Covert Water-to-Air Optical Wireless Communication Based on an Adversarial Autoencoder

Qingqing Hu, Nuo Huang, Yuwei Chen, and Chen Gong, *Senior Member, IEEE*

Abstract—This paper proposes a covert scheme based on an adversarial autoencoder for water-to-air (W2A) optical wireless communication (OWC) systems. We develop optimization algorithms to train the proposed covert scheme, such that a watchful adversary cannot distinguish the generated covert signal from artificial noise (AN), and a legitimate receiver can demodulate the covert signal correctly. The validation set results show that a high level of covertness performance can be achieved in W2A channels. In addition, the proposed scheme is tested under different receiver noise intensities, block lengths and code rates, where linear block codes (LBCs) are adopted as the performance comparison benchmark. The minimum block error rate (BLER) under hard decision and the minimum Jensen-Shannon (JS) divergence between the covert signal and AN among all LBC groups are both higher than those of the proposed scheme, while the minimum BLER under soft decision with pilot-free channel estimation is close to that of the proposed scheme. We further establish an experimental W2A-OWC system to verify the scheme performance under different wave intensities and position offsets. It is shown that the BLER of the proposed scheme is slightly higher than that of LBCs, while the JS divergence of the proposed scheme is significantly lower than that of LBCs. The related code will be available at https://github.com/qingqinghu-Ricky/AAE-for-covert-communication upon publication.

Index Terms—Covert scheme, water-to-air, optