



---

*Research article*

## **An optimized LSTM-based equalizer for 100 Gigabit/s-class short-range fiber-optic communications**

**Vuong Quang Phuoc<sup>1,2,3,\*</sup>, Nguyen Van Dien<sup>4</sup>, Ho Duc Tam Linh<sup>3</sup>, Nguyen Van Tuan<sup>2</sup>, Nguyen Van Hieu<sup>2</sup>, Le Thai Son<sup>5</sup> and Nguyen Tan Hung<sup>1</sup>**

<sup>1</sup> The University of Danang - Advanced Institute of Science and Technology, Danang 50000, Vietnam

<sup>2</sup> The University of Danang - University of Science and Technology, Danang 50000, Vietnam

<sup>3</sup> University of Sciences, Hue University, Hue 49000, Vietnam

<sup>4</sup> FPT University, Danang 50000, Vietnam

<sup>5</sup> Nubis Communications, New Providence NJ 07974, US

\* **Correspondence:** Email: [vqphuoc@hueuni.edu.vn](mailto:vqphuoc@hueuni.edu.vn).

**Abstract:** Intensity modulation/direct detection (IM/DD) remains to be the preferred optical transmission scheme for short-range applications for its simplicity of application, inexpensiveness, and small footprint. However, the impairments of low-cost device and fiber chromatic dispersion lead to the limitation of system performance when the data rate rises to 100 Gbps or higher. In this paper, we demonstrated that an equalizer using neural networks can effectively improve the transmission performance of high-speed IM/DD systems. An optimization of a long short-term memory (LSTM) structure in terms of network depth and distribution of neurons in hidden layers leads to an enhancement of the overall performance of the 50 Gbaud PAM4 communications. Furthermore, the results for a system using a LSTM-based equalizer give the better outcome than the traditional feed-forward equalizer (FFE) or artificial neural network (ANN)-based equalizer.

**Keywords:** fiber-optic communications; intensity modulation; direct detection; signal equalization; long short-term memory; artificial neural network

---

## 1. Introduction

The increasing demand for data is putting a strain on optical network capacity. The primary cause is the sharp rise in the use of internet-based applications including cloud computing, video-on-demand services, 5G deployment, and other developing technologies [1]. To meet specific requirements, researchers have conducted thorough into short-reach optical links for applications such as data center interconnects (DCI), 5G fronthaul, optical access, etc [2,3]. However, the primary objective is always to increase transmission speed. Some newer standards, such as NG-PON2 [4] and High Speed PON [5], allow only up to 40 or 50 Gbps/wavelength, respectively. Therefore, the field of 100G+ transmission lines is quite attractive considering the current demand's rapid development.

In contrast to long-haul transmission, the widespread deployment of short-reach optical links places significant emphasis on cost, size, and complexity that are considered critical factors. The most suitable approach for achieving these critical factors is direct intensity modulation with direct detection (IM/DD) technique rather than coherent detection one [6]. However, it is difficult for conventional IM/DD optical links, which are established with the NRZ-OOK, to get greater transmission rates due to the need of high-speed, high-cost components. Various advanced modulation techniques, including discrete multi tone (DMT), carrierless amplitude phase modulation (CAP), and pulse amplitude modulation (PAM), are employed to reduce the bandwidth limitation of components through increasing the spectrum efficiency (SE). Among these formats, high order PAM, e.g., four-level PAM (PAM4), have been shown as the most suitable for short-reach scenarios due to its low complexity and high-power efficiency [7].

Short-reach optical systems utilizing PAM4 are subject to both linear and nonlinear impairments, such as chromatic dispersion, limited bandwidth, nonlinearity of the low-cost devices. To overcome these challenges while remaining simple and flexible for short-reach schemes, digital signal processing (DSP) plays a critical role in compensating for these imperfections [8]. Thus, advanced equalization techniques have been proposed to improve the overall performance. In short-reach systems, feed-forward equalizers (FFE) are commonly used for equalization of linear impairments including chromatic dispersion and inter-symbol interference (ISI). However, due to their linearity, FFEs become ineffective in mitigating nonlinear distortions caused by the low-cost opto-electronic components, the complex interactions of laser chirps, chromatic dispersion and square-law detection [8]. In recent years, nonlinear equalization schemes, including Volterra nonlinear equalizers (VNLEs) [9] and machine learning (ML) based equalizers using artificial neural networks (ANNs) [10,11], have been demonstrated to be useful in enhancing the performance of high data rate IM/DD systems. While the VNLEs have been shown to be very effective in mitigating system nonlinearities provided sufficient polynomial order and memory depth [12–14], the architectural complexity of VNLEs increases exponentially due to the excessive extension for multiplications between symbols, which needs the use of multipliers [15]. On the other hand, once the ANN has been trained, which can be done offline, the ANN only requires multiplications with fixed numbers (weights), which can be efficiently implemented using lookup tables (LUTs). LUTs can be also used to implement simple nonlinear activation functions, such as tanh or hard tanh function [15].

Despite these advances, standard ANNs have limitations for several reasons. First, “vanishing gradient” problem limits the predictive ability of an ANN model. This issue occurred when the gradients became too small, resulting in the deeper layer’s weights not being updated properly during training and preventing the model from learning features from data [16]. Second, ANNs are unable to

recall information from earlier steps, reducing their effectiveness in handling time series data [17]. Signal transmission within the system can be affected by many factors such as noise, dispersion, or nonlinearity. These effects often have a complex relationship and span multiple time steps, so using an ANN-based equalizer may not be optimal in this scenario.

