transmission process can be described using Manakov equation [28] derived from NLSE, as follows:

$$\frac{\partial \boldsymbol{A}}{\partial z} = -\frac{i}{2}\beta_2\frac{\partial^2 \boldsymbol{A}}{\partial t^2} + \frac{8}{9}i\gamma|A|^2\boldsymbol{A} - \frac{\alpha}{2}\boldsymbol{A}, \tag{2}$$

Optics EXPRESS

Table 1. Parameter setup of simulation system

Parameters	Value
Roll-off factor of RRC	0.1
WDM channel number	5
Symbol rate and WDM channel space	[30 GBaud 50 GHz]
Launch power	3.5 dBm/channel
Span length	80 km
Carrier wavelength	1550 nm
Attenuation	0.2 dB/km
Dispersion	17 ps/(nm·km)
Nonlinear coefficient	1.3 /(W·km)
EDFA noise figure	5 dB

where $\boldsymbol{A}$ denotes the optical field of the two orthogonal polarization modes $\boldsymbol{A_x}$ and $\boldsymbol{A_y}$. β_2 denotes the group velocity dispersion parameter. γ represents the nonlinear Kerr parameter. α is the loss parameter. t is the time, and z is the transmission distance. SSFM is the most used numerical method for solving the NLSE. In this paper, SSFM is used to generate data for NN training and serves as the comparative baseline in performance evaluation. The SSFM divides long-haul optical fiber into numerous small steps, allowing linear and nonlinear operators to be considered independently. The symmetric SSFM operation [29] at each step is expressed as:

$$\boldsymbol{A}(z+h,t) \approx \exp\left(\frac{h}{2}\hat{\boldsymbol{D}}\right)\exp\left\{h\hat{\boldsymbol{N}}\left[\boldsymbol{A}\left(z+\frac{h}{2},t\right)\right]\right\}\exp\left(\frac{h}{2}\hat{\boldsymbol{D}}\right), \tag{3}$$

where $\widehat{\boldsymbol{D}}$ is the linear operator that accounts for the effects of attenuation and chromatic dispersion (CD), and $\widehat{\boldsymbol{N}}$ is the nonlinear operator. h is the length in each step. The step size configuration of SSFM includes both constant variable step size methods. In this paper, the constant step size SSFM is set with a step size of 0.01 km. For the variable step size SSFM, the nonlinear phase rotation method [30] is used, with the maximum phase rotation set to 0.005. The sampling rate of SSFM is set to four times the number of WDM channels. An erbium-doped fiber amplifier (EDFA) with amplifier spontaneous emission (ASE) noise is deployed at the end of to compensate the fiber attenuation.

The channel aCD refers to the CD accumulated along the transmission link. Varying aCD can lead to various nonlinear effects accumulation within the fiber link, due to the strong interplay between linear CD and fiber nonlinearities [31–34]. The adjustment of aCD can be achieved by introducing dispersion management (DM) unit after each span. To demonstrate the generalization ability of NN for aCD in the following text, DM units are introduced in simulation system, as shown in Fig. 1(a). The ratio of dispersion management (R_{DM}), defined as the ratio of the management intensity to the aCD generated in a single-span SSMF transmission. $\widehat{\boldsymbol{A}}$ denotes the signals after DM, which can be expressed as:

$$\widehat{A}(z,\omega) = A(z,\omega)(-i\frac{\beta_2}{2}\omega^2 L_{span})R_{DM}. \tag{4}$$

At the receiver, the Rx DSP resamples the signal to two samples per symbol (SPS) and compensates for system impairments. The signal is initially processed through a low-pass filter and IQ balancing, followed by chromatic dispersion compensation (CDC). The CDC can be replaced by DBP [5], which compensates for both linear and nonlinear impairments. The signals are then downsampled to 1SPS, followed by frequency offset estimation (FOE) [35] and carrier

Optics EXPRESS

phase recovery (CPR) [36]. FOE eliminates FO between the laser and local oscillator (LO), while CPE recovers PN caused by the laser and LO. Finally, demodulation and transmission performance calculation are carried out.

2.2. FDD modeling scheme

The basic NN-based fiber channel waveform-level modeling scheme is the FDD scheme proposed in [20]. To enhance accuracy and reduce complexity, an improved version of FDD, the sequence-to-sequence (Seq2Seq)-FDD framework [25], is employed. This scheme incorporates three key characteristics: distributed modeling, feature decoupling, and the Seq2Seq framework.

Distributed modeling refers to the structure of the overall channel model, which mirrors the practical transmission link by modeling one span at a time and performing iterative calculations for multi-span long-haul transmission.

Feature decoupling is applied to each single-span model, where the linear and nonlinear channel effects are treated separately. The single-span FDD model consists of a linear one-step model, a nonlinear NN model, and additive random noise, as shown in Fig. 1(b). CD is modeled using a physical model identical to the linear model in SSFM, given by:

$$\boldsymbol{A}(z+L,\omega)=\boldsymbol{A}(z,\omega)(-i\frac{\beta_2}{2}\omega^2 L). \tag{5}$$

Nonlinear channel effects addressed using a NN. The effect of the EDFA is modeled as an additive random noise with known distributions. A bi-directional long-short term memory (BiLSTM) network is employed to model the nonlinear effects, leveraging its excellent capability to extract temporal features. Additionally, the BiLSTM employs a Seq2Seq framework during inference, producing multiple-symbol outputs per inference. Compared to traditional BiLSTM [20] or traditional Transformer [37], this significantly reduces repetitive calculations, saves computational time, and is particularly advantageous for multi-channel WDM systems. The training of Seq2Seq involves two stages. In the first stage, the model's input window is 60 symbols (20 + 20 + 20) with 20 padding symbols at front and back. In the second training stage, the input window is extended to 380 symbols (40 + 300 + 40), with 40 overlapping symbols. During inference, to further minimize redundant computations from overlapping symbols and reduce overall complexity, we adopt an input window of 1080 symbols (40 + 1000 + 40).

The FDD scheme is chosen as the foundation for this modeling approach due to its flexibility in adapting to various parameters, including modulation format and span number, as well as its ability to generalize to a certain degree across launch powers. Furthermore, by using a physical model for linear effects, it offers greater flexibility in generalizing linear parameters. As a result, FDD serves as an excellent basis for achieving more flexible and large-scale generalization across multiple dimensions.

