Diffusion Model Enabled Data-Physics Hybrid Waveform Modeling for Experimental Optical Transmission System

Minghui Shi, Zekun Niu, Junzhe Xiao, Lyu Li, Nianheng Jiang, Mingzhe Chen, Weisheng Hu, and Lilin Yi

***Abstract*—Waveform-level modeling of optical transmission systems serves as a pivotal tool for characterizing system properties, enabling cost-effective system design and optimization. However, conventional physics-based methods, such as the split-step Fourier method (SSFM), suffer from high computational complexity and exhibit performance discrepancies when compared with experimental systems due to parameter mismatches. Meanwhile, although purely data-driven neural network (NN) approaches are widely employed in simulations, they often lack rigorous validation within real-world experimental setups. To address these challenges, this paper proposes a diffusion model-enabled data-physics hybrid framework to achieve accurate and generalizable modeling of optical transmission systems. By leveraging the inherent iterative denoising process of diffusion models, the proposed method effectively characterizes system noise distributions. Compared to traditional generative adversarial networks (GANs), it demonstrates superior training stability and enhanced generalization capabilities. The effectiveness of the proposed model is experimentally validated over a 40-channel, 60 GBaud, 1600 km full C-band transmission system. Experimental results indicate that the diffusion model exhibits consistently higher stability and accuracy across multiple training trials than the GAN-based baseline. In generalization tests across varying launch powers and transmission distances, the diffusion model achieves average Q-factor errors of merely 0.06 dB and 0.09 dB. These results correspond to error reductions of 50% and 78%, respectively, compared to the GAN baseline. Furthermore, the model demonstrates exceptional cross-modulation format generalization, enabling it to be trained on a single modulation format and directly applied to others. Leveraging this framework, we successfully achieve efficient transmission performance prediction, modulation format selection, and parameter optimization for nonlinear compensation (NLC) algorithms within our experimental platform. Ultimately, this proposed framework holds promise for lowering the barrier to experimental validation and expediting the intelligent evolution of next-generation optical networks.**

***Index Terms*—Deep learning (DL), diffusion model, waveform-level modeling, wavelength division multiplexing (WDM) system.**

Received 20 March 2026; revised 18 May 2026; accepted 28 May 2026. Date of publication 2 June 2026; date of current version 3 August 2026. This work was supported in part by the National Key R&D Program of China under Grant 2023YFB2905400, in part by the National Natural Science Foundation of China under Grant 62501388, and in part by Shanghai Jiao Tong University 2030 Initiative. *(Corresponding authors: Lilin Yi; Zekun Niu.)*

The authors are with the State Key Laboratory of Photonics and Communications, School of Integrated Circuits (School of Information Science and Electronic Engineering), Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: minghuishi@sjtu.edu.cn; zekunniu@sjtu.edu.cn; xiaojunzhe@sjtu.edu.cn; boyeren_lilv@sjtu.edu.cn; jnh0907@sjtu.edu.cn; cmz1461404680@sjtu.edu.cn; wshu@sjtu.edu.cn; lilinyi@sjtu.edu.cn).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/JLT.2026.3698936.

Digital Object Identifier 10.1109/JLT.2026.3698936

I. Introduction

Driven by the exponential growth of data-intensive applications such as virtual reality, high-definition video, and artificial intelligence, the demand for optical network capacity continues to surge [1], [2]. However, current optical transmission systems are approaching the nonlinear Shannon capacity limit [3], [4]. To further expand capacity, advanced digital signal processing (DSP) techniques including high-order modulation formats [5], [6], [7], probabilistic shaping (PS) [8], [9], [10], and optical fiber nonlinearity compensation algorithms [11], [12], [13], [14] are employed to mitigate signal-to-noise ratio (SNR) degradation. Consequently, the incorporation of these techniques has increased the complexity of system design and optimization. Conventional design methodologies rely on repetitive, labor-intensive experiments that demand substantial technical expertise for performance tuning [15], [16]. This approach not only increases operational costs but also slows the evaluation toward intelligent optical transmission system. To address these challenges, system modeling has emerged as a cost-effective and efficient alternative [17], [18], [19]. By accurately replicating the physical layer's operating status, modeling serves as a reliable tool for performance prediction and parameter optimization, thereby facilitating the automated and intelligent management of optical transmission systems.

Developing an accurate transmission system model is a prerequisite for realizing intelligent system management. Existing modeling approaches are generally categorized into physics-driven and data-driven methods. Physics-driven methods are formulated upon rigorous physical equations and parameters, with examples including the Gaussian Noise (GN) model [20], [21], [22] and the Split-Step Fourier Method (SSFM) [23], [24], [25]. The GN model and its variants treat fiber nonlinearity as Gaussian noise, enabling rapid and precise estimation of the generalized signal-to-noise ratio (GSNR). However, they are incapable of providing detailed signal waveform information. The SSFM is a waveform-level modeling technique that solves the Nonlinear Schrödinger Equation (NLSE) governing optical fiber transmission through the iterative application of linear and nonlinear operators. By capturing full-field signal details, this waveform-level approach is indispensable for the development and optimization of DSP algorithms. However, the accuracy of physics-driven models is highly contingent upon precise physical parameters. In practical experimental systems, factors such as fiber bending, aging, and inaccurate measurement often lead to parameter mismatch, yielding inevitable discrepancies between modeled and experimental performance [26], [27], [28], [29].

Data-driven models—particularly those utilizing neural networks (NNs)—have garnered substantial attention for accurate and rapid waveform-level modeling of optical fiber channels, owing to their superior nonlinear fitting [30] and parallel computing [31] capabilities. By directly learning underlying characteristics from experimental data, these models effectively circumvent errors induced by inaccurate physical parameters. In our previous work, we successfully achieved high-precision channel modeling [32], [33], [34], [35] based on NN and demonstrated its outstanding generalization ability [36]. The majority of current NN-based research is confined to ideal simulation environments [37], [38], [39], [40], [41]. Despite their theoretical potential to directly model physical systems, the transition from simulation to real-world experiments frequently exposes severe performance gaps. As demonstrated in an FNO-based study [42], the modeling accuracy achieved in an experimental setup was inferior to that in the simulated domain. This disparity clearly underscores the difficulties of implementing high-precision data-driven models under real experimental conditions. The fundamental reason for this performance discrepancy lies in the complex coupling of concurrent physical impairments within experimental data, encompassing linear and nonlinear fiber effects, as well as random optoelectronic noise. Moreover, the dynamic, time-varying characteristics of linear effects and random noise fundamentally hinder conventional NNs from accurately characterizing the overall system behavior.

To overcome these limitations, our previous work proposed a data-physics hybrid framework [43] that leverages the synergistic strengths of both paradigms. Within this architecture, a physical model captures time-varying linear effects, circumventing the difficulty for neural networks in modeling such dynamics, while a bidirectional long short-term memory (BiLSTM) network handles nonlinear effects, and a generative adversarial network (GAN) models random noise. Nevertheless, GANs necessitate the adversarial training of a generator and discriminator [44]. This competitive dynamic frequently provokes training instability and convergence issues [45], [46], leading to inconsistent results that limit their practical application in optical transmission system modeling. Diffusion models [47] represent an emerging class of generative frameworks that reconstruct data distributions through a sequential forward noise-injection and reverse denoising process. Crucially, this mechanism bypasses the adversarial training paradigm, thereby eliminating the instability issues inherent to GANs. Furthermore, unlike the single-step generation process of typical GANs, the iterative denoising nature of diffusion models enables a more precise reconstruction of complex data distributions, leading to higher generation accuracy. Diffusion models have demonstrated superior generation fidelity and diversity in fields such as computer vision and image synthesis [48], [49], [50]. However, their application to the modeling of optical transmission systems remains unexplored. Motivated by this, we propose the integration of a diffusion model into the waveform-level modeling of a full C-band, 60 GBaud, 1600 km optical transmission system.

The primary contributions of this work are summarized as follows:

- We introduce a diffusion model into the data-physics hybrid framework for waveform-level optical transmission system modeling. By replacing the unstable adversarial training of GANs with an iterative denoising process, this architectural mitigates convergence issues and endows the model with superior accuracy and generalization.
- The proposed model undergoes comprehensive experimental validation on a full C-band, 60 GBaud, 1600 km optical transmission platform. It demonstrates superior modeling accuracy and generalization capabilities across varying launch powers, transmission distances and modulation formats.
- We demonstrate the practical application of the proposed model in assisting system design and optimization through three specific tasks: transmission performance prediction, modulation format selection, and DSP parameter optimization.

The remainder of this paper is organized as follows: Section II describes the methodology for system design assisted by optical transmission system modeling. Section III details the proposed data-physics modeling architecture, alongside the theoretical principles and training procedures of the diffusion model. Section IV presents the experimental setup and discusses the results. Finally, Section V concludes the paper.

II. Model-Assisted Optical Transmission System Design and Optimization

Currently, the introduction of advanced DSP technologies—such as high-order modulation formats, PS, and fiber nonlinearity compensation—has enhanced the capacity of optical transmission systems. However, this progress simultaneously escalates the complexity of system design and optimization. Traditional design paradigms, which heavily rely on manual tuning and physical experiments, struggle to meet the demanding requirements of modern complex networks. In this context, waveform-level modeling of optical transmission systems emerges as a powerful tool. By generating detailed signal waveform information, it facilitates critical tasks including transmission performance prediction, system configuration selection, and DSP parameter optimization. Consequently, this modeling technology provides a viable solution to improve design efficiency of next-generation optical networks. The proposed model-assisted optical transmission system design framework is illustrated in Fig. 1, which comprises three hierarchical layers from bottom to top: the physical layer, the model layer, and the application layer.

The physical layer establishes the hardware and software foundation of the optical transmission system, encompassing optoelectronic transceivers, optical fibers, amplifiers, wavelength selective switches (WSS) and DSP algorithms. This layer is responsible for actual optical signal transmission and supplies the digital signals acquired by transceivers as training data for NN models. Moreover, the receiver (Rx) DSP algorithms operate based on physical principles, enabling the adaptive estimation

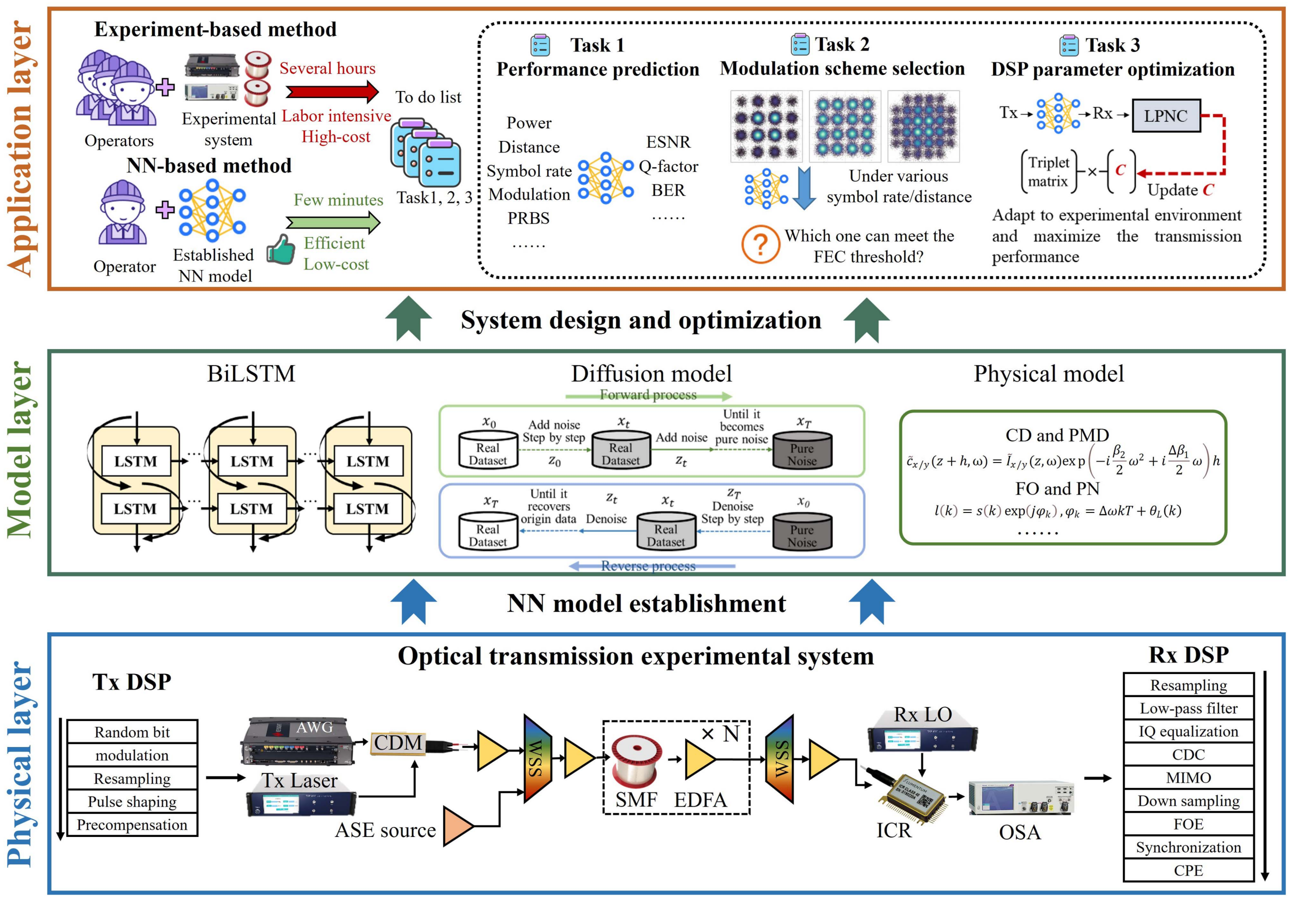

Fig. 1. Schematic diagram of the model-assisted optical transmission system design and optimization.

and compensation of system impairments, such as frequency offset (FO), phase noise (PN), and chromatic dispersion (CD). By utilizing these estimated physical parameters and performing inverse physical modeling, the linear effects of the system can be effectively characterized.

