Optics Communications 591 (2025) 132099

Contents lists available at ScienceDirect

Optics Communications

journal homepage: www.elsevier.com/locate/optcom

Machine learning-based models for optical fiber channels

Yulin Wang [a], Mark Leeson [a], Zheng Liu [c], Sander Wahls [b], Tongyang Xu [e], Sergei Popov [d], Gan Zheng [a], Tianhua Xu [a,e,*]

[a] School of Engineering, University of Warwick, Coventry CV4 7AL, United Kingdom
[b] Karlsruhe Institute of Technology, Karlsruhe 76187, Germany
[c] Tianjin University, Tianjin 300072, China
[d] KTH Royal Institute of Technology, Stockholm SE-11419, Sweden
[e] Department of Electronic and Electrical Engineering, University College London, London WC1E 6BT, United Kingdom

ARTICLE INFO

Keywords:
Optical fiber communication
Channel modeling
Machine learning

ABSTRACT

This paper presents a comprehensive review of machine learning (ML) in optical fiber communications, particularly in channel modeling. It discusses the evolution from conventional methods to ML-based approaches that aim to enhance predictive and computational efficiency. Specifically, the discussions categorize ML methodologies into data-driven and principle-driven approaches. The former treats channel modeling as a "black box" providing rapid modeling capabilities at the expense of transparency and substantial data requirements. In contrast, the latter integrate physical principles into the ML-based system, enhancing model interpretability and reducing data dependency. In addition, the emergence of hybrid models that combine the strengths of both approaches is explored. This classification provides a structured overview of how ML is reshaping channel modeling in optical fiber communications, underscoring its potential to improve system design and exploring advanced nonlinear dynamics in optical fiber communication systems.

1. Introduction

The accuracy of channel modeling is critical in optical fiber communications for system simulation and design. Optical fiber channel dynamics are governed by the Nonlinear Schrödinger Equation (NLSE) [1]. The interplay of impairments such as chromatic dispersion (CD), nonlinearity, and amplified spontaneous emission (ASE) noise complicates the analytical solution of this equation, increasing the complexity of channel modeling [2]. Conventional channel modeling is predominantly conducted via the split-step Fourier Method (SSFM) [3], modeling the channel effects by numerically solving the NLSE. Nonetheless, SSFM involves numerous iteration steps, resulting in significant computational complexity. Moreover, as accuracy requirements, transmission distances, and distortion orders increase, the computational complexity is expected to increase substantially.

Machine learning (ML) has become ubiquitous across diverse domains, and its application in optical communications is rapidly expanding in response to escalating global internet traffic [4]. ML-based applications in optical communication now span software-defined networking, nonlinear fiber compensation, physical layer design and network management [5–7]. Motivated by the concurrent trends of rising system complexity, advancements in ML, and the requirements of channel modeling, a growing number of studies focus on utilizing ML-based algorithms to provide models for optical fiber channels. These models provide output pulses from given input pulses. Present methodologies are primarily classified into two categories based on their underlying approach: data-driven and principle-driven.

The data-driven approach regards channel modeling as a "black-box" problem, wherein the calculation operation of the data-driven method contains only multiplications and additions and no sophisticated operations. The strength of the data-driven method is that when it trains well, it can model the optical fiber channel much faster than the conventional method while reaching near accuracy. However, the data-driven method requires massive amounts of input–output data during training and the resulting models are hard to interpret for humans. The principle-driven method merges ML with the principles of physics, thus enhancing the interpretability and computational efficiency of ML through the incorporation of domain expertise. Distinct principle-driven methods utilize principles in different ways to satisfy varying requirements. For instance, the Physics-informed Neural Network (PINN) considers physics prior knowledge during training, Learned Digital Back-Propagation (LDBP) modifies SSFM within the framework of NNs and optimizes the parameters through them. Although the generalizability of the principle-driven method is acceptable after adjustment, fiber channels cannot be directly modeled

* Corresponding author at: School of Engineering, University of Warwick, Coventry CV4 7AL, United Kingdom.
E-mail address: tianhua.xu@ieee.org (T. Xu).

https://doi.org/10.1016/j.optcom.2025.132099
Received 13 February 2025; Received in revised form 28 May 2025; Accepted 9 June 2025
Available online 21 June 2025

Table 1
Comparison of features in existing reviews and our review.

Feature	Nev. [8]	Mus. [6]	Gau. [9]	Kha. [5]	Jia. [14]	Gen. [11]	Fre. [12]	Wan. [15]	Ours
Detailed classification	No	Partial	No	Partial	Partial	No	Partial	Partial	**Yes**
Systematic comparison	No	Limited	Limited	Limited	Limited	Limited	No	Limited	**Extensive**
Focus on interpretability	No	Limited	No	Limited	Yes	No	No	Yes	**Extensive**
Hybrid method coverage	No	No	No	No	Limited	No	No	No	**Yes**
Real-time modeling	No	Limited	No	Limited	No	No	No	Limited	**Yes**
Performance benchmarking	No	Partial	Yes (limited)	Partial	Partial	No	Partial	Partial	**Comprehensive**
Emerging challenges	Partial	Partial	No	Partial	No	Partial	No	Partial	**Extensive**

by existing principle-driven methods outside the situations they were adjusted. Furthermore, there are also hybrid data-principle methods, which learn from both input–output data and physical knowledge or split the channel effects and model them through a data-driven method and a physical method (not ML), respectively. For clarity in categorization, we distinctly classify Hybrid approaches into data-driven and principle-driven methods, focusing on their primary properties within specific approaches.

1.1. Related work and scope

Previous comprehensive reviews have extensively covered ML applications in optical communications from various perspectives. Nevin et al. provide an introductory overview, discussing key ML methods and their broad applications across optical fiber communications without extensive methodological classifications [8]. Musumeci et al. offer insights into ML applications covering optical networking and system-level aspects, yet lack detailed comparative evaluations and interpretability considerations [6]. Gautam et al. focus specifically on nonlinear optical pulse propagation, offering comparative studies that primarily evaluate computational performance rather than comprehensive methodological categorization [9]. Khan et al. highlight ML applications at the physical layer, emphasizing system-level issues but offering limited comparative analyses of model performances [5]. Jiang et al. explore Physics-Informed Neural Networks (PINNs), thoroughly validating their use for various physical phenomena without extensively addressing data-driven or hybrid approaches [10]. Genty et al. review ML applications within ultrafast photonics, primarily targeting laser system optimization without extensive consideration of channel modeling methods in optical fiber communications [11]. Freire et al. introduce Artificial Neural Networks (ANNs) specifically for photonic applications, providing insights into recent developments but lacking comprehensive comparative analyses [12]. Lastly, Wang et al. focus extensively on Physics-Informed Machine Learning (PIML) methods for optical fiber communications, with limited coverage of data-driven methods and hybrid methodologies [13].

In contrast, our review paper provides detailed classifications of ML-based channel modeling methodologies, explicitly differentiating between data-driven, principle-driven, and hybrid approaches. Unlike previous reviews, we provide systematic comparative evaluations across multiple dimensions, including temporal and spectral accuracy, nonlinear modeling capability, generalization potential, and computational complexity. Additionally, our paper extensively covers interpretability and introduces detailed discussions of hybrid methods, combining the strengths of both data-driven and principle-driven approaches. We further discuss emerging challenges, notably real-time modeling and adaptive model updates required by systems such as Digital Twins. The differences between our review and other existing reviews are summarized in Table 1.

Fig. 1 presents an overview of ML-based channel modeling methods. The left side of the figure details principle-driven approaches, while the right side outlines a variety of data-driven methods. The central portion of the figure identifies hybrid methods.

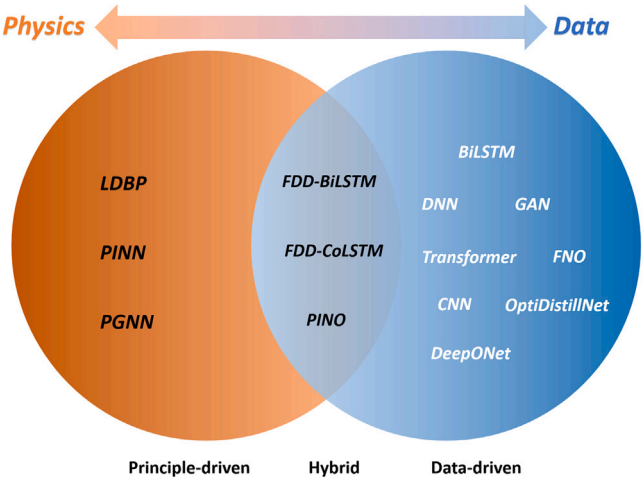

Fig. 1. Up-to-date ML-based channel modeling.

1.2. Contents and structure

The rest of the review is organized as follows. In Section 2, we will first describe the motivation of ML-based fiber channel modeling and then outline the existing ML applications in the subject. Section 2 also illustrates the fundamental concept of channel modeling and ML techniques, which are utilized by the methods discussed subsequently. In Sections 3–5, we will discuss data-driven, principle-driven and hybrid methods for channel modeling with respect to the literature review, framework, quantitative performance, and complexity. Section 6 first quantitatively compares existing papers in multi-dimensions and highlights their effectiveness for varying scenarios; then, it discusses open issues and obstacles, outlining future directions in this field. Finally, Section 7 concludes the paper. All the abbreviations in this review are listed in Table Appendix.

2. Background to channel modeling

In this section, we first illustrate the theory of the conventional channel modeling method and then introduce the motivation for ML-based channel modeling.

2.1. Theory

The transmission of light through a single-mode fiber is governed by the NLSE [1], expressed as

$$\frac{\partial A}{\partial z} = -\frac{\alpha}{2}A + i\frac{\beta_2}{2}\frac{\partial^2 A}{\partial t^2} + \frac{\beta_3}{6}\frac{\partial^3 A}{\partial t^3} + i\gamma|A|^2 A \tag{1}$$

wherein A denotes the complex envelope of the gradually changing optical field and z represents the propagation distance. The parameters α, β_2, β_3, γ are the propagation attenuation, group velocity dispersion, dispersion slope, and nonlinear coefficient, respectively. Eq. (1) serves as the foundation for modeling optical fiber communication systems. Nevertheless, due to the coexistence of dispersion and nonlinearity, it

Table 2
Computational efficiency comparison of ML-based fiber channel modeling methods.

Method	Ref.	Time/Complexity vs. SSFM	Dist. (km)	Key metrics
BiLSTM	[18]	~20% of SSFM time	80	nMSE < 0.008 for amplitude
FDD-BiLSTM	[19]	1%–4% (vs. C-SSFM), 13%–40% (vs. NP-SSFM)	1040	Maintains accuracy over 41 ch.
BiLSTM-NNSpan	[20]	Not specified	1120	Best among four NNSpan variants
FDD-CoLSTM	[21]	Reduced complexity vs. BiLSTM[a]	2400	Single calculation for LSTM layers
Transformer	[22]	Not specified	1600	EVM < 4%
Simpl. Transformer	[23]	Reduced vs. Transformer[a]	100	Removed self-attention mechanism
Improved Transformer	[24]	Not specified	50	Tailored for Digital Twin systems
GAN/cGAN	[25]	2% of SSFM multiplications	1000	5.57×10^{9} vs. 2.85×10^{11} ops
Generic DNN	[26]	3.5% of SSFM time	240	nMSE < 0.02 across formats
FNO	[27,28]	Not specified	1200	nMSE < 0.0025 at 5 dBm
DeepONet	[29,30]	Not specified	800	MSE $< 1.4\times10^{-3}$
LDBP	[31]	77 taps vs. 1540+ for DBP	2000	2.1 dB SNR improvement
PINN	[14]	14.6% of SSFM multiplications	1000	Efficiency is architecture-dependent
PINO	[32]	1000× faster than SSFM	320	MSE $< 5\times10^{-4}$

[a] Specific numerical values not given in original papers.

is intractable using analytical methods. Although NLSE can be solved analytically using the Nonlinear Fourier under the presence of perfect distributed amplification [16], SSFM is still recognized as the most effective numerical method to obtain a solution, in the presence of nonlinearity. To explain SSFM, Eq. (1) can be simplified to

$$\frac{\partial A(z,t)}{\partial t} = (\hat{D} + \hat{N}(A(z,t)))[A(z,t)] \tag{2}$$

in which $\hat{D}$ and $\hat{N}$ represent the linear and nonlinear operators, respectively. These operators are presumed to function independently with a narrow interval h. By employing the symmetric split-step scheme, an approximate solution is derived as [17]

$$\begin{aligned} A(z+h,t) \approx & \exp\left(\frac{h}{2}\hat{D}\right) \\ & \times \exp\left(\int_{z}^{z+h} \hat{N}(z')\,dz'\right) \\ & \times \exp\left(\frac{h}{2}\hat{D}\right) A(z,t) \end{aligned} \tag{3}$$

In Eq. (3), $\exp(\frac{h}{2}\hat{D})$ is computed using the Fourier transform operation, thereby enabling a fast numerical evaluation of the linear operator.

2.2. Motivation of ML-based channel modeling

- The computational cost of the conventional method is relatively high compared with the ML-based models, due to the hierarchical function architecture and the continuous iterative processes [33]. These theoretical advantages have been empirically validated in recent studies. For instance, Wang et al. demonstrated that their BiLSTM-based model achieved approximately 80% reduction in computation time compared to SSFM for modeling 80 km fiber transmission [18]. Specifically, for a data size of 2^{15} bits, the BiLSTM model required only 20% of the computational time needed by SSFM, while maintaining comparable accuracy with nMSEs below 0.008 for temporal amplitude modeling. Similarly, Yang et al. reported that their computation time ratio of FDD-BiLSTM model to Constant-Step-Size SSFM (C-SSFM) ranged from 1% to 4% depending on the number of channels, representing a 96%–99% reduction in computational overhead [19]. Table 2 illustrates the computational efficiency comparison of ML-based modeling methods and SSFM.
- Although the optical fiber channel is consistently complex and highly nonlinear, the implementation of coherent technologies significantly exacerbates this situation as various elements are introduced, including channel impairments, modulation formats and multi-carrier systems [34–36].
- The conventional SSFM method only works effectively in static and singular scenarios, while an increasing body of research focuses on dynamic or adaptive systems [6,37]. Moreover, some of the latest systems, such as Digital Twins or large-scale wavelength-division multiplexed systems, where model updates must occur within milliseconds and the iterative SSFM cannot update fast enough [38,39].

3. Data-driven methods

In this section, we will present an overview of each data-driven method, focusing on network architectures and their corresponding results. For the network architectures subsection, we will provide a brief background on each method and explain how it works within optical fiber systems. In the results subsection, we will discuss the results of each method across various dimensions to demonstrate their modeling performance.

3.1. Bidirectional Long Short-Term Memory (BiLSTM)

3.1.1. Network architectures

Given the sequential nature of the propagation problem in optical fibers, recurrent neural networks (RNNs) [40] form the basis of one data-driven approach since they feed results back into the network, providing memory and time-dependence. However, RNNs evaluate each sample independently, enabling accurate predictions from recent information but often resulting in imprecise predictions from long-term information. Therefore, in fiber optic systems, RNNs are unable to effectively capture the time dependencies between adjacent samples. Long short-term memory (LSTM) [41], a RNN variant, was proposed to solve this problem since in them, the output of the previous step is fed back into the network as input for the current step, improving retainment of information and comprehending the entire process. LSTM has been widely applied in the field of optical fiber communication. In particular, the bidirectional LSTM (BiLSTM) model is capable of capturing information from both forward and backward directions of a sequence, thereby facilitating a more comprehensive representation of the underlying channel characteristics, especially considering inter-symbol-induced interference (ISI). The gating mechanism within LSTM cells facilitates effective retention and propagation of long-range dependencies, allowing for accurate modeling of cumulative dispersion and nonlinear impairments encountered over extended transmission distances.

Wang et al. constructed a BiLSTM to model the fiber channels for on–off keying (OOK) and pulse amplitude modulation 4 (PAM4) signals [18]. In contrast to the conventional SSFM method, BiLSTM is capable of learning the fiber channel transfer function while reducing the computation time by 80%. Deligiannidis et al. implemented an LSTM network into the compensation of nonlinearities in a coherent system [42]. Kotlyar et al. proposed a BiLSTM-based equalizer in a one polarization conventional nonlinear Fourier transform (NFT) based system [43]. Salmela et al. utilized an RNN with LSTM to learn the nonlinear propagation of ultrafast pulses in optical fibers [44]. Fang et al.

