

# The Digital Twin Framework for the Physical Wideband and Long-Haul Optical Fiber Communication Systems

Hang Yang, Zekun Niu, Qirui Fan, Lyu Li, Minghui Shi, Chuyan Zeng, Shilin Xiao, Weisheng Hu, and Lilin Yi\*

Digital twin (DT) modeling is essential to optical fiber communication systems, particularly for enhancing system performance, controlling the system in real time, and understanding signal nonlinearity. Conventional split-step Fourier method -based simulations, however, struggle with wide-band transmissions, plagued by increasing complexity and inherent biases due to inconsistent link parameter availability. Addressing these challenges, a hybrid data-driven and model-driven DT approach for the wide-band and long-haul physical systems with various system effects is developed. The approach utilizes a neural network (NN) to capture fiber nonlinear features as well as biased perturbations as “lumped” stochastic noises while offloading the linear effects to modules described by physical models of link elements. The model, tested in a 30.5-Tbps 1200 km fiber transmission link with 40 channels, achieves a mean Q factor error of less than 0.1 dB and a maximum runtime of 1.3 s for NN processing under various launch powers, transmission lengths, and optical signal-to-noise ratios. Furthermore, the study has implemented a nonlinear compensation algorithm on the DT model, yielding a consistent enhancement in experimental data. The accuracy and adaptability of the DT model underline its suitability for planning, design, and optimization within the physical optical fiber communication systems.

performance issues, maintaining ongoing monitoring, and facilitating decision-making processes, subsequently optimizing the corresponding physical object.<sup>[4–12]</sup> In the field of optical fiber communication, with the capacity increase, the intricate design of transmission transceivers, lines, amplifiers, and related components becomes increasingly critical. Traditional design approaches, relying on experimental systems for optical transmission, encounter formidable challenges due to their prohibitive costs and the depth of expertise and debugging required, particularly in the contexts of wideband and long-haul transmission systems.<sup>[13–16]</sup> In this case, DT technology is regarded as a crucial tool, offering a cost-effective and efficient simulation alternative that not only aligns with but also leverages the capabilities of existing experimental systems. In addition, due to the channel features provided in real-time by DT, it can be a reliable tool in system control and optimization, promoting intelligent and automated management of the physical optical fiber communication system.

## 1. Introduction

Digital twin (DT) technology, a groundbreaking innovation, creates a virtual model that mirrors a real physical entity or process.<sup>[1–3]</sup> This model, enriched with real-world data, becomes instrumental in conducting simulations, evaluating

However, the development of a suitable DT model for long-haul optical transmission experimental systems, particularly wideband transmission, is an urgent yet unresolved problem. Although the classical split-step Fourier method (SSFM) is a recognized modeling technique in light propagation modeling,<sup>[17,18]</sup> its limitations become apparent in wideband transmission scenarios. The SSFM struggles to scale effectively due to its complexity, which exponentially increases with bandwidth increase. Moreover, the absence of a reliable method to calibrate SSFM under variable link parameters often leads to biased results, underscoring the need for more adaptable and accurate modeling techniques in this evolving field.<sup>[18]</sup> The well-known Gaussian noise (GN) models, offering fast simulations, are considered another candidate technique for DT.<sup>[19,20]</sup> They are used as the quality-of-transmission estimator to calculate the optical and generalized signal-noise ratio (SNR). Nevertheless, GN models are limited in accurate DT for waveform analysis, system optimization, and digital signal processing (DSP) verification, due to the underlying Gaussian noise assumption.

H. Yang, Z. Niu, L. Li, M. Shi, C. Zeng, S. Xiao, W. Hu, L. Yi  
State Key Lab of Advanced Optical Communication Systems and Networks

School of Electronic Information and Electrical Engineering  
Shanghai Jiao Tong University  
Shanghai 200240, China  
E-mail: [lilinyi@sjtu.edu.cn](mailto:lilinyi@sjtu.edu.cn)

Q. Fan  
Huawei Hong Kong Research Center  
Hong Kong, SAR 999077, China

 The ORCID identification number(s) for the author(s) of this article can be found under <https://doi.org/10.1002/lpor.202400234>

DOI: 10.1002/lpor.202400234

Recent advancements in data-driven models, particularly those utilizing neural networks (NN), have shown considerable promise for fast and accurate optical fiber channel modeling through enough training data.<sup>[21–26]</sup> These models, not reliant on system parameters and physical models, offer a novel approach to channel modeling and DT. However, the models are primarily developed based on simulations, posing significant challenges for direct application in experimental systems with complex time-varying random effects. This data-driven model suffers from limited generalization ability, which does not scale well in the real experiment scenario. Even if the NN is fed with a large amount of channel parameters and data, accurately modeling the full spectrum of channel effects remains a challenging task. This difficulty primarily stems from the complexities associated with training large-scale NNs.

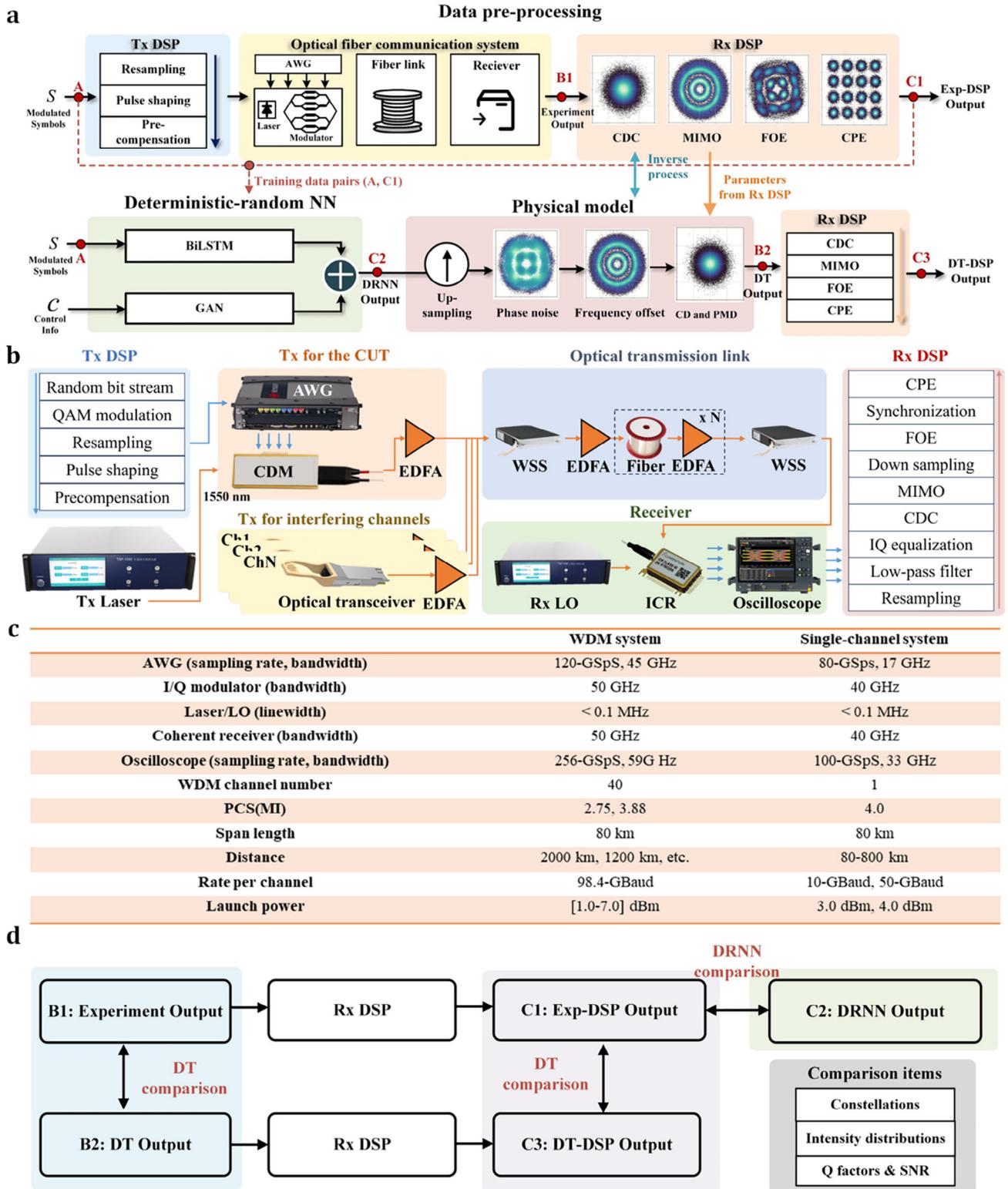
Here, for the first time, we present a DT framework for a physical optical fiber communication system, including three modules: data preprocessing, deterministic-random NN, and physical model. In the data preprocessing, the mature DSP algorithms are utilized to separate data features for assisting NN training and provide physical parameters for feature tracking. Based on the feature simplification, a data-driven approach is proposed to capture fiber nonlinear features as well as the interferences from other bands, which are considered perturbations and modeled as “lumped” stochastic noises. Both deterministic and stochastic features can be obtained simultaneously in a dual-parallelism mode. After the NN modeling, the compensated linear effects by DSP can be described by physical models of link elements. Finally, the complete channel features can be obtained by the hybrid data-driven and model-driven models, which serve as a gray-box DT framework for a physical channel with composable elements. In our experiments, two experimental systems of different transmission conditions are studied, one for wideband long-haul transmission and one for high-nonlinear single-channel transmission, which can determine the wide applicability of the DT framework. A mean Q factor error of less than 0.09 dB is achieved for various conditions, including a 30.5-Tbps 1200 km transmission with 40 channels and 400-Gbps 800 km transmission with a single channel. The proposed NN model exhibits fast modeling with a maximum running time of 1.3 s across various conditions. Compared with the 40-channel SSFM transmission, DT achieves 1200 times complexity reduction. Besides, strong generalization capabilities are demonstrated, including a relatively wide range of launch powers, optical signal-noise ratio (OSNR) levels, and transmission distances. The results highlight the high accuracy, low complexity, and strong generalization ability of the gray-box DT framework. In addition, a nonlinear compensation algorithm is implemented on the DT model and has achieved 0.3 dB Q factor improvement, which is consistent with the experiment. Such a technique opens up new possibilities for nonlinear DSP design and verification. We believe that the proposed DT model provides a reliable physical system model for the development of high-rate and long-haul optical fiber communications. In other fields, this DT framework can also be applied as a universal paradigm for DT studies of other physical systems with complex effects.

