A MIMO Detector With Deep-Neural-Network for Faster-Than-Nyquist Optical Wireless Communications

Minghua Cao [ID], Ruifang Yao [ID], Qinxue Sun [ID], Yue Zhang [ID], Qing Yang [ID], and Huiqin Wang [ID]

Abstract—Conventional multiple input multiple output (MIMO) detection algorithms face challenges related to computational complexity and limited performance when handling high-dimensional inputs and complex channel conditions. In order to enhance signal recovery accuracy in atmospheric turbulence channels for faster-than-Nyquist (FTN) optical wireless communication (OWC) systems, a deep learning (DL) based MIMO detector is proposed. By leveraging a deep neural network (DNN), it becomes possible to learn nonlinear mappings within MIMO systems, resulting in improved detection performance while reducing computational overheads. Simulation results validate that our proposed DNN detector achieves comparable performance to the maximum likelihood (ML) method, while reducing complexity by 40%.

Index Terms—Deep neural network, faster-than-nyquist, multiple input multiple output, optical wireless communication.

I. INTRODUCTION

THE implementation of fast multiplexing and demultiplexing for non-orthogonal frequency-division multiplexing (NOFDM) is achieved through the utilization of inverse fast Fourier transform and fast Fourier transform [1]. The authors propose a novel approach called non-orthogonal discrete multi-tone (NODMT), which combines the advantages of NOFDM and discrete multi-tone (DMT). This integration has the potential to significantly enhance spectral efficiency in future optical networks. The multiple input multiple output (MIMO) system, renowned for its capability to improve spectral efficiency and enhance link reliability, has emerged as a prevalent technology in contemporary wireless communication standards [2]. Compared to single input single output (SISO) systems, MIMO systems effectively utilize spatial resources and augment channel capacity without requiring additional bandwidth because each receiver antenna can simultaneously receive signals transmitted by all transmitting antennas [3]. With the growing demand for high data rate communication networks, Mazo demonstrated in 1975 that higher transmission rates could be achieved using faster-than-Nyquist (FTN) technology [4]. In [5], the authors introduced the concept of non-orthogonal wavelength division multiplexing (WDM) and provided a comprehensive review of the principle of FTN. A performance comparison between FTN and constellation shaping was presented. The authors extensively discussed the underlying principles of FTN signaling, emphasizing its orthogonality and distinguishability relative to Nyquist and Mazo limits. Furthermore, they demonstrated single carrier time domain FTN signals using the cascaded binary-phase-shift-keying iterative detection (CBID) algorithm [6]. In [7], the authors investigated the capacity of FTN signaling for both frequency flat and frequency selective (FS) MIMO channels. Considering that FTN introduces additional frequency selectivity, it was found that precoding in time (or equivalent frequency on the spectrum) combined with waterfilling in spatial domain achieves capacity for frequency flat MIMO channels with FTN. This combination of FTN technology and MIMO technology further enhances spectrum efficiency by artificially compressing symbol intervals to transmit more symbols. However, it also increases the complexity of signal detection. Maximum likelihood (ML) detection [8] represents the optimal choice capable of achieving peak performance. However, its computational grows exponentially with both modulation order and the number of antennas, posing challenges for practical implementation. On the other hand, algorithms with lower computational complexity such as zero forcing [9] (ZF) and Minimum Mean Square Error [10] (MMSE) achieve detection through straightforward linear transformations but exhibit a significant gap compared to the ML algorithm and have unsatisfactory performance for FTN signals with a high acceleration factor.

In recent years, deep learning (DL) has garnered significant attention from both academia and industry due to its powerful learning capabilities. The main advantage of DL lies in its ability to extract crucial information from premarked training data [11]. Consequently, DL approaches have become increasingly popular for solving MIMO signal detection problems. For example, [12] explores an unsupervised DL-based MIMO detection method that utilizes an autoencoder to learn the entire system. Moreover, DetNet [13] represents one of the earliest DL-based detection approaches for MIMO detection by employing a model-based algorithm that unfolds the iterations of the projected gradient descent method. In [14], authors proposed

a simplified version of DetNet by reducing inputs dimensions and simplifying network structure. Corlay et al. [15] suggest replacing the sigmoid activation function used in DetNet with a multi-plateau version and implementing two networks with distinct initial values to simultaneously detect transmitted signals, resulting in improved detection performance through selection based on a smaller loss function solution. The authors conducted a comprehensive evaluation of four prominent model-based deep learning techniques, namely DetNet, MMNet, GEPNet, and Recurrent Equivariant MIMO (RE-MIMO), based on different working principles. They assessed the reliability, complexity, and robustness of these techniques against the practical Massively Parallel Non-Linear (MPNL) processing detection approach [16]. Similar to DetNet, orthogonal approximate message passing network (OAMP-Net) was developed by unfolding the OAMP algorithm [17], [18]. It has been proven that OAMP-Net requires minimal training time and is capable of adapting to varying channels. However, it necessitates prior estimation of noise variance. Additionally, two data-driven approaches called deep neural network (DNN) and convolutional neural network (CNN) are employed for MIMO detection over a fixed channel case in [19]. Another data-driven approach is presented in [20], where conventional DL network structures are utilized for signal detection in a typical MIMO system over an erroneous channel scenario. In [21], authors propose a DL-based detection method utilizing neural networks to obtain optimal decision regions for multi-user MIMO systems. Furthermore, in [22], a MIMO detection method employing DNN is proposed for an optical transmission system, indicating the growing trend of applying DL to communication technology. The authors propose the modified expectation propagation network (MEPNet), which employs the DL scheme and unfolds iterative the modified expectation propagation detector (MEPD) to provide the best damping factor and initial variance in [23]. The authors in [24] propose a parallel detection network (PDN) that achieves a significant diversity effect by incorporating a tailored loss function and minimizing the similarity between detection networks. Notably, the performance of PDN exhibits substantial improvement with an increasing number of parallel detection networks in time-varying MIMO channels. The authors [25] introduce the learn iterative search algorithm (LISA), which treats the signal detection problem as a tree-based decision problem with the objective of learning the optimal decision strategy. The authors develop a model-driven DL detector based on variational Bayesian inference, where their proposed unfolded DL architecture is inspired by the non-invertible variational Bayesian learning framework, effectively avoiding matrix inversion by maximizing the relaxed evidence lower bound in [26]. In [27], authors propose an efficient data-driven detection network, i.e., accelerated multiuser interference cancellation network (AMIC-Net), for uplink massive MIMO systems.

However, the design and implementation of DL-based MIMO detection algorithms present their own challenges. These challenges encompass the selection of an appropriate network architecture, optimizing of hyperparameters, mitigation of overfitting issues, and management of the computational complexity associated with training large-scale DNNs. Nonetheless, numerous research endeavors have demonstrated the feasibility and efficacy of DL-based MIMO detection in various wireless communication scenarios. In this research context, the paper is dedicated to overcoming the challenges of high complexity in DL detection algorithms, particularly focusing on time complexity. Therefore, a DNN-based detection method is proposed for pulse position modulation (PPM) signal detection in MIMO-FTN OWCs.

The remaining part of this paper is organized as follows: Section II presents the system model, followed by the detection scheme is in Section III. Numerical results and complexity analysis are presented in Section IV. Finally, Section V concludes the paper.

II. SYSTEM MODEL

Traditionally, intensity modulation/direct detection based on an on–off keying (OOK) is widely accepted in OWC owing to its easy implementation and lower cost [28]. To further improve spectrum efficiency and anti-interference capability, PPM has been considered for OWC communications. Compared with OOK, PPM significantly increases the data transmission rate and system reliability [29]. Fig. 1 illustrates a DNN-based MIMO-FTN OWC system utilizing 4PPM modulation. User data is initially encoded using Gray code and then mapped into 4PPM format. Subsequently, the mapped signal undergoes FTN shaping via a filter [30] before being converted from digital to analog through a digital-to-analog converter (DAC), and transmitted through multiple optical antennas in the atmospheric channel. At the receiver end, the optical signals are received by multiple optical antennas and converted into electrical signals which are then forwarded for analog-to-digital conversion (ADC), matched filtering, and sampling. Finally, the signal is sent to the DL module for data recovery.

According to the Nyquist criterion, in a bandwidth-constrained channel, the maximum code rate for high-speed data transmission should not exceed twice the channel bandwidth if we want to avoid inter-symbol interference (ISI). If this limit is exceeded, severe ISI will occur and result in a degradation of the system's bit error rate (BER) performance. This maximum rate for ISI-free transmission is also referred to as the Nyquist rate. Therefore, for an ideal low-pass channel with the bandwidth of W, the symbol transmission rate must be less than $2W$ Bd, where $2W$ Bd is defined as the Nyquist rate [31]. For an ideal band pass channel, the corresponding Nyquist rate is W Bd. In FTN rate communication, the symbol transmission rate exceeds $2W$ Bd while maintaining a low-pass channel of just W. With the same symbol rate, FTN technology directly improves spectrum efficiency [32]. Assuming that $h(t)$ represents a band-limited pulse with finite energy and $H(f)$ is its Fourier transform, $h(t)$ is considered T-orthogonal when it satisfies (1).

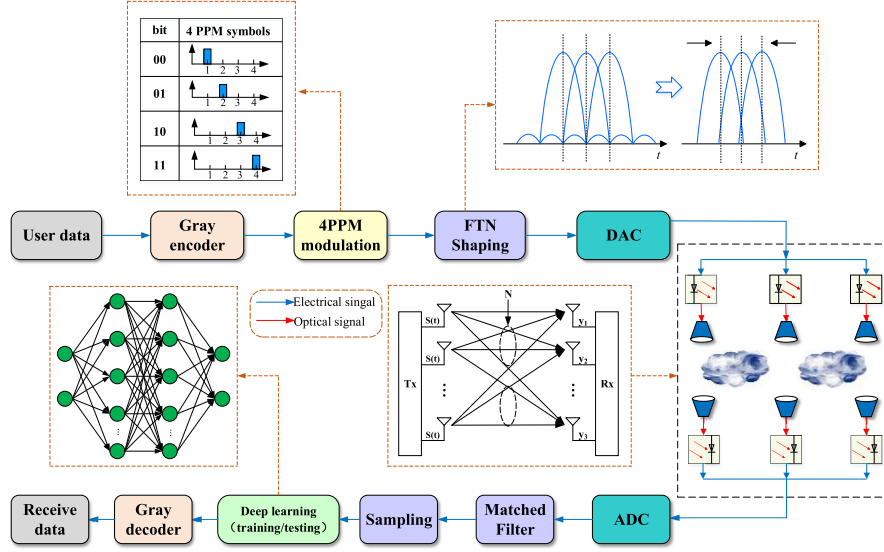$$\frac{1}{T} \sum_{k=-\infty}^{+\infty} \left| H\left(f + \frac{k}{T}\right) \right|^2 = 1 \qquad (1)$$

Fig. 1. Schematic of DNN-based MIMO-FTN OWC system.

Since the $\sin c(t) = \frac{\sin(\pi t/T)}{(\pi t/T)}$ pulse, which is a T-orthogonal pulse, can be obtained by normalizing its energy:

$$g_T(t) = \frac{\sin(\pi t/T)}{(\pi t/\sqrt{T})} \qquad (2)$$

Using the pulse $g_T(t)$ to transmit binary data with a symbol period T, assuming the transmitted data is an all "1" sequence and there is no ISI, this transmission mode is referred to as orthogonal transmission. In this case, the transmitted signal can be represented as:

$$s(t) = \sum_{n=N_1}^{N_2} a_n g_T(t - nT) \qquad (3)$$

where $a_n \in \{-1, 1\}$, N_1 and N_2 are integers that satisfy $N_2 > N_1$. Assuming transmission through an Additive White Gaussian Noise (AWGN) channel with a noise two-sided power spectral density of $N_0/2$, the BER at the receiver using optimal detection can be expressed as:

$$P_e = Q\left(\sqrt{\frac{2E}{N_0}}\right) \qquad (4)$$

where $Q(x) = \frac{1}{2}\text{erfc}(\frac{x}{\sqrt{2}})$, $\text{erfc}(\cdot)$ is the complementary error function, E is the energy of pulse $g_T(t)$. Sending data with smaller time intervals while keeping the transmission power constant can be expressed as:

$$s(t) = \sum_{n=N_1}^{N_2} a_n g_T(t - nT') = \sum_{n=N_1}^{N_2} a_n g_T(t - n\tau T) \qquad (5)$$

where a_n denotes the information carried by the $n - th$ symbol, τ signifies the time acceleration factor ($0 < \tau < 1$), which characterize the Nyquist compression ratio, and T corresponds the symbol period, $T' = \tau T < T$. It should be noted that when $\tau = 1$, the entire process is equivalent to Nyquist transmission
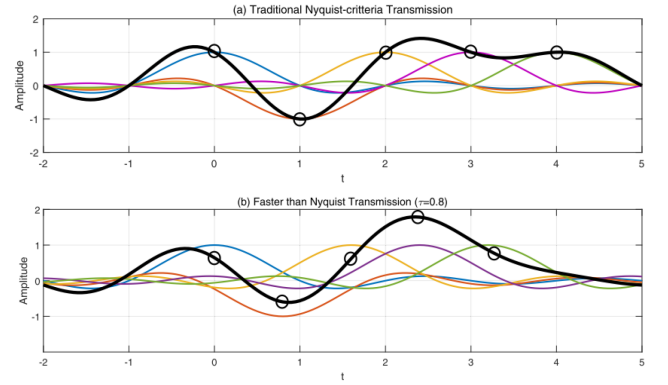

Fig. 2. Transmitted symbol waveforms with Nyquist and FTN signaling.

without any ISI. When $0 < \tau < 1$, the system can be considered as an FTN transmission system. FTN improves the spectral efficiency by reducing distance between adjacent symbols within limited bandwidths.

Fig. 2 illustrates the transmitted symbols with Nyquist and FTN signaling, respectively. Each symbol's waveform is subject to interference from other symbols, thereby introducing challenges in demodulation and symbol detection.

The atmospheric channel fading coefficient h follows Gamma–Gamma distribution [33], and its probability density function can be expressed as:

$$\varphi = \frac{2(\alpha\beta)^{\frac{\alpha+\beta}{2}}}{\Gamma(\alpha)\Gamma(\beta)} \cdot h^{\frac{\alpha+\beta}{2}-1} \cdot Q_{\alpha-\beta}(2\sqrt{\alpha\beta h}), h > 0 \qquad (6)$$

where $Q_{\alpha-\beta}(\cdot)$ is the second class modified Bessel function of order $\alpha - \beta$; $\Gamma(\cdot)$ is the Gamma function, α and β are the large and small scale scattering coefficients, respectively. The values
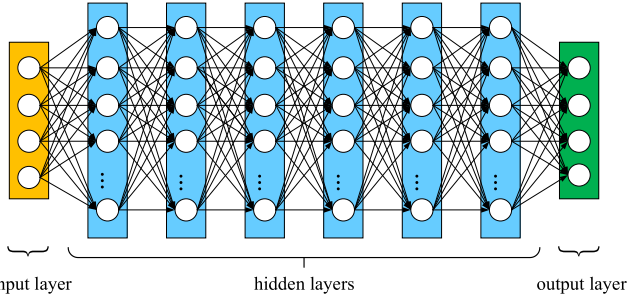
Fig. 3.   Schematic diagram of an 8-layer deep neural network.

Fig. 4.   Structure of m-th hidden layer.

of α and β can be determined as

$$\alpha = \left\{ \exp\left[\frac{0.49\sigma^2}{\left(1 + 0.18d^2 + 0.56\sigma^{12/5}\right)^{7/6}} \right] - 1 \right\}^{-1} \quad (7)$$

$$\beta = \left\{ \exp\left[\frac{0.51\sigma^2}{\left(1 + 0.9d^2 + 0.62\sigma^{12/5}\right)^{5/6}} \right] - 1 \right\}^{-1} \quad (8)$$

where $\sigma^2 = 0.5C_n^2 k^{7/6} L^{11/6}$ is the Rytov variance, C_n^2 is the refractive-index structure constant, L is the transmission distance, $k = 2\pi/\lambda$, λ is the wavelength, $d = \sqrt{kD^2/(4L)}$, D is the receiver aperture diameter.

The considered system is a MIMO configuration, where the number of receiving antennas (N_r) and transmitting antennas (N_t) are defined. The received signal can be mathematically expressed as

$$\bar{y} = \bar{H}\bar{x} + \bar{n} \quad (9)$$

where $\bar{x}$ denotes the transmitted symbol vector, $\bar{H}$ denotes the channel matrix and $\bar{n}$ denotes the noise.

When employing ML detection, the estimated signal can be mathematically represented as follows:

$$x_{ML} = \arg\min \|y - Hx\|^2 \quad (10)$$

where y, H, and x denote the received signal, channel matrix, and transmitted signal respectively. The ML detection technique achieves optimal performance when there is perfect knowledge of the channel state information (CSI). It is widely acknowledged as the most effective approach for detecting signals in MIMO systems. However, its exponential complexity renders it impractical for real-world applications. Therefore, DL-based algorithms are being considered for signal detection.

III. DEEP NEURAL NETWORK DETECTION SCHEME

A. The Structure of the DNN

The schematic diagram in Fig. 3 illustrates a fully connected DNN with 8-layer, consisting of an input layer, 6 hidden layers, and an output layer. The primary function of the input layer is to receive initial input data and propagate it through the network. The hidden layers play a crucial role in learning and extracting pertinent features from the input data. The output layer servers as the final stage of the DNN, responsible for generating predictions
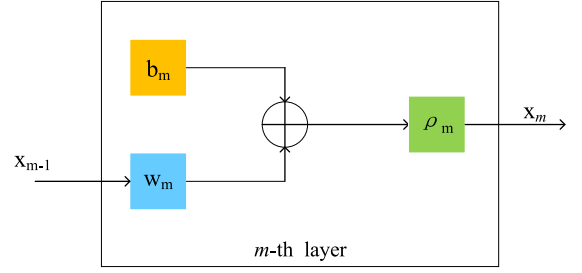
or outputs based on the processed input data. To achieve multi-classification, the Softmax activation function is employed to produce the final output of the network. These interconnected layers are weighted connections that are learned during training process. As illustrated in Fig. 3, we can describe this complete DNN as a function $f(\cdot)$ that maps input vectors to output vectors at each layer through neuron calculations. The mapping function of the entire DNN can be represented as:

$$x_{out} = f(x_{in}; \theta) \quad (11)$$

where x_{in} denotes the input vector, x_{out} denotes the output vector, and $\theta = \{\theta_1, \theta_2, \ldots, \theta_L\}$ denotes parameter set of DNN which includes subset parameters in each layer.

As depicted in Fig. 4, for an M-layered DNN, the x_{M-1} refers to $m-th$ layer's input vector which acts as $(m-1)-th$ layer's output vector. The mapping function can be expressed as:

$$x_m = f_m(x_{m-1}; \theta_m)$$
$$= \rho_m(w_m x_{m-1} + b_m) \quad (12)$$

where x_m denotes $m-th$ layer's output vector, $w_m \in \theta_m$ and $b_m \in \theta_m$ denote weight and bias respectively for $m-th$ layer, $\rho_m(\cdot)$ signifies activation function adding non-linearity to enable arbitrary fitting capability [34]. Furthermore, we have opted to utilize the Sigmoid function [35] as the activation function for output layers and the rectified linear unit (ReLU) [36] activation function for hidden layers. Specifically, these functions can be expressed as:

$$f_S(x) = \text{Sigmoid}(x) = \frac{1}{1 + e^{-x}} \quad (13)$$

$$f_R(x) = \text{ReLU}(x) = \begin{cases} x, & \text{if } x \geq 0 \\ 0, & \text{if } x < 0 \end{cases} \quad (14)$$

The Sigmoid function yields an output value ranging from 0 to 1, rendering is suitable for binary classification methods. On the other hand, the ReLU function exhibits characteristics of simplicity, low computation cost and fast convergence speed.

B. The Structure of the DNN-based MIMO Detector

The proposed DNN-based MIMO-FTN detector is illustrated in Fig. 5. At the receiving end, the signal is sampled and fed into a DL module, which can be considered as a "black box" that decodes the signal using a neural network, thereby achieving demodulation of the signal. In the DL module, the model is
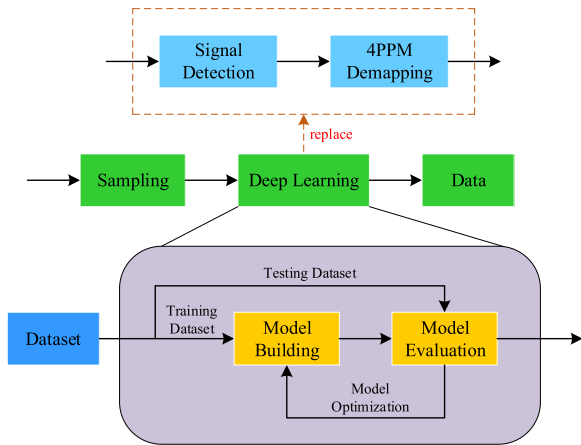
Fig. 5. Schematic diagram of the signal detector.

Fig. 6. Testing process. (a)Accuracy curve, (b) loss curve.

optimized through training the dataset. Each hidden layer within the network consists of a fully connected (FC) sublayer and a ReLU. The numbers of neurons in each layer are 16, 100, 200, 300, 300, 200, 100 and 16, respectively. The output layer only contains a FC sublayer. To train this model, we utilize the cross-entropy function as the loss function within the Tensorflow framework. Initially, output values are mapped to the interval (0, 1) using a Sigmoid function and then we calculate the cross entropy between these practical output values and training targets.

C. Training Procedure and Details

The training process of DNN consists of the following steps and procedures: (a) Data preparation. Initially, Matlab software is utilized to generate a random 1×10^6 (0–1) sequence. Subsequently, the data sequence is mapped into 4PPM signal, followed by dividing the data into training and testing dataset. (b) Parameter initialization. The weights and biases of the network are randomly initialized to introduce an element of randomness. (c) Forward propagation. Data is input from the input layer and propagated through the hidden layer, with the output layer responsible for generating predictions or final result. This involves utilizing the output of each preceding layer as input of subsequent layers until reaching the output layer itself. (d) Loss Calculation. The network's output is compared with true labels in order to calculate loss function values accordingly. (e) Backpropagation. Data flows from the output layer back to the input layer while gradients of each parameter with respect to loss function are calculated using chain rule principles. The parameters are then updated iteratively via an optimization algorithm that minimizes loss. (f) Parameter update. Network parameters such as weights and biases are updated using an optimization algorithm based on gradients. (g) Iteration. By repeatedly feeding data into the network and performing aforementioned steps, network parameters continue being updated until a desired criterion is achieved. Once all iteration has been completed, DNN detector can serve as a MIMO system detector equipped with trained parameters. Specifically, we employ adaptive moment estimation optimizer during our training process.
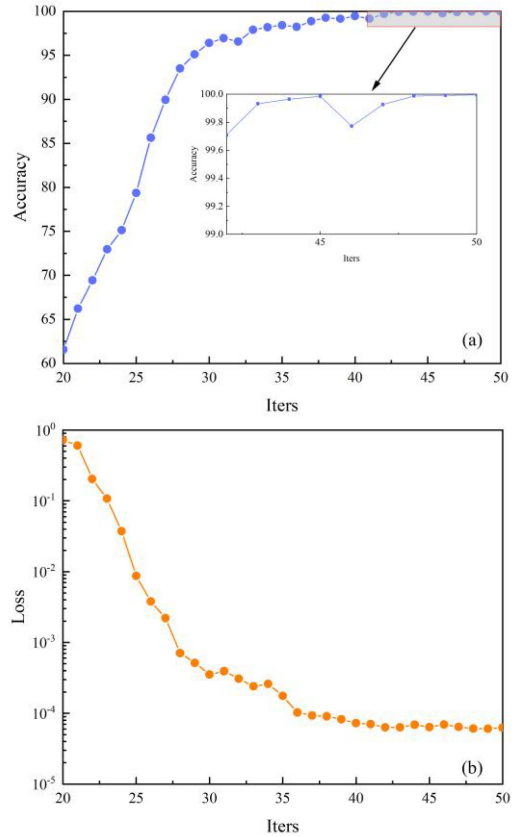
D. Testing Process

The testing process of a trained DNN involves evaluating the model's performance and generalization ability using an independent testing dataset. By utilizing the learned weights and biases, the testing data is propagated through the trained network via forward propagation to obtain predictions or outputs from the output layer of the network. Finally, the SoftMax function classification can be utilized for signal detection.

The accuracy and Loss curves are depicted in Fig. 6. As illustrated in Fig. 6(a), it becomes apparent that the network's testing accuracy tends to stabilize after undergoing 40 epochs of training. Furthermore, as observed from Fig. 6(b), the loss exhibits no further reduction beyond the 40th epoch. Ultimately, the network achieved an accuracy rate of approximately 99.99%, with a loss value below 10^{-4}.

IV. SIMULATION ANALYSIS

The choice of training or test dataset size contingent upon the complexity of the system and the employed DL algorithm. Utilizing a small dataset may lead to subpar detection performance, as it might not adequately capture the diverse characteristics inherent to the system. Conversely, using a large dataset can escalate computational complexity [37]. Therefore, multiple simulations are conducted to determine the appropriate dataset size and parameters that can yield optimal BER performance. It should be noted that the accuracy of the neural network is

TABLE I
SUMMARY OF SIMULATION PARAMETERS

Parameter	Value
size of data	2×10^6
training dataset	1.6×10^6 (80%)
test dataset	4×10^5 (20%)
batch size	100
dropout	0.05
cycle index	20
SNR range	0dB-30dB
learning rate	0.001
layer	8
activation function	Softmax

TABLE II
RELATIONSHIP BETWEEN LEARNING RATE AND ACCURACY

Learning Rate	Accuracy Rate
0.00001	89.36548953%
0.00010	98.46000061%
0.00100	99.96199799%
0.01000	99.33684231%
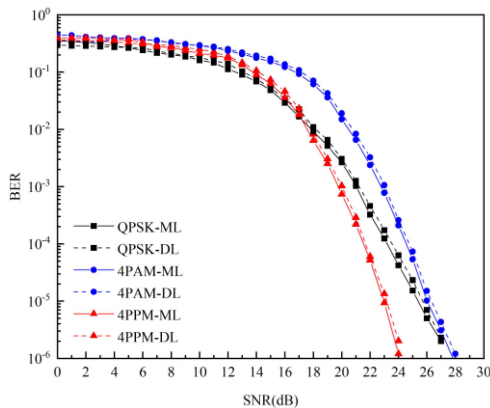
Fig. 7. Relationship of BER and modulation format.

Fig. 8. BER under different atmospheric turbulence channels.

affected by the DL algorithm itself, which plays a crucial role in solving nonlinear problems. Hence, evaluating the performance and robustness of our proposal is essential. The simulation parameters used are listed in Table I.

Table II displays the accuracy of the network under different learning rates. A validated system necessitates an appropriate learning rate, as convergence becomes unattainable with excessively large values and slow or nonexistent convergence occurs with excessively low values. Furthermore, increasing the learning rate may cause a transition from underfitting to overfitting [38]. It is evident from the Table that a learning rate of 0.001 yields optimal performance.

Fig. 7 shows the BER curves under different modulation formats. Pulse amplitude modulation (PAM) commonly employed for intensity modulation, where the pulse amplitude is adjusted according to a specific law to regulate the output. However, achieving idea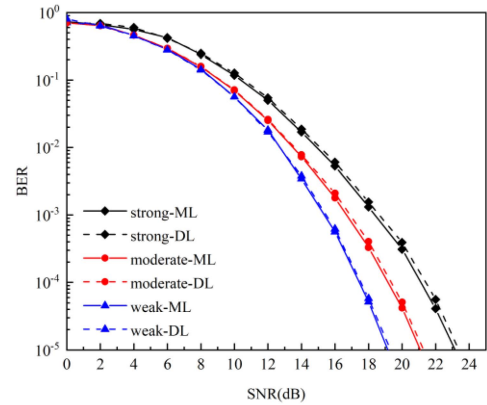l sampling of the impulse sequence proves challenging in practical scenarios. Quadrature phase shift keying (QPSK), a form of quadrature modulation, enables transmission of two bits per symbol by dividing the carrier signal into in-phase (I) and quadrature (Q) components that are modulated independently. PPM modulation stands out due to its simplicity, ease of implementation, and high robustness, leading to significant improvements in data transmission rate and system reliability. As shown in Fig. 7, it can be observed that QPSK outperforms 4PPM modulation when the signal to noise ratio (SNR) falls below 17.5 dB. However, when SNR exceeds this threshold value, 4PPM surpasses QPSK as a result of DL being consistently performed within the real-valued domain with consideration given to treating imaginary part signals as equivalent real-valued representations. Henceforth, we adopt 4PPM modulation due to its superior performance at high SNR levels.

As depicted in Fig. 1, the signal is transmitted via the antennas and directed towards the Gamma-Gamma atmospheric channel. It is widely acknowledged that various atmospheric factors, such as rain, snow, sleet, fog, haze, pollution, etc., significantly impact laser beams by inducing reflection, refraction, scattering, and attenuation [39], [40], [41]. The accuracy of modeling atmospheric turbulence using the Gamma-Gamma distribution has been demonstrated with distinct levels of turbulence intensity (weak, moderate, and strong) categorized by refractive-index structure constant values [42] of $C_n^2 = 1 \times 10^{-17}$, $C_n^2 = 1 \times 10^{-14}$, and $C_n^2 = 1 \times 10^{-13}$, respectively. The curves depicted in Fig. 8 illustrate the correlation between different levels of atmospheric turbulence intensities and BER, with a roll-off factor of 0.5, $\tau =$ 0.8, and a transmission distance of 1000 m. It is evident from the figure that weak turbulence conditions result in superior BER performance. Specifically, when BER$= 3.8 \times 10^{-3}$, the BER performance under weak turbulence is approximately 1 dB and 2 dB better than that under moderate and strong turbulence, respectively. Moreover, as the intensity of turbulence increases, the adaptability of DL gradually deteriorates.