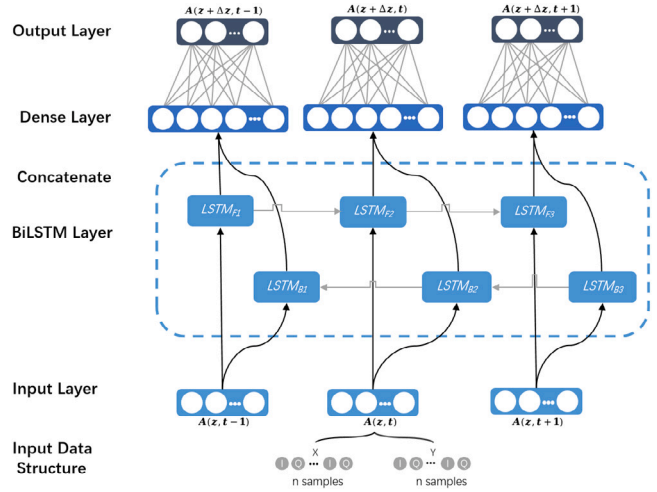

Fig. 2. Framework of BiLSTM-based fiber channel include input, BiLSTM model, hidden and output layers.

employed a BiLSTM with an attention mechanism to explore the soliton generation process in an ultra-fast laser, transitioning from the detuned steady state to the stable mode-locked state [45]. Ding et al. introduced a perturbation-based neural network (P-NN) algorithm, incorporating a BiLSTM to compensate for Kerr fiber nonlinearity, which is similar to reverse channel modeling [46]. Liu et al. evaluated the performance of BiLSTM-based methods against approaches based on conditional generative adversarial networks (cGANs) (discussed in Section 3.2.1), noting that the BiLSTM model captures more dynamic characteristics of the optical fiber communication channel than the cGAN [47]. Pu et al. proposed a model with various LSTM layers and a dimension expansion-based prior information feeding method, which enables fast prediction for the complex nonlinear dynamics of the mode-locked fiber laser [48]. Jiang et al. proposed four types of ML-based models called NNSpans for fiber channel modeling, and each type of NNSpan employed different algorithms, including deep neural networks (DNNs), convolutional neural networks (CNNs), BiLSTM, and cGAN. After comprehensive simulations using varying transmission distances, launched powers, and modulation formats, BiLSTM-NNSPan achieved the best modeling performance [20].

Our discussion here is focused on a representative work of Wang et al. [18]. The framework of the BiLSTM network in the model is depicted in Fig. 2. In the input layer, three temporally consecutive optical pulses $A(z,t-1)$, $A(z,t)$, and $A(z,t+1)$ are generated as the input vectors, where z is the propagation distance. Subsequently, these are transmitted to the BiLSTM network, which comprises both forward and backward layers. Following this, two state vectors generated at the same time step are integrated into the same vector within the hidden layer. Finally, after processing through one fully connected layer, three corresponding optical pulses $A(z+\Delta z,t-1)$, $A(z+\Delta z,t)$, $A(z+\Delta z,t+1)$ are consecutively output.

3.1.2. Results

In the work of Wang et al. [18], an intensity modulated direct detection (IM-DD) system employing 20 Gbps PAM4 and 10 Gbps OOK signals was simulated. The modeling performance was assessed across three dimensions: temporal, spectral, and phase characteristics; generalization against variations in transmission distance, nonlinearity, dispersion, and computational complexity. Initially, the performance of time-domain optical waveform modeling over a 50 km transmission distance with varying epochs was investigated. At 200 epochs, the generated waveforms closely matched those modeled by SSFM ([18] Fig. 6), demonstrating similar modeling capabilities in the time domain. Referring to phase waveforms for both OOK and PAM4, the phase waveforms at a transmission distance of 50 km were highly overlapped ([18] Fig. 7). Furthermore, with respect to the BiLSTM-based modeling capacity in frequency domains, the BiLSTM-based spectral components within the main lobe and central frequency closely resembled those in SSFM-based signals. However, significant differences were observed in the high-frequency regions around the side lobes ([18] Fig. 8), attributable to the discrete nature of the BiLSTM-based data in the final layer, which caused minor amplitude jitters in the time domain and resulted in additional high-frequency noise. To mitigate this issue, an optical low-pass filter was implemented.

Secondly, the study evaluated performance across various transmission distances ranging from 20 to 80 km. The normalized mean squared errors (nMSEs) for both OOK and PAM4 signals remained below 0.008 and 0.005 for temporal amplitude and below 0.011 and 0.019 for spectral power, respectively. These results suggested that the BiLSTM network could effectively generate signals for short-haul transmission. Fiber nonlinearity was controlled by varying the optical launch power from 0 to 15 dBm via an Erbium doped fiber amplifier (EDFA) after the transmitter. Furthermore, the nMSEs with the launch power from 0 to 15 dBm at varying transmission distances from 20 to 80 km for both OOK and PAM4 signals were determined. BiLSTM-generated waveforms maintained good fitting capacity even at 15 dBm launch power, with nMSEs well below 0.02. This result demonstrated BiLSTM's capability to accurately simulate fiber nonlinearity within the 0 to 15 dBm launch power range in short-haul transmissions. Moreover, the dispersion coefficient varied from 16 to 32 ps (km nm)$^{-1}$ and even under the extreme dispersion condition, the nMSEs remained below the acceptable threshold of 0.02 for satisfactory waveform fitting.

In accessing complexity, Wang et al. focused on the relative computation time rather than the absolute computation time, as the latter is highly dependent on the hardware performance and computing resources. For a data size of 2^{15} bits, the computation times of BiLSTM were found to be approximately 20% of those required by the SSFM, remaining stable across various transmission distances with only minor increases.

3.2. Generative Adversarial Networks (GAN)

3.2.1. Network architectures

The GAN was first introduced by Goodfellow et al. in 2014 [49]. It contains two distinct NNs, the *generator* and the *discriminator*, which are trained in an adversarial way. To elaborate, the generator produces fake outputs that are similar to real input data, and the discriminator evaluates this compared to the original data, behaving similarly to binary classification. Ideally, the training process converges to the situation where the generator produces perfect replicas of real data, and the discriminator cannot distinguish between real and generated data. Mirza et al. introduced conditional information to constrain the generated outputs of a GAN [50] producing *conditional* GAN (cGAN), which improves the training efficiency and reduces training failure probability. Fig. 3 illustrates the framework of cGAN.

Up to now, GAN-based models have been widely used in communication channel modeling. Since GANs are a versatile method, once trained, the GAN-based model can quickly generate synthetic 'real' data that accurately satisfies any given channel and system characteristics. Ye et al. utilized a cGAN to depict channel effects within an end-to-end wireless communication system [51,52]. Meanwhile, Y. Yang et al. introduced a GAN-based approach specifically for wireless channel modeling [53]. H. Yang et al. proposed a conditional GAN-based model to simulate the EDFA and fiber channel, including effects of CD, self-phase modulation (SPM), and attenuation [25]. Additionally, Sun et al. advanced a novel GAN-based model, a significant adaptation of ViterbiNet [54], designed to deduce channel transition probabilities from receiver observations [55]. Zou et al. developed a GAN-based underwater optical channel generator called the underwater optical

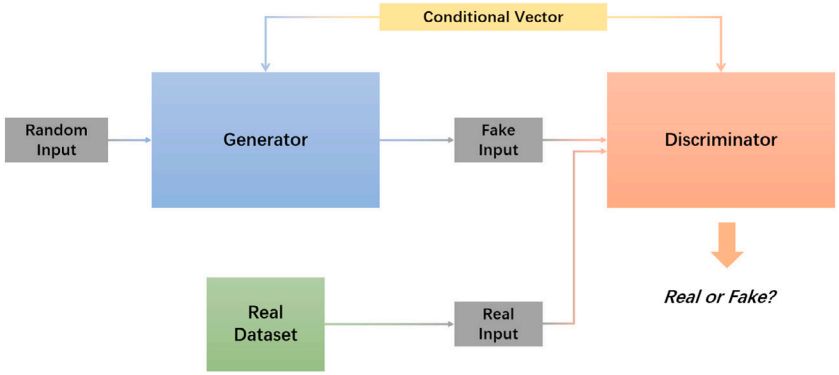

Fig. 3. Structure of conditional GAN.

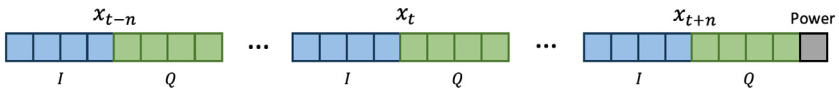

Fig. 4. Structure of conditional vector in the research of Yang et al..

Table 3
Transmission distance, launched power and nMSE values.

Transmission distance (km)	Launched power (dBm)	nMSE × 10^{-4}
50	10	2.8
100	10	7.2
200	−2	9.0
200	−1	91
200	0	9.3
200	1	9.5
200	2	9.8
200	3	13.8
200	4	18
200	4.5	21
1000	0	92.5

Table 4
CGAN-based model results for modulation formats and distributions within (50 km, 10 dBm) transmission.
Source: Adapted from [25].

Modulation format	MSE ×10^{-4}
16QAM	72.0
QPSK	15.6
32APSK	10.7
64QAM	10.6
16QAM with Gaussian noise	7.2

channel impulse response generator (UCIRG) framework, enhancing both generalizability and flexibility [56]. Chen et al. conducted tests and predictions to evaluate the modeling efficacy of GAN-based models [57]. Wei et al. proposed a cGAN-based channel simulator, which accurately models both deterministic and stochastic noise components in optical wireless communication systems [58]. Furthermore, Ye et al. investigated the impact of input optical SNR (OSNR) on the performance of a cGAN-based channel modeling technique [59]. Our discussion centers on the research conducted by H. Yang et al. [25], which serves to elucidate the application of GANs in optical fiber channel modeling. In an optical fiber system, the signal data transmitted before and after the fiber was collected as the condition vector and real data, respectively. The condition vector contained several consecutive samples, which assisted the GAN in learning the dependency within these and further learning the ISI caused by CD. The structure of the condition vector is shown in Fig. 4, where $x = [x_{t-n}, \ldots, x_t, \ldots, x_{t+n}]$ is the condition vector, x_t is the current transmitted samples, the left-hand side represents the samples transmitted before the current one, while the right-hand side represents the samples transmitted afterwards. The number of surrounding samples n is proportional to the delay caused by CD and varies with the transmission length, and can be obtained based on the calculation of symbols affected by ISI and ensuring alignment with the transmission rate. Additionally, they incorporate the value of optical launch power at the end of the condition vector to enhance flexibility. This approach requires that the training dataset includes not only input–output data but also the launch power.

3.2.2. Results

H. Yang et al. tested the cGAN-based model in a single-channel long-haul transmission with CD, SPM, attenuation, and ASE. Notably, they changed the dispersion and nonlinear effect through transmission distance and optical launched power instead of the dispersion and nonlinearity coefficients. They demonstrated the accurate modeling capacity of the cGAN-based model in long-haul transmission (500, 1000 km) with low launched power (0 dBm) and short-haul transmission (50, 100 km) and high launched power (10 dBm). Additionally, they also illustrated the modeling ability in 200 km transmission with varying launched power (−2, 0, 2, 4 dBm). Furthermore, they demonstrated the generalization ability of the model to slight variations in launch power by adjusting the launch power to the set (−1, 1, 3, 4.5 dBm) that were not included in the training stage. The results of the above investigation are shown in Table 3[25].

Concurrently, H. Yang et al. predicted the generalization capacity of the modulation formats and distributions, including QPSK, 32APSK, 64QAM, and 16QAM with Gaussian noise. The results are shown in Table 4[25]

H. Yang et al. analyzed the complexity by dividing it into the number of multiplications within the modeling and computational time. When the transmission distance was 1000 km, and the step length was 0.01 km, the number of multiplications of the SSFM-based model was 2.85×10^{11}, while that of the GAN-based model was 5.57×10^9, approximately 2% that of the SSFM-based model. Moreover, they recorded the computing time when the data size was risen from 1000 to 7×10^7. And when the data size reached 7×10^7, the GAN-based model execution time was under 0.1 s, whereas that of the SSFM-based model was some 1204 s on a CPU and 71 s on a GPU. The above investigation demonstrated a considerable improvement in the modeling efficiency of cGAN.

3.3. Transformer

3.3.1. Network architectures

The Transformer, an ML algorithm based on self-attention, was initially introduced by Google Brain in 2017 [60]. A fundamental tenet of the Transformer is its utilization of multi-head attention mechanisms, allowing it to comprehensively capture dependencies within input sequences and process these sequences in parallel. In particular, an attention function can be described as a mechanism that maps a query and a set of key–value pairs to an output, with the query, keys, values, and output in vectors, where the output vector is computed by taking a weighted sum of the features. These are two mechanisms that are beneficial in modeling optical channels characterized by intricate nonlinear dynamics and dispersion effects. Moreover, their adaptive weighting capability provides robustness against varying channel conditions, thus enhancing the accuracy and stability of the optical channel modeling process.

Zhang et al. first introduced the Transformer into optical Orthogonal Frequency-Division Multiplexing (OFDM) systems to achieve fiber channel modeling [22]. Zang et al. introduced a model that is composed of a Transformer encoder and a deep, fully connected structure [23]. It is noted that both of these only utilized the encoder part of the Transformer, which enabled parallel data output by removal of the decoder to improve efficiency. Zhu et al. introduced the transformer-based high-fidelity modeling method for radio over fiber links, which has found considerable application for Digital Twins in microwave photonics [24].

In the study of Zhang et al. a simplified Transformer encoder was developed for optical OFDM channel modeling [22]. Fig. 5(a) depicts their model. To reduce computational complexity, the self-attention mechanism and Decoder were eliminated, while the multi-head mechanism, FFN, residual connections, and layer normalization were retained as essential components. The number of heads in the multi-head mechanism was set at 8, enabling effective extraction of diverse nonlinear channel features. Each Transformer layer included a two-layer FFN with hidden dimensions and ReLU activations, followed by residual connection and normalization. Apart from these adjustments, traditional Transformer models include Embedding layers at both the input and output of the Encoder, originally designed for word vectors in Neural Machine Translation. However, these layers proved to be unstable when applied to waveform inputs. As a result, the Embedding layers were substituted with linear layers containing an increased number of neurons, a change aimed at enhancing feature extraction and fitting the model more effectively to the data. The model was trained using the ADAM optimizer, and with a learning rate of 1e−4 at the beginning. They cancel the self-attention mechanism according to the work of [61], self-attention was substituted by two cascaded linear layers and two normalization layers in the Transformer architecture and the multi-head mechanism was integrated into the complete MLP (Multilayer Perceptron) model. Extensive experiments demonstrate that this modified approach achieved better performance than the original self-attention and some of its variants. Moreover, Domhan [62] suggested that the multi-head mechanism, residual connection, and FFN were cores to the Transformer, and self-attention was unnecessary for it. Self-attention introduced unwarranted computational complexity related to the number of input feature vectors and could not bring any benefit.

In the study of Zang et al. a model combining a Transformer encoder with a multi-head mechanism and a fully connected structure was developed. Fig. 5(b) depicts their model. The number of heads in the multi-head attention mechanism is set at 17 to allow the extraction of diverse features related to signal distortion and transmission effects. Initially, the input data, representing pre-transmission signals, were duplicated. The first part was processed through the multi-head attention mechanism, while the second bypassed it and entered the addition and normalization layer directly. These outputs were merged using a residual structure to mitigate gradient vanishing during training, as proposed in ResNet [63]. The normalized data were then duplicated again, with one copy entering a feed-forward layer configured with hidden dimensions of 512 → 1024 → 512 and ReLU activations. This output was merged with the bypassed data in a second addition and normalization layer. Finally, the output from the Transformer encoder was passed into a deep, fully connected structure consisting of three layers (512 → 1024 → 1024 → 512) for further feature refinement and signal regression. The model was trained using the ADAM optimizer and optimized via nMSE, enabling robust performance across various bit rates, modulation formats, and distances up to 100 km.

The study of Zhu et al. diverged notably from the preceding two analyses by adapting the transformer in structure. They modified the transformer from three perspectives: data formats, model structure, and loss functions. The input data format was adjusted to align with the sampled signal because the transformer was originally employed in machine translation instead of channel modeling tasks. Additionally, a conditional vector was implemented as encoder input to learn the influence of CD in the signal. Similar to the conditional vector in [25], this vector encompassed adjacent symbolic data along with conditional variables illustrating factors such as optical fiber transmission distance, launched power, and carrier frequency. Referring to structure, the output layer was reconfigured with a linear layer, while the conventional *Softmax* activation function was substituted with a tanh function, and the Dropout was eliminated to avoid the undesirable induction of instabilities in channel modeling. The model was trained using the ADAM optimizer with a learning rate of 1e−4 and the loss function was derived from the MSE between the ground truth and the predicted signal. Experimental results demonstrated accuracy and generalization under varying RoF conditions, including different fiber lengths, launch powers, carrier frequencies, and SNRs, with all normalized MSEs maintained below 0.02. The framework is depicted in Fig. 6(a).

3.3.2. Results

Our discussion primarily centers on the research of Zhang et al., which is presented in four dimensions, including the capacity for fitting nonlinearity, accuracy, and robustness in the context of 16QAM-OFDM signals with varying numbers of subcarriers and Peak-to-Average Power Ratio (PAPR) values, transmission rates, and complexity. For comparison, Zhang et al. also implemented Bi-LSTM and trained it using the same training dataset, training methodology, and loss function.