## 2. The Digital Twin Framework

The signal propagated through optical fiber communication systems may be susceptible to intricate and diverse system interferences, instigated by various devices and optical fibers. These coupling disturbances include linear and nonlinear effects, along with random and time-varying noise, complicating the task of accurately modeling the entire process. However, advancements in DSP have enabled detailed investigation of certain interference types, especially linear types. This encourages the flexible utilization of mature DSP algorithms to separate system features and track parameters in real optical fiber transmission systems. Leveraging DSP for feature decoupling effectively mitigates time-varying impacts and simplifies channel characteristics. Subsequently, a data-driven methodology can be directly employed to model the real transmission system based on the collected data. The integration of parameter tracking via DSP with the data-driven model ensures the DT model remains congruent with the experimental system’s dynamics. Here, we introduce these technologies briefly and detailed information can be found in the Experimental section.

The proposed DT framework for optical fiber communication systems comprises three core modules: data preprocessing, deterministic-random NN (DRNN), and physical model, as depicted in **Figure 1a**. The DRNN training data is obtained from a comprehensive process combining the transmission system and DSP referred to as data preprocessing. The transmitted data, labeled as point A, consists of 16-quadrature amplitude modulation (QAM) modulated symbols with probability constellation shaping (PCS), which also serve as the NN input. The modulated symbols are first processed by the Tx DSP, which includes resampling, pulse shaping by root-raised cosine filter, and precompensation. Precompensation is used to compensate for frequency bandwidth limits and time delay of high-speed cable skew introduced by the transmitter devices.<sup>[27]</sup> Subsequently, the processed signal enters the optical fiber communication systems, traversing the transmitter, fiber link, and receiver in sequence. Upon receiving the experimental output, marked as point B1, Rx DSP processes the data, involving resampling to two samples per symbol, low-pass filtering, and IQ balancing. Chromatic dispersion compensation (CDC) and multiple-input multiple-output (MIMO) equalizer are then applied, addressing chromatic dispersion (CD) and polarization mode dispersion (PMD), respectively.<sup>[28–30]</sup> The signals are downsampled to one sample per symbol, followed by frequency offset estimation (FOE), synchronization, and carrier phase recovery (CPE). FOE eliminates frequency offset between the laser and local oscillator (LO), while CPE recovers phase noise (PN) caused by the laser and LO.<sup>[31]</sup> The synchronized transmitter and receiver signals generate aligned training data pairs. The final received signal post-DSP, referred to Exp-DSP output and denoted by point C1, forms the training data pairs (A, C1) for DRNN training in the subsequent stage.

The data preprocessing module in our framework mitigates time-varying effects, such as PMD and PN, which pose challenges for NNs processing. After the compensation, the residual characteristics include uncompensated linear and nonlinear effects, along with random noise. These can be categorized into



**Figure 1.** The DT framework of the optical fiber communication system. a) The proposed digital twin framework. The framework has two main parts: deterministic-random neural network (DRNN) and physical model. Rx DSP is used for data processing. CDC, Chromatic dispersion compensation; MIMO, multiple-input multiple-output; FOE, frequency offset estimation; CPE, carrier phase recovery; PN, phase noise; PMD, polarization mode dispersion. b) The real experiment system of optical fiber communications. CUT, channel under test. c) The conditions and signal parameters for experimental systems. d) Schematic diagram of comparison schemes between experiment output and DT output.

deterministic and random features. Despite their nonanalytic and random nature, the robust nonlinear fitting capability of NNs enables data-driven methods to achieve accurate modeling with sufficient training data.

Accordingly, we have designed the DRNN, employing Bidirectional Long Short-Term Memory (BiLSTM) for deterministic features modeling and Generative Adversarial Network (GAN) for random features modeling. Specifically, after linear compensation, the BiLSTM mainly handles deterministic nonlinearity, which is not fully addressed by DSP. This approach is grounded in perturbation theory, where nonlinearities are considered perturbations to the linearly compensated signals.<sup>[32,33]</sup> The BiLSTM excels in this role due to its ability to capture temporal correlations within signal data, enabling it to predict signal behavior over time from both past and future contexts, which has been utilized in fiber nonlinear compensation.<sup>[34]</sup> Here, it primarily models data-related nonlinear effects from noisy signals, including self-phase modulation (SPM) and average nonlinearity induced by random noise and cross-phase modulation (XPM). Regarding the GAN, it models the remaining distortions after DSP and BiLSTM, focusing on random noise. Stochastic features typically arise from noise sources such as amplified spontaneous emission (ASE) and nonlinear interactions like XPM, which introduce random variations into the signal. The underlying theory is based on statistical models, like the Gaussian noise (GN) model, which considers uncompensated effects, especially for XPM under long-haul transmission.<sup>[19,34,35]</sup> GANs are ideal for this application because they can generate new data samples that mimic the distribution of real, observed stochastic disturbances.

Following training with data pairs (A, C1), the DRNN output indicated as C2, encompasses both deterministic and random features in Figure 1a. The DRNN is specifically focused on modeling the nonanalytic and random features after DSP decoupling. However, to establish a complete DT system, it is imperative to also model the compensated features by DSP. These are typically linear effects and can be effectively modeled by physical models of the link elements. As illustrated in Figure 1a, the signal modifications induced by these physical models are essentially the inverse of those applied in the Rx DSP. Consequently, the DT outputs, denoted as B2, are acquired, and the same Rx DSP is executed for signal compensation, facilitating comparison with the Exp-DSP output. This compensated signal termed the DT-DSP output, is marked as point C3. By integrating the DRNN and physical models, and tracking the parameters of post-DSP, the proposed DT model can align with the real transmission system. This alignment is effective across both single-band and wideband transmission scenarios. The model not only achieves high accuracy but also boasts remarkably low time consumption, making it a highly efficient tool in optical fiber communication system analysis.

### 3. Result

#### 3.1. Experiment Setup

Here, we use a coherent polarization-division-multiplexed and wavelength-division-multiplexed (WDM) experiment system to validate the effectiveness of the DT framework, as shown in Figure 1b. At the transmitter side, the signals are transmitted to

the arbitrary waveform generator to generate the analog electrical waveforms. These waveforms are then utilized to drive the IQ modulators for the channel under test (CUT), while the other interfering channels are generated by real optical transceivers. A wavelength selective switch (WSS) is employed to achieve WDM transmission. Subsequently, WDM waveforms are launched into the fiber link that comprises several spans of standard single-mode fiber, each spanning 80 km, along with erbium-doped fiber amplifiers. A WSS is employed after every eight spans to suppress ASE noise accumulation out of the WDM band and adjust the flatness of the WDM signal. At the receiver side, the selected CUT signal and the Rx LO are sent to the integrated coherent receiver. Finally, the received electrical signals are digitized by an oscilloscope to perform offline Rx DSP.