2.3. Parameter encoding structure

The parameter encoding structure is crucial for enhancing generalizability. Parameters such as launch power, transmission distance, and sample interval, which characterize both the signals and the fiber channels, serve as valuable side information to improve waveform-level modeling performance under varying system setups. A straightforward approach might be to directly concatenate these parameters with the input signals; however, this results in performance degradation. This occurs because, compared to the analog-like waveform signal samples, parameters are fundamentally different in nature—they are lower-dimensional, discrete, and often in a different format. To address this issue, a pre-processing step is introduced to obtain an implicit, higher-dimensional representation of the parameters, known as parameter encoding. This encoding process allows the NN to effectively learn the relationships between signals and the various system parameters. Specifically, parameter encoding unifies the influence of different

Optics EXPRESS

physical parameters on the output waveform and enhances feature compatibility between the parameter and signal segments.

The schematic of the single-span model is shown in Fig. 1(b) and is described as follows. The encoding procedure is implemented by a small network, where a fully connected layer is employed in this work. Compared to convolutional or attention-based encoders, the fully connected layer is well-suited for learning direct mappings from low-dimensional parameter vectors to high-dimensional latent representations, a design principle widely adopted in large language models (LLMs) [38] for parameter embedding and feature expansion. The specific input forms and normalization methods for the different parameters are detailed in Table 2. The preprocessing includes unit changes and scale adjustments, so that the parameter information has a clear physical meaning and has the same magnitude as the normalized input signal, which is convenient for model training convergence. The dimension-raising factor is defined as the ratio of the output dimension to the input dimension of this network, which is set to 4 in this work. This value provides an effective balance between modeling accuracy and computational complexity. The encoded output is then concatenated with the signal matrix that has undergone pre-processing steps, including feature decoupling, normalization, and data arrangement. This concatenated matrix is subsequently fed into the BiLSTM for modeling the nonlinear characteristics. After processing through the parameter encoding structure, the input dimension of BiLSTM is 96.

Table 2. Preprocessing of the encoded parameters

Parameters	Actual input form
Launch power	Mean power of WDM signals, in milliwatts (mW).
Span length	Length of a single span in kilometers (km) and divided by 80.
Symbol rate	Sampling interval, in picoseconds (ps).
Channel space	Ratio of the transmission rate in GBaud to the channel space in GHz.

2.4. Modeling performance evaluation

The performance evaluation of NN-based optical fiber channel waveform modeling encompasses both accuracy assessment and complexity assessment. Accuracy is evaluated based on waveform modeling accuracy and transmission performance prediction accuracy. The results of SSFM serve as the reference value.

Waveform modeling accuracy is quantified using the normalized mean square error (NMSE), defined as:

$$NMSE = \sum_{i=1}^{N_{data}} |\hat{y}_i - y_i|^2 / \sum_{i=1}^{N_{data}} |y_i|^2, \tag{6}$$

where y_i represents the output samples of SSFM and $\widehat{y}_i$ represents the output of NN. NMSE can be immediately calculated from the modeling results.

To assess transmission performance prediction accuracy, the output signals are processed through receiver DSP, and the Q-factor is calculated based on the bit error rate (BER):

$$Q = 20\log_{10}(\sqrt{2}erfc^{-1}(2BER)). \tag{7}$$

The Q-factor error between SSFM and NN is the metric of transmission performance prediction accuracy.

Complexity assessment is performed by comparing the computational time required for SSFM and NN to transmit the same signal. The acceleration ratio of NN relative to SSFM can be expressed as:

$$Ratio = \frac{Time_{SSFM} - Time_{NN}}{Time_{SSFM}} \tag{8}$$

Optics EXPRESS

3. Performance gain of parameter encoding structure

To demonstrate the generalization performance gain achieved by the parameter encoding structure, we compare it with two other schemes: one without parameter input and one with direct parameter input. For clarity in subsequent discussions, the parameter encoding scheme is referred to as "Encode," the scheme without parameter input as "W/o," and the scheme with direct parameter input as "Direct." We evaluate the performance of these schemes under four distinct generalization conditions to comprehensively highlight the advantage of the Encode scheme. These test conditions include generalization of launch power, span length, symbol rate and WDM channel space. These parameters are selected because they have a significant impact on system transmission performance, and strong generalization capability with respect to these parameters is essential for the effective design and optimization of optical transmission systems. For example, under varying symbol rates, different optimal launch powers need to be selected to balance ASE noise and nonlinear noise. Additionally, in practical optical transmission links, the lengths of each fiber span are often different, and NNs need to be capable of adapting to configurations with varying fiber span lengths.

The training and testing datasets are identical for all three schemes, as detailed in Table 3. To mitigate overfitting and reduce pattern prediction [39,40] by NNs, different random number seeds are used to generate the bit sequences during the training and testing processes. Each condition in the training set contains a sequence of 61,440 symbols generated from four different random number seeds. For each scheme, the BiLSTM consists of 3 layers with a hidden size of 80. The Smooth L1 loss is used as the loss function. The number of training epochs is set at 500. Adam is employed as the optimizer. The learning rate (LR) is 2×10^{-3} initially and decreases during the training process. The LR for each parameter group follows a cosine annealing schedule. During the testing process, for each configuration, a sequence including 2^{18} symbols per WDM channel is employed. FO and PN are not included in the results described in Section III, as their influence on the model's performance are found to be negligible, as shown in Fig. 8(e) later. This simplification does not affect the validity of the comparative results.

3.1. Launch power

To evaluate the performance gain of Encode scheme for the launch power generalization, three models are trained using a dataset with varying launch power. The training dataset contains 6 launch power values, specifically, 1.5, 3.54, 4.68, 5.19, 5.64 6.05 dBm/channel, covering a range of approximately 4.5 dB. These values are more concentrated in the higher power region, where nonlinearities are more pronounced and can be easier captured by NN. The test launch power conditions range from -2 to 7 dBm/channel, with a 0.5 dB interval, covering a 9 dB range. This range spans both linear and nonlinear regions, which is beneficial for identifying the optimal launch power and for verifying the performance of nonlinear compensation algorithms in the high nonlinear regions. Additionally, this test set extends beyond the training range to assess the model's generalization performance outside the training dataset.

Figure 2 compares the NMSE of the three models across the launch power range of -2 to 7 dBm/channel after 800 km transmission with the span length of 80 km. It can be observed that the NMSE of the Encode scheme consistently outperforms the other two schemes across all power conditions. The average NMSE of the Encode scheme, under different power levels, is reduced by 49.9% and 39.1% compared to the Direct and W/o schemes, demonstrating its superior waveform modeling performance. Additionally, outside the range of the training dataset, specifically between -2 to 1 dBm and 6 to 7 dBm, both the Direct and W/o scheme exhibit significant accuracy degradation, indicating that these models have only been fitted to the training set features. In contrast, the Encode scheme shows only slight degradation outside the training range, highlighting its ability to better capture the influence of launch power on channel characteristics and can be

Optics EXPRESS

effectively applied to scenarios beyond the training set. This demonstrates a significant advantage of the Encode scheme in large-scale launch power generalization.

