Check for updates

Research Article | Vol. 26, No. 14 | 9 Jul 2018 | OPTICS EXPRESS 18131

Optics EXPRESS

Binary signaling design for visible light communication: a deep learning framework

HOON LEE,[1] INKYU LEE,[2] TONY Q. S. QUEK,[1] AND SANG HYUN LEE[2,*]

[1] *Information Systems Technology and Design Pillar, Singapore University of Technology and Design, 487372, Singapore*
[2] *School of Electrical Engineering, Korea University, Seoul 02841, South Korea*
[*] *sanghyunlee@korea.ac.kr*

Abstract: This paper develops a deep learning framework for the design of on-off keying (OOK) based binary signaling transceiver in dimmable visible light communication (VLC) systems. The dimming support for the OOK optical signal is achieved by adjusting the number of ones in a binary codeword, which boils down to a combinatorial design problem for the codebook of a constant weight code (CWC) over signal-dependent noise channels. To tackle this challenge, we employ an autoencoder (AE) approach to learn a neural network of the encoder-decoder pair that reconstructs the output identical to an input. In addition, optical channel layers and binarization techniques are introduced to reflect the physical and discrete nature of the OOK-based VLC systems. The VLC transceiver is designed and optimized via the end-to-end training procedure for the AE. Numerical results verify that the proposed transceiver performs better than baseline CWC schemes.

OCIS codes: (060.4510) Optical communications; (230.3670) Light-emitting diodes; (220.2945) Illumination design; (010.1690) Color.

References and links

1. T. Komine and M. Nakagawa, "Fundamental analysis for visible light communication system using LED light," IEEE Trans. Consum. Electron. **56**(1), 100-107 (2004).
2. S. H. Lee, S.-Y. Jung, and J. K. Kwon, "Modulation and coding for dimmable visible light communication," IEEE Commun. Mag. **53**(2), 136-143 (2015).
3. S. Kim and S.-Y. Jung, "Novel FEC coding scheme for dimmable visible light communication based on the modified Reed.Muller codes," IEEE Photon. Techno. Lett. **23**(20), 1514-1516 (2011).
4. S. H. Lee and J. K. Kwon, "Turbo code-based error correction scheme for dimmable visible light communication systems," IEEE Photon. Techno. Lett. **24**(17), 1463-1465 (2012).
5. S. Zhao, "A serial concatenation-based coding scheme for dimmable visible light communication systems," IEEE Commun. Lett. **20**(10), 1951-1954 (2016).
6. C. E. Mejia, C. N. Georghiades, M. M. Abdallah, and Y. H. Al-Badarneh, "Code design for flicker mitigation in visible light communications using finite state machines," IEEE Trans. Commun. **65**(5), 2091-2100 (2017).
7. P. Ostergard, "Classification of binary constant weight codes," IEEE Trans. Inf. Theory **56**(8), 3779-3785 (2010).
8. H. Lee, I. Lee, and S. H. Lee, "Deep learning based transceiver design for multi-colored VLC systems ," Opt. Express **26**(5), 6222-6238 (2018).
9. T. O'Shea and J. Hoydis, "An introduction to deep learning for the physical layer," IEEE Trans. Cogn. Commun. Netw. **3**(4), 563-575 (2017).
10. S. Dorner, S. Cammer, J. Hoydis, and S. Brink, "Deep learning based communication over the air," IEEE J. Sel. Topics Signal Process. **12**(1), 132-143 (2017).
11. O. Vinyals, S. Bengio, and M. Kudlur, "Order matters: sequence to sequence for sets," in Proccedings of International Conference on Learning Representations (ICLR, 2016).
12. A. Siddique and M. Tahir, "Joint error-brightness control coding for LED based VLC link," in Proccedings of Wireless Communications and Networking Conference (IEEE, 2014), pp. 400-404.
13. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature **512**(7553), 436-444 (2015).
14. Z. Cao, M. Long, J. Wang, and P. Yu, "HashNet: deep learning to hash by continuation," in Proccedings of IEEE International Conferenece on Computer Vision (ICCV, 2017).
15. D. Kingma and J. Ba, "Adam: a method for stochastic optimization," in Proccedings of International Conference on Learning Representations (ICLR, 2015).

#331278 https://doi.org/10.1364/OE.26.018131
Journal © 2018 Received 10 May 2018; revised 26 Jun 2018; accepted 26 Jun 2018; published 28 Jun 2018

Optics EXPRESS

1. Introduction

Visible light communication (VLC) in use of the visible light spectrum of light-emitting diodes (LEDs) for optical wireless communication has attracted great research interest [1,2]. The lighting purpose of LEDs gives rise to an intensity constraint that the average intensity adapts to the dimming requirement imposed by external preference. For the adaptation to this requirement, the intensity modulation along with direct detection (IM/DD) is typically used in VLC. Various modulation techniques have been developed based on IM/DD, such as on-off keying (OOK), pulse position modulation (PPM), color shift keying (CSK) and orthogonal frequency division multiplexing (OFDM), and OOK is practical from the simplicity point of view. To ensure the reliability of the OOK-based VLC systems, various coding techniques have been developed [3–6].

