The Extended SLM Combined Autoencoder of the PAPR Reduction Scheme in DCO-OFDM Systems

Lili Hao 1,2,*,†, Dongyi Wang 2,†, Yang Tao 2, Wenyong Cheng 3, Jing Li 4 and Zehan Liu 1

1 School of Information and Electrical Engineering, Shandong Jianzhu University, Jinan 250101, China; haoxh-888@163.com
2 Bio-Imaging and Machine Vision Lab, Fischell Department of Bioengineering, University of Maryland, College Park, MA 20740, USA; dywang@umd.edu (D.W.); ytao@umd.edu (Y.T.)
3 Advanced Research Center for Optics, Shandong University, Jinan 250100, China; cwy@sdu.edu.cn
4 CETC key laboratory of aerospace information applications, Shijiazhuang 050081, China; lijing030326@foxmail.com
* Correspondence: lilihao81@foxmail.com; Tel.: +1-571-267-8985
† These authors contributed equally to this work.

Received: 2 January 2019; Accepted: 19 February 2019; Published: 27 February 2019

Abstract: End-to-end learning in optical communication systems is a promising technique to solve difficult communication problems, especially for peak to average power ratio (PAPR) reduction in orthogonal frequency division multiplexing (OFDM) systems. The less complex, highly adaptive hardware and advantages in the analysis of unknown or complex channels make deep learning a valid tool to improve system performance. In this paper, we propose an autoencoder network combined with extended selected mapping methods (ESLM-AE) to reduce the PAPR for the DC-biased optical OFDM system and to minimize the bit error rate (BER). The constellation mapping/de-mapping of the transmitted symbols and the phase factor of each subcarrier are acquired and optimized adaptively by training the autoencoder with a combined loss function. In the loss function, both the PAPR and BER performance are taken into account. The simulation results show that a significant PAPR reduction of more than 10 dB has been achieved by using the ESLM-AE scheme in terms of the complementary cumulative distribution function. Furthermore, the proposed scheme exhibits better BER performance compared to the standard PAPR reduction methods.

Keywords: orthogonal frequency division multiplexing; autoencoder; end-to-end learning; peak-to-average power ratio

1. Introduction

Visible light communication (VLC) based on light emitting diodes (LEDs) is a promising technology for indoor wireless access [1–3]. To overcome the multipath distortion caused by reflections from different sources inside a room and enhance the communication efficiency, the optical orthogonal frequency division multiplexing (OFDM) has been widely adopted in VLC systems [4–7]. However, the high peak to average power ratio (PAPR) associated with the OFDM signals is one of the main limitations for VLC systems due to the constraints on the average radiated optical power and the limited dynamic range of the front-end devices, like digital to analog converters and power amplifiers [8]. High PAPR makes the VLC system more susceptible to non-linear distortions and consequently drastically degrade the system’s performance [9–11].
Several PAPR reduction techniques for DC-biased optical orthogonal frequency division multiplexing (DCO-OFDM) systems have been investigated in the literature. The authors proposed a genetic algorithm and peak-value optimization algorithm [12] to mitigate the high PAPR in VLC systems with lower complexity. The study in [13] used a semidefinite relaxation method for tone injection to reduce the PAPR for the DCO-OFDM. The branch and bound method [14] and tone reservation method [15] were also proposed to lower the impact of the high PAPR in VLC systems. Nevertheless, these methods reduce the PAPR at the sacrifice of computational complexity and channel resources. Moreover, the authors in [16] introduced a pilot-assisted approach to achieve improved PAPR performance over the select mapping scheme for high-level constellations. However, it resulted in a data rate loss according to the density of pilot symbols. In addition, a subcarrier grouping scheme for the OFDM-based VLC system [17] is proposed to reduce the PAPR, but at a lower signal-to-noise ratio (SNR) the bit error rate (BER) performance was inferior to that of DCO-OFDM.

To mitigate the effects of these limitations, deep learning offers an efficient option for its good generalization properties with flexible modeling and learning capabilities. Deep learning is a promising technique to solve difficult communication problems [18] for its minimal complicity, adaptive hardware and robustness in the analysis of the unknown or complex channels. Use of deep learning methods to solve difficult communication problems has been reported and has demonstrated better performances than conventional communication methods in the end-to-end learning of encoding and decoding application [19]. Another approach reported in [20] is the end-to-end learning of a prototype consisting of two software-defined radios that communicate over an actual wireless channel. Deep learning techniques for channel estimation and symbol detection in an end-to-end manner [21] are investigated which have wide application in many communication systems. Among them, a special network architecture named autoencoder (AE) which is usually used for denoising corrupted data is suitable to dealing with the non-linear distortions caused by the high PAPR [22]. It optimizes reconstruction loss through a series of representations typically using a mean squared error objective and a stochastic gradient descent solver to find network weights achieving an effective regression [23]. An AE-based system [24] that was solely composed of Neural Networks communicating over-the-air was extended in the OFDM scheme. Kim and Cho [22] proposed a PAPR-reducing network and discussed the PAPR behavior in the RF system. Sohn and Kim [25,26] applied an artificial neural network to reduce the complexity in solving the PAPR reduction problem. However, the focus of these studies was limited to the clipping and filtering technique and active constellation extension signals, which may not acquire better performance than conventional methods when extended to optical OFDM systems.

A novel deep neural network combined with extended Selected Mapping (ESLM), namely ESLM-AE, is proposed in this paper to mitigate the high PAPR issue of DCO-OFDM signals. It uses an AE structure to represent the constellation mapping and de-mapping of the transmitted symbols. In the network, the ESLM method is added after the constellation mapping to reduce the high PAPR of the DCO-OFDM system. By designing the loss function of neural network and considering both the BER and PAPR performance, autoencoder and SLM can be combined organically. Further, the phase factor of SLM can be determined and optimized accordingly in the network training process. Thus, it is expected that the proposed ESLM-AE method is more efficient in reducing the PAPR without deterioration of the BER performance.

The remainder of this paper is organized as follows. Section 2.1 gives an overview of the DCO-OFDM system model. Section 2.2 presents the detailed architecture of the proposed ESLM-AE scheme. Section 3 contains the simulation results and a discussion of the proposed scheme compared to the standard methods in different channels. Finally, conclusions are reported in Section 4.
2. Methods

2.1. An Overview of the DCO-OFDM System Model

Figure 1 shows an overview of the DCO-OFDM transmitter and receiver structure based on the proposed ESLM-AE scheme. Different from the conventional DCO-OFDM system, an AE is applied in the whole model. AE is a special neural network which can represent the mapping from the input to itself because of universal approximation theorem [27,28]. The neurons in the hidden layer of autoencoder can be considered as a coding representation of input.

In our scheme, the input signals are firstly fed into the encoder and phase rotator modules before Hermitian symmetry and inverse fast Fourier transform (IFFT) to get the in-phase and quadrature (I-Q) constellation mapping and generate the alternative output sequence. The outputs of IFFT are then converted into unipolar by adding DC bias and clipping. Signal clipping is performed in order to fit the real time-domain OFDM symbols into a limited range of the LED. In the optical channel, the transmitted signals are influenced by the noise sources in a real scenario. At the receiver side, phase recovery and the decoder part aim to recover the distorted signals. The network can be trained with the lowest PAPR and BER.

It is assumed that the OFDM signal is transformed by 2N subcarriers. Let \( x, f(x) \) and \( g(x) \) be the input of encoder, encoder, and decoder of the AE, respectively. According to the DCO-OFDM system, the transmitted data stream is mapped into complex-value symbols. As shown in Figure 1, after serial to parallel conversion, the input data sequence is divided into 2N messages \( x \), where \( x = [x_0, x_1, \ldots, x_{2N-1}]^T \). Then \( x \) is mapped into I-Q constellation according to the encoder part which will be detailed discussed in Section 2.2. We define the output of the encoder \( X = f(x) \), where \( X \in \mathbb{R}^{2N} \) consists of 2N real values and among them pairwise combination in a certain order forms \( N \) complex symbols \( A, A = [A_0, A_1, \ldots, A_{N-1}]^T, A \in \mathbb{C}^N \).

However, the classic AE is only designed to minimize the BER. In practice, the transceiver usually suffers from a high PAPR. To mitigate the high PAPR, each \( A_k, k = 0,1,\ldots,N-1 \), is multiplied by a phase factor \( a_k \) which can be expressed as \( \overline{A_k} = A_k \cdot a_k \), where \( a_k = e^{i\theta_k} \) and \( \theta_k \in [0, 2\pi) \) [29]. For the VLC-OFDM system, the intensity modulation requires the transmitted signals of the LED to be nonnegative and real-valued [5]. Thus, Hermitian symmetry is imposed on \( \overline{A_k} \) to form the frequency domain OFDM symbols \( S = [\overline{A_0}, \overline{A_1}, \ldots, \overline{A_{N-1}}, \overline{A_{N-1}^*}, \ldots, \overline{A_1^*}] \) and the DC components are \( \overline{A_0} = \overline{A_{N-1}} = 0 \). This results in a 2N-point IFFT output of the OFDM symbols. Subsequently the time domain OFDM signal can be obtained by feeding \( S(k), k = 0,1,\ldots,2N-1 \), into an inverse discrete Fourier transform using Equation (1).

\[
s(n) = \frac{1}{\sqrt{2N}} \sum_{k=0}^{2N-1} S(k)e^{j2\pi nk/2N},
\]
where \( n = 0, 1, \ldots, 2N - 1 \).

Consequently, the PAPR of the DCO-OFDM signal is calculated using Equation (2).

\[
PAPR\{s(n)\} = \frac{\max_{0 \leq n \leq 2N - 1} |s(n)|^2}{E[|s(n)|^2]},
\]

where \( n = 0, 1, \ldots, 2N - 1 \). When the high amplitudes of different subcarriers with the same phase appear at the same time, the high PAPR will appear. Complementary cumulative distribution function (CCDF) is used to denote the probability that the PAPR of signals will exceed a given threshold value, \( PAPR_0 \), i.e., \( CCDF = \text{Pr}(PAPR > PAPR_0) \).

After the parallel to serial conversion and adding a cyclic prefix (CP), the DC bias and clipping are added to the time-domain discrete signals \( s(n) \) to ensure all the signal amplitudes are nonnegative. In VLC systems, the transmitted signal has to be constrained in the linear range due to the nonlinear characteristics of LED [30]. Accordingly, \( s(n) \) is subjected to amplitude clipping at given upper (\( \xi_{upper} \)) and lower (\( \xi_{lower} \)) levels before fed into LED. Assume the linear range of LED is \( [0, 2\xi_{upper}] \) and the symmetric clipped signal of \( s(n) \) is given in Equation (3) [31].

\[
x_{\text{Clipped}}(n) = \begin{cases} 
\xi_{upper}, & s(n) > \xi_{upper} \\
\xi_{lower} \leq s(n) \leq \xi_{upper}, & \\
\xi_{lower}, & s(n) < \xi_{lower}
\end{cases}
\]

where \( n = 0, 1, \ldots, 2N - 1 \). We define \( \xi_{upper} = -\xi_{lower} = \sqrt{E[|s(n)|^2]} \) and \( \gamma \) is referred to the clipping ratio. To assure a steady light intensity and maximize the modulation depth, the DC bias is set as \( B_{DC} = \xi_{upper} \), Then the DCO-OFDM signals fed into LED are expressed as Equation (4).

\[
x_{\text{DCO}}(n) = x_{\text{Clipped}}(n) + B_{DC} = \begin{cases} 
2\xi_{upper}, & s(n) > \xi_{upper} \\
\xi_{upper}, & s(n) + \xi_{upper} \leq s(n) \leq \xi_{upper}, \\
0, & s(n) < -\xi_{upper}
\end{cases}
\]

where \( n = 0, 1, \ldots, 2N - 1 \). Essentially, clipping noise will degrade the performance of the DCO-OFDM system due to the non-linear distortion.

Subsequently the unipolar signal drives the LED to converts the electrical signals to the optical signals. In the optical channel, the line-of-sight (LOS) links are assumed to dominate over all multipath components from the wall and ceiling reflections. The received signal is influenced by the noise sources in a real scenario. The dominant noise source in an indoor wireless optical channel is the ambient light induced shot noise [32], which is modeled as the additive white Gaussian noise (AWGN) given by \( Z = [Z_0, Z_1, \ldots, Z_{2N-1}]^T \). It has been considered that there is no additional clipping introduced by the photodetector (PD). Thus, after converting the optical signals to electrical signals using PD, the received signal, \( y = [y_0, y_1, \ldots, y_{2N-1}]^T \) can be computed using Equation (5).

\[
y = x_{\text{DCO}} \otimes q + Z,
\]

where \( x_{\text{DCO}} \) is the vector of DCO-OFDM signal \( x_{\text{DCO}}(n) \), \( q \) is the channel response, \( Z \) is AWGN with zero mean and variance of \( \sigma_Z^2 \) and \( \otimes \) denotes the convolution operation. We assume that the channel state information is perfectly known in advance.

Then a reverse process can be implemented to demodulate the data. After removing the DC bias, the corresponding vector \( y \) passes through the fast Fourier transform (FFT) operation. A simplified representation of the output \( Y \) is shown in Equation (6).

\[
Y = FFT\{Q \circ IFFT[f(x)] + \epsilon\},
\]
where \( Q \) denotes the effects of the optical channel and \( \varepsilon \) is the noise at the receiver. Finally \( Y \) is transformed to the decoder \( g(Y) \), which functions as the constellation mapper to get the recovered symbol \( \hat{x} \).

### 2.2. The ESLM-AE PAPR Reduction Scheme

The neural network, which is an important statistic tool, can be used for describing the relationship between the input and the output. Its parameters can be determined automatically using backpropagation given a particular loss function. The proposed ESLM-AE uses an AE structure combined with ESLM method to mitigate the high PAPR issue of DCO-OFDM signals.

#### 2.2.1. Autoencoder Network

In our scheme, an AE network is applied to the DCO-OFDM system to optimize the end-to-end performance.

Usually a feedforward neural network (NN) with \( L \) layers describes a mapping \( f(r_0; \theta) : R^{N_0} \rightarrow R^{N_L} \) of an input vector \( r_0 \in R^{N_0} \) to an output vector \( r_L \in R^{N_L} \) through \( L \) iterative processing steps [33]:

\[
\begin{align*}
\ell & = f_\ell(r_{\ell-1}; \theta_\ell), \ell = 1, \ldots, L
\end{align*}
\]

Where \( f_\ell(r_{\ell-1}; \theta_\ell) : R^{N_{\ell-1}} \rightarrow R^{N_\ell} \) is the mapping carried out by the \( \ell \)th layer and \( \theta = \{\theta_1, \ldots, \theta_L\} \) is used to denote the set of all parameters of the network. The \( \ell \)th layer is called dense or fully-connected (FC) layer if \( f_\ell(r_{\ell-1}; \theta_\ell) \) has the form

\[
\begin{align*}
f_\ell(r_{\ell-1}; \theta_\ell) &= \rho(W_\ell r_{\ell-1} + b_\ell),
\end{align*}
\]

Where \( W_\ell \in R^{N\ell \times N^{\ell-1}}, b_\ell \in R^{N_\ell} \), are the weights and bias for the \( \ell \)th layer respectively. \( \rho(\cdot) \) is an activation function and the set of parameters for this layer is \( \theta_\ell = \{W_\ell, b_\ell\} \).

For the classic AE, its expected output is the input, which is different from other neural networks. Therefore, the AE can be trained from scratch without supervision, and a multilayer network can represent a mapping from the input to the expected output, identity as the input. The AE has been applied in many communication fields such as channel encoding and decoding, channel compensation and modulation recognition [22–24].

A brief illustration of the proposed AE system is shown in Figure 2 [20]. The transmitter part is called the encoder and it maps the input signals into the I-Q constellation. We assume both encoder and decoder are composed of \( L_L = L_R = 3 \) sub-blocks. Each of the sub-blocks is composed of dense layer, batch normalization (Batchnorm), activation function and dropout.

![Figure 2. Illustration of the proposed autoencoder system.](image)

Let \( r^{\ell}_{\ell} \) be the input of the \( \ell \)th dense layer of the encoder and the output can be expressed as \( h^{\ell}_{\ell} = W^{\ell}_{\ell} r^{\ell}_{\ell} + b^{\ell}_{\ell} \), where \( W^{\ell}_{\ell} \) and \( b^{\ell}_{\ell} \) are the weights and bias for the \( \ell \)th layer. The output of each dense
layer passes through the batch normalization (Batchnorm) layer to minimize the internal covariate shift. The Batchnorm can be mathematically expressed as $\| h^f_\ell \|_{\text{norm}} = \alpha \frac{h^f_\ell - E[h^f_\ell]}{\sqrt{\text{Var}[h^f_\ell]} + \nu} + \beta$, where $\alpha$ and $\beta$ are the scaling and shift factors, respectively. $\nu = 0.001$ is a constant which prevents the division by zero [22].

Then the normalized value is fed into the activation function to make the data features nonlinear. The activation functions used in our scheme are the rectifier linear unit (ReLU) [33] and sigmoid [34] which are defined as $\max(\| h^f_\ell \|_{\text{norm}}, 0)$ and $\frac{1}{1+e^{-h^f_\ell}}$ respectively. In the encoder part the activation function used in each of the sub-blocks is ReLU.

Finally, dropout is used for addressing the overfitting problem for the proposed AE network, which has large number of parameters. The key idea is to randomly drop units from the neural network during training, which significantly reduces overfitting and gives major improvements over other regularization methods [35].

As is shown in Figure 2, the output of the encoder is then transmitted to the simulated channel and decoder. Decoder has a similar structure as the encoder network. The only difference is the activation function of the last sub-block is sigmoid. It aims to recover the original binary information from the distorted signal after the complicated channel transmission.