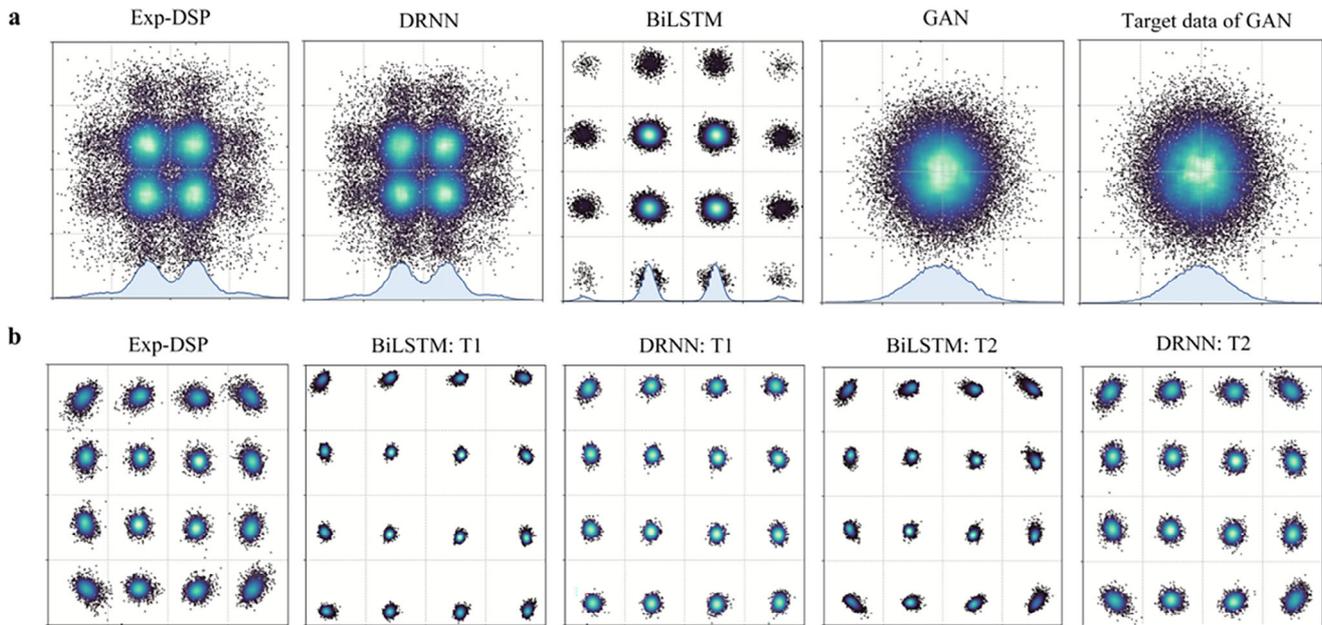
In practice, we use two experimental systems with different device parameters to verify the applicability of the scheme in multiple system scenarios, as shown in Figure 1c. One of the systems has better device conditions, so it is used to transmit WDM high-speed signals. Another experimental system is used for single-channel signals. In the WDM setup, we configured 40 channels, each transmitting at a rate of 98.4-GBaud, with an inter-channel spacing of 100 GHz. The PCS 16 QAM is employed, with the mutual information (MI) of 2.75 and 3.88 for transmission distances of 2000 and 1200 km, respectively. The total transmission rate reaches 21.6-Tbps (40 channels  $\times$  98.4-GBaud  $\times$  2.75  $\times$  2 polarizations) and 30.5-Tbps (40 channels  $\times$  98.4-GBaud  $\times$  3.88  $\times$  2 polarizations). The launch power is varied from 1.0 to 7.0 dBm per channel, encompassing the optimal power point and exhibiting varying degrees of nonlinearity. In the single-channel system, no PCS is performed with 80 to 800 km transmission. The symbol rate is 10-GBaud and 50-GBaud, and the total transmission rate is 80-Gbps (1 channel  $\times$  10-GBaud  $\times$  4.0  $\times$  2 polarizations) and 400-Gbps (1 channel  $\times$  50-GBaud  $\times$  4.0  $\times$  2 polarizations). Corresponding launch powers for these rates were set at 3.0 and 4.0 dBm per channel, respectively.

As shown in Figure 1d, a schematic diagram of the comparison scheme to comprehensively present the DT modeling results. The constellations, intensity distributions, and system performance, including Q factor and SNR, can be presented as comparison items quantitatively. In order to save space and avoid repetition, only X-polarized constellation maps are drawn in the results. In this paper, the signal performance errors are calculated through  $\text{perf}(\text{Exp-DSP}) - \text{perf}(\text{DRNN or DT-DSP})$  for one condition, where perf denotes either the Q factor or SNR. When testing multiple conditional values, we calculate the mean performance error (MPE) using absolute performance errors:

$$\text{MPE} = \mathbb{E} \left\| \text{perf}(\text{Exp-DSP}) - \text{perf}(\text{DRNN or DT-DSP}) \right\| \quad (1)$$

#### 3.2. The Accuracy and Generalization Capabilities of DRNN

The DRNN model, foundational to the DT system, underwent the performance validation of the accuracy and generalization capabilities. We first demonstrate the DRNN output constellations in Figure 2, compared with the CUT results for WDM transmission under specific conditions: MI of 2.75, launch power at 4.0 dBm, and fiber length of 2000 km. The chosen power value is the optimized launch power. Beyond constellations, we plot the intensity distribution of in-phase symbols, depicted in blue histograms



**Figure 2.** The constellations of the Exp-DSP output and DRNN output. a) CUT results of the WDM transmission. WDM, wavelength-division-multiplexed. The blue histogram and envelope represent the intensity distribution of the in-phase symbols. b) High nonlinearity results of the single-channel transmission. T1:30 k training symbols; T2:150 k training symbols.

and envelopes, for a more specific comparison in Figure 2a. The third column illustrates the BiLSTM output, where data distributions align with PCS principles. The fifth column presents the target data for the GAN, derived by subtracting the BiLSTM output from the Exp-DSP output. The variances of the generated data and GAN's target data are 0.44 and 0.439, respectively, with both sets bearing a mean value of 0, showing high consistency. The experimental output and DRNN model exhibit high similarity in both constellations and histograms, where the BiLSTM captures the PCS configuration and the GAN successfully models the random noise of systems.

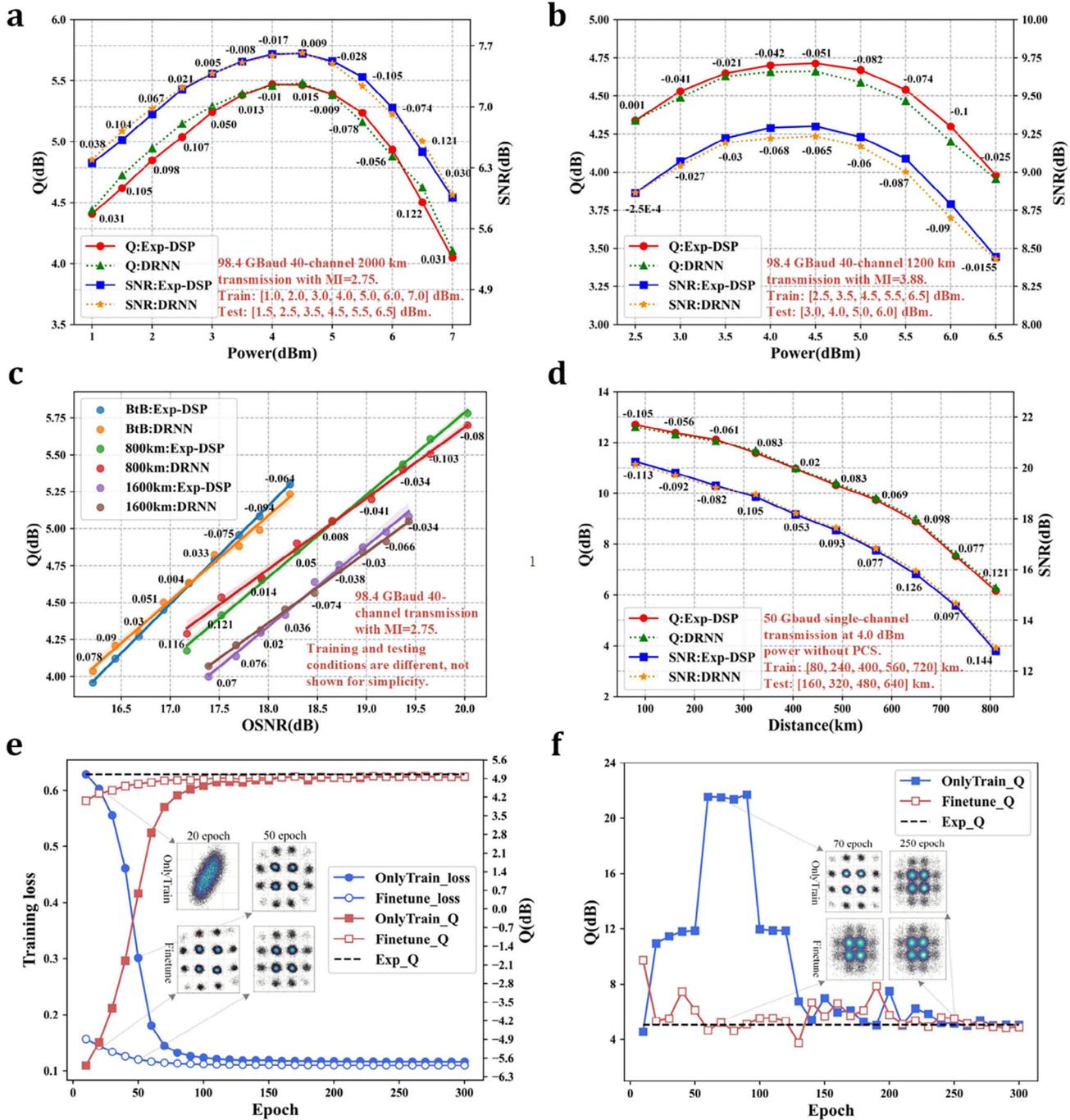
Figure 2b shows the performance at a highly nonlinear channel condition with 3.0 dBm for 400 km single-channel transmission. The first column, showcasing the Exp-DSP output, reveals a nonlinear phase rotation correlated with signal amplitudes. Employing the BiLSTM to capture the nonlinearities, specifically SPM, we observe a significant divergence between the BiLSTM and Exp-DSP outputs with 30k training data. However, increasing the training dataset to 150k enables the BiLSTM to accurately learn nonlinear effects, as manifested in constellations with amplitude-related phase rotations. With enough training, DRNN can model nonlinear characteristics, closely resembling the Exp-DSP output. “30k/150k” data references the size of the training dataset, with each input comprising several consecutive symbols as depicted in Figure 6a.