The BER performance of the 2×2 MIMO-FTN system gradually deteriorates with increasing transmission distance at a specific wavelength, as depicted in Fig. 9, where the acceleration factor is set to 0.8 and SNR is maintained at 20 dB.
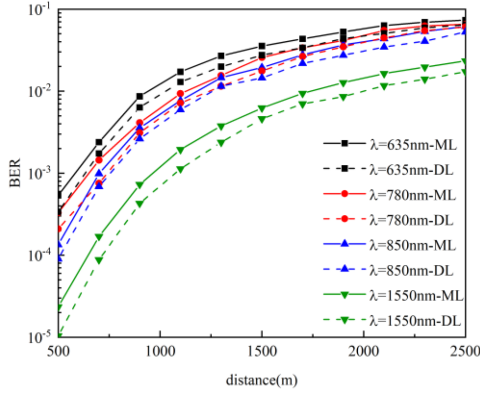
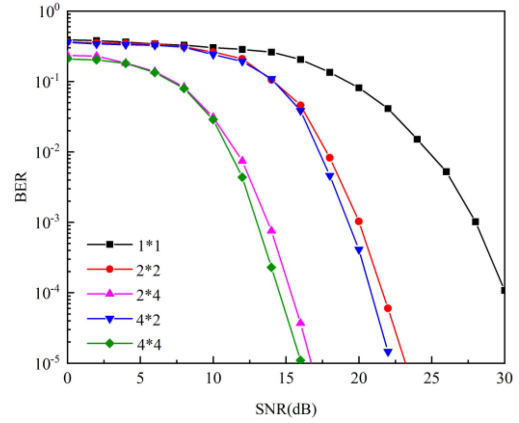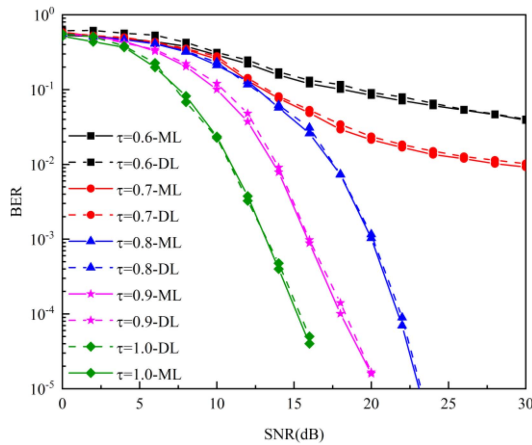Fig. 9. Curve of transmission distance and BER at different wavelengths.

Fig. 11. Correlation between BER and the number of antennas.

Fig. 10. Relationship of BER and acceleration factor τ.

Fig. 12. Correlation between the BER and training time as well as the number of network layers.

This degradation can be attributed to atmospheric turbulence within the channel, leading to varying degrees of reflection, refraction, scattering, and attenuation of the laser signal. Furthermore, a comparison among different wavelengths reveals that DL method and longer laser wavelengths exhibit superior BER performance and enhanced ISI resistance for the system.

Fig. 10 shows the impact of the acceleration factor on the system's BER performance. As the BER reaches 10^{-4}, reducing the acceleration factor from 1 to 0.9 and further to 0.8 results in a degradation of approximately 2.5 dB and 5 dB, respectively. Similarly, when the BER is 10^{-3}, decreasing the acceleration factor from 1 to 0.9 and then to 0.8 leads to a decline in BER performance by about 3 dB and 4 dB, respectively. It can be inferred from this figure that as acceleration factor decreases, there is a rapid decrease in BER curves observed. Moreover, it is evident that with an acceleration factor of only 0.8, our proposed approach can still ensure satisfactory communication quality. Additionally, the spectrum efficiency has enhanced by approximately 25%.

The BER curves for different numbers of antennas is depicted in Fig. 11. The MIMO-FTN system exhibits a significant reduction in BER compared to the SISO-FTN system. 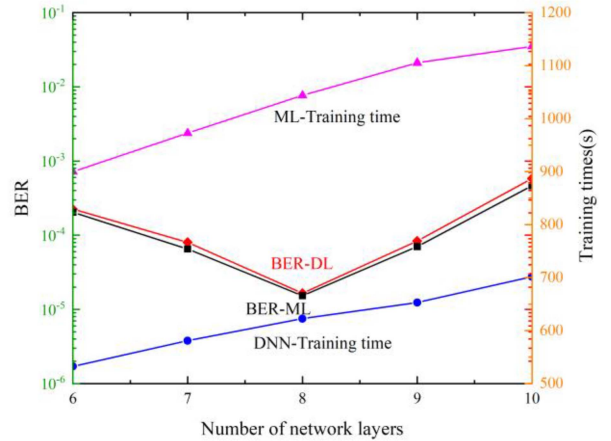At a BER is 10^{-4}, the SNR for 2×2 and 2×4 MIMO-FTN system is reduced by approximately 7.5 dB and 14 dB, respectively, relative to the SISO-FTN system. Hence, increasing the number of antennas can effectively mitigate the impact of atmospheric turbulence on BER. In addition, augmenting the number of antennas at the receiving end confers greater advantages in mitigating the effects of atmospheric turbulence compared to increasing them at the transmitting end. Compared to the massive MIMO method proposed in [26], our proposal shows comparable BER performance in 4×4 MIMO-FTN systems. Additionally, the trained neural network by the [26] can only accommodate a particular channel realization.

Fig. 12 shows the BER performance and required training time of the DNN versus the number of network layers. We can find that when the network is with an increasing number of network layers, the training time becomes larger indicating that it becomes more difficult to train the parameters in the network. Moreover, the BER performance is improved when the number of network layers increases from 6 to 8, since more layers can help capture the interference correlation characteristics among different symbols. However, too many parameters will cause the training result inaccurate due to inefficient back propagation

of gradients in deep networks [43], [44], which leads to a poorer BER. Additionally, studies in [45], [46] have shown that increasing the number of network layers results in a significant rise in computational complexity and overfitting. Overfitting can be categorized into three causes [47]. Firstly, when the training dataset consists of a small number of samples, it may not accurately represent all possible scenarios, resulting in less accurate predictions by the trained network. Therefore, it is important for the training dataset to encompass various types of data as much as possible. Secondly, the network is unable to precisely estimate the relationship between input and output due to excessive interference from training data. Lastly, high complexity within the network requires processing numerous parameters to fit every data point in the training dataset accurately; consequently, this prevents generalization to test datasets. Henceforth, selecting an appropriate number of network layers plays a crucial role in system performance optimization. By taking into account both the BER performance and training time, we can validate that it is reasonable to set the number of network layers to 8. It becomes evident that the performance of DNN approaches that of ML and the complexity reduced by 40%. This can be attributed to ML's exhaustive search over all possible transmitted signals in order to find the global optimal solution, resulting in prohibitively high time complexity. Furthermore, compared to the method presented in [27], our approach exhibits improved performance and achieves a certain level of complexity reduction.