The model layer constructs the optical transmission system model by integrating the collected data and physical parameters derived from the physical layer. To leverage the complementary strengths of both paradigms, this layer employs a hybrid framework that combines analytical physical models with NNs. Specifically, physical models provide interpretability and explicit characterization of linear effects, whereas NNs utilize their robust nonlinear fitting and distribution generating capabilities to model complex nonlinearities and random noise. This data-physics hybrid approach effectively overcomes the accuracy limitations frequently encountered by purely data-driven models when applied to experimental environments. Comprehensive details regarding the model implementation are presented in Section III.

The application layer utilizes the high-precision and generalizable models developed in the preceding layer to establish an efficient, cost-effective framework for system design and optimization. By leveraging the waveform data generated by the model layer to accurately mirror the characteristics of the physical system, this top layer enables the rapid evaluation of transmission performance under varying operational conditions. Furthermore, it facilitates optimal modulation format selection and the precise optimization of DSP parameters. This model-assisted approach circumvents the high costs and extended development cycles associated with traditional experimental design methodologies, thereby providing a practical pathway toward the automated design and intelligent operation of next-generation optical networks.

III. Data-Physics Hybrid Framework for Optical Transmission System Modeling

The prerequisite for effective model-assisted system design lies in developing accurate and generalizable models. In practical optical fiber transmission systems, signals suffer from a complex interplay of linear effects, nonlinear effects, and random noise. The intricate coupling and time-varying nature of these impairments make it highly challenging for purely data-driven NNs to directly maintain modeling accuracy in experimental environments. To address this challenge, we propose a data-physics hybrid modeling framework, as illustrated in Fig. 2. This architecture decouples system impairments to leverage the complementary strengths of physical principles and NNs. Grounded in established physical models, the physical model explicitly accounts for linear impairments such as FO, PN and

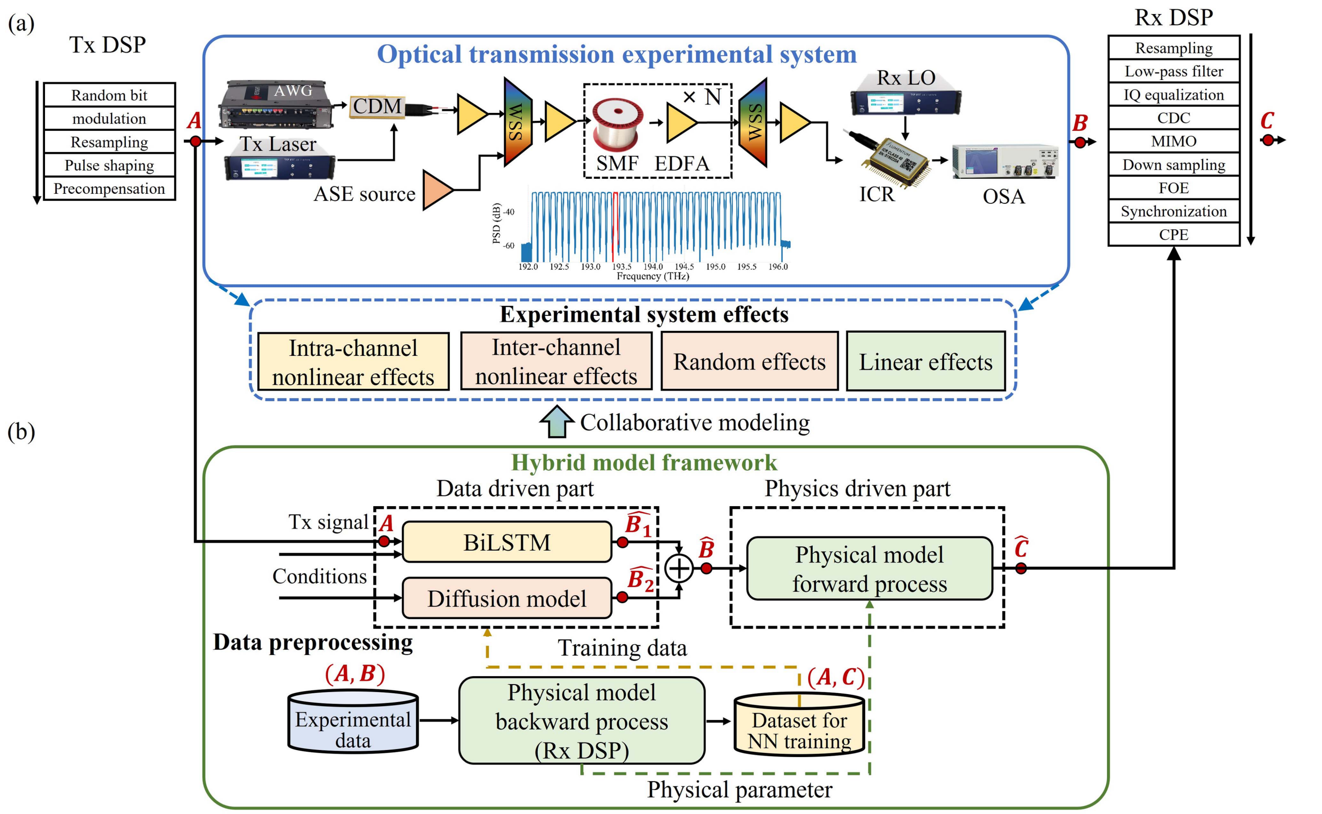

Fig. 2. Schematic diagram of data-physics hybrid model for optical transmission system modeling. (a) Optical transmission experimental system. (b) Data-physics hybrid model framework.

CD. This physical component consists of a backward process and a forward process. The backward process, implemented via Rx DSP, isolates linear effects to extract essential physical parameters and provide a cleaner dataset for NN training. Conversely, the forward process reconstructs these isolated linear effects during inference. By analytically managing the linear components, this approach alleviates the representation burden on the NNs. The NN component employs a BiLSTM network to characterize deterministic intra-channel nonlinearities, while leveraging a diffusion model to capture residual inter-channel nonlinearities and random noise. Operationally, the transmitted (Tx) signal is initially processed by the NN component, and the intermediate output is subsequently passed through the forward physical model to reconstruct the overall waveform. This hybrid approach enables accurate, robust, and interpretable waveform-level modeling. Detailed implementations of both the physical and NN components are presented in the subsequent subsections.

A. Physical model

The physical model addresses linear effects through two distinct procedures: a backward compensation process and a forward modeling process. As illustrated in Fig. 2, the backward process employs the Rx DSP algorithm to generate the training dataset for the NNs and extract physical parameters. The input to the Rx DSP consists of the signals at point C, which have propagated through the experimental transmission system and have been acquired using an oscilloscope. Initially, the Rx signal is resampled to two samples per symbol, subjected to low-pass filtering, and corrected for in-phase and quadrature imbalance. Subsequently, chromatic dispersion compensation (CDC) and a multiple-input multiple-output (MIMO) equalizer are applied to mitigate CD and polarization mode dispersion (PMD), respectively. The signal is then downsampled to one sample per symbol, followed by frequency offset estimation (FOE), synchronization, and carrier phase recovery (CPR). The output of the CPR module serves as the target signal for training the NNs, forming the training data pair consisting of the Tx signal at point A and the processed Rx signal at point B. The Rx DSP isolates time-varying effects, such as PMD and PN, from the raw Rx signals. This isolation yields a dataset that primarily contains deterministic nonlinearities and residual random distortions, free from time-varying linear effects. Consequently, this refined dataset enables more stable and accurate NN training.

Conversely, the forward modeling process reconstructs these compensated linear effects to enable complete system modeling, as depicted in Fig. 2(b). Functionally, the forward process reverses the sequence of the Rx DSP chain. Its input is the predicted Rx signal at point $\hat{B}$, which represents the NN output after the modeling of nonlinear and random effects. The physical model initially applies the laser-induced PN and FO. Following this step, the linear interferences induced by the optical fiber channel, specifically PMD and CD, are modeled. During the prior backward compensation process, PMD is mitigated using

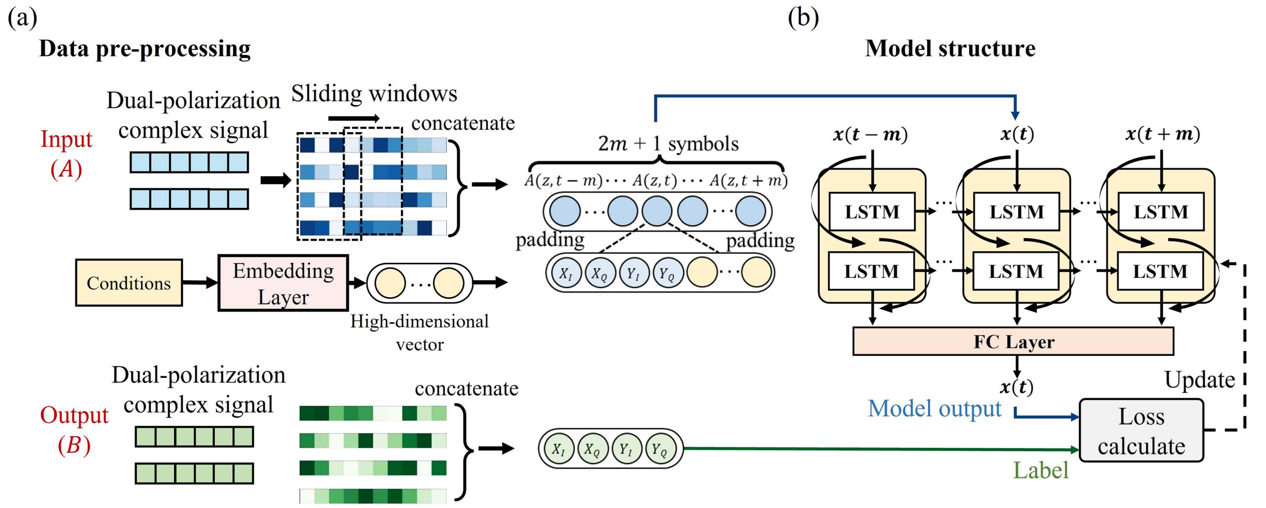

Fig. 3. Schematic diagram of the BiLSTM. (a) Data pre-processing. (b) The model structure.

a MIMO filter; however, the parameters of this filter lack an explicit physical interpretation. Therefore, in the forward process, PMD is mathematically modeled using the Manakov-PMD equations. The numerical simulation is executed via the SSFM, which is widely recognized as a highly accurate technique for linear channel feature modeling. In the Manakov-PMD equation, the PMD effect is characterized by the accumulated differential group delay (DGD) of the link, which can be estimated from the converged coefficients of the MIMO equalizer [51]. In addition, due to the decoupled design of the physical model and the NN, other advanced PMD estimation or modeling methods can also be readily incorporated into the proposed framework to further improve the modeling accuracy. Based on this sequence of operations, the physical model successfully reconstructs the cumulative linear effects experienced during experimental transmission.

B. BiLSTM

The BiLSTM is employed to capture deterministic nonlinear effects, as depicted in Fig. 3(b). In optical transmission, nonlinear effects exhibit strong correlations among symbols across different time instances. With its recurrent architecture and specialized memory and forget gates, the BiLSTM is particularly well-suited for modeling such time-dependent features within sequential data. This architecture has been extensively adopted in optical fiber nonlinearity compensation [52], [53] and waveform modeling tasks [34], [37]. Specifically, the network models the nonlinear distortion by establishing a mapping from the clean Tx signal at point A to the distorted Rx signal at point B. In the experimental setup, due to receiver bandwidth limitations, only the signal from a single channel under test (CUT) can be captured at a time, rendering signals from adjacent channels unavailable. Consequently, self-phase modulation (SPM) emerges as the dominant intra-channel nonlinearity and the primary target for modeling. Following the construction of this training dataset, the raw input data undergoes a preprocessing procedure prior to network training, as illustrated in Fig. 3(a). Given that the BiLSTM operates more effectively on real-valued inputs, the original dual-polarization complex signals are decomposed into four real-valued components: X_I, X_Q, Y_I, Y_Q. To enable the BiLSTM to capture inter-symbol correlations, a sliding window strategy is implemented to construct input sequences that incorporate both past and future symbols. Each input window can be represented as $[\boldsymbol{A}(z,t-m), \boldsymbol{A}(z,t), \boldsymbol{A}(z,t+m)]$, where $\boldsymbol{A}(z,t)$ denotes the symbol at distance z and time t, and m is the half-window length. To ensure robust generalization across varying transmission scenarios, the conditions are incorporated as a control parameter. A parameter encoding structure [36] is introduced to process these conditions, enhancing the model's ability to generalize beyond the training distribution. Within this encoding structure, the raw control parameters are initially mapped through an encoding layer to generate a higher-dimensional latent vector. This encoded vector is subsequently concatenated with the signal segment and fed into the BiLSTM for recurrent processing. Following this step, the hidden states extracted from multiple time steps of the BiLSTM are aggregated and concatenated into a single vector. This aggregated vector is then processed by a fully connected (FC) layer to generate the final output, which represents the predicted symbols corrupted by intra-channel nonlinear distortions.

