



Article

# A Hybrid Network Integrating MHSA and 1D CNN–Bi-LSTM for Interference Mitigation in Faster-than-Nyquist MIMO Optical Wireless Communications

Minghua Cao <sup>1,\*</sup> , Qing Yang <sup>1</sup>, Genxue Zhou <sup>1</sup>, Yue Zhang <sup>1</sup>, Xia Zhang <sup>2,3</sup> and Huiqin Wang <sup>1</sup> 

<sup>1</sup> School of Computer and Communication, Lanzhou University of Technology, Lanzhou 730050, China; 222085402027@lut.edu.cn (Q.Y.); 222085402011@lut.edu.cn (G.Z.); zyue940209@163.com (Y.Z.); whq1222@lut.edu.cn (H.W.)

<sup>2</sup> School of Physical Science and Information Engineering, Liaocheng University, Liaocheng 252000, China; wenerzhang2002@163.com

<sup>3</sup> Shandong Provincial Key Laboratory of Optical Communication Science and Technology, Liaocheng 252000, China

\* Correspondence: caominghua@lut.edu.cn

**Abstract:** To mitigate inter-symbol interference (ISI) caused by Faster-than-Nyquist (FTN) technology in a multiple input multiple output (MIMO) optical wireless communication (OWC) system, we propose an ISI cancellation algorithm that combines multi-head self-attention (MHSA), a one-dimensional convolutional neural network (1D CNN), and bi-directional long short-term memory (Bi-LSTM). This hybrid network extracts data features using 1D CNN and captures sequential information with Bi-LSTM, while incorporating MHSA to comprehensively reduce ISI. We analyze the impact of antenna numbers, acceleration factors, wavelength, and turbulence intensity on the system's bit error rate (BER) performance. Additionally, we compare the waveform graphs and amplitude–frequency characteristics of FTN signals before and after processing, specifically comparing sampled values of four-pulse-amplitude modulation (4PAM) signals with those obtained after ISI cancellation. The simulation results demonstrate that within the Mazo limit for selecting acceleration factors, our proposal achieves a 7 dB improvement in BER compared to the conventional systems without deep learning (DL)-based ISI cancellation algorithms. Furthermore, compared to systems employing a point-by-point elimination adaptive pre-equalization algorithm, our proposal exhibits comparable BER performance to orthogonal transmission systems while reducing computational complexity by 31.15%.

**Keywords:** Faster-than-Nyquist; Bi-LSTM; deep learning; interference cancellation; optical wireless communication



**Citation:** Cao, M.; Yang, Q.; Zhou, G.; Zhang, Y.; Zhang, X.; Wang, H. A Hybrid Network Integrating MHSA and 1D CNN–Bi-LSTM for Interference Mitigation in Faster-than-Nyquist MIMO Optical Wireless Communications. *Photonics* **2024**, *11*, 982. <https://doi.org/10.3390/photonics11100982>

Received: 24 September 2024

Revised: 12 October 2024

Accepted: 17 October 2024

Published: 19 October 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The rapid development of mobile communication technology has greatly improved the performance of OWC systems, which are now widely used in various applications such as intelligent transportation, the Internet of Things, and 5G and future 6G networks [1]. To achieve higher transmission rates and better signal quality, advanced technologies like MIMO have become essential. MIMO maximizes system gain by deploying multiple antennas at both ends to improve spectral efficiency, enhance interference immunity, and maintain high-quality signal transmission in complex environments [2]. Additionally, FTN technology offers new opportunities for increasing OWC system rates by enabling approximately 25% higher data transmission within the same bandwidth through shorter symbol intervals and increased signaling rates [3]. FTN is a non-orthogonal transmission technique that offers distinct advantages over Power Domain Non-Orthogonal Multiple Access (PD-NOMA) and Code Domain Non-Orthogonal Multiple Access (CD-NOMA). FTN achieves higher data transmission rates within the same time by reducing symbol interval, thereby

improving spectral efficiency. In contrast to PD-NOMA, which relies on power allocation and user interference cancellation, FTN primarily focuses on optimizing information transmission in high-bandwidth-demanding scenarios. Additionally, CD-NOMA differentiates users through various coding techniques, which can enhance capacity; however, its performance may be limited by coding complexity and the number of users in the applications that require high spectral efficiency. Therefore, FTNs demonstrate superior performance in specific scenarios which necessitate high-speed data transmission and thus warrant further research and application [4,5]. The implementation of FTN technology, however, poses challenges regarding ISI. When symbol transmission rates exceed the Nyquist rate, non-orthogonal heterodyne pulses can occur, resulting in degraded demodulation performance and impaired signal identification. In high data rate transmissions, signal quality instability may significantly impact communication effectiveness [6]. Therefore, effectively mitigating the ISI introduced by FTNs in MIMO-FTN-OWC systems is crucial as it directly affects signal reliability and user experience [7].

A low-complexity equalization algorithm based on the variational method suggested in [8] exhibits linearly increasing complexity with longer block lengths due to its limited ability to eliminate the ISI caused by strong mutual interference. However, the FTN technique allows for full knowledge of the introduced ISI at the transmitter, eliminating the need for feedback of the channel state information. An alternative approach proposed in [9] is linear pre-equalization, which transforms FTN signals into orthogonal signals at higher baud rates for improved spectral efficiency and to achieve optimal BER performance. For the complete elimination of ISI, ref. [10] proposed an adaptive point-by-point elimination pre-equalization algorithm that calculates interference weights using shaping filters and transmitted symbols for various acceleration factors. It then offsets interference values after FTN shaping on a point-by-point basis based on these weights. However, this method requires increased computational complexity in exchange for higher spectral efficiency and BER performance. In recent years, there has been growing attention towards integrating DL techniques into communications. Notably, ref. [11] utilized a recurrent neural network (RNN) structure based on LSTM networks to eliminate ISI introduced by FTN, achieving similar performance to the Bahl–Cocke–Jelinek–Raviv (BCJR) algorithm. Additionally, ref. [12] introduced an equalization algorithm based on Bi-LSTM networks to address successive ISI elimination between neighboring signals, effectively enhancing system BER performance. Meanwhile, ref. [13] devised a pre-equalization network using the backpropagation (BP) algorithm to eliminate ISI caused by FTN shaping, resulting in a 5 dB improvement compared to the traditional equalization algorithms.

Researching the existing literature revealed that conventional FTN systems' ISI cancellation methods often encounter high computational complexity, necessitating intricate algorithms and substantial computational resources, which limits their practical applications. Although the recent advancements in DL-based approaches have somewhat mitigated this complexity, they still fall short of completely eliminating ISI. Therefore, effectively eradicating ISI while minimizing computational complexity remains a critical challenge. To address this issue, we propose an algorithm that combines 1D CNN, Bi-LSTM, and MHSA to eliminate ISI. This algorithm aims to uncover the hidden features in interfering data and utilize its powerful computational capabilities to directly tackle the ISI introduced by FTN [14–16]. Simulation results indicate that our proposed algorithm achieves up to a 7 dB improvement in BER within the Mazo limit compared to a conventional system without this ISI cancellation algorithm. Furthermore, our algorithm not only maintains comparable BER performance but also reduces computational complexity by 31.15% compared to the point-by-point cancellation adaptive pre-equalization algorithm, nearly matching the performance of an orthogonal system. These results demonstrate an effective enhancement of system performance through our method and provide new insights for signal processing in complex OWC links.

This manuscript is structured as follows: The first section provides a brief introduction to the study's background and significance; the second section explains the system model

and the relevant theoretical foundations; the third section offers a detailed description of the experimental and simulation processes; the fourth section presents a thorough analysis of the results; and finally, the manuscript concludes with a summary.