Most coding schemes stick to a strict dimming constraint, thereby leading to the use of constant weight codes (CWCs) as a component binary signaling technique. However, the design of an efficient CWC is a challenging task [5], and it is mostly replaced by a randomly generated CWC [5] or a punctured version of a uniform weight code [3], which are obviously far from the optimum [7]. Existing studies on the optimal CWCs have focused on the characterization of mathematical properties such as Hamming weight and minimum Hamming distance, whereas the optimal and practical encoding and decoding methods have not been adequately studied in generic system configurations. The generation of the optimal CWC typically relies on exhaustive search processes, which does not provide any directions for designing efficient OOK-based VLC systems. In addition, existing schemes are prone to the vulnerability to channel imperfection, such as channel crosstalk and signal-dependent shot noise, since these methods are designed without taking into consideration of such impairments.

To handle these shortcomings, this paper addresses the design of the CWC-based binary signaling scheme and the corresponding transmitter/receiver pair using a deep learning (DL) framework. In particular, an autoencoder (AE) is employed to learn the set of binary signaling codewords subject to constant weight as well as to minimize the decoding error probability of the learned codebook under practical optical channels. The AE techniques have been exploited for the joint design of transmitter/receiver pair in multi-colored VLC [8] and radio frequency communication systems [9,10], focusing on the training of real-valued modulation constellation points. In the real-valued VLC systems [8], lighting constraints are given as affine functions that can be handled by adding convex projection operations to the AE. In contrast, the OOK-based VLC systems need to obtain a set of binary codewords to encode the transmitted message, incurring discrete (non-convex) lighting constraints for the binary restriction. Therefore, conventional convex projection algorithms in [8] are not straightforwardly applied to the OOK systems.

For designing the binary signaling, it requires to learn to map message indices to set of binary output sequences corresponding to the OOK codewords. One naive approach is to employ a hard thresholding layer which directly forces the output of the AE either to zero or one. From a learning point of view, however, this incurs a *vanishing gradient* problem, which is the most challenging issue for training deep neural networks [11]. For this reason, conventional neural network training algorithms in [8–10] would fail to handle the binary signaling design problem.

To tackle this challenge, we develop a training strategy to calculate the gradient for binary inputs. To be specific, the AE is trained with a soft binarization procedure, which is gradually *annealed* to a hard binarization during the training step. Based on the proposed training approach, the AE successfully learns the nature of the binary VLC systems and yields an efficient OOK transceiver over general optical channels which cannot be achieved by the previous techniques in [8]. Numerical results verify the impact of the proposed training strategy over conventional algorithms and show the optimality of the resulting AE codebook as well as the superiority of the corresponding transceiver over baseline CWC schemes.

Notations: Throughout this paper, uppercase boldface letters, lowercase boldface letters, and normal letters represent matrices, vectors, and scalar quantities, respectively. We denote a set of

Optics EXPRESS

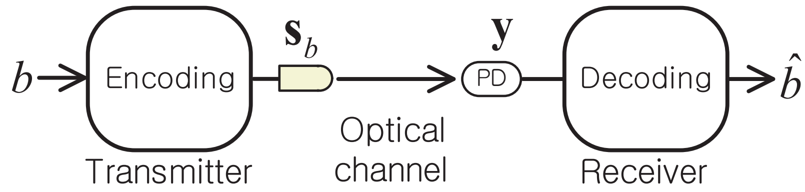

Fig. 1. System model for OOK-based dimmable VLC.

all real matrices of size P-by-Q by $\mathbb{R}^{P\times Q}$. Also, $[\mathbf{x}]_p$ indicates the p-th element of a vector $\mathbf{x}$.

2. System model and autoencoder basics

We describe the system model for a dimmable OOK-based VLC and give a brief introduction to the concept of AE.

2.1. System model

Figure 1 shows an OOK-based VLC system where a transmitter equipped with a white LED sends $M \triangleq 2^k$ different messages $b \in \mathcal{M} \triangleq \{1, \cdots, M\}$ to a receiver that estimates transmitted message $\hat{b}$, where $k \triangleq \log_2 M$ denotes the number of transmitted bits. Message b is encoded by binary codeword $\mathbf{s}_b \in \mathcal{S}$ of length N, where $\mathcal{S}$ accounts for a codebook. For dimming control, $\mathcal{S}$ is designed such that the Hamming weight of the codewords, i.e., the number of ones in $\mathbf{s}_b$, is equal to the desired dimming target $W \in \{1, \cdots, N\}$. This can be expressed as

$$\sum_{i=1}^{N}[\mathbf{s}_b]_i = W, \text{ for } \mathbf{s}_b \in \mathcal{S}. \tag{1}$$

Such codebook $\mathcal{S}$ is called a CWC with length N and weight W [7]. The received intensity vector denoted by $\mathbf{y} \in \mathbb{R}^{N\times 1}$ is given by

$$\mathbf{y} = \mathbf{s}_b + \mathbf{n}_{th} + \mathbf{n}_{sh}, \tag{2}$$

where $\mathbf{n}_{th} \in \mathbb{R}^{N\times 1}$ with $[\mathbf{n}_{th}]_i \sim \mathcal{N}(0, \sigma^2)$ $(i = 1, \cdots, N)$ and $\mathbf{n}_{sh} \in \mathbb{R}^{N\times 1}$ with $[\mathbf{n}_{sh}]_i \sim \mathcal{N}(0, s_i\psi^2\sigma^2)$ are thermal noise and signal-dependent shot noise, respectively, and ψ^2 denotes the shot noise variance scaling factor.