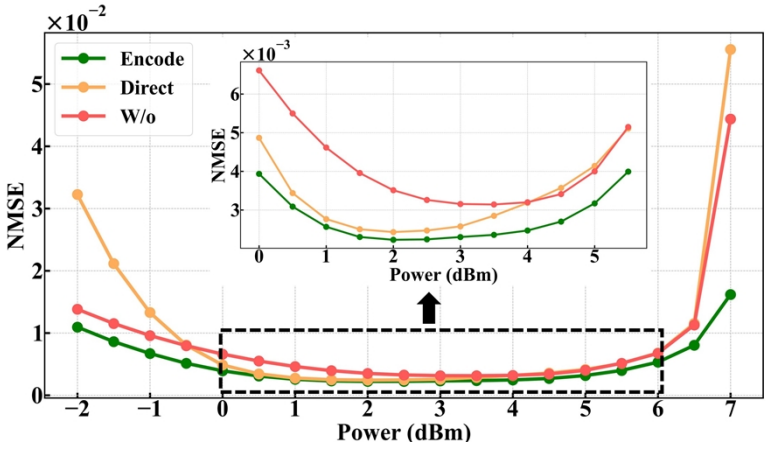

Fig. 2. NMSE of the Encode, Direct and W/o schemes for the launch power generalization, evaluated across launch powers ranging from -2 to 7 dBm/channel after 800 km transmission.

3.2. *Span length*

The distributed structure of the FDD scheme allows for coarse-grained distance generalization at integer multiples of the fixed span length, such as 80 km [20,21]. However, to achieve more flexible distance generalization—where the span length can be arbitrarily adjusted over a wide range—the model must predict outputs for various span lengths. To evaluate the performance gain of Encode scheme for the span length generalization, we construct a training dataset that includes span lengths of 30, 50, and 80 km, and train models with three different parameter input design. During testing, the span lengths range from 30 km to 100 km, with an interval of 10 km. Figure 3 (a) shows the single-span NMSE and Fig. 3(b) corresponds to the NMSEs after 10-span iteration at each test span length. Similar to the results in the launch power generalization, the other two models without parameters encoding represent higher NMSEs and an overfitting problem at the training conditions. This issue becomes more pronounced as the iteration number increases, further emphasizing the benefits of the Encode scheme in enhancing generalization.

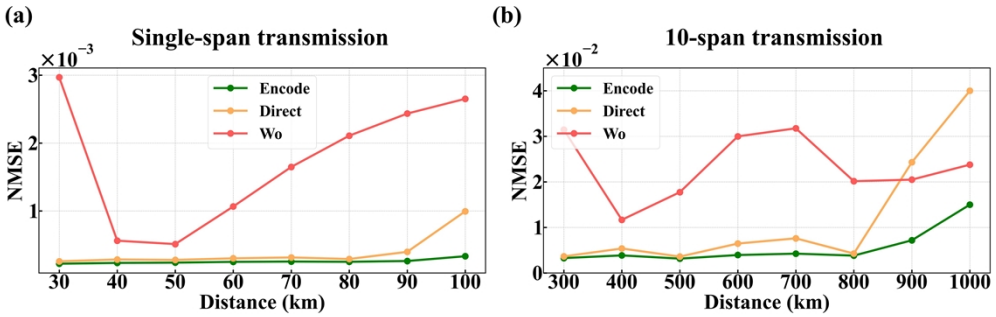

Fig. 3. NMSE at different span lengths of the Encode, Direct and W/o schemes with span length generalization training. (a) Single-span transmission. (b) 10-span transmission.

Optics EXPRESS

The long-haul transmission modeling performance is tested under 5 random span length configurations, where the length of each span varies from 30 km to 100 km, with a total of 15 spans. The total transmission distance ranges from 911 km to 1015 km. The detailed span length configurations are depicted in Fig. 4 (a). Figure 4(b) and (c) compare the NMSE and the Q-factor errors, respectively, of the three models at 3.5 dBm/channel launch power. For all 5 test configurations, the maximum Q-factor error of the Encode scheme remains below 0.25 dB. The average Q-factor prediction error and the NMSE of Encode scheme after 15-span transmission are reduced by 65.10% and 54.81%, respectively, compared to the W/o scheme, demonstrating the significant improvement in both accuracy and generalization performance offered by the Encode scheme.

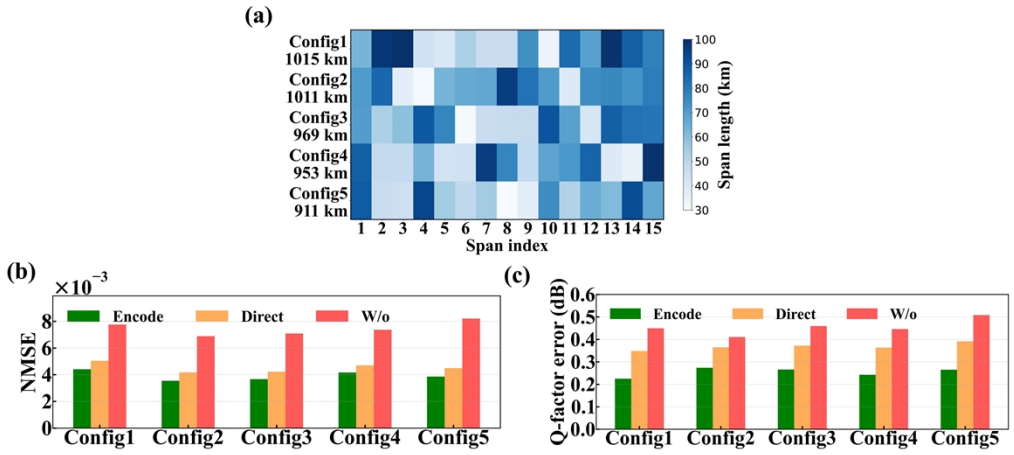

Fig. 4. Generalization results of the Encode, Direct and W/o scheme under random span length configurations with a total of 15 spans at the launch power of 3.5 dBm/channel. (a) The detailed span lengths and the total distances of the 5 test configurations. (b) NMSE. (c) Q-factor error.