To assess nonlinear fitting capacity, they chose 13 sub-channels from among the 256 channels within the 16QAM-OFDM signal. They measured the Four-Wave Mixing (FWM) noise power generated in each of these sub-channels during transmissions at distances of 400 km, 800 km, 1200 km, and 1600 km. The FWM noise power is measured by leaving a specific sub-channel empty before applying the IFFT at the transmitter first. Afterwards, the signal power can be assessed on the corresponding sub-channel after applying the FFT at the receiver. The FWM noise power modeled by the Transformer exhibited remarkable consistency with the FWM power modeled by SSFM. The average relative errors modeled between Transformer and SSFM for the 13 selected sub-channels were found to be 0.38%, 0.64%, 0.97%, and 1.24% at distances of 400 km, 800 km, 1200 km, and 1600 km, respectively.

Secondly, regarding accuracy and robustness, they adopted the Error Vector Matrix (EVM) as a criterion. EVM is defined as the relative error between the signal simulated by the model and the signal simulated by SSFM. The EVMs consistently maintained a low level, remaining below 4% even at the maximum transmission distance. For signals with 2048 sub-carriers at 1600 km, the EVM of the Transformer-based model was 2.06%, while the EVM of the BiLSTM-based model was 3.97%, underscoring the outperforming modeling accuracy of the Transformer-based model, even when trained without ASE noise. For signals with 2048 sub-carriers at 2048 km, the EVMs for the Transformer and Bi-LSTM were 3.97% and 5.33%, respectively. It is evident that all EVMs increase by approximately 2% compared to EVMs without

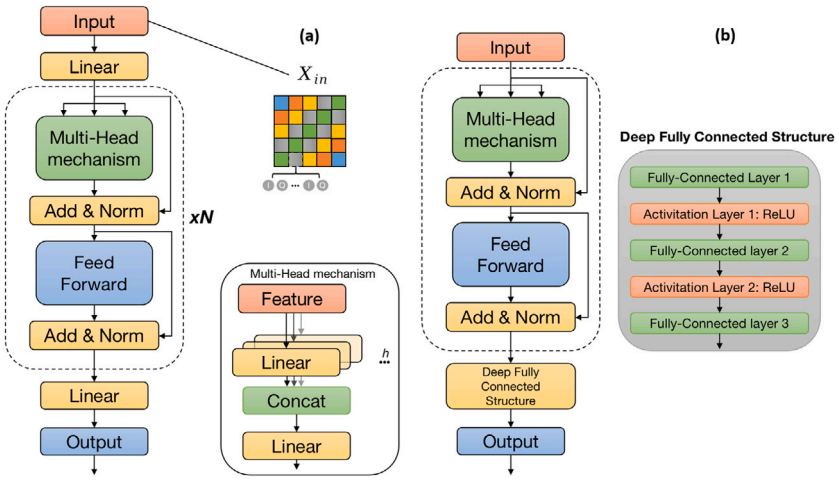

Fig. 5. Structure of simplified Transformer in the research of (a) Zhang et al. [22] and (b) Zang et al. [23].

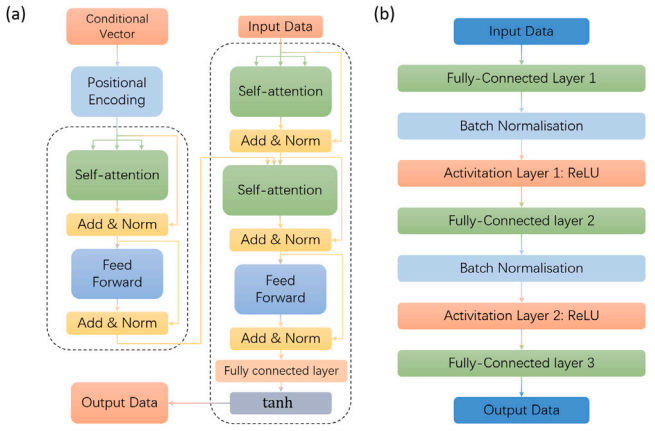

Fig. 6. Framework of (a) Improved Transformer proposed in Zhu et al. (2023) [24] and (b) DNN proposed in Jiang et al. (2022) [26].

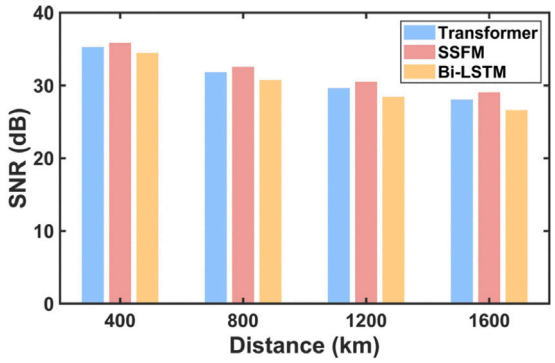

Fig. 7. SNRs of optical OFDM signals transmitted over Transformer-based (blue), SSFM (red) and BiLSTM-based (yellow) channel models for distances of 400 km, 800 km, 1200 km, and 1600 km, with ASE noise [22].

ASE noise. Nevertheless, the EVMs with ASE noise remained at a low level. EVMs exhibited a linear increase with transmission distance due to accumulated errors from iterations. For both the Transformer and BiLSTM-based models, the number of sub-carriers and PAPR only had a minor effect on the performance. Additionally, measurements of the SNR following sub-carrier de-mapping and Digital Backpropagation (DBP) were conducted. Fig. 7 illustrates the SNR variation between signals transmitted at distances of 400 km, 800 km, 1200 km, and 1600 km, as observed between the Transformer, SSFM and BiLSTM. And the SNRs difference between Transformer and SSFM ranged from 0.6 dB to 1.02 dB, which further underscored the accuracy of the Transformer-based channel modeling.

Thirdly, they conducted an investigation into the modeling using different transmission rates, specifically at 30 GBd, 160 GBd, and 500 GBd. When transmitting 1600 km, EVMs increased from 3.93% to 8.84% while the transmission rates varied from 30 GBd to 160 GBd, which meant that the modeling performance declined significantly due to the higher CD effect caused by a higher transmission rate. Notably, one-step lumped CD compensation did not completely eliminate all linear effects. Therefore, these residual linear effects manifested as temporal correlations among adjacent samples, which posed challenges for the model during the learning process. In contrast to the previous results, from 160 GBd to 500 GBd, the EVMs decreased from 8.84% to 5.54%. This was attributed to the phase mismatch induced by the CD effect, which diminished the efficiency of third-order nonlinear interactions [64]. Consequently, the damage inflicted by FWM was mitigated.

3.4. Generic Deep Neural Network (DNN)

3.4.1. Network architectures

The research of Jiang et al. delved into the efficacy of the DNN-based approach in modeling channels affected by attenuation, CD, ASE noise, SPM, and XPM [26]. Fig. 6(b) illustrates the DNN framework in their model.

3.4.2. Results

The investigation of Jiang et al. started with the prediction of model performance in nMSE over a 240 km transmission featuring a launched power of 0 dBm with 16QAM signals. The nMSEs in both the time and frequency domains were deemed acceptable, which are 0.0032 in the time domain and 0.0004 in the frequency domain, respectively. Moreover, the generalizability of the DNN across different constellations and waveform schemes was examined using parameters from the training dataset. However, the performance was influenced when the wavelength schemes were farther from the wavelength scheme (1549.32, 1550.12 nm) in the training dataset, with all tested nMSEs being below 0.02. In terms of complexity, the DNN-based model demonstrated a relative efficiency of 3.5% in comparison with the SSFM. The simulation results of the DNN-based model with distinct constellations and wavelength schemes are shown in Table 5.

Table 5
DNN-based model performance with distinct constellations under transmission 1 (80 km, 4 dBm) and transmission 2 (160 km, 4 dBm)[26].

Modulation format	Wavelength schemes (nm)	nMSE in time-domain under transmission 1	nMSE in frequency-domain under transmission 1	nMSE in time-domain under transmission 2	nMSE in frequency-domain under transmission 2
QPSK	1549.32, 1550.12	0.0017	0.00067	0.0017	0.00070
8PSK	1549.32, 1550.12	0.0017	0.00066	0.0017	0.00069
8QAM	1549.32, 1550.12	0.0020	0.00094	0.0023	0.0013
16QAM	1542.94, 1543.73	0.0034	0.00042	0.0026	0.00038
16QAM	1546.12, 1546.92	0.0019	0.00041	0.0023	0.00037
16QAM	1547.72, 1548.51	0.0016	0.00040	0.0018	0.00036
16QAM	1550.92, 1551.72	0.0018	0.00040	0.0015	0.00036
16QAM	1552.52, 1553.33	0.0021	0.00041	0.0017	0.00036
16QAM	1555.75, 1556.55	0.0030	0.00042	0.0038	0.00038

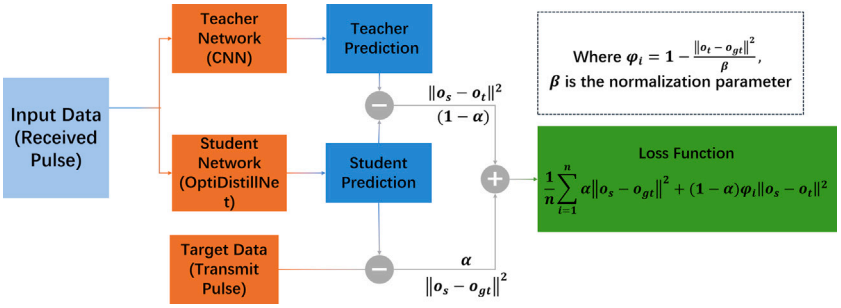

Fig. 8. Structure of OptiDistillNet.

3.5. OptiDistillNet (knowledge-distillation CNNs)

3.5.1. Network architectures

Sui et al. proposed a lightweight convolutional neural network (CNN) with optimized convolutional blocks and parameters to model multi-pulse propagation in fiber optics [65]. Their investigation included forward mapping and inverse mapping under the coexisting influence of GVD and SPM. The maximum absolute error is 0.026 and 0.01 for forward mapping and inverse mapping, respectively. Gautam et al. [66] proposed OptiDistillNet, which utilized knowledge distillation (KD) (first proposed by Hinton et al. [67] and good for scenarios with limited resources) to form a CNN. In the network, a complicated CNN was trained to solve the NLSE as a teacher model while a reduced CNN as a student learned from this. The network structure of OptiDistillNet is depicted in Fig. 8. In this network, the teacher model was trained on a substantial dataset and achieved considerable modeling precision, which was assumed to be as accurate as the SSFM-based model. Due to this assumption, an extra term was considered in the loss function of the student model, which represented the difference between the prediction of the student and teacher model. Furthermore, to prevent inaccurate modeling by teacher networks, the empirical error between the teacher prediction and the target data was considered to reduce the weight of the former. The loss function of OptiDistillNet was expressed as

$$L_{\text{OptiDistillNet}} = \frac{1}{n}\sum_{i=1}^{n} \alpha \left\| o_s - o_{gt} \right\|^2 + (1-\alpha)\left\| o_s - o_t \right\|^2 \tag{4}$$

where α represented a scaling factor employed to equilibrate the loss terms, while o_s, o_{gt} and o_t corresponded to the outputs of the student, the ground truth, and the teacher, respectively. ψ_i denoted the empirical error obtained by assessing the variance between the maximal and minimal absolute squared discrepancies between the teacher predictions and the target data.

3.5.2. Results

The investigation was mainly about the design of the structure of networks [66]. First, the choice of teacher network through a comparative test within FCNN, LSTM Network, and CNN was explored. After comparisons in terms of MSE and the number of trainable parameters, CNN was capable of learning the NLSE with the lowest MSEs. Subsequently, the optimal structure of the teacher and student model, including the number of output channels for each convolutional layer and the number of convolutional layers was explored. Further, the loss function of OptiDistillNet was enhanced and minor enhancements in MSEs were achieved. According to the results, the number of trainable parameters of OptiDistillNet was 91.2% smaller than those of the teacher network, and the required training data were decreased approximately 57% from 7×10^5 to 3×10^5.

3.6. Fourier Neural Operator (FNO)

3.6.1. Network architectures

This method is a variety of neural operator learning, which was first proposed by Li et al. [28], and has been widely applied in solving physical problems [68–70]. The FNO aims to learn the mappings between infinite-dimensional spaces of functions using the function-to-function architecture. This approach addresses the mesh-dependent limitations of traditional finite-dimensional approximation methods, by providing a single set of network parameters that can be applied across diverse discretizations [71,72]. The FNO was introduced to optical fiber channel modeling by He et al. [27,28], which enabled high robustness and precision modeling in long-haul transmission. In their simulation, the continuous signal must be discretized into a set of finite, uniform time intervals. The transmitted signal is typically obtained by down-sampling from a high-dimensional vector space. $x(t)$ and $y(t)$ were the transmitted signal and the received signal, which were sampled from two unrecognized function spaces X and Y, respectively. Therefore, the nonlinear map S_θ can be seen as the operator $S : X \to Y$. Subsequently, if S_θ, the approximate solution of S can be learned with a finite collection of input–output signal pairs, and thus the received signal could be approximately expressed by $\approx S_\theta(x)$. According to universal approximation theorem [71], any nonlinear continuous functional, as well as both linear and nonlinear operators, can be approximated by the NN. The FNO derives the kernel of an integral Hilbert–Schmidt operator through paired observations, which enables the formation of a new solution map for data-driven PDEs, including those related to

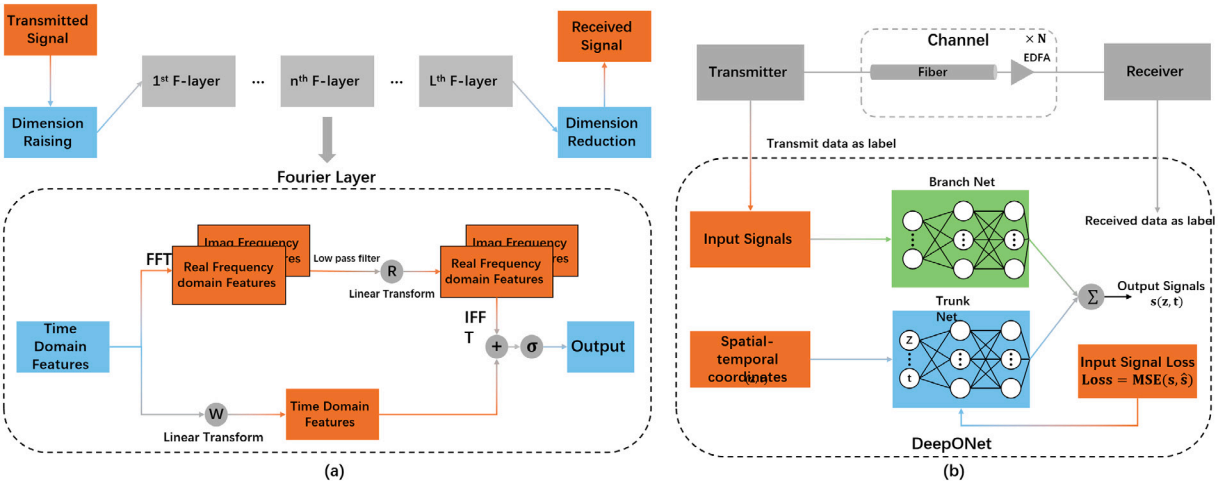

Fig. 9. (a) Framework of FNO-based Fiber Channel. (b) Simply framework of DeepONet-based fiber model.

the transmission of signals in optical fibers. The learning mechanism of FNO resolves empirical risk minimization issues using a given set of signals, and the loss function of FNO is given by

$$L_{FNO} = \frac{1}{n}\sum_{i=1}^{n} \left\| y_i(t) - S_\theta(x_i(t)) \right\|_2 \tag{5}$$

where S_θ contains the number of linear and nonlinear operators, which can be expressed by

$$S_\theta := \left(\mathcal{Q}\circ\mathcal{L}^{(L)}\circ\mathcal{L}^{(L-1)}\circ\cdots\circ\mathcal{L}^{(1)}\circ\mathcal{P}\right)(x(t)), \tag{6}$$

wherein $\circ$ and L represent the operator composition and number of layers, respectively. $\mathcal{P} \in \mathbb{R}^{1\times h}$ and $\mathcal{Q} \in \mathbb{R}^{h\times 1}$ are the lifting and projection operators, which raise and reduce the dimension of the signal. h is the size of the hidden layer. $\mathcal{L}$ is the Fourier layer, which is composed of the linear combination of linear and nonlinear operators. It can be represented as

$$z_t^{(\ell+1)} = \mathcal{L}^{(l)}\left(z_t^{(\ell)}\right) = \sigma\left(W^{(\ell)}z_t^{(\ell)} + \mathcal{K}^{(\ell)}\left(z_t^{(\ell)}\right)\right), \tag{7}$$

where σ denotes a point-wise nonlinear activation function, while W_ℓ denotes the one-dimensional convolution in the time domain for the ℓ th layer. K_ℓ is identified as the integral kernel operator, which can be expressed as

$$\mathcal{K}^{(\ell)}(z_t^{(\ell)}) = \mathrm{IFFT}\left(\mathcal{R}^{(\ell)}\cdot(\mathrm{FFT}(z_t^{(\ell)}))_*\right), \tag{8}$$

which FFT and IFFT represent the fast Fourier transform and the inverse fast Fourier transform, respectively. $(\cdot)_*$ indicates a low-pass filter, $\mathcal{R}$ represents matrices of weights in the frequency domain. The hidden representation $z_t \in \mathbb{R}_{d_t\times h}$ can be regarded as the basis function for the initially transmitted signal, more features can be integrated by both time- and frequency-domain information of the basis function. Thus a more comprehensive set of features can be obtained compared to relying solely on time-domain information. At the final stage of the network, the dimension reduction block maps the final hidden representation to the received signal. The framework of FNO in fiber is shown in Fig. 9(a)

3.6.2. Results

He et al. investigated the modeling performance at 28 GBd data rate, a variety of modulation formats, including 4QAM, 16QAM, and 64QAM, and varying launch power within 1200 km long-haul transmission standard single-mode fiber with and without ASE noise. Firstly, with respect to channel modeling without considering ASE, the FNO-based model achieved effective performance in three different launch powers (−1, 2, 5 dBm). Despite a high launch power set at 5 dBm, all nMSEs remained below 0.0025 (transmission distance within 1200 km). This was attributable to the capacity of the FNO-based model to simultaneously learn features in both the time and frequency domains, while high nonlinearity significantly changed the features of signals in the frequency domain. Further, the investigation was expanded to three distinct training datasets, which were (−3, 0, 3 dBm) launch powers and 4QAM, 16QAM, 64 QAM modulation formats; −5 to 5 dBm launch powers and 16 QAM modulation format; −5 to 5 dBm launch powers and 4QAM, 16QAM, 64QAM modulation formats. According to the observations of He et al., implementing different modulation formats does not significantly affect the nMSEs, the nMSEs of all tests are well below 0.02, which suggested modulation format generalization capability, implying a wider range of launch powers only led to minor improvements in the performance within high nonlinear regions, while the trends in nMSEs remained consistent.