## 2. System Model

The schematic diagram of the MIMO-FTN-OWC system is illustrated in Figure 1. Assuming the presence of  $N_t$  transmitting antennas and  $N_r$  receiving antennas, the binary data undergo Gray coding and subsequent mapping into a PAM signal with a modulation order of 4. The modulated 4PAM signal  $g(t)$  is then passed through an FTN shaping filter to obtain the FTN signal  $S_{FTN}(t)$ , which is subsequently fed into a trained neural network for ISI mitigation. Following this, the signal traverses a digital-to-analog converter (DAC) and drives a laser diode (LD) to generate an optical signal  $S_T(t)$  that is transmitted via multiple optical antennas through the atmospheric channel. In this system, FTN signals are directly converted to optical signals through the LD after passing through the DAC, eliminating the need for an optical amplifier (e.g., EDFA). The utilization of an optical amplifier can introduce undesirable effects such as nonlinear distortion, increased noise, and gain saturation, which may result in signal distortion, reduced signal-to-noise ratio, and compromised demodulation performance and signal reliability. Therefore, although opting out of an optical amplifier may lead to insufficient signal strength during long-distance transmission or in interference-prone environments, it enhances anti-interference capability by employing a high-precision DAC to minimize signal distortion, selecting a stable and linear laser diode and utilizing a 4PAM modulation method suitable for FTN transmission. Additionally, we apply the MHSA-1D CNN-Bi-LSTM algorithm to mitigate ISI effectively while ensuring high-quality output of optical signals. The optical signal  $S_R(t)$  received by multiple optical antennas at the receiver end are converted into an electrical signal through a photodetector (PD). These electrical signals then go through an analog-to-digital converter (ADC), followed by matched filtering, sampling, and subsequent transmission to a maximum likelihood detector. Finally, the data are recovered after Gray decoding.

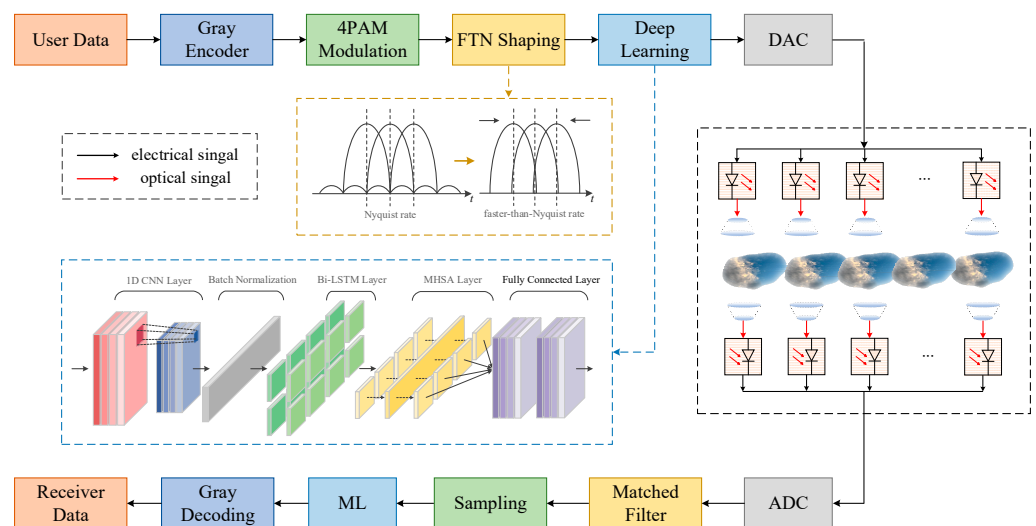


Figure 1. Schematic of MIMO-FTN-OWC system.

In the MIMO-FTN-OWC system, the PAM signal is denoted as  $P = [p_1 p_2 \dots p_n]$ . The system employs a rising cosine filter  $g(t)$  with a roll-off factor  $\psi$ ,  $r$  denotes the number of truncation segments,  $p$  denotes the number of samples per segment,  $T = p$  denotes the Nyquist transmission interval and filter length is  $L = r \times p + 1$ . The vector form of the rising cosine filter  $g(t)$  can be expressed as follows [17]:

$$G = [g_1, g_2, \dots, g_{L-1}, g_L] \tag{1}$$

where  $G$  denotes the peak point of the filter coefficients, which corresponds to the sampling point of the received signal. By multiplying PAM signal values with  $G$  and superimposing them after a delay, the FTN signal  $S_{FTN}(t)$  can be expressed as follows [18]:

$$S_{FTN}(t) = \sum_{i=0}^n p_i g(t - (i - 1)\tau T), 0 < \tau < 1 \tag{2}$$

where  $n$  denotes the number of modulated symbols transmitted per frame,  $p_i \in P$  denotes the  $i$ -th symbol,  $P = \{a_m = 2m - 1 - U, m = 1, 2, \dots, U\}$ ,  $U$  denotes the modulation order of PAM,  $\tau T$  is the delay interval,  $\tau = p'/p$  denotes the acceleration factor,  $p'$  denotes the sampling factor when shaping FTN signal.

The FTN signal is subsequently input into a trained neural network that combines 1D CNN, the Bi-LSTM network, and the MHSA mechanism. The 1D CNN extracts crucial features from the data and arranges them into sequences for the Bi-LSTM network to learn from. Additionally, the MHSA mechanism assigns extra weight to enhance learning outcomes and ensures that the pre-processed de-interference signal  $S_{FTN-DL}(t)$  closely resembles the original signal. Subsequently,  $S_{FTN-DL}(t)$  is converted to the digital-to-analog format before driving the LDs to generate an optical signal  $S_T(t)$ . Upon propagation through the atmospheric channel, the received signal can be written as follows [19]:

$$S_R(t) = h \cdot S_T(t) + n(t) \tag{3}$$

where  $h$  denotes the channel fading coefficient, and  $n(t)$  denotes the additive noise.

The channel fading coefficients exhibit a Gamma–Gamma distribution and their probability density function can be mathematically formulated as follows [20]:

$$\varphi_h = \frac{2(\alpha\beta)^{\frac{\alpha+\beta}{2}}}{\Gamma(\alpha)\Gamma(\beta)} \cdot h^{\frac{\alpha+\beta}{2}-1} \cdot N_{\alpha-\beta} \cdot (2\sqrt{\alpha\beta h}) \tag{4}$$

where  $N_{\alpha-\beta}(\cdot)$  denotes the second-class modified Bessel function of order  $\alpha - \beta$ ,  $\Gamma(\cdot)$  denotes the gamma function,  $\alpha$  denotes the large-scale scattering coefficient, and  $\beta$  denotes the small-scale scattering coefficient, which can be respectively expressed as follows:

$$\alpha = \left\{ \exp \left[ \frac{0.49\sigma^2}{(1 + 1.11\sigma^{12/5})^{7/6}} \right] - 1 \right\}^{-1} \tag{5}$$

$$\beta = \left\{ \exp \left[ \frac{0.51\sigma^2}{(1 + 0.69\sigma^{12/5})^{5/6}} \right] - 1 \right\}^{-1} \tag{6}$$

where  $\sigma^2 = 0.5C_n^2 k^{7/6} l^{11/6}$  denotes the Rytov variance,  $C_n^2$  denotes the atmospheric refractive index structure constant,  $k = 2\pi/\lambda$ ,  $\lambda$  denotes the laser wavelength, and  $l$  denotes the laser transmission distance.