3.3. *Symbol rate and WDM channel space*

The symbol rate is a critical signal parameter that greatly influences both the linear and nonlinear channel effects. For WDM transmission, channel space is another key parameter determining inter-channel interferences. To evaluate the generalization ability of both symbol rate and WDM channel space, we consider both the change of them. The model is only trained at 30 GBaud and 50 GHz channel space, while is tested at three groups of conditions, including [30 GBaud, 50 GHz], [40 GBaud, 50 GHz] and [30 GBaud, 60 GHz]. The corresponding output power spectrums of the model are illustrated in Fig. 5, with time-domain NMSE after 800 km presented below. In Fig. 5(b) and (c), significant abnormalities are observed in the output signal spectra, accompanied by large NMSE values, even with small changes in one of the two parameters. These results indicate that the model trained under a single condition is not flexible enough to adapt to change in both parameters simultaneously. Consequently, such limited generalization hinders the ability of the NN to support flexible parameter tuning during optical system optimization.

To evaluate the performance gain of Encode scheme for the symbol rate and WDM channel space generalization, we construct a dataset containing 5 groups of the symbol rate and channel space combinations. Specifically, the symbol rates are 20, 25, 30, 40, 50 GBaud, paired with the channel spaces of 25, 30, 35, 45, 60 GHz, respectively. The dataset composition is detailed in Table 3. In this case, the input parameters are the sample time interval and the ratio of symbol rate to channel space. Figure 6(a) and (b) compare the NMSE for 10 span, 800 km transmission

Optics EXPRESS

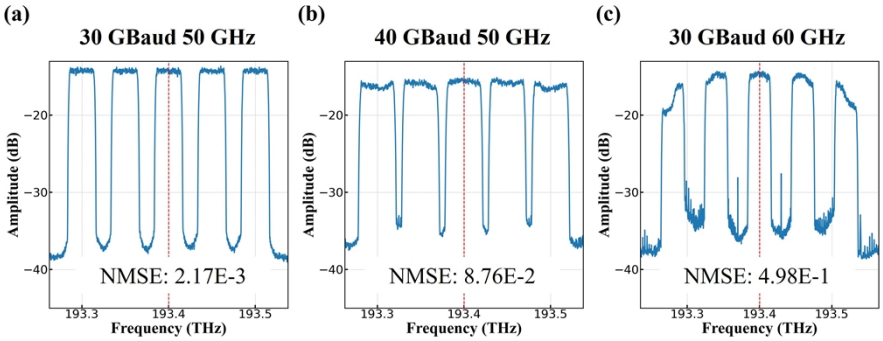

Fig. 5. Output power spectrums and NMSEs of model trained only at condition of [30 GBaud, 50 GHz], and tested at three groups of conditions. (a) [30 GBaud, 50 GHz] (b) [40 GBaud, 50 GHz] (c) [30 GBaud, 60 GHz].

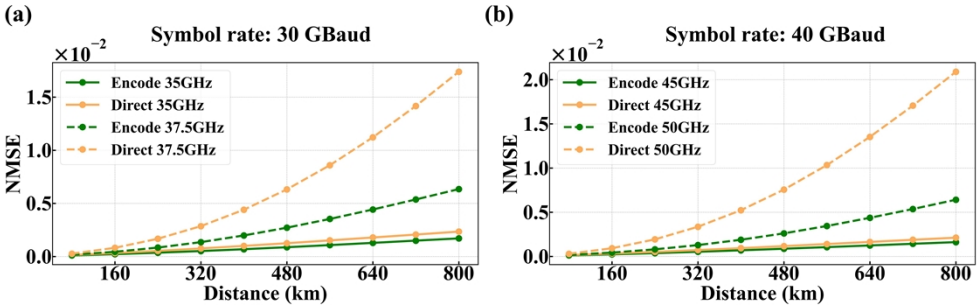

Fig. 6. NMSE of the Encode and Direct schemes under symbol rate and channel space generalization training, evaluated on channel space conditions both included in and excluded from the training dataset, for 3.5 dBm/channel, 800 km transmission. (a) 30 GBaud. (b) 40 GBaud.

Table 3. Training and testing setup under launch power, symbol rate, WDM channel space and span length generalization

Generalization type	Train or test	Launch power (dBm/channel)	Symbol rate and WDM channel space (GBaud, GHz)	Span length (km)
Generalization of **launch power**	Train	[1.5, 3.54, 4.68, 5.19, 5.64, 6.05]	[30, 50]	80
	Test	From -2 dBm to 7 dBm with 0.5 dB step		
Generalization of **span length**	Train	3.5	[30, 50]	[30, 50, 80]
	Test			From 10 km to 100 km with 10 km step
Generalization of **symbol rate and WDM channel space**	Train	3.5	[20, 25], [25, 30], [30, 35], [40, 45], [50, 60]	80
	Test		[30, 35], [30, 37, 5] [40, 45], [40, 50]	

Optics EXPRESS

of the two models (Encode and Direct) at 3.5 dBm/channel, under two symbol rates: 30 GBaud and 40 GBaud. For each test symbol rate, we evaluate both channel space conditions that are included in the training dataset and those that are not, to thoroughly test generalizability. The Direct scheme represents significantly larger NMSEs under conditions not covered by the training dataset, suggesting its poor generalizability. In contrast, the Encode scheme maintains NMSEs no larger than 7×10^{-3}. Compared to the Direct scheme, the Encode scheme reduces NMSEs by 27.2% and 23.8%, respectively, for the two training conditions of [30 GBaud 35 GHz] and [40 GBaud 45 GHz]. For the two out-of-training-range test scenarios, the NMSE is reduced by 63.5% and 69.3%, demonstrating a significant improvement in generalization of symbol rate and WDM channel space.

4. Enhanced generalized NN-based channel modeling for optical transmission systems

4.1. Transmission performance prediction across various system parameters

The Encode scheme has demonstrated excellent generalization performance. Furthermore, to enable the design and optimization of optical fiber transmission system, it is essential to flexibly adjust various system parameters and accurately predict the transmission performance [27,36]. By integrating the Encode and FDD scheme, for the first time, we achieve simultaneous generalization across night different system parameters through a single NN. This enhanced generalized NN can accurately predict transmission performance across diverse system configurations while maintaining low computational complexity compared to SSFM. A schematic diagram illustrating these generalizable parameters in the optical transmission system is shown in Fig. 7. These parameters include modulation format, symbol rate, WDM channel space, FO and PN of lasers, launch power, span length, total transmission distance and channel aCD. To achieve the generalization across launch power, symbol rate, WDM channel space and span length, we need to construct a dataset that combines these parameters in various combinations for model training. Each condition contains 61,440 symbols generated by two different random number seeds. Based on NVIDIA GeForce RTX 3090 Graphics Processing Unit (GPU) with 24 GB of memory, the time required for SSFM-based data generation is approximately 220 minutes, which is acceptable for practical applications. The BiLSTM consists of 3 layers with a hidden size of 240. The training epoch is set to 1000 with an initial LR of 2×10^{-3}.