3.7. Deep Operator Network (DeepONet)

3.7.1. Network architectures

The first neural operators, proposed by Lu et al. [73], were DeepONets. They could learn mappings between infinite-dimensional function spaces based on the universal approximation theorem [71], addressing challenges in modeling complex systems governed by differential equations. A DeepONet is constructed with two sub-networks, which are branch and trunk nets. Depending on the number of branch nets, DeepONet can be categorized into unstacked and stacked networks. The former only has one branch net, whilst the latter contains a number of branch nets, enabling multiple mappings simultaneously. Generally, DeepONet divides input data into initial functions and spatial–temporal coordinates, whereas initial functions are the input of branch nets, and spatial–temporal coordinates are the input of trunk nets. Subsequently, DeepONet processes data distinctly with subnetworks and combines the outputs in the final layer. Compared with NNs, which learn nonlinear mappings from input–output pairs, DeepONet is flexible regarding the number of inputs. When the DeepONet is trained, it is capable of predicting high-resolution outputs from sparse discrete values of input functions within the spatial–temporal range in the training dataset. Notably, if the spatial–temporal coordinates are beyond the range or the training dataset only concludes part of the domains, the predicting performance will be influenced considerably. Zhang et al. proposed DeepONet-based models for channel modeling, which demonstrated high flexibility in both single channel and WDM systems extending beyond the C band [29,30]. In their simulation, initially, they designed an unstacked DeepONet model for learning the Manakov-polarization mode dispersion equation in Polarisation-Division Multiplexing (PDM) single-channel fiber modeling. The simple DeepONet-based channel framework is demonstrated in Fig. 9(b). The

top portion of the figure illustrates the dataset collection setup, while the bottom portion depicts the architecture of the DeepONet. The waveforms transmitted before and after propagation through optical fibers were collected as features and labels, respectively, for training the DeepONet.

Additionally, Zhang et al. proposed an inverse Fourier transform (IFT) DeepONet with a modified training method and model structure according to the characteristics of IFT, which minimized data requirements by implementing spectra and waveforms into the labels. Further, they proposed two improved models, based on stacked DeepONet and loading condition-aware DeepONet (LCA-DeepONet), for fully loaded condition and random loaded condition wideband WDM system modeling, respectively.

3.7.2. Results

With respect to the single-channel system, Zhang et al. investigated a system in the transmission of PMD-16QAM signals with a launch power set (−3, 0, 3 dBm) within long-haul distance, and the 10th DeepONet (720 km to 800 km) was measured as evidence. Their results demonstrated that the MSEs were $< 1.4\times10^{-3}$ within all measurements. Additionally, they expanded the measurements to the transmission after 800 km with 0 dBm launch power, the SNR calculated SNRs of signal modulated by SSFM and DeepONet are 26.1 and 25.9, respectively. According to these observations, including similar waveform components and constellation distributions compared between DeepONet and SSFM-based modeling and acceptable low MSEs, the effectiveness of DeepONet-based modeling in single-channel systems could be fully observed.

Subsequently, the generalization capacity regarding launch powers, modulation formats, and transmission distances was demonstrated. The datasets of the training and test are shown in Table 6. The MSEs for all datasets were $< 1.2\times10^{-3}$, which predicted the adequate generalization ability of DeepONet.

Referring to the fully loaded condition, the performance of stacked DeepONet was investigated. The nMSEs increased slightly with the transmission distance, however, they remained below 1.5×10^{-3} even in the later span. In randomly loaded conditions, they adjusted the loading ratio between 60% and 90%. The nMSEs rose marginally as the loading ratio decreased and the transmission increased. After 800 km transmission, the error distribution remained uniform across varying loading ratios, with the mean error being less than 0.001. These two observations predicted the modeling performance in a wideband WDM system.

4. Principle-driven methods

The previous section reviewed the performance of data-driven methods across multiple dimensions. This section introduces and analyzes principle-driven methods, highlighting the distinct reliance on physical laws and fundamentally different modeling theories in these approaches in contrast to data-driven methods.

4.1. Learned digital back-propagation

4.1.1. Network architectures

DBP is designed to invert the optical channel effects of forward propagation at the receiver [74]. In DBP, all signs of parameters are reversed compared to those used in forward propagation which is specified in Eq. (1). Häger et al. proposed LDBP for Nonlinear Interference (NLI) mitigation [75–77]. The core idea of LDBP is to transfer SSFM into the multi-layer NN form due to their similar functional form. By parameterizing iterative linear and nonlinear operations, a parameterized multi-layer network can be obtained. The weights of this network can then be optimized using the training method for NNs. LDBP has been shown to reduce complexity significantly, a finding detailed in [31] and later experimentally validated for PDM and WDM systems in [78]. Furthermore, in [79], triplets derived from an NLSE perturbation analysis were integrated as input features in NN-based equalizers that reduce complexity by constructing tensor weights and evaluating their contributions.

Table 6
The training dataset and test dataset in the investigation of generalization of DeepONet-based model.

	Training dataset	Test dataset
Launch power (dBm)	0	±1, ±3
Modulation format	16QAM	16QAM, 32QAM PS-16QAM
Transmission distance (km)	740, 760, 780, 800	730, 740, 750, 760 770, 780, 790, 800

In contrast to data-driven methods, LDBP has three main strengths. Initially, a notable difficulty associated with conventional NN models is the absence of clear directives for decisions on network architecture. Comparatively, the LDBP provides us with clear hyperparameter guidance, such as the number of layers/steps and activation function, while in LDBP, the activation function corresponds to the Kerr effect. The second strength is the effective initialization of the parameters. Generally, it is difficult to find proper parameter initialization strategies that enhance the training process and the convergence to optical solutions. On the contrary, it has been proven that initial parameters close to SSFM are suitable initialization parameters for LDBP. Third, unlike data-driven models, LDBP is based on the linear step of FIR filters, which enables the interpretation of learned frequency responses to yield novel theoretical insights into the solutions achieved. Notably, although this method has been widely applied for channel equalization, it has yet to be specifically tested for channel modeling, despite the fact that they are essentially the same problem in the absence of noise [12].

4.1.2. Results

First, the method has been tested in single channel transmission of a 10.7 GBd signal at the distance of 25×80 km fiber [34]. LDBP achieved a peak SNR of 23.5 dB, utilizing a total impulse response of 77 taps. Relative to 1-StPS DBP, the peak SNR experienced a decrease of around 2 dB. Furthermore, for the filter design for 1-StPS DBP in [74], achieving performance comparable to 2 dB required 77 taps/step, which is more than 20 times the number of taps needed for LDBP. Additionally, as the number of taps increases, the performance improves rapidly and slightly exceeds that of frequency-domain DBP. Subsequently, performance at higher baud rates was predicted with 32 GBd in single-channel 10×100 km transmission. Compared to DBP in the same setting, notable enhancements in performance were achieved. Specifically, an increase of 2.1 dB was observed in the peak SNR for two steps per span (StPS), and 0.6 dB for 4 StPS. Finally, the performance of WDM transmission was investigated. The performance enhancements achievable with static single-channel equalizers are constrained by nonlinear interference from adjacent channels [80]. Additionally, the precision demands for resolving the NLSE become substantially less stringent, in contrast to single-channel transmission, due to the lower effective SNRs attainable. Consequently, the complexity of LDBP was diminished by removing superfluous filter taps. Their findings demonstrated that the efficacy of the 2-StPS LDBP model outperforms DBP with 500 StPS. In particular, an additional reduction in the model by 36% resulted in a minimal reduction in peak SNRs of approximately 0.2 dB.

4.2. Physics informed neural network

4.2.1. Network architectures

PINNs were first proposed by Raissi et al. [81], which are trained to integrate learning with consideration of the principles defined by general nonlinear partial differential equations representing physical laws.

In optical fiber communication, by embedding the initial conditions of pulses or signals prior to transmission alongside the NLSE within its loss functions, a PINN is capable of gradually learning the parameters of a given PDE plus a representation of the training pulse. However, PINNs still face several critical challenges in practical applications. One prominent issue is spectral bias, wherein the network exhibits a preference for learning low-frequency components and often fails to capture sharp gradients or high-frequency dynamics. Another concern is causality violation, where residual minimization at later time steps may proceed before earlier dynamics are accurately learned, leading to erroneous temporal evolution. Additionally, the optimization process is often hindered by gradient imbalance across different loss components such as initial conditions, boundary constraints and PDE residuals, which compromises convergence and accuracy. PINNs also exhibit a strong dependency on variable scaling and initialization, necessitating careful pre-processing strategies such as non-dimensionalization and the incorporation of architectural enhancements like Fourier feature embeddings and curriculum learning. These limitations highlight that the effective deployment of PINNs requires carefully designed training pipelines and significant empirical tuning. Recent studies, including the comprehensive work by Wanget al. [82], have proposed a suite of best practices such as adaptive loss weighting, causal training and residual-based sampling to address these pathologies and improve model robustness and fidelity.

Many researchers have extensively explored the capabilities of PINNs in modeling tasks and optimized them to fit practical environments. Zang et al. modified a PINN and proposed a PINN-based fiber transmission model for modeling pulse evolution, signal transmission, and fiber birefringence [83,84]. Jiang et al. proposed a PINN-based system to model complex nonlinear dynamics and various physical effects such as CD, SPM, and higher-order nonlinearities in optical fiber [10,14]. Jiang et al. proposed a PINN-based approach for system parameter estimation (SPE) and predicted its performance under three scenarios: single Gaussian pulse propagation, ultrashort pulse propagation in high nonlinear fiber, and optical signal transmission in single standard fiber [85]. Wang et al. investigated the feasibility of PINNs within the field of optical fiber communication first, then proposed and discussed five potential directions of PINN in the field, including forward and backward optical waveform propagation models, fiber parameters identification method, wideband spectrum evolution model and electric field distribution model [13]. Jiang et al. introduced an advanced method based on PINNs to simulate ultrafast nonlinear dynamics in optical fiber through the solution of the Generalized NLSE (GNLSE), which considered two general situations of high-order soliton compression and supercontinuum generation [86]. Uduagbomen et al. found the loss landscape of the PINN becomes increasingly complex as the coefficients of the NLSE increase. They proposed a modified fail-proof PINN-based method and achieved higher accuracy [87–89]. Chen et al. proposed the generative pre-trained PINN (GPT-PINN), which mitigates the cost of training and over-parameterization by utilizing the whole pre-trained networks as the activation functions of another network [90]. Zang et al. demonstrated that the GPT-PINN can efficiently process input signals with varying bit rates, ranging from 2 Gbps to 100 Gbps, without requiring complete re-training [91,92]. In addition, there are numerous studies using PINNs to analyze the high-order nonlinear dynamic characteristics of optical pulses [93–98], further demonstrating the potential of PINNs in modeling complex phenomena in the optical domain.

Most PDEs can be delineated in a universal format, incorporating the principal equation alongside initial and boundary conditions as

$$\frac{\partial A}{\partial t} + N_x(A) = 0, \quad x \in \Omega, \quad t \in [0, T] \tag{9}$$

$$A(x, 0) = A_0(x, 0), \quad x \in \Omega, \quad t = 0 \tag{10}$$

$$A(x, t) = A_b(x, t), \quad x \in \partial\Omega, \quad t \in [0, T] \tag{11}$$

in which $A(x,t)$ signifies the PDE solution, $\frac{\partial A}{\partial t}$ denotes the temporal derivative, N_x is a generalized nonlinear differential operator, encompassing spatial derivatives and additional nonlinear elements such as $|A|^2$. The computational domain and its boundary are Ω and $\partial\Omega$, respectively. Moreover, $A_0(x,0)$ and $A_b(x,t)$ signify the solutions at $t = 0$ and the boundary.

To address the PDE in Eq. (9) along with its initial conditions Eq. (10) and boundary conditions Eq. (11), as shown in Fig. 10a PINN consisting of a DNN is implemented to approximate the solution A. The PINN receives temporal–spatial coordinates (x,t) as input and yields the forecast solution $A(x,t)$. To satisfy the constraints of the governing equation, a precise calculation of the temporal derivatives, spatial derivatives, and nonlinear terms is crucial for deducing the value of the equation. Additionally, the network predicts solutions at the initial and boundary, subsequently aligning these predictions with precise data to confirm compliance with the initial and boundary stipulations. Thus, the optimization of network parameters is achieved through the reduction of the MSE loss tied to the governing equation alongside the initial and boundary conditions, delineated as

$$\begin{aligned} \text{Loss} &= \text{MSE}_f + \text{MSE}_{h_0} + \text{MSE}_{h_b} \\ &= \frac{1}{N_f}\sum_{i=1}^{N_f}\left[\frac{\partial A}{\partial t} + N(h)\right]^2 \\ &+ \frac{1}{N_0}\sum_{i=1}^{N_0}\left[A_{\text{pred}}(x_i', 0) - A_0(x_i', 0)\right]^2 \\ &+ \frac{1}{N_b}\sum_{i=1}^{N_b}\left[A_{\text{pred}}(x_i', t_i') - A_b(x_i', t_i')\right]^2 \end{aligned} \tag{12}$$

where $MSE_f, MSE_{h_0}, MSE_{h_b}$ are utilized to penalize the residuals linked to the governing equation, initial, and boundary conditions, respectively. The number of auxiliary coordinates required to compute each MSE term, denoted as N_f, N_0, and N_b. The predicted solution at (x_i', t_i') is represented as $A_{\text{pred}}(x_i', t_i')$. It is noteworthy that the boundary condition specified in Eq. (12) corresponds to the Dirichlet boundary condition. If the PDE involves mixed boundary conditions(such as the Neumann condition), the boundary derivative should also be considered. After numerous iterations, when the loss function is minimized to a sufficiently small value, this means that the network is capable of solving this PDE as a complex operator.