At the receiver end, the signal received  $S_R(t)$  from multiple optical antennas is converted into an electrical signal by PDs. Subsequently, it passes through a matched filter and is sampled at intervals of  $\tau T$ . The vector representation of the sampled signal can be expressed as follows:

$$Y = [y_1, y_2, \dots, y_n] = [y_{r \times p/2+1}, y_{r \times p/2+1+p'}, \dots, y_{r \times p/2+1+(n-1) \times p'}] \tag{7}$$

where  $y_n$  denotes the sampled symbol,  $Y = y_1 + y_2 + \dots + y_n$  indicates that the signals from all the receiving antennas are combined with equal gain. Subsequently, the sampled signal  $Y$  is subjected to maximum likelihood sequence detection, followed by the demapping process for the restoration of user data.

### 3. MHSA-1D CNN-Bi-LSTM Interference Cancellation Scheme

#### 3.1. 1D CNN

In the MIMO-FTN-OWC system, a computational method can be employed to determine the ISI value for each signal pulse when  $\tau$  is known. Utilizing this information, we initially utilize the ISI-free signal  $S_{FTN}'(t)$  as a labeled sequence. Subsequently, a one-dimensional convolutional (Conv1D) layer is employed to train the FTN signal sequence and derive the optimal convolutional kernels that minimize loss. These kernels are then utilized to automatically extract intricate FTN-ISI features. The maximum pooling layer further consolidates and preserves these complex features extracted by the convolutional layers. Finally, the highest values from each convolutional layer's feature vectors are merged to form the ultimate feature values, as depicted in Figure 2.

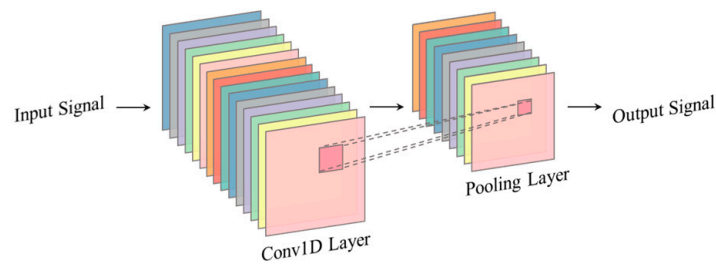


Figure 2. The structure of 1D CNN (Different colors represent different filters).

#### 3.2. Bi-LSTM

The Bi-LSTM is an extension of the LSTM architecture designed to handle data with temporal dependencies by utilizing two independent input sequences that are fed into the same LSTM layer. The forward and reverse LSTMs of the Bi-LSTM are trained using FTN signal sequence data, resulting in a structure consisting of an input layer, a forward hidden layer, a reverse hidden layer, and an output layer as depicted in Figure 3.

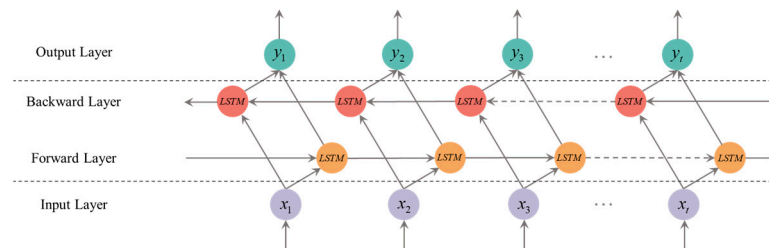


Figure 3. The structure of Bi-LSTM.

The specific implementation of the Bi-LSTM algorithm is detailed in Table 1 [21]. The Bi-LSTM algorithm processes time-series data through an input sequence  $X = [x_1, x_2, \dots, x_T]$ , where each feature vector  $x_t$  is mapped to both forward and backward LSTM cells at time step  $t$ . The core of the LSTM is its gating mechanism, which includes input gates  $i$ , forget gates  $f$ , output gates  $o$ , and candidate cell states  $g$ . These gates utilize weight matrices  $W$  and  $U$  along with bias  $b$  to regulate the flow and storage of information. The forward and backward hidden states  $h_{forward}[t]$  and  $h_{backward}[t]$  together generate the output  $h[t]$  of the bidirectional LSTM, effectively capturing both forward and backward dependencies in the sequence data while handling nonlinear relationships through activation functions such as Sigmoid and hyperbolic tangent (tanh).

**Table 1.** Bi-LSTM computational procedure.

Algorithm	Bi-LSTM
<b>Start</b>	
	Input: Input Sequence $X = [x_1, x_2, \dots, x_T]$ , $T$ is the length of the sequence. Parameters: LSTM cell parameters $(W, U, b)$ , $W, U$ is weights and $b$ is biases. Output: Output Sequence $H = [h_1, h_2, \dots, h_T]$
	<b>Step 1: Initialize hidden and cell states.</b> $h_{forward}[0] = 0,$ $c_{forward}[0] = 0,$ $h_{backward}[0] = 0,$ $c_{backward}[0] = 0.$
	<b>Step 2: Forward propagation.</b> for $t = 1$ to $T$ do, Loop statement that iterates from 1 to $T$ . Calculate the output of forward LSTM: $i_{forward}[t] = \sigma(W_i * x_t + U_i * h_{forward}[t-1] + b_i),$ $f_{forward}[t] = \sigma(W_f * x_t + U_f * h_{forward}[t-1] + b_f),$ $o_{forward}[t] = \sigma(W_o * x_t + U_o * h_{forward}[t-1] + b_o),$ $g_{forward}[t] = \tanh(W_g * x_t + U_g * h_{forward}[t-1] + b_g),$  $c_{forward}[t] = f_{forward}[t] * c_{forward}[t-1] + i_{forward}[t] * g_{forward}[t],$ $h_{forward}[t] = o_{forward}[t] * \tanh(c_{forward}[t]).$
	<b>Step 3: Backward propagation.</b> for $t = T$ to 1 do, Loop statement that iterates from $T$ to 1. Calculate the output of backward LSTM: $i_{backward}[t] = \sigma(W_i * x_t + U_i * h_{backward}[t+1] + b_i),$ $f_{backward}[t] = \sigma(W_f * x_t + U_f * h_{backward}[t+1] + b_f),$ $o_{backward}[t] = \sigma(W_o * x_t + U_o * h_{backward}[t+1] + b_o),$ $g_{backward}[t] = \tanh(W_g * x_t + U_g * h_{backward}[t+1] + b_g),$  $c_{backward}[t] = f_{backward}[t] * c_{backward}[t+1] + i_{backward}[t] * g_{backward}[t],$ $h_{backward}[t] = o_{backward}[t] * \tanh(c_{backward}[t]).$
	<b>Step 4: Combine forward and backward outputs.</b> for $t = 1$ to $T$ do, $h[t] = \text{concat}(h_{forward}[t], h_{backward}[t]),$ Splice forward and backward hidden states. return to $H = [h_1, h_2, \dots, h_T]$ . <b>End</b>

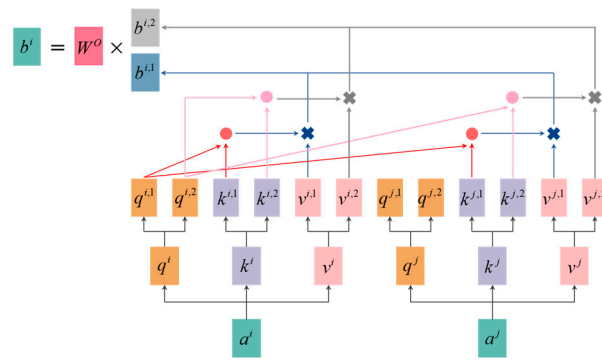
### 3.3. Multi-Head Self-Attention

The MHSA mechanism is a variant of the self-attention mechanism, designed to enhance the expressive and generalization abilities of the model. Multiple attention computations are performed in parallel on the signal sequence, organized by a “head” structure. If there are  $h$  “head” structures,  $Q$ ,  $K$ , and  $V$  undergo linear transformations and are allocated to each head, with  $1/h$  of the original dimension during the initial modeling stage. Subsequently, each “head” independently calculates attention using its allocated data. The computed results are then concatenated along a specific dimension and transformed through a linear operation to get the final attention distribution. The process can be expressed as in the following equations [22]:

$$head_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) \quad (8)$$

$$\text{MultiHead}(Q, K, V) = \text{Concat}(head_1, head_2, \dots, head_h)W^O \quad (9)$$

where  $W_i^Q$ ,  $W_i^K$ ,  $W_i^V$ , and  $W^O$  are learnable linear transformation parameter matrices. The two-head self-attention mechanism is depicted in Figure 4.