Mathematically, the output of the encoder can be expressed as Equation 9 [22].

$$X = f(x) = \rho_{L_f}(W^f_{L_f} \rho_{L_f-1} \ldots \rho_{1}(W^f_{1} x + b^f_{1} |_{\text{norm}}) \ldots) + b^f_{L_f} |_{\text{norm}},$$

(9)

where $W^f_{L_f}$ and $b^f_{L_f}$ are the weights and bias for the $L_f$th dense layer of the encoder respectively.

Similarly, the output of the decoder can be expressed as Equation (10).

$$\hat{x} = g(Y) = \rho_{L_d}(W^g_{L_d} \rho_{L_d-1} \ldots \rho_{1}(|W^g_{1} Y + b^g_{1} |_{\text{norm}}) \ldots) + b^g_{L_d} |_{\text{norm}},$$

(10)

where $W^g_{L_d}$ and $b^g_{L_d}$ are the weights and bias for the $L_d$th dense layer of the decoder respectively.

As mentioned, the noise channel could distort the signal during the transmission. The AE aims to find an adequate encoding and decoding strategy to eliminate the complex optical channel and noise interference. To achieve the objective, the first network loss function can be set as the reconstruction error given in Equation (11).

$$\text{Loss}_1(x, \hat{x}) = \| x - \hat{x} \|_2^2,$$

(11)

For the training of the AE, the stochastic gradient descent (SGD) [33] optimization method is popular used which starts with some random initial values of $\theta = \theta_0$ and update $\theta$ iteratively as

$$\theta^+ = \theta - \lambda \nabla_\theta \text{Loss}_1(x, \hat{x}),$$

(12)

where $\lambda > 0$ is the learning rate, $\theta$ denotes parameters of the AE and $\nabla_\theta$ denotes the Gradient operation $\frac{\partial \text{Loss}_1(x, \hat{x})}{\partial \theta}$.

2.2.2. Extended Selected Mapping Technique

For the DCO-OFDM, a high signal peak value implies a need for a large DC bias that causes serious degradation of the system’s power efficiency. Therefore, inspired by SLM, a reduction scheme, named extended Selected Mapping method is given in this section.

The Selected Mapping (SLM) method is one popular scheme to reduce the PAPR, because it is simple to implement without introducing any distortion to the signal and it can be used with any subcarrier number and modulation style [36]. The principle is that $u$ copies of the complex data
vector $C_u = [C_{0u}^u, C_{1u}^u, \cdots, C_{N-1}^u]$ are multiplied by $u$ distinct phase vectors $B_u = [B_{0u}^u, B_{1u}^u, \cdots, B_{N-1}^u]$, $u = 1, 2, \cdots, U$. The corresponding time domain data vector after IFFT is shown in Equation (13).

$$c_n^u = IFFT\{C_k^u \cdot B_k^u\},$$  \hspace{1cm} (13)$$

where $n = 0, 1, \cdots, N - 1$.

The objective of the SLM technique is to determine the transmitted $c_n^u$ using Equation (14) [29].

$$\min_{(B^u)} \left( \max \left\{ \frac{|c_n^u|^2}{N\sum_{n=0}^{N-1}|c_n^u|^2}, n = 0, 1, \ldots, N - 1 \right\} \right),$$  \hspace{1cm} (14)$$

To improve the PAPR reduction performance of the SLM scheme, the SLM technique requires an increase in the number of phase sequences. However, the computational complexity of the SLM scheme linearly increases as the number of phase sequences increases. The alternative method is an extensive search for the optimal sequence that achieves the minimum PAPR. Consequently, in the SLM, the PAPR reduction is primarily dependent on the chosen phase sequence candidates [37].

In this paper, we extended the SLM technique to AE to get the adaptive phase sequence shown in Figure 3. Each phase factor $a_k$ of $A_k$ no longer needs to be artificially arranged since it can be trained and continuously optimized in the deep learning network. Simultaneously, in the test, once the phase sequence is determined, the calculation of the IFFT is needed only once.

![Figure 3. Partial block diagram of an OFDM transmitter with the extended Selected Mapping (SLM) technique.](image)

In our proposed scheme, the network is trained to reduce the PAPR without reducing the BER performance [38]. Consequently, two distinct factors must be taken into the account at the same time. To reduce PAPR value, we define the second loss component $Loss_2(x)$ which is given as Equation (15).

$$Loss_2(x) = PAPR\{s(n)\},$$  \hspace{1cm} (15)$$

Based on simulation, the $Loss_2(x)$ is helpful to reduce the high PAPR and lower distortion will in turn improve the BER performance in the training process.