The NLSE can characterize multiple physical effects in fiber optics, as may be seen in Eq. (1). For ultrashort optical pulses with pulse width< 0 ps, high-order nonlinear effects such as self-steepening (SS) [99] and intrapulse Raman scattering (IRS) [100] need to be considered. Therefore, Eq. (1) takes on the following form

$$\begin{aligned} \frac{\partial A}{\partial z} &= -\frac{\alpha}{2}A + i\frac{\beta_2}{2}\frac{\partial^2 A}{\partial t^2} + \frac{\beta_3}{6}\frac{\partial^3 A}{\partial t^3} + iN^2[|A|^2 A \\ &+ is\frac{\partial(|A|^2 A)}{\partial t} - \tau_R\frac{\partial(|A|^2)}{\partial t}A] \end{aligned} \tag{13}$$

where $s = \frac{1}{\omega_0 T_0}$ and $\tau_R = \frac{T_R}{T_0}$ denote the normalized parameters of SS and IRS, respectively. $N^2 = \frac{L_D}{L_N L} = \frac{\gamma P_0 T_0^2}{|\gamma_2|}$, represents the ratio of L_D over the nonlinear length $L_N L$, which governs the comparative influence of GVD and SPM on pulse propagation within the fiber. P_0, T_0, ω_0, and T_R denote the peak power, pulse width, central angular frequency of the pulse, and the initial moment of the Raman response function correspondingly. As a result of the complex nature of the NLSE, the governing equation should be decomposed into real and imaginary parts, which enables the implementation of PINNs in optical fibers. Therefore, the solution $A(z,t)$ is expressed in rectangular notation as $A(z,t) = u(z,t) + iv(z,t)$, where $u(z,t), v(z,t)$ represents the real and imaginary components of $A(z,t)$, respectively. Consequently, the NLSE as depicted in Eq. (13) can be reformulated to express separately its real and imaginary components as follows

$$f(u, v) + ig(u, v) = 0 \tag{14}$$

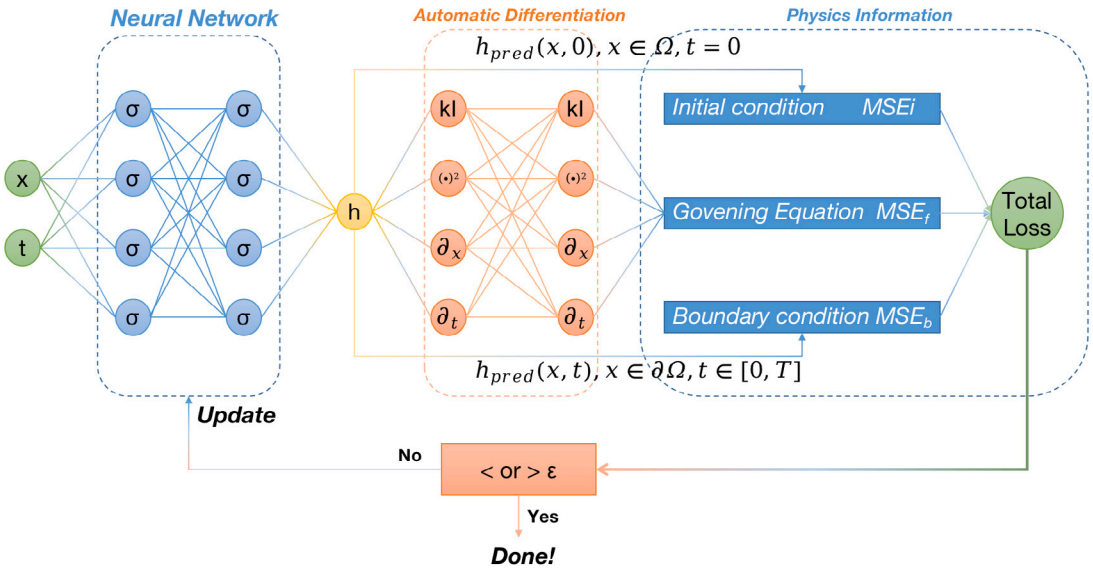

Fig. 10. Structure of Physics-Informed Neural Networks (PINNs).

in which f and g define the real and imaginary components functions of the NLSE, can be articulated as

$$f: \frac{\partial u}{\partial z} + \frac{\alpha}{2}u - \frac{\beta_2}{2}\frac{\partial^2 v}{\partial t^2} - \frac{\beta_3}{6}\frac{\partial^3 u}{\partial t^3} + N^2\left((u^2+v^2)u\right) + s\frac{\partial\left((u^2+v^2)u\right)}{\partial t} - \tau_R\frac{\partial\left((u^2+v^2)v\right)}{\partial t} \tag{15}$$

$$g: \frac{\partial v}{\partial z} + \frac{\alpha}{2}v + \frac{\beta_2}{2}\frac{\partial^2 u}{\partial t^2} - \frac{\beta_3}{6}\frac{\partial^3 v}{\partial t^3} - N^2\left((u^2+v^2)v\right) - s\frac{\partial\left((u^2+v^2)v\right)}{\partial t} - \tau_R\frac{\partial\left((u^2+v^2)u\right)}{\partial t} \tag{16}$$

According to Eqs. (9)–(11), the loss function of PINNs for fiber optics nonlinearity is expressed as

$$\begin{aligned}\text{Loss} &= \text{MSE}_u + \text{MSE}_{v_0} + \text{MSE}_f + \text{MSE}_g \\ &= \frac{1}{N_0}\sum_{i=1}^{N_0}\left|u_{\text{pred}}(0,t_i) - u_0(0,t_i)\right|^2 \\ &+ \frac{1}{N_0}\sum_{i=1}^{N_0}\left|v_{\text{pred}}(0,t_i) - v_0(0,t_i)\right|^2 \\ &+ \frac{1}{N_f}\sum_{i=1}^{N_f}\left|f(z_i,t_i)\right|^2 + \frac{1}{N_f}\sum_{i=1}^{N_f}\left|g(z_i,t_i)\right|^2\end{aligned} \tag{17}$$

where MSE_{u_0} and MSE_{v_0} are the real and imaginary component constraints imposed on the initial pulse. Concurrently, MSE_f and MSE_g are functions to penalize training outcomes that fail to meet the constraints established by the governing equations. In more detail, $u_0(0,t_i)$ and $v_0(0,t_i)$ signify the real and imaginary components of the initial pulse, while $u_{pred}(0,t_i)$ and $v_{pred}(0,t_i)$ are the real and imaginary elements forecast by the PINN for the input $(0,t_i)$. The functions $f(z_i,t_i)$ and $g(z_i,t_i)$ constitute the outcomes derived by the PINN from Eqs. (15) and (16) using the input (z_i,t_i). N_0 and N_f denote the number of auxiliary coordinates necessary to compute the MSE terms associated with the initial conditions and the governing equations.

Fig. 11 depicts the whole process of modeling nonlinear dynamics in an optical fiber by a PINN. As shown in Fig. 11(a), a two-dimensional temporal–spatial domain is constructed to learn the evolution dynamics of pulses within the fiber. Consequently, N_0 data points from the initial pulse and the auxiliary coordinates N_f are sampled within the 2D region, which satisfies the stipulations of the initial condition and the governing equations. It is significant that the precise values at these sampled auxiliary coordinates remain undetermined. As illustrated in Fig. 11(c), utilizing the PINN and automatic differentiation enables the calculation of the four MSE terms in Eq. (12) using temporal–spatial coordinates(z,t). After multiple iterations, the system can model the nonlinear dynamics in fiber accurately and effectively.

4.2.2. Results

Jiang et al. first investigated the performance of modeling the combined influence of linear effects and SPM in fiber [14]. According to their observation, the PINN-based model can effectively model pulse evolution in scenarios where GVD and SPM coexist, as well as those where third-order dispersion (TOD) and SPM are both present. Although the error observed is slightly larger than in the previous PINN work, the evolution within scenarios that feature simultaneous TOD and SPM effects remains effectively modeled.

Furthermore, the study was expanded to model more complex nonlinear effects such as SS and IRS. Quantitatively, the modeling error associated with these advanced nonlinear phenomena compared with SSFM was of magnitude 10^{-2} and could be further improved through optimization. Jiang et al. indicated that while modeling errors were present, the key characteristics of these effects were well-represented, and the approach also avoided the discrete meshing issues typical of traditional numerical methods.

Furthermore, the scope of pulse types was broadened to include optical solitons and multi-pulses. Pertaining to soliton propagation, the analysis encompassed both fundamental and second-order solitons, with the fundamental soliton exhibiting a maximal error of 3.9×10^{-3}. Notable errors were observed for the second-order soliton in two peaks of high pulse power, whereas other regions and the periodic case were accurately characterized by the PINN-based model. For multipulse propagation, two scenarios comprising five individual Gaussian pulses were analyzed: one featuring regular waveforms of uniform amplitudes and another with random waveforms of varying amplitudes. In multipulse propagation of equal amplitude, the modeling error approached a magnitude of 10^{-2}, with significant errors at the peak, especially at the central one. Each pulse was superimposed, broadening, and crosstalking at these peaks, and it was most pronounced at the central peaks when all pulses had equal amplitudes. In contrast, for the different amplitude multipulse propagation, the influences of broadening and crosstalk changed with pulse amplitude. The error was notably substantial at the peaks around the leading edge during evolution, correlating with the lower initial pulse amplitudes at the leading edge. Numerically, the greatest error observed in the time-domain outcomes was 2.16×10^{-2}, slightly exceeding that of single pulse propagation. The study further suggested that augmenting the quantity of coordinate

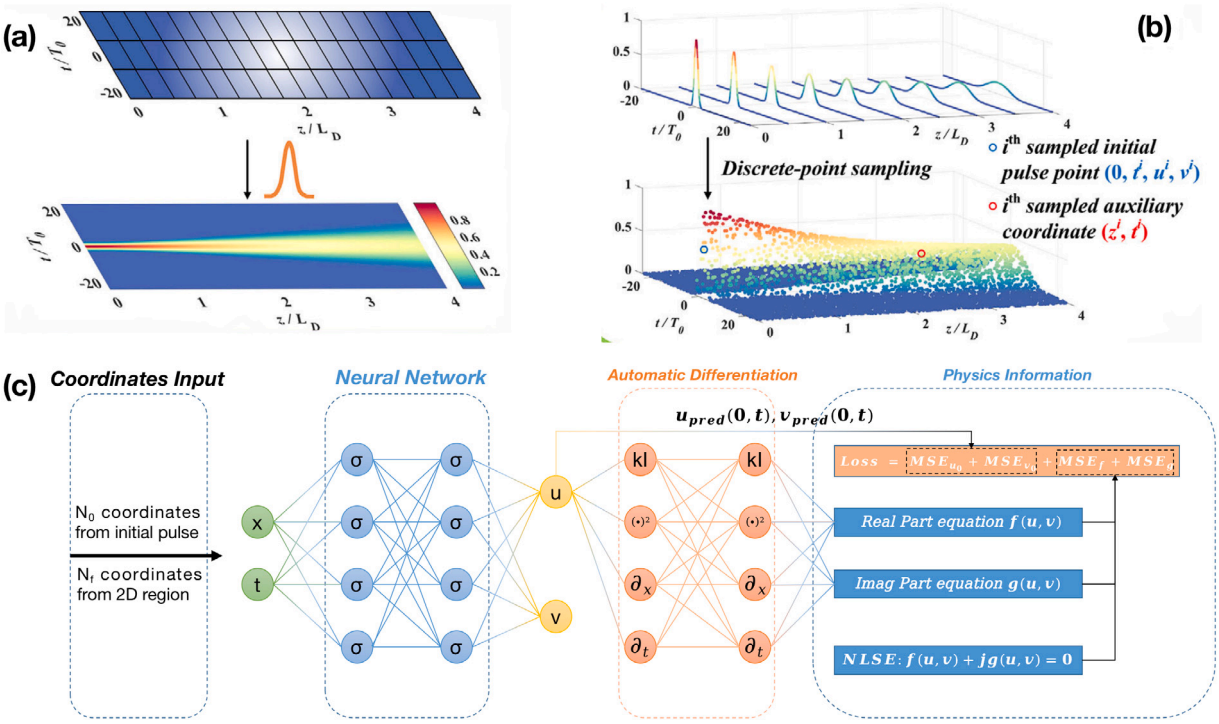

Fig. 11. Process of PINN-based fiber optics modeling: (a) Learning the evolution dynamics of pulses in fiber over a temporal–spatial 2D region [14], (b) Pulse discrete-point sampling [14], (c) Structure of PINN for solving the NLSE in fiber.

points and data could significantly improve the performance of the model. Subsequently, the generalizability of the PINN-based model was validated across different powers and waveforms by adjusting the peak power of the initial pulse and employing random waveforms, respectively. Notably, the robustness for random waveforms illustrates the potential of PINNs to characterize stochastic signal evolution across diverse modulation formats. Regarding complexity, a comparative analysis was carried out between the number of multiplications required by the SSFM-based and PINN-based models, of which the multiplication count for the PINN-based model constituted merely 14.6% of that required by the SSFM-based model. Additionally, the number of PINN multiplications depended mainly on the architecture of the NN, as opposed to the volume of data, making it particularly suitable for scenarios involving extremely large datasets.

4.3. Physics-guided NNs (PGNNs)

4.3.1. Network architectures

PGNNs (or physics-based NNs) embedding the physics-based model as guidance within training to reduce the requirement of training data. This method is common and effective for modeling systems or components governed by physical laws, such as [101–103]. The physics model and NN framework are flexible based on the requirement. The PGNN model was first proposed by Karpatne et al. for lake temperature modeling [104]. Liu et al. built a PGNN by employing the incoherent Gaussian Noise (IGN) model into a semi-supervised NN for fiber nonlinear noise modeling [105–107]. The architecture of the model is shown in Fig. 12(a). Firstly, they integrate the estimation results from physics-based models, such as incoherent or coherent GN models, into the input features. Additionally, they incorporate physical laws directly into the loss function as a form of physical regularization. Specifically, they account for the relationship between nonlinear noise variance and signal power, embedding this physical law into the neural network using a semi-supervised learning framework. Sui et al. proposed a physical-based DL model called Phynet, which models the nonlinear pulse propagation without requiring the true pulse dynamics [108]. Their simulation also examined the inverse propagation problem, demonstrating that Phynet attained performance comparable to conventional DL methods while requiring significantly fewer training datasets.

4.3.2. Results

Liu et al. conducted a comprehensive investigation into two PGNN models: 1D-PGNN and 2D-PGNN. As shown in Fig. 12(a), the former utilized a physics-based model solely for input features, which means only implementing setup 1. While the latter incorporated a physics-based model both as input features and within the loss functions, which means incorporating two setups. Their simulations assessed various metrics, including accuracy, loss curves, the impact of different data sizes, framework configurations, and generalizability.

For accuracy assessment, the modeling precision of CGN, IGN, a generic black-box NN model, and PGNN models using the RMSE was evaluated. The RMSEs were 2.59 dB, 2.02 dB, 1.13 dB, and 0.74 dB for the CGN, IGN, black-box NN, and PGNN models, respectively.

Regarding loss curves, compared to the black-box NN model, the 2D-PGNN exhibited the best performance, closely followed by the 1D-PGNN. This result underscored the enhanced accuracy and faster convergence of PGNN. In scenarios involving different data sizes and structures, the performance of PGNN was also examined with only the physical constraints added for comparison. When data size was limited, the advantage of physics-based input features was more pronounced than that of physical constraints alone, although the benefits tended to converge as data size increased. This pattern suggested that in data-scarce environments, the 1D-PGNN can effectively leverage physics-based model estimations to support learning the mapping relations, whereas the 2D-PGNN significantly outperformed other models due to the dual advantages of physics-based inputs and constraints offering distinct feature perspectives.

In terms of generalizability, the models were tested on datasets with signal power ranging from 1 dBm to 4 dBm, while the training datasets ranged from −4 dBm to 0 dBm. The lowest observed RMSEs for the black-box NN, 1D-PGNN, and 2D-PGNN were 5.87 dB, 2.91 dB, and 0.89 dB, respectively, indicating the robust modeling capabilities of the PGNN across varying operational conditions.

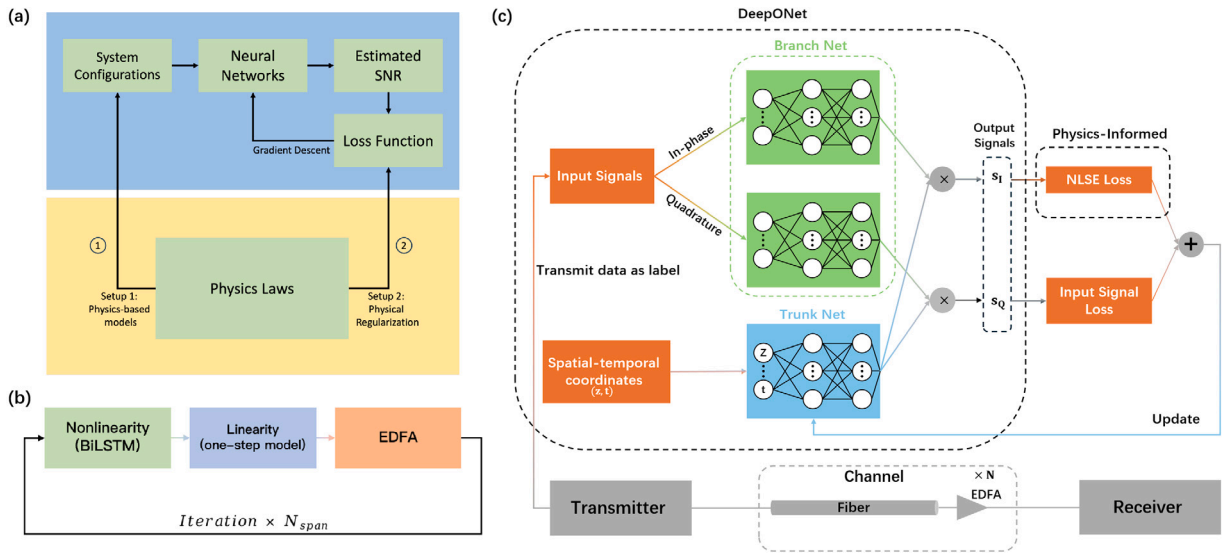

Fig. 12. (a) Architecture of Physics-guided NNs-based model. (b)Transmission processes of FDD-BiLSTM-based optical fiber channel model. (c)Framework of a Physics-Informed Neural Operator (PINO)-based channel.