C. Diffusion model

While BiLSTM effectively models SPM, its capacity to capture residual random noise is inherently limited due to the stochastic nature of such features. To address this, we introduce a generative model designed to handle the remaining random noise. These random features primarily stem from noise sources such as ASE and inter-channel nonlinearities such as cross-phase modulation (XPM), both of which introduce random perturbations into signals. The theoretical basis for modeling these effects is rooted in statistical models, like the GN model, which characterizes uncompensated nonlinear effects, especially for XPM accumulation in long-haul transmission systems [20], [43], [52], [54].

The generative model is suited for this task due to its ability to learn the underlying probability distribution of the residual

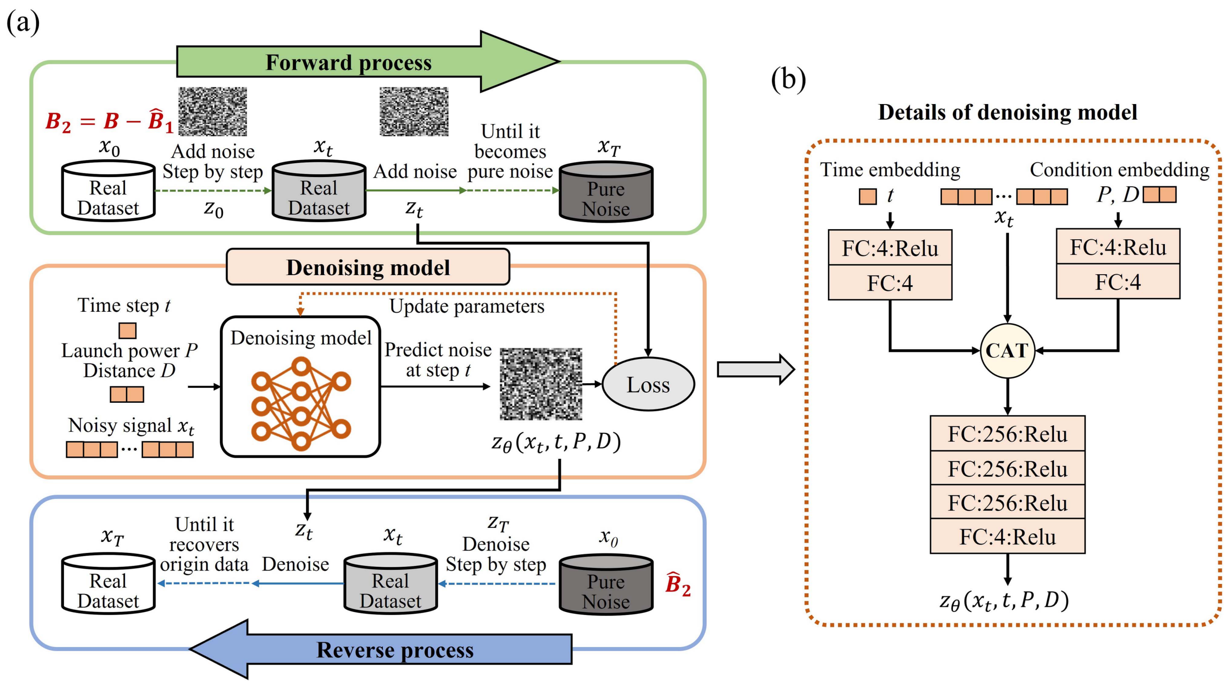

Fig. 4. Schematic diagram of the diffusion model. (a) Details of diffusion model, including three parts: forward diffusion process, denoising model and reverse denoise process. (b) The structure of the denoising model.

noise and generate new samples that replicate the statistical characteristics. The training label for this random component is defined as the discrepancy between the experimentally measured signal at point B and the BiLSTM-predicted output $\hat{B}_1$, formally expressed as $B_2 = B - \hat{B}_1$ The residual B_2 serves as the ground-truth label for training the generative model. Among various generative architecture, GANs and diffusion models are commonly used. GANs rely on an adversarial training paradigm involving a generator and a discriminator that compete to improve sample fidelity. However, this min-max optimization process often suffers from instability and poor reproducibility, which hinders reliable deployment in random noise modeling tasks. To overcome these challenges, we adopt the diffusion model in this study. The Diffusion model generates high-quality samples through a gradual denoising process, avoiding the unstable dynamics of adversarial training and delivering superior generation performance with stable convergence.

The diffusion model is inspired by the physical diffusion process. In machine learning, this concept is implemented via two phases: a forward diffusion process and a reverse denoising process [47], as shown in Fig. 4(a). During the forward process, Gaussian noise is progressively added to the data over T time steps, gradually transforming the original data distribution into a pure Gaussian noise. Given an input sample $\boldsymbol{x_0}$ drawn from true residual noise distribution at point B_2, the forward diffusion process is defined as:

$$q\left(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1}\right) = \mathcal{N}\left(\boldsymbol{x}_t; \sqrt{1-\beta_t}\boldsymbol{x}_{t-1}, \beta_t \mathbf{I}\right), \tag{1}$$

where $\boldsymbol{x_t}$ denotes the noisy samples at step t, $\beta_t \in (0,1)$ is the noise level at step t, determined by a cosine noising schedule, and $\mathcal{N}$ represents a Gaussian distribution with zero mean and unit variance scaled by β_t. As $t \to T$, $\boldsymbol{x_t}$ approaches a standard normal distribution $\mathcal{N}(0, \mathbf{I})$. Instead of sequentially applying (1) for all T steps—which would be computationally expensive—we exploit the additive property of Gaussian variables to directly sample $\boldsymbol{x_t}$ from $\boldsymbol{x_0}$ using:

$$\boldsymbol{x}_t = \sqrt{\bar{\alpha}_t}\boldsymbol{x}_0 + \sqrt{1-\bar{\alpha}_t} z_0 \tag{2}$$

where $\alpha_t = 1 - \beta_t$, $\bar{\alpha}_t = \prod_{i=1}^{t} \alpha_t$, and $\boldsymbol{z}_0 \sim \mathcal{N}(0, \mathbf{I})$. This reparameterization reduces computational complexity while preserving statistical equivalence.

The key idea behind the diffusion model is that if we can reverse this noising chain, we can generate realistic data by starting from pure Gaussian noise $\boldsymbol{x_T} \sim \mathcal{N}(0, \mathbf{I})$ and iteratively denoising it back to $\boldsymbol{x}_0$. Under mild assumptions, the reverse process $p(x_{t-1} \mid x_t)$ also allows a Gaussian distribution, enabling tractable sampling. However, we cannot easily estimate the distribution, since it requires knowing the distribution of all possible data samples to be able to calculate the conditional probability. Therefore, we introduce a learned denoising model parameterized by a NN, to approximate the noise added at each step. Specifically, during the forward process, the true noise at step t is give by:

$$z_{\text{true}} = \frac{\boldsymbol{x}_t - \sqrt{\bar{\alpha}_t}\boldsymbol{x}_0}{\sqrt{1-\bar{\alpha}_t}} \tag{3}$$

which we aim to predict using the denoising model:

$$\hat{z} = z_\theta\left(x_t, t, P, D\right) \tag{4}$$

where $\boldsymbol{z_\theta}$ denotes the NN with parameters θ, taking as input the noisy signal $\boldsymbol{x_t}$, the current time step t, and generalized system conditions (e.g. launch power P, transmission distance D). The predicted noise $\hat{z}$ corresponds to the estimated mean of the forward process noise. Accordingly, the training objective at each timestep t is defined as the expected squared error between the true and predicted noise:

$$L_t = \mathbb{E}_{t\,\mathrm{Unif[T]}}\left[\left\|z_0 - z_\theta\left(x_t, t, P, D\right)\right\|^2\right], \tag{5}$$

whose minimization enables accurate reconstruction of the reverse process. In this work, the denoising model adopts a FC network as its backbone, illustrated in Fig. 4(b). This design is motivated by both the task requirement and the need for a lightweight implementation. From the task perspective, the diffusion model in our framework is used to characterize the residual random distortions after the deterministic nonlinear effects have been modeled by the BiLSTM. Accordingly, the denoising task does not require modeling multi-scale spatial features or long-range temporal dependencies, making a simple FC architecture sufficient for this problem. In addition, from the implementation perspective, a lightweight FC network reduces computational complexity and training cost, while avoiding unnecessary architectural overhead compared with more sophisticated models such as U-Net or Transformer. A time embedding layer preprocesses the timestep t to enhance temporal awareness and enable precise noise prediction across different diffusion steps. Additionally, a condition embedding layer maps system parameters (such as launch power) into a high-dimensional latent space, allowing the model to generalize across varying operating conditions. The complete training procedure is summarized in Algorithm 1. The latter implements the reverse diffusion inference: starting from pure Gaussian noise $\boldsymbol{x_T} \sim \mathcal{N}(0, \mathbf{I})$, the trained denoising model predicts and removes the estimated noise at each step, successively recovering $\boldsymbol{x_{T-1}}, \boldsymbol{x_{T-2}}, \cdots, \boldsymbol{x_0}$. The inference (sampling) process is detailed in Algorithm 2. The final output $\boldsymbol{x_0}$ represents the generated residual noise at point $\hat{B}_2$. Finally, the generated stochastic component is added to the BiLSTM output $\hat{B}_1$ to reconstruct the total nonlinear and random distortion: $\hat{B} = \hat{B}_1 + \hat{B}_2$. Subsequently, the forward physical model reintroduces the compensated linear effect, ultimately generating signals at point $\hat{C}$ that closely matches the experimental received waveform C.

Compared to conventional GANs, the proposed diffusion model exhibits three advantages for modeling random system noise, specifically concerning training stability, modeling precision, and generalization capability:

- *Enhanced training stability via non-adversarial optimization:* The diffusion model structurally circumvents the adversarial training paradigm inherent to GANs [44]. Traditional GANs require the simultaneous optimization of a generator and a discriminator through a min-max game, a fragile process frequently plagued by inherent instability and convergence failures. In contrast, the diffusion model relies on a well-defined, single-objective reverse denoising optimization (5). By eliminating the need for competing networks, this deterministic approach guarantees stable convergence and highly reproducible training outcomes, making it more reliable for characterizing complex physical systems.

Algorithm 1: Training Process.

Require: Target residual noise label x_0 sampling from point B_2, time step t, launch power P, transmission distance D and Gaussian noise $\mathbf{z}_0$
Output: The predicted noise $z_\theta(x_t, t)$
1: **repeat**
2: Sampling x_0
$x_0 \sim q(x_0)$
3: Random sampling t
$t \sim \text{Uniform}(\{1, ..., T\})$
4: Sampling Gaussian noise z_0
$\mathbf{z}_0 \sim \mathcal{N}(0, \mathbf{I})$
5: Calculate the noisy signal x_t at time step t
$x_t = \sqrt{\bar{\alpha}_t} x_0 + \sqrt{1 - \bar{\alpha}_t} \mathbf{z}_0$
6: Take gradient descent step on
$\nabla_\theta \|z_0 - z_\theta(x_t, t, P, D)\|^2$
7: **until** converged

- *Superior modeling precision through iterative refinement:* Unlike the single-step generation mechanism of GANs, the diffusion model employs a multi-step, progressive denoising process to synthesize the noise distribution. This continuous, fine-grained updating mechanism ensures that each reverse step systematically moves toward the high-density regions of the true data distribution. Consequently, this iterative nature allows for the constant correction of localized errors during generation, resulting in higher modeling precision. Conversely, the single-step paradigm of GANs lacks this intermediate error-correction capability, rendering the final output more susceptible to unrecoverable distribution mismatches.
- *Robust generalization driven by objective function differences:* The two generative architectures differ in their underlying optimization goals. As established in (5), the diffusion model minimizes the expected error between the true and predicted noise, which is mathematically equivalent to minimizing the forward Kullback-Leibler (KL) divergence from the empirical data distribution to the model distribution [47]. Conversely, the adversarial objective of GANs typically approximates the reverse KL divergence [55], [56]. Consequently, the diffusion model is incentivized to capture all modes of the data distribution. In contrast, GANs tend to focus on the most prominent high-density regions, potentially neglecting low-probability events. This fundamental theoretical difference endows the diffusion model with superior robustness and generalization capabilities across diverse transmission conditions.