We aim at identifying efficient encoding and decoding strategy of the OOK-based VLC systems, such that the receiver successfully recovers the transmitted message, i.e., to minimize the symbol error rate (SER) $\Pr\{b \neq \hat{b}\}$. This results in a joint optimization task of the design of codebook $\mathcal{S}$ satisfying the dimming constraint (1) and of the decoding technique for the shot noise channel (2), which usually requires a computationally demanding search process to determine the optimal solution.

The mathematical properties of the optimal CWC, which achieves the maximum of the minimum Hamming distance performance over all possible CWCs, have been studied extensively in terms of the codebook size for each N and W [7]. However, a practical CWC generation method has not been adequately investigated with a few configurations of N and W available [12]. Furthermore, conventional CWC constructions are based on the maximization of the minimum Hamming distance and are, in general, not efficient over a signal-dependent shot noise channel, where construction criterion varies depending on the signal dependency of noise. Although a randomly generated CWC codebook can be employed for simple dimming control as done in [5], it will incur the SER degradation. To tackle this, we develop a DL based design method for a VLC transceiver that can adapt to general configuration.

Optics EXPRESS

2.2. Autoencoder basics

Let us briefly recap the basics of the AE concept. We consider a neural network with L layers which maps an input vector $\mathbf{x}_0 \in \mathbb{R}^{D\times 1}$ to an output vector $\mathbf{x}_L \in \mathbb{R}^{N_L\times 1}$, where D represents the input dimension and N_l for $l = 1, \cdots, L$ equals the number of neurons at layer l. Each layer l computes its output vector $\mathbf{x}_l \in \mathbb{R}^{N_l\times 1}$ as

$$\mathbf{x}_l = f_l(\mathbf{W}_l\mathbf{x}_{l-1} + \mathbf{b}_l), \text{ for } l = 1, \cdots, L, \tag{3}$$

where $\mathbf{W}_l \in \mathbb{R}^{N_l\times N_{l-1}}$ and $\mathbf{b}_l \in \mathbb{R}^{N_l\times 1}$ denote a weight matrix and a bias vector at layer l, respectively, and a vector function $f_l(\mathbf{x}) : \mathbb{R}^{N_l\times 1} \rightarrow \mathbb{R}^{N_l\times 1}$ represents an activation at layer l which adds non-linearity to the neural network.

The goal of the AE training is to reconstruct the input vector $\mathbf{x}_0$ from cascaded neural network computations in (3), such that the output becomes close to the input as $\mathbf{x}_L \simeq \mathbf{x}_0$ with $N_L = D$. Let $\mathbf{x}_0^{(j)}$ for $j = 1, \cdots, J$ denote the j-th training sample in a training set $\mathcal{J} \triangleq \{\mathbf{x}_0^{(1)}, \cdots, \mathbf{x}_0^{(J)}\}$ and $\mathbf{x}_L^{(j)}$ as the corresponding output. Then, the AE training process solves an optimization problem as follows:

$$\min_{\Theta} C(\Theta) \triangleq \frac{1}{J}\sum_{j\in\mathcal{J}} c(\mathbf{x}_0^{(j)}, \mathbf{x}_L^{(j)}), \tag{4}$$

where $\Theta \triangleq \{\mathbf{W}_l, \mathbf{b}_l\}$ indicates the parameter of the AE, $C(\Theta)$ stands for a cost function, and a function $c(\mathbf{x}_0, \mathbf{x}_L)$ characterizes the affinity between the input and the output vectors. A popular choice for the cost function is the mean-square error $c(\mathbf{x}_0, \mathbf{x}_L) = \|\mathbf{x}_0 - \mathbf{x}_L\|^2$ for a real-valued input or the categorial cross entropy $c(\mathbf{x}_0, \mathbf{x}_L) = -\sum_{d=1}^{D}[\mathbf{x}_0]_d \log[\mathbf{x}_L]_d$ for a binary input.

A solution of (4) can be obtained via a stochastic gradient descent (SGD) algorithm that updates parameter Θ_t at the t-th iteration as follows:

$$\Theta_t = \Theta_{t-1} - \eta \bigtriangledown C(\Theta_{t-1}), \tag{5}$$

where $\bigtriangledown C(\Theta)$ indicates the gradient of the cost function, and the positive constant η denotes the learning rate. We can efficiently compute the gradient $\bigtriangledown C(\Theta)$ by the back propagation algorithm based on the chain rule [13].

Training such a deep neural network would be non-trivial for the vanishing gradient problem. This is induced by an improper choice of activation functions which have small gradient values. The network parameter Θ may not be effectively optimized by the SGD algorithm (5) if the gradient $\bigtriangledown C(\Theta)$ is of a small value. As a result, the AE training step converges very slowly or ends up getting stuck within a few iterations. Therefore, the development of a proper learning strategy is an important issue for the AE training, which can vary with types of training data and applications.

3. Proposed autoencoder approach

Figure 2 presents the proposed AE structure for training the OOK-based dimmable VLC system, comprised of a transmitter, an optical channel, and a receiver as in Fig. 1.