V. CONCLUSION

We propose a DNN decoder for MIMO-FTN signal detection, which utilizes the backpropagation algorithm to compute gradients from the output layer to the input layer and updates each parameter using the chain rule to minimize loss. Simulation results demonstrate that this network exhibits comparable BER performance to the ML method while reducing complexity by 40%. Consequently, this scheme effectively reduces system complexity while ensuring spectral efficiency. In addition, the DL-aided detection method holds promise as a candidate for high-speed FTN-MIMO OWC systems.

REFERENCES

[1] J. Zhou, Q. Sui, and Z. Li, "Non-orthogonal discrete multi-tone: Toward higher spectral efficiency for optical networks," *IEEE Commun. Mag.*, vol. 59, no. 10, pp. 70–75, Oct. 2021.

[2] R. S. Kshetrimayum, M. Mishra, S. Aïssa, S. K. Koul, and M. S. Sharawi, "Diversity order and measure of MIMO antennas in single-user, multiuser, and massive MIMO wireless communications," *IEEE Antennas Wireless Propag. Lett.*, vol. 22, no. 1, pp. 19–23, Jan. 2023.

[3] C. Chen et al., "2D generalized optical spatial modulation for MIMO-OWC systems," *IEEE Photon. J.*, vol. 14, no. 4, Aug. 2022, Art. no. 7343606.

[4] J. E. Mazo, "Faster-than-Nyquist signaling," *Bell Syst. Tech. J.*, vol. 54, no. 8, pp. 1451–1462, Oct. 1975.

[5] L. Li, Z. Xiao, L. Liu, and Y. Lu, "Non-orthogonal WDM systems with faster than Nyquist technology," in *Proc. Opt. Fiber Commun. Conf. Exhib.*, 2019, pp. 1–3.

[6] P. Wang, Y. Zhu, and F. Zhang, "Single carrier time domain FTN signaling enabled by cascaded BPSK iterative detection," in *Proc. Asia Commun. Photon. Conf.*, 2018, pp. 1–3.

[7] Z. Zhang, M. Yuksel, and H. Yanikomeroglu, "Faster-than-Nyquist signaling for MIMO communications," *IEEE Trans. Wireless Commun.*, vol. 22, no. 4, pp. 2379–2392, Apr. 2023.

[8] G. Forney, "Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference," *IEEE Trans. Inf. Theory*, vol. 18, no. 3, pp. 363–378, May, 1972,.

[9] G. Wang, D. Wang, and D. Li, "An efficient ZF-SIC detection algorithm in MIMO CDMA system," in *Proc. IEEE 14th Proc. Pers., Indoor Mobile Radio Commun.*, 2003, pp. 1708–1711.

[10] T.-H. Liu, "Some results for the fast MMSE-SIC detection in spatially multiplexed MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 8, no. 11, pp. 5443–5448, Nov. 2009.

[11] H. Huang et al., "Deep learning for physical-layer 5G wireless techniques: Opportunities, challenges and solutions," *IEEE Wirel Commun*, vol. 27, no. 1, pp. 214–222, Feb. 2020.

[12] T. J. O'Shea, T. Erpek, and T. C. Clancy, "Physical layer deep learning of encodings for the MIMO fading channel," in *Proc. 55th Annu. Allerton Conf. Commun. Control Comput. (Allerton)*, Monticello, IL, USA, 2017, pp. 76–80.

[13] N. Samuel, T. Diskin, and A. Wiesel, "Deep MIMO detection," in *Proc. IEEE 18th Int. Workshop Signal Process. Adv. Wireless Commun.*, 2017, pp. 1–5.

[14] G. Gao, C. Dong, and K. Niu, "Sparsely connected neural network for massive MIMO detection," in *Proc. IEEE 4th Int. Conf. Comput. Commun.*, 2018, pp. 397–402.

[15] V. Corlay, J. J. Boutros, P. Ciblat, and L. Brunel, "Multilevel MIMO detection with deep learning," in *Proc. 52nd Asilomar Conf. Signals, Syst., Comput.*, 2018, pp. 1805–1809.

[16] J. C. D. L. Ducoing, C. Jayawardena, and K. Nikitopoulos, "An assessment of deep learning versus massively parallel, non-linear methods for highly-efficient MIMO detection," *IEEE Access*, vol. 11, pp. 97493–97502, 2023.

[17] H. He, C.-K. Wen, S. Jin, and G. Y. Li, "A model-driven deep learning network for MIMO detection," *IEEE Commun. Lett.*, vol. 24, no. 8, pp. 1724–1728, Aug. 2020.

[18] H. He, C.-K. Wen, S. Jin, and G. Y. Li, "Model-driven deep learning for MIMO detection," *IEEE Trans. Signal Process.*, vol. 68, pp. 1702–1715, 2020.

[19] M. - S. Baek, S. Kwak, J. - Y. Jung, H. M. Kim, and D. - J. Choi, "Implementation methodologies of deep learning-based signal detection for conventional MIMO transmitters," *IEEE Trans. Broadcast.*, vol. 65, no. 3, pp. 636–642, Sep. 2019.

[20] Q. Chen, S. Zhang, S. Xu, and S. Cao, "Efficient MIMO detection with imperfect channel knowledge-A deep learning approach," in *Proc. IEEE Wireless Commun. Netw. Conf.*, 2019, pp. 1–6.

[21] T. Faghani, A. Shojaeifard, K.-K. Wong, and A. H. Aghvami, "Deep learning-based decision region for MIMO detection," in *Proc. IEEE 30th Annu. Int. Symp. Pers., Indoor Mobile Radio Commun.*, 2019, pp. 1–5.

[22] B. Poudel, J. Oshima, H. Kobayashi, and K. Iwashita, "MIMO detection using a deep learning neural network in a mode division multiplexing optical transmission system," *Opt. Commun.*, vol. 440, no. 2019, pp. 41–48, Jun. 2019.