IV. Experiment results and discussion

A. Experiment setup

We evaluate the effectiveness of the diffusion model enabled hybrid model within a coherent polarization-division-multiplexed and wavelength-division-multiplexed (WDM) experiment system, as detailed in Fig. 2(a). At the Tx side, a 120 GSa/s arbitrary waveform generator (AWG) with 50 GHz

Algorithm 2: Inference Process.

Require: Gaussian noise $\mathbf{x}_T$, launch power P, transmission distance D and hyper-parameter α_T
Output: The target signals x_0
1: Sampling $\mathbf{x}_T$ from Gaussian distribution
   $\mathbf{x}_T \sim \mathcal{N}(0, \mathbf{I})$
2: **for** $t = T, \ldots, 1$ **do**
3: Sampling $\mathbf{z}$
   $\mathbf{z} \sim \mathcal{N}(0, \mathbf{I})$ if $t > 1$, else $\mathbf{z} = \mathbf{0}$
4: Remove noise
   $\mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}} z_\theta(\mathbf{x}_t, t, P, D) \right) + \sigma_t \mathbf{z}$
5: **end for**
6: **return**x_0

bandwidth produces 60 GBaud 16-QAM symbols, shaped by squared raised cosine filter with roll-off factor of 0.1. The electrical waveforms are then fed into a 40 GHz bandwidth coherent driver modulator (CDM) to modulate the optical signals and generate the signals for the CUT. Fiber lasers with an approximate 100 kHz linewidth are utilized at both Tx and Rx side. The modulated optical waveform is subsequently amplified by erbium-doped fiber amplifiers (EDFA) with a 5.0 dB noise figure. Other interfering channels are generated from an ASE noise. A wavelength selective switch (WSS) is employed to multiplex these signals into a WDM signal for optical fiber transmission. The fiber link comprises several spans for standard single mode fiber (SSMF), each approximately 80 km in length. An EDFA is deployed at the end of each optical fiber span to compensate for fiber loss. Additionally, WSS is deployed every 8 spans to suppress the cumulative out-of-band ASE noise and to flatten the signal power spectrum. At the Rx side, the CUT signals and local oscillator (LO) are sent to an integrated coherent receiver (ICR) with a 40 GHz bandwidth for coherent detection. The Rx electrical signals are sampled by a 100 Gsa/s digital oscilloscope with 30 GHz electrical bandwidth. Finally, the Rx DSP is employed to compensate for system impairments.

B. Dataset and training setup

To train the proposed hybrid model, experimental data are acquired from the coherent optical transmission system. To mitigate overfitting and reduce pattern prediction by NNs [57], [58], multiple random seeds are employed to generate diverse pseudo-random binary sequences (PRBS), which are then transmitted through the optical fiber link. Furthermore, to ensure robust generalization across various operating conditions, the training data are collected under a wide range of launch power and distance levels. For each configuration, three distinct random seeds are employed to construct the training dataset, while a separate seed is reserved for constructing the test set. Each PRBS sequence is mapped to a 16-QAM constellation, yielding transmitted signals that comprise 153,600 symbols per polarization. The detailed architecture of the hybrid model is summarized in Table I, and the corresponding training hyperparameters are provided in Table II. Specifically, the BiLSTM comprises three hidden layers, with the hidden state dimension set to three times the input feature size to further enhance the nonlinear fitting capacity of the model. For the diffusion model, the forward diffusion process spans $T = 1{,}000$ time steps. To further justify these design choices, additional ablation studies on the diffusion step number T and the model size of the GAN baseline are presented in Appendices B and C, respectively. The noise level β_t at each step scheduled using a cosine annealing strategy over the range [0.0001,0.02]. The denoising network, responsible for executing the reverse diffusion process, adopts a FC architecture consisting of two hidden layers and ReLU activation functions. To evaluate the performance gains introduced by the diffusion model, a GAN is trained as a comparative baseline, employing FC networks for both the generator and discriminator components. During optimization, both the BiLSTM and the denoising network are trained utilizing the Smooth L1 (SL1) loss function. The GAN is optimized using Binary Cross-Entropy (BCE) loss for its adversarial objectives. All models are trained using the Adam optimizer with a batch size of 512 and an initial learning rate (LR) of 1×10^{-3}. Throughout the training process, the LR is gradually decayed following a cosine annealing schedule to ensure convergence stability and optimize the final modeling performance. In the following results presentation, for each experimental condition, five independent measurements are performed and the reported Q-factor is obtained by averaging the corresponding results. Different PRBS sequences are used in different measurement trials to reduce the influence of experimental fluctuations and data-pattern dependence.

C. Performance gain of diffusion model

This section experimentally validates the performance gains achieved by incorporating the diffusion model. Specifically, we demonstrate its superior training stability, higher accuracy, and enhanced generalization capabilities compared to a GAN baseline.

Prior to evaluating the diffusion model's performance, a BiLSTM network is initially trained to characterize the deterministic intra-channel nonlinear effects and generates the training labels required for both the diffusion model and the GAN. Fig. 5(a) illustrates the experimental baseline and BiLSTM modeling outputs: the first column depicts the constellations of the experimental data, the second column presents the BiLSTM-generated constellations, and the third column displays the residual noise obtained by subtracting the BiLSTM output from the experimental signal. It is observed that the BiLSTM successfully reconstructs the deterministic characteristics of the 16-QAM constellation distribution. However, due to its inability to capture random effects and inter-channel nonlinearities, the predicted distribution deviates from the true experimental distribution. This residual noise serves as the target labels for training both the diffusion model and GAN.

With the training labels established, we conduct a comparative evaluation of the diffusion model and GAN under identical conditions—using the same training dataset and configuration—to assess their accuracy and training stability. Five independent training trials are performed for both models. Fig. 5(b) presents the generated noise constellations from these trials: the first

TABLE I
THE SPECIFIC PARAMETER CONFIGURATION OF NN

	BiLSTM	Diffusion model		Denoise model		Generator/Discriminator in GAN
Layers	3	T	1000	Input size	24	104/20
Input size	36	β range	[0.0001, 0.02]	Hidden layer	2	2/2
Hidden neuron	108	β schedule	Cosine	Hidden neuron	[256, 256]	[256, 64]/[256, 64]
Time step	31			Output size	16	16/1
Learned parameters	0.69M			Learned parameters	0.14M	0.16M/0.14M

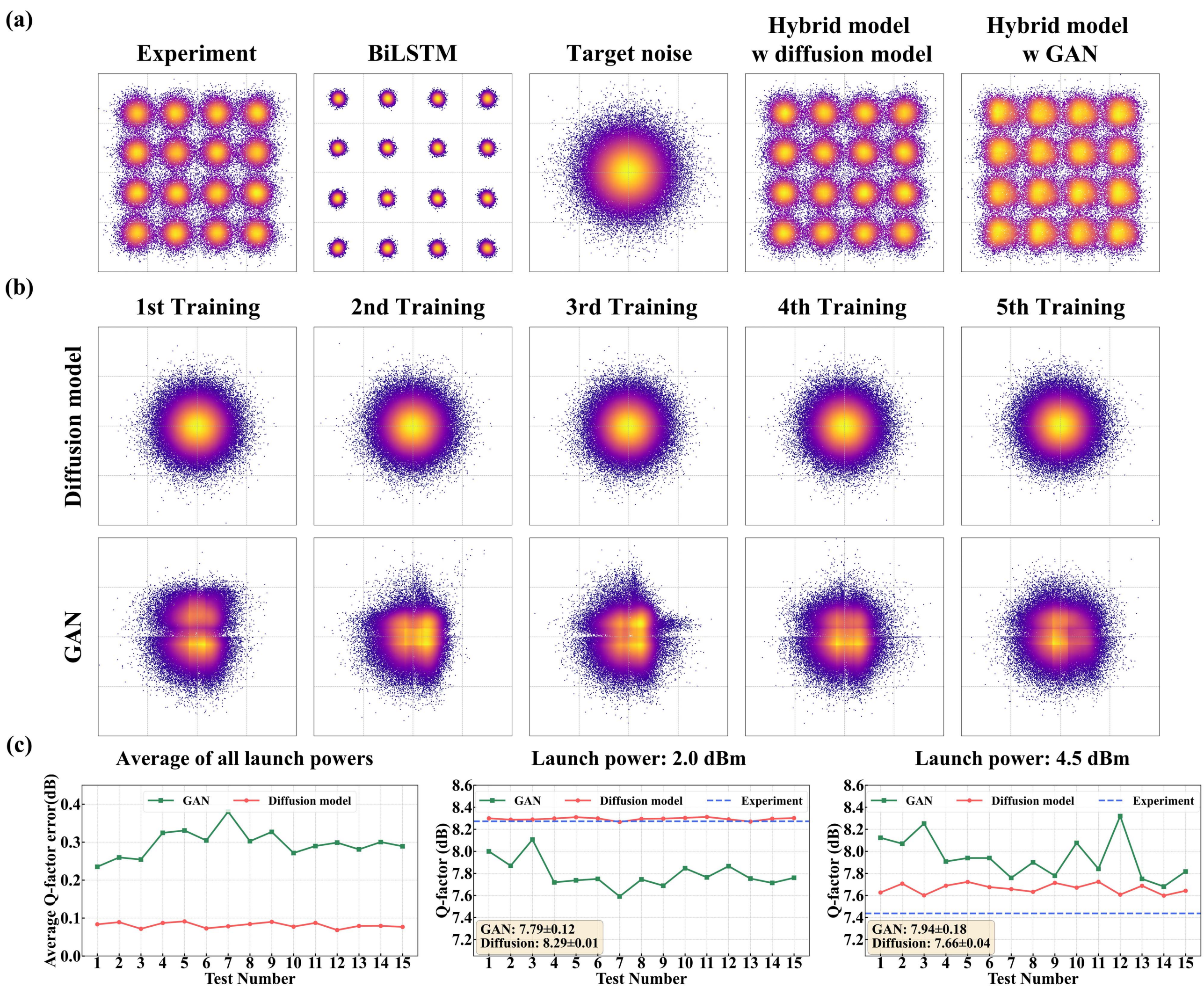

Fig. 5. Constellations and quantitative performance comparison of different models and training runs. (a) Constellations from the experiment, BiLSTM, and the hybrid model. (b) Constellations of the diffusion model and GAN across five independent training trials. (c) Quantitative comparison of the training stability of the diffusion model and GAN over 15 independent runs, including the average Q-factor prediction error over all launch powers, and the Q-factor distributions at launch powers of 2.0 dBm and 4.5 dBm.

row corresponds to the diffusion model, and the second row to the GAN. Notably, the GAN outputs exhibit variation across these trials, indicating poor reproducibility and the inherent instability of the adversarial training. Furthermore, none of the GAN-generated noise patterns accurately match the random characteristics of the true labels, demonstrating its limited capacity to capture the noise distribution of the real physical system. Conversely, the outputs of the diffusion model across all trials are highly consistent and closely mirror the ground-truth noise distribution, thereby confirming its superior training stability and modeling fidelity. To further validate the overall performance of the hybrid model, the final predicted waveform is reconstructed by summing the BiLSTM output and the noise generated by either the GAN or the diffusion model from the

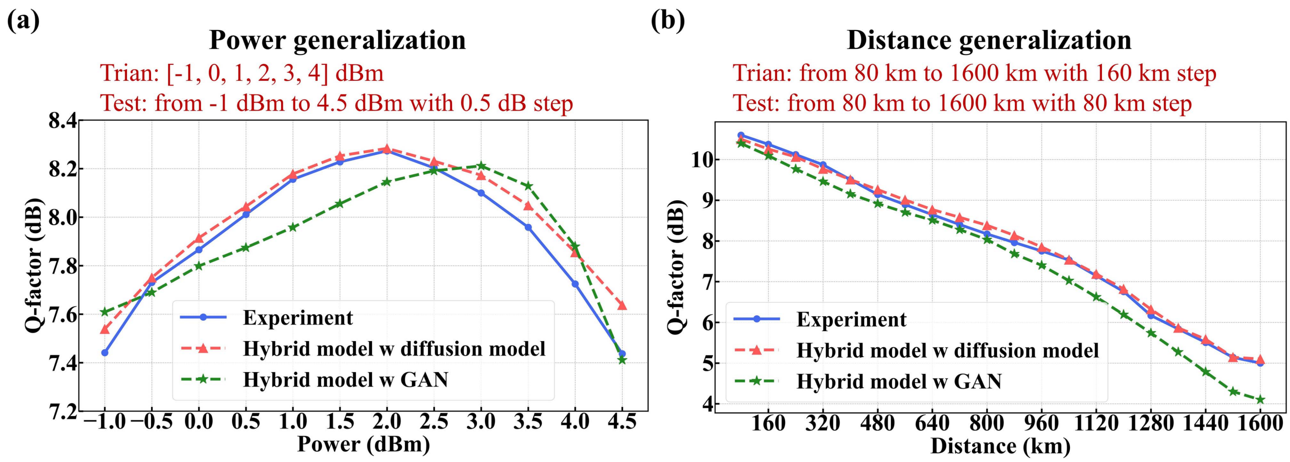

Fig. 6. Generalization performance comparison of hybrid models incorporating the diffusion model and GAN. (a) Q-factor curves versus launch power. (b) Q-factor curves versus transmission distance.