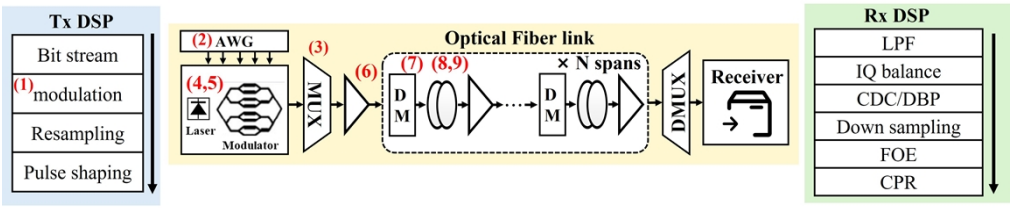

Fig. 7. A schematic diagram illustrating the generalization parameters of a single NN in the optical fiber transmission system.

To evaluate the NN-based predicted results of transmission performance across various parameters, we independently change each system parameter and compare the results between NN and SSFM. Figure 8(a) presents the constellations for the generalization of modulation format. The training set of the NN includes the 16QAM modulation format. During the testing process, probabilistic shaping 16QAM (PS-16QAM), 64QAM, and geometric shaping 64QAM

Optics EXPRESS

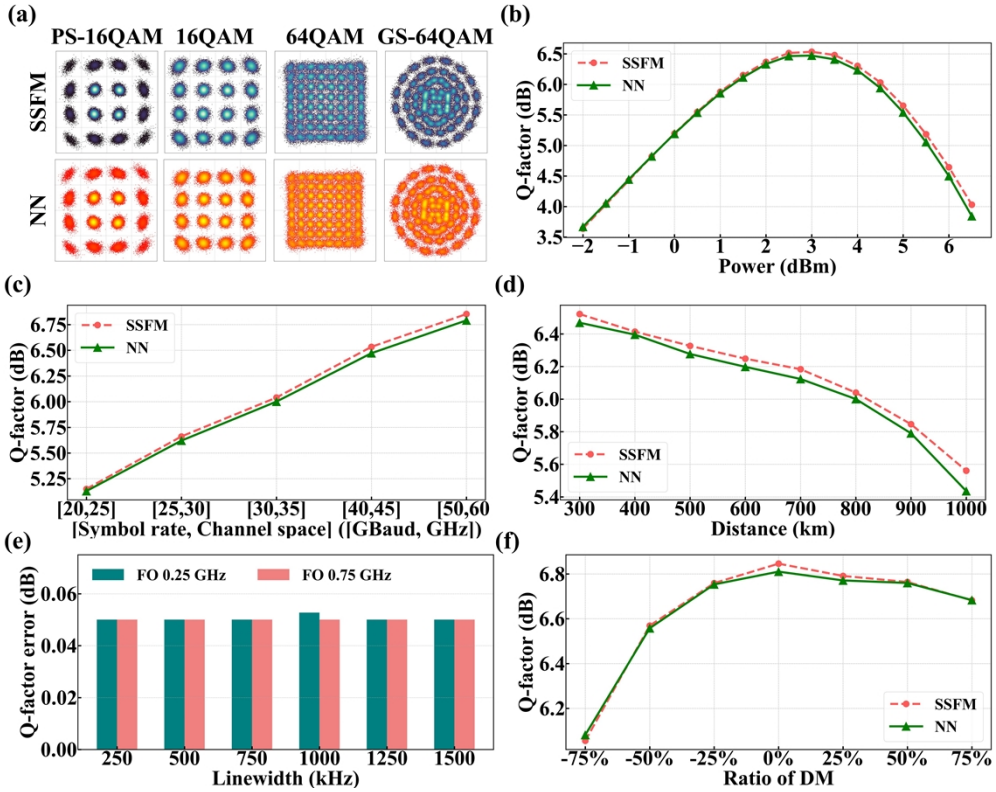

Fig. 8. Generalization performance of the NN across various system parameters in comparison with the SSFM. (a) Constellations for modulation format generalization. (b) Q-factor results for launch power generalization. (c) Q-factor results for symbol rate and WDM channel space generalization. (d) Q-factor results for transmission distance generalization. (e) Q-factor error results across varying FO and PN of laser. (f) Q-factor results for ratio of DM generalization.

(GS-64QAM) are adopted. Under these different modulation formats, the constellations predicted by the NN are similar to those of SSFM, demonstrating its excellent generalization ability of modulation formats. This capability allows the NN to be used as a differential surrogate channel [6–9] to achieve E2E optimization for constellation shaping, thereby improving the performance of the optical transmission system.

Figure 8(b), (c) and (d) show the Q-factor error curves of NN and SSFM under varying launch power, symbol rate, WDM channel space, and transmission distance. Under these different parameter configurations, the average Q-factor errors between SSFM and NN are 0.06 dB, 0.05 dB, and 0.06 dB, respectively, demonstrating that NN can accurately predict transmission performance across different parameter configurations. These favorable results are attributed to the Encode scheme, which enhances the generalization ability of the NN, enabling it to learn the channel feature changes under different parameter configurations. Thanks to this generalization ability, the NN can be used to search for the optimal launch power at varying symbol rates or transmission distances, significantly improving the efficiency of system parameter optimization.

Figure 8(e) shows the Q-factor errors between NN and SSFM under different FO and linewidths of lasers. According to formula (1), linewidth directly affects the PN of the laser. The green and red bar graphs represent the Q-factor errors obtained when the linewidth varies from 250 to 1500

Optics EXPRESS

kHz under FO values of 0.25 and 0.75 GHz, respectively. Regardless of the FO and linewidth combinations, the Q-factor errors of NN and SSFM remain below 0.06 dB, with very small fluctuations. This confirms that NN has excellent generalization ability for PN and FO of the laser. The main reasons for the successful generalization of PN and FO lie in the linear decoupling mechanism of the FDD and its waveform-level modeling capability. In the FDD scheme, linear effects are modeled using a physical model, allowing the system to generalize across different conditions by adjusting the parameters of the physical model. Furthermore, PN and FO are laser-induced impairments that only perturb the input waveform of the optical fiber channel and do not alter its intrinsic nonlinear response. Therefore, the nonlinear response function of the optical fiber fitted by BiLSTM is still applicable in different PN and FO conditions. For changes in the input waveform, the waveform-level modeling scheme can achieve the generalization of different input waveform characteristics. This feature allows researchers to flexibly adjust laser parameters and use NN to quickly verify the performance of FOE and CPR algorithms [27,35,36], assisting in the rapid design of those algorithms.