[23] H. Chen, G. Yao, and J. Hu, "Algorithm parameters selection method with deep learning for EP MIMO detector," *IEEE Trans. Veh. Technol.*, vol. 70, no. 10, pp. 10146–10156, Oct. 2021.

[24] X. Jin and H.-N. Kim, "Parallel deep learning detection network in the MIMO channel," *IEEE Commun. Lett.*, vol. 24, no. 1, pp. 126–130, Jan. 2020.

[25] J. Sun, Y. Zhang, J. Xue, and Z. Xu, "Learning to search for MIMO detection," *IEEE Trans. Wireless Commun.*, vol. 19, no. 11, pp. 7571–7584, Aug. 2020.

[26] Q. Wan, J. Fang, Y. Huang, H. Duan, and H. Li, "A variational bayesian inference-inspired unrolled deep network for MIMO detection," *IEEE Trans. Signal Process.*, vol. 70, pp. 423–437, Jan. 2022.

[27] Y. Yu, J. Ying, P. Wang, and L. Guo, "A data-driven deep learning network for massive MIMO detection with high-order QAM," *J. Commun. Netw.*, vol. 25, no. 1, pp. 50–60, Feb. 2023.

[28] Y. Xu, L. Huang, W. Jiang, L. Xue, W. Hu, and L. Yi, "Automatic optimization of volterra equalizer with deep reinforcement learning for intensity-modulated direct-detection optical communications," *J. Lightw. Technol.*, vol. 40, no. 16, pp. 5395–5406, Aug. 2022.

[29] G. A. Mahdiraji and E. Zahedi, "Comparison of selected digital modulation schemes (OOK, PPM and DPIM) for wireless optical communications," in *Proc. IEEE 4th Student Conf. Res. Develop.*, 2006, pp. 5–10.

[30] J. X. Pan, N. Ye, A. H. Wang, and X. M. Li, "A deep learning-aided detection method for FTN-based NOMA," *Wireless Commun. Mobile Comput.*, vol. 2020, pp. 1–11, Jan. 2020.

[31] O. Andrisano and M. Chiani, "The first nyquist criterion applied to coherent receiver design for generalised MSK signals," *IEEE Trans. Commun.*, vol. 42, no. 234, pp. 449–457, Feb-Apr. 1994.

[32] X. Su, S. Wang, H. Yang, J. Jin, and Q. Wang, "Faster-than-nyquist sig-naling: Value and challenges of 6G-oriented applications," *Telecommun. Sci.*, vol. 37, no. 09, pp. 38–47, Sep. 2021.

[33] Y. Kim, H. Lim, and D. Yoon, "On the optimal moment-based estimation of gamma-gamma fading channel parameters," *IEEE Trans. Veh. Technol.*, vol. 71, no. 10, pp. 11240–11244, Oct. 2022.

[34] X. Jin and H.-N. Kim, "Parallel deep learning detection network in the MIMO channel," *IEEE Commun. Lett.*, vol. 24, no. 1, pp. 126–130, Oct. 2020.

[35] L. K. Jones, "Constructive approximations for neural networks by sig-moidal functions," *Proc. IEEE*, vol. 78, no. 10, pp. 1586–1589, Oct. 1990, doi: 10.1109/5.58342.

[36] S. Ou, Z. Guo, and J. Wang, "Multistability of fuzzy neural networks with rectified linear units and state-dependent switching rules," *IEEE Trans. Fuzzy Syst.*, vol. 31, no. 5, pp. 1518–1530, May 2023.

[37] F. Tian, L. Wang, and M. Xi., "Signals recognition by CNN based on attention mechanism," *Electronics*, vol. 11, no. 13, Jul. 2022, Art. no. 2100.

[38] P. Song, F. Gong, Q. Li, G. Li, and H. Ding, "Receiver design for faster-than-Nyquist signaling: Deep-learning-based architectures," *IEEE Access*, vol. 8, pp. 68866–68873, Apr. 2020.

[39] M. Cao, X. Wu, H. Wang, and Q. Peng, "Performance of faster-than-nyquist optical communication system under gamma-gamma atmospheric turbulence," *Chin. J. Lasers*, vol. 47, no. 9, pp. 245–254, Sep. 2020.

[40] F. Wang, W. Du, Q. Yuan, D. Liu, and S. Feng, "A survey of structure of atmospheric turbulence in atmosphere and related turbulent effects," *Atmosphere*, vol. 12, no. 12, Dec. 2021, Art. no. 1608.

[41] N. Badar and R. K. Jha, "Performance comparison of various modu-lation schemes over free space optical (FSO) link employing gamma–Gamma fading model," *Opt. Quantum Electron.*, vol. 49, no. 5, pp. 1–10, May 2017.

[42] A. Garcia-Zambrana, "Bit detect and forward relaying for FSO links using equal gain combining over gamma-gamma atmospheric turbulence chan-nels with pointing errors," *Opt. Exp.*, vol. 20, no. 15, pp. 16394–16409, 2012.

[43] D. Deng, C. Wang, Z. Zhang, L. Li, and W. Wang, "Joint flexible duplexing and power allocation with deep reinforcement learning in cell-free massive MIMO system," *China Commun.*, vol. 20, no. 4, pp. 73–85, Apr. 2023.

[44] Y. Wang, M. Liu, J. Yang, and G. Gui, "Data-driven deep learning for automatic modulation recognition in cognitive ra-dios," *IEEE Trans. Veh. Technol.*, vol. 68, no. 4, pp. 4074–4077, Apr. 2019.

[45] G. S. Yadav, C. Y. Chuang, K. M. Feng, J. Chen, and Y. K. Chen, "Computation efficient sparse DNN nonlinear equalization for IM/DD 112 Gbps PAM4 inter-data center optical interconnects," *Opt. Lett.*, vol. 46, no. 9, pp. 1999–2002, May 2021.

[46] J. Guo et al., "Deep neural network-aided Gaussian message passing detection for ultra-reliable low-latency communications," *Future Gener. Comput. Syst.*, vol. 95, no. 2019, pp. 629–638, Jun. 2019.

[47] M. Bejani and M. Ghatee, "A systematic review on overfitting control in shallow and deep neural networks," *Artif. Intell. Rev.*, vol. 54, no. 8, pp. 6391–6438, Mar. 2021.