Considering the two factors, we use a hyperparameter $\eta$ to balance the two different loss components. Thus, the total loss function can be expressed in Equation (16).

$$Loss(x, \hat{x}) = Loss_1(x, \hat{x}) + \eta Loss_2(x),$$  \hspace{1cm} (16)$$

Notice that each phase factor $a_k$ of $A_k$ can be trained in terms of $\frac{dLoss}{da_k}$ based on the propagation algorithm [39].
3. Results and Discussion

In this section, simulation results are presented to demonstrate the PAPR and BER performances of the proposed scheme in different channels. The parameters of the network are shown in Table 1.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transmitter (TX) dimensions</td>
<td></td>
</tr>
<tr>
<td>Input</td>
<td>128</td>
</tr>
<tr>
<td>Dense (ReLU)</td>
<td>2048</td>
</tr>
<tr>
<td>Dense (ReLU)</td>
<td>2048</td>
</tr>
<tr>
<td>Dense (ReLU)</td>
<td>128</td>
</tr>
<tr>
<td>Receiver (RX) dimensions</td>
<td></td>
</tr>
<tr>
<td>Input</td>
<td>128</td>
</tr>
<tr>
<td>Dense (ReLU)</td>
<td>2048</td>
</tr>
<tr>
<td>Dense (ReLU)</td>
<td>2048</td>
</tr>
<tr>
<td>Dense (sigmoid)</td>
<td>128</td>
</tr>
<tr>
<td>Optimizer</td>
<td>SGD with Adam [40]</td>
</tr>
<tr>
<td>Learning rate $\lambda$</td>
<td>0.0001</td>
</tr>
<tr>
<td>Batch size</td>
<td>32</td>
</tr>
<tr>
<td>Weight parameter $\eta$</td>
<td>0.01</td>
</tr>
<tr>
<td>Dropout probability</td>
<td>0.1</td>
</tr>
<tr>
<td>Number of subcarriers</td>
<td>128</td>
</tr>
<tr>
<td>Length of cyclic prefix</td>
<td>32</td>
</tr>
</tbody>
</table>

In the training of the proposed network, a total of 64,000,000 independent random bits are used for training, 12,800,000 bits for validation and 12,800,000 bits for testing, respectively. Taking $\text{SNR} = 15$ dB for an example, the corresponding average PAPR and BER results of training set, validation set, and test set are given in Table 2. Note that all the following simulation and discussion of the proposed ESLM-AE and AE schemes are based on the results of test set.

<table>
<thead>
<tr>
<th>Results</th>
<th>Average PAPR</th>
<th>BER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training set</td>
<td>2.0198</td>
<td>0.0003976</td>
</tr>
<tr>
<td>Validation set</td>
<td>2.0499</td>
<td>0.0004077</td>
</tr>
<tr>
<td>Test set</td>
<td>2.0439</td>
<td>0.0004059</td>
</tr>
</tbody>
</table>

For comparison, we also investigate the performances of other PAPR reduction schemes such as basic AE network without ESLM method, classical SLM [41] using $U = 128, B_{DC} = 7$ dB and amplitude clipping [30] with a different clipping ratio $\gamma$. All the simulation results are taken from 100,000 OFDM symbols, and 4-quadrature amplitude modulation is adopted. The number of phase sequences for SLM considered here are significantly high to maintain the SLM performance depicted here close to the upper-bound on the scheme’s performance. The schemes are listed as follows:

Clipped DCO-OFDM and Clipping with clipping ratio $\gamma$ refer to the conventional DCO-OFDM sample values which are larger than the upper levels or less than the lower levels are directly upper or lower clipped without any other recovery methods.

ESLM-AE and AE with clipping ratio $\gamma$ refer to the ESLM-AE and AE sample values are symmetrically clipped similar to the direct upper and lower clipping implementation.