5. Hybrid method

The previous two sections reviewed data-driven and principle-driven methods. Hybrid methods leverage the strengths of both. This section introduces two such methods, each merging these strengths in distinct ways.

5.1. Feature Decoupling Distributed (FDD) neural network

5.1.1. Network architectures

Yang et al. proposed a distributed FDD modeling approach for long-haul wavelength division multiplexing (WDM) fiber channels [19]. FDD segregates channel effects into linearities, nonlinearities, and random noise sources. Distributed modeling refers to the iterative signal processing operation combining the one-step model, BiLSTM model, and random noise. The linear effects of the channels are modeled using a method derived from the NLSE, while the nonlinear effects of the channels are modeled by the BiLSTM-based model. The simplified diagram of the FDD-BiLSTM modeling scheme, illustrated in Fig. 12(b), demonstrates this concept.

Zheng et al. proposed an FDD-Centre Oriented LSTM (CoLSTM) channel modeling scheme [21], which dramatically reduced the computational complexity compared with SSFM and BiLSTM, by monitoring the ASE noise at the output of each fiber span. This was attributed to the CoLSTM model only requiring a single calculation for all LSTM layers, while the BiLSTM model processed the output data from each LSTM layer sequentially [109]. Specifically, Co-LSTM employs two unidirectional LSTMs to process the input symbols x_{z-t} to x_z (LSTM_L) and x_z to x_{z+t} (LSTM_R), respectively. LSTM_L and LSTM_R independently process the symbol information to the left and right of the target symbol z_t. Finally, the hidden states from the last time step of both LSTMs are concatenated and passed through a fully connected layer (FCL).

5.1.2. Results

Yang et al. focused on evaluating the performance of FDD-BiLSTM in three aspects: modeling accuracy, generalization, and complexity. They assessed modeling accuracy by considering factors including long-haul transmission distance, launch power, and WDM channel numbers. In testing optical time-domain and spectral waveforms, they employed both SSFM and FDD-BiLSTM in a three-channel system at a power of 5 dBm per channel over an 800 km distance. A significant overlap between the waveforms generated by SSFM and FDD-BiLSTM was observed, demonstrating the modeling accuracy of long-haul transmission.

Their investigation into system performance across varying optical launch powers (−10, −6, −3, 0, 3 dBm per channel) for a setup of 5 channels and 400 km showed increasing nMSEs due to intensified distortion from high-intensity nonlinearity. The nMSEs for varying launch powers were $1.32\times10^{-4}, 2.92\times10^{-4}, 4.21\times10^{-4}, 1.43\times10^{-3}, 6.99\times10^{-3}$, respectively. However, the nMSEs remained substantially below the acceptable upper limit of 0.02, affirming the effectiveness of the model at different launch powers.

They also examined configurations of 9, 21, and 41 WDM channels, with launch powers set at (5, −1, −1 dBm per channel) over 80 km, respectively. The spectra of FDD-BiLSTM-generated channel outputs presented a high degree of similarity to the channel outputs generated by SSFM, demonstrating the modeling accuracy across various channel numbers. Additionally, the comparable signal-to-noise ratios (SNRs) for both polarizations under varying channel numbers further confirmed the modeling accuracy.

The generalization capabilities of FDD-BiLSTM was evaluated under a variety of conditions that had never occurred in the training process, including different launch powers, WDM channel numbers, and modulation formats. Launch powers were set to (−7.5 −4.25, −1.25 and 1.75 dBm per channel), yielding low and consistent nMSEs of $2.28\times10^{-4}, 3.8\times10^{-4}, 1.53\times10^{-3}, 3.21\times10^{-3}$, respectively.

Given the variability in WDM channel numbers, the input layer dimension fluctuated due to differing sampling requirements for each channel number. Therefore, it was difficult for a model to adapt to all channels. Consequently, Yang et al. implemented a model designed for the majority number of channels for systems with fewer channels by transmitting all-zero signals in the unused channels. The model was trained under five-channel conditions, and when the signals of the first channel, second channel, second, and fourth channels were transmitted as all-zero signals, the similarity between the optical spectra waveforms generated by SSFM and FDD-BiLSTM underscored the generalizability to varying channel numbers.

For modulation format generalization, they expanded their test datasets to quadrature phase shift keying (QPSK), 32 quadrature amplitude modulation (32QAM), 64QAM, and 128QAM. The nMSEs for these input modulation formats were recorded as 1.04×10^{-4} for QPSK, 1.47×10^{-4} for 32QAM, 1.66×10^{-4} for 64QAM, and 1.50×10^{-4} for 128QAM, respectively. Additionally, the remarkable similarity in the output constellations across varying modulation formats and launched powers effectively further emphasized generalization.

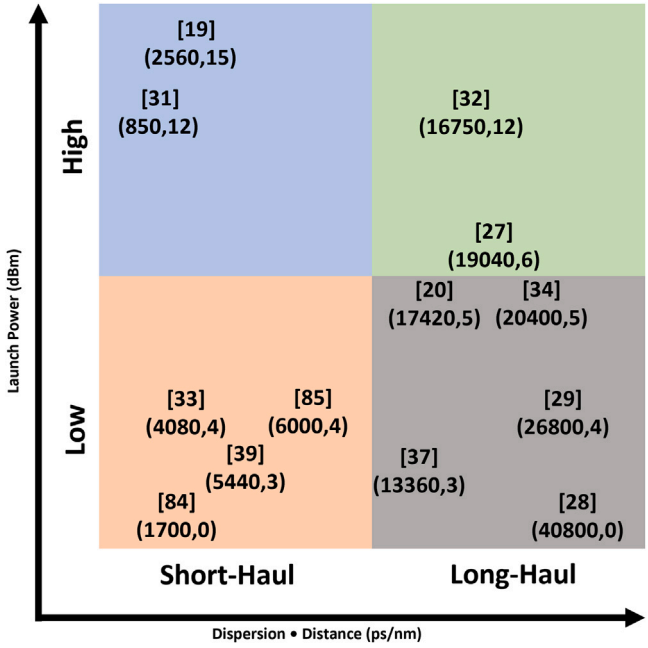

Fig. 13. Comparison of dispersion and nonlinearity in selected systems based on total dispersion (dispersion coefficient multiplied by distance) and the launch power in dBm used in their simulations.

The explicit complexity metrics for FDD-BiLSTM were not detailed. However, Yang et al. evaluated the complexity by considering the computing time ratio between FDD-BiLSTM and SSFM. They focused on two varieties of SSFM algorithms: one with a constant step size of 0.01 km per step, referred to as Constant-Step-Size SSFM (C-SSFM), and the other with a maximum nonlinear phase shift of 0.005 rad per step, known as Nonlinear Phase SSFM (NP-SSFM). The time ratio of FDD-BiLSTM to C-SSFM increased gradually with the number of channels, reaching a minimum of about 1% tested under 5 channels and a maximum of nearly 4% tested under 41 channels. Conversely, the time ratio of FDD-BiLSTM to NP-SSFM decreased with increasing channel numbers, with a minimum of close to 13% tested under 41 channels and a maximum of about 4% tested under 5 channels.

Zheng et al. evaluated the modeling capabilities of FDD-BiLSTM with respect to transmission distance and modulation formats. Their findings indicated that in long-haul transmission scenarios, the nMSEs for both model configurations – with and without the ASE noise – remained below 0.02, affirming its substantial modeling proficiency.

Further analysis was conducted on different modulation formats, specifically testing the model using QPSK, 16QAM, and 64QAM signals. Remarkably, the nMSEs across these tests were consistent, and the variations in SNRs for all tests, regardless of changes in transmission distances and modulation formats, were less than 0.5 dB. This consistency in performance across varying conditions highlights the robustness of the FDD-CoLSTM model to transmission distance and modulation formats in optical communications applications.

5.2. Physics-informed Neural Operator (PINO)

5.2.1. Network architectures

Neural operators are adept at efficiently learning mapping relationships between infinite-dimensional function spaces, which means that the same neural operator model can be applied to different discretizations of function spaces without needing to be retrained. Compared with numerical methods, it has demonstrated superior performance in solving PDEs such as Burger's and the Navier–Stokes Equations [110]. The Physics-informed DeepONet (PI-DeepONet), initially introduced by Wang et al. [111], enhances modeling accuracy and generalizability by constraining that the outputs approximately adhere to the governing physical laws. Following a similar conceptual framework, the PINO was first developed by Li et al., employing the FNO as the neural operator, with physical constraints akin to those in PI-DeepONet. Song et al. utilized PI-DeepONet for modeling optical fiber channels [32], incorporating the NLSE into the loss functions to regulate outputs. Furthermore, they explored NLSE backward learning using PINO, achieving an improvement of 0.17 dB over the 2StPS DBP method [112]. Notably, Song et al. refer to their system as PINO because they regard it as a neural operator with physics-informed constraints and select DeepONet as the neural operator framework, which is different from the PINO of Li et al.. The structure of the PINO-based system is depicted in Fig. 12(c).

5.2.2. Results

In the research conducted by Song et al. [32], system performance was evaluated over an 80 km transmission distance employing the same fiber settings as [84]. Transmission of a 16QAM signal at a line rate of 14 GBd was executed with launch powers of −3, 0, and 3 dBm. Observations indicated that the majority of the MSE per symbol consistently remained below 5×10^{-4}, evidencing the effective modeling performance of PINO. Further, the investigation extended to a 320 km system yielded high-fidelity results, with efficiency over 1000 times faster than the traditional SSFM.

6. Discussion, challenges and future orientation

6.1. Discussion

Table 7 provides a comparative overview of the results from ML-based fiber channel modeling in five dimensions: modulation formats, transmission distance, distortion, channel number, and polarization. Instances marked with an asterisk (*) indicate dispersion and nonlinearity parameters not specified in the respective studies. Table 8 presents reference data for comparing distortions, presenting the dispersion coefficient, maximum launch power, and nonlinear coefficient. Fig. 13 illustrates the fitting scenarios for dispersion and nonlinearity of various systems presented in Table 8, based on transmission distance and launch power. The scenarios are categorized as follows: low nonlinearity, short-haul; high nonlinearity, short-haul; low nonlinearity, long-haul; and high nonlinearity, long-haul.

Research articles [18–21] focus on the capabilities of the BiLSTM-based model, demonstrating its effectiveness in all dimensions. The Transformer model benefits significantly from its multi-head mechanism, showing notable efficiency and broad applicability in diverse modeling scenarios. However, current transformer-based models do not consider multi-channel and PDM scenarios. Although GAN-based models show promising performance, they encounter training challenges such as mode collapse [23–25]. The DNN-based model has proven its modeling capabilities in WDM systems and XPM, yet the performance in longer transmission systems requires further exploration [26]. The FNO excels by concurrently learning time and frequency domain features [70]. Additionally, the neural operator-based approach offers greater flexibility in network outputs compared to conventional NN-based learning. DeepONet has shown substantial modeling proficiency in long-haul, WDM, and PDM systems [29]. Moreover, two modified DeepONet systems have been proposed for wideband systems, addressing both fully loaded and randomly loaded conditions effectively. PINN-based systems primarily focus on modeling nonlinear dynamics [14,83,84]; however, only [85] extends to practical optical signal transmission in fiber systems, challenged by the complexity of data handling. PINN models depend heavily on prior knowledge, such as initial and boundary conditions, governing equations, and physical parameters, which complicates the accuracy assurance in practical settings. Furthermore, while PINN-based models reduce reliance

Table 7
Systematic comparison of ML-based modeling methods.

Method	Research	Modulation formats	Transmission distance (Distance per span)	Distortion	Channel	Polarization
BiLSTM	[18]	OOK, PAM4	80 km	CD, SPM	1	Single
	[19]	16QAM, QPSK, 32QAM, 64QAM, 128QAM	1040(80) km	CD, SPM, PMD, XPM, FWM	41	Dual
	[20]	QPSK, 8QAM, 16QAM, GS-16QAM, 64QAM	1120(80) km	CD, SPM, ASE	1	Single
	[21]	QPSK, 16QAM, 64QAM	2400(80) km	CD, ASE, nonlinearity[a]	1	Single
Transformer	[22]	16QAM, OFDM	1600(80) km	CD, SPM, XPM, FWM, ASE	256	Single
	[23]	OOK, PAM, QPSK, 16QAM	100 km	CD, nonlinearity[a]	1	Single
	[24]	QPSK, 16QAM	50 km	CD, nonlinearity[a]	1	Single
GAN	[25]	16QAM, 16QAM with GN, QPSK, 32APSK, 64QAM	1000(50) km	CD, SPM, ASE	1	Single
DNN	[26]	QPSK, 8PSK, 8QAM, 16QAM	240(80) km	CD, SPM, XPM, ASE	2	Single
FNO	[27]	4QAM, 16QAM, 64QAM	1200(80) km	CD, SPM, ASE	1	Single
DeepONet	[30]	16QAM, PS-16QAM, 32QAM	800(80) km	CD, SPM, XPM, SRS, ASE	96	Dual
PINN	[83,84]	N/A	100 km	Dispersion[a], nonlinearity[a]	1	Single
	[14]	N/A	1000 km	CD, SPM, SS, IRS	1	Single
	[85]	OOK	80 km	dispersion[a], nonlinearity[a]	1	Single
PINO	[32]	16QAM, OOK, QPSK	320(80) km	CD, nonlinearity[a], ASE	1	Single

[a] Represents non-specified.

on training data, random noise factors like ASE and thermal noise from transceivers introduce additional uncertainties. The PINO framework combines the strengths of PINN and DeepONet. However, current studies have not fully demonstrated the advantages of PINO over each individual approach, indicating a requirement for more extensive research to elucidate its benefits.

6.2. Challenges and outlook

Although ML has proven its feasibility in optical fiber channel modeling, a plethora of challenges remain that need to be tackled before extending it to more practical and complex scenarios. In this section, we identify several critical challenges that warrant further exploration in future research.

6.2.1. Generalizability

Generalizability has long been a challenge for various methods. For data-driven approaches, performance significantly relies on the diversity and representativeness of the training data. These models may perform suboptimally under practical conditions or in environments that deviate markedly from those represented in the training dataset. Additionally, acquiring adequate data for optical fiber channel modeling in systems that are newly implemented or not widely used can be costly and challenging. On the other hand, principle-driven methods, while not dependent on extensive datasets, are limited to scenarios that align with their underlying principles, restricting their general applicability. Thus, enhancing the generalizability of ML-based channel modeling methods remains a critical area for improvement.

6.2.2. Interpretability and complexity

Despite the impressive performance of data-driven models, they often provide limited insight into their operational logic, which is valuable for further research of optical communication systems. Physics-informed methods can mitigate issues of interpretability but face challenges in universal applicability. Furthermore, practical scenarios will necessitate real-time or near-real-time modeling support. However, most ML models are computationally demanding and may not efficiently support real-time applications like digital twins [113–116]. Therefore, balancing accuracy and computational complexity will be pivotal for the viability of future applications.

6.2.3. Training difficulty

A well-known issue about ML is the difficulty of training. For instance, GAN faces mode collapse, PINN is generally influenced by training pathologies and suboptimal decisions made by users who may not have deep learning expertise [82,117].

6.2.4. Real-world deployment

Despite theoretical and simulation results, the deployment of ML-based fiber channel models in operational optical networks also remains challenging. First, robust data acquisition remains essential due to the inherent variability of operational environments, including factors such as temperature fluctuations, component aging, and dynamic traffic conditions. Field experiments have emphasized the necessity for continuous monitoring frameworks that collect comprehensive, high-quality channel data across diverse operational conditions [118]. Second, ensuring model adaptability is critical, as ML models trained offline frequently experience performance degradation under previously unseen scenarios. Implementing online learning approaches, which facilitate continuous model updates, has demonstrated the ability to sustain performance within approximately 0.5 dB of optimal levels over extended deployment periods [119]. Nevertheless, incremental model updates must be precisely managed to prevent instability, particularly in multi-carrier systems with complex inter-channel interactions. Third, computational efficiency remains a critical requirement for real-time implementation. This challenge can be mitigated to a certain extent through hardware optimization. FPGA-based implementations of quantized neural networks have achieved inference latencies on the order of microseconds and throughput rates of up to 400 Gbps [120]. Moreover, GPUs enable high-throughput parallel computation and have been effectively applied in accelerating data-driven optical network applications [121].