TABLE II
THE SPECIFIC TRAINING HYPERPARAMETERS CONFIGURATION

	BiLSTM	Denoise model	Generator/Discriminator in GAN
Optimizer		Adam	
Batch size		512	
Initial LR		1×10^{-3}	
LR schedule		Cosine Annealing	
Loss function	SL1 loss		BCE loss

fifth training trial. The resulting constellations are shown in the fourth and fifth columns of Fig. 5(a). The output of the hybrid model enabled by the diffusion model yields a constellation that closely matches the experimental measurements. In contrast, the result obtained by the hybrid enabled by the GAN exhibits a pronounced elliptical distortion in the 16 constellation points. This discrepancy primarily arises because the GAN fails to learn an accurate representation of the underlying noise distribution, and the resulting error propagates into the final reconstruction. To further quantify the training stability, we evaluate the Q-factor prediction errors over 15 independent training runs for both the GAN and the diffusion model. As shown in the first column of Fig. 5(c), the average Q-factor prediction error over different launch powers is plotted for each run. The diffusion model consistently maintains an error below 0.1 dB with only small fluctuations, whereas the GAN shows larger variation, with errors ranging from approximately 0.2 dB to 0.4 dB. We further compare the Q-factor distributions at two representative operating points, including the optimal launch power of 2.0 dBm and the high-nonlinearity condition of 4.5 dBm, as shown in the second and third columns of Fig. 5(c), respectively. At 2.0 dBm, the GAN achieves 7.79 $\pm$ 0.12 dB, while the diffusion model reaches 8.29 $\pm$ 0.01 dB. At 4.5 dBm, the corresponding results are 7.94 $\pm$ 0.18 dB for the GAN and 7.66 $\pm$ 0.04 dB for the diffusion model. These quantitative results further confirm that the diffusion model exhibits both lower prediction error and better run-to-run stability than the GAN baseline. By circumventing the unstable adversarial process, the diffusion model learns the data distribution through a iterative denoising process, ultimately achieving highly reliable and precise modeling.

Beyond training stability, the diffusion model demonstrates robust generalization capabilities compared to the GAN—a critical attribute for the design and optimization of optical communication systems. We quantitatively evaluated the generalization performance of both models with respect to launch power and transmission distance. To assess generalization across varying launch powers, a training dataset is constructed using experimental data at power levels of $[-1, 0, 1, 2, 3, 4]$ dBm. To evaluate performance beyond the training distribution, testing is conducted at out-of-distribution power levels of $[-0.5, 0.5, 1.5, 2.5, 3.5, 4.5]$ dBm, spanning both linear and nonlinear operating regimes. Fig. 6(a) compares the Q-factor values obtained from the experimental system against those predicted by the hybrid models incorporating either the diffusion model or the GAN. The results show that the hybrid model with diffusion model aligns closely with experimental measurements, reproducing the characteristic Q-factor trends across both linear and nonlinear regions. The average Q-factor error between the hybrid model with diffusion model and the experiment is merely 0.06 dB, with the maximum error of 0.2 dB occurring at the highly nonlinear 4.5 dBm. In contrast, the hybrid model with GAN exhibits deviations from the experimental results, particularly failing to capture the nonlinear power knee point observed in the experimental measurements. Its average Q-factor error reaches 0.12 dB. This strong consistency of the diffusion model in power generalization lays a solid foundation for its application in parameter optimization for nonlinear compensation (NLC) algorithms.

Subsequently, we evaluate generalization over transmission distance. The training set encompasses distances range from 80 to 1600 km in 160-km intervals. Testing is performed at out-of-distribution distance values across 80 to 1600 km range at finer 80-km intervals. As shown in Fig. 6(b), the hybrid model with diffusion model shows closer agreement with the experimental system, achieving an average Q-factor error of just 0.09 dB and a maximum error of 0.21 dB across the entire transmission distance. Conversely, the hybrid with GAN suffers significant degradation, yielding an average error of 0.41 dB and a maximum error of 0.91 dB. This consistent performance across varying distances supports the development of the hybrid model for system configuration design under diverse transmission scenarios.

The high accuracy and strong generalization of the diffusion model can be attributed to three key factors. First, the consistency observed across independent experimental trials is a direct consequence of its iterative generation framework, which structurally bypasses the inherently unstable min-max adversarial dynamics of GANs. Second, the superior modeling precision is achieved through a multi-step, iterative denoising process. This progressive generation mechanism enables the fine-grained refinement of localized errors at each step, ensuring the generated noise closely aligns with the target distribution. In contrast, the single-step generation paradigm of GANs inherently lacks any intermediate correction capability. Third, the robust generalization capability of the diffusion model stems from its optimization objective. By minimizing a metric equivalent to the forward KL divergence [47], the model is mathematically incentivized to capture the entire data distribution, demonstrating robust generalization. Conversely, the reverse KL divergence objective of GANs [55], [56] inherently forces a mode-seeking behavior, causing the model to overfit high-density regions while systematically neglecting critical low-probability variations.

D. Comparison with a parameterized SSFM

To further examine whether the proposed hybrid framework provides gains beyond a calibrated physics-only model, we additionally construct a parameterized SSFM baseline. After Rx DSP, the residual impairment mainly consists of device noise, link ASE noise, and nonlinear noise. Accordingly, the parameterized SSFM is calibrated through a three-stage procedure, as illustrated in Fig. 7(a). First, the device noise under the back to back (B2B) condition is identified by matching the OSNR and Rx Q-factor, as shown in Fig. 7(b). Second, the link ASE noise is calibrated by matching the post-transmission OSNR, as shown in Fig. 7(c). Third, the launch power in SSFM is set to the experimentally measured value obtained using a power meter, and the nonlinear-related parameters γ, β_2 and α are optimized by grid search within 0.7$\sim$1.3 times the reference values of standard single-mode fiber [29], using the Rx waveform NMSE as the metric [59], as shown in Fig. 7(d). The optimized γ, β_2 and α are kept identical across all spans. Based on this calibrated baseline, Fig. 8 compares the Q-factor prediction errors of the parameterized SSFM and the proposed hybrid model under different launch powers. The average error of the parameterized SSFM is 0.61 dB, while that of the hybrid model is only 0.07 dB. These results show that, even after parameter calibration, the SSFM still cannot fully reproduce the impairments in the real experimental system. A possible reason is that the parameterized SSFM still suffers from parameter uncertainties, unaccounted signal distortions, and the Gaussian assumption used for device

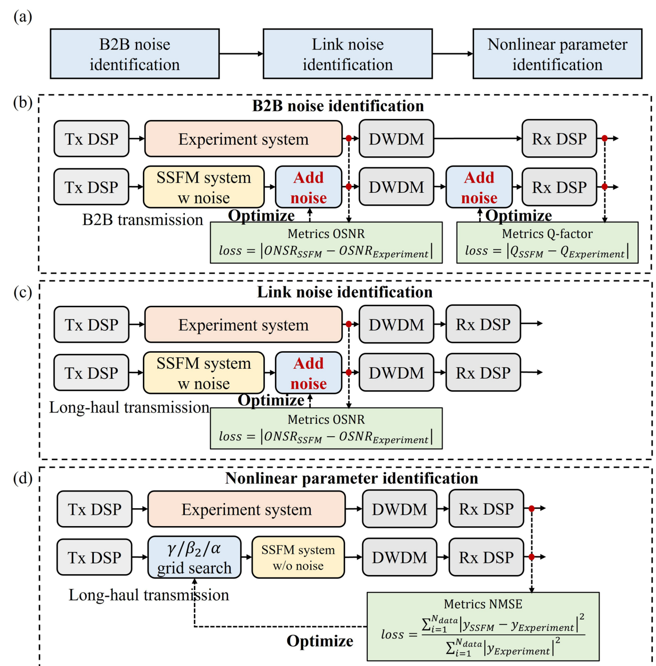

Fig. 7. Parameter calibration procedure for the parameterized SSFM baseline. (a) Overall three-stage framework. (b) Device-noise calibration under the B2B condition using OSNR and Rx Q-factor matching. (c) Link-noise calibration using OSNR matching. (d) Nonlinear parameter identification using Rx waveform NMSE.

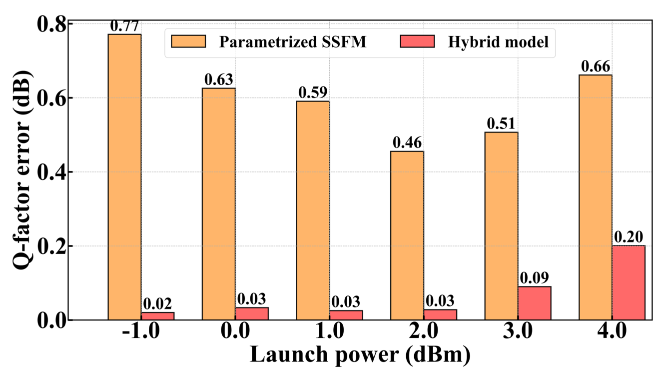

Fig. 8. Q-factor prediction error of the parameterized SSFM and the proposed hybrid model under different launch powers.

noise modeling. In contrast, the diffusion model in the proposed hybrid framework can learn more general stochastic distributions from data, leading to more accurate experimental system modeling.

E. Complexity analysis

To compare the inference efficiency of different modeling approaches, we conduct a computational cost evaluation under a 40-channel, 60-GBaud, 1600-km transmission system, where the launch power varies from -1 to 4.5 dBm with a 0.5-dB interval. All models run on a single RTX 3090 GPU with 24 GB memory. For inference-time evaluation, the total runtime required to process a Tx sequence containing 15,360 symbols per channel per polarization is measured under each tested condition. Table III summarizes the learned parameters,

TABLE III
COMPLEXITY COMPARISON OF SSFM, GAN BASELINES, AND DIFFUSION MODELS WITH AND WITHOUT DPM-SOLVER++ ACCELERATION

	Parameterized SSFM	Small-scale GAN	Large-scale GAN	Diffusion model w/o acceleration	Diffusion model w DPM-solver++
Training time	/	3.3 h	4.1 h	2.9 h	2.9 h
Inference time	27.5 h	1.32 s	3 s	81.6 s	2.52 s
Model structure	/	FC network [104,256,256,16]	FC network [104,1024,1024,16]	FC network [24,256,256,16] T = 1000 during inference	FC network [24,256,256,16] T = 18 during inference
Parameter count	/	0.16M	2.23M	0.14M	0.14M
Average Q-factor error	0.6 dB	0.29 dB	0.27 dB	0.065 dB	0.07 dB

training time, and inference time of the compared models. For the SSFM baseline, the propagation step size is selected according to the nonlinear phase rotation criterion [60], with the maximum nonlinear phase set to 0.005. The SSFM baseline is slower due to its numerical propagation process, requiring 27.5 h to complete all tested conditions. For the diffusion model, the training time is about 2.9 h, which remains relatively low due to the lightweight fully connected backbone. The inference is accelerated by diffusion probabilistic models (DPM)-Solver++ [61], which treats the reverse diffusion process as a differential equation and uses a high-order solver to approximate the reverse trajectory with fewer sampling steps. Under this setting, the diffusion model requires only 18 sampling steps, achieves an average Q-factor prediction error of 0.07 dB, and has an inference time of 2.52 s. Even when the training time is also taken into account, the overall computational cost of the diffusion model remains much lower than that of SSFM. By comparison, the small-scale and large-scale GANs require 3.3 h/1.32 s and 4.1 h/3 s for training/inference, respectively, but both exhibit lower prediction accuracy. These results indicate that the proposed diffusion model achieves a favorable trade-off between modeling accuracy and computational efficiency.

F. Hybrid model enabled modulation scheme selection

In modern optical transmission systems, the implementation of high-order modulation formats increases the information capacity per symbol. PS techniques are frequently introduced to mitigate the impact of ASE noise and fiber nonlinearities on signal quality. However, the adoption of these advanced technologies introduces additional tunable parameters, such as modulation order [5] and the PS degree [8] (e.g., mutual information, MI). As signals propagate over varying transmission distances and symbol rates, they experience diverse system impairments. Consequently, parameter optimization is essential to satisfy forward error correction (FEC) thresholds and guarantee transmission quality. Conventional design approaches rely on iterative experiments prior to system deployment to identify optimal configurations, which inflates both design cost and time. Waveform-level modeling offer a promising alternative by exhibiting exceptional generalization capabilities across modulation formats. A key advantage of this approach is that the hybrid model trained on a single modulation format can directly predict transmission performance under unseen formats, thereby eliminating the need for additional training overhead or experimental costs. This property enables efficient modulation format selection and accelerates the system design cycle.