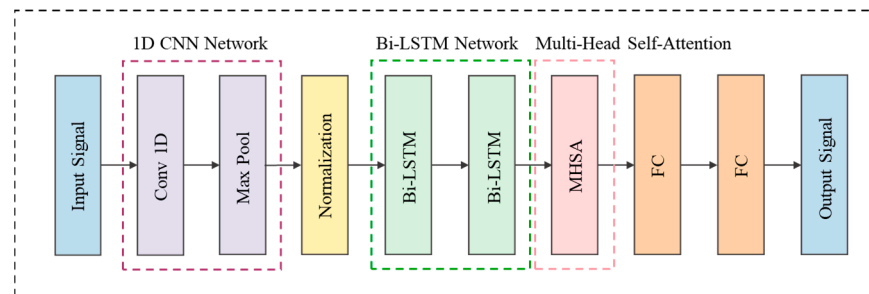


**Figure 4.** Calculation process of the double-head self-attention mechanism.

The distinction between the MHSA mechanism and the self-attention mechanism lies in the fact that the latter only necessitates a single  $W^Q$ ,  $W^K$ , and  $W^V$  for computing Q, K, and V, whereas the former requires an additional computation of computed Q, K, and V. Taking  $q^i$  as an example,  $q^{i,1} = W^{q,1}q^i$ ,  $q^{i,2} = W^{q,2}q^i$  is initially computed, followed by the computation of  $b^{i,1}$  using  $q^{i,1}, k^{i,1}, v^{i,1}$  and  $q^{i,1}, k^{i,1}, v^{i,1}$ . Subsequently,  $b^{i,2}$  is computed using  $q^{i,2}, k^{i,2}, v^{i,2}$  and  $q^{i,2}, k^{i,2}, v^{i,2}$ . Then,  $b^{i,1}, b^{i,2}$  are concatenated to obtain  $b^i$  through matrix multiplication with  $W^O$ . Similarly,  $b^j$  can be obtained following this procedure, thereby completing the computation of two-head self-attention mechanism [23].

### 3.4. MHSA-1D CNN-Bi-LSTM Interference Cancellation Algorithm

The signal sequence typically allows for extracting only a limited number of features by a single model. To effectively mitigate the ISI introduced by the FTN, we propose a hybrid model, namely MHSA-1D CNN-Bi-LSTM, which combines 1D CNN, a Bi-LSTM network, and an MHSA mechanism. The network structure is illustrated in Figure 5.



**Figure 5.** The structure of MHSA-1D CNN-Bi-LSTM network (Blue represents the input and output signals; purple represents the 1D-CNN; light yellow represents the normalization layer; green represents the Bi-LSTM network; red represents the multi-head self-attention mechanism layer; and orange represents the fully connected layer).

Initially, the FTN signal sequence  $S_{FTN}(t)$  is utilized as input for the neural network. A 1D CNN is utilized to extract and organize feature vectors associated with information such as ISI and noise into sequences. Subsequently, the Bi-LSTM network is split into two independent LSTM networks, where these identical LSTMs generate forward and reverse hidden state sequences from the input sequences. The LSTM selectively forgets information through the forgetting gate using the previous time node’s output  $h_{forward}[t]$  and the current time node’s input  $x_t$ . It then determines which new information will be stored in the current candidate memory cell  $c_{forward}[t]$  to obtain the latest memory cell  $c_{backward}[t]$ . The predecessor memory cell  $c_{backward}[t - 1]$  is updated accordingly. By combining output value  $o_t$  of the output gate with the current moment memory cell state  $c_{backward}[t]$ , we get the LSTM’s output value  $h_{backward}[t]$ . This is followed by a linear transformation with a weight matrix dimension matching that of the input data dimension, ensuring consistent dimensions for subsequent operations. An MHSA layer is introduced for a “head-splitting”

operation after this step. Through the MHSA mechanism, distinct weights are assigned to each LSTM unit’s output  $[h_1, h_2, \dots, h_T]$ , enabling selective attention to different time steps during sequence processing. This facilitates the efficient selection of prediction information relevant to achieving current task goals. Finally, operations are performed within each self-attention mechanism and their outputs are concatenated for linear transformations before normalization.

#### 4. Simulation and Analysis

The system is implemented in Python 3.8, utilizing Keras framework with TensorFlow serving as the backend for tensor operations. Both the training and test datasets are generated randomly using MATLAB R2022a, while the simulation parameters are presented in Table 2.

**Table 2.** Simulation parameters.

Parameter	Value
size of data	$2 \times 10^6$
training dataset	$1.6 \times 10^6$
test dataset	$4 \times 10^5$
batch size	256
cycle index	100
dropout	0.3
learning rate	0.0001
loss function	MSE
optimizer	Adam
atmospheric refractive [24]	strong: $1.13 \times 10^{-13} \text{ m}^{-2/3}$
	medium: $1.13 \times 10^{-14} \text{ m}^{-2/3}$
	weak: $1.13 \times 10^{-17} \text{ m}^{-2/3}$

The MHSA–1D CNN–Bi-LSTM hybrid model consists of a Conv1D layer with 64 filters, kernel size of six, stride of one, and the “same” padding. ReLU activation is applied post-convolution to enhance learning performance. L2 regularization with a weight decay factor 0.01 is utilized to mitigate overfitting and enhance generalization during the training process. A subsequent max pooling layer follows the convolutional layer with a window size of two and stride of one. Power normalization is utilized to compress symbols to simulate hardware constraints accurately. Two Bi-LSTM layers ensue, where the first consists of 100 neurons and the second has only 20 neurons. In the MHSA–CNN–Bi-LSTM hybrid model, we set the dropout rate to 0.3 based on our experimental observations, striking a favorable balance between enhancing model performance and mitigating overfitting. Our cross-validation results indicate that lower dropout rates fail to adequately suppress overfitting, while higher rates may result in information loss. Therefore, by retaining adequate information and reducing the risk of overfitting, a dropout rate of 0.3 ensures the model’s ability to generalize effectively to unknown data. Subsequently, an attention layer employing a two-headed self-attention mechanism employs a total of 64 neurons for enhanced representation capability. Finally, the output layer encompasses two fully connected layers, with one comprising 10 neurons and another consisting solely of 1 neuron responsible for generating the final result. During the training process, we load the labeled dataset from a CSV file, which includes both the pre-FTN shaping data and the disturbed post-FTN shaping data. To enhance the effectiveness of model training, we normalize the disturbed data to ensure a mean of 0 and a standard deviation of 1. Subsequently, we divide the dataset into training and test sets in an 80/20 ratio. Mean square error is employed as the loss function during training, optimized using the Adam optimizer. The model undergoes 100 training epochs while monitoring both training loss and validation loss to assess convergence and generalization capability. Upon completion of training, we evaluate predictive performance using metrics such as  $R^2$ , root mean square error (RMSE), and mean absolute error (MAE) to demonstrate its efficacy in handling disturbed signals



and further validate its potential. Finally, for future analysis and application purposes, we save both the trained model and its normalizer.

