

Deep Learning-Aided Optical IM/DD OFDM Approaches the Throughput of RF-OFDM

Thien Van Luong, Xiaoyu Zhang, Luping Xiang, Tiep M. Hoang, Chao Xu, Periklis Petropoulos, and Lajos Hanzo

Abstract—Deep learning-aided optical orthogonal frequency division multiplexing (O-OFDM) is proposed for intensity modulated direct detection transmissions, which is termed as O-OFDMNet. In particular, O-OFDMNet employs deep neural networks (DNNs) for converting a complex-valued signal into a non-negative signal in the time-domain at the transmitter and vice versa at the receiver. The associated frequency-domain signal processing remains the same as in conventional radio frequency (RF) OFDM. As a result, our scheme achieves the same spectral efficiency as the RF scheme, which has never been attained by the existing O-OFDM schemes, because they have relied on the Hermitian symmetry of the spectral-domain signal to guarantee that the time-domain signal becomes real-valued. We show that O-OFDMNet can be viewed as an autoencoder architecture, which can be trained in an end-to-end manner in order to simultaneously improve both the bit error ratio (BER) and the peak-to-average power ratio (PAPR) for transmission over both additive white Gaussian noise and frequency-selective channels. Furthermore, we intrinsically integrate a soft-decision aided channel decoder with our O-OFDMNet and investigate its coded performance relying on both convolutional and polar codes. The simulation results show that our scheme improves both the uncoded and coded BER as well as a reducing the PAPR compared to the benchmarks at the cost of a moderate additional DNN complexity. Furthermore, our scheme is capable of approaching the throughput of RF-OFDM, which is notably higher than that of conventional O-OFDM. Finally, our complexity analysis shows that O-OFDMNet is suitable for real-time operation.

I. INTRODUCTION

Orthogonal frequency division multiplexing (OFDM) has been widely adopted in wireless communication systems [1], [2], where the radio frequency (RF) hardware is capable of supporting complex-valued signal processing [3], [4]. As an evolution from RF-OFDM, optical OFDM (O-OFDM) [5] has been proposed both for visible light communications (VLC)

and optical fibers in order to exploit its robustness to inter-symbol interference (ISI) and its simplicity of single-tap channel equalization under dispersive channels. In VLC, as well as in free-space optical (FSO) and fiber-optic communications, low-complexity intensity modulation paired with direct detection (IM/DD) [6] has been commonly used, which requires the transmitted signal to be real- and positive-valued. Explicitly, in IM/DD systems, the intensity of the emitted light-wave is used for modulating the signal, while a photo detector is employed at the receiver for converting the optical signal back into the electronic domain. In this context, substantial efforts have been invested in modifying RF-OFDM in order to create a real and positive signal in the time domain (TD), i.e. after inverse fast Fourier transform (IFFT), which is a bipolar and complex-valued operation [5].

The constraint of frequency-domain (FD) Hermitian symmetry has been commonly exploited in conjunction with quadrature amplitude modulation (QAM) to make the TD modulated signal after IFFT real-valued. However, we still have to eliminate the negative parts of the post-IFFT signal. Motivated by this, the DC-biased optical OFDM (DCO-OFDM) concept was proposed by Carruthers and Kahn in [7], which adds a DC bias to the bipolar, real-valued post-IFFT signal, so that most samples of the resultant signal become positive-valued, while the remaining low-amplitude negative samples can be clipped at zero. The DC bias must be carefully selected so that the non-linear clipping distortion is minimized. In [8], the DC bias of DCO-OFDM is suggested to be at least 10 dB higher than the data-bearing bipolar signal. Hence, DCO-OFDM has a low power-efficiency due to the wasteful DC bias. Thus, asymmetrically clipped OFDM (ACO-OFDM) [9] was proposed by Armstrong and Lowery for improving the power efficiency (PE) at the cost of reduced spectral efficiency (SE). Explicitly, the ACO-OFDM signals only convey useful information on the odd-indexed subcarriers, whose complex-valued M -ary symbols are arranged to obey the Hermitian symmetry, while the even-indexed subcarriers do not carry useful payload. This design allows ACO-OFDM to clip the post-IFFT signal at zero without any information loss, since the resultant non-linear clipping distortion products fall only on the inactive subcarriers. A similar scheme termed as the pulse-amplitude modulated discrete multitone (PAM-DMT) arrangement was proposed in [10], which modulates the imaginary part of subcarriers and leaves their real part blank. Note that although ACO-OFDM and PAM-DMT completely remove the DC bias, they only achieve half the SE of DCO-OFDM. Hence, various hybrid schemes, which combine the

The financial support of the Engineering and Physical Sciences Research Council projects EP/P034284/1 and EP/P003990/1 (COALESCE), EP/S002871/1 as well as of the European Research Council's Advanced Fellow Grant QuantCom (Grant No. 789028) is gratefully acknowledged.

T. V. Luong is with the Faculty of Computer Science, Phenikaa University, Hanoi 12116, Vietnam, and also with the Phenikaa Research and Technology Institute (PRATI), A&A Green Phoenix Group JSC, Hanoi 11313, Vietnam, (e-mail: thien.luongvan@phenikaa-uni.edu.vn).

L. Xiang, T. M. Hoang, C. Xu and L. Hanzo are with the School of Electronics and Computer Science, University of Southampton, Southampton, SO17 1BJ, U.K. (e-mails: {l.xiang, tiep.hoang, cx1g08, lh}@soton.ac.uk).

X. Zhang is with the school of Electronics and Computer Science, and with the Optoelectronics Research Centre, University of Southampton, Southampton, SO17 1BJ, U.K. (e-mail: xy.zhang@soton.ac.uk).

P. Petropoulos is with the Optoelectronics Research Centre, University of Southampton, Southampton, SO17 1BJ, U.K. (e-mail: pp@orc.soton.ac.uk).

forementioned methods, have been proposed in [11], [12]. These hybrid schemes have a fixed SE, which is dependent on the specific choice of their component schemes, but they cannot exceed the SE of DCO-OFDM. With the objective of striking a flexible trade-off between the SE and the PE, layered ACO-OFDM (LACO-OFDM) has been proposed by Wang *et al.* [13], which exploits the inactive subcarriers in ACO-OFDM to employ additional ACO layers. By increasing the number of ACO layers, LACO-OFDM improves the SE approaching half that of RF-OFDM at the cost of an increased transceiver complexity, since more IFFT/FFT operations are required [14]. Additionally, LACO-OFDM combined with channel coding has been intensively investigated in [15]–[17] and its capacity has been derived in [18].

Historically speaking, learning-aided wireless transceivers have been investigated for 30 years [19]–[21]. However, at the time of writing deep learning (DL) [22] has found numerous further applications in communication systems. For example, deep neural networks (DNNs) have been used for channel estimation and signal detection in classical RF-OFDM [23] and in OFDM combined with index modulation [24], respectively. An application of DL in noncoherent energy detection-based systems was reported in [25], where both the transmitter and receiver are modelled by DNNs based on an autoencoder (AE) structure [22], facilitating joint transceiver optimization via end-to-end training. Similarly, a coherent multicarrier AE-based system has been proposed in [26], which maximizes the diversity and coding gains over fading channels. DL has also been applied to optical communications. For instance, in [27], a DL-aided detection scheme was designed by Amran *et al.* for an indoor VLC system using DCO-OFDM, offering a bit error ratio (BER) close to that of the optimal maximum likelihood detector. An AE structure was applied to an on-off keying (OOK) transceiver of a dimmable VLC systems in [28]. Similarly, an AE based on recurrent neural networks was employed for optical IM/DD fiber-based systems in [29]. Additionally, an end-to-end AE-based VLC system was proposed in [30] both for flicker reduction and dimming. An ACO-OFDM-based VLC system using a hybrid AE scheme was advocated in [31] for improving the BER and peak-to-average power ratio (PAPR) of classical ACO-OFDM, while DL was adopted in multi-colored VLC systems in [32] for reducing the symbol error probability.

It is worth noting that all existing O-OFDM systems achieve at most half of the SE of RF-OFDM due to the FD Hermitian symmetry constraint. Moreover, they suffer from high PAPR inherited from the IFFT based modulation of OFDM, which causes severe nonlinear signal distortion, thereby degrading the system performance. Note that because the TD signal of O-OFDM must be real-valued and positive, some of the PAPR reduction methods conceived for RF schemes [33], [34] are not applicable. Recently, LACO-OFDM has been shown to be able to reduce the PAPR compared to ACO-OFDM through increasing the number of layers [14]. However, this comes at the cost of significantly increased transceiver complexity. Furthermore, recall that the DCO-OFDM schemes have a low power-efficiency owing to their excessive DC bias requirement. Hence, in this contribution, we circumvent these

TABLE I
CONTRIBUTION COMPARISON OF DEEP LEARNING-AIDED OPTICAL COMMUNICATIONS

Papers Topics	[27]	[28]	[29]	[31]	[30]	Our Work
OFDM framework	✓			✓		✓
Autoencoder structure		✓	✓	✓	✓	✓
BER improvement	✓	✓	✓	✓	✓	✓
PAPR reduction				✓		✓
Channel coding						✓
Non-Hermitian symmetry					✓	✓
Full spectral efficiency						✓
Time-domain neural networks						✓
Signal-dependent noise	✓					✓
Clipping distortion				✓		✓

impediments by applying DL with the aid of the AE structure to O-OFDM. Our new contributions are boldly and explicitly contrasted to the literature in Table I, and are summarized as follows:

- We propose a DL-aided O-OFDM scheme termed as O-OFDMNet, where the post-IFFT TD modulated signal of O-OFDM is further encoded and detected by DNNs based on an AE structure. In order to produce real- and positive-valued symbols in the TD, the classic sigmoid activation [22] is utilized as the output layer of the transmitter's DNN. In contrast to the existing O-OFDM schemes [7]–[14], this design completely dispenses with the requirement of Hermitian symmetry, hence matching the ultimate SE of RF-OFDM. In other words, all available subcarriers are actively used for carrying QAM data symbols, since the FD signal processing of O-OFDMNet is the same as that of classical RF-OFDM.
- The proposed O-OFDMNet is trained for simultaneously minimizing the BER and the PAPR for transmission over both additive white Gaussian noise (AWGN) and over frequency-selective (FS) channels. Our simulation results demonstrate that our proposed scheme simultaneously improves both the BER and the PAPR. Moreover, our scheme approaches the throughput of RF-OFDM, which is significantly higher than that of the existing O-OFDM schemes. Another unique contribution is that we intrinsically amalgamate O-OFDMNet with both convolutional and polar codes, using both hard and soft decoders.
- We analyze the complexity of O-OFDMNet in comparison to the existing O-OFDM schemes, which shows that given its moderate complexity, our scheme is eminently suitable for real-time applications.
- Our numerical results quantify the benefits of our O-OFDMNet in terms of its throughput, uncoded/coded BER, PAPR and runtime complexity over the existing O-OFDM schemes, under both AWGN and frequency-selective channels. Furthermore, our results show that our scheme is more robust to practical optical channel conditions, such as signal-dependent noise and nonlinear clipping distortion, than the benchmarks.

The rest of the paper is organized as follows. Section II

introduces the overview of O-OFDM systems, which motivates our new O-OFDMNet philosophy presented in Section III. Our simulation results are reported in Section IV and we conclude the paper in Section V.

Notations: The FFT and IFFT transforms are denoted by \mathcal{F} and \mathcal{F}^{-1} , respectively. The transpose operation and the Frobenius norm are represented by $(\cdot)^T$ and $\|\cdot\|$, respectively. \mathbb{R}^n and \mathbb{C}^n denote the real and complex coordinate spaces of dimension n , respectively. The convolution operation is denoted by \otimes . The expectation and variance operations are denoted by $\mathbb{E}[\cdot]$ and $\text{Var}[\cdot]$, respectively. The complex conjugate of complex number Z is denoted by Z^* .