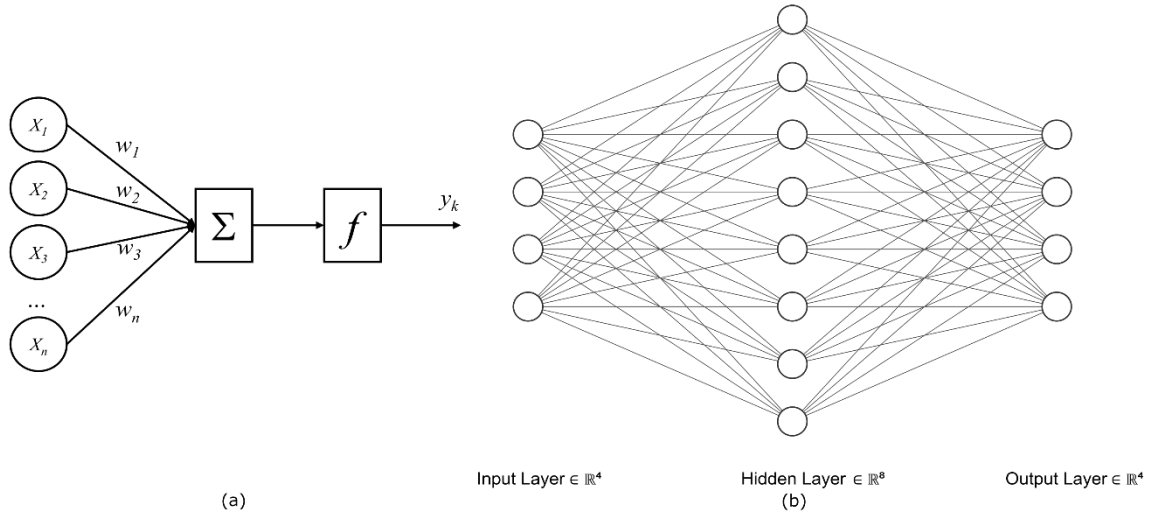
Long short-term memory (LSTM) architecture, a variant of recurrent neural network (RNN), was developed to address previous drawbacks of ANNs. LSTMs have long-term memory and selectively retain relevant information, which allows them to perform well with time series data. The LSTM gating mechanism also provides context awareness and adaptive learning, while mitigating the “vanishing gradient” problem. These make LSTM suitable for applications to predict and analysis of transmission performance and anomaly detection or applying LSTM to traffic generation and prediction [16]. LSTM has been used for distortion equalization in optical communications [17]. In [18], the author proposed a nonlinear equalizer based on LSTM for 50 Gbps (25 Gbaud) PAM4 transmission system to compensate and directly classify the received signal. Unfortunately, using dispersion compensation fiber (DCF) and transmission power of 10-12 dBm for optimal cases made it difficult for short-reach communication applications. LSTM has also been applied to short-reach communication for ISI extraction [19] or nonlinear equalization [20]; however, these approaches use coherent detection, causing a significant increase in the complexity and the cost for the system.

In this study, we analyze and evaluate the impact of LSTM-based equalizers (using feed-forward equalizer combined with LSTM structure) for mitigating signal distortions in an optical fiber communication system. This process is achieved by adding an LSTM network just behind the traditional feed-forward equalizer. The output of the equalizer is then fed into the LSTM network through delayers, where LSTM cells extract features and relationships between past and future data to estimate the corresponding PAM4 levels at the output. For deeper evaluation, we also perform a comparison of results in cases using only the traditional equalizer and in cases combining with an ANN architecture. The system performance is assessed by using the BER and the optical signal-to-noise ratio (OSNR) penalty. The results show that the use of LSTM leads to an OSNR penalty improvement of about 1.5 dB, doubling compared to the case using ANN-based equalizer, when compared with the traditional equalizer at the same BER. This OSNR penalty value rises to 2.5 dB when considering a BER of  $10^{-3}$  and tends to increase as the OSNR of the signal increases. In addition, to determine the optimal architecture when using LSTM, we also consider changes in network depth and size. Research results indicate that, with the same number of hidden neurons, using two hidden layers in the neural network architecture yields better results than one layer, while increasing the depth of the model does not further improve signal quality and may decrease performance due to overfitting. Using an uneven distribution for hidden layers also helps to improve the quality of the output signal. For example, in a simple model with 16 hidden neurons, allocating 12 neurons to the first layer and 4 neurons to the second hidden layer produces better results than that of an 8:8 ratio. These results are promising for future applications in high-speed optical fiber communication systems.

The structure of the paper is as follows. In Section 2, we describe the ML-based equalizers, emphasizing the ANN, LSTM architects, and ML-based equalizer structures. In Section 3, we explain the configuration of the simulation system. In Section 4, we present the results and discussions, and draw conclusions in Section 5.

## 2. ML-based equalizers

### 2.1. Artificial Neuron Network – ANN



**Figure 1.** (a) A simple architecture of Perceptron, and (b) a particular MLP network with only one hidden layer.

ANNs have been suggested in recent years to mitigate impairments in optical communication systems. The main benefit of ANNs is their ability to generalize any input-output sequence with a few hidden neurons/layers. Basically, standard ANNs, also known as Multi-Layers Perceptrons (MLPs), arranged in a series of layers. This structure includes 3 layer types: (i) The input layer contains the independent variables  $x_n$  (symbols) that should be considered; (ii) the output layer provides an estimated output symbol  $y_n$  based on the values of the previous layer; and (iii) the hidden layer(s) takes the role of collecting information from previous layer, each neurons use weights, bias, and then incorporates with the activation function to defines the output of neuron according to certain conditions. The output of a neuron in hidden layer can be illustrated in Figure 1 and formulated as follows [16]:

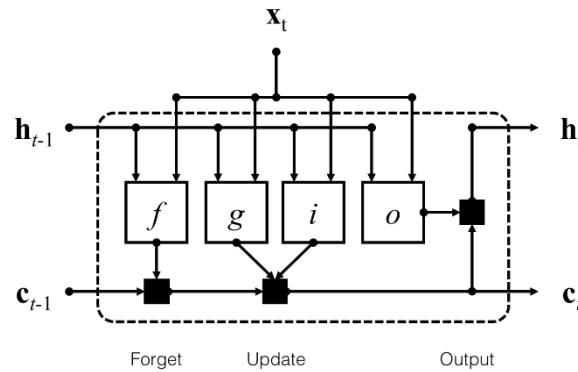
$$y_k = f \left( \sum_i^n x_i w_i + b \right) \quad (1)$$

where weight parameters ( $w$ ) indicate the strength of the connection between inputs ( $x$ ) and output ( $y$ ) of neurons in the current hidden layer. Bias value ( $b$ ) is a constant value that allows controlling the output. To prevent the model from becoming linear, ANNs applies a nonlinear activation function ( $f$ ) to each hidden neuron.

### 2.2. Long Short-Term Memory – LSTM

Different from traditional ANNs, LSTM networks were designed to address drawbacks of ANNs based on cell state (memory cell) and gate units [17,21,22]. First, using cell state allows LSTMs to remember information for long periods, making them more effective at detecting

long-term data relationships than ANNs. Second, LSTMs use three gates (input, forget, and output) to control the flow of information into and out of the LSTM cell. Due to this gating mechanism, LSTMs can selectively keep or discard information over long periods of time. Moreover, it helps to avoid the problem of vanishing gradients, ensuring that important information is retained even across long sequences. Furthermore, LSTMs are robust against noise and data variations because of their ability to maintain and update states selectively, since long-term context enables LSTMs to distinguish meaningful patterns from random fluctuations.



**Figure 2.** Basic structure of a LSTM cell with three types of gates: input gate  $i$ , forget gate  $f$  and output gate  $o$ .

Figure 2 illustrated basic structure of a LSTM cell, where  $i$  (input gate) handles the level of cell state update;  $f$  (forget gate) decides the level of cell state reset (forget);  $g$  (cell candidate) determines which information is added to cell state; and  $o$  (output gate) controls the level of cell state added to the hidden state. The cell state can be expressed as

$$\mathbf{c}_t = f_t \odot \mathbf{c}_{t-1} + i_t \odot g_t \quad (2)$$

where  $\odot$  stands for the element-wise multiplication of vectors and the hidden state  $\mathbf{h}_t$  is provided by

$$\mathbf{h}_t = o_t \odot \sigma_c(\mathbf{c}_t) \quad (3)$$

where  $\sigma_c$  denotes the cell activation function, here we use tanh function.

The process of LSTM cell is given by

Input gate:

$$i_t = \sigma_g(W_i \mathbf{x}_t + R_i \mathbf{h}_{t-1} + b_i) \quad (4)$$

Forget gate:

$$f_t = \sigma_g(W_f \mathbf{x}_t + R_f \mathbf{h}_{t-1} + b_f) \quad (5)$$

Cell candidate:

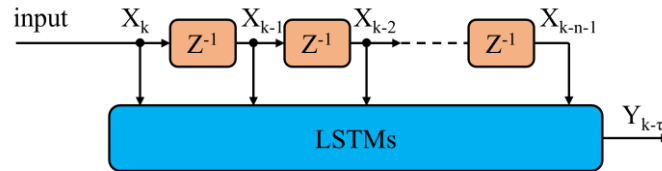
$$g_t = \sigma_g(W_g \mathbf{x}_t + R_g \mathbf{h}_{t-1} + b_g) \quad (6)$$

Output gate:

$$o_t = \sigma_g(W_o \mathbf{x}_t + R_o \mathbf{h}_{t-1} + b_o) \quad (7)$$

In these calculations,  $W$  indicates the input weights,  $R$  is the recurrent weights, the bias is  $b$ , and  $\sigma_g$  denotes the gate activation function with the sigmoid function, given by  $\sigma(x) = (1 + e^{-x})^{-1}$ .

### 2.3. Approach and modelling process for ML-based equalizer



**Figure 3.** Structures of evaluated LSTMs-based equalization schemes with  $n$  inputs.

In this study, the data preparation for LSTM-based equalizers is preprocessed by FFE using an FIR filter with the LMS algorithm, and then fed into the LSTM network through  $n$  delays, as shown in Figure 3. For evaluation, we define the LSTM structure with  $2^m$  ( $m \in \mathbb{N}$ ,  $m > 0$ ) hidden neurons and 4 neurons at the output corresponding to 4 levels of PAM4 signals. We use cross-entropy loss function, which is simple and sensitive to the change of network. Furthermore, the combination of cross-entropy and SoftMax function gives a smooth and stable derivative, making the gradient not increase too large or decrease too small, helping to reduce the “Exploding Gradient” or “Vanishing Gradient” problems.

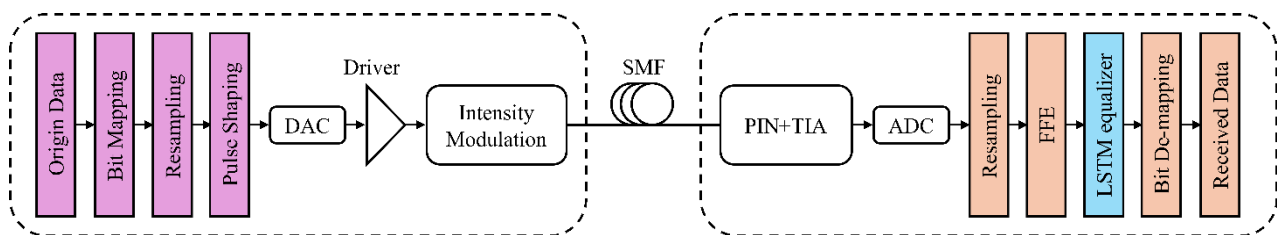
Determining the optimal structure of LSTMs in terms of hidden layers and the quantity of hidden units requires a lot of experimentation. Some research suggests that increasing the number of neurons or layers may improve performance, but the trade-off is the increase in complexity. In detail, we investigate the LSTMs-based equalizer through three scenarios:

- i. Increasing complexity: In this case, we use a single hidden layer model and vary the number of neurons to evaluate the impact of neuron’s quantity to equalization.
- ii. Increasing the depth: Using one or more layers of the LSTM model while keeping the total number of neurons constant, with the number of neurons in each hidden are the same.
- iii. Tuning the structure for feature learning: Like case (ii), when using a hierarchical structure with two hidden layers, varying the number of neurons in the hidden layers for more flexibility in adjusting the model’s capacity.

## 3. Simulation setup

Figure 4 illustrates the simulation setup of a typical system for short-reach optical communication based on PAM4 modulation in band C. First, the raw data is generated by a random binary sequence of  $2^{19}$  bits and then mapped into PAM4 format. After resampling with 8 samples per symbol, the signal is pulse-shaped to optimize bandwidth utilization. Here, the raised cosine (RC) was configured with a roll-off factor of 0.1. The processed signal is then fed into a digital-to-analog converter (DAC) to obtain the baseband electrical signal. Subsequently, the DAC output signals are

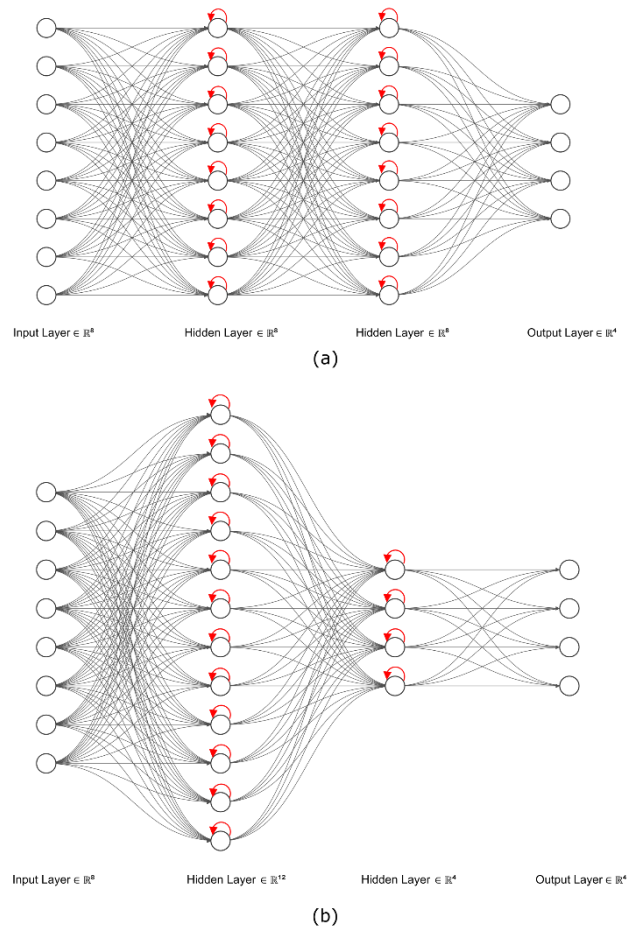
amplified by a linear control circuit and then directly modulated by a DML at 1550 nm to generate PAM4 optical signals at the rate of 50 Gbaud. The signals are transmitted through optical fiber with the dispersion configuration, SSMF (17.6 ps/nm/km). At the receiver, the signal is directly detected by a photodetector (PIN photodiode) with PD responsivity set at 0.7 A/W. To improve receiver sensitivity, a transimpedance amplifier (TIA) is used to amplify the electrical signal. In this evaluation, both DML and PIN-TIA are set with a 3-dB bandwidth of 25 GHz. After being processed by an analog-to-digital converter (ADC), the signal is resampled to 1 sample-per-symbol. Next, an FFE equalizer, using a FIR filter with 15 taps, which are enough for FFE to achieve the best performance, is employed to equalize the received signal. Then, LSTM equalizer is used to mitigate the impairments caused by the distortion of channel and low-cost devices. Finally, after mapping the output of LSTM to PAM4 symbol, it then converts to bits. Here the system also calculates BER of the received data.



**Figure 4.** Simulation model of 50 Gbaud PAM4 IM-DD transmission at wavelength 1550 nm.

In the schemes proposed, the number of hidden layers are set with 1, 2, 3, and 4 layers, respectively. The number of total hidden units is exponential function of 2. The activation function we use in the hidden layers is the tanh. The model is trained with the Adam optimizer, and the initial learning rate was set to  $10^{-3}$  with a reduction of factor  $10^{-1}$  after each 20 epochs. For the model evaluation,  $2^{18}$  symbols data set is split into two parts: 70% for training and the rest for testing and validation. The training data is fed through the network until they reach the output layer and classified with a SoftMax activation to receive the symbol class.

Figure 5 shows two sample architectures of network, both structures have an input layer and an output layer with 8 and 4 neurons, respectively. The difference is the distribution of neurons in two hidden layers. In which case (upper) is deployed with the same number of neurons in each hidden layer, while case (lower) uses the architecture with an uneven distribution. This investigation allows us to evaluate the feature learning when using a hierarchical model. The first layer may learn more basic temporal patterns of data, and then the second hidden layer can capture abstract and higher-level temporal dependencies. In above cases, we use BER to evaluate system quality. The lower the BER index is, compared to other methods (traditional FFE and ANNs-based equalizer), the more reliable the LSTM-based equalizer.



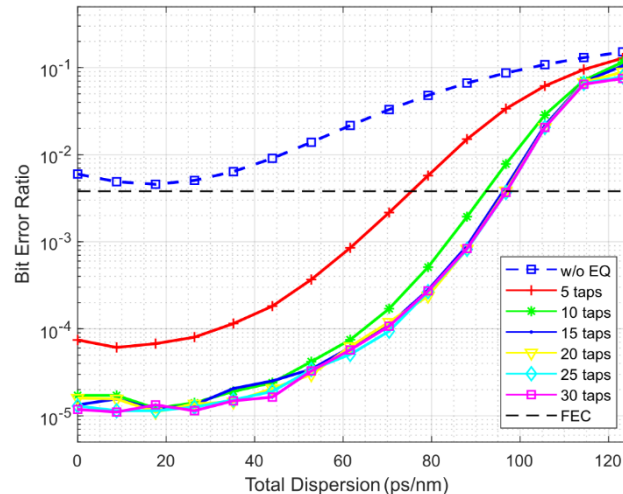
**Figure 5.** Two structures with 8 input neurons, 4 output neurons, and 2 hidden layers with 16 hidden neurons in total: (a) Using even distribution and (b) using uneven distribution with the same number of hidden units.

#### 4. Simulation results and discussion

In this session, for a better understanding and more accurate evaluation of the LSTM equalizers, we first examine the undertest system with or without the conventional FFE. Then, we demonstrate the combination of the FFE and ANN equalizer/LSTM equalizer to determine how much the BER improves. Finally, several different LSTM equalizer structures are also investigated and evaluated.

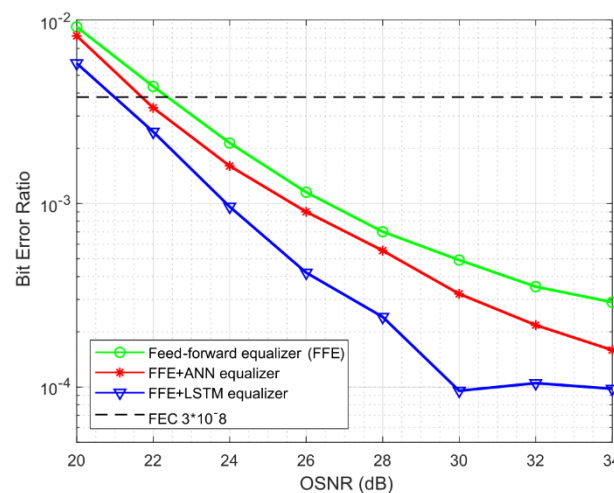
First, we investigate the performance of FFE equalizers. In this case, only a simple FFE with a FIR filter is established at the receiver for ISI compensation and channel equalization. The system's performance is measured via BER. Equalizer's coefficients are adjusted by the least mean square (LMS) algorithm. Figure 6 demonstrates that the FFE significantly improves system performance. In the case of a back-to-back (B2B) system (at the dispersion of 0 ps/nm), the BER decrease from nearly  $6 \times 10^{-3}$  to  $7.5 \times 10^{-5}$  with just five taps. We can see that while the number of taps increases, the performance increases too. Note that, with the forward error correction (FEC) threshold, the system reaches its limit at the dispersion of 96.8 ps/nm. We also find that 15 taps are suitable for the FFE to achieve the best performance in all given cases; other cases of 20, 25, and 30 taps give better performance but not much while increasing the complexity of the FFE equalizer.



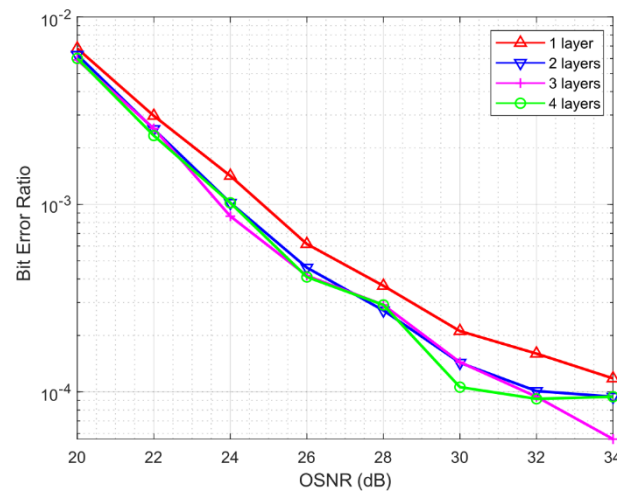


**Figure 6.** BER performance as functions of total dispersion with different number of taps required for FFE for the 50 GBaud PAM4 system.

We then go over the function and effects of the ML-based equalizer on the system. In this scenario, ANN or the LSTM model, established with two hidden layers of 16 total neurons, is set following the 15 taps FFE to evaluate at 96.8 ps/nm dispersion with the OSNR range of 20 dB to 34 dB. The outcomes of three cases are shown in Figure 7. Use only FFE with 15 taps FIR filter; combine FFE and ANN for equalization (FFE + ANN equalizer) and use the LSTM instead of ANN structure (FFE + LSTM equalizer). It could be observed that the system has the highest BER when applying only FFE (the green line). If ANN is set up along with FFE for equalization, the BER outcome (the red line) has somewhat decreased, but the gain has been only about 1 dB of OSNR penalty. The system achieves the best when the LSTM equalizer and FFE are combined. Additionally, BER improved with increasing OSNR. For example, at FEC threshold BER =  $3.8 \times 10^{-3}$ , for the case FFE + LSTM equalizer, the OSNR penalty is roughly 1.5 dB increase to 2.5 dB at BER =  $10^{-3}$ . The results show that the system performance is enhanced better by the FFE + LSTM equalizer than by the FFE + ANN equalizer. This can be described as the improvement of LSTM over the drawbacks of conventional ANN, like gradient vanish.

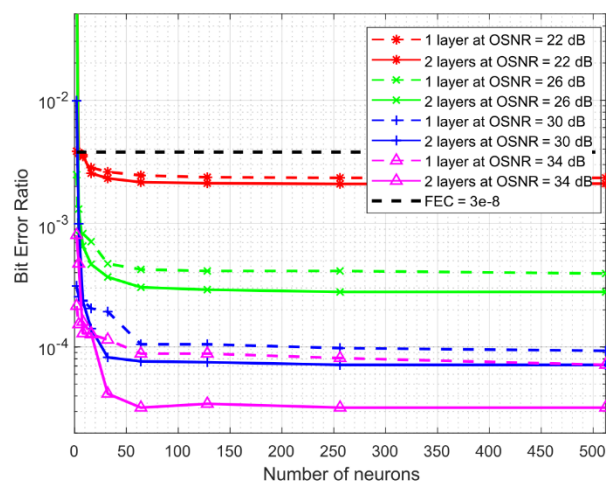


**Figure 7.** BER performance as functions of OSNR as ANN and LSTM equalizer used with FFE.



**Figure 8.** BER performance as functions of OSNR for LSTM equalizer with different model depths, each hidden layer contains 8 neurons.

To look deeper, we perform investigations by varying the depth (number of layers) and the complexity (number of neurons) of LSTM models to improve the equalization system's accuracy. Figure 8 illustrates the performance of the system when applying with an LSTM equalizer. LSTM models are configured with 1/2/3 or 4 hidden layers, each with 8 neurons. Four lines represent various configurations - the red line for one layer, the blue line for two layers, the purple line for three layers, and the green line for four layers. On the OSNR scale, markers are placed at various points along each line. Found that BER decreases as OSNR rises by all lines. Besides, as the number of layers rises, the system performance does not significantly improve, this might be the result of overfitting with the simple data in this case. Adding additional layers to the model, which raises its complexity and makes it too close to the training data, which causes the model cannot make accurate predictions from other data than the training data. In this case, an LSTM model with one or two hidden layers may be the best choice.

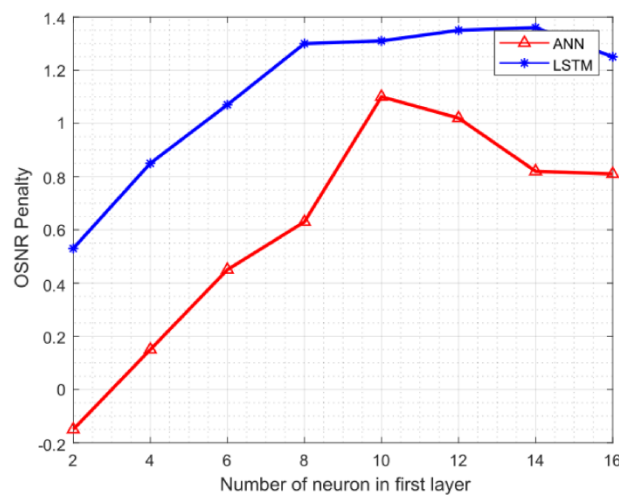


**Figure 9.** BER performance as functions of number of neurons in cases of one and two layer(s) at different OSNR levels.

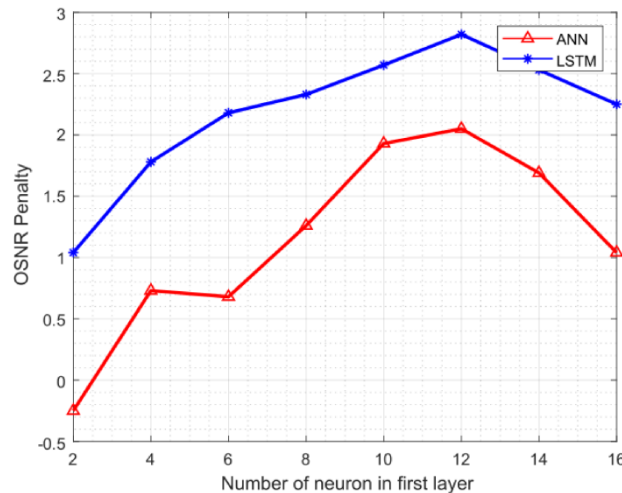
To determine the optimal number of neurons for the LSTM model, the authors ran tests with the LSTM equalizer while varying the number of hidden neurons. The investigated model has one

hidden layer (the dash line) or two hidden layers (the solid line) and  $2^m$  ( $m \in \mathbb{N}$ ,  $0 \leq m \leq 9$ ) total neurons. Thus, the results are obtained under various OSNR conditions. Figure 9 shows that, like increasing the depth of the model, increasing the number of neurons also leads to saturation of output results. Based on simulation results, the model's neuron count can be optimized with different OSNR values, trading-off complexity and performance. For example, the system produces the best results with 64 neurons, and more neurons make the system more complex without significantly improving performance. However, even with only 16 or 32 neurons, the system can improve computational time and reduce power consumption while maintaining the expected BER ratio. To be able to come up with an appropriate configuration, we must consider the system's complexity as well as cost requirements. Figure 9 shows that equalizers with two LSTM layers using the same number of neurons reduce BER; thus, the higher the OSNR, the better the result. It could be clear that a hierarchical structure allows the network to organize information between layers and easier to understand relationships and dependencies from the data, leading to better results than using only one hidden layer.

In this paper, we propose a LSTM equalizer with two hidden layers and sixteen hidden neurons. This choice ensures both simplicity for the balancing system and inherits advantages of a hierarchical structure. To determine the best structure in this case, we implemented several surveys that varied the structure of the two hidden layers, by making the number of neurons in the first hidden layer multiple of two while keeping a total of 16 hidden neurons. Survey cases were also carried out using the same ANN model structure for comparison. Evaluation results are compared using the OSNR penalty (the difference between the required OSNR for a given BER of the ANN/LSTM equalizer result and the FFE result), with a higher penalty indicating a more optimal system.



**Figure 10.** OSNR penalties of the system using ANN and LSTM with respect to the case using only FFE for FEC threshold at BER level of  $3.8 \times 10^{-3}$ .

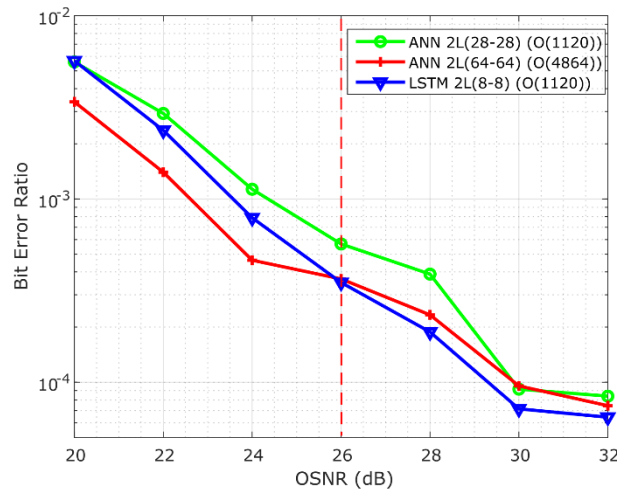


**Figure 11.** OSNR penalties of the system using ANN and LSTM with respect to the case using only FFE for FEC threshold at BER level of  $10^{-3}$ .

Figures 10 and 11 illustrate OSNR penalties of the system using two neural network models - ANN and LSTM – with respect to the case using only FFE for FEC threshold at BER levels of  $3.8 \times 10^{-3}$  and  $10^{-3}$ , respectively. The results show that the OSNR penalty in case of LSTM equalizer is always better than that in case of ANN one at both BER =  $3.8 \times 10^{-3}$  and  $10^{-3}$ . This indicates that in an optical communication system, the equalizer using the LSTM model with two hidden layers performed is better at minimizing BER than the ANN model. So, LSTM can be an attractive choice for equalization in optical communication systems. In detail, the OSNR penalty increases with a growing number of neurons in the first layer. Moreover, with 12 neurons in the first layer, the LSTM equalizer can improve the OSNR penalty by approximate 1.4 dB at the FEC threshold and nearly 2.8 dB at BER of  $10^{-3}$  when compared to a 15 taps FFE operating alone. We also observed that the model becomes a single hidden layer when the first layer's number of neurons reaches 16, and that the two-layer model produces a better BER than the one-layer model.

In terms of complexity, we carried out a complexity comparison between ANN and LSTM-based equalizers. According to Ian Goodfellow's Deep Learning [17], the computational complexity of these models can be calculated as follows. For the ANN model, the computational complexity is determined by the number of operations needed to compute the outputs of each layer. For an ANN with  $x$  inputs,  $y$  outputs, and  $l$  hidden layers, where each hidden layer has  $h$  hidden units, the total number of parameters includes those from the input to the first hidden layer ( $x \times h$ ), between hidden layers  $h \times h$  for each layer, and from the last hidden layer to the output  $h \times y$ . For the LSTM model, which is a variant of recurrent neural networks (RNNs), each LSTM cell has four sets of weights corresponding to its gates and cell candidate. For an LSTM with  $x$  inputs,  $y$  outputs,  $l$  hidden layers, and  $h$  hidden units in each layer, the total number of parameters is  $4 \times (x \times h + h \times h + h \text{ (bias)})$ , and the last hidden layer to the output is  $(h \times y)$ . For the same network structure, the complexity of LSTM model is always larger than that of ANN model because each LSTM cell requires an implementation of four sets of weights. For example, for a network of 8 inputs, 4 outputs, 2 hidden layers (8 hidden units per each layer), the complexity of ANN model is  $O(8 \times 8 + 8 \times 8 + 8 \times 4) = O(160)$ ; and the complexity of LSTM model is  $O(544 + 544 + 32) = O(1120)$ . However, as can be seen in Figure 7, with the same network of 8 inputs, 4 outputs, and 2 hidden layers (8 hidden units per layer), the performance of the LSTM model is much better

than that of the ANN. As a result, the complexity of LSTM and ANN models should be compared given the same performance.



**Figure 12.** Performance comparison of LSTM and ANN models using 8 inputs, 4 outputs, 2 hidden layers, and different units per hidden layer.

Figure 12 shows the BER performance comparison between the LSTM and ANN models using 8 inputs, 4 outputs, 2 hidden layers, and different units per hidden layer. As shown in Figure 12, to achieve the same BER of  $3.6 \times 10^{-4}$  at OSNR of 26 dB, the LSTM needs only 8 units per hidden layer with a complexity of  $O(1120)$  while the ANN requires 64 units per hidden layer, causing a higher complexity of  $O(4864)$ . Furthermore, with 28 units per hidden layer the ANN model gives the same complexity of  $O(1120)$  and a larger BER ( $5.7 \times 10^{-4}$ ) at OSNR of 26 with respect to the LSTM with 8 units per hidden layer. These results indicate that to achieve the same system performance, the LSTM model offers a lower complexity than the ANN.

Furthermore, since the LSTM-based equalizer is implemented in the DSP blocks after the optical/electrical conversion at the receiver, in principle, it can also be used for multiple-wavelength short-reach optical communication systems. One major concern is the nonlinear crosstalk between wavelengths, which impacts the effectiveness of the LSTM-based equalizer. However, in short-reach optical communication systems, wavelengths are often allocated with large channel spacing via coarse wavelength division multiplexing (CWDM) technology [23]. According to ITU-T, CWDM wavelength channel spacing is 20 nm. Therefore, the nonlinear crosstalks between such CWDM channels are negligible and the LSTM-based equalizer can work effectively for all wavelengths.

In summary, the proposed LSTM-based equalizer has shown robustness in our simulations, performing well under various signal conditions, including different OSNR levels and optical nonlinearities. However, for practical applications, a few issues need further research: First, the complexity of LSTMs could be a challenge. Second, the LSTM based equalizer may require more optimization techniques for real-time processing. Finally, while our results are promising, further testing under different scenarios is important to ensure consistent performance of the transmission system. Hence, there are possible areas for further research, including optimizing the LSTM architecture for reducing computational complexity and improving the generalization of the model to different and unpredictable channel conditions.

## 5. Conclusions

In this paper, we demonstrated the effectiveness of LSTM-based equalizers in improving the performance of high-speed PAM4 transmission systems. Specifically, for a 50 Gbaud PAM-4 transmission, our results reveal that both ANN and LSTM equalizers, each with only 2 hidden layers and 16 neurons, can significantly enhance the BER at the Forward Error Correction (FEC) limit. For a given chromatic dispersion of 96.8 ps/nm, the LSTM equalizer improves by approximate 1.4 dB at the FEC threshold and nearly 3 dB at BER of  $10^{-3}$  when compared to a 15 taps FFE operating alone. These results demonstrate the ability of LSTM architectures to reduce the impact of chromatic dispersion while improving the overall performance of PAM4 transmission systems. Furthermore, our findings show that the LSTM equalizer scheme outperforms the ANN equalizer in terms of performance when both the number of hidden units and network depth are kept constant. This suggests that LSTM architecture could offer a more effective equalization solution in high-speed PAM4 transmission systems, providing better sensitivity and outperforming the traditional equalization techniques.

## Author contributions

Vuong Quang Phuoc, Nguyen Tan Hung, Nguyen Van Tuan: Writing - Original Draft, Writing - Review & Editing. All authors: Writing - Review & Editing, Conceptualization, Formal analysis, Investigation, Software, Validation. All authors have read and agreed to the submitted version of the manuscript.

## Acknowledgments

This research was supported by Vietnam Ministry of Education and Training under Grant B2022-DNA-10.

## Conflict of interest

All authors declare no conflicts of interest in this paper.

## References

1. Cisco (2020) Cisco Annual Internet Report (2018–2023) White Paper. Available from: <https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html>
2. Zhong K, Zhou X, Wang Y, Gui T, Yang Y, Yuan J, et al. (2017) Recent advances in short reach systems. *Optical Fiber Communication Conference*, Tu2D.7. <https://doi.org/10.1364/OFC.2017.Tu2d.7>
3. Kachris C, Kanonakis K, Tomkos I (2013) Optical interconnection networks in data centers: recent trends and future challenges. *IEEE Commun Mag* 51: 39–45. <https://doi.org/10.1109/mcom.2013.6588648>

4. Telecommunication Standardization Sector of ITU, G.989.2: 40-Gigabit-capable passive optical networks 2 (NG-PON2): Physical media dependent (PMD) layer specification, Telecommunication Standardization Sector of ITU 2019. Available from: <https://www.itu.int/rec/T-REC-G.989.2>
5. Telecommunication Standardization Sector of ITU, G.9804.3: 50-Gigabit-capable passive optical networks (50G-PON): Physical media dependent (PMD) layer specification, Telecommunication Standardization Sector of ITU 2021. Available from: <https://www.itu.int/rec/T-REC-G.9804.3-202109-I/en>
6. Wei J, Cheng Q, Penty RV, White IH, Cunningham DG (2015) 400 Gigabit Ethernet using advanced modulation formats: Performance, complexity, and power dissipation. *IEEE Commun Mag* 53: 182–189. <https://doi.org/10.1109/MCOM.2015.7045407>
7. Zhong K, Zhou X, Gui T, Tao L, Gao Y, Chen W, et al. (2015) Experimental study of PAM-4, CAP-16, and DMT for 100 Gb/s Short Reach Optical Transmission Systems. *Opt Express* 23: 1176–1189. <https://doi.org/10.1364/OE.23.001176>
8. Zhou H, Li Y, Liu Y, Yue L, Gao C, Li W, et al. (2019) Recent Advances in equalization Technologies for ShortReach Optical Links based on PAM4 modulation: A review. *Applied Sciences* 9: 2342. <https://doi.org/10.3390/app9112342>
9. Stojanovic N, Karinou F, Zhang Q, Prodaniuc C (2017) Volterra and Wiener Equalizers for Short-Reach 100G PAM-4 applications. *J Lightwave Technol* 35: 4583–4594. <https://doi.org/10.1109/JLT.2017.2752363>
10. Yi L, Tao L, Huang L, Xue L, Li P, Hu W (2019) Machine Learning for 100 Gb/s/ $\lambda$  Passive Optical Network. *J Lightwave Technol* 37: 1621–1630. <https://doi.org/10.1109/JLT.2018.2888547>
11. Estaran J, Rios-Müller R, Mestre MA, Jorge F, Mardoyan H, Konczykowska A, et al. (2016) Artificial Neural Networks for Linear and Non-Linear Impairment Mitigation in High-Baudrate IM/DD Systems. *42nd European Conference on Optical Communication*, 1–3. VDE.
12. Giacomidis E, Le ST, Aldaya I, Wei JL, McCarthy M, Doran NJ, et al. (2016) Experimental Comparison of Artificial Neural Network and Volterra based Nonlinear Equalization for CO-OFDM. *Optical Fiber Communication Conference*, W3A-4. <https://doi.org/10.1364/OFC.2016.W3A.4>
13. Kyono T, Otsuka Y, Fukumoto Y, Owaki S, Nakamura M (2018) Computational Complexity Comparison of Artificial Neural Network and Volterra Series Transfer Function for Optical Nonlinearity Compensation with Time- and Frequency-Domain Dispersion Equalization. *European Conference on Optical Communication (ECOC)*, 1–3. IEEE. <https://doi.org/10.1109/ECOC.2018.8535153>
14. Hung NT, Stainton S, Le ST, Haigh PA, Tien HP, Vien ND, et al. (2023) High-speed PAM4 transmission using directly modulated laser and artificial neural network nonlinear equaliser. *Opt Laser Technol* 157: 108642. <https://doi.org/10.1016/j.optlastec.2022.108642>
15. Schädler M, Böcherer G, Pachnicke S (2021) Soft-Demapping for Short Reach Optical Communication: A Comparison of Deep Neural Networks and Volterra Series. *J Lightwave Technol* 39: 3095–3105. <https://doi.org/10.1109/JLT.2021.3056869>
16. Nielsen MA (2019) *Neural networks and deep learning*. Determination Press.
17. Goodfellow I, Bengio Y, Courville A (2016) *Deep Learning*. MIT Press, 800 pp.



18. Dai X, Li X, Luo M, You Q, Yu S (2019) LSTM networks enabled nonlinear equalization in 50-Gb/s PAM-4 transmission links. *Appl Optics* 58: 6079–6084. <https://doi.org/10.1364/AO.58.006079>
19. Peng CW, Chan DW, Chow CW, Hung TY, Jian YH, Tong Y, et al. (2023) Long Short Term Memory Neural Network (LSTMNN) and inter-symbol feature extraction for 160 Gbit/s PAM4 from silicon micro-ring transmitter. *Opt Commun* 529: 129067. <https://doi.org/10.1016/j.optcom.2022.129067>
20. Wu Q, Xu Z, Zhu Y, Zhang Y, Ji H, Yang Y, et al. (2023) Machine learning for Self-Coherent detection Short-Reach optical communications. *Photonics* 10: 1001. <https://doi.org/10.3390/photonics10091001>
21. Hochreiter S, Schmidhuber J (1997) Long Short-Term memory. *Neural Comput* 9: 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
22. Jozefowicz R, Zaremba W, Sutskever I (2015) An empirical exploration of recurrent network architectures. *Proceedings of the 32nd International Conference on Machine Learning*, 2342–2350. PMLR.
23. Telecommunication Standardization Sector of ITU. G.694.2: Spectral Grids for WDM Applications: CWDM Wavelength Grid. Telecommunication Standardization Sector of ITU 2003. Available from: <https://www.itu.int/rec/T-REC-G.694.2-200312-I/en>



AIMS Press

© 2024 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0>)