To compare the computation complexity of SLM and the proposed ESLM-AE, the Big O notations of the two methods are given below. For SLM with U groups of phase factors. For each group, the most time-consuming part is the process of the IFFT, with time complexity of $O(N^2)$, where $N$ is the number of subcarriers. To select the best phase factor for PAPR reduction, the time complexity of the SLM algorithm is $O(U N^2)$, where $U$ is the total number of phase vectors. For the ESLM-AE algorithm,
the training process is offline. Once the network weights are determined, when new messages come, only the forward propagation of network are needed. Assume there are \( L_f \) hidden layers in encoder network, and \( L_g \) hidden layers in the decoder network. All the hidden layers are assumed to have \( M \) neurons. The computation complexity of encoder network is \( O(2 * N * M + (L_f - 1) * M * M) \). Similarly, the computation complexity of decoder network is \( O(2 * N * M + (L_g - 1) * M * M) \). In the ESLM-AE, the phase factor is also determined in the training phase, and only one IFFT is needed to be computed which computation complexity is \( O(N^2) \). Therefore, the computation complexity of ESLM-AE is \( O(4 * N * M + (L_f + L_g - 2) * M * M + N * N) \). In practice, \( U, N \) and \( M \) are in the same scale. The time complexity of original SLM is \( O(N^3) \), while the time complexity of ESLM-AE is \( O(N^2) \), the same as the clipping method.

3.1. PAPR Comparison

First, we evaluated the PAPR performance of the presented system compared to AE, SLM, DCO-OFDM with different clipping ratio \( \gamma \) and the DCO-OFDM with no PAPR reduction scheme. CCDF curves are presented to illustrate the PAPR comparison results depicted in Figure 4. It can be observed from Figure 4 that the PAPR of the proposed ESLM-AE method outperforms other methods with a PAPR reduction gain of 10.8 dB compared to DCO-OFDM, while the SLM \( U = 128 \) gives the least PAPR reduction gain of 4.9 dB. To reach a CCDF of \( 10^{-3} \), for example, the PAPR is 2.5 dB for the ESLM-AE, 3.1 dB for the AE, 8.2 dB for the SLM \( U = 128 \), 5.7 dB and 6.0 dB for the clipped DCO-OFDM with \( \gamma = 1.2 \) and \( \gamma = 1.5 \). Additionally, the AE method exhibits a higher PAPR than the proposed ESLM-AE, because the ESLM-AE signals are jointly processed with the extended SLM technique in the deep learning network. Therefore, the proposed ESLM-AE method enjoys a significant PAPR reduction in terms of CCDF, which can diminish the linearity requirement of the LED.

3.2. BER Analysis

Then, we investigated the BER performance of the DCO-OFDM system when the optical channel is assumed to be a LOS channel modeled by an AWGN channel. As seen in Figure 5, the comparison

![Figure 4. Complementary cumulative distribution function (CCDF) of peak to average power ratio (PAPR) comparison of ESLM-AE, Autoencoder, clipped DCO-OFDM, DCO-OFDM and SLM. The clipping ratios, \( \gamma \), used for ESLM-AE and Autoencoder is 1.5.](image-url)
results show that the proposed scheme can result in a lower BER compared to the conventional methods in the whole SNR range. At BER $=10^{-3}$, the SNR requirements of the ESLM-AE are 4 dB, 4.7 dB and 7 dB less than that of the SLM $U = 128, B_{DC} = 7$ dB, clipped DCO-OFDM $\gamma = 1.5$ and clipped DCO-OFDM $\gamma = 1.2$ respectively. Additionally, the BER performance of ESLM-AE and AE is almost same. Since the error due to the clipping results in an increased BER in the DCO-OFDM at high SNR values, a larger DC Bias and a lower BER are achieved by using the proposed scheme in the high SNR range.

![Figure 5](image-url)  
**Figure 5.** Bit error rate (BER) comparison of the clipped DCO-OFDM, SLM, Autoencoder and the proposed scheme under the line-of-sight (LOS) channel.

In VLC systems, the channel is generally modeled as a multi-path propagation environment. Consequently, in our scenario, Rician distribution is taken into account to simulate the multipath effects with the LOS path to be dominant [42–44]. The Rician $K$-factor gives the ratio of the squared signal power of the LOS link over that of the signal from the non-LOS link [45]. Here $K = 1$ and 5 is taken into consideration and the BER performance of the proposed method with Rician effects and AWGN is given in Figure 6. Notably compared to SLM $U = 128$, signal deterioration of the proposed ESLM-AE scheme is much smaller. Simultaneously, the system inside an environment with Rician fading $K = 5$ needs an additional 2.8 dB SNR in order to obtain the same performance (BER $=10^{-4}$) as when it operates inside an AWGN channel due to the multipath fading. With the increase of $K$, the probability of encountering a deep fade reduces. On the contrary, if $K$ decreases, the dominant path degenerates in amplitude. When $K$ is reduced to 0, the Rician distribution reverts to Rayleigh.