Figure 8(f) shows the Q-factor error curves of NN and SSFM under different R_{DM}. As the R_{DM} varies from -75% to 75%, both NN and SSFM exhibit similar Q-factor variation trends, with an average Q-factor error of only 0.02 dB. This similarity arises from the distributed architecture of FDD and the separate modeling of its linear and nonlinear characteristics, allowing it to flexibly handle the changes in dispersion intensity before each span by adjusting the parameters of the linear model. This advantage enables NN to quickly predict transmission performance under different DM configurations, providing convenience for large-scale DM optimization [34].

4.2. Computational complexity

To evaluate the low complexity advantage of NN, we compare the running times between NN and SSFM. SSFM employs both the nonlinear phase rotation step size configuration method (NP-SSFM) and the constant step size configuration method (C-SSFM). At the transmitter, each WDM channel generates 2^{20} symbols for transmission, with a total transmission distance of 800 km. The codes for SSFM and NN are executed on the same server, equipped with an NVIDIA GeForce RTX 3090 GPU with 24 GB of memory. Table 4 shows the running times of different models under various launch powers. Figure 9 illustrates the acceleration ratio of NN relative to NP-SSFM under different launch powers. It can be observed that the acceleration ratio increases with the launch power, as the stronger nonlinear effects in the channel at higher power levels require more iterations for NP-SSFM, while NN's running time remains unaffected by changes in launch power. At launch powers of -1.5 dBm and 6.0 dBm, the running time of NN is 6.02 s, while NP-SSFM runs for 148 s and 812 s, respectively. The running time of C-SSFM is 6339 s. The acceleration ratios of NN relative to NP-SSFM are 95.9% and 99.3%, and the acceleration ratio relative to C-SSFM is 99.9%. These results clearly demonstrate that NN significantly reduces running time compared to SSFM, particularly in high nonlinear regions, due to its efficient parallel computing capacities on GPUs. The reduced computational complexity allows the NN to overcome the complexity limitations of SSFM, positioning it as a fast and efficient simulation tool for design and optimization of optical transmission systems.

Table 4. Running time of NN, NP-SSFM, and C-SSFM under various launch powers

Launch power	NN	NP-SSFM	C-SSFM
-1.5 dBm/channel	6.02 s	148 s	6339 s
6.0 dBm/channel	6.02 s	812 s	6339 s

Optics EXPRESS

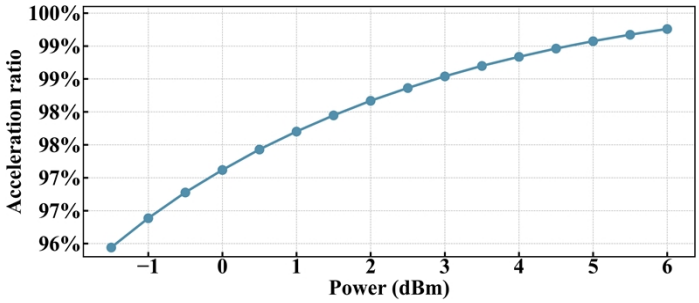

Fig. 9. Acceleration ratio of NN relative to NP-SSFM across different launch powers.

4.3. Application for nonlinear compensation algorithm

To further enhance the capacity of optical transmission systems, developing advanced nonlinear compensation algorithms to mitigate the effects of fiber nonlinearity is crucial. It is crucial to test their compensation results under different launch powers and nonlinear intensities. Furthermore, balancing performance and complexity across different system parameter configurations requires fine optimization of hyperparameters in the algorithm, such as the step size configuration in DBP [1,5]. This optimization process typically involves continuous system simulations to evaluate the algorithm's performance, which can be time-consuming, especially when using traditional channel simulation methods. The low computational complexity of NN helps alleviate the long processing times associated with traditional SSFM. Furthermore, the robust nonlinear modeling and generalization capabilities of NNs enable them to yield results comparable to SSFM, significantly improving the efficiency of nonlinear compensation algorithm design.

To evaluate the effectiveness of the enhanced generalized NN for nonlinear compensation algorithms, we compare results between NN and SSFM across the CDC algorithm and DBP algorithm with different step size configurations. The step size configurations of DBP algorithm are set to 3 and 10 steps per span (StPS), respectively, which represents different nonlinear compensation capabilities. The simulation system operates with a symbol rate of 40 GBaud, a WDM channel spacing of 45 GHz, and launch power varying between -2.0 to 6.5 dBm/channel. Figure 10 shows the Q-factor values of signals modeled by NN and SSFM after processing by different algorithms. It is evident that, across various launch power levels, the average Q-factor errors between signals modeled by NN and SSFM after processing by CDC, 3StPS-DBP, and 10StPS-DBP are only 0.06 dB, 0.04 dB, and 0.04 dB, respectively. This clearly demonstrates that, under different launch powers, the NN model replicates the channel characteristics of SSFM with high accuracy, making it an effective tool for nonlinear compensation algorithm design.

To further validate the effectiveness of NN in nonlinear compensation algorithms across varying system configurations, we change the symbol rate and modulation format and perform performance verification for the 10StPS-DBP algorithm. The tested symbol rates are 30, 40, and 45GBaud, with modulation formats of 8QAM, PS-16QAM with a MI of 3.6 bits/symbols, and standard 16QAM, respectively. Figure 10(b) shows the Q-factor of NN and SSFM after processing by 10StPS-DBP under different parameter configurations. Under the configurations of [30GBaud, 8QAM], [40GBaud, PS-16QAM], and [45GBaud, 16QAM], the average Q-factor errors between NN and SSFM are only 0.04 dB, 0.07 dB, and 0.04 dB, respectively. These results demonstrate that DBP algorithms with varying step size configurations exhibit similar compensation performance differences in both NN and SSFM models under varying launch power, modulation format, and symbol rate configurations. This further confirms the effectiveness of applying NN to nonlinear compensation algorithms.