To quantitatively compare the accuracy and generalization performance of the Exp-DSP and DRNN outputs, we calculate the Q factor and SNR under varying conditions, including launch powers, OSNRs, and transmission distances. The training and testing results are presented in Figure 3a–d. Each figure's results are derived from the same DRNN. The generation of random symbols is facilitated using Matlab's random number generator, applying diverse methods, each with unique seeds, to ensure a

broad range of training scenarios. The methods utilized include: 1) Mersenne Twister. 2) Combined Multiple Recursive. 3) Multiplicative Lagged Fibonacci. 4) Philox  $4 \times 32$  generators with ten rounds. 5) Threefry  $4 \times 64$  generators with 20 rounds. During the training phase, we employ four distinct types, each configured with different seeds, to enrich the model's learning experience. For the testing phase under the same condition selected in training, we opt for a different method, altering the seed to assess the robustness and generalizability of the model.

Figure 3a,b illustrate the generalization performance versus launch power under different transmission settings. The ranges of powers used in the training and testing processes are distinctly different, as illustrated in the figure. These power levels have been specifically chosen to include a broad spectrum of nonlinearity levels, ensuring a comprehensive evaluation of the model's performance across diverse scenarios. This strategic selection of launch powers plays a critical role in accurately capturing the intricate dynamics of nonlinearity within the optical fiber communication system. In summary, to ensure robust generalization, the dataset must encompass a diverse range of scenarios. Similarly, our framework is capable of generalizing to various types of parameters and can extend across different channel numbers by utilizing data collected from various CUTs for training.

In Figure 3a, the model demonstrates impressive accuracy with an average Q-factor error of just 0.041 dB across these power scenarios and an SNR error of 0.037 dB. During testing, the mean error for the Q-factor slightly increases to 0.073 dB, and the SNR to 0.061 dB. Figure 3b reveals a mean Q-factor error of 0.035 dB and an SNR error of 0.040 dB under training conditions. In the testing phase, these errors also exhibit a modest increase, with the mean Q error reaching 0.066 dB and the SNR error 0.061 dB. These results underscore the model's robustness and its ability to maintain high accuracy even under varying conditions. No-



**Figure 3.** Training performance comparisons of the Exp-DSP and DRNN output under different conditions. a–c) WDM transmission. d) Single-channel transmission. e, f) DRNN convergence progress from a single channel with 10G baud 400 km link to WDM channel with 98.4G baud and 2000 km link. (e) and (f) are the convergence progress of BiLSTM, and of GAN, respectively.

tably, the maximum errors in Q factor and SNR across different conditions are confined within 0.122 dB. Compared to the 1-dB error benchmark set by GNpy, an open-source platform,<sup>[35]</sup> the DRNN demonstrates exceptional accuracy and generalization abilities, effectively representing symbols across various MI, link lengths, and power conditions with a single DRNN.

Figure 3c illustrates performance against OSNR under varying conditions: back-to-back (BtB), 800, and 1600 km, with an MI of 2.75 and a launch power of 4.5 dBm. A noise source is integrated before the coherent receiver, allowing adjustable OSNR levels. OSNR values during training range from 16.2 to 20.1 dB, with discrete selection ensuring bit error rates (BER) within the range

of 2.8E-2 to 5.8E-2. During testing, different OSNR values are selected to assess the generalization ability of GAN. The recorded results, shown as dots, indicate a smaller slope for the DRNN curve compared to the Exp-DSP output, attributable to data-driven schemes and dataset settings. NN training performance tends to meet the best at the average level of the training datasets, so there are generally greater errors at both ends of the OSNR value range. The maximum Q error in testing is only 0.121 dB, underscoring the GAN's capability in tracking noise distributions and achieving high accuracy across OSNR levels. Note that even for the conditions in the training dataset, the performances at the fringe conditions are slightly worse than at the central ones. The core of this phenomenon is the distribution of the training data conditions. From a data-driven perspective, the primary objective is to minimize the total dataset loss, leading the NN to focus on reducing loss in the densely sampled middle conditions, which results in slightly poorer performance at the edges. Although this phenomenon exists, the performance at the edges still maintains a low error under 0.1 dB, as shown in Figure 3c. By adjusting the dataset distribution, one can achieve a more balanced performance and enhance modeling accuracy at the boundaries by increasing data density at specific conditions.

The wide-ranging applicability of the DRNN is further demonstrated under single-channel transmission conditions in Figure 3d. In the training condition, the mean error of the Q factor is only 0.066 dB, and SNR is 0.084 dB. The transmission distances under tested have mean Q factor and SNR errors of 0.088 and 0.112 dB, respectively. These results, slightly higher in testing than training conditions, still maintain low error levels, evidencing the model's strong distance generalization capability and potential applicability in single-channel, short-distance scenarios with high-intensity nonlinearity.

To address the variability in transmission conditions and devices, we also explore the finetuning of the model between different scenarios. DRNN's finetuning is performed from a single-channel, 10G baud, 400 km link to a WDM channel, 98.4G baud, 2000 km link scenario. Figure 3e,f presents the BiLSTM and GAN loss during this adaptation process. The NN parameters are typically initialized randomly, a process referred to as Only-Train mode. To accelerate training when system conditions undergo substantial changes, we keep the parameters from the previous system configuration as the initial parameters for the NN. Subsequently, all layers of the NN undergo training without any layers being frozen. This method is designated as Finetune mode. Employing this finetune operation enables the NN model to rapidly adapt to and track changes in system characteristics. For instance, with initial training on 150 thousand symbols for the single-channel case, only 6000 symbols are needed to finetune the model for the WDM scenario. The BiLSTM in Finetune mode converges at 60 epochs, as opposed to 120 epochs in Only-Train mode. The constellation diagrams and Q-factor curves of the modeled signal also exhibit the Finetune mode's fast convergence and enhanced accuracy. Specifically, the GAN shows minimal jitter during Finetune convergence, quickly stabilizing at the accurate Q value, in contrast to significant jitter in the Only-Train mode, as shown in Figure 3f. The adaptability of the proposed DRNN is particularly robust, allowing it to be efficiently fine-tuned to specific channel conditions with minimal learning

overhead. This characteristic is crucial for real-world transmission systems, where channel conditions and photoelectric devices may vary significantly over time or across different locations. Employing the finetuning scheme, the DRNN model can effectively track these changing channel conditions over time, significantly aiding subsequent analysis and optimization efforts.