3.1. Transmitter

At the transmitter block, an input message b is first represented by an one-hot vector $\mathbf{e}_b \in \mathbb{R}^{N\times 1}$, an all zero vector except the b-th element being replaced with one. This one-hot vector is processed by two subsequent layers 1 and 2. For activations, a rectified linear unit (ReLU) $f_1(\mathbf{x}) = \max\{\mathbf{0}, \mathbf{x}\}$ is used at layer 1, which is a common non-linear activation in recent DL applications [13], whereas a linear activation $f_2(\mathbf{x}) = \mathbf{x}$ is adopted at layer 2. The output vectors

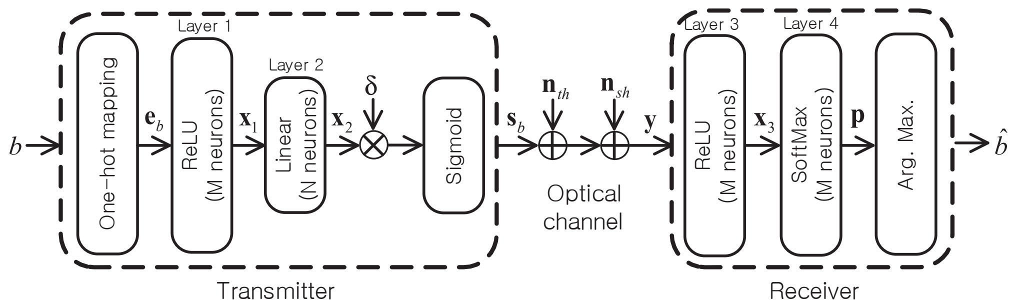

Fig. 2. AE structure for OOK-based dimmable VLC.

$\mathbf{x}_1 \in \mathbb{R}^{M\times 1}$ and $\mathbf{x}_2 \in \mathbb{R}^{N\times 1}$ at layers 1 and 2 can be computed from (3) by setting $\mathbf{x}_0 = \mathbf{e}_b$. Here, the number of neurons at layers 1 and 2 are fixed as $N_1 = M$ and $N_2 = N$, respectively, to map M different message vectors $\mathbf{e}_b$ into binary codewords $\mathbf{s}_b$ of length N.

To obtain a binary output vector $\mathbf{s}_b$ from $\mathbf{x}_2$, one may wish to employ a simple binary activation function, e.g., a unit step function, whose gradient is zero for all domain. Unfortunately, training neural networks with such activations is quite challenging in practice since the gradient of the cost function with respect to the parameters $\{\mathbf{W}_l, \mathbf{b}_l\}_{l=1}^2$ becomes zero [13]. In the DL, this is a well-known vanishing gradient issue where neural network parameters do not get updated with the SGD algorithm (5) and are usually stuck at a point of poor performance, as will be shown by numerical simulation in Section 4. Thus, the AE construction for the OOK-based VLC turns out to be inherently more difficult than the real-valued AE learning [8–10].

To overcome this difficulty, we employ a sigmoid activation layer at the end of the transmitter block as follows:

$$[\mathbf{s}_b]_i = \frac{1}{1 + e^{-\delta[\mathbf{x}_2]_i}}, \text{ for } i = 1, \cdots, N. \tag{6}$$

where the non-negative number δ controls the hardness of the output vector $\mathbf{s}_b$. As illustrated in Fig. 3, the sigmoid in (6) changes smoothly and converges to the unit step function providing a binary output, while it yields real-valued output $[\mathbf{s}_b]_i \in [0, 1]$ for a finite δ. A method for adjusting the hardness parameter δ will be clearly explained later.

3.2. Receiver

To reflect the properties of the optical channel in the AE training, we add noise vectors $\mathbf{n}_{th}$ and $\mathbf{n}_{sh}$ to the transmitter block output $\mathbf{s}_b$ to obtain the received intensity vector $\mathbf{y}$ in (2). Then, the receiver block computes its output vector $\mathbf{p} \in \mathbb{R}^{M\times 1}$ via layers 3 and 4 of which activations are given by ReLU $f_3(\mathbf{x}) = \max\{\mathbf{0}, \mathbf{x}\}$ and softmax function $[f_4(\mathbf{x})]_m = e^{[\mathbf{x}]_m} / \sum_{k=1}^{M} e^{[\mathbf{x}]_k}$, respectively. By defining $\mathbf{z} \triangleq \mathbf{W}_4\mathbf{x}_3 + \mathbf{b}_4 \in \mathbb{R}^{M\times 1}$ with the output vector of layer 3 $\mathbf{x}_3 \in \mathbb{R}^{M\times 1}$, the output is calculated as

$$[\mathbf{p}]_m = \frac{e^{[\mathbf{z}]_m}}{\sum_{q=1}^{M} e^{[\mathbf{z}]_q}}, \text{ for } m = 1, \cdots, M, \tag{7}$$

where the number of neurons at layers 3 and 4 are, respectively, selected as $N_3 = N_4 = M$ in order to recover M different messages from the received intensity $\mathbf{y}$. Note that the softmax activation in (7) has been widely utilized in DL applications for classification since each element of $\mathbf{p}$ characterizes the probability of the class to which the network input belongs. In our case, $[\mathbf{p}]_b$ indicates the probability of each message $b \in \mathcal{M}$ being transmitted. Thus, the transmitted message can be estimated as $\hat{b} = \arg\max_{1\le m\le M}[\mathbf{p}]_m$.