#### 4.1. Evaluation Index

The performance evaluation encompasses the following three selected metrics: RMSE, MAE, and  $R^2$ , as presented in Table 3.

**Table 3.** Model evaluation indicators.

Indicator Name	Indicator Meaning	Formula
RMSE	The square root of the ratio of the square of the deviation of the predicted value from the true value to the number of observations [25].	$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$
MAE	The average of the absolute values of the deviations of all individual observations from the arithmetic mean [26].	$MAE = \frac{1}{n} \left( \sum_{i=1}^n  y_i - \hat{y}_i  \right)$
$R^2$	The degree of fit of the regression line to the observations [27].	$R^2 = 1 - \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (\bar{y}_i - y_i)^2}$

where  $y_i$  denotes the true value,  $\hat{y}_i$  denotes the predicted value,  $\bar{y}_i$  denotes the average of all true values in the test set and  $n$  denotes the number of samples in the test set within  $[0, +\infty]$ . A smaller RMSE and MAE indicate better prediction accuracy for the model. Similarly, a higher  $R^2$  value indicates enhanced discriminative ability for the model.

Additionally, the comparative analysis encompasses LSTM, Bi-LSTM, CNN–Bi-LSTM models, in addition to the proposed MHSA–1D CNN–Bi-LSTM hybrid model. The specific outcomes are presented in Table 4.

**Table 4.** Performance comparison of different models.

Model	RMSE	MAE	$R^2$
LSTM	0.68733	0.56313	0.62213
Bi-LSTM	0.64597	0.55962	0.66623
CNN–Bi-LSTM	0.36303	0.31812	0.89458
MHSA–1D CNN–Bi-LSTM	0.31499	0.24220	0.92060

The superiority of the proposed MHSA–1D CNN–Bi-LSTM hybrid model over the LSTM, Bi-LSTM, CNN–Bi-LSTM models is evident, as demonstrated by the significantly improved results presented in Table 4.

#### 4.2. System Performance Analysis

Figure 6 illustrates the amplitude–frequency characteristics of various signals. Upon observing Figure 6, it is evident that the interference cancellation algorithm significantly improves the amplitude–frequency characteristics of the signal. However, there is a slight deviation from those of the 4PAM signal. This discrepancy arises because, although the MHSA–1D CNN–Bi-LSTM interference cancellation algorithm effectively eliminates interference values at signal sampling points, residual interference values still exist at other moments outside these sampling points. Therefore, it can be concluded that the proposed algorithm exclusively eliminates interference affecting the target value of signal amplitude during transmission without altering the overall duration of FTN signals.

The sampled values of three different signals are presented in Figure 7. Figure 7a illustrates the sampled values of the 4PAM signal (prior to FTN shaping filtering), while Figure 7b displays the sampled values of the signal after FTN shaping filtering. Additionally, Figure 7c presents the sampled values of the FTN signal following the application of our proposed interference cancellation algorithm. A comparison between Figures 7a and 7b clearly reveals the presence of ISI introduced by employing FTN technique for modulation purposes. In contrast, a point-by-point comparison between Figures 7a and 7c demonstrates consistent amplitudes for almost all corresponding points, indicating that our

interference cancellation algorithm effectively eliminates the ISI caused by the FTN modulation technique, thereby showcasing its potential as a robust solution for enhancing reliability in FTN communication systems.

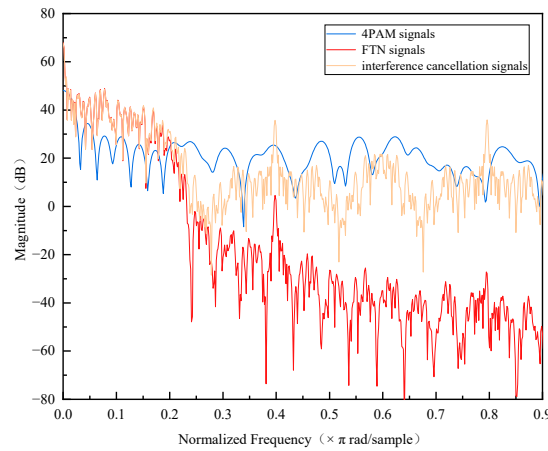


Figure 6. Magnitude–frequency characteristics.

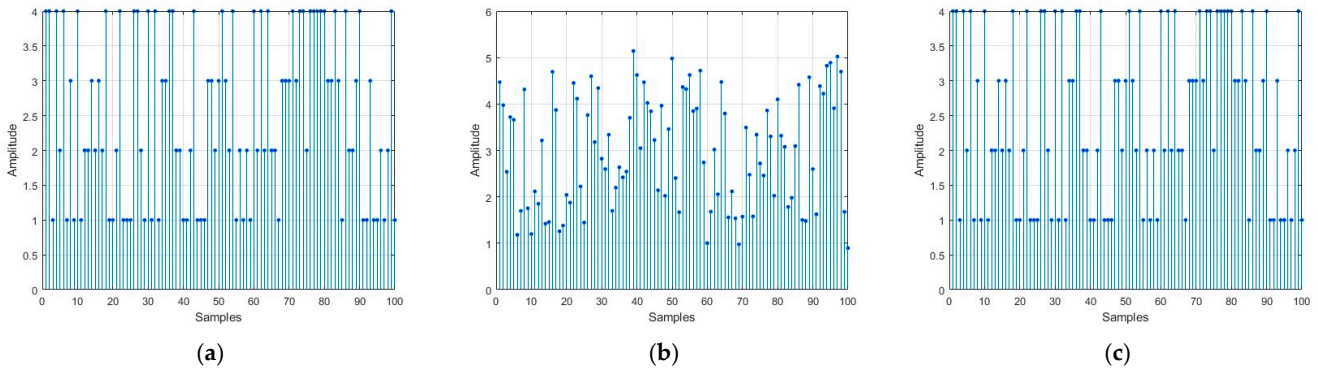
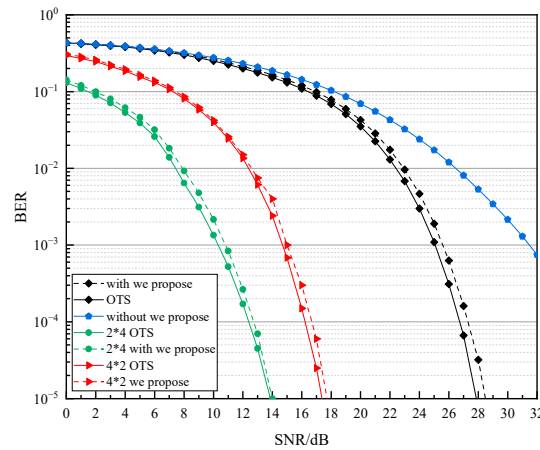


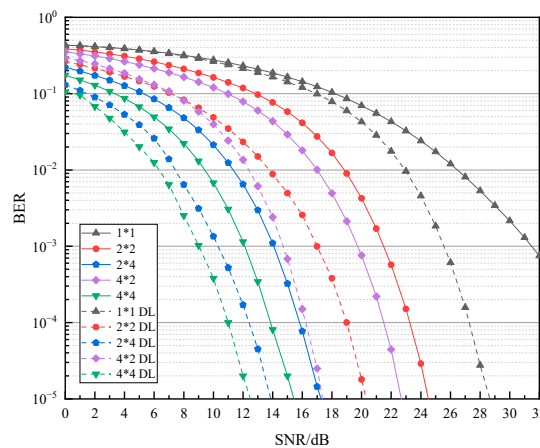
Figure 7. Comparison of signal sampling values (a) 4PAM signals (b) FTN signals (c) signals after ISI cancellation.