Optics EXPRESS

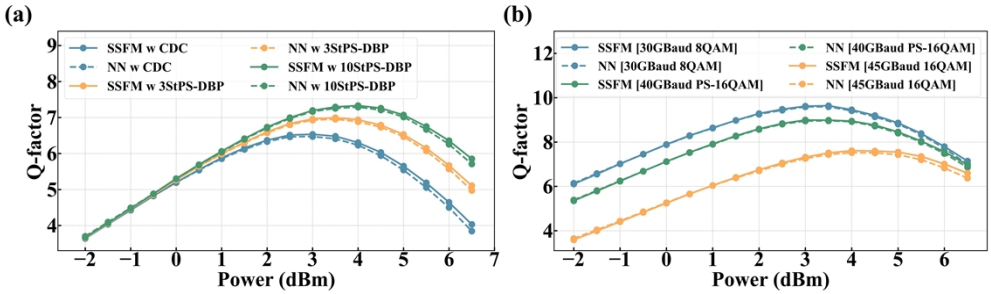

Fig. 10. Q-factor performance of SSFM and NN under various launch power after 800 km transmission. (a) Results for 30 GBaud and 16 QAM signals after CDC, 3StPS-DBP and 10StPS-DBP processing. (b) Results for signals with different symbol rates and modulation formats—specifically, [30 GBaud 8QAM], [40 GBaud PS-16QAM], and [45 GBaud 16QAM]—after 10StPS-DBP processing.

5. Conclusion

In this paper, we propose a novel parameter encoding structure integrated with the FDD scheme, enhancing the generalization capability of the NN-based channel waveform modeling scheme. In the generalized scenario, with a wide range of launch power from -2 to 7 dBm and arbitrary transmission distances, the proposed parameter encoding structure improves waveform modeling accuracy by 49.9% and 69.7% compared to the non-encoded scheme. Furthermore, the model maintains low computational complexity, achieving a 99.3% reduction in computation time relative to SSFM. To the best of our knowledge, this work presents the first NN-based model capable of generalizing across multiple system parameters simultaneously—modulation format, symbol rate, WDM channel spacing, PN and FO of lasers, launch power, aCD, span length, and total transmission distance. This enhanced generalized NN holds significant potential for the design and optimization of optical transmission systems. This enhanced generalized NN holds significant potential for the design and optimization of optical transmission systems. It can accurately predict transmission performance across various system parameter configurations, with a Q-factor error of less than 0.12 dB compared to SSFM. Additionally, the NN demonstrates comparable performance to the SSFM implemented in DBP algorithms with varying step size configurations, highlighting its potential in the design and optimization of nonlinear compensation algorithms. The superior generalization ability and computational efficiency of NN-based method overcome the complexity limitations of traditional waveform modeling methods, offering a promising solution for the design and optimization of next-generation optical communication systems.

Funding. National Key Research and Development Program of China (2023YFB2905400); National Natural Science Foundation of China (62025503); Shanghai Jiao Tong University 2030 Initiative.

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. Q. Fan, G. Zhou, T. Gui, *et al.*, "Advancing theoretical understanding and practical performance of signal processing for nonlinear optical communications through machine learning," Nat. Commun. **11**(1), 3694 (2020).
2. S. Deligiannidis, A. Bogris, C. Mesaritakis, *et al.*, "Compensation of Fiber Nonlinearities in Digital Coherent Systems Leveraging Long Short-Term Memory Neural Networks," J. Lightwave Technol. **38**(21), 5991–5999 (2020).
3. X. Lin, S. Luo, S. K. O. Soman, *et al.*, "Perturbation Theory-Aided Learned Digital Back-Propagation Scheme for Optical Fiber Nonlinearity Compensation," J. Lightwave Technol. **40**(7), 1981–1988 (2022).