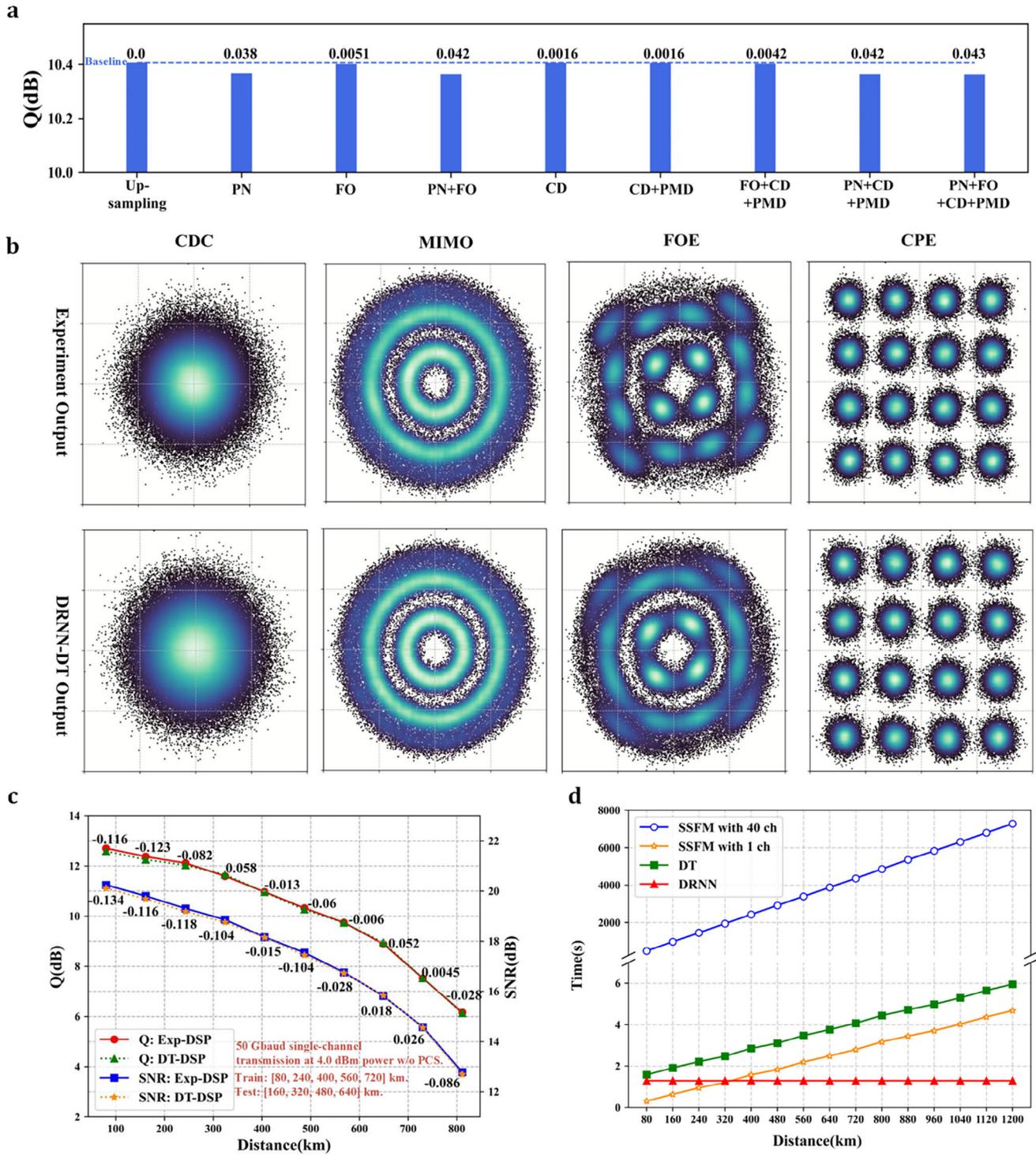
### 3.3. Performance of the Complete DT System with Physical Models

The complete DT system integrates the physical model with an accurate DRNN, ensuring that the modeled signals replicate the channel effects experienced in real-world transmission systems, as captured by the DSP. In evaluating the performance of this complete DT system, our attention centers on both its accuracy and efficiency in time consumption. Given the similarity in features during physical modeling for both WDM and single-channel transmission, we focused our results on the latter, specifically employing a single-channel system with a 50-GBaud rate and 4.0 dBm launch power to avoid redundancy.

Note that equalization-enhanced phase noise (EPPN) arises from the interaction between the local oscillator's laser phase noise and chromatic dispersion equalizers in coherent optical systems. In our system, the EPPN effect is not pronounced, so we have not incorporated this effect into our DT model. However, given its random characteristics,<sup>[36,37]</sup> we can utilize GAN to model this impairment. Additionally, cycle slips in CPE occur due to significant, abrupt changes in the estimated carrier phase, leading to a loss of lock in the phase tracking loops. This phenomenon can severely impair the performance of both DSP and DT models. To mitigate this, DSP employs pilot-aided CPE to reduce the effect of cycle slips. Moreover, during the training process of the DT model, we monitor the phase curve and discard any symbols where the phase curve is discontinuous.

Although the physical models can be considered as the inverse process of the Rx DSP, some modeled features should not be completely the same as the compensated features, such as the PMD and PN. One can extract the channel effects, such as the linewidth of the laser, to achieve flexible configurations. However, these differences may introduce extra performance degradation during the physical modeling process. In order to demonstrate whether the modeling performance will decrease and analyze the impact of different channel effects, we select nine kinds of combinations of different channel effects during the modeling process, denoted by plus signs (e.g., PN+FO signifies the inclusion of only PN and FO effects). After the DSP processing, we can obtain the Q-factor of the DT-DSP output (referred to as  $Q_{DT-DSP}$ ). By comparison with the Q factor of DRNN output ( $Q_{DRNN}$ ) without any channel effects, we obtained the degradation performance of the signal, which can be calculated as  $Q_{DRNN} - Q_{DT-DSP}$ .

Our findings, depicted in Figure 4a, reveal that for combinations including up-sampling, FO, CD, and PMD, the DT-DSP performance closely mirrors the DRNN output with minimal degradation. This suggests these effects can be set flexibly in forward modeling, and can be fully compensated by DSP, not impacting DT performance. However, PN and combinations containing PN slightly deteriorate signal performance, reaching up to 0.043 dB. This slight degradation is attributed to the Wiener



**Figure 4.** Results of the complete DT system. a) Q factor results of the DT-DSP outputs under different channel effects conditions. The baseline is the Q factor of the DRNN output, and the performance differences between DRNN output and DT-DSP output caused by channel effects are listed above the bar chart. b) The constellation changes of the experiment output and DT output during the same Rx DSP process. c) Q and SNR versus distance of the Exp-DSP output and DT-DSP output. d) Running time of DRNN and DT model with 40-channels 98.4 G Baud transmission.

process-based modeling of PN, different from the real data, which CPE cannot fully compensate to the same level. Fortunately, the degree of performance degradation is very small and does not affect the final performance of the DT system. Enhancing the DT model's predictive accuracy may be feasible by extracting the phase noise curve from the Rx DSP. The Rx DSP processes the actual received signal, encompassing all impairments and noises incurred during transmission, including those not perfectly modeled analytically. By analyzing the received signal, a phase curve that more accurately represents the actual state of the signals can be derived. To achieve this, the CPE must be meticulously designed to avoid cycle slip effects. More crucially, an accurate phase noise curve should be extracted, taking into account slow time-varying frequency offsets and EEPN. Investigating how to accurately extract the phase noise from CPE algorithms—and assessing its impact on modeling performance and potentially aiding in reverse-engineering CPE design—is a promising area for detailed research.

We then undertake a comprehensive comparison of the DT output accuracy with complete channel effects, analyzed through constellations and system performance. Both the experimental system and DT system signals are compensated using the same Rx DSP, with resulting constellations from each DSP step displayed in Figure 4b. The similarity in constellation characteristics post-DSP steps, including CDC, MIMO, FOE, and CPE, underscores the DT system's accuracy. The observed differences between the FOE constellations of the experimental output and the DRNN-DT output stem from variations in the PN values. As previously mentioned, the PN values employed in our study are not derived from the experimental DSP; rather, they are modeled using the Wiener process. Quantitatively, the Q factor and SNR of the DT-DSP and Exp-DSP are recorded under varying distances, as shown in Figure 4c. The mean error for the Q factor was a mere 0.054 dB, and for SNR, 0.075 dB. These errors are smaller than those of the DRNN output, which, originally limited in interference modeling, displayed slightly higher performance than the experimental signal. The introduction of PN-induced performance degradation brings the DT interference level closer to the experimental output, reducing the discrepancy with the Exp-DSP output.

Finally, we record the running times of the DRNN, the DT model (combination of DRNN and physical model with a single channel), and the SSFM across different distances for both single channel and 40 channels, as illustrated in Figure 4d, demonstrating that the DT system is not only accurate but also efficient, meeting the basic requirements for practical application. Both the 40-channel transmission and single-channel transmission employ 4.0 dBm signal power, with 131 072 transmitted symbols. The DRNN model's complexity is independent of distance, leading to consistent running times, all below 1.3 s on an RTX3090. The physical model uses the linear model of SSFM. As the distance increases, the modeling time of physical models will gradually increase with the span number. The running time of the whole DT process, depending on the physical model, is roughly equivalent to the sum of the DRNN and SSFM times for a single channel, which at 80 km is  $\approx 1.6$  s. As for the WDM transmission, both the DRNN and physical model have the same structure with the single-channel transmission case, resulting in a significant reduction of running time. For example, compared

to the SSFM in 40-channel 1200 km transmission, DT takes  $\approx 6$  s versus SSFM's 7280 s, achieving a reduction by over 1200 times. As the number of transmitted symbols increases, the acceleration of the DT model is expected to be further amplified, owing to high parallelization on GPU platforms.

Compared to SSFM and GN models, the proposed DT model presents significant advantages. The SSFM simulates waveform propagation in optical fibers, governed by the Nonlinear Schrödinger Equation (NLSE). Although it provides highly accurate solutions to the NLSE, SSFM is not ideal for establishing DTs of experimental systems, primarily due to its complexity. In such instances, while the DT model experiences only a modest increase in time consumption, the SSFM becomes impractical due to its complexity, which scales at least to the fourth power of the bandwidth. Furthermore, SSFM struggles to align closely with real experimental environments due to discrepancies in parameters between the NLSE and SSFM, as well as dynamic effects that are difficult to capture and accurately model using the basic NLSE framework.

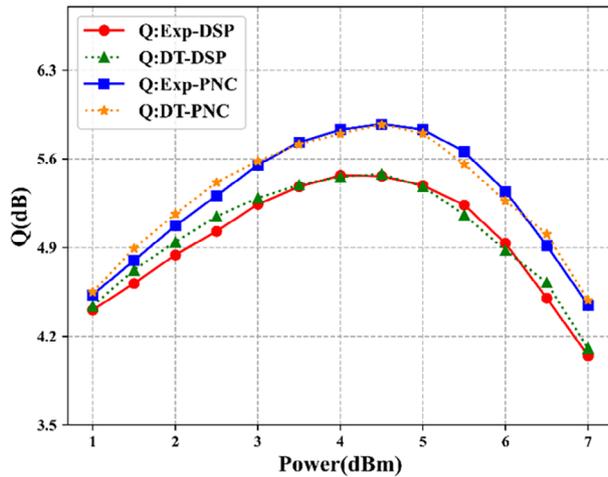
GN models, which focus on statistical characteristic modeling, are employed to predict optical network performance. Due to their reliance on statistical models, GN models do not provide waveform information, which limits their utility in DSP design. Additionally, these models are based on Gaussian assumptions and often require modifications for practical applications. This is particularly true in scenarios characterized by high nonlinearity where signals are non-Gaussian, such as in short-distance, high-power transmissions.

The proposed DT framework, by directly collecting training data from the experimental platform, incorporates all channel and component effects of the real system, establishing a highly accurate DT for the CUT. This framework combines the benefits of both SSFM and GN models, achieving high accuracy in waveform modeling with significantly greater computational efficiency. Moreover, the DT model demonstrates remarkable scalability for more complex scenarios involving multiple channels. As optical communication is reaching its limit of spectrum efficiency, ultra-wide-band transmission is being adopted to further push forward the optical fiber capacity. The proposed DT model's ability to adapt and scale in response to the evolving demands of optical communication positions it as an essential tool for driving future advancements in the field.

### 3.4. Nonlinear DSP Verifications using DT Generated Data

DT model's ability to generate datasets that closely mimic those from experimental systems stands out as a significant advantage, particularly for the validation of DSP algorithms. In addition to the linear compensation algorithms previously mentioned, the DT model also facilitates the design and validation of nonlinear compensation (NLC) algorithms, a crucial aspect of modern optical communication systems.

One notable NLC algorithm we incorporated is the classic perturbation-based nonlinearity compensation (PNC) architecture,<sup>[32]</sup> which is implemented subsequently to linear compensation. The PNC algorithm is specifically designed to counteract the SPM effect, a common nonlinear impairment in optical fibers. A key advantage of this approach is its



**Figure 5.** Nonlinear compensation results of the experiment and DT system. Q:EXP-DSP: Q factor performance after DSP using experiment data; Q:DT-DSP: Q factor performance after DSP using DT data; Q:EXP-PNC: Q factor performance after perturbation nonlinear compensation using experiment data; Q:DT-PNC: Q factor performance after perturbation nonlinear compensation using DT data.

compatibility with the existing generalized DSP framework, where the PNC can be deployed following the CPE algorithm without necessitating any alterations to the established DSP structure. This seamless integration makes the implementation of NLC not only effective but also convenient, allowing for more straightforward adaptation and optimization in real-world optical communication systems.

**Figure 5** provides a detailed comparison of the transmission performance of a system before and after the implementation of PNC, operating at a transmission rate of 98.4G Baud over a 2000 km distance, with power levels varying from 1.0 to 7.0 dBm. The PNC employed features 81 taps, indicating the level of complexity and precision in the compensation process. We design the PNC based on the DT model and then perform the PNC on the experiment data.

At a low power range, the SPM effect is weak, and there is no significant performance improvement in PNC. At a high power range, PBC can achieve a better Q-factor improvement. Based on the DT model, the designed PNC is tested in the experiments. Taking into account all power factors, the experimental data showed an average improvement of 0.33 dB in the Q-factor, and the DT data showed an average improvement of 0.31 dB in the Q-factor. The optimal power changed from 4.0 dBm before PNC compensation to 4.5 dBm after PNC compensation, both in the experiments and DT outputs. The high consistency between the DT results and the experimental data underscores the DT model's accuracy in nonlinear modeling. In addition, with the capacity limited by the nonlinearities, one of the potential applications of DT model can be nonlinear compensation design and verification. The DT model, therefore, not only enhances the accuracy and efficiency of simulating real-world transmission scenarios but also serves as a robust platform for testing and refining advanced DSP algorithms, including those aimed at mitigating nonlinear impairments. This capability is particularly

valuable in the pursuit of higher performance and more reliable optical communication networks.

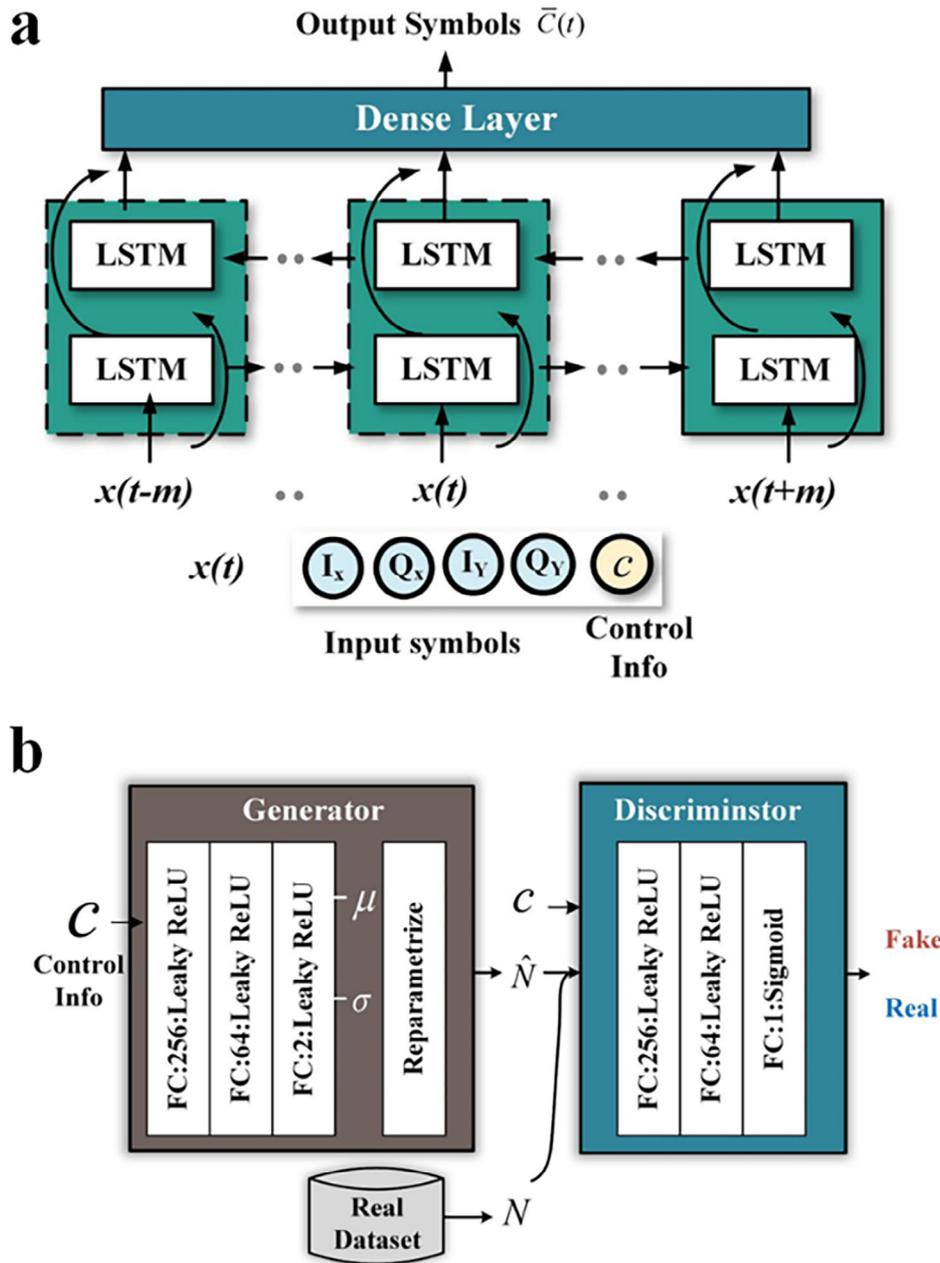
Looking forward, the ability to generate data that aligns closely with experimental outcomes using the DT model opens up new possibilities. It allows for rapid DSP verifications without the need for extensive physical experimentation, significantly accelerating the development and optimization of nonlinear DSP algorithms in optical communication. This capability is especially beneficial for exploring and validating new techniques in a more efficient and cost-effective manner.