Optics EXPRESS

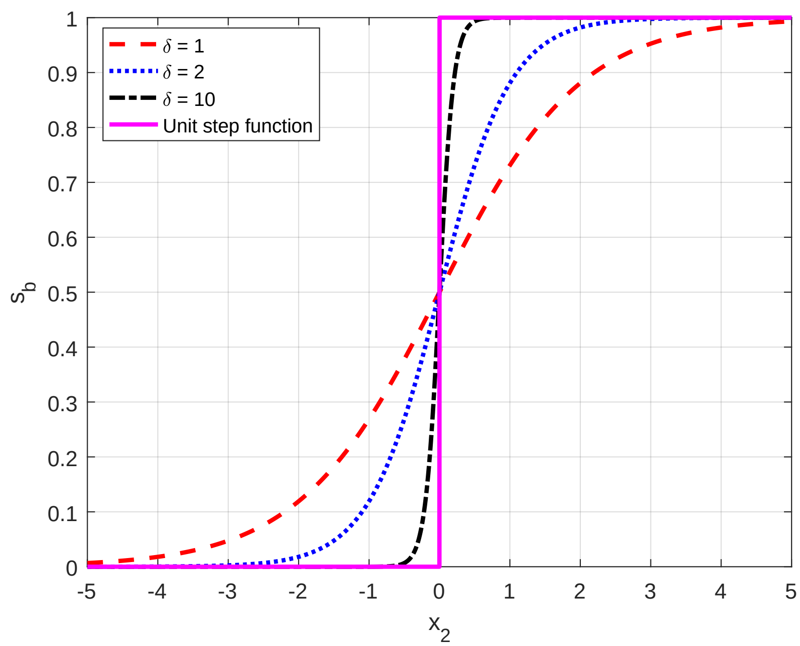

Fig. 3. Sigmoid activation with different δ.

3.3. Proposed training strategy

We present a training strategy for the proposed AE which yields the binary output vector $\mathbf{s}_b$ at the transmitter block while satisfying the dimming constraint in (1). Since the AE structure accepts the message index $b \in \mathcal{M}$ as an input, we construct the training data set $\mathcal{J} = \{b^{(1)}, \cdots, b^{(J)}\}$ with randomly generated message index $b^{(j)} \in \mathcal{M}$ for $j = 1, \cdots, J$. Then, we can define the cost function $C(\Theta)$ as

$$C(\Theta) = \frac{1}{J}\sum_{j=1}^{J}\log([\mathbf{p}]_{b^{(j)}}) + \frac{\lambda}{J}\sum_{j=1}^{J}\left(\sum_{i=1}^{N}[\mathbf{s}_{b^{(j)}}]_i - W\right)^2, \tag{8}$$

where the first term in (8) accounts for the categorial cross entropy between the one-hot representation $\mathbf{e}_{b^{(j)}}$ and the classification probability $\mathbf{p}$ for a training data $b^{(j)} \in \mathcal{J}$, and the second term is included to satisfy the target dimming constraint (1). Here, the non-negative parameter λ controls the tradeoff between the symbol recovery performance and the dimming constraint. The choice of λ is a critical issue in the AE training. If λ is too small, the dimming constraint would not be fulfilled. Otherwise, for a large λ, the categorial cross entropy term would not be sufficiently minimized, and thus the SER performance may be degraded. By intensive simulations, we have found that $\lambda = 0.03$ achieves a good tradeoff for the cost function in (8).

Next, we consider the binary constraint for $\mathbf{s}_b$ by controlling the hardness parameter δ in (6). It is emphasized that, if δ is set to a large number, we can easily attain the binary vector $\mathbf{s}_b$ since the sigmoid activation in (6) becomes the unit step function. However, as we mentioned before, it will lead to the gradient vanishing problem that makes the AE training difficult. To this end, we adopt a *multi-stage* training approach illustrated in Fig. 4, which increases the hardness parameter δ gradually at each stage of the AE training [14]. Let $\delta[g]$ and $\Theta[g]$ be the hardness parameter