We first experimentally validate the modulation format generalization capability of the proposed hybrid model. The model is trained exclusively on 16-QAM data. During testing, we evaluate its predictions under quadrature phase shift keying (QPSK), PS-16QAM (MI = 2.7 bits/symbol), 16-QAM, PS-32QAM (MI = 3.5 bits/symbol) and 64-QAM. These test cases encompass modulation schemes with both lower and higher orders than the training set, as well as complex PS constellations. As shown in Fig. 9(a), the first row presents experimental constellations, while the second row displays the corresponding outputs generated by the hybrid model. It is evident that across all five modulation formats, the NN-predicted constellations exhibit high similarity to experimental baseline, particularly in terms of probabilistic characteristics and noise distributions. Furthermore, we quantitatively assess the model's generalization performance across varying MI values, a critical metric for the design of PS schemes. As shown in Fig. 9(b), the Q-factor trends predicted by the hybrid model closely track the experimental measurements across the entire evaluated MI range. The average Q-factor error is a mere 0.11 dB, with the maximum error of 0.21 dB occurring at MI of 3.1 bits/symbol. This marginal error increase is attributed to the larger distributional discrepancy between this test condition and the 16-QAM training baseline. These results demonstrate that the hybrid model achieves excellent modulation-format generalization without necessitating additional training data. This advantage stems from the waveform-level modeling paradigm: because different modulation formats merely alter the features of the input waveforms without changing the underlying physical channel response function, a waveform-level NN—trained on continuous input-output temporal signals—inherently learns the physical mapping of the channel. Consequently, it possesses a robust capability to generalize across varying input waveform features. This property has also been corroborated in prior simulation-based waveform modeling studies.

Building on this capacity, we combine modulation format generalization with transmission distance generalization to demonstrate flexible modulation format selection under diverse system configurations. In this comprehensive evaluation, we test three distinct symbol rates (40, 50, and 60 GBaud), noting that higher baud rates typically induce more severe system impairments. The evaluated transmission distances span from 80 to 1600 km (corresponding to 1 to 20 spans). The hybrid model is

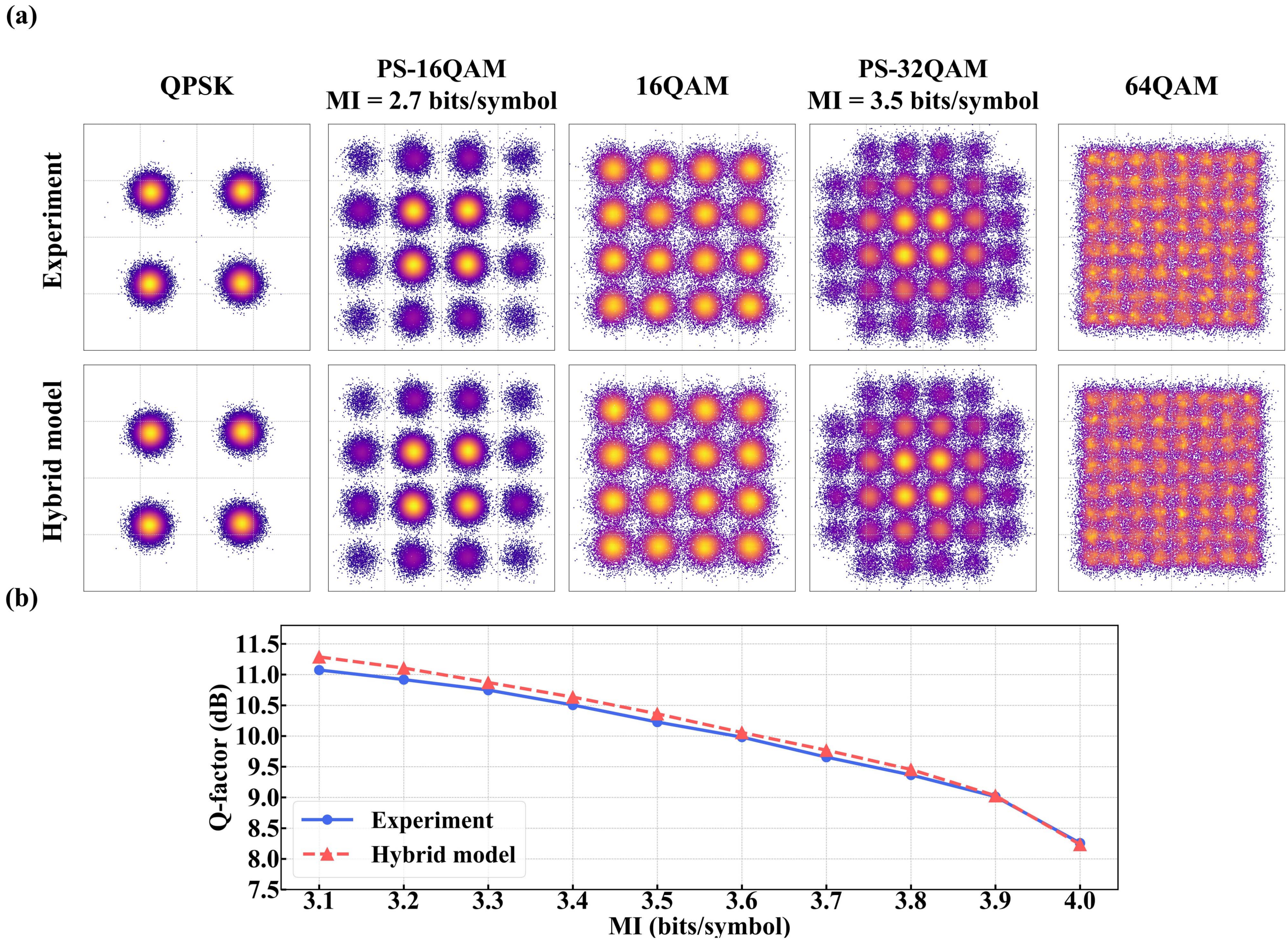

Fig. 9. Experimental validation of modulation format generalization using the proposed hybrid model (trained exclusively on 16-QAM). (a) Comparison of experimentally acquired constellations (top row) and hybrid model-predicted constellations (bottom row) under varying modulation formats. (b) Quantitative Q-factor performance comparison between experimental measurements and hybrid model predictions for PS-16QAM signals across varying MI levels.

trained solely on 16-QAM data, while testing employed out-of-distribution formats including PS-16QAM, 16-QAM, 32-QAM, and 64-QAM. As illustrated in Fig. 10, the performance predictions generated by the NN align tightly with the experimental results across all combinations of symbol rates, transmission distances, and modulation formats, accurately reproducing the consistent degradation trends of the physical system. The black dashed line in the figure denotes the BER threshold of 2×10^{-2} for 20% overhead FEC [62]. By leveraging NN predictions, system designers can rapidly and reliably determine the maximum permissible modulation format for any given rate and distance condition to satisfy stringent FEC requirements. Importantly, utilizing this hybrid model for modulation format selection yields a substantial reduction in experimental overhead. Conventional experimental optimization would require data collection under 218 distinct conditions. As a contrast, the hybrid model enabled approach requires merely 30 training conditions. This translates to an 86.2% reduction in overall data acquisition and experimental effort. Notably, this operational advantage becomes exponentially more pronounced in modern optical networks, where finer-grained modulation format and baud rate selections are routinely required.

G. Hybrid model enabled nonlinear compensation algorithm parameter optimization

Beyond accurate performance prediction, the proposed data-physics hybrid architecture is capable of generating high-fidelity waveform data, which serves as an invaluable resource for the parameter optimization of DSP algorithms. In next-generation optical communication systems, fiber nonlinearity remains a primary bottleneck limiting further capacity scaling. While traditional NLC algorithms are deployed to mitigate these distortions, their effectiveness is hindered by the fact that physical parameters in practical experimental systems are often inaccurate due to inevitable component imperfections and aging. To overcome this limitation, advanced NLC algorithms frequently adopt a learnable paradigm—such as learnable digital back propagation (LDBP) [12] or learnable perturbative nonlinearity compensation (LPNC) [11]—where algorithmic parameters are adaptively optimized based on the inherent characteristics of the experimental data to maximize transmission performance. Traditionally, the development and parameter tuning of such data-driven NLC algorithms rely on extensive physical validations utilizing high-speed optical transmission testbeds. However,

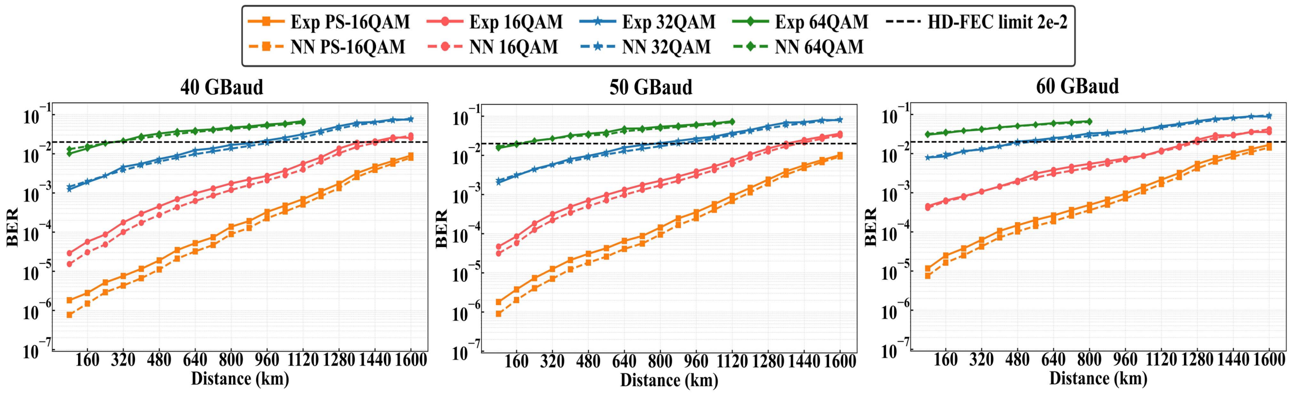

Fig. 10. System performance prediction for modulation format selection under diverse transmission configurations. BER obtained from the experimental system against the hybrid model predictions across varying symbol rates, transmission distances, and modulation formats. The black dashed line indicates the 20% overhead forward error correction (FEC) threshold at a BER of 2×10^{-2}.

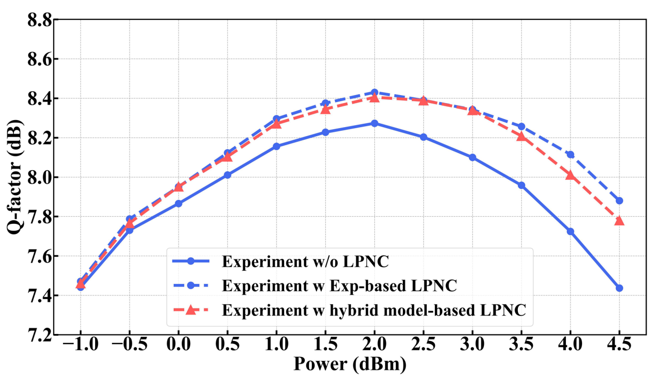

Fig. 11. Variation of the experimental Q-factor with launch power under three algorithm configurations: 1) without LPNC, 2) LPNC based on the experimental system, 3) LPNC based on the proposed hybrid model.

these experimental systems are expensive to operate and demand manual expertise for repetitive data collection and calibration, thereby posing a barrier to iterative research and rapid algorithm deployment. Functioning as a high-accuracy virtual replica of the physical transmission link, our proposed hybrid model offers a solution to lower this barrier by enabling fast and efficient algorithm parameter optimization offline.

To validate the practical utility of the proposed hybrid model for algorithm parameter optimization, we employ the LPNC algorithm as a representative case study. Rooted in perturbation theory, the LPNC algorithm offers relatively low computational complexity and has recently garnered considerable research interest. In our experiment, we independently optimize the LPNC parameters using waveform data generated from two distinct sources: 1) the physical experimental system, and 2) the proposed hybrid model. Subsequently, both optimized parameter sets are applied to the actual experimental system to evaluate their compensation effectiveness. A detailed description of the LPNC coefficient optimization procedure is provided in Appendix A. Fig. 11 presents the Q-factor performance under varying launch powers. The solid line represents the baseline experimental system performance without any LPNC applied. The two dotted lines illustrate the system performance when employing LPNC coefficients optimized by data from the aforementioned two sources. The LPNC performance optimized directly via the physical experimental system serves as the ideal comparison baseline. The results reveal that the average Q-factor error between the LPNC designed via experimental data and the one designed via the hybrid model is a mere 0.032 dB across the entire power sweep. In the linear region (launch power below 1.0 dBm), the discrepancy between the hybrid model-designed LPNC and the experimentally designed LPNC is virtually nonexistent. Even in the highly nonlinear operating regime (launch power at 4.5 dBm), the maximum Q-factor error between the two remains confined to 0.1 dB. These results verify that the hybrid model accurately captures the complex nonlinear characteristics, exhibiting features equivalent to those of the physical experimental system. Consequently, it can be reliably deployed for the data-driven design of NLC algorithms. The slight, 0.1 dB deviation observed in the highly nonlinear region is primarily attributable to the inherent challenge of the BiLSTM attempting to disentangle and learn purely deterministic nonlinear characteristics from noisy experimental data. Overall, this comprehensive demonstration not only confirms the superior accuracy of the hybrid model in emulating empirical systems but also highlights its potential to reduce reliance upon expensive physical testbeds during the algorithm development cycle.