Table 8
Comparison of dispersion and nonlinearities settings between above-mentioned research in Table 7.

Research	Dispersion coefficient (ps/(nm km))	Nonlinearity (Launch power: dBm, Nonlinear coefficient: W^{-1} km^{-1})
BiLSTM [18]	32	Launch power: 15 Nonlinear coefficient: 1.3
FDD-BiLSTM [19]	16.75	Launch power: 5 Nonlinear coefficient: 1.3
NNSpan [20]	17	Launch power: 6
FDD-CoLSTM [21]	17	Launch power: 0 Nonlinear coefficient: 1.3
Transformer [22]	16.75	Launch power: 4 Nonlinear coefficient: 1.3
Transformer [23]	*	Launch power: 0
Transformer [24]	17	Launch power: 12
GAN [25]	16.75	Launch power: 12 Nonlinear coefficient: 1.3
Generic DNN [26]	17	Launch power: 4 Nonlinear coefficient: 2
FNO [27]	17	Launch power: 5 Nonlinear coefficient: 1.3
DeepONet [30]	16.7	Launch power: 3 Nonlinear coefficient: 1.3
PINN [84]	17	Launch power: 0 Nonlinear coefficient: 1.3
PINN [85]	16	Launch power: 4 Nonlinear coefficient: 1.3
PINO [32]	17	Launch power: 3 Nonlinear coefficient: 1.3

6.2.5. Outlook

The data-driven method demonstrates promising modeling capacity in practical transmission scenarios by neglecting complex channel effects. In contrast, the principle-driven method excels in nonlinear dynamics and pulse propagation due to the use of specific physical equations. Benefits from merging the strength of both data-driven and principle-driven methods, the hybrid method shows significant potential across all three scenarios: practical transmission modeling, nonlinear dynamics, and pulse propagation. Our review indicates a substantial increase in the number of studies focusing on the hybrid method. Additionally, future research should also consider hardware-aware training methodologies, and federated learning strategies to enhance cross-domain model generalization. Such efforts are pivotal for the transition of ML-based approaches from controlled experimental environments to robust, scalable deployments in next-generation optical fiber networks.

7. Conclusions

This review has systematically explored the implementation of ML techniques within the realm of optical fiber channel modeling. Our classification segregates ML approaches into data-driven and principle-driven methods, each catering to distinct aspects of channel modeling challenges. Data-driven techniques, while rapid and efficient, often grapple with issues of low dataset quality and huge time consumption in the training stage, necessitating large datasets for effective training. Conversely, principle-driven techniques merge the rigor of physical laws with the adaptability of ML, enhancing interpretability and reducing excessive data needs, but generalizability remains a concern. The emergence of hybrid models aims to bridge these strengths and limitations by leveraging both physical priors and data-driven flexibility, thereby offering a balanced solution but often demanding careful integration and system design. Quantitative comparisons reveal that no single approach universally outperforms others across all dimensions. Instead, the selection of methods should be context-dependent, guided specifically by application requirements, available data, and interpretability constraints. This review thus serves as a guide for researchers to evaluate, compare, and implement ML-based fiber channel models, and to explore their applications and challenges in increasingly sophisticated and practical scenarios.

CRediT authorship contribution statement

Yulin Wang: Writing – original draft, Visualization, Investigation. **Mark Leeson:** Writing – review & editing. **Zheng Liu:** Writing – review & editing. **Sander Wahls:** Writing – review & editing. **Tongyang Xu:** Writing – review & editing. **Sergei Popov:** Writing – review & editing. **Gan Zheng:** Writing – review & editing. **Tianhua Xu:** Writing – review & editing, Supervision, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work is supported by EU Horizon 2020 Grant 101008280 (DIOR), UK Royal Society Grant (IES\R3\223068), EU Horizon Europe Grant 101131146 (UPGRADE), EU Horizon Europe Grant 101236637 (SPAR), UK EPSRC (EP/Y000315/2, EP/X04047X/2 and EP/Y037243/1).

Appendix. Table of abbreviations

Abbreviation	Definition
ASE	Amplified Spontaneous Emission
BiLSTM	Bidirectional Long Short-Term Memory
C-SSFM	Constant-SSFM
CD	Chromatic Dispersion
CNN	Convolutional Neural Network
DNN	Deep Neural Network
EDFA	Erbium Doped Fiber Amplifier
EVM	Error Vector Magnitude
FDD	Feature Decoupling Distributed
FFT	Fast Fourier Transform
FNO	Fourier Neural Operator
FWM	Four-Wave Mixing
GAN	Generative Adversarial Network
GN	Gaussian Noise
GVD	Group Velocity Dispersion
IFFT	Inverse Fast Fourier Transform
IRS	Intrapulse Raman Scattering
ISI	Inter-Symbol Interference
LDBP	Learned Digital Back-Propagation
LSTM	Long Short-Term Memory
ML	Machine Learning
nMSE	Normalized Mean Squared Error
NL	Nonlinearity
NLSE	Nonlinear Schrödinger Equation
NP-SSFM	Nonlinear Phase SSFM
OOK	On-Off Keying
OFDM	Orthogonal Frequency-Division Multiplexing
PAM4	Pulse Amplitude Modulation 4-level
PGNN	Physics-Guided Neural Network
PINN	Physics-Informed Neural Network
PINO	Physics-Informed Neural Operator
QPSK	Quadrature Phase Shift Keying
RNN	Recurrent Neural Network
SNR	Signal-to-Noise Ratio
SPM	Self-Phase Modulation
SS	Self-Steepening
SSFM	Split-Step Fourier Method
XPM	Cross-Phase Modulation

Data availability

No data was used for the research described in the article.

References

[1] G.P. Agrawal, Nonlinear fiber optics, in: Nonlinear Science at the Dawn of the 21st Century, Springer, 2000, pp. 195–211.

[2] T. Xu, N.A. Shevchenko, Y. Zhang, C. Jin, J. Zhao, T. Liu, Information rates in Kerr nonlinearity limited optical fiber communication systems, Opt. Express 29 (2021) 17428–17439.

[3] J. Shao, X. Liang, S. Kumar, Comparison of split-step Fourier schemes for simulating fiber optic communication systems, IEEE Photonics J. 6 (4) (2014) 1–15.

[4] P.J. Winzer, D.T. Neilson, A.R. Chraplyvy, Fiber-optic transmission and networking: the previous 20 and the next 20 years, Opt. Express 26 (18) (2018) 24190–24239.

[5] F.N. Khan, Q. Fan, C. Lu, A.P.T. Lau, An optical communication's perspective on machine learning and its applications, J. Lightwave Technol. 37 (2) (2019) 493–516.

[6] F. Musumeci, C. Rottondi, A. Nag, I. Macaluso, D. Zibar, M. Ruffini, M. Tornatore, An overview on application of machine learning techniques in optical networks, IEEE Commun. Surv. Tutor. 21 (2) (2018) 1383–1408.

[7] F. Musumeci, C. Rottondi, G. Corani, S. Shahkarami, F. Cugini, M. Tornatore, A tutorial on machine learning for failure management in optical networks, J. Lightwave Technol. 37 (16) (2019) 4125–4139.

[8] J.W. Nevin, S. Nallaperuma, N.A. Shevchenko, X. Li, M.S. Faruk, S.J. Savory, Machine learning for optical fiber communication systems: An introduction and overview, Apl Photonics 6 (12) (2021).

[9] N. Gautam, A. Choudhary, B. Lall, Comparative study of neural network architectures for modelling nonlinear optical pulse propagation, Opt. Fiber Technol. 64 (2021) 102540.

[10] X. Jiang, D. Wang, Q. Fan, M. Zhang, C. Lu, A.P.T. Lau, Solving the nonlinear Schrödinger equation in optical fibers using physics-informed neural network, in: Optical Fiber Communication Conference, Optica Publishing Group, 2021, pp. M3H–8.

[11] G. Genty, L. Salmela, J.M. Dudley, D. Brunner, A. Kokhanovskiy, S. Kobtsev, S.K. Turitsyn, Machine learning and applications in ultrafast photonics, Nat. Photonics 15 (2) (2021) 91–101.

[12] P. Freire, E. Manuylovich, J.E. Prilepsky, S.K. Turitsyn, Artificial neural networks for photonic applications—from algorithms to implementation: tutorial, Adv. Opt. Photonics 15 (3) (2023) 739–834.

[13] D. Wang, X. Jiang, Y. Song, M. Fu, Z. Zhang, X. Chen, M. Zhang, Applications of physics-informed neural network for optical fiber communications, IEEE Commun. Mag. 60 (9) (2022) 32–37.

[14] X. Jiang, D. Wang, Q. Fan, M. Zhang, C. Lu, A.P.T. Lau, Physics-informed neural network for nonlinear dynamics in fiber optics, Laser Photonics Rev. 16 (9) (2022) 2100483.

[15] D. Wang, X. Jiang, Y. Song, M. Fu, Z. Zhang, X. Chen, M. Zhang, Applications of physics-informed neural network for optical fiber communications, IEEE Commun. Mag. 60 (9) (2022) 32–37.

[16] S.K. Turitsyn, J.E. Prilepsky, S.T. Le, S. Wahls, L.L. Frumin, M. Kamalian, S.A. Derevyanko, Nonlinear Fourier transform for optical data processing and transmission: advances and perspectives, Optica 4 (3) (2017) 307–322.

[17] O.V. Sinkin, R. Holzlöhner, J. Zweck, C.R. Menyuk, Optimization of the split-step Fourier method in modeling optical-fiber communications systems, J. Lightwave Technol. 21 (1) (2003) 61.

[18] D. Wang, Y. Song, J. Li, J. Qin, T. Yang, M. Zhang, X. Chen, A.C. Boucouvalas, Data-driven optical fiber channel modeling: A deep learning approach, J. Lightwave Technol. 38 (17) (2020) 4730–4743.

[19] H. Yang, Z. Niu, H. Zhao, S. Xiao, W. Hu, L. Yi, 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) (2022) 4571–4580.

[20] R. Jiang, Z. Wang, T. Jia, Z. Fu, C. Shang, C. Wu, Flexible optical fiber channel modeling based on a neural network module, Opt. Lett. 48 (16) (2023) 4332–4335.

[21] J. Zheng, T. Zhang, F. Zhang, Co-LSTM-based fiber link modeling with ASE noise tracking for long-haul coherent optical transmission, Opt. Lett. 49 (7) (2024) 1848–1851.

[22] N. Zhang, H. Yang, Z. Niu, L. Zheng, C. Chen, S. Xiao, L. Yi, Transformer-based long distance fiber channel modeling for optical OFDM systems, J. Lightwave Technol. 40 (24) (2022) 7779–7789.

[23] Y. Zang, Z. Yu, K. Xu, M. Chen, S. Yang, H. Chen, Data-driven fiber model based on the deep neural network with multi-head attention mechanism, Opt. Express 30 (26) (2022) 46626–46648.

[24] Y. Zhu, J. Ye, L. Yan, T. Zhou, P. Li, X. Zou, W. Pan, Transformer-based high-fidelity modeling method for radio over fiber link, J. Lightwave Technol. 41 (9) (2023) 2657–2665.

[25] H. Yang, Z. Niu, S. Xiao, J. Fang, Z. Liu, D. Fainsin, L. Yi, Fast and accurate optical fiber channel modeling using generative adversarial network, J. Lightwave Technol. 39 (5) (2020) 1322–1333.

[26] R. Jiang, Z. Fu, Y. Bao, H. Wang, X. Ding, Z. Wang, Data-driven method for nonlinear optical fiber channel modeling based on deep neural network, IEEE Photonics J. 14 (4) (2022) 1–8.

[27] X. He, L. Yan, L. Jiang, A. Yi, Z. Pu, Y. Yu, H. Chen, W. Pan, B. Luo, Fourier neural operator for accurate optical fiber modeling with low complexity, J. Lightwave Technol. 41 (8) (2022) 2301–2311.

[28] Z. Li, N. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhattacharya, A. Stuart, A. Anandkumar, Fourier neural operator for parametric partial differential equations, 2020, arXiv preprint arXiv:2010.08895.

[29] X. Zhang, D. Wang, Y. Song, X. Jiang, J. Li, M. Zhang, Neural operator-based fiber channel modeling for WDM optical transmission system, in: 2023 Opto-Electronics and Communications Conference, OECC, IEEE, 2023, pp. 1–4.

[30] X. Zhang, M. Zhang, Y. Song, X. Jiang, F. Zhang, D. Wang, Deeponet-based waveform-level simulation for a wideband nonlinear wdm system, J. Lightwave Technol. 41 (22) (2023) 6908–6922.

[31] C. Häger, H.D. Pfister, Physics-based deep learning for fiber-optic communication systems, IEEE J. Sel. Areas Commun. 39 (1) (2020) 280–294.

[32] Y. Song, D. Wang, Q. Fan, X. Jiang, X. Luo, M. Zhang, Physics-informed neural operator for fast and scalable optical fiber channel modelling in multi-span transmission, in: 2022 European Conference on Optical Communication, ECOC, IEEE, 2022, pp. 1–4.

[33] G.P. Agrawal, Fiber-Optic Communication Systems, John Wiley & Sons, 2012.

[34] E. Ip, A.P.T. Lau, D.J. Barros, J.M. Kahn, Coherent detection in optical fiber systems, Opt. Express 16 (2) (2008) 753–791.

[35] J. Zhao, Y. Liu, T. Xu, Advanced DSP for coherent optical fiber communication, Appl. Sci. 9 (19) (2019) 4192.

[36] J. Liu, Y. Yi, T. Xu, Coherent model for cross-polarization coupling from a single point-like perturbation in optical fibers, J. Lightwave Technol. (2024).

[37] I. De Miguel, R.J. Durán, T. Jiménez, N. Fernández, J.C. Aguado, R.M. Lorenzo, A. Caballero, I.T. Monroy, Y. Ye, A. Tymecki, et al., Cognitive dynamic optical networks, J. Opt. Commun. Netw. 5 (10) (2013) A107–A118.

[38] D. Wang, Z. Zhang, M. Zhang, M. Fu, J. Li, S. Cai, C. Zhang, X. Chen, The role of digital twin in optical communication: fault management, hardware configuration, and transmission simulation, IEEE Commun. Mag. 59 (1) (2021) 133–139.

[39] Y. Song, M. Zhang, Y. Zhang, Y. Shi, S. Shen, X. Tang, S. Huang, D. Wang, Lifecycle management of optical networks with dynamic-updating digital twin: a hybrid data-driven and physics-informed approach, IEEE J. Sel. Areas Commun. (2025).

[40] O.I. Abiodun, A. Jantan, A.E. Omolara, K.V. Dada, N.A. Mohamed, H. Arshad, State-of-the-art in artificial neural network applications: A survey, Heliyon 4 (11) (2018).

[41] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (8) (1997) 1735–1780.

[42] S. Deligiannidis, A. Bogris, C. Mesaritakis, Y. Kopsinis, Compensation of fiber nonlinearities in digital coherent systems leveraging long short-term memory neural networks, J. Lightwave Technol. 38 (21) (2020) 5991–5999.

[43] O. Kotlyar, M. Kamalian-Kopae, M. Pankratova, A. Vasylchenkova, J.E. Prilepsky, S.K. Turitsyn, Convolutional long short-term memory neural network equalizer for nonlinear Fourier transform-based optical transmission systems, Opt. Express 29 (7) (2021) 11254–11267.

[44] L. Salmela, N. Tsipinakis, A. Foi, C. Billet, J.M. Dudley, G. Genty, Predicting ultrafast nonlinear dynamics in fibre optics with a recurrent neural network, Nat. Mach. Intell. 3 (4) (2021) 344–354.

[45] Y. Fang, H.-B. Han, W.-B. Bo, W. Liu, B.-H. Wang, Y.-Y. Wang, C.-Q. Dai, Deep neural network for modeling soliton dynamics in the mode-locked laser, Opt. Lett. 48 (3) (2023) 779–782.

[46] J. Ding, T. Liu, T. Xu, W. Hu, S. Popov, M.S. Leeson, J. Zhao, T. Xu, Intra-channel nonlinearity mitigation in optical fiber transmission systems using perturbation-based neural network, J. Lightwave Technol. 40 (21) (2022) 7106–7116.

[47] J. Liu, X. Huang, X. Ren, J. Li, J. Zhang, Q. Zhang, Accurate data-driven fiber channel modeling based on BiLSTM and conditional GAN, in: 2023 Opto-Electronics and Communications Conference, OECC, IEEE, 2023, pp. 1–4.

[48] G. Pu, R. Liu, H. Yang, Y. Xu, W. Hu, M. Hu, L. Yi, Fast predicting the complex nonlinear dynamics of mode-locked fiber laser by a recurrent neural network with prior information feeding, Laser Photonics Rev. 17 (6) (2023) 2200363.