Optics EXPRESS

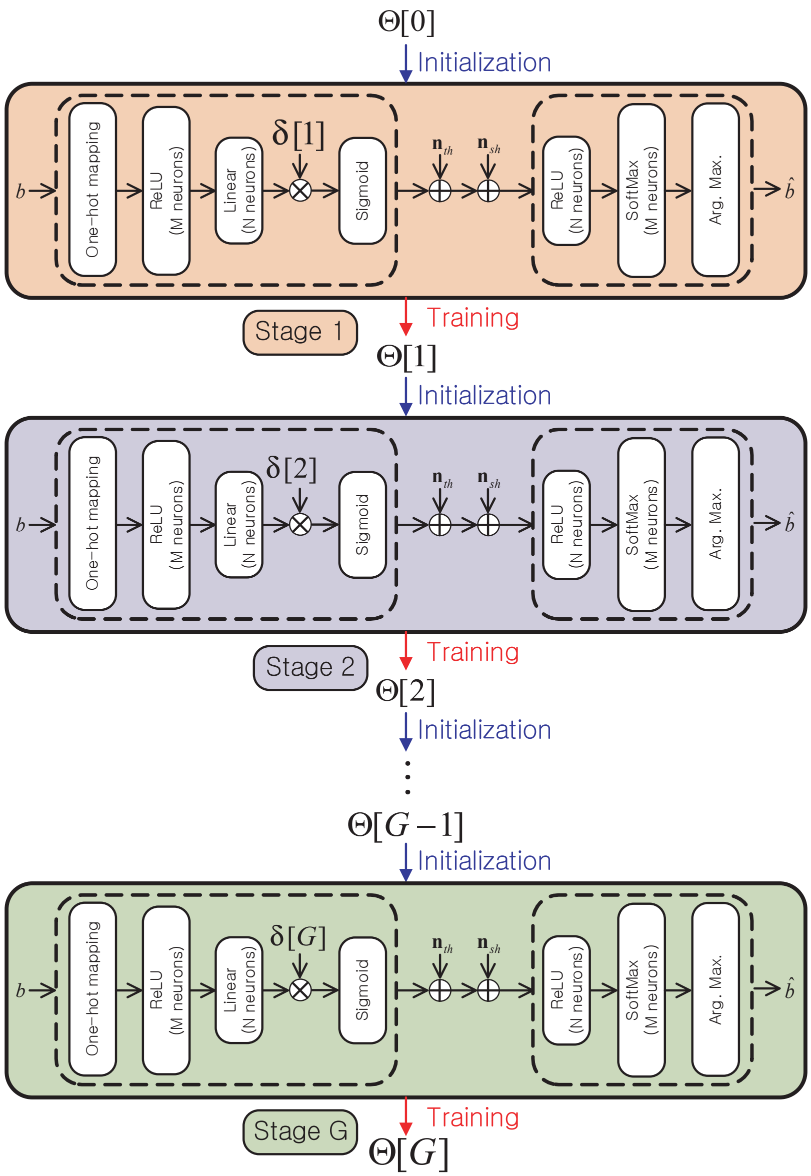

Fig. 4. Multi-stage training strategy for dimming constraint.

and the trained AE parameter at stage $g = 1, \cdots, G$, respectively. We first initialize $\Theta[0]$ and $\{\delta[g]\}$ such that $\delta[1] < \cdots < \delta[G]$. To ensure the binary output $\mathbf{s}_b$, $\delta[G]$ must be sufficiently large, while $\delta[g]$ for $g = 1, \cdots, G-1$ is chosen to be a moderate number to avoid the gradient vanishing problem. At each stage g, the AE is trained with $\delta = \delta[g]$ based on (5) from the initial point $\Theta_0 = \Theta[g-1]$. Then, we obtain the learned AE parameter $\Theta[g]$, which is utilized for the initialization of the consecutive stage with increased hardness parameter. Hence, at each stage, we can carefully *fine-tune* the AE trained at the previous stage, which is easier than training the AE directly with the unit step activation [13, 14]. Finally, the AE learned with $\delta = \delta[G]$ to produce the binary codewords. It is found by simulation that $\delta[G] = 1000$ is sufficient to produce the binary output. The overall training strategy is summarized in Algorithm 1.

Table 1. **Proposed multi-stage training strategy.**

Multi-stage training strategy
Initialize $\Theta[0]$ and $\{\delta[g]\}$ such tat $\delta[1] < \cdots < \delta[G]$.
for $g = 1 : G$
Set the initial point of the SGD algorithm to $\Theta_0 = \Theta[g-1]$.
Train (8) with $\delta = \delta[g]$ from (5) until convergence.
Obtain the learned parameter $\Theta[g]$ of the AE.
end

With the aid of state-of-the-art DL techniques and general-purpose GPUs, the training of the AE can be efficiently carried out in advance before the transmission [8, 9]. Thus, the training overhead can be handled by the offline SGD algorithms. After the training step, the transmitter and receiver blocks can be utilized as a VLC transceiver pair by storing the learned AE parameter in the memory units. As a result, the modulation and the demodulation processes of the trained AE can be implemented with simple matrix multiplications and additions in (3), which are valid in real-time VLC systems. The performance of the trained AE is then evaluated with $\delta = \delta[G]$ over unseen additive noise vectors during the training.

4. Numerical results

4.1. *Implementation details*

We present numerical results for evaluating the performance of the designed VLC transceiver in Fig. 2 with $N = 12$. The signal-to-noise ratio (SNR) is defined as $1/\sigma^2$. The Adam optimizer [15], one of the most powerful SGD algorithms, is adopted to optimize the AE parameter Θ with learning rate $\eta = 0.001$. We utilize the set of $J = 10^6$ samples for the training step, and another set of 10^6 samples is employed for the validation step such as searching the hyper-parameters λ and $\{\delta[g]\}$. Then, the average SER of the trained AE is evaluated over a new test data set with 10^9 samples. The message indices and the noise vectors are generated independently at random for the training, the validation, and the testing step for the AE with different random seeds. We implement the neural network in Python 3.5.2 with TensorFlow 1.4.0.

For simulation, we consider two different cases of $\psi^2 = 0$ and 5. In the absence of shot noise, i.e., $\psi^2 = 0$, the AE is trained at a certain SNR value and subsequently is applied to all SNR range in the testing step, which proves efficient for a thermal noise case [8, 9]. However, this would not be suitable for general signal-dependent shot noise channels $\psi^2 = 5$ where the optimal transmission strategy strongly depends on the noise power. To this end, in the case of $\psi^2 = 5$, we train two AEs at different SNR values SNR_{low} and SNR_{high}, each of which is employed for low and high SNR regimes in the testing step, respectively. Notice that the SNR values utilized in the training are hyper-parameters optimized via the validation step.