H. Future research directions

Future work may proceed in two directions: sampling acceleration and multi-channel joint modeling. First, the iterative reverse sampling of DDPM is a main source of computational complexity. In this work, this issue is alleviated by adopting DPM-Solver++[61], which reduces the required sampling steps through a high-order approximation of the reverse diffusion dynamics. Further advances in accelerated sampling may improve inference efficiency. Second, the proposed framework may be extended to multi-channel joint modeling for a more physical treatment of inter-channel nonlinear effects such as XPM. A practical challenge is that neighboring-channel waveforms are usually not simultaneously accessible in experiments. If available, multi-channel waveforms can be concatenated as the input to the BiLSTM for joint nonlinear modeling, while the channel center frequency can be introduced as an additional condition in the diffusion model to characterize channel-dependent residual stochastic distortions.

V. Conclusion

In this paper, we propose a diffusion model enabled data-physics hybrid framework for the high-precision waveform-level modeling of optical transmission systems, aimed at facilitating efficient system design and optimization. By leveraging a progressive denoising generation process, the diffusion model effectively circumvents the adversarial training instability and poor generalization inherent to GANs, thereby achieving superior distribution-fitting precision and robustness. The proposed framework is validated on a 41-channel, 60-GBaud, 1600-km full C-band experimental optical transmission platform. Compared with the traditional GAN baseline, the diffusion model not only demonstrates highly reproducible stability across multiple training trials but also exhibits out-of-distribution generalization. Specifically, under conditions of varying launch powers and transmission distances, it achieves average Q-factor prediction errors of merely 0.06 dB and 0.09 dB. These results correspond to error reductions of 50% and 78%, respectively, compared to the GAN baseline. Furthermore, the model demonstrates robust cross-modulation format generalization. Leveraging this high-fidelity modeling capability, we successfully demonstrate its practical utility in rapid system performance prediction, flexible modulation format selection, and the parameter optimization of LPNC. Ultimately, this research substantially lowers the experimental barriers and operational costs associated with physical optical testbeds, providing a highly viable technical pathway to accelerate the automated design and intelligent evolution of next-generation optical networks.

Appendix A

The NLC algorithm employed in this study is the LPNC algorithm [11], a data-driven variant of the conventional PNC method. LPNC enhances adaptability and performance in practical transmission scenarios by enabling the optimization of key model parameters directly from experimental data. As illustrated in Fig. 12, LPNC is implemented after CPR within the Rx DSP chain. The PNC originates from the first-order perturbation of the nonlinear Schrödinger equation. When CD is fully compensated by the Rx DSP, the nonlinear perturbation terms for the X polarization can be represented by:

$$\Delta A_X = \sum_{m,n} P_0^{3/2} \left(A_{n,X} A^*_{m+n} A_{m,X} \right. \\ \left. + A_{n,Y} A^*_{m+n,Y} A_{m,X} \right) C_{m,n}, \tag{6}$$

where $A_{n,X}$ and $A_{n,Y}$ represent the Tx symbols for X and Y polarizations at index n, respectively; P_0 denotes the launch power,

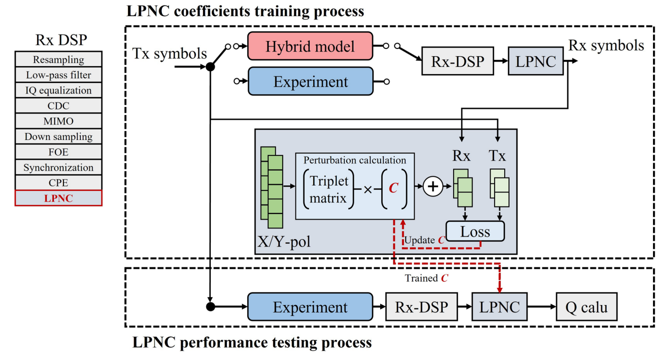

Fig. 12. Placement of the LPNC algorithm within the Rx DSP chain and its parameter optimization process using three methods: the experimental system and the proposed data-physical hybrid model.

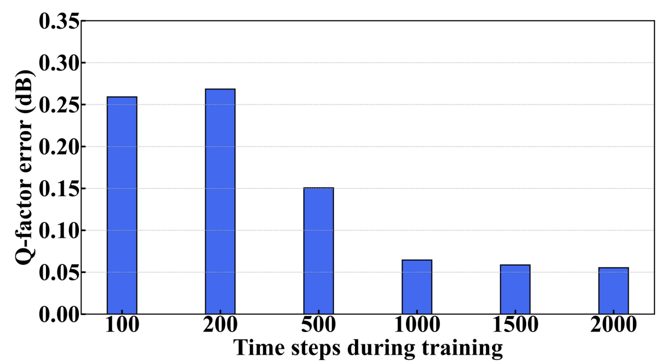

Fig. 13. Ablation study of the diffusion step number T during training process.

and $C_{m,n}$ is the matrix of first-order perturbation coefficients, which can be optimized with respect to the link parameters in LPNC. The corresponding perturbation term for the Y polarization can be derived analogously interchanging the X and Y polarization indices. To compensate for the nonlinear interactions using PNC in the Rx side, it is necessary to pre-determine the products of three symbols. In this paper, we replace the Tx symbols by the Rx symbols after DSP and decision. The coefficient matrix is optimized via machine learning techniques, utilizing a training set of known Tx symbols and the given link parameters. Following the computation of perturbation terms as (7), LPNC effectively removes nonlinear impairments by:

$$\hat{A}_X = \tilde{A}_X - \Delta A_X, \tag{7}$$

where the $\hat{A}_X$ and $\tilde{A}_X$ are PNC output and input symbols, respectively. To verify the accuracy of the grey-box model for fiber nonlinearity modeling and the effectiveness of the LPNC algorithm designed based on it, we optimize the perturbation coefficient matrix $C_{m,n}$ using data from a physical experimental system, the grey-box model, and physics-based SSFM, respectively. Subsequently, these optimized coefficients are applied to the experimental system to measure Q-factor improvement, evaluating optimization effectiveness. The performance of LPNC optimized with the experimental system serves as the baseline for comparison.

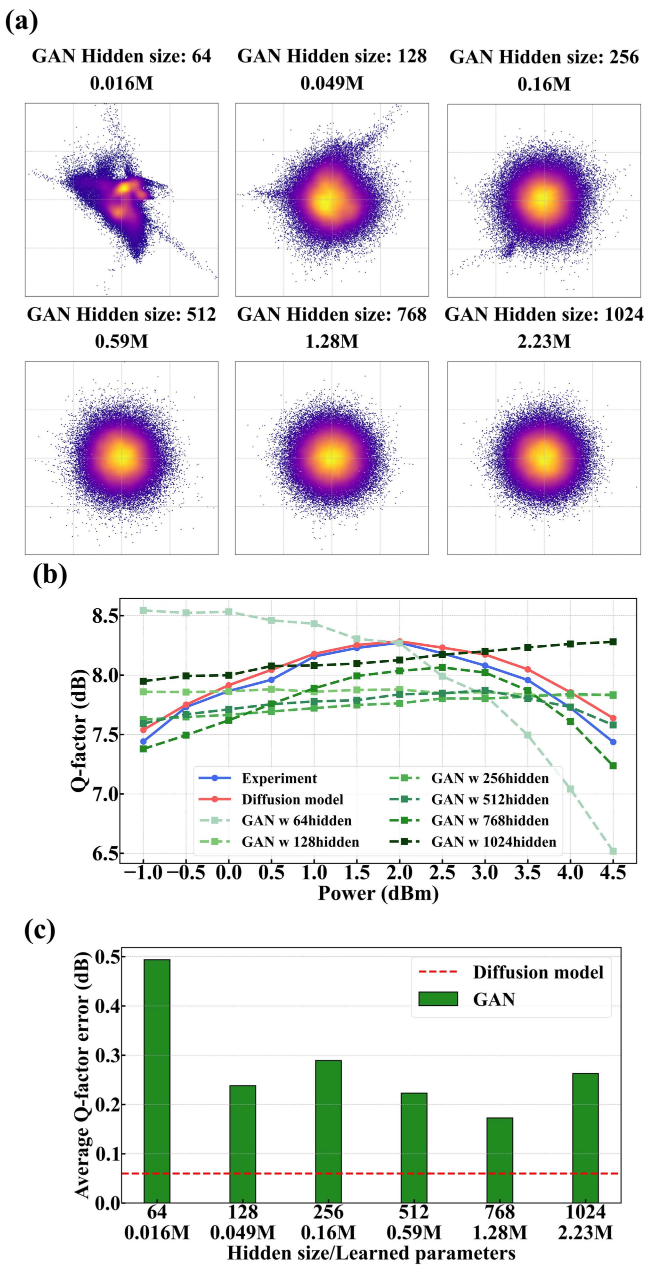

Fig. 14. Comparison between the diffusion model and scaled GAN baselines. (a) Constellations generated by GANs with different hidden sizes, compared with the experiment and the diffusion model. (b) Q-factor versus launch power for the experiment, diffusion model, and GANs with different hidden sizes. (c) Average Q-factor prediction error over all tested launch powers for the diffusion model and GANs with different hidden sizes.

Appendix B

To justify the choice of the diffusion step number T, we further conducted an ablation study under the 40-channel, 1600-km transmission condition with different launch powers. Specifically, we evaluate $T \in 100, 200, 500, 1000, 1500, 2000$ and compared the corresponding Q-factor prediction errors. Here, we would like to clarify that T denotes the number of diffusion steps used in the training process. As shown in Fig. 13, the prediction error is noticeably larger when $T < 1000$, indicating that the diffusion process is insufficient for accurate residual noise modeling in this case. As T increases to 1000, the prediction error is significantly reduced, whereas further increasing T beyond 1000 brings only marginal improvement. Therefore, $T = 1000$ is selected in this work as the diffusion-step setting in training, providing a reasonable trade-off between modeling accuracy and computational cost.

Appendix C

To further examine the fairness of the comparison from the perspective of model complexity, we additionally scale up the GAN baseline by increasing its hidden size from 64 to 1024 and explicitly evaluated the corresponding performance-complexity trade-off. The tested GAN models span a parameter range from 0.016 M to 2.23 M, whereas the diffusion model used in this work contains 0.14 M parameters. As shown in Fig. 14(a), with increasing GAN size, the generated constellations gradually become closer to the experimentally observed noise distribution. However, this improved fitting does not translate into comparable generalization capability. As shown in Fig. 14(b), when evaluated across different launch powers, most GAN configurations fail to accurately reproduce the experimentally observed Q-factor variation trend. Although the GAN with hidden size 512 provides a relatively closer trend, its overall prediction error remains larger than that of the diffusion model. This observation is further quantified in Fig. 14(c), where the diffusion model achieves an average Q-factor error of 0.07 dB, while all tested GAN configurations exhibit higher errors. These results indicate that the advantage of the proposed diffusion model is not simply due to a larger computational budget, but is also related to its stronger ability to learn the residual stochastic distribution and generalize across different operating conditions. A possible reason is that GAN-based adversarial training may place relatively more emphasis on dominant modes of the data distribution [55], [56], whereas diffusion models may be better suited to capturing a broader range of the distribution [47], which could lead to more robust residual noise modeling in the present task.

Acknowledgment

The authors acknowledge the funding provided by the National Key R&D Program of China (2023YFB2905400), National Natural Science Foundation of China (62501388), and Shanghai Jiao Tong University 2030 Initiative.

References

[1] A. Ferrari et al., "Assessment on the achievable throughput of multi-band ITU-T G.652.D fiber transmission systems," *J. Lightw. Technol.*, vol. 38, no. 16, pp. 4279–4291, Aug. 2020.

[2] D. Soma et al., "25-THz O+S+C+L+U-band digital coherent DWDM transmission using a deployed fibre-optic cable," in *Proc. 49th Eur. Conf. Opt. Commun.*, 2023, pp. 1658–1661.

[3] P. P. Mitra and J. B. Stark, "Nonlinear limits to the information capacity of optical fibre communications," *Nature*, vol. 411, no. 6841, pp. 1027–1030, 2001.

[4] R. J. Essiambre and R. W. Tkach, "Capacity trends and limits of optical communication networks," *Proc. IEEE*, vol. 100, no. 5, pp. 1035–1055, May 2012.

[5] V. Bajaj, F. Buchali, M. Chagnon, S. Wahls, and V. Aref, "Single-channel 1.61 Tb/s optical coherent transmission enabled by neural network-based digital pre-distortion," in *Proc. Eur. Conf. Opt. Commun.*, 2020, pp. 1–4.