II. OVERVIEW OF OPTICAL OFDM SYSTEMS

In this section, we review a number of important O-OFDM schemes and discuss their limitations, which motivate us to propose our new deep learning-aided optical OFDM scheme doubling the SE. In particular, the conventional RF-OFDM is specifically designed for transmitting complex-valued signal over frequency-selective channels by using single-tap equalization. Thus, most of the O-OFDM schemes, such as DCO-OFDM [7], ACO-OFDM [9] and LACO-OFDM [13], rely on FD Hermitian symmetry for accommodating the complex M -ary symbols in the FD, so that the resultant TD representative becomes real-valued and compatible for IM/DD transmission. Unfortunately, the requirement of Hermitian symmetry halves the SE.

A. DCO-OFDM

In DCO-OFDM having N subcarriers, all the subcarriers carry non-zero M -ary symbols, except for the first and $N/2$ -th subcarriers, which are zeros. In particular, the FD signal is given by

$$\mathbf{X} = [0, X_1, \dots, X_{N/2-1}, 0, X_{N/2-1}^*, \dots, X_1^*], \quad (1)$$

where $X_i \in \mathcal{S}$ with \mathcal{S} being the M -ary modulation constellation for $i = 1, \dots, N/2 - 1$, and the Hermitian symmetry is imposed on the FD symbols of DCO-OFDM, i.e., $X_i = X_{N-i}^*$, in order to make its TD signal denoted by $\mathbf{u} = \mathcal{F}^{-1}(\mathbf{X})$ after IFFT real-valued. Then, a DC bias B_{DC} is added to the post-IFFT signal \mathbf{u} and the resultant signal will be clipped at zero to produce the positive signal in the TD. The DC bias can be determined as follows [8]: $B_{\text{DC}} = \kappa \sigma_u$, where $\sigma_u^2 = \mathbb{E}[|u|^2]$ with u being the element of \mathbf{u} and κ is a non-negative proportionality constant. The bias-index can be defined as $\xi = 10 \log_{10}(1 + \kappa^2)$ (dB), which represents the power ratio between the DC-biased and non-DC-biased signal.

Again, the main disadvantage of DCO-OFDM is its low power efficiency due to the presence of DC bias. If the DC bias is too low, the severe clipping distortion will significantly degrade the performance. Moreover, this scheme suffers from a SE loss compared to classical RF-OFDM, since less than half of the total subcarriers are used for conveying useful information.

B. ACO-OFDM

In ACO-OFDM, the data bits are carried on the odd-indexed subcarriers, while all the even-indexed subcarriers are set to zero. The Hermitian symmetry is applied to the M -ary symbols loaded onto the odd-indexed subcarriers as follows:

$$\mathbf{X} = [0, X_1, 0, X_3, \dots, X_{N/2-1}, 0, X_{N/2-1}^*, \dots, X_1^*], \quad (2)$$

which results in an anti-symmetric TD signal \mathbf{u} associated with $u_i = -u_{i+N/2}$, $0 \leq i < N/2$. Then, \mathbf{u} is clipped at zero in order to obtain the non-negative signal \mathbf{x} without any loss of information, since the clipping noise falls only on the even-indexed subcarriers.

It has been shown from [8] that ACO-OFDM is more power-efficient than DCO-OFDM, since no DC bias is required. However, ACO-OFDM suffers from a significant SE loss, since only a quarter of the subcarriers are used for carrying useful information. In order to improve the SE of ACO-OFDM, LACO-OFDM was proposed in [13] that will be discussed next.

C. LACO-OFDM

In LACO-OFDM [14], there are several parallel ACO layers used for forming the transmitted signal \mathbf{x} , in which the 1-st layer is the same as in ACO-OFDM that produces the non-negative TD signal \mathbf{x}_1 . The 2-nd layer applies ACO-OFDM signal processing to the $N/2$ null subcarriers, i.e., to the even-indexed ones, obtaining the non-negative TD signal \mathbf{x}_2 . Similarly, in the j -th layer, we employ ACO-OFDM for the $N/2^{j-1}$ blank subcarriers left from the former layers, resulting in the non-negative TD signal \mathbf{x}_j . As such, the transmitted LACO-OFDM signal is obtained by superimposing the output signals of all layers, formulated as $\mathbf{x} = \sum_{j=1}^J \mathbf{x}_j$, where J is the total number of layers. Note that in the j -th layer, there are $N/2^{1+j}$ M -ary symbols carrying useful data, leading to a total of $\frac{N}{2} \left(\sum_{j=1}^J 2^{-j} \right)$ useful M -ary symbols, which tends to $N/2$ upon increasing J .

At the receiver, the data bits of each layer are sequentially decoded as follows. The FD signal of the 1-st layer is decoded first, which is then converted back into the TD signal using the IFFT. We remove the contribution of the 1-st layer in the received TD signal, and then apply FFT to it in order to obtain the new received FD signal for decoding the signal of the 2-nd layer. This decoding process is repeated until the signal of the last layer is decoded.

It is worth noting that LACO-OFDM is capable of approaching the SE of DCO-OFDM, i.e., half of the SE of RF-OFDM, upon increasing the number of layers J . However, this obviously also increases the complexity and latency at both the transmitter and receiver of LACO-OFDM, especially owing to its layer-by-layer iterative detection process, which involves J IFFT/FFT operations.

D. Motivation of Our Work

The SE of the above-mentioned schemes are summarized in Table II. At a glance, advanced schemes such as LACO-OFDM succeed in improving the SE of ACO-OFDM, but at the cost

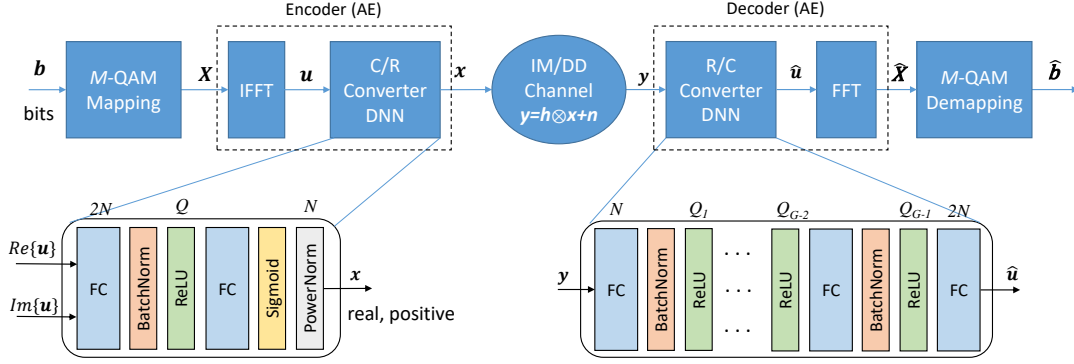


Fig. 1. The block diagram of the proposed O-OFDMNet system.

TABLE II
SPECTRAL EFFICIENCY ANALYSIS OF OPTICAL OFDM SYSTEMS

Scheme	Null symbols	Repetition	DC bias	SE
DCO-OFDM [7]	-	✓	✓	$< 1/2$
ACO-OFDM [9]	✓	✓	-	$1/4$
LACO-OFDM [13]	✓	✓	-	$< 1/2$
Our O-OFDMNet	-	-	-	1
RF-OFDM	-	-	-	1

of increased transceiver complexity. Moreover, these schemes suffer from a high PAPR, which causes a severe nonlinear distortion of the signal.

As a remedy, we propose a novel DL-aided O-OFDM scheme termed as O-OFDMNet, which employs DNNs for converting the TD signal from the complex-valued modulated symbols to non-negative transmit symbols, and vice versa at the receiver, while the FD signal processing remains the same as in RF-OFDM. As a benefit, O-OFDMNet becomes capable of approaching the SE of RF-OFDM, since the Hermitian symmetry is no longer required. We note that the throughput figures seen in Table II justify the title of the paper, namely that our O-OFDMNet approaches the throughput of RF-OFDM. As a further benefit, our proposed scheme can be trained for jointly minimizing the BER and the PAPR, hence potentially outperforming its Hermitian symmetry-based conventional counterparts.

III. THE PROPOSED O-OFDMNET

A. O-OFDMNet System Model

The block diagram of O-OFDMNet is illustrated in Fig. 1. In particular, O-OFDMNet has N subcarriers, which are used for conveying a total of $p = N \log_2 M$ bits per OFDM symbol, where M is the size of the QAM modulation alphabet. At the transmitter, the sequence of p bits denoted by the vector \mathbf{b} enters the M -QAM mapper, resulting in the FD signal vector $\mathbf{X} = [X_0, X_1, \dots, X_{N-1}] \in \mathbb{C}^N$, where $X_i \in \mathcal{S}$ and \mathcal{S} is the M -QAM modulation constellation. Similar to RF-OFDM, the IFFT transform is applied to \mathbf{X} to obtain the TD signal vector $\mathbf{u} = \mathcal{F}^{-1}(\mathbf{X}) \in \mathbb{C}^N$. Note that the post-IFFT signal \mathbf{u} is a complex-valued vector, which is the same as that in RF-OFDM. In order to produce the real- and positive-valued TD signal, O-OFDMNet uses a complex-to-real (C/R) converter represented by a DNN as shown in Fig. 1. The positive-valued

output vector of the C/R converter $\mathbf{x} \in \mathbb{R}^N$ is sent through the optical IM/DD channel.

The received electrical signal $\mathbf{y} \in \mathbb{R}^N$ obtained after an optical-to-electrical signal conversion is given by

$$\mathbf{y} = \mathbf{h} \otimes \mathbf{x} + \mathbf{n}, \quad (3)$$

where $\mathbf{h} = [h_0, h_1, \dots, h_{L-1}] \in \mathbb{R}^L$ is the multi-path channel response with L being the number of paths, $\mathbf{n} \in \mathbb{R}^N$ is the AWGN vector with zero mean and variance $N_0 = \sigma^2/2$. Note that in this work we also consider the AWGN channel, which often occurs when the line of sight (LOS) path becomes dominant over the non-line of sight (NLOS) paths. More specifically, for AWGN channels, we set up $L = 1$ and $h_0 = 1$, hence the model of (3) reduces to $\mathbf{y} = \mathbf{x} + \mathbf{n}$.

At the receiver, the real-to-complex (R/C) converter designed by another DNN is used for converting the real-valued received signal \mathbf{y} into the complex signal $\hat{\mathbf{u}}$, which is regarded as an estimate of \mathbf{u} . Then, the FFT transform and the M -QAM demapper are applied to recover the FD signal $\hat{\mathbf{X}}$ and the bit sequence $\hat{\mathbf{b}}$, respectively. In what follows, we will focus on the novel components of O-OFDMNet, namely on the C/R and R/C converters.

1) *C/R Converter*: As shown in Fig. 1, the C/R converter is fed with a $2N$ -dimensional real-valued vector $\mathbf{u}_0 \in \mathbb{R}^{2N}$ formed by the real and imaginary parts of the post-IFFT signal \mathbf{u} . Then, two fully-connected (FC) layers are employed to process \mathbf{u}_0 , in which the first FC layer is followed by the batch normalization (BatchNorm) and ReLU activation [35], while the second FC layer is followed by the sigmoid activation and power normalization (PowerNorm).

In particular, let us denote the weights and the bias of the first FC layer by $\mathbf{W}_1^a \in \mathbb{R}^{Q \times 2N}$ and $\mathbf{b}_1^a \in \mathbb{R}^Q$, respectively, whose output is given by $\mathbf{u}_1^{FC} = \mathbf{W}_1^a \mathbf{u}_0 + \mathbf{b}_1^a \in \mathbb{R}^Q$, where Q is the number of hidden nodes of the C/R converter. Then, the BatchNorm layer is applied as follows:

$$\mathbf{u}_1^{BN} = f_{BN}(\mathbf{u}_1^{FC}) = \gamma^a \frac{\mathbf{u}_1^{FC} - \mathbb{E}[\mathbf{u}_1^{FC}]}{\sqrt{\text{Var}[\mathbf{u}_1^{FC}] + \nu}} + \beta^a, \quad (4)$$

where γ^a and β^a are the shift and scaling factors of the C/R converter, respectively, and ν is a constant close to

zero, used for preventing the division by zero [36].¹ Note that γ^a and β^a are also 'learnable' parameters, which are updated through training. Then, the ReLU activation function [22] is applied to \mathbf{u}_1^{BN} in order to get the final output of the first FC layer as $\mathbf{u}_1 = f_{\text{ReLU}}(\mathbf{u}_1^{BN}) = \max[0, \mathbf{u}_1^{BN}]$. Similarly, let us denote the weights and the bias of the second FC layer by $\mathbf{W}_2^a \in \mathbb{R}^{N \times Q}$ and $\mathbf{b}_2^a \in \mathbb{R}^N$, respectively. Then, the final output of this layer after sigmoid activation is given by $\mathbf{u}_2 = f_{\text{sigmoid}}(\mathbf{W}_2^a \mathbf{u}_1 + \mathbf{b}_2^a)$, where we note that $f_{\text{sigmoid}}(t) = 1/(1 + e^{-t}) \in (0, 1)$ aims for making the output of the C/R converter real- and positive-valued. Finally, the PowerNorm is applied to \mathbf{u}_2 for ensuring that the average transmit power per symbol remains constant, namely E_s , as follows [25]:

$$\mathbf{x} = f_{\text{PN}}(\mathbf{u}_2) = \frac{\sqrt{E_s} \mathbf{u}_2}{\sqrt{\mathbb{E}[\|\mathbf{u}_2\|^2]}}, \quad (5)$$

where the values of the elements of \mathbf{x} are still positive and no longer limited to the range $(0, 1)$. As such, the average signal-to-noise ratio (SNR) per bit is given by $\mu = 2E_b/\sigma^2$, where $E_b = E_s/m$ is the average power per bit and $m = \log_2 M$.

In summary, the mapping from the complex signal \mathbf{u} to the real and positive signal \mathbf{x} through the C/R converter is expressed by

$$\mathbf{x} = f_{\text{PN}}(f_{\text{sigmoid}}\{\mathbf{W}_2^a f_{\text{ReLU}}[f_{\text{BN}}(\mathbf{W}_1^a \mathbf{u}_0 + \mathbf{b}_1^a)] + \mathbf{b}_2^a\}), \quad (6)$$

where we recall that \mathbf{u}_0 is formed by the real and imaginary parts of \mathbf{u} . We denote the C/R converter by the function $\mathbf{x} = \mathcal{G}_{\text{C/R}}(\mathbf{u})$.

2) *R/C Converter*: As illustrated in Fig. 1, the R/C converter has four FC layers, in which the first three FC layers are followed by the BatchNorm layer and ReLU activation similar to the C/R converter, while the final FC layer has no nonlinear activation, i.e. no ReLU or sigmoid. Note that the use of linear activation for the output layer is indeed plausible, since there is no restriction on the range of elements of \mathbf{u} , which will be recovered by the R/C converter.

In particular, let us denote the weights and bias of the i -th FC layer by $\mathbf{W}_i^b \in \mathbb{R}^{Q_i \times Q_{i-1}}$ and $\mathbf{b}_i^b \in \mathbb{R}^{Q_i}$, where Q_i and Q_{i-1} represent its input and output sizes, respectively, for $i = 1, \dots, G$. Here, G is the number of FC layers of the R/C converter, and thus, there are $G-1$ BatchNorm and ReLU layers in this converter, as shown in Fig. 1. It can be seen that the input size and the output size of the R/C converter are $Q_0 = N$ and $Q_4 = 2N$, respectively. The output of the C/R converter can be expressed by

$$\hat{\mathbf{u}}_0 = \mathbf{W}_G^b f_{\text{ReLU}}\{\dots f_{\text{ReLU}}[f_{\text{BN}}(\mathbf{W}_1^b \mathbf{y} + \mathbf{b}_1^b)] \dots\} + \mathbf{b}_G^b, \quad (7)$$

where $\hat{\mathbf{u}}_0 \in \mathbb{R}^{2N}$ is then reshaped to produce the complex vector $\hat{\mathbf{u}}$ as the estimate of \mathbf{u} . We denote the R/C converter of Fig. 1 by the function of $\hat{\mathbf{u}} = \mathcal{G}_{\text{R/C}}(\mathbf{y})$.

Note that experiments will be carried out to find the minimum number of FC layers required at the R/C converter

¹As pointed out in [36], the BatchNorm can accelerate DNN training by reducing internal covariate shift. Particularly, it allows us to use much higher learning rates and be less careful about initialization.

TABLE III
THE DNN STRUCTURE PARAMETERS OF C/R AND R/C CONVERTERS

Parameter	C/R converter	R/C converter
Input size	$2N$	N
Output size	N	$2N$
FC layers	2	4
Hidden nodes	Q	$Q_1 - Q_2 \dots Q_{G-1}$
Output activation	sigmoid	linear

for reliably detecting the signals, as demonstrated in Subsection IV-A. The DNN structure parameters of the two converters are summarized in Table III.

Let us now summarize the key benefits of the proposed scheme in the two following remarks.

Remark 1. Since the C/R and R/C converters are designed by DNNs for efficiently converting the TD signal between the real and complex domains, O-OFDMNet does not have to rely on the Hermitian symmetry to generate the M-QAM symbols in the FD. As a result, the proposed scheme achieves the full spectral efficiency of RF-OFDM, which has never been achieved by the existing IM/DD O-OFDM schemes, as already shown in Table II of Section II. In particular, in O-OFDMNet, the FD signal processing is the same as in RF-OFDM. Therefore, in contrast to the existing O-OFDM schemes such as DCO-OFDM and ACO-OFDM, our scheme does not suffer from any SE reduction due to transmitting null or repeated M-ary data symbols. This benefit will be further confirmed by the achievable throughput analysis of Subsection IV-B.

Remark 2. In addition to its spectral efficiency and throughput benefits, our proposed O-OFDMNet also provides the following advantages. Firstly, compared to the DCO-OFDM-based schemes, O-OFDMNet is more power-efficient since no DC bias is needed. Secondly, DNNs allow our O-OFDMNet to jointly optimize multiple objectives, such as the BER and PAPR, as it will be mentioned in the next subsection. Finally, the subcarrier orthogonality is retained by O-OFDMNet, which allows us to use low-complexity single-tap equalization for signal detection.

B. Training Procedure

In order to efficiently train O-OFDMNet, we regard the IFFT transform and C/R converter as an amalgamated encoder, while the R/C converter and FFT transform as a decoder of an autoencoder (AE) architecture, as illustrated in Fig. 1. As such, the encoder of the AE accepts the FD signal vector \mathbf{X} as its input data, while the decoder outputs its reconstructed signal denoted by $\hat{\mathbf{X}}$, resulting in the input-output relationship of the AE represented by

$$\hat{\mathbf{X}} = \mathcal{F}(\mathcal{G}_{\text{R/C}}\{\mathbf{h} \otimes \mathcal{G}_{\text{C/R}}[\mathcal{F}^{-1}(\mathbf{X})] + \mathbf{n}\}). \quad (8)$$

Hence, the training dataset is a set of \mathbf{X} , whose entries are the M-QAM symbols randomly generated from the incoming bits \mathbf{b} . Based on the AE model, the proposed scheme can be trained in an end-to-end manner, minimizing both the BER and PAPR of our O-OFDMNet. In particular, the loss functions for

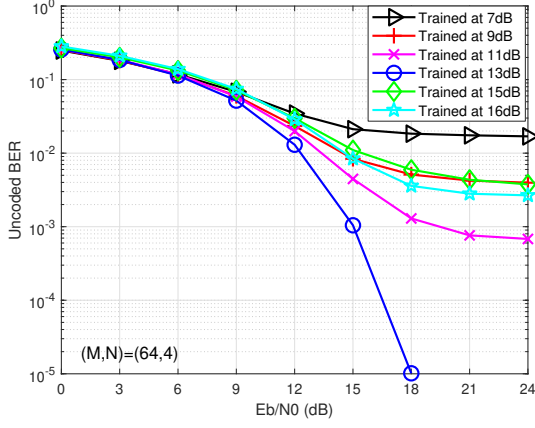


Fig. 2. BER performance of O-OFDMNet with $(N, M) = (64, 4)$, trained at different training SNRs μ_{tr} , under AWGN channels. Here, the training parameters for this figure are detailed in Table IV in Subsection IV-A.

the BER (\mathcal{L}_1) and PAPR (\mathcal{L}_2) objectives are respectively given by

$$\mathcal{L}_1(\theta) = \frac{1}{T} \sum_{j=1}^T \left\| \mathbf{X}_j - \hat{\mathbf{X}}_j \right\|^2, \quad (9)$$

and

$$\mathcal{L}_2(\theta) = \frac{1}{T} \sum_{j=1}^T \frac{\max_{0 \leq i \leq N-1} x_{i,j}^2}{\mathbb{E}[x_{i,j}^2]}, \quad (10)$$

where θ denotes the 'learnable' parameters of both the C/R and of the R/C converters (including weights, biases, as well as shift and scaling factors of BatchNorm layers), T is the batch size, i.e., the number of data samples feeding the AE model at each iteration, $x_{i,j}$ is the i -th element of \mathbf{x}_j , which is the output of the C/R converter corresponding to the input \mathbf{X}_j , and $\hat{\mathbf{X}}_j$ is the reconstruction of \mathbf{X}_j . Note that $\mathcal{L}_1(\theta)$ is often termed as the reconstruction loss, which is minimized for ensuring that the DNN-based receiver reliably decodes the transmitted signal.

Based on (9) and (10), the joint loss is expressed by

$$\mathcal{L}(\theta) = \mathcal{L}_1(\theta) + \lambda \mathcal{L}_2(\theta), \quad (11)$$

where λ is the loss scaling factor, which can be adjusted to strike a BER vs PAPR trade-off. For example, if λ is large, the training will put more effort into reducing the PAPR rather than the BER. Using (11), the model parameter θ is updated for each batch of T samples based on the stochastic gradient decent (SGD) algorithm as follows:

$$\theta := \theta - \eta \nabla \mathcal{L}(\theta), \quad (12)$$

where η is the learning rate and $\nabla \mathcal{L}(\theta)$ is the gradient of $\mathcal{L}(\theta)$ with respect to θ . In this work, we adopt an advanced SGD method termed as the adaptive moment estimation (Adam) [37] for updating θ . Moreover, Xavier's initialization method of [38] is used for initializing the weights and biases. Both methods are available in various DL libraries such as TensorFlow [39].

Remark 3. In order to reduce the training overhead, we do not have to train O-OFDMNet with multiple SNRs. Instead, it is

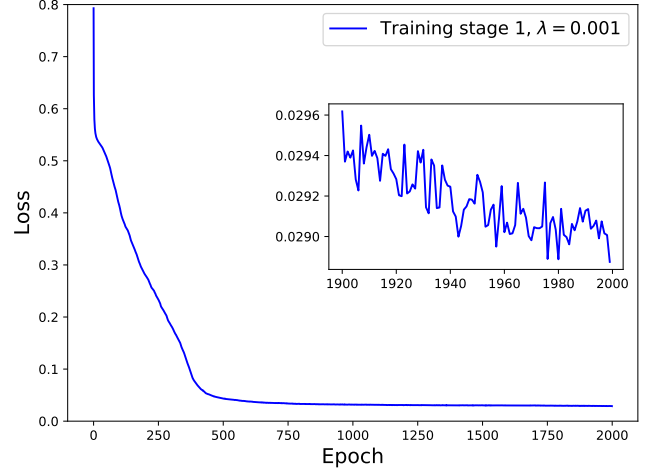


Fig. 3. Loss convergence behavior of Stage 1 of our O-OFDMNet when $(N, M) = (64, 4)$, under AWGN channels. Here, the training parameters for this figure are detailed in Table IV in Subsection IV-A.

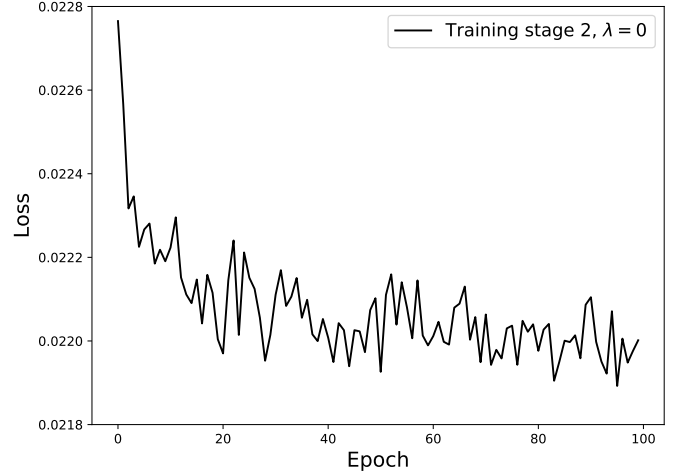


Fig. 4. Loss convergence behavior of Stage 2 of our O-OFDMNet when $(N, M) = (64, 4)$, under AWGN channels. Here, the training parameters for this figure are detailed in Table IV in Subsection IV-A.

trained only at a single training SNR denoted by μ_{tr} , which will be carefully selected via experiments so as to provide the best performance in the testing SNR range of interest. This is shown in Fig. 2, where training our scheme at $\mu_{tr} = 13$ dB gives the best performance compared to other values of μ_{tr} . As we can see from this figure, training at exactly the same testing SNR does not always provide the best performance at that specific SNR level. For example, observe at 15 dB in Fig. 2 that our O-OFDMNet trained at 13 dB achieves much better BER than that trained at 15dB. The selection of a single training SNR is also common in the previous contributions [24], [26].

Remark 4. Aiming for improving the BER without significantly increasing the PAPR, we propose to train O-OFDMNet in two stages as follows. In the first stage, the model is trained with $\lambda > 0$, e.g., $\lambda = 0.001$. Then, the model obtained is retrained with $\lambda = 0$ using fewer epochs, smaller learning rate and larger batch size. Thus, the second stage requires much less

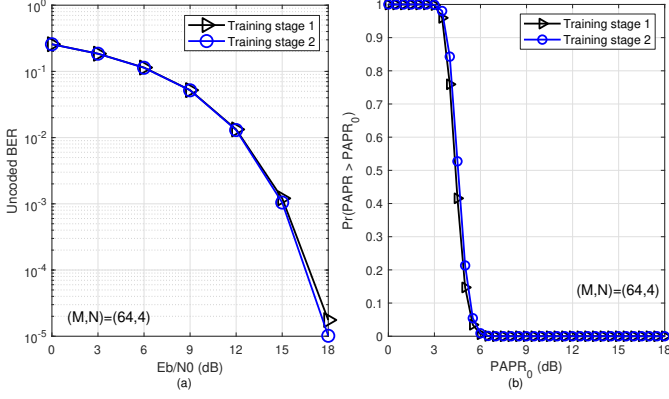


Fig. 5. BER and PAPR performance of the two training stages of our O-OFDMNet when $(N, M) = (64, 4)$, under AWGN channels. Here, the training parameters for this figure are detailed in Table IV in Subsection IV-A.

training time than the first one. All these parameters will be detailed in Subsection IV-A. In order to get an insight into our proposed training strategy, we illustrate the loss convergence behavior of training stages 1 and 2 in Fig. 3 and Fig. 4, which show that the loss of Stage 1 converges after about 1000 epochs, while that of Stage 2 converges quicker, as expected, after just 20 epochs. Furthermore, Fig. 4 shows that by setting $\lambda = 0$ in Stage 2, the joint loss (11) becomes the reconstruction loss (9), which can be further reduced, leading to an improvement in the BER performance at a negligible increase of the PAPR compared to Stage 1 as shown in Fig. 5.

C. Soft Decoder for O-OFDMNet

While it is straightforward to apply a hard-decision decoder to O-OFDMNet by directly estimating data bits from the estimated M -ary symbols $\hat{\mathbf{X}}$, the application of a soft-decision decoder is less intuitive, since the relationship between $\hat{\mathbf{X}}$ and \mathbf{X} is rather complex due to the presence of two neural networks between them. To address this, we simply assume

$$\hat{\mathbf{X}} = \mathbf{X} + \mathbf{W}, \quad (13)$$

where $\mathbf{W} \in \mathbb{C}^N$ is a complex Gaussian noise process with zero mean and variance of $\sigma_w^2 = 1/\mu$. Based on this assumption, we can compute the log-likelihood ratio (LLR) for the i -th bit b_i of an M -QAM symbol as follows [40]:

$$L_i = \log \frac{P(b_i = 0 | \hat{X})}{P(b_i = 1 | \hat{X})} = \log \frac{\sum_{X \in \mathcal{S}_i^{(0)}} p(\hat{X} | X)}{\sum_{X \in \mathcal{S}_i^{(1)}} p(\hat{X} | X)}, \quad (14)$$

where X is the element of \mathbf{X} , \hat{X} is its corresponding estimate in $\hat{\mathbf{X}}$, $\mathcal{S}_i^{(0)}$ and $\mathcal{S}_i^{(1)}$ are the sets of the M -QAM symbols containing $b_i = 0$ and 1, respectively, $\mathcal{S}_i^{(0)} \cup \mathcal{S}_i^{(1)} = \mathcal{S}$, and $p(\hat{X} | X) = \frac{1}{\pi \sigma_w^2} \exp\left(-|\hat{X} - X|^2 / \sigma_w^2\right)$ is the probability density function (PDF) of \mathbf{W} . In order to prevent numerical

overflow, we can use the max-sum approximation of $\sum_j u_j \approx \max_j u_j$, $\forall u_j \geq 0$, to calculate the LLR from (14) as follows:

$$\begin{aligned} L_i &\approx \log \frac{\max_{X \in \mathcal{S}_i^{(0)}} p(\hat{X} | X)}{\max_{X \in \mathcal{S}_i^{(1)}} p(\hat{X} | X)} \\ &= -\frac{1}{\sigma_w^2} \left(\min_{X \in \mathcal{S}_i^{(0)}} |\hat{X} - X|^2 - \min_{X \in \mathcal{S}_i^{(1)}} |\hat{X} - X|^2 \right). \end{aligned} \quad (15)$$

Finally, the approximated LLRs are fed to the soft decoder to decode the encoded bits. In Section IV, both convolutional and polar codes will be applied to the proposed O-OFDMNet.

IV. SIMULATION RESULTS

In this section, various simulation results are provided for characterizing the capacity as well as the BER and PAPR performance of O-OFDMNet in comparison to the baselines under both AWGN and FS channels. Moreover, the forward error correction (FEC) coded BER of convolutional and polar codes [41] using both hard and soft decoders is investigated. We also investigate the performance of O-OFDMNet under several realistic optical channel conditions. The complexity of O-OFDMNet is discussed at the end of this section. In the following, we first introduce the parameters of our O-OFDMNet and of the baselines selected for comparison.

A. Parameter Settings

1) *Model and training parameters:* The model and training parameters of O-OFDMNet are summarized in Table IV, where we consider two SE scenarios, namely 1 and 2 bps/Hz. As observed via this table, at 2 bps/Hz O-OFDMNet needs more hidden nodes and more epochs for the first training stage than at 1 bps/Hz. Through experiments, we find that selecting $\lambda = 0.001$ for stage 1 strikes an attractive BER vs PAPR trade-off.² We only need 20 epochs for stage 2 in order to further improve BER performance without increasing the PAPR, as already shown in Fig. 5 in Subsection III-B. Moreover, the batch size of stage 2 is 2000, i.e. twice as large as that of stage 1. Thus, the training time of stage 2 is kept very short compared to stage 1. To keep the training stable and avoid struggling far from the model obtained in stage 1, the learning rate of stage 2 is selected to be 0.0001, i.e. lower than that of stage 1. Based on our experience, a good training SNR μ_{tr} should be the value at which the BER ranges from 10^{-3} to 10^{-2} . Specifically, we select μ_{tr} to be 8 dB and 13 dB for the two scenarios considered, as shown in Table IV.

Next, we note that the hyper-parameters of the DNNs given in Table IV, such as the learning rate, training SNR, and number of nodes or layers, are selected based on the trial-and-error method. For example, we now justify our selection of the number of FC layers in the R/C converter given in Table IV.

²We opt for such a small value for the loss scaling factor λ , since intuitively the PAPR loss \mathcal{L}_2 would be much larger than the reconstruction BER loss \mathcal{L}_1 . More particularly, \mathcal{L}_2 is always greater than 1, while \mathcal{L}_1 tends to 0 while training. For example, consider O-OFDMNet with $(M, N) = (64, 4)$ under AWGN channels, after 2000 training epochs of stage 1, we obtain $\mathcal{L}_1 = 0.024$, while $\mathcal{L}_2 = 3.5$. This result indicates that \mathcal{L}_1 and $\lambda \mathcal{L}_2$ are comparable to each other in the training.

TABLE IV
MODEL AND TRAINING PARAMETERS OF O-OFDMNET

Parameter	Definition	Value (1 bps/Hz)	Value (2 bps/Hz)
N	Number of subcarriers	64	
M	Modulation order	2	4
Q	Hidden nodes of C/R converter	256	512
$Q_1 - Q_2 \dots Q_{G-1}$	Hidden nodes of R/C converter	256	512-512
N/A	Training data size	10^5	
N/A	Testing data size	2×10^5	
N/A	Number of epochs of stage 1	1000	2000
N/A	Number of epochs of stage 2	20	
η	Learning rate of stage 1	0.001	
	Learning rate of stage 2	0.0001	
λ	Loss scaling factor of stage 1	0.001	
	Loss scaling factor of stage 2	0	
T	Batch size of stage 1	1000	
	Batch size of stage 2	2000	
μ_{tr}	Training SNR	8 dB	13 dB

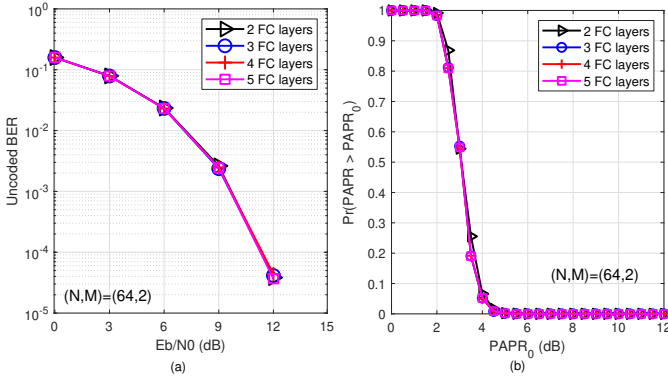


Fig. 6. BER and PAPR performance of our O-OFDMNet having $(N, M) = (64, 2)$, under AWGN channels, when the R/C converter uses different numbers of fully-connected (FC) layers. Here, the SE is 1 bps/Hz.

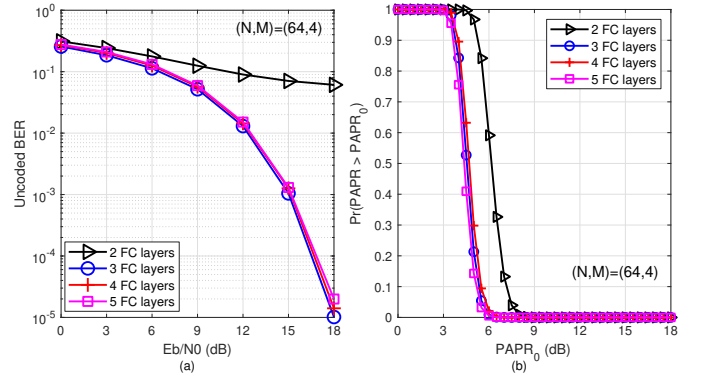


Fig. 7. BER and PAPR performance of our O-OFDMNet having $(N, M) = (64, 4)$, under AWGN channels, when the R/C converter uses different numbers of fully-connected (FC) layers. Here, the SE is 2 bps/Hz.

More explicitly, Figs. 6 and 7 portray the BER and PAPR performance of our O-OFDMNet using different numbers of FC layers for the R/C converter, when $(N, M) = (64, 2)$ and $(64, 4)$, respectively. Here, the number of hidden nodes for $(N, M) = (64, 2)$ and $(64, 4)$ is given by 256 and 512, respectively. Observe in Fig. 6 that at the SE of 1 bps/Hz, O-OFDMNet only needs 2 FC layers at the receiver for achieving the desired performance, which cannot be substantially improved upon using more FC layers. Furthermore, at the SE of 2 bps/Hz characterized in Fig. 7, our receiver requires 3 FC layers for achieving the desired performance, while using more FC layers does not improve the performance, despite increasing the DNN complexity. Similarly, other hyper-parameters are also selected based on this trial-and-error approach, in which another example for selecting the training SNR has already been demonstrated in Fig. 2 of Subsection III-B.

2) *Channel types*: We consider both AWGN and FS channels. In particular, under an AWGN channel, the received TD signal is simply written by $\mathbf{y} = \mathbf{x} + \mathbf{n}$. For the FS channel, we assume that there are $L = 5$ paths, whose gains follow a negative exponential decay with $h_i = \frac{0.4^i}{\alpha}$ for $i = 0, \dots, L-1$, where $\alpha = \sqrt{\sum_{i=0}^{L-1} (0.4^i)^2}$ is used to make the total power of L paths be unity, i.e., $\sum_{i=0}^{L-1} h_i^2 = 1$. Such a FS channel

is illustrated in Fig. 8. Note that the FS channel envelope is assumed to be the same during both the training and testing phases, since in visible light communications, the channel tends to be static for a certain physical configuration of the receiver, reflectors, and transmitter, as documented in [42], [43]. However, if the channel changes dramatically, we have to retrain the DNNs to adapt to the new channel. It is worth designing DNNs that can be near-instantaneously adaptive to time-varying channels without retraining. One can apply an impressive technique of in [44], where a dedicated DNN is designed for estimating the channel-related information that is then fed to another DNN for detecting the desired signal. We will consider this investigation in our future work.³

3) *Baselines*: We consider three well-known O-OFDM schemes, namely, ACO-OFDM [9], LACO-OFDM [13] and DCO-OFDM [7] to be the baselines of O-OFDMNet. Moreover, in order to demonstrate the importance of the IFFT/FFT transforms of OFDM in combating the ISI caused by the frequency-selective channel, we also include on-off keying (OOK) and M -ary pulse-amplitude modulation (M -PAM) [6],

³We assume that the cyclic prefix inserted into the time-domain signal \mathbf{x} in (3) is sufficiently long for eliminating the inter-symbol interference inflicted by interference among the LED wavelengths of different colors, which is widely adopted in OFDM-based systems [1].

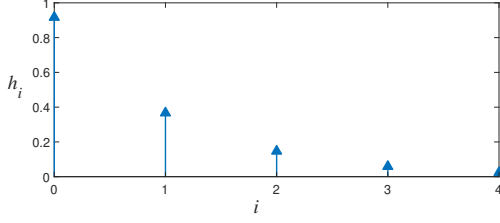


Fig. 8. Frequency-selective channel.

TABLE V
PARAMETERS OF THE BASELINES FOR TWO SPECTRAL EFFICIENCIES

Baseline	SE = 1 bps/Hz	SE = 2 bps/Hz
ACO-OFDM [9]	$M = 16$	$M = 256$
LACO-OFDM [13]	5 layers, $M = 4$	5 layers, $M = 16$
DCO-OFDM [7]	7 dB DC, $M = 4$	13 dB DC, $M = 16$
OOK or M -PAM [6]	$M = 2$	$M = 4$

which do not rely on those transforms, for comparison at the SEs of 1 and 2 bps/Hz, respectively. The parameters of those baselines are summarized in Table V for two SE scenarios similar to Table IV. Note that since OOK sends only 0 and 1 symbols, it cannot support a SE higher than 1 bps/Hz. Additionally, because the O-OFDM baselines rely on the Hermitian symmetry in the arrangement of M -QAM symbols in the FD, they require higher QAM modulation orders M in order to achieve the same SE as the proposed O-OFDMNet scheme. Also note that LACO-OFDM needs 5 ACO layers to achieve the SEs of 0.96875 and 1.9375 bps/Hz with $M = 4$ and 16, respectively, while DCO-OFDM employs 7 dB and 13 dB DC biases, as recommended in [8].

B. Capacity Analysis

We first study the channel capacity of the proposed O-OFDMNet. Explicitly, in order to compare the capacity of different optical schemes, we study the discrete-input continuous-output memoryless channel (DCMC) capacity [18] related to the bit-to-symbol mapping. In particular, the calculation of the O-OFDMNet capacity follows the mutual information based method of [15]. Explicitly, for each of the symbols \hat{X}_k ($k = 1, \dots, N$) at the output of the FFT-based demodulator the probability of being originally transmitted as the m -th symbol in the M -QAM constellation \mathcal{S} - namely that we have $X_k = \mathcal{S}_m$ - is formulated as $P(\hat{X}_k | X_k = \mathcal{S}_m) = \frac{1}{\pi\sigma_w^2} \exp\left(-|\hat{X}_k - \mathcal{S}_m|^2 / \sigma_w^2\right)$. Then its normalized counterpart is used for representing the corresponding *a posteriori* probability as

$$P_{k,m} \triangleq P(X_k = \mathcal{S}_m | \hat{X}_k) = \frac{P(\hat{X}_k | X_k = \mathcal{S}_m)}{\sum_{m'=1}^M P(\hat{X}_k | X_k = \mathcal{S}_{m'})}. \quad (16)$$

The DCMC capacity can then be expressed as the entropy of the source $H(\mathbf{X})$ minus the average information lost $H(\mathbf{X}|\hat{\mathbf{X}})$,

i.e.

$$C = H(\mathbf{X}) - H(\mathbf{X}|\hat{\mathbf{X}}) \\ = \log_2 M - \mathbb{E}_k \left[- \sum_{m=1}^M P_{k,m} \log_2 P_{k,m} \right], \quad (17)$$

which can be approximated using the classic Monte-Carlo method relying on a large number of signal realizations. This method is also used for evaluating the capacities of the baselines for comparison.

Fig. 9 plots the DCMC capacities of O-OFDMNet both for 1 bps/Hz and 2 bps/Hz bit-to-symbol mapping, i.e. for (a) $M = 2$ and (b) $M = 4$, respectively. These results are compared to both ACO-OFDM, as well as to 5-layer LACO-OFDM and DCO-OFDM using the same bit-to-symbol mapping constellations, while also benchmarked against the RF capacity. It is shown in both cases that our O-OFDMNet approaches twice the SEs of both LACO-OFDM as well as of DCO-OFDM and four time that of ACO-OFDM. This is in line with our SE figures presented in Table II of Section II, although the O-OFDMNet curve requires slightly higher power to reach its maximum than its RF counterpart. This can be further observed in Fig. 10, where the conventional O-OFDM and RF-OFDM schemes having either the same or a similar SE are also included for reference. It is observed that O-OFDMNet reaches the maximum capacity at a significantly lower E_b/N_0 (around 0 dB in both cases) than the O-OFDM baselines, which require around 10 dB in Fig. 10(a) and more than 15 dB in Fig. 10(b). Finally, it is worth noting that the capacity trends observed in Figs. 10(a)-(b) match well with the BER trends seen in Figs. 11(a)-(b) in the next subsection.

C. BER Performance

Fig. 11 depicts the uncoded BER comparison between O-OFDMNet and the baselines under AWGN channels, at the SEs of (a) 1 and (b) 2 bps/Hz. It is shown that our scheme outperforms the baselines right across the entire SNR region. For example, in Fig. 11(a), O-OFDMNet achieves about 3dB SNR gain at a BER of 10^{-3} over the O-OFDM baselines. Interestingly, the proposed scheme has the same BER as OOK or 4-PAM, where we note that under AWGN channels, these two schemes are the optimal designs for IM/DD communications at the SEs of 1 and 2 bps/Hz. This confirms that our scheme is indeed capable of learning the optimal mapping from the complex signal to the real, non-negative TD signal via the C/R converter, in the presence of IFFT/FFT of OFDM. This cannot be achieved by the conventional Hermitian-based baselines such as LACO-OFDM and DCO-OFDM.

In Fig. 12, we compare the uncoded BER of O-OFDMNet to the baselines under the FS channel, at the SEs of (a) 1 and (b) 2 bps/Hz. Again, the BERs of our scheme are better than that of the baselines right across the whole SNR region. Specifically, at a BER of 10^{-4} , SNR gains of 5 dB and 3 dB are achieved by O-OFDMNet over LACO-OFDM in Fig. 12(a) and (b), respectively. Observe furthermore in Fig. 12(a) and (b) that OOK and 4-PAM operating without IFFT/FFT totally fails due

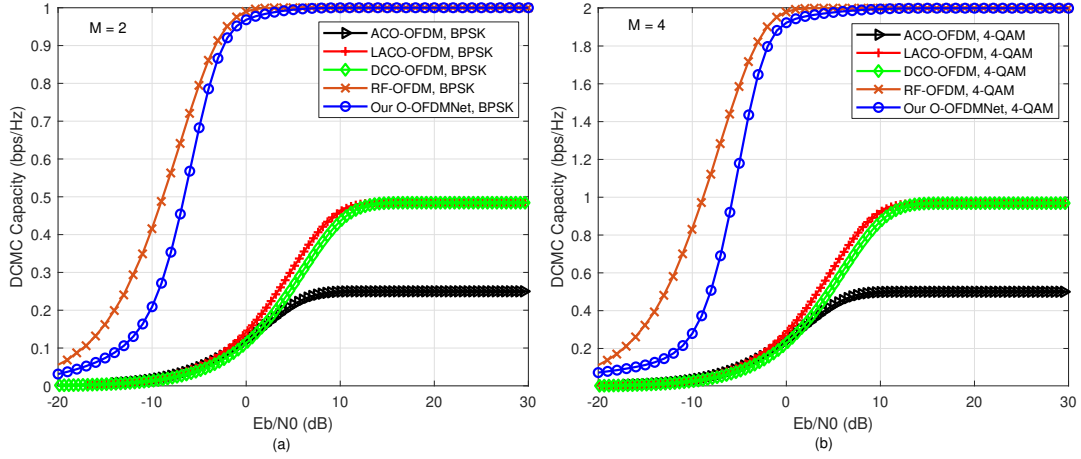


Fig. 9. DCMC capacity comparison between the proposed O-OFDMNet and the baselines under AWGN channels, using the same modulation order of (a) $M = 2$ and (b) $M = 4$. The parameters of Tables IV and V are used.

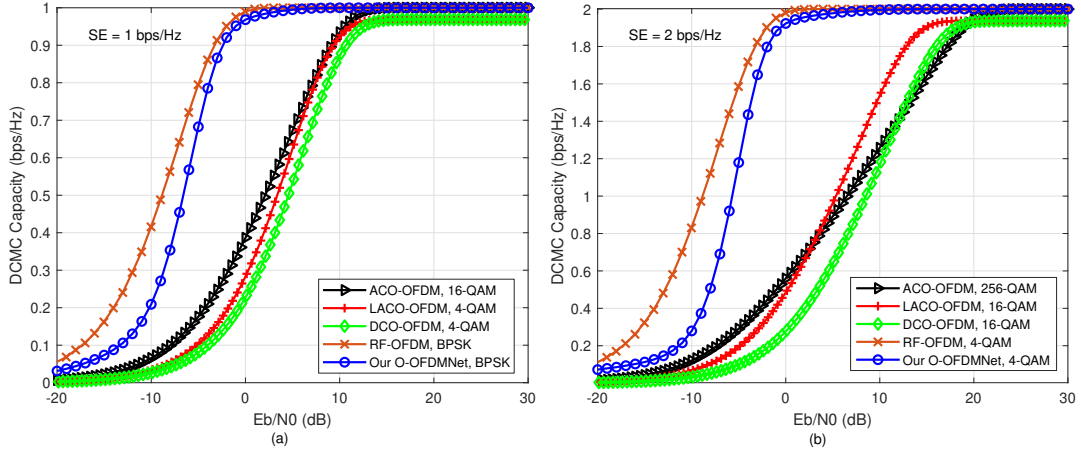


Fig. 10. DCMC capacity comparison between the proposed O-OFDMNet and the baselines under AWGN channels, at the same SE of (a) 1 bps/Hz and (b) 2 bps/Hz, using the parameters of Tables IV and V.

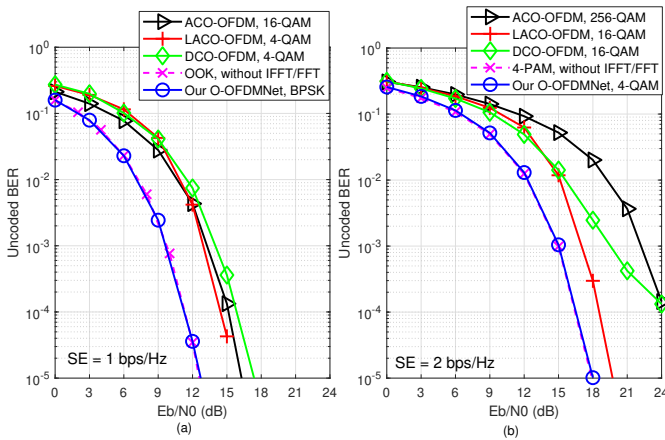


Fig. 11. Uncoded BER comparison between the proposed O-OFDMNet and the baselines at the SEs of 1 and 2 bps/Hz under AWGN channels, using the parameters of Tables IV and V.

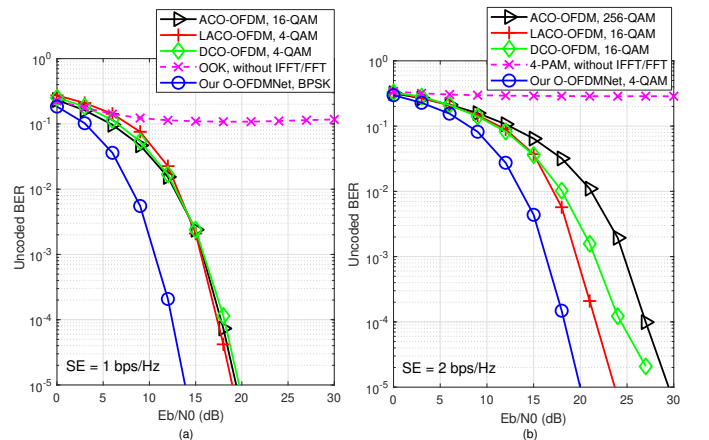


Fig. 12. Uncoded BER comparison between the proposed O-OFDMNet and the baselines at the SEs of 1 and 2 bps/Hz under the FS channel, using the parameters of Tables IV and V.

to the severe ISI caused by the multipath effect. This confirms the importance of using OFDM in combating the ISI in optical communications experiencing multipath effects.

Fig. 13 shows the FEC-coded BER comparison between our O-OFDMNet and the baselines under AWGN channels at the SE of 1 bps/Hz, where the 1/2-rate convolutional code

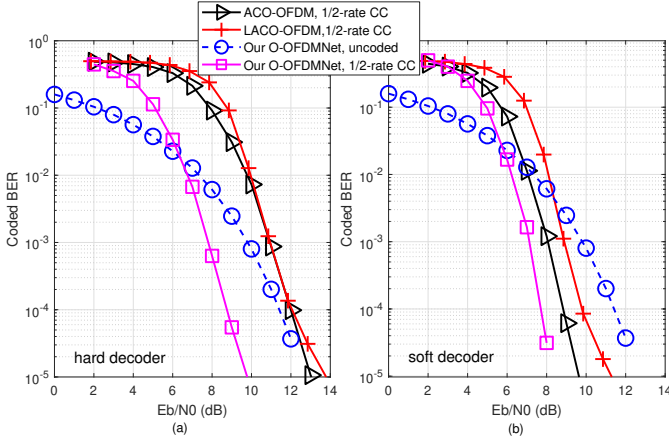


Fig. 13. Coded BER comparison between the proposed O-OFDMNet and the baselines at the SE of 1 bps/Hz under AWGN channels, using the parameters of Tables IV and V, where 1/2-rate CC employing (a) hard decoder and (b) soft decoder is used.

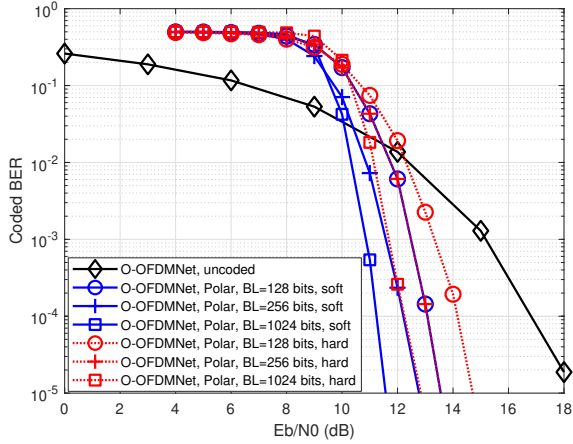


Fig. 14. Coded BER of the proposed O-OFDMNet at the SE of 2 bps/Hz under AWGN channels, using the parameters of Table IV, where the 1/2-rate polar code with different block lengths and hard/soft decoder is used.

(CC) employing (a) hard and (b) soft decoders is used. More specifically, we employ the same CC with the octal code generator polynomials [171, 133] for all schemes having a memory of 7. As observed via Fig. 13, using the CC, the BER of O-OFDMNet is significantly improved at the SNRs above 6 dB. Moreover, the BER of soft decoder is better than that of hard decoder, as expected. Compared to the baselines, the coded BER of our scheme is remarkably better. For instance, at a BER of 10^{-4} , our scheme achieves coding gains of above 3 dB and 2 dB over LACO-OFDM in Figs. 13(a) and (b), respectively.

In order to further improve the coded BER of O-OFDMNet, we can use stronger codes such as polar code relying on a hard/soft decoder as shown in Fig. 14. More particularly, we employ a 1/2-rate polar code having block lengths (BL) of 128, 256 and 1024 bits, corresponding to 1, 2 and 8 OFDM symbols relying on 4-QAM. As expected, the BER of O-OFDMNet using the soft polar decoder of [45] is improved upon increasing the block length and it is always better than that of a hard decoder. More importantly, the polar-coded BER

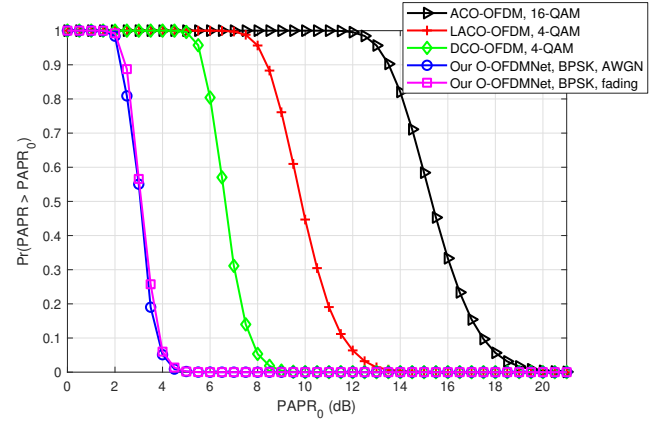


Fig. 15. PAPR comparison between the proposed O-OFDMNet and the baselines at the SE of 1 bps/Hz, using the parameters of Tables IV and V. Our proposed scheme is trained for both AWGN and FS channels.

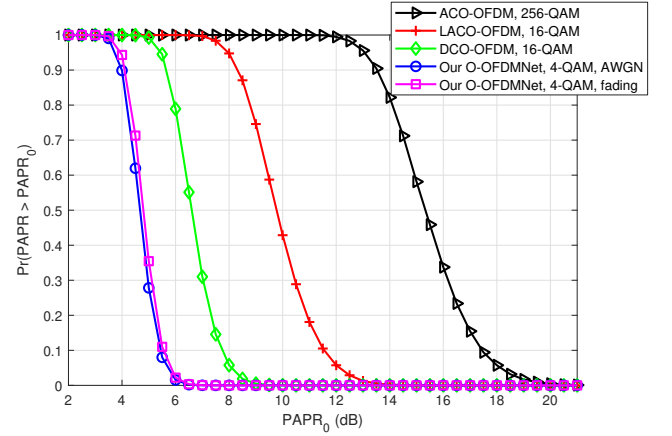


Fig. 16. PAPR comparison between the proposed O-OFDMNet and the baselines at the SE of 2 bps/Hz, using the parameters of Tables IV and V. Our proposed scheme is trained for both AWGN and FS channels.

is much better than the uncoded BER in the high SNR region. For example, at a BER of 10^{-5} , using a polar code and a soft decoder at BL = 1024 bits, our coded scheme achieves 7 dB and 2 dB coding gains over the uncoded scheme and over the same coded scheme having BL = 128 bits, respectively.

D. PAPR Performance

Fig. 15 compares the PAPR of O-OFDMNet to the baselines at the SE of 1 bps/Hz, where our scheme is trained for both AWGN and FS channels. It is shown in Fig. 15 that the proposed O-OFDMNet achieves a lower PAPR than the baselines. For example, at $\Pr(\text{PAPR} > \text{PAPR}_0) = 0.1$, the PAPR of O-OFDMNet is about 4, 8 and 14 dB lower than that of DCO-OFDM, LACO-OFDM and ACO-OFDM, respectively. This is because by employing DNNs for the C/R and R/C converters, our scheme only requires a low M -ary modulation order of $M = 2$, compared to the baselines, which use $M = 4$ or 16. Moreover, the PAPR of our scheme trained for the FS channel is similar to that trained for AWGN channels. Similar observations can be made from Fig. 16 at 2 bps/Hz, where

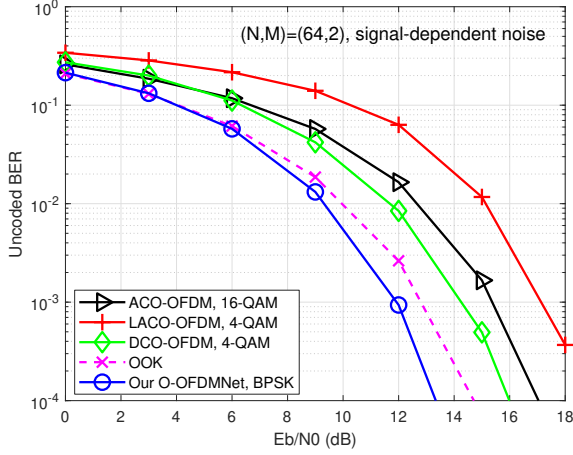


Fig. 17. BER comparison between O-OFDMNet when $(N, M) = (64, 2)$ and the baselines in the presence of signal-dependent noise, using the parameters of Tables IV and V. Here, the noise scaling factor of $\rho = 1$ and AWGN channels are used.

O-OFDMNet yields significantly lower PAPR than all the baselines.

As such, we can conclude that with the aid of deep learning, the proposed O-OFDMNet outperforms the baselines in terms of both its BER and PAPR. In what follows, we will investigate the performance of our O-OFDMNet in several realistic optical channel conditions, followed by the analysis of its computational complexity in comparison to the benchmarks.

E. O-OFDMNet under Realistic Channel Conditions

We now investigate the performance of O-OFDMNet in the presence of both signal-dependent noise, as well as nonlinear distortion caused by clipping and a practical indoor VLC channel [43]. Here, we keep the DNN structure of O-OFDMNet unchanged, as specified in Table IV. We train our scheme with these realistic channels which are different from the AWGN and FS channels presented in Subsection IV-A. Hence, the learnable parameters of DNNs such as the weights, biases, as well as the shift and scaling factor of the BatchNorm layers (γ, β) are also different from those trained with the time-invariant channels specified in Subsection IV-A.

Firstly, the BER performance of our O-OFDMNet using $(N, M) = (64, 2)$ is shown in Fig. 17 in the presence of signal-dependent noise, where the baselines from Table V are also included for comparison. In particular, the variance of the noise \mathbf{n} in the input-output channel model (3) is given by

$$\sigma_i^2 = x_i \rho^2 \sigma_0^2 + \sigma_0^2, \quad i = 1, 2, \dots, N, \quad (18)$$

where σ_0^2 is the variance of the signal-independent Gaussian thermal noise, while $x_i \rho^2 \sigma_0^2$ refers to the variance of the signal-dependent noise, where ρ is the noise scaling factor. Observe in Fig. 17 that our O-OFDMNet noticeably outperforms the baselines. In contrast to the case of having only signal-independent-thermal noise in Fig. 11, our scheme achieves better BER performance than OOK, especially at high SNRs. This is because by training the DNNs of O-OFDMNet, our scheme becomes capable of jointly optimizing both the

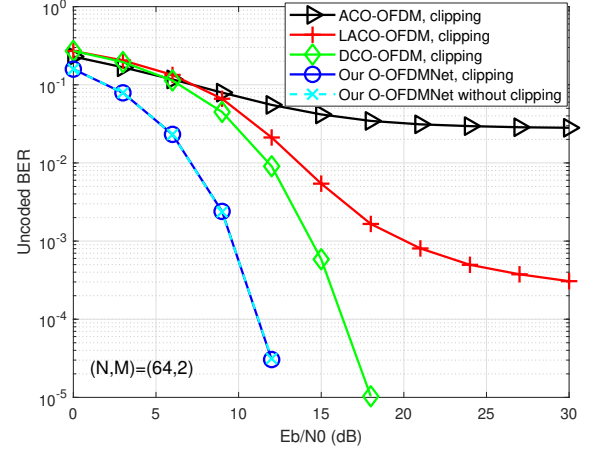


Fig. 18. BER comparison between our O-OFDMNet when $(N, M) = (64, 2)$ and the baselines in the presence of clipping distortion, under AWGN channels, using the parameters of Tables IV and V. Here, the clipping factor ζ is selected to be 5 dB.

transmitter and receiver to achieve the best performance under practical channel conditions, while the transceiver designs for OOK as well as other O-OFDM schemes are not optimized for optical channels experiencing signal-dependent noise.

Secondly, we quantify the impact of nonlinear clipping distortion on our O-OFDMNet in comparison to the existing O-OFDM schemes, as demonstrated in Fig. 18. Particularly, the unipolar signals x_i ($i = 1, 2, \dots, N$) of all schemes are clipped at the same upper limit of $\Psi_{\text{upper}} = \zeta \sqrt{\mathbb{E}[|x_i|^2]}$ defined as follows:

$$x_i^{\text{clipped}} = \begin{cases} \Psi_{\text{upper}}, & x_i > \Psi_{\text{upper}}, \\ x_i, & x_i \leq \Psi_{\text{upper}}, \end{cases} \quad (19)$$

where ζ is the clipping factor given in [dB]. Accordingly, Fig. 18 shows the BER comparison between our scheme and the O-OFDM baselines when clipped at $\zeta = 5$ dB. It is observed from this figure that our O-OFDMNet does not incur any performance loss caused by clipping. By contrast, upon observing both Fig. 11(a) and Fig. 18, we can find that the BER performance of the baselines significantly degrades due to clipping, and becomes much worse than that of our scheme. This observation validates the superiority of our O-OFDMNet in terms of achieving lower PAPR than the existing O-OFDM schemes. In other words, by offering lower PAPR, our scheme is more robust to hostile nonlinear distortion, such as clipping, than the benchmarks. We also note that the BER trend in Fig. 18 matches well with the corresponding PAPR trend seen in Fig. 15, where a scheme having higher PAPR will have worse BER in the presence of clipping distortion.

Finally, we investigate the performance of our O-OFDMNet under a practical FS VLC channel measured in an indoor environment, as described in [43]. In particular, an empty room having the dimensions of $(5.0 \text{ m} \times 5.0 \text{ m} \times 3.0 \text{ m})$ with plaster walls is considered, where an LED is located in the center of the ceiling $(2.5 \text{ m}, 2.5 \text{ m}, 3.0 \text{ m})$, and a photo-detector is placed in the corner of the floor $(0.5 \text{ m}, 1.0 \text{ m}, 0.0 \text{ m})$ pointing upward. Three paths were observed,

TABLE VI
COMPLEXITY COMPARISON BETWEEN OUR O-OFDMNET AND THE BASELINES

Scheme	Transmitter complexity		Receiver complexity	
	IFFT/FFT	DNN	IFFT/FFT	DNN
DCO-OFDM [7]	$\mathcal{O}(N \log_2 N)$	N/A	$\mathcal{O}(N \log_2 N)$	N/A
ACO-OFDM [9]	$\mathcal{O}(N \log_2 N)$	N/A	$\mathcal{O}(N \log_2 N)$	N/A
LACO-OFDM [13]	$\mathcal{O}(KN \log_2 N)$	N/A	$\mathcal{O}(2KN \log_2 N)$	N/A
Our O-OFDMNet	$\mathcal{O}(N \log_2 N)$	$\mathcal{O}(3QN)$	$\mathcal{O}(N \log_2 N)$	$\mathcal{O}(3QN + (G-2)Q^2)$

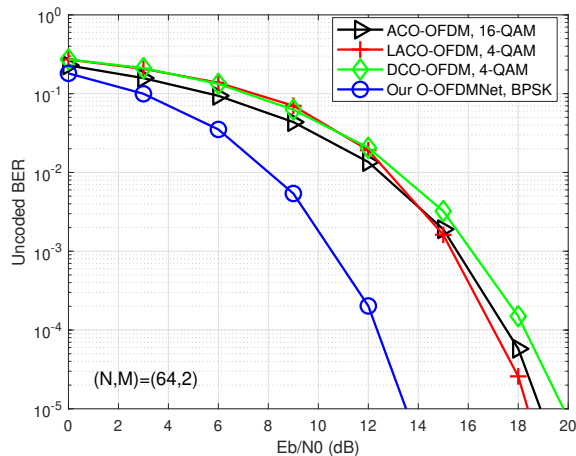


Fig. 19. BER comparison between our O-OFDMNet and the O-OFDM baselines under a practical frequency selective VLC channel [43], at the SE of 1 bps/Hz, when $(N, M) = (64, 2)$. Here, the parameters of Tables IV and V are used.

including the line-of-sight (LOS) path and two reflected paths, whose power delay profile is given in Fig. 3 of [43]. More explicitly, the amplitudes and time delays of these three paths are $(3.8, 1.4, 0.5) \times 10^{-5}$ and $(11, 14, 17)$ (ns), respectively. The BER performance of our O-OFDMNet compared to the baselines under this channel is illustrated in Fig. 19. It is observed that our scheme outperforms the baselines using this FS channel. For instance, at a BER of 10^{-5} , O-OFDMNet provides SNR gains of more than 5 dB over the baselines.

F. Complexity Analysis

The complexity of O-OFDMNet includes two main parts, namely the IFFT/FFT and DNN complexity. Specifically, since our scheme needs one IFFT at the transmitter and one FFT at the receiver, the IFFT/FFT complexity order is given by $\mathcal{O}(N \log_2 N)$, which is similar to that of ACO-OFDM and DCO-OFDM. For LACO-OFDM having K ACO layers, there are K IFFT transforms at the transmitter, and $2K$ IFFT/FFT transforms at the receiver. Hence, its complexity order is $\mathcal{O}(KN \log_2 N)$ and $\mathcal{O}(2KN \log_2 N)$ at the transmitter and receiver, respectively. In contrast to the baselines, O-OFDMNet involves the extra DNN complexity, which is on the order of $\mathcal{O}(3QN)$ and $\mathcal{O}(3QN + (G-2)Q^2)$ at the transmitter and receiver, respectively, where for simplicity we assume that the number of hidden nodes is Q for all hidden layers of the C/R and R/C converters, as shown in Table IV. Table VI summarizes the complexity of O-OFDMNet in comparison to the baselines. It is observed that our scheme

TABLE VII
RUNTIME COMPARISON BETWEEN OUR O-OFDMNET AND BASELINES

Scheme	Runtime (ms)	Runtime/0.033
DCO-OFDM [7]	0.482	16
ACO-OFDM [9]	0.581	18
LACO-OFDM [13]	3.8	115
O-OFDMNet (1 sample)	0.577	17
O-OFDMNet (10 samples)	0.140	4
O-OFDMNet (10^2 samples)	0.045	1.4
O-OFDMNet (10^3 samples)	0.033	1

has a lower IFFT/FFT complexity than LACO-OFDM, while its DNN complexity is additional to that of the baselines.

As an indication of O-OFDMNet's ability to support real-time communications, we additionally include some runtime comparisons between O-OFDMNet and the baselines in Table VII, where all schemes operate at the SE of 2 bps/Hz. Here, the runtime is measured for signal processing of a data sample at both the transmitter and receiver. For a fair comparison, we convert our learning scheme from Tensorflow into MATLAB and measure the runtime of all schemes in MATLAB on the same computer. Since DNNs can process data in parallel, we can feed them with a batch of multiple data samples, i.e., OFDM symbols, at once. Hence, it is easy to speed up our scheme by increasing the number of samples feeding the DNNs, as shown in Table VII. In particular, with 1 sample, our scheme has a runtime similar to that of ACO-OFDM and DCO-OFDM, while LACO-OFDM has a much longer runtime than our scheme. This is because the serial interference cancellation (SIC) receiver of LACO-OFDM involves several iterations of IFFT/FFT operations, which causes a significant delay compared to the other schemes operating without SIC. More importantly, O-OFDMNet achieves an even much shorter runtime than the benchmarks, when the number of samples feeding the DNNs is increased to 10, 10^2 or 10^3 . For example, for 10^3 samples, the runtime of O-OFDMNet is 0.033 ms, which is 16, 18 and 115 times shorter than those of DCO-OFDM, ACO-OFDM and LACO-OFDM, respectively. This is mainly because our scheme benefits from the parallel processing capability of DNNs [46], which is not available in the baselines.

V. CONCLUSION

We proposed the new DL-aided O-OFDMNet concept, which employs DNNs for C/R and R/C converters for processing the TD signal at the transmitter and receiver, respectively. This new design enables O-OFDMNet to dispense with the Hermitian symmetry, thereby achieving the same SE as classical RF-OFDM, which has never been achieved

by the existing O-OFDM schemes. Furthermore, in contrast to DCO-OFDM-based systems, our scheme is more power-efficient, since it does not require the DC bias for generating the positive transmitted signal. O-OFDMNet was trained for jointly minimizing both the BER and PAPR over both AWGN and frequency selective channels. We also derived the soft decoder for our scheme, which was then adopted to both convolutional and polar codes for further improving the BER. Our simulation results showed that O-OFDMNet reduces both the BER and the PAPR of the baselines in both uncoded and coded scenarios. Furthermore, our scheme achieves the DCMC capacity approaching that of RF-OFDM, which is superior to that of the existing O-OFDM schemes. Our O-OFDMNet is also more robust to hostile signal-dependent noise and nonlinear clipping distortion. Finally, our complexity analysis showed that our scheme imposes the extra DNN complexity, while its IFFT/FFT complexity is similar to that of DCO/ACO-OFDM and lower than that of LACO-OFDM. Our analysis also showed that O-OFDMNet achieves a shorter runtime than the baselines, since it can make use of the parallel processing capability of DNNs. In our future work, we plan to carry out the experimental verification for our scheme in indoor environments.

REFERENCES

- [1] T. Hwang, C. Yang, G. Wu, S. Li, and G. Y. Li, "OFDM and its wireless applications: A survey," *IEEE Trans. Veh. Technol.*, vol. 58, no. 4, pp. 1673–1694, May 2009.
- [2] L. Hanzo, M. Munster, B. Choi, and T. Keller, *OFDM and MC-CDMA for Broadband Multi-User Communications, WLANs and Broadcasting*. Hoboken, NJ, USA: Wiley, 2005.
- [3] T. Keller and L. Hanzo, "Adaptive multicarrier modulation: A convenient framework for time-frequency processing in wireless communications," *Proc. IEEE*, vol. 88, no. 5, pp. 611–640, 2000.
- [4] M. Jiang and L. Hanzo, "Multiuser MIMO-OFDM for next-generation wireless systems," *Proc. IEEE*, vol. 95, no. 7, pp. 1430–1469, 2007.
- [5] J. Armstrong, "OFDM for optical communications," *J. Lightwave Technol.*, vol. 27, no. 3, pp. 189–204, Feb 2009.
- [6] J. M. Kahn and J. R. Barry, "Wireless infrared communications," *Proc. IEEE*, vol. 85, no. 2, pp. 265–298, 1997.
- [7] J. B. Carruthers and J. M. Kahn, "Multiple-subcarrier modulation for nondirected wireless infrared communication," *IEEE J. Sel. Areas Commun.*, vol. 14, no. 3, pp. 538–546, 1996.
- [8] J. Armstrong and B. J. C. Schmidt, "Comparison of asymmetrically clipped optical OFDM and DC-biased optical OFDM in AWGN," *IEEE Commun. Lett.*, vol. 12, no. 5, pp. 343–345, 2008.
- [9] J. Armstrong and A. J. Lowery, "Power efficient optical OFDM," *Electron. Lett.*, vol. 42, no. 6, pp. 370–372, 2006.
- [10] S. C. J. Lee, S. Randel, F. Breyer, and A. M. J. Koonen, "PAM-DMT for intensity-modulated and direct-detection optical communication systems," *IEEE Photon. Technol. Lett.*, vol. 21, no. 23, pp. 1749–1751, 2009.
- [11] S. D. Dissanayake, K. Panta, and J. Armstrong, "A novel technique to simultaneously transmit ACO-OFDM and DCO-OFDM in IM/DD systems," in *IEEE GLOBECOM*, 2011, pp. 782–786.
- [12] B. Ranjha and M. Kavehrad, "Hybrid asymmetrically clipped OFDM-based IM/DD optical wireless system," *J. Opt. Commun. Netw.*, vol. 6, no. 4, pp. 387–396, 2014.
- [13] Q. Wang, C. Qian, X. Guo, Z. Wang, D. G. Cunningham, and I. H. White, "Layered ACO-OFDM for intensity-modulated direct-detection optical wireless transmission," *Opt. Express*, vol. 23, no. 9, pp. 12 382–12 393, May 2015.
- [14] X. Zhang, Q. Wang, R. Zhang, S. Chen, and L. Hanzo, "Performance analysis of layered ACO-OFDM," *IEEE Access*, vol. 5, pp. 18 366–18 381, 2017.
- [15] X. Zhang, Z. Babar, R. Zhang, S. Chen, and L. Hanzo, "Multi-class coded layered asymmetrically clipped optical OFDM," *IEEE Trans. Commun.*, vol. 67, no. 1, pp. 578–589, 2019.
- [16] Z. Babar, X. Zhang, P. Botsinis, D. Alanis, D. Chandra, S. X. Ng, and L. Hanzo, "Near-capacity multilayered code design for LACO-OFDM-aided optical wireless systems," *IEEE Trans. Veh. Technol.*, vol. 68, no. 4, pp. 4051–4054, 2019.
- [17] X. Zhang, Z. Babar, P. Periklis, H. Haas, and L. Hanzo, "The evolution of optical OFDM," *IEEE Commun. Surv. Tuts.*, 2021, to be published.
- [18] X. Zhang, S. Chen, and L. Hanzo, "On the discrete-input continuous-output memoryless channel capacity of layered ACO-OFDM," *J. Lightwave Technol.*, vol. 38, no. 18, pp. 4955–4968, 2020.
- [19] J. Wang, C. Jiang, H. Zhang, Y. Ren, K. C. Chen, and L. Hanzo, "Thirty years of machine learning: The road to Pareto-optimal wireless networks," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 3, pp. 1472–1514, 2020.
- [20] M. S. Yee, B. L. Yeap, and L. Hanzo, "Radial basis function-assisted Turbo equalization," *IEEE Trans. Commun.*, vol. 51, no. 4, pp. 664–675, 2003.
- [21] S. X. Ng, M. S. Yee, and L. Hanzo, "Coded modulation assisted radial basis function aided Turbo equalization for dispersive Rayleigh-fading channels," *IEEE Trans. Wireless Commun.*, vol. 3, no. 6, pp. 2198–2206, 2004.
- [22] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, <http://www.deeplearningbook.org>.
- [23] H. Ye, G. Y. Li, and B. Juang, "Power of deep learning for channel estimation and signal detection in OFDM systems," *IEEE Wireless Commun. Lett.*, vol. 7, no. 1, pp. 114–117, Feb. 2018.
- [24] T. V. Luong, Y. Ko, N. A. Vien, D. H. N. Nguyen, and M. Matthaiou, "Deep learning-based detector for OFDM-IM," *IEEE Wireless Commun. Lett.*, vol. 8, no. 4, pp. 1159–1162, Aug. 2019.
- [25] T. V. Luong, Y. Ko, N. A. Vien, M. Matthaiou, and H. Q. Ngo, "Deep energy autoencoder for noncoherent multicarrier MU-SIMO systems," *IEEE Trans. Wireless Commun.*, vol. 19, no. 6, pp. 3952–3962, 2020.
- [26] T. V. Luong, Y. Ko, M. Matthaiou, N. A. Vien, M. T. Le, and V. D. Ngo, "Deep learning-aided multicarrier systems," *IEEE Trans. Wireless Commun.*, vol. 20, no. 3, pp. 2109–2119, 2021.
- [27] N. A. Amran, M. D. Soltani, M. Yaghoobi, and M. Safari, "Deep learning based signal detection for OFDM VLC systems," in *IEEE ICC*, 2020, pp. 1–6.
- [28] H. Lee, I. Lee, T. Q. S. Quek, and S. H. Lee, "Binary signaling design for visible light communication: A deep learning framework," *Opt. Express*, vol. 26, no. 14, pp. 18 131–18 142, Jul 2018.
- [29] B. Karanov, D. Lavery, P. Bayvel, and L. Schmalen, "End-to-end optimized transmission over dispersive intensity-modulated channels using bidirectional recurrent neural networks," *Opt. Exp.*, vol. 27, no. 14, pp. 19 650–19 663, 2019.
- [30] M. G. Ulkar, T. Baykas, and A. E. Pusane, "VLCnet: Deep learning based end-to-end visible light communication system," *J. Lightw. Technol.*, vol. 38, no. 21, pp. 5937–5948, 2020.
- [31] L. Hao, D. Wang, W. Cheng, J. Li, and A. Ma, "Performance enhancement of ACO-OFDM-based VLC systems using a hybrid autoencoder scheme," *Opt. Commun.*, vol. 442, pp. 110–116, Jul. 2019.
- [32] H. Lee, I. Lee, and S. H. Lee, "Deep learning based transceiver design for multi-colored VLC systems," *Opt. Express*, vol. 26, no. 5, pp. 6222–6238, Mar 2018.
- [33] C. Xu, P. Zhang, R. Rajashekar, N. Ishikawa, S. Sugiura, Z. Wang, and L. Hanzo, "Near-perfect finite-cardinality generalized space-time shift keying," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 9, pp. 2146–2164, 2019.
- [34] C. Xu, T. Bai, J. Zhang, R. G. Maunder, S. Sugiura, Z. Wang, and L. Hanzo, "Constant-envelope space-time shift keying," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 6, pp. 1387–1402, 2019.
- [35] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Netw.*, vol. 61, pp. 85–117, 2015.
- [36] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *ICML*, 2015.
- [37] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *ICLR*, 2014.
- [38] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. AISTATS*, 2010, pp. 249–256.
- [39] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin *et al.*, "Tensorflow: Large-scale machine learning on heterogeneous distributed systems," *arXiv preprint arXiv:1603.04467*, 2016.
- [40] C. Xu, N. Ishikawa, R. Rajashekar, S. Sugiura, R. G. Maunder, Z. Wang, L. Yang, and L. Hanzo, "Sixty years of coherent versus non-coherent tradeoffs and the road from 5G to wireless futures," *IEEE Access*, vol. 7, pp. 178 246–178 299, 2019.

- [41] E. Arikan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3051–3073, 2009.
- [42] J. Carruthers and J. Kahn, "Modeling of nondirected wireless infrared channels," *IEEE Trans. Commun.*, vol. 45, no. 10, pp. 1260–1268, 1997.
- [43] K. Lee, H. Park, and J. R. Barry, "Indoor channel characteristics for visible light communications," *IEEE Commun. Lett.*, vol. 15, no. 2, pp. 217–219, 2011.
- [44] S. Dörner, S. Cammerer, J. Hoydis, and S. t. Brink, "Deep learning based communication over the air," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 1, pp. 132–143, 2018.
- [45] I. Tal and A. Vardy, "List decoding of polar codes," *IEEE Trans. Inf. Theory*, vol. 61, no. 5, pp. 2213–2226, 2015.
- [46] N. Samuel, T. Diskin, and A. Wiesel, "Learning to detect," *IEEE Trans. Signal Process.*, vol. 67, no. 10, pp. 2554–2564, 2019.