Optics EXPRESS

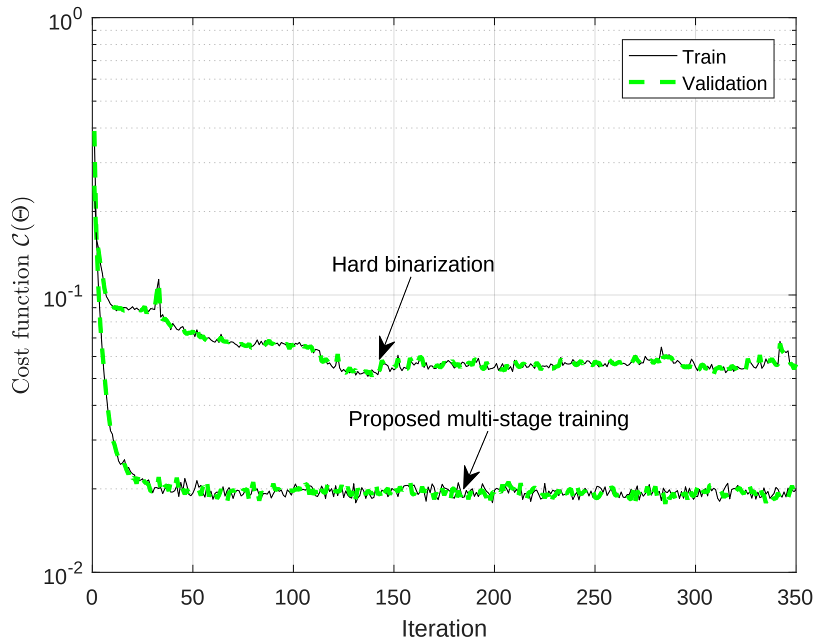

Fig. 5. Convergence behavior of AE with different binarization methods for $W = k = 5$, $\psi^2 = 0$, and SNR = 10 dB.

4.2. Impact of multi-stage training strategy

Figure 5 illustrates the convergence behavior of the proposed multi-stage training strategy. We evaluate the AE cost function $C(\Theta)$ in (8) over both training data and validation data with $W = k = 5$, $\psi^2 = 0$, and SNR = 10 dB. Here, the number of stages G and the number of iterations of the SGD algorithm at each stage are set to $G = 7$ and 50, respectively, thereby resulting in total 350 iterations. To see the impact of the multi-stage training strategy, we also present the convergence of a conventional single-stage training approach that trains the AE with $G = 1$ and $\delta[1] = 1000$ during 350 iterations, i.e., the sigmoid function (6) is replaced with the hard binarization function. It is observed that the cost function decreases gradually in the proposed training method, indicating that the AE parameters are efficiently optimized for successful estimation of the transmitted message. This is because the proposed training strategy smoothly grows the hardness parameter δ at each stage such that the sigmoid activation in (6) approaches to the unit step function during the training step. As a result, the AE can gradually adapt to the binary signaling environment without incurring the vanishing gradient problem. On the other hand, if the AE is trained with the unit step function, the cost function no longer decreases after 130 iterations and exhibits a poor performance due to the vanishing gradient issue. Note that the minimum Hamming distances of the CWCs leaned by the AE with the proposed training algorithm and the unit step function are given by 4 and 2, respectively. This reveals that conventional AE transceiver schemes in [8–10] which designs real-valued transmitted signals cannot be straightforwardly applied to binary signaling systems.

4.3. Performance comparison

We compare the performance of the proposed AE transceiver with following baseline schemes.

Optics EXPRESS

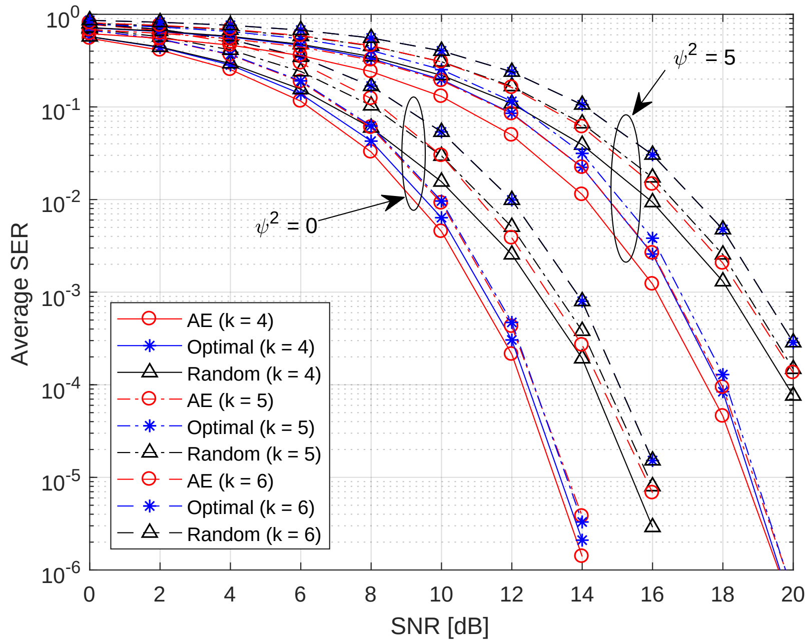

Fig. 6. Average SER performance as a function of SNR with $\psi^2 = 0$ and 5 for $W = 4$.

- *Random [5]:* Random generation of CWC is applied for each experiment.
- *Optimal [7]:* The optimal CWC, which achieves the theoretical maximum of the minimum Hamming distance, is found by exhaustive search.

For both schemes, the maximum likelihood detector is employed for the demodulation at the receiver. We present the minimum Hamming distance of the optimal CWC and the proposed

Table 2. **Minimum Hamming distance for $\psi^2 = 0$.**

	Optimal [7]				AE			
$k\backslash W$	3	4	5	6	3	4	5	6
4	4	4	4	6	4	4	4	6
5	2	4	4	4	2	4	4	4
6	2	2	4	4	2	2	4	4

AE in Table 2 without shot noise $\psi^2 = 0$ for different number of bits $k \in \{4, 5, 6\}$ and dimming target $W = \{3, \cdots, 6\}$. Note that the minimum Hamming distance of the random CWC is two for all k and W. It can be seen that, for all simulated cases, the proposed AE achieves the theoretical minimum Hamming distance performance investigated in [7]. This implies that the AE can learn the optimal CWC generation rules by itself for successful message transmission.

Figures 6 and 7 depict the average SER performance of the AE and the baseline schemes as a function of the SNR for $\psi^2 = 0$ and 5 with $k \in \{4, 5, 6\}$ and $W = \{4, 5\}$. The randomly searched CWC shows the worst performance consistently in all configurations. The proposed AE transceiver outperforms the CWC for all simulated setups. It is interesting to see that, in the

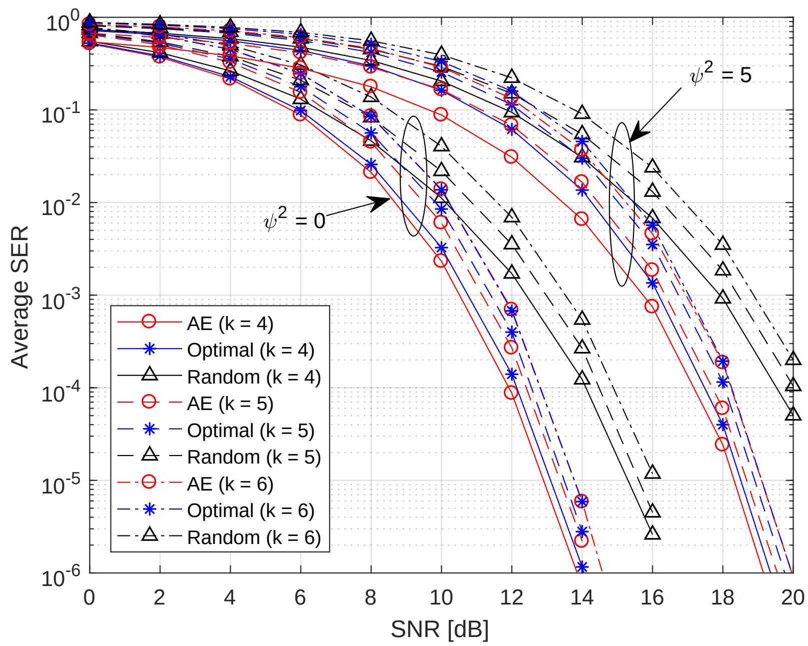

Fig. 7. Average SER performance as a function of SNR with $\psi^2 = 0$ and 5 for $W = 5$.

thermal noise case $\psi^2 = 0$, the proposed AE performs better than the CWC which optimally maximizes the minimum Hamming distance, regardless of the dimming constraints. This is because the AE method directly minimizes the erroneous detection probability via the end-to-end training step, while the optimal CWC focuses on the maximization of the minimum Hamming distance among the codewords. Also, we can observe that the proposed AE transceiver is effective for the shot noise channel $\psi^2 = 5$, in particular, in the low SNR regime where the shot noise becomes dominant. This implies that the AE could efficiently learn the complicated nature of the shot noise channel in which the optimal CWC cannot achieve sufficiently good SER performance. Therefore, we can conclude that the proposed AE framework allows identifying a more efficient OOK transceiver over general VLC channels compared to the conventional CWC designs.

5. Conclusion

This paper has studied OOK-based dimmable VLC transceiver design problems for reliable symbol recovery, which requires computationally demanding search process to optimize a binary CWC encoding strategy as well as the decoding rule for signal-dependent shot noise channels. To combat with complicated optical nature of the dimming control and channel imperfection, the AE based DL approach has been applied to this challenge. We have presented soft binarization procedures for efficient AE training such that the output of the transmitter gradually converges to binary codewords without the vanishing gradient problem. As compared to conventional CWC designs which rely on a brute-force search for the determining the optimal codebook, the proposed AE framework optimizes the OOK transceiver in a systematic way over arbitrary VLC channels including a simple thermal noise channel as well as a signal-dependent shot noise channel. As a result, the learned AE can yield efficient CWCs that achieves theoretically the best minimum Hamming distance and improves the average SER performance. Numerical results have

Optics EXPRESS

demonstrated that the proposed AE approach provides superior performance over conventional CWC designs.

This work has focused on a setup where the channel between the transmitter and the receiver remains unchanged. For a future research direction, the mobility of the transceiver can be taken into account for the OOK-based VLC system design. In this configuration, the channel may vary during one symbol duration, and thus the channel knowledge may not be available at the transceiver. The AE framework can be extended to the mobile setup by including a time-varying channel effect to the optical channel block of the proposed AE structure. With random channel samples generated based on accurate models, the AE can learn the effect of the mobility by itself via the end-to-end training procedure. In this sense, developing efficient learning algorithms and investigating the impact of the AE framework for the time-varying VLC channels are worthwhile.

Funding

National Research Foundation of Korea (NRF) (2015R1C1A1A01052529, 2017R1A2B3012316); Korea University Grant.