Simultaneously, a comparison with the experiment results is carried out to verify the BER performance of the proposed system under a diffused optical wireless (DOW) channel. The DOW channel is modeled by the sum of a set of positive taps as Equation (17) [46].

$$q(t) = \sum_{n=0}^{V-1} p_n \delta(t - \tau_n),$$  \hspace{1cm} (17)

where $q(t)$ is the channel impulse response at the time slot $t$, $p_n$ and $\tau_n$ are the amplitude and time delay of the $n$th path, $V$ is the number of channel taps.
Figure 6. BER comparison of the ESLM-AE and SLM under the Rician fading channel with additive white Gaussian noise (AWGN).

The diffuse fading follows the exponentially decaying and ceiling bounce models as described in [47] for DOW channels. In the simulation, 32-sample long CP is inserted. We set $V = 11$ and suppose that the time delay is uniformly distributed from 2 to 20 ns. Figure 7 demonstrates the improvement in the BER performance by adopting the proposed scheme under the DOW channel. Note that the results in Figure 6 are achieved by setting a CP larger than the multipath delay where no inter-symbol interference (ISI) exists. Numerically, the BER of the proposed ESLM-AE scheme can reach and even be less than the order of magnitude of $10^{-3}$, which is much better than the clipped DCO-OFDM and SLM methods, although the BER performances of all schemes degrade due to the multipath fading.

Figure 7. BER comparison of the clipped DCO-OFDM, SLM, Autoencoder and the proposed scheme under the DOW channel.
To demonstrate the influence of ISI, we set a shorter CP of 4-samples while the maximum delay of the multipath channel is 11. Figure 8 gives the comparison of BER performance of the above methods in DOW channel with and without ISI. SLM $U = 128$ and clipping ratio $\gamma = 1.5$ are adopted in the simulation. We can observe that the proposed scheme can even adapt to the ISI with only smaller performance degradation than the curve without ISI. That is because trainable parameters of the network can be used to compensate the multiple effects of the optical channel. Moreover, in the presence of ISI, the performance of the proposed ESLM-AE is still superior as compared to SLM and Clipping methods in the whole SNR range. From the performance results, we can conclude that the ESLM-AE scheme outperforms conventional PAPR reduction methods in terms of both the BER and PAPR. In addition, the decrease in the PAPR transforms to a BER performance gain both in the LOS channel and the DOW channel due to the mitigation of nonlinear distortion caused by the nonlinearities of the LED.

![Figure 8. BER comparison of the clipped DCO-OFDM, SLM, Autoencoder and the proposed scheme under the DOW channel with/without ISI. The clipping ratios, $\gamma$, used for clipping, ESLM-AE and Autoencoder are 1.5.](image)

4. Conclusions

This paper proposes an ESLM-AE network to reduce the PAPR value for the DCO-OFDM system without reducing the BER performance. The constellation mapping and de-mapping of the symbols and phase factor of each subcarrier are trained adaptively using a combined loss function of the AE with two different loss components. The simulation results show that our proposed scheme significantly outperforms conventional schemes in terms of both the PAPR and BER. By using our proposed technique, a distinct PAPR reduction of more than 10 dB is achieved for the VLC system, which significantly relieves the linear requirement of the front-end devices. Moreover, the decrease in the PAPR transforms to a BER performance gain both in the LOS channel and the DOW channel. The proposed scheme is a primary study on the application of deep learning networks in VLC systems. Its simulation results are very promising, and we are working on its implementation in digital signal processing. It can give insights on combining a deep learning framework with conventional communication methods for different applications.
Author Contributions: Data curation, L.H.; Formal analysis, L.H. and D.W.; Funding acquisition, W.C.; Investigation, L.H.; Methodology, L.H. and D.W.; Project administration, J.L.; Software, L.H.; Supervision, Y.T.; Visualization, Z.L.; Writing—original draft, L.H.; Writing—review & editing, D.W. and Y.T.

Funding: This research was funded by the Fundamental Research Funds of Shandong University (Grant No. 2015JC043) and the Independent Innovation Program of Universities and Institutes in Jinan City (Grant No. 201303007).

Conflicts of Interest: The authors declare no conflict of interest.

References


34. Ito, Y. Representation of functions by superpositions of a step or sigmoid function and their applications to neural network theory. Neural Netw. 1991, 4, 385–394. [CrossRef]


43. Yang, B.; Letaief, K.B.; Cheng, R.S.; Cao, Z. Channel estimation for OFDM transmission in multipath fading channels based on parametric channel modeling. IEEE Trans. Commun. 2001, 49, 467–479. [CrossRef]


© 2019 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 (http://creativecommons.org/licenses/by/4.0/).