## 4. Conclusion

In conclusion, we propose a gray-box DT framework including a data preprocessing module, DRNN, and physical models, for a physical optical communication system. The above results verify the wide generalization capability, high accuracy, fast calculation speed, and DSP verification capabilities of the DT structure, demonstrating the effectiveness and efficiency of the proposed DT framework. The DRNN model demonstrates robust and versatile modeling capabilities across a wide spectrum of link conditions and signal configurations. This includes various symbol distributions such as PCS and uniform distribution, WDM and single-channel scenarios, as well as long-haul and short distances. It also accommodates a wide range of launch power values and OSNR levels. Moreover, the model's adaptability is enhanced by fast finetuning techniques, enabling real-time tracking of channel conditions through online training. In testing scenarios, the DT model has shown remarkable accuracy, with the mean Q factor and SNR error below 0.088 dB, which closely align with real experimental data. The DRNN's running time of less than 1.3 s highlights its efficiency and low complexity. As the industry shifts toward ultra-wide-band transmission to further augment the capacity of optical fibers, the role of the proposed DT model becomes increasingly pivotal. In addition, we achieve a PNC based on the DT model, and the same nonlinear performance improvement is validated on the experiment data. Based on the excellent results, we believe the proposed DT architecture reduces both the cost and the time for the development cycle since the system modeling prior to the implementation and field trials. Building upon these advancements, the DT model can be more effectively applied in the realms of device design, system optimization, and management, thereby contributing significantly to the advancement of communication technologies.

## 5. Experimental Section

**Deterministic-Random Neural Network:** The data preprocessing module eliminates time-varying effects that neural networks were not good at processing, including PMD and PN. After the compensation, residual characteristics consist of uncompensated linearity, nonlinearity, and random noise, which could be divided into deterministic features and random features. There were two challenges in such data acquisition. First, to enhance the model's generalization capabilities, it was essential to alter system conditions to collect a large volume of data under various conditions. Second, data at the receiver end must undergo DSP before it can be utilized for network training. Both modifying experimental conditions and performing DSP operations were feasible, given the abundance of data available in communication systems. Thus, the challenges associated with data acquisition could be effectively addressed.



**Figure 6.** Deterministic-random NN structure. a) BiLSTM structure for deterministic features. b) GAN structure for random features. FC, fully connected.

Considering the time memory between the channel data and the perfect internal memory ability of BiLSTM, we choose it for deterministic effects modeling with temporal correlation.<sup>[38]</sup> As shown in **Figure 6a**, in the input layer, the symbols at different times  $[x(t-m), \dots, x(t), \dots, x(t+m)]$  were fed to the BiLSTM as an input sequence, which means that the input layer contains  $2m+1$  time steps. At each time step,  $x(t)$  contains one input symbol and control information, such as the launch power, distance, and OSNR value. The units of these values were dBm, km, and dB, respectively. The control information relates to the system performance and brings a strong generalization ability of NN. Note that the symbols in a coherent transmission system were generally represented by complex numbers. Complex numbers were converted into real numbers for NN input. Then the input data were fed to the LSTM cells to realize the recurrent

connection. The outputs from multiple time steps were concatenated to form a vector and this vector was then fed into a dense layer to obtain the output  $\bar{C}$ . The hyperparameters of the DT model are detailed in **Table 1**, which includes configurations for the BiLSTM, generator, and discriminator models.

GAN was a generative model consisting of two parts: a generator and a discriminator.<sup>[39]</sup> The generator aims to capture the training data distribution and generate new data with the same distribution to fool the discriminator. The discriminator classifies the real data and fake data. Here, a GAN structure was designed to learn the random distribution and generate random features of the channel. The structure is shown in **Figure 6b**. The training data, also the real dataset, of GAN was the difference between the Exp-DSP outputs and the BiLSTM outputs, which could be

**Table 1.** The hyperparameters of our digital twin model.

Bi-directional Long Short-Term Memory		Generator in Generative Adversarial Network (GAN)		Discriminator in GAN	
layers	3	input size	1	input size	5
input size	5	hidden layers	2	hidden layers	2
hidden neuron	5	hidden neuron	[256,64]	hidden neuron	[256,64]
time step	101	output size	1	output size	1

represented as  $C1 - \bar{C}$ . In the coherent system without dispersion management, the distribution of the combined effects of other channels and noise on the CUT could be considered as Gaussian distribution.<sup>[40,41]</sup> GAN itself could learn arbitrary random distribution, but this Gaussian approximation was made on the generator to reduce the training time and improve the modeling accuracy. The designed generator contains four fully connected layers. The control information was input to the generator to control the output random data corresponding to the system setting condition. The output of the third layer was the mean value  $\mu$  and standard deviation  $\sigma$  of random Gaussian noise and the random noise could be reparameterized by these learned values. That was, random data satisfying the corresponding distribution was obtained through sampling. The discriminator contains three fully connected layers. The control information and the generated data  $\hat{N}$  or real data  $N$  from the dataset were input. The output of the discriminator represents the probability that  $\hat{N}$  or  $N$  was the real data. In the final training stage, the discriminator cannot distinguish the generated data and the real data, which means they have similar distributions, and the model training was completed. In practice, training GAN involves an adversarial process that might introduce some distribution errors in the modeled data. These errors could be mitigated by increasing the volume of training data, which was discussed in the previous work.<sup>[25]</sup>

In the training process, the loss function of the BiLSTM is the smooth  $L1$  loss, and the GAN was the modified adversarial loss. The number of training data with each condition is 150 thousand. One training data reference one input and output symbol pair. The epoch number of BiLSTM was set at 200 and GAN was set at 300. The weights of both NNs were initialized by He initialization.<sup>[42]</sup> The optimizer was Adam.<sup>[43]</sup> The learning rate was  $1e-3$  initially and it decreases during the training process using a cosine annealing schedule.<sup>[44]</sup> The decreasing learning rate could improve training accuracy. After the training, the DRNN output, marked as  $C2$ , contains deterministic and random features and could be represented by  $\bar{C} + \hat{N}$ .

**Physical Model:** The DRNN only models the nonanalytic and random features preserved after DSP. The compensated features should also be modeled to form a complete DT system. The compensated features could be considered as linear effects, which could be described by physical models of link elements. As shown in Figure 1, the signal changes by the physical models were basically inverse to that of Rx DSP. The signal is first up-sampled to two samples per symbol because the linear effects modeling process should satisfy the Nyquist sampling theorem. Then PN and FO caused by laser on the  $k$ th symbol could be modeled as

$$l(k) = s(k) \exp(j\varphi_k), \varphi_k \Delta \omega kT + \theta_L(k) \quad (2)$$

where  $s(k)$  is the  $k$ th symbol after up-sampling,  $l(k)$  represents the  $k$ th symbol after the laser model,  $\varphi_k$  is the total phase caused by the PN and FO,  $T$  is the sampling period.  $\Delta \omega$  is the additional phase caused by the FO, which can be obtained from the FOE algorithm.  $\theta_L(k)$  is the PN from the laser linewidth, which is generally considered a Wiener process.<sup>[45]</sup>

Then the linear interferences from optical fiber are modeled. Since the parameters of MIMO adaptive filters were not explicable and had no physical meanings, the physical polarization model based on the Manakov-PMD equation was utilized to represent the PMD.<sup>[46,47]</sup> The numerical simulation method was SSFM, which was considered a highly accurate method for linear channel feature modeling. Differential group delay (DGD) values relate to the PMD intensity, and most PMD effects could

be compensated by MIMO adaptive filters. Therefore, the PMD intensity could be defined by users flexibly, and the DT performance would not be affected. With the nonlinear interplay with PMD effects along the link, nonlinearity becomes a random effect that could be averaged over the Poincaré sphere with the well-known  $8/9$  factor.<sup>[46]</sup> This suggests that the intensity of nonlinearity with PMD remains stable. The GAN was used to capture these random effects as a “lumped” noise, which was challenging to separate for any channel modeling scheme. The effectiveness of this approach was validated by the experimental results shown in Figure 3.

For convenience, CD and PMD effects were performed together using the SSFM in some steps. In each calculation step, the operation consists of three steps. First, rotate two local principal states of polarizations (SOP) of the polarization beam by a rotation  $R(\theta, \varphi)$ , which can be expressed as:

$$R(\theta, \varphi) = \begin{pmatrix} \cos \theta \sin \theta \exp(i\varphi) \\ -\sin \theta \exp(i\varphi) \cos \theta \end{pmatrix} \quad (3)$$

where  $\theta$  and  $\varphi$  represent the rotation angle and phase.<sup>[43]</sup> These rotation-related values are random corresponding to the random polarization states during the propagation. Then the effects of the CD and PMD can be implemented on the current SOP as

$$\begin{aligned} \bar{c}_x(z+h, \omega) &= \bar{c}_x(z, \omega) \exp\left(-i\frac{\beta_2}{2}\omega^2 + i\frac{\Delta\beta_1}{2}\omega\right) h, \bar{c}_y(z+h, \omega) \\ &= \bar{c}_y(z, \omega) \exp\left(-i\frac{\beta_2}{2}\omega^2 - i\frac{\Delta\beta_1}{2}\omega\right) h \end{aligned} \quad (4)$$

where  $\bar{c}_x$  and  $\bar{c}_y$  represent the frequency-domain signal output from the laser model over two arbitrary orthogonal polarization modes,  $\bar{c}_x$  and  $\bar{c}_y$  represent the frequency-domain signal output from the dispersion model,  $z$  is the distance coordinate and  $h$  is the step length.  $\beta_2$  is the group velocity dispersion (GVD) parameter, and this parameter is the same as that in CDC.  $\Delta\beta_1$  is DGD caused by PMD.<sup>[48]</sup> After modeling the linear effects defined by GVD and DGD, the SOP is rotated back to the original states. With the integration of additional measurements, users can acquire further parameters such as the device’s bandwidth limit and the delay between the in-phase (I) and quadrature (Q) branches through the optical spectrum.<sup>[49]</sup> These physical parameters are not included in the BiLSTM model, as they are explicitly defined within the physical model itself. To incorporate these parameters, an upsampling operation and a Tx physical model would need to be implemented prior to the BiLSTM model. This inclusion allows for the consideration of Tx distortions in the model. Although this adjustment may deviate from what is illustrated in Figure 1a, it does not alter the fundamental principle of the DT model. Future research can aim to enhance the flexibility of the DT model, thereby accommodating the evolving user needs and incorporating additional features.