[6] H. Wang, X. Yi, J. Zhang, and F. Li, "Extended study on equalization-enhanced phase noise for high-order modulation formats," *J. Lightw. Technol.*, vol. 40, no. 24, pp. 7808–7813, Dec. 2022.
[7] H. Huang et al., "Distortion-aware phase retrieval receiver for carrierless high-order QAM transmission," *J. Lightw. Technol.*, vol. 42, no. 13, pp. 4372–4385, Jul. 2024.
[8] F. Buchali, F. Steiner, G. Böcherer, L. Schmalen, P. Schulte, and W. Idler, "Rate adaptation and reach increase by probabilistically shaped 64-QAM: An experimental demonstration," *J. Lightw. Technol.*, vol. 34, no. 7, pp. 1599–1609, Apr. 2016.
[9] T. Fehenberger, A. Alvarado, G. Böcherer, and N. Hanik, "On probabilistic shaping of quadrature amplitude modulation for the nonlinear fiber channel," *J. Lightw. Technol.*, vol. 34, no. 21, pp. 5063–5073, Nov. 2016.
[10] J. Cho and P. J. Winzer, "Probabilistic constellation shaping for optical fiber communications," *J. Lightw. Technol.*, vol. 37, no. 6, pp. 1590–1607, Mar. 2019.
[11] S. Zhang et al., "Field and lab experimental demonstration of nonlinear impairment compensation using neural networks," *Nature Commun.*, vol. 10, no. 1, 2019, Art. no. 3033.
[12] Q. Fan, G. Zhou, T. Gui, C. Lu, and A. P. T. Lau, "Advancing theoretical understanding and practical performance of signal processing for nonlinear optical communications through machine learning," *Nat. Commun.*, vol. 11, no. 1, 2020, Art. no. 3694.
[13] L. Li et al., "Low complexity exponential pruning learned digital backpropagation method for fiber nonlinearity mitigation," *Opt. Lett.*, vol. 50, no. 6, pp. 1893–1896, 2025.
[14] L. Li, Z. Niu, J. Xiao, W. Hu, and L. Yi, "Joint pre- and post-learned perturbation nonlinearity compensation optimization for long-haul optical fiber transmission based on end-to-end deep learning," in *Proc. Opt. Fiber Commun. Conf. Exhib.*, 2025, pp. 1–3.
[15] K. Christodoulopoulos et al., "Toward efficient, reliable, and autonomous optical networks: The orchestra solution [invited]," *IEEE J. Opt. Commun. Netw.*, vol. 11, no. 9, pp. C10–C24, Sep. 2019.
[16] B. Karanov et al., "End-to-end deep learning of optical fiber communications," *J. Lightw. Technol.*, vol. 36, no. 20, pp. 4843–4855, Oct. 2018.
[17] C. Sun, X. Yang, G. Charlet, P. A. Stavrou, and Y. Pointurier, "Digital twin-enabled optical network automation: Power re-optimization," in *Proc. Opt. Fiber Commun. Conf. Exhib.*, 2024, pp. 1–3.
[18] R. Vilalta et al., "Applying digital twins to optical networks with cloud-native SDN controllers," *IEEE Commun. Mag.*, vol. 61, no. 12, pp. 128–134, Dec. 2023.
[19] D. Wang et al., "The role of digital twin in optical communication: Fault management, hardware configuration, and transmission simulation," *IEEE Commun. Mag.*, vol. 59, no. 1, pp. 133–139, Jan. 2021.
[20] P. Poggiolini, "The GN model of non-linear propagation in uncompensated coherent optical systems," *J. Lightw. Technol.*, vol. 30, no. 24, pp. 3857–3879, Dec. 2012.
[21] A. Carena, G. Bosco, V. Curri, Y. Jiang, P. Poggiolini, and F. Forghieri, "EGN model of non-linear fiber propagation," *Opt. Exp.*, vol. 22, no. 13, pp. 16335–16362, Jun. 2014.
[22] D. Semrau, R. I. Killey, and P. Bayvel, "The Gaussian noise model in the presence of inter-channel stimulated raman scattering," *J. Lightw. Technol.*, vol. 36, no. 14, pp. 3046–3055, 2018.
[23] D. Gloge, "Optical fibers for communication," *Appl. Opt.*, vol. 13, no. 2, pp. 249–254, Feb. 1974.
[24] P. Serena, C. Lasagni, S. Musetti, and A. Bononi, "On numerical simulations of ultra-wideband long-haul optical communication systems," *J. Lightw. Technol.*, vol. 38, no. 5, pp. 1019–1031, Mar. 2020.
[25] O. Sinkin, R. Holzlohner, J. Zweck, and C. Menyuk, "Optimization of the split-step Fourier method in modeling optical-fiber communications systems," *J. Lightw. Technol.*, vol. 21, no. 1, pp. 61–68, Jan. 2003.
[26] E. Torrengo et al., "Experimental validation of an analytical model for nonlinear propagation in uncompensated optical links," *Opt. Exp.*, vol. 19, no. 26, pp. B790–B798, 2011.
[27] M. Cantono et al., "On the interplay of nonlinear interference generation with stimulated raman scattering for QoT estimation," *J. Lightw. Technol.*, vol. 36, no. 15, pp. 3131–3141, Aug. 2018.
[28] Q. Zhuge et al., "Application of machine learning in fiber nonlinearity modeling and monitoring for elastic optical networks," *J. Lightw. Technol.*, vol. 37, no. 13, pp. 3055–3063, Jul. 2019.
[29] J. Lu et al., "Performance comparisons between machine learning and analytical models for quality of transmission estimation in wavelength-division-multiplexed systems [invited]," *IEEE J. Opt. Commun. Netw.*, vol. 13, no. 4, pp. B35–B44, Apr. 2021.
[30] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Netw.*, vol. 2, no. 5, pp. 359–366, 1989.
[31] J. Wei et al., "Deploying and scaling distributed parallel deep neural networks on the Tianhe-3 prototype system," *Sci. Rep.*, vol. 11, 2021, Art. no. 20244.
[32] M. Shi et al., "Deep learning waveform channel modeling for wideband optical fiber transmission: Model comparisons, challenges and potential solutions," *J. Lightw. Technol.*, vol. 44, no. 9, pp. 3357–3382, May 2026.
[33] H. Yang et al., "Fast and accurate optical fiber channel modeling using generative adversarial network," *J. Lightw. Technol.*, vol. 39, no. 5, pp. 1322–1333, Mar. 2021.
[34] M. Shi et al., "Fast and accurate waveform modeling based on sequence-to-sequence framework for multi-channel and high-rate optical fiber transmission," *Opt. Lett.*, vol. 50, no. 7, pp. 2286–2289, 2025.
[35] H. Yang, Z. Niu, H. Zhao, S. Xiao, W. Hu, and L. Yi, "Fast and accurate waveform modeling of long-haul multi-channel optical fiber transmission using a hybrid model-data driven scheme," *J. Lightw. Technol.*, vol. 40, no. 14, pp. 4571–4580, Jul. 2022.
[36] M. Shi et al., "Enhancing generalization in neural network-based waveform-level channel modeling for optical fiber transmission through parameter encoding structures," *Opt. Exp.*, vol. 33, no. 19, pp. 40851–40867, 2025.
[37] D. Wang et al., "Data-driven optical fiber channel modeling: A deep learning approach," *J. Lightw. Technol.*, vol. 38, no. 17, pp. 4730–4743, Sep. 2020.
[38] Y. Zang, Z. Yu, K. Xu, M. Chen, S. Yang, and H. Chen, "Multi-span long-haul fiber transmission model based on cascaded neural networks with multi-head attention mechanism," *J. Lightw. Technol.*, vol. 40, no. 19, pp. 6347–6358, Oct. 2022.
[39] X. Zhang, M. Zhang, Y. Song, X. Jiang, F. Zhang, and D. Wang, "Deeponet-based waveform-level simulation for a wideband nonlinear WDM system," *J. Lightw. Technol.*, vol. 41, no. 22, pp. 6908–6922, Nov. 2023.
[40] J. Liu et al., "UVL-based hybrid neural network model for generalized optical fiber channel modeling," *Opt. Exp.*, vol. 33, no. 13, pp. 28233–28242, 2025.
[41] T. Zhang et al., "Fourier neural operator accelerated fast and accurate optical fiber transmission modeling," *J. Lightw. Technol.*, vol. 43, no. 8, pp. 3592–3604, Apr. 2025.
[42] X. He et al., "Fourier neural operator for accurate optical fiber modeling with low complexity," *J. Lightw. Technol.*, vol. 41, no. 8, pp. 2301–2311, Apr. 2023.
[43] H. Yang et al., "The digital twin framework for the physical wideband and long-haul optical fiber communication systems," *Laser Photon. Rev.*, vol. 18, no. 10, 2024, Art. no. 2400234.
[44] I. J. Goodfellow et al., "Generative adversarial networks," 2014, *arXiv:1406.2661*.
[45] D. Saxena et al., "Generative adversarial networks (GANs): Challenges, solutions, and future directions," *ACM Comput. Surv.*, vol. 54, no. 3, pp. 1–42, 2022.
[46] J. Gui, Z. Sun, Y. Wen, D. Tao, and J. Ye, "A review on generative adversarial networks: Algorithms, theory, and applications," 2020, *arXiv:2001.06937*.
[47] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," 2020, *arXiv:2006.11239*.
[48] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 10684–10695.
[49] S. Lim et al., "Score-based generative modeling through stochastic evolution equations in hilbert spaces," in *Proc. Adv. Neural Inf. Process. Syst.*, A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, Eds., 2023, vol. 36, pp. 37799–37812. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2023/file/76c6f9f2475b275b92d03a83ea270af4-Paper-Conference.pdf
[50] P. Dhariwal and A. Nichol, "Diffusion models beat GANs on image synthesis," in *Proc. Adv. Neural Inf. Process. Syst.*, M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, and J. W. Vaughan, Eds., 2021, vol. 34, pp. 8780–8794. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2021/file/49ad23d1ec9fa4bd8d77d02681df5cfa-Paper.pdf

[51] F. N. Hauske et al., "Optical performance monitoring from fir filter coefficients in coherent receivers," in *Proc. Conf. Opt. Fiber Commun./Nat. Fiber Optic Eng. Conf.*, 2008, pp. 1–3.

[52] S. Deligiannidis, A. Bogris, C. Mesaritakis, and Y. Kopsinis, "Compensation of fiber nonlinearities in digital coherent systems leveraging long short-term memory neural networks," *J. Lightw. Technol.*, vol. 38, no. 21, pp. 5991–5999, Nov. 2020.

[53] H. Ming, X. Chen, X. Fang, L. Zhang, C. Li, and F. Zhang, "Ultralow complexity long short-term memory network for fiber nonlinearity mitigation in coherent optical communication systems," *J. Lightw. Technol.*, vol. 40, no. 8, pp. 2427–2434, Apr. 2022.

[54] A. Ferrari et al., "GNPy: An open source application for physical layer aware open optical networks," *IEEE J. Opt. Commun. Netw.*, vol. 12, no. 6, pp. C31–C40, Jun. 2020.

[55] S. Nowozin, B. Cseke, and R. Tomioka, "f-GAN:-Training generative neural samplers using variational divergence minimization," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, vol. 29, pp. 271–279.

[56] I. Goodfellow, "NIPS 2016 Tutorial: Generative adversarial networks," 2017, *arXiv:1701.00160*.

[57] T. A. Eriksson, H. Bülow, and A. Leven, "Applying neural networks in optical communication systems: Possible pitfalls," *IEEE Photon. Technol. Lett.*, vol. 29, no. 23, pp. 2091–2094, Dec. 2017.

[58] L. Yi, T. Liao, L. Huang, L. Xue, P. Li, and W. Hu, "Machine learning for 100 $Gb/s/\lambda$ passive optical network," *J. Lightw. Technol.*, vol. 37, no. 6, pp. 1621–1630, Mar. 2019.

[59] P. d. Koster, O. Schulz, J. Koch, S. Pachnicke, and S. Wahls, "Fast single-mode fiber nonlinearity monitoring: An experimental comparison between split-step and nonlinear Fourier transform-based methods," *IEEE Photon. J.*, vol. 15, no. 6, 2023, Dec. 2023, Art. no. 7202313.

[60] O. V. Sinkin, R. Holzlohner, J. Zweck, and C. R. Menyuk, "Optimization of the split-step fourier method in modeling optical-fiber communications systems," *J. Lightw. Technol.*, vol. 21, no. 1, pp. 61–68, Jan. 2003.

[61] C. Lu, Y. Zhou, F. Bao, J. Chen, C. Li, and J. Zhu, "DPM-Solver : Fast solver for guided sampling of diffusion probabilistic models," *Mach. Intell. Res.*, vol. 22, no. 4, pp. 730–751, 2025.

[62] J. Zhang, X. Li, Y. Hu, M. S. Alam, and D. V. Plant, "Spectrally efficient integrated silicon photonic phase-diverse DD receiver with near-ideal phase response for C-band DWDM transmission," *Laser Photon. Rev.*, vol. 18, no. 3, 2024, Art. no. 2300623.