Figure 8 illustrates the BER plots for various transceiver antenna configurations, with and without the interference cancellation algorithm, under medium turbulence channel conditions with  $\tau = 0.8$ ,  $\lambda = 1550 \text{ nm}$ ,  $ipce = 0.5$ .  $ipce$  denotes the photoelectric conversion efficiency and is set to 0.5 to represent typical photodetector performance across diverse environmental conditions. This selection effectively accounts for optical signal attenuation and noise effects, ensuring system reliability under varying channel conditions. Moreover, opting for a value of 0.5 strikes a balance between signal quality and system complexity, thereby enhancing overall performance. Consequently, this setting guarantees stable and efficient model operation in real-world applications. Firstly, our proposed algorithm demonstrates comparable performance to the orthogonal system across all antenna configurations, indicating its effectiveness in eliminating ISI. Experimental results on BER performance with  $1 \times 1$ ,  $2 \times 4$ , and  $4 \times 2$  transceiver antenna configurations confirm that both the algorithms exhibit nearly identical performances, thereby showcasing their equivalent diversity gain. This equivalence is primarily attributed to their effective utilization of the spatial resources provided by multiple antennas under similar channel conditions while maintaining signal independence and employing comparable signal processing strategies. Consequently, they achieve similar BER performance with reduced signal fading effects. Notably, in the  $1 \times 1$  transceiver antenna configuration, the MIMO-FTN-OWC system without the interference cancellation algorithm achieves an approximate signal-to-noise ratio (SNR) of 31 dB compared to around 27 dB when utilizing the algorithm at a  $BER = 10^{-3}$ . Therefore, the proposed algorithm yields a coding gain of 4 dB.



**Figure 8.** Relationship between BER and SNR for our proposal and orthogonal system (OTS refers to Orthogonal Transmission System).

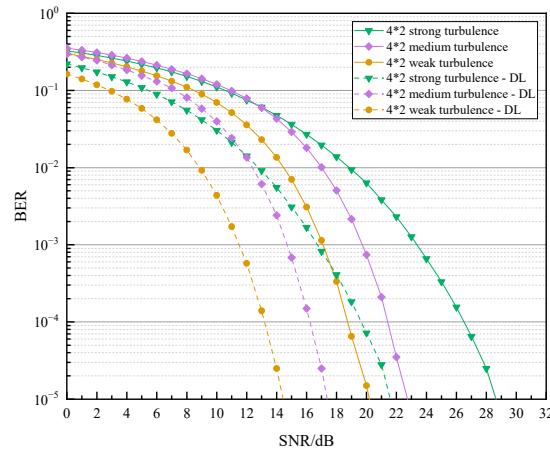
Figure 9 demonstrates the BER performance under medium turbulence conditions within the Gamma–Gamma channel for different numbers of transceiver antennas. It compares scenarios with and without the utilization of the proposed algorithm, at  $\tau = 0.8$ ,  $ipce = 0.5$ , and  $\lambda = 1550$  nm. Firstly, in the Single Input Single Output (SISO) system, employing the DL ISI cancellation algorithm improves system performance by 6 dB at  $BER = 10^{-3}$  compared to its absence. In the MIMO system, with a fixed number of two transmitting antennas and varying receiving antennas set to two and four, respectively, utilizing the proposed algorithm improves system performance by 4.5 dB and 3 dB at  $BER = 10^{-5}$  compared to its absence. Similarly, when four transmitting antennas are fixed while the receiving antennas are set to two and four, respectively, applying the proposed ISI cancellation algorithm improves system performance by 5.5 dB and 3 dB at  $BER = 10^{-5}$  compared to its non-usage scenario. All aforementioned simulation results effectively demonstrate that our proposed MHSA–1D CNN–Bi-LSTM hybrid network successfully mitigates eliminates ISI introduced by FTN across various numbers of transceiver antennas, thereby significantly enhancing overall system performance.



**Figure 9.** Relationship between BER and SNR under different numbers of antennas.

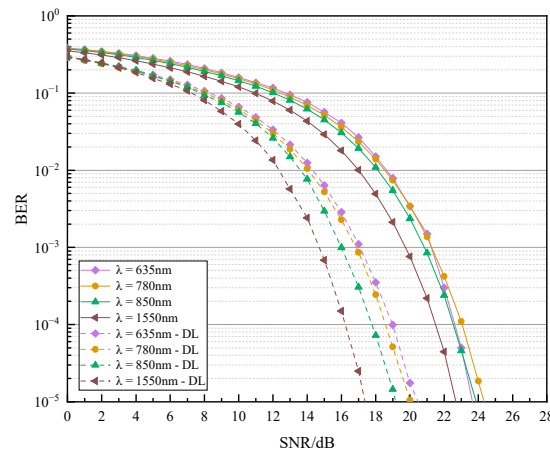
The optimal system performance is observed with a transceiver antenna configuration of  $4 \times 2$ , as shown in Figure 9. Therefore, this configuration is selected to investigate the algorithm’s performance under varying turbulence intensities. Figure 10 presents a comparison of the BER performance with and without the proposed algorithm across weak, medium, and strong turbulence intensities, at  $\tau = 0.8$ ,  $ipce = 0.5$ , and  $\lambda = 1550$  nm. From the figure, it can be observed that utilizing the proposed ISI cancellation algorithm leads to an improvement in the system performance by 6.5 dB, 5.5 dB, and 7 dB, respectively, at

$BER = 10^{-5}$ . This clearly demonstrates that the BER performance consistently surpasses that of the system without employing this algorithm.



**Figure 10.** Relationship between BER and SNR under different turbulence intensities.

Under Gamma–Gamma medium turbulence conditions with an  $4 \times 2$  antenna configuration, Figure 11 presents a comparison of the BER performance at different laser wavelengths with and without the proposed algorithm, when  $\tau = 0.8$  and  $ipce = 0.5$ . At laser wavelengths of 1550 nm, 850 nm, 780 nm, and 635 nm, respectively, the system’s performance is improved by 5.5 dB, 5 dB, 4.2 dB, and 3.4 dB at  $BER = 10^{-5}$ . The results indicate that as the laser wavelength increases, there is a decrease in BER along with an improved ability of the algorithm to eliminate interference.



**Figure 11.** Relationship between BER and SNR under different laser wavelengths.

The effectiveness of the proposed ISI cancellation algorithm is verified at different acceleration factors in Figure 12, comparing the system performance with and without the proposed algorithm for four acceleration factors (0.5, 0.6, 0.7, and 0.8) under Gamma–Gamma medium turbulence conditions with a  $4 \times 2$  antenna configuration at  $ipce = 0.5$  and  $\lambda = 1550$  nm. Firstly, in the absence of the ISI cancellation algorithm and for  $\tau < 0.8$ , a significant occurrence of ISI arises due to the close overlap of waveforms, resulting in a degradation of normal communication performance. However, upon application of the algorithm, there is a notable enhancement in the BER performance. Specifically, when  $\tau = 0.8$ , there is an improvement in system performance by 5.5 dB; and at  $\tau = 0.7$  and  $\tau = 0.6$ , the ISI cancellation algorithm elevates the system SNR to 18 dB and 19.4 dB, respectively, at  $BER = 10^{-5}$ , thereby restoring normal communication capability. It is also noted that when  $\tau = 0.5$ , employing the ISI cancellation algorithm has minimal impact on system performance. The simulation results indicate that this algorithm significantly

enhances system performance particularly within a range of  $\tau$  values between 0.6 and 0.8, with smaller values yielding better interference cancellation.