Optics EXPRESS

4. Z. Niu, H. Yang, L. Li, *et al.*, "Learnable digital signal processing: a new benchmark of linearity compensation for optical fiber communications," Light: Sci. Appl. **13**(1), 188 (2024).
5. E. Ip and J. M. Kahn, "Compensation of Dispersion and Nonlinear Impairments Using Digital Backpropagation," J. Lightwave Technol. **26**(20), 3416–3425 (2008).
6. B. Karanov, M. Chagnon, V. Aref, *et al.*, "Concept and Experimental Demonstration of Optical IM/DD End-to-End System Optimization using a Generative Model," in *Optical Fiber Communication Conference (OFC) 2020*, OSA Technical Digest (Optica Publishing Group, 2020), Th2A.48.
7. Z. Niu, H. Yang, H. Zhao, *et al.*, "End-to-End Deep Learning for Long-haul Fiber Transmission Using Differentiable Surrogate Channel," J. Lightwave Technol. **40**(9), 2807–2822 (2022).
8. L. Li, Z. Niu, J. Xiao, *et al.*, "Joint Pre- and Post-Learned Perturbation Nonlinearity Compensation Optimization for Long-haul Optical Fiber Transmission Based on End-to-end Deep Learning," in *Optical Fiber Communication Conference (OFC) 2025*, Technical Digest Series (Optica Publishing Group, 2025), W1 K.6.
9. Z. Li, C. Wang, J. Jia, *et al.*, "Model-Driven Deep-Learning for End-to-End Optimization in Fiber-Terahertz Communication Systems," J. Lightwave Technol. **43**(7), 3099–3117 (2025).
10. G. P. Agrawal, "Chapter 2 - Pulse propagation in fibers," in *Nonlinear Fiber Optics (Sixth Edition)*, G. P. Agrawal, ed. (Academic Press, 2019), pp. 27–55.
11. P. Poggiolini, "The GN Model of Non-Linear Propagation in Uncompensated Coherent Optical Systems," J. Lightwave Technol. **30**(24), 3857–3879 (2012).
12. P. Poggiolini, G. Bosco, A. Carena, *et al.*, "The GN-Model of Fiber Non-Linear Propagation and its Applications," J. Lightwave Technol. **32**(4), 694–721 (2014).
13. A. Carena, G. Bosco, V. Curri, *et al.*, "EGN model of non-linear fiber propagation," Opt. Express **22**(13), 16335–16362 (2014).
14. D. Semrau, R. I. Killey, and P. Bayvel, "A Closed-Form Approximation of the Gaussian Noise Model in the Presence of Inter-Channel Stimulated Raman Scattering," J. Lightwave Technol. **37**(9), 1924–1936 (2019).
15. P. Serena, C. Lasagni, S. Musetti, *et al.*, "On Numerical Simulations of Ultra-Wideband Long-Haul Optical Communication Systems," J. Lightwave Technol. **38**(5), 1019–1031 (2020).
16. D. Wang, Y. Song, J. Li, *et al.*, "Data-driven Optical Fiber Channel Modeling: A Deep Learning Approach," J. Lightwave Technol. **38**(17), 4730–4743 (2020).
17. Y. Song, D. Wang, Q. Fan, *et al.*, "Physics-Informed Neural Operator for Fast and Scalable Optical Fiber Channel Modelling in Multi-Span Transmission," in *European Conference on Optical Communication (ECOC) 2022*, Technical Digest Series (Optica Publishing Group, 2022), We5.32.
18. H. Yang, Z. Niu, S. Xiao, *et al.*, "Fast and Accurate Optical Fiber Channel Modeling Using Generative Adversarial Network," J. Lightwave Technol. **39**(5), 1322–1333 (2021).
19. Q. Qiu, H. Lun, X. Liu, *et al.*, "Fourier Neural Operator Based Fibre Channel Modelling for Optical Transmission," in *European Conference on Optical Communication (ECOC) 2022*, Technical Digest Series (Optica Publishing Group, 2022), We3B.2.
20. H. Yang, Z. Niu, H. Zhao, *et al.*, "Fast and Accurate Waveform Modeling of Long-Haul Multi-Channel Optical Fiber Transmission Using a Hybrid Model-Data Driven Scheme," J. Lightwave Technol. **40**(14), 4571–4580 (2022).
21. X. He, L. Yan, L. Jiang, *et al.*, "Fourier Neural Operator for Accurate Optical Fiber Modeling With Low Complexity," J. Lightwave Technol. **41**(8), 2301–2311 (2023).
22. X. Zhang, M. Zhang, Y. Song, *et al.*, "DeepONet-Based Waveform-Level Simulation for a Wideband Nonlinear WDM System," J. Lightwave Technol. **41**(22), 6908–6922 (2023).
23. T. Zhang, Y. Pan, J. Zheng, *et al.*, "Fourier Neural Operator Accelerated Fast and Accurate Optical Fiber Transmission Modeling," J. Lightwave Technol. **43**(8), 3592–3604 (2025).
24. J. Liu, R. Wang, H. Lin, *et al.*, "UVL-based hybrid neural network model for generalized optical fiber channel modeling," Opt. Express **33**(13), 28233–28242 (2025).
25. M. Shi, Z. Niu, H. Yang, *et al.*, "Fast and accurate waveform modeling based on sequence-to-sequence framework for multi-channel and high-rate optical fiber transmission," Opt. Lett. **50**(7), 2286–2289 (2025).
26. C. Zeng, Z. Niu, H. Yang, *et al.*, "Enhancing Generalization in Neural Channel Model for Optical Fiber WDM Transmission through Learned Encoding of System Parameters," in *Optical Fiber Communication Conference (OFC) 2024*, Technical Digest Series (Optica Publishing Group, 2024), M4 K.1.
27. T. Pfau, S. Hoffmann, and R. Noe, "Hardware-Efficient Coherent Digital Receiver Concept With Feedforward Carrier Recovery for M -QAM Constellations," J. Lightwave Technol. **27**(8), 989–999 (2009).
28. D. Marcuse, C. R. Manyuk, and P. K. A. Wai, "Application of the Manakov-PMD equation to studies of signal propagation in optical fibers with randomly varying birefringence," J. Lightwave Technol. **15**(9), 1735–1746 (1997).
29. J. Shao, X. Liang, and S. Kumar, "Comparison of Split-Step Fourier Schemes for Simulating Fiber Optic Communication Systems," IEEE Photonics J. **6**(4), 1–15 (2014).
30. O. V. Sinkin, R. Holzlohner, J. Zweck, *et al.*, "Optimization of the split-step Fourier method in modeling optical-fiber communications systems," J. Lightwave Technol. **21**(1), 61–68 (2003).
31. V. A. Konyshev, A. V. Leonov, O. E. Nanii, *et al.*, "Correlation of nonlinear noises from different spans in 100Gb/s multi-span fiber optic lines," Opt. Commun. **381**, 352–359 (2016).
32. P. Gou, L. Zhao, K. Wang, *et al.*, "Nonlinear Look-Up Table Predistortion and Chromatic Dispersion Precompensation for IM/DD PAM-4 Transmission," IEEE Photonics J. **9**(5), 1–7 (2017).

Optics EXPRESS

33. K. Keykhosravi, M. Secondini, G. Durisi, *et al.*, "How to Increase the Achievable Information Rate by Per-Channel Dispersion Compensation," J. Lightwave Technol. **37**(10), 2443–2451 (2019).
34. C. Zeng, Z. Niu, H. Yang, *et al.*, "Neural Channel Model-Aided Distributed Dispersion Manage Optimization for Mitigation of Nonlinearities and Dispersion in Fibre Transmission," in *2024 Asia Communications and Photonics Conference (ACP) and International Conference on Information Photonics and Optical Communications (IPOC)*, 2024), 1–4.
35. A. Leven, N. Kaneda, U. V. Koc, *et al.*, "Frequency Estimation in Intradyne Reception," IEEE Photonics Technol. Lett. **19**(6), 366–368 (2007).
36. X. Zhou, "An Improved Feed-Forward Carrier Recovery Algorithm for Coherent Receivers With M -QAM Modulation Format," IEEE Photonics Technol. Lett. **22**(14), 1051–1053 (2010).
37. M. Shi, H. Yang, Z. Niu, *et al.*, "Accurate and Efficient Optical Fiber WDM Transmission Modeling Using the Encoder-Only Transformer with Feature Decoupling Distributed Method," in *2023 Asia Communications and Photonics Conference/2023 International Photonics and Optoelectronics Meetings (ACP/POEM)*, 2023), 1–5.
38. T. B. Brown, B. Mann, N. Ryder, *et al.*, "Language models are few-shot learners," in *Proceedings of the 34th International Conference on Neural Information Processing Systems*, (Curran Associates Inc., Vancouver, BC, Canada, 2020), p. Article 159.
39. T. A. Eriksson, H. Bülow, and A. Leven, "Applying Neural Networks in Optical Communication Systems: Possible Pitfalls," IEEE Photonics Technol. Lett. **29**(23), 2091–2094 (2017).
40. L. Yi, T. Liao, L. Huang, *et al.*, "Machine Learning for 100 Gb/s/λ Passive Optical Network," J. Lightwave Technol. **37**(6), 1621–1630 (2019).