With advancements in DSP, more accurate dynamic effects could be captured and modeled inversely. However, the challenge remains that time-varying effects may change over time, making real-time tracking a critical limitation in capturing such effects accurately. To address this limitation, it was essential to develop a real-time tracking model that consistently aligns with experimental conditions. This development might involve three key steps: 1) extraction of DSP parameters, 2) estimation of parameters for the physical model, and 3) construction using the DT model.

Integrating these steps into a comprehensive framework represents an intriguing and promising area for future research based on the DT model.

## Acknowledgements

H.Y. and Z.N. contributed equally to this work. The authors acknowledge the funding provided by the National Key R&D Program of China (2023YFB2905400), the National Natural Science Foundation of China (62025503), and the Shanghai Jiao Tong University 2030 Initiative.

## Conflict of Interest

The authors declare no conflict of interest.

## Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## Keywords

data-driven and model-driven, digital twin, generative neural networks, optical fiber communications

Received: February 19, 2024

Revised: May 7, 2024

Published online: June 3, 2024

- [1] D. Jones, C. Snider, A. Nassehi, J. Yon, B. Hicks, *CIRP. J. Manuf. Sci. Tec.* **2020**, 29, 36.
- [2] M. Liu, S. Fang, H. Dong, C. Xu, *J. Manuf. Syst.* **2021**, 58, 346.
- [3] S. Haag, A. Reiner, *Manuf. Lett.* **2018**, 15, 64.
- [4] F. Tao, H. Zhang, A. Liu, A. Nee, *IEEE Trans. Industr. Inform.* **2018**, 15, 2405.
- [5] Benjamin, S., Nabil, A., Luc, M., *Cirp Ann.-Manuf. Techn.* **2017**, 66, 141.
- [6] I. Errandonea, B. Sergio, A. Saioa, *Comput. Ind.* **2020**, 123, 103316.
- [7] R. Sacks, I. Brilakis, E. Pikas, H. Xie, M. Girolami, *Data-Centr. Engin.* **2020**, 1, e14.
- [8] F. Tao, F. Sui, A. Liu, Q. Qi, M. Zhang, B. Song, Z. Guo, C. Lu, A. Nee, *Int. J. Prod. Res.* **2019**, 57, 3935.
- [9] W. Kritzing, M. Karner, G. Traar, J. Henjes, W. Sihn, *Ifac-PapersOnline* **2018**, 51, 1016.
- [10] J. Vachálek, L. Bartalský, O. Rovný, D. Šišmišová, M. Morháč, M. Lokšík, *IEEE international conf. on PC* **2017**.
- [11] Gartner's Top10 strategic technology trends for 2018, <https://www.gartner.com-smarterwithgartner/gartners-top-10-technology-trends-2018>, (accessed: October 2017).
- [12] Gartner's Top 10 strategic technology trends for 2019, <https://www.gartner.com-smarterwithgartner/gartners-top-10-technology-trends-2019>, (accessed: October 2018).
- [13] D. Wang, Z. Zhang, M. Zhang, M. Fu, X. Chen, *IEEE Com. M.* **2021**, 59, 133.
- [14] N. A. Kudryashov, *Optik* **2019**, 189, 42.
- [15] I. S. Amiri, Z. R. Ahmed, P. Yupapin, *J. Opt. Commun.* **2020**.
- [16] B. Karanov, M. Chagnon, F. Thouin, T. A. Eriksson, H. Bülow, D. Lavery, P. Bayvel, *J. Lightwave. Technol.* **2018**, 36, 4843.
- [17] G. P. Agrawal, *Nonlinear Fiber Optics*, 4th ed., Academic Press, UK, **2006**.
- [18] O. V. Sinkin, R. Holzlöhner, J. Zweck, C. Menyuk, *J. Lightwave. Technol.* **2003**, 21, 61.
- [19] P. Poggiolini, *J. Lightwave. Technol.* **2012**, 30, 3857.
- [20] A. Carena, G. Bosco, V. Curri, Y. Jiang, P. Poggiolini, F. Forghieri, *Opt. Express* **2014**, 22, 16335.
- [21] D. Wang, Y. Song, J. Li, J. Qin, A. Boucouvalas, *J. Lightwave. Technol.* **2020**, 38, 4730.
- [22] H. Yang, Z. Niu, S. Xiao, J. Fang, L. Yi, *J. Lightwave. Technol.* **2021**, 39, 1322.
- [23] H. Yang, Z. Niu, H. Zhao, S. Xiao, W. Hu, L. Yi, *J. Lightwave. Technol.* **2022**, 40, 4571.
- [24] Y. Zang, Z. Yuan, K. Xu, M. Chen, S. Yang, H. Chen, *Opt. Express* **2022**, 30, 46626.
- [25] H. Yang, H. Zhao, Z. Niu, G. Pu, S. Xiao, W. Hu, L. Yi, *Opt. Express* **2022**, 30, 43691.
- [26] X. Jiang, D. Wang, Q. Fan, M. Zhang, C. Lu, A. Lau, *Laser Photonics Rev.* **2022**, 16, 2100483.
- [27] S. S. Haykin, *Digital Communications*, Wiley, Hoboken, NJ **1988**.
- [28] C. Xie, *IEEE Photonic. Tech. L.* **2013**, 25, 992.
- [29] E. Ip, M. Joseph, *J. Lightwave. Technol.* **2008**, 26, 3416.
- [30] K. Oh, O. Yong, Presented at *Proceedings IEEE ICC*, June, **1995**.
- [31] A. Leven, N. Kaneda, U. V. Koc, Y. K. Chen, *IEEE Photonic. Tech. L.* **2007**, 19, 366.
- [32] Z. Tao, L. Dou, W. Yan, L. Li, T. Hoshida, C. Rasmussen, *J. Lightwave. Technol.* **2011**, 29, 2570.
- [33] A. Amari, O. A. Dobre, R. Venkatesan, O. S. S. Kumar, P. Ciblat, Y. Jaouën, *IEEE Commun. Surv.* **2017**, 19, 3097.
- [34] S. Deligiannidis, A. Bogris, C. Mesaritis, Y. Kopsinis, *J. Lightwave. Technol.* **2020**, 38, 5991.
- [35] A. Ferrari, M. Filer, K. Balasubramanian, Y. Yin, E. L. Rouzic, J. Kundrat, G. Grammel, G. Galimberti, V. Curri, *J. Opt. Commun. Netw.* **2020**, 12, C31.
- [36] A. Kakkar, J. R. Navarro, R. Schatz, H. Louchet, X. Pang, O. Ozolins, G. Jacobsen, S. Popov, *J. Lightwave. Technol.* **2015**, 33, 4834.
- [37] H. Wang, X. Yi, J. Zhang, F. Li, *J. Lightwave. Technol.* **2022**, 40, 7808.
- [38] S. Hochreiter, J. Schmidhuber, *Neural. Comput.* **1997**, 9, 1735.
- [39] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, *Proc. Adv. Neural Inf. Process. Syst.* **2014**, 27, 2672.
- [40] R. Dar, M. Feder, A. Mecozzi, M. Shtaif, *Opt. Express.* **2013**, 21, 25685.
- [41] D. Dahan, M. Zarubinsky, Y. Liang, O. Golani, M. Shtaif, *J. Lightwave. Technol.* **2022**, 40, 2441.
- [42] K. He, X. Zhang, S. Ren, J. Sun, *Proc. IEEE Int. Conf. Comput. Vis.* **2015**, 1026.
- [43] D. P. Kingma, J. Ba, *arXiv* **2014**, 1412.6980.
- [44] I. Loshchilov, F. Hutter, *arXiv* **2016**, 1608.03983.
- [45] T. Pfau, H. Sebastian, N. Reinhold, *J. Lightwave. Technol.* **2009**, 27, 989.
- [46] D. Marcuse, C. Menyuk, P. Wai, *J. Lightwave. Technol.* **1997**, 15, 1735.
- [47] S. Evangelides, L. Mollenauer, J. P. Gordon, N. S. Bergano, *J. Lightwave. Technol.* **1992**, 10, 28.
- [48] X. Zhou, C. Xie, *Enabling Technologies for High Spectral-Efficiency Coherent Optical Communication Networks*, John Wiley & Sons, Hoboken, NJ **2016**.
- [49] H. Chen, X. Su, Z. Tao, T. Oyama, H. Nakashima, T. Hoshida, K. Kato, 2017 European Conf. on Optical Communication IEEE, Sweden **2017**.