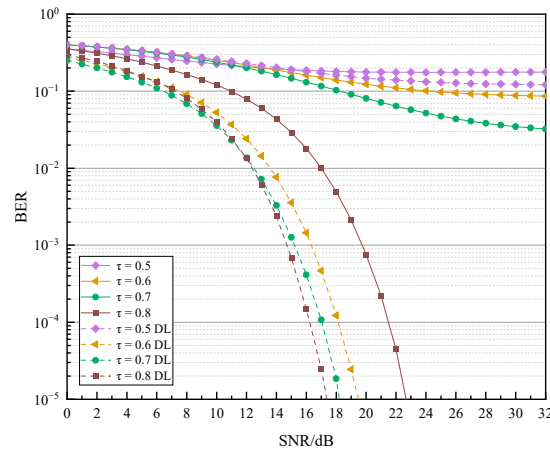


Figure 12. Relationship between BER and SNR under different acceleration factors.

To further validate the impact of the acceleration factor on the interference cancellation capability, Figure 13 shows the system BER with and without the ISI cancellation algorithm under Gamma–Gamma moderate turbulence conditions, using a  $4 \times 2$  antenna configuration at  $ipce = 0.5$ ,  $\lambda = 1550$  nm, and SNR of 15 dB. The results demonstrate that for  $\tau \geq 0.5$ , there is an improvement in the cancellation effect of the algorithm as  $\tau$  decreases. However, when  $\tau < 0.5$ , it is observed that the algorithm becomes less effective in mitigating ISI introduced by the FTN, resulting in significant performance degradation. The acceleration factor below 0.5 significantly increases ISI due to increased symbol overlap. A critical threshold is reached at an acceleration factor of 0.5, where the wider pulse shape reduces the symbol overlap and weakens the effect of ISI on demodulation. Conversely, when the acceleration factor drops below 0.5, the narrower pulse width leads to abnormally high ISI with weakened effectiveness of the interference cancellation algorithm. On the other hand, within the range of 0.5 to 1, our proposed algorithm effectively eliminates ISI, and its effectiveness improves as ISI increases.

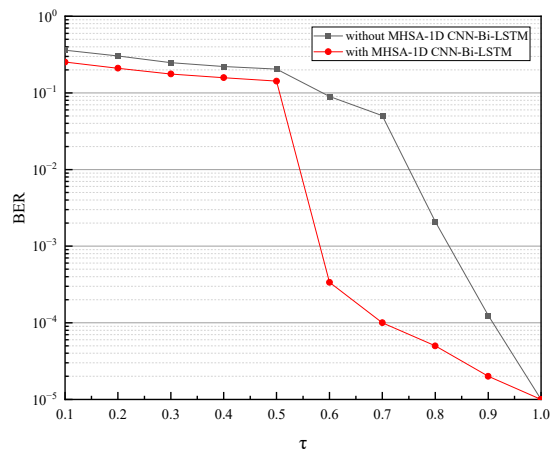


Figure 13. Relationship between acceleration factor and BER.

### 4.3. Computational Complexity

The number of floating-point operations serves as a complexity metric for comparing the computational complexity between the proposed algorithm and the point-by-point cancellation adaptive pre-equalization algorithm [10]. The computational complexity of the point-by-point cancellation adaptive pre-equalization algorithm is as follows:

$$Com_1 = n \times L + (L - \tau \times p' + n \times \tau \times p') \times 2 + n \times 4 + [v + (2 \times v - 1)] \times v + (n - 2 \times v) \times 2 \times v + n \times L + (L - \tau \times p' + n \times \tau \times p') + n \times 2 + [v + (2 \times v - 2)] \times v + (n - 2 \times v) \times (2 \times v - 1) \tag{10}$$

where  $v$  denotes the number of signs of the trailing influence.

The computational complexity of our proposal is as follows:

$$Com_2 = O(n \times k) + O(3n) + O(n \times d_1) + O(n \times d_2) + O(n \times m_1) + O(n \times m_2) + O(n \times m_3) + v \times (2 \times n - 1 - v) + v \times (3 \times v - 2) + (n - 2 \times v) \times (2 \times v - 1) \tag{11}$$

where  $k$  denotes the size of the convolution kernel of the Conv1D,  $d_1$  and  $d_2$  denote the number of neurons of the two Bi-LSTM, respectively.  $m_1$  denotes the number of neurons of the multi-head attentional layer,  $m_2$  denotes the number of neurons of the fully connected layer, and  $m_3$  denotes the number of neurons of the output layer. When  $f = 6$ ,  $p' = 10$ ,  $L = 61$ ,  $k = 6$ ,  $d_1 = 100$ ,  $d_2 = 20$ ,  $m_1 = 64$ ,  $m_2 = 10$ , and  $m_3 = 1$ , the relationship between the acceleration factor  $\tau$  and computational complexity is shown in Table 5.

**Table 5.** Computational complexity.

Algorithm	$\tau$			
	0.6	0.7	0.8	0.9
Algorithm proposed in [10]	$4.21 \times 10^8$	$4.17 \times 10^8$	$4.12 \times 10^8$	$4.09 \times 10^8$
MHSA-1D CNN-Bi-LSTM interference cancellation algorithm	$2.98 \times 10^8$	$2.95 \times 10^8$	$2.77 \times 10^8$	$2.75 \times 10^8$

The results in Table 5 demonstrate a reduction in the computational complexity of the proposed algorithm by 32.74% compared to the point-by-point elimination adaptive pre-equalization algorithm for  $\tau = 0.9$ , 33.33% for  $\tau = 0.8$ , 29.28% for  $\tau = 0.7$ , and 29.24% for  $\tau = 0.6$ . On average, the overall computational complexity decreased by 31.15%.

### 5. Conclusions

A novel MHSA-1D CNN-Bi-LSTM algorithm is proposed to mitigate the ISI caused by the FTN in MIMO-FTN-OWC systems. The algorithm integrates 1D CNN, Bi-LSTM, and MHSA to effectively capture interference signal features, handle long sequence data, and assign importance weights to features, respectively. Simulation results demonstrate that the proposed ISI cancellation algorithm achieves a BER improvement of up to 7 dB within the Mazo limit compared to a conventional system without the algorithm. This significant enhancement validates the effectiveness of the algorithm in addressing ISI. Moreover, the algorithm not only maintains comparable BER performance with the point-by-point cancellation adaptive pre-equalization algorithm but also reduces computational complexity by 31.15%. These findings offer new insights and practical guidance for signal processing in complex OWC channels and highlight its potential for improving system performance. We will further investigate the algorithm’s applicability in more intricate environments, such as dynamic channel conditions and multi-user scenarios. Moreover, we will explore emerging approaches that integrate DL techniques to enhance system robustness and efficiency. These efforts aim to provide stronger technical support and broaden application potential for advancements in OWC systems.

**Author Contributions:** Conceptualization, M.C., Q.Y., G.Z. and H.W.; Methodology, M.C., H.W. and Q.Y.; Validation, M.C., Q.Y. and G.Z.; Investigation, G.Z. and Y.Z.; Resources, Q.Y. and X.Z.; Data Curation, X.Z.; Writing—Original Draft Preparation, G.Z. and Q.Y.; Writing—Review and Editing, M.C., H.W., Q.Y. and Y.Z.; Visualization, Q.Y. and H.W.; Supervision, M.C. and H.W.; Project Administration, M.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by NSFC Program under Grant 62265010, 62261033 and Natural Science Foundation of Gansu Province, China (24JRRA183) and Youth Science and Technology Foundation of Gansu Province, China (24JRRA974) and Natural Science Foundation of Shandong Province, China (ZR2022MF284).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

**Acknowledgments:** We gratefully acknowledge the assistance of Genxue Zhou in preparing and debugging the program. We also thank Ruifang Yao and Yue Zhang for help with the English writing and details. All authors have consented to and are informed by this acknowledgement.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Dash, S.P.; Kaushik, A. Ris-assisted 6g wireless communications: A novel statistical framework in the presence of direct channel. *IEEE Commun. Lett.* **2024**, *28*, 717–721. [[CrossRef](#)]
2. Jeyapal, R.; Matrouk, K.; Purushothaman, D. An efficient heuristic-aided adaptive autoencoder-based dilated DNN with attention mechanism for enhancing the performance of the MIMO system in 5G communication. *Multimed. Syst.* **2024**, *30*, 116. [[CrossRef](#)]
3. Gameda, M.D.; Han, M.S.; Abebe, A.T.; Kang, C.G. Design of Deep Learning-based One-bit Transceiver with Oversampling and Faster-Than-Nyquist (FTN) Signaling. *IEEE Access* **2024**, *12*, 43968–43983. [[CrossRef](#)]
4. Chen, C.; Zhong, W.D.; Yang, H.; Du, P. On the performance of MIMO-NOMA-based visible light communication systems. *IEEE Photonics Technol. Lett.* **2018**, *30*, 307–310. [[CrossRef](#)]
5. Li, Y.T.; Geng, T.W.; Gao, S.J. On the signal combinations for a uniquely decodable coded MIMO-FSO communication system. *Opt. Laser Technol.* **2024**, *172*, 110533. [[CrossRef](#)]
6. Zhang, Z.; Yuksel, M.; Yanikomeroglu, H. Faster-than-Nyquist signaling for MIMO communications. *IEEE Trans. Wirel. Commun.* **2023**, *22*, 2379–2392. [[CrossRef](#)]
7. Tong, M.; Huang, X.; Zhang, J.A. Faster-Than-Nyquist Transmission with Frame-by-Frame Decision-Directed Successive Interference Cancellation. *IEEE Trans. Commun.* **2023**, *71*, 4851–4861. [[CrossRef](#)]
8. Yuan, W.; Wu, N.; Wang, H.; Kuang, J. Variational inference-based frequency-domain equalization for faster-than-Nyquist signaling in doubly selective channels. *IEEE Signal Process Lett.* **2016**, *23*, 1270–1274. [[CrossRef](#)]
9. Jana, M.; Medra, A.; Lampe, L.; Mitra, J. Pre-equalized faster-than-Nyquist transmission. *IEEE Trans. Commun.* **2017**, *65*, 4406–4418. [[CrossRef](#)]
10. Cao, M.; Zhang, W.; Wang, H.; Lv, J. Adaptive pre-equalisation algorithm for point-by-point cancellation in super-Nyquist wireless optical communications. *Acta Opt. Sin.* **2020**, *40*, 46–53.
11. Baek, M.S.; Park, W.; Lee, Y.T. Deep learning-based signal detection technique for ftn signaling-based emergency alert communication system. In Proceedings of the 2021 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), Chengdu, China, 4–6 August 2021; pp. 1–3.
12. Lai, S.; Li, M. Recurrent neural network assisted equalization for FTN signaling. In Proceedings of the ICC 2020—2020 IEEE International Conference on Communications (ICC), Dublin, Ireland, 7–11 July 2020; pp. 1–6.
13. Cao, M.; Wu, Z.; Wang, H.; Xia, J.; Zhang, J.; Li, W. Deep Learning Assisted Pre-equalization Scheme for Faster-than-Nyquist Optical Wireless Communications. In Proceedings of the 2021 13th International Conference on Advanced Infocomm Technology (ICAIT), Yanji, China, 15–18 October 2021; pp. 118–121.
14. Oh, S.; Yu, M.; Cho, S.; Noh, S.; Chun, H. Bi-LSTM-Augmented deep neural network for multi-Gbps VCSEL-based visible light communication link. *Sensors* **2022**, *22*, 4145. [[CrossRef](#)] [[PubMed](#)]
15. Palitharathna, K.W.; Wickramasinghe, N.D.; Vegni, A.M.; Suraweera, H.A. Neural Network-Based Optimization for SLIPT-Enabled Indoor VLC Systems with Energy Constraints. *IEEE Trans. Green. Commun. Netw.* **2024**, *8*, 839–851. [[CrossRef](#)]
16. Li, L.; Zhu, Z.; Zhang, J. Neural Network-Based Transceiver Design for VLC System over ISI Channel. *Photonics* **2022**, *9*, 190. [[CrossRef](#)]
17. Zou, D.; Wang, W.; Li, F.; Sui, Q.; Yi, X.; Lu, C.; Li, Z. Colored Noise Suppressed FTN vs. PS in Peak Power Constrained IM-DD PAM Systems. *J. Light. Technol.* **2024**, *42*, 550–559. [[CrossRef](#)]

18. Che, D.; Chen, X. Modulation format and digital signal processing for IM-DD optics at post-200G era. *J. Light. Technol.* **2024**, *42*, 588–605. [[CrossRef](#)]
19. Ke, J.; Gao, M.; Chen, B.; Shao, W. Partial response THP for high-speed PAM-4 transmission. *Opt. Fiber Technol.* **2023**, *81*, 103487. [[CrossRef](#)]
20. Zedini, E.; Ata, Y.; Alouini, M.S. Improving Performance of Integrated Ground-HAPS FSO Communication Links with MIMO Application. *IEEE Photonics J.* **2024**, *16*, 1–14. [[CrossRef](#)]
21. Li, Q.; Wang, Y.; Dong, J.; Zhang, C.; Peng, K. Multi-node knowledge graph assisted distributed fault detection for large-scale industrial processes based on graph attention network and bidirectional LSTMs. *Neural Netw.* **2024**, *173*, 106210. [[CrossRef](#)]
22. Liang, R.; Xie, Y.; Cheng, J.; Pang, C.; Schuller, B. A non-invasive speech quality evaluation algorithm for hearing aids with multi-head self-attention and audiogram-based features. *IEEE/ACM Trans. Audio Speech Lang. Process* **2024**, *32*, 2166–2176. [[CrossRef](#)]
23. Guo, R.; Liu, H.; Xie, G.; Zhang, Y.; Liu, D. A Self-Interpretable Soft Sensor Based on Deep Learning and Multiple Attention Mechanism: From Data Selection to Sensor Modeling. *IEEE Trans. Ind. Inform.* **2023**, *19*, 6859–6871. [[CrossRef](#)]
24. Cao, M.; Wang, R.; Zhang, Y.; Zhang, X.; Wang, H. End-to-end performance of a CNN-AE based faster-than-Nyquist rate free space optical communication system. *J. Chongqing Univ. Posts Telecommun. (Nat. Sci. Ed.)* **2024**, *36*, 181–190.
25. Zhou, H.; Huang, L.; Yang, J.; Zhou, J.; Song, D.; Chen, X.; Dong, M. Low Complexity Double Vector Model Predictive Control with Minimum Root Mean Square Error for Three-phase Three-Level Inverters. *IEEE J. Emerg. Sel. Top. Power Electron.* **2023**, *11*, 5809–5819. [[CrossRef](#)]
26. Zafar, M.H.; Khan, N.M.; Abou Houran, M.; Mansoor, M.; Akhtar, N.; Sanfilippo, F. A novel hybrid deep learning model for accurate state of charge estimation of Li-Ion batteries for electric vehicles under high and low temperature. *Energy* **2024**, *292*, 130584. [[CrossRef](#)]
27. Huang, F.; Zhang, X. A new interpretable streamflow prediction approach based on SWAT-BiLSTM and SHAP. *Environ. Sci. Pollut. Res.* **2024**, *31*, 23896–23908. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.