[49] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial networks, 2020.

[50] M. Mirza, S. Osindero, Conditional generative adversarial nets, 2014, arXiv preprint arXiv:1411.1784.

[51] H. Ye, G.Y. Li, B.-H.F. Juang, K. Sivanesan, Channel agnostic end-to-end learning based communication systems with conditional GAN, in: 2018 IEEE Globecom Workshops, GC Wkshps, IEEE, 2018, pp. 1–5.

[52] H. Ye, L. Liang, G.Y. Li, B.-H. Juang, Deep learning-based end-to-end wireless communication systems with conditional GANs as unknown channels, IEEE Trans. Wirel. Commun. 19 (5) (2020) 3133–3143.

[53] Y. Yang, Y. Li, W. Zhang, F. Qin, P. Zhu, C.-X. Wang, Generative-adversarial-network-based wireless channel modeling: Challenges and opportunities, IEEE Commun. Mag. 57 (3) (2019) 22–27.

[54] N. Shlezinger, N. Farsad, Y.C. Eldar, A.J. Goldsmith, ViterbiNet: A deep learning based viterbi algorithm for symbol detection, IEEE Trans. Wirel. Commun. 19 (5) (2020) 3319–3331.

[55] L. Sun, Y. Wang, A.L. Swindlehurst, X. Tang, Generative-adversarial-network enabled signal detection for communication systems with unknown channel models, IEEE J. Sel. Areas Commun. 39 (1) (2020) 47–60.

[56] C. Zou, F. Yang, J. Song, Z. Han, Underwater optical channel generator: A generative adversarial network based approach, IEEE Trans. Wirel. Commun. 21 (11) (2022) 9394–9403.

[57] W. Chen, M. Zhang, D. Wang, Y. Zhan, S. Cai, H. Yang, Z. Zhang, X. Chen, D. Wang, Deep learning-based channel modeling for free space optical communications, J. Lightwave Technol. 41 (1) (2022) 183–198.

[58] Y. Wei, C. Chen, F. Li, L. Yao, H. Zhang, Y. Liu, Z. Li, C. Shen, J. Zhang, N. Chi, et al., An accurate and realistic channel simulator of optical wireless communication systems combining deterministic and random noise, J. Lightwave Technol. 42 (8) (2024) 2666–2682.

[59] G. Ye, J. Xiang, G. Zhou, M. Xiang, J. Li, Y. Qin, S. Fu, Impact of the input OSNR on data-driven optical fiber channel modeling, J. Opt. Commun. Netw. 15 (2) (2023) 78–86.

[60] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, Adv. Neural Inf. Process. Syst. 30 (2017).

[61] M.-H. Guo, Z.-N. Liu, T.-J. Mu, S.-M. Hu, Beyond self-attention: External attention using two linear layers for visual tasks, IEEE Trans. Pattern Anal. Mach. Intell. 45 (5) (2022) 5436–5447.

[62] T. Domhan, How much attention do you need? A granular analysis of neural machine translation architectures, in: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2018, pp. 1799–1808.

[63] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.

[64] M. Nazarathy, J. Khurgin, R. Weidenfeld, Y. Meiman, P. Cho, R. Noe, I. Shpantzer, V. Karagodsky, Phased-array cancellation of nonlinear FWM in coherent OFDM dispersive multi-span links, Opt. Express 16 (20) (2008) 15777–15810.

[65] H. Sui, H. Zhu, H. Jia, Q. Li, M. Ou, B. Luo, X. Zou, L. Yan, Predicting nonlinear multi-pulse propagation in optical fibers via a lightweight convolutional neural network, Opt. Lett. 48 (18) (2023) 4889–4892.

[66] N. Gautam, V. Kaushik, A. Choudhary, B. Lall, OptiDistillNet: Learning nonlinear pulse propagation using the student-teacher model, Opt. Express 30 (23) (2022) 42430–42439.

[67] G. Hinton, O. Vinyals, J. Dean, Distilling the knowledge in a neural network, 2015, arXiv preprint arXiv:1503.02531.

[68] G. Wen, Z. Li, K. Azizzadenesheli, A. Anandkumar, S.M. Benson, U-FNO—An enhanced Fourier neural operator-based deep-learning model for multiphase flow, Adv. Water Resour. 163 (2022) 104180.

[69] Z. Li, M. Liu-Schiaffini, N. Kovachki, B. Liu, K. Azizzadenesheli, K. Bhattacharya, A. Stuart, A. Anandkumar, Learning dissipative dynamics in chaotic systems, 2021, arXiv preprint arXiv:2106.06898.

[70] J. Feng, L. Jiang, L. Yan, A. Yi, S.-s. Li, W. Pan, B. Luo, Y. Pan, B. Xu, L. Yi, et al., Modeling of a multi-parameter chaotic optoelectronic oscillator based on the Fourier neural operator, Opt. Express 30 (25) (2022) 44798–44813.

[71] T. Chen, H. Chen, Approximations of continuous functionals by neural networks with application to dynamic systems, IEEE Trans. Neural Netw. 4 (6) (1993) 910–918.

[72] T. Chen, H. Chen, Approximation capability to functions of several variables, nonlinear functionals, and operators by radial basis function neural networks, IEEE Trans. Neural Netw. 6 (4) (1995) 904–910.

[73] L. Lu, P. Jin, G. Pang, Z. Zhang, G.E. Karniadakis, Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators, Nat. Mach. Intell. 3 (3) (2021) 218–229.

[74] E. Ip, J.M. Kahn, Compensation of dispersion and nonlinear impairments using digital backpropagation, J. Lightwave Technol. 26 (20) (2008) 3416–3425.

[75] C. Häger, H.D. Pfister, Nonlinear interference mitigation via deep neural networks, in: Optical Fiber Communication Conference, Optica Publishing Group, 2018, pp. W3A–4.

[76] M. Lian, C. Häger, H.D. Pfister, What can machine learning teach us about communications? in: 2018 IEEE Information Theory Workshop, ITW, IEEE, 2018, pp. 1–5.

[77] C. Häger, H.D. Pfister, Deep learning of the nonlinear Schrödinger equation in fiber-optic communications, in: 2018 IEEE International Symposium on Information Theory, ISIT, IEEE, 2018, pp. 1590–1594.

[78] Q. Fan, G. Zhou, T. Gui, C. Lu, A.P.T. Lau, Advancing theoretical understanding and practical performance of signal processing for nonlinear optical communications through machine learning, Nat. Commun. 11 (1) (2020) 3694.

[79] S. Zhang, F. Yaman, K. Nakamura, T. Inoue, V. Kamalov, L. Jovanovski, V. Vusirikala, E. Mateo, Y. Inada, T. Wang, Field and lab experimental demonstration of nonlinear impairment compensation using neural networks, Nat. Commun. 10 (1) (2019) 3033.

[80] R. Dar, P.J. Winzer, Nonlinear interference mitigation: Methods and potential gain, J. Lightwave Technol. 35 (4) (2017) 903–930.

[81] M. Raissi, P. Perdikaris, G.E. Karniadakis, Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, J. Comput. Phys. 378 (2019) 686–707.

[82] S. Wang, S. Sankaran, H. Wang, P. Perdikaris, An expert's guide to training physics-informed neural networks, 2023, arXiv preprint arXiv:2308.08468.

[83] Y. Zang, Z. Yu, K. Xu, M. Chen, S. Yang, H. Chen, Universal fiber models based on PINN neural network, in: Asia Communications and Photonics Conference, Optica Publishing Group, 2020, pp. M4A–266.

[84] Y. Zang, Z. Yu, K. Xu, X. Lan, M. Chen, S. Yang, H. Chen, Principle-driven fiber transmission model based on PINN neural network, J. Lightwave Technol. 40 (2) (2021) 404–414.

[85] X. Jiang, D. Wang, X. Chen, M. Zhang, Physics-informed neural network for optical fiber parameter estimation from the nonlinear Schrödinger equation, J. Lightwave Technol. 40 (21) (2022) 7095–7105.

[86] X. Jiang, M. Zhang, Y. Song, H. Chen, D. Huang, D. Wang, Predicting ultrafast nonlinear dynamics in fiber optics by enhanced physics-informed neural network, J. Lightwave Technol. 42 (5) (2023) 1381–1394.

[87] J. Uduagbomen, S. Lakshminarayana, M.S. Leeson, T. Xu, Physics-informed neural network modeling of soliton pulses in optical communication systems, in: 2022 IEEE Photonics Society Summer Topicals Meeting Series, SUM, IEEE, 2022, pp. 1–2.

[88] J. Uduagbomen, S. Lakshminarayana, Z. Liu, M.S. Leeson, T. Xu, Physics-informed neural network for fibre channel modelling in optical communication systems, in: 2023 23rd International Conference on Transparent Optical Networks, ICTON, IEEE, 2023, pp. 1–4.
[89] J. Uduagbomen, M.S. Leeson, Z. Liu, S. Lakshminarayana, T. Xu, Modified failproof physics-informed neural network framework for fast and accurate optical fiber transmission link modeling, Appl. Opt. 63 (14) (2024) 3794–3802.
[90] Y. Chen, S. Koohy, Gpt-pinn: Generative pre-trained physics-informed neural networks toward non-intrusive meta-learning of parametric pdes, Finite Elem. Anal. Des. 228 (2024) 104047.
[91] Y. Zang, B. Hua, Z. Lin, F. Zhang, S. Li, Z. Zhang, H. Chen, Fiber transmission model with parameterized inputs based on GPT-PINN neural network, 2024, arXiv preprint arXiv:2408.09947.
[92] Y. Zang, B. Hua, Z. Lin, F. Zhang, S. Li, Z. Zhang, H. Chen, Fiber transmission model with parameterized inputs based on generative pre-trained physics-informed neural networks, Opt. Express 33 (1) (2025) 50–61.
[93] J. Pu, J. Li, Y. Chen, Solving localized wave solutions of the derivative nonlinear Schrödinger equation using an improved PINN method, Nonlinear Dynam. 105 (2021) 1723–1739.
[94] Y. Fang, G.-Z. Wu, Y.-Y. Wang, C.-Q. Dai, Data-driven femtosecond optical soliton excitations and parameters discovery of the high-order NLSE using the PINN, Nonlinear Dynam. 105 (1) (2021) 603–616.
[95] Y. Fang, G.-Z. Wu, X.-K. Wen, Y.-Y. Wang, C.-Q. Dai, Predicting certain vector optical solitons via the conservation-law deep-learning method, Opt. Laser Technol. 155 (2022) 108428.
[96] B.-W. Zhu, Y. Fang, W. Liu, C.-Q. Dai, Predicting the dynamic process and model parameters of vector optical solitons under coupled higher-order effects via WL-tsPINN, Chaos Solitons Fractals 162 (2022) 112441.
[97] G.-Z. Wu, Y. Fang, N.A. Kudryashov, Y.-Y. Wang, C.-Q. Dai, Prediction of optical solitons using an improved physics-informed neural network method with the conservation law constraint, Chaos Solitons Fractals 159 (2022) 112143.
[98] Y.-H. Yin, X. Lü, Dynamic analysis on optical pulses via modified PINNs: Soliton solutions, rogue waves and parameter discovery of the CQ-NLSE, Commun. Nonlinear Sci. Numer. Simul. 126 (2023) 107441.
[99] F. DeMartini, C. Townes, T. Gustafson, P. Kelley, Self-steepening of light pulses, Phys. Rev. 164 (2) (1967) 312.
[100] K.J. Blow, D. Wood, Theoretical description of transient stimulated Raman scattering in optical fibers, IEEE J. Quantum Electron. 25 (12) (1989) 2665–2673.
[101] Q. Zhuge, X. Zeng, H. Lun, M. Cai, X. Liu, L. Yi, W. Hu, Application of machine learning in fiber nonlinearity modeling and monitoring for elastic optical networks, J. Lightwave Technol. 37 (13) (2019) 3055–3063.
[102] Y. Pointurier, Machine learning techniques for quality of transmission estimation in optical networks, J. Opt. Commun. Netw. 13 (4) (2021) B60–B71.
[103] S. Zhu, C. Gutterman, A.D. Montiel, J. Yu, M. Ruffini, G. Zussman, D. Kilper, Hybrid machine learning EDFA model, in: Optical Fiber Communication Conference, Optica Publishing Group, 2020, pp. T4B–4.
[104] A. Karpatne, W. Watkins, J. Read, V. Kumar, Physics-guided neural networks (pgnn): An application in lake temperature modeling, 2017, arXiv preprint arXiv:1710.11431. 2.
[105] P. Poggiolini, The GN model of non-linear propagation in uncompensated coherent optical systems, J. Lightwave Technol. 30 (24) (2012) 3857–3879.
[106] X. Liu, L. Liu, H. Lun, Y. Zhang, L. Yi, W. Hu, Q. Zhuge, A grey-box model for estimating nonlinear SNR in optical networks based on physics-guided neural networks, in: 2021 Asia Communications and Photonics Conference, ACP, IEEE, 2021, pp. 1–3.
[107] X. Liu, Y. Fan, Y. Zhang, M. Cai, L. Liu, L. Yi, W. Hu, Q. Zhuge, Fusing physics to fiber nonlinearity model for optical networks based on physics-guided neural networks, J. Lightwave Technol. 40 (17) (2022) 5793–5802.
[108] H. Sui, H. Zhu, B. Luo, S. Taccheo, X. Zou, L. Yan, Physics-based deep learning for modeling nonlinear pulse propagation in optical fibers, Opt. Lett. 47 (15) (2022) 3912–3915.
[109] H. Ming, X. Chen, X. Fang, L. Zhang, C. Li, F. Zhang, Ultralow complexity long short-term memory network for fiber nonlinearity mitigation in coherent optical communication systems, J. Lightwave Technol. 40 (8) (2022) 2427–2434.
[110] N. Kovachki, Z. Li, B. Liu, K. Azizzadenesheli, K. Bhattacharya, A. Stuart, A. Anandkumar, Neural operator: Learning maps between function spaces with applications to pdes, J. Mach. Learn. Res. 24 (89) (2023) 1–97.
[111] S. Wang, H. Wang, P. Perdikaris, Learning the solution operator of parametric partial differential equations with physics-informed DeepONets, Sci. Adv. 7 (40) (2021) eabi8605.
[112] Y. Song, X. Jiang, X. Luo, X. Zhang, M. Zhang, D. Wang, Physics-informed neural operator-based full wavefield back-propagation for multi-span optical transmission, in: Optical Fiber Communication Conference, Optica Publishing Group, 2023, pp. M2F–5.
[113] D. Wang, Z. Zhang, M. Zhang, M. Fu, J. Li, S. Cai, C. Zhang, X. Chen, The role of digital twin in optical communication: fault management, hardware configuration, and transmission simulation, IEEE Commun. Mag. 59 (1) (2021) 133–139.
[114] Y. Song, M. Zhang, Y. Zhang, Y. Shi, S. Shen, B. Guo, S. Huang, D. Wang, Implementing digital twin in field-deployed optical networks: Uncertain factors, operational guidance, and field-trial demonstration, IEEE Netw. (2023).
[115] Q. Zhuge, X. Liu, Y. Zhang, M. Cai, Y. Liu, Q. Qiu, X. Zhong, J. Wu, R. Gao, L. Yi, et al., Building a digital twin for intelligent optical networks [invited tutorial], J. Opt. Commun. Netw. 15 (8) (2023) C242–C262.
[116] M.S. Faruk, S.J. Savory, Measurement informed models and digital twins for optical fiber communication systems, J. Lightwave Technol. 42 (3) (2023) 1016–1030.
[117] A. Srivastava, L. Valkov, C. Russell, M.U. Gutmann, C. Sutton, Veegan: Reducing mode collapse in gans using implicit variational learning, Adv. Neural Inf. Process. Syst. 30 (2017).
[118] S. Shen, R. Yang, H. Li, Y. Teng, R. Wang, R. Nejabati, S. Yan, D. Simeonidou, Field-trial demonstration of ML deployment in optical networks using telemetry and AI engine, in: 49th European Conference on Optical Communications, ECOC 2023, Vol. 2023, IET, 2023, pp. 1126–1129.
[119] S. Lennard, F.A. Barbosa, F.M. Ferreira, Zero-shot ML equalization in coherent optical transmission: A first experimental demonstration, J. Lightwave Technol. (2025).
[120] P.J. Freire, S. Srivallapanondh, M. Anderson, B. Spinnler, T. Bex, T.A. Eriksson, A. Napoli, W. Schairer, N. Costa, M. Blott, et al., Implementing neural network-based equalizers in a coherent optical transmission system using field-programmable gate arrays, J. Lightwave Technol. 41 (12) (2023) 3797–3815.
[121] P. Castoldi, R.A. Bakar, A. Sgambelluri, J.J.V. Olmos, F. Paolucci, F. Cugini, Programmable packet-optical network security and monitoring using DPUs with embedded GPUs, J. Opt. Commun. Netw. 17 (2) (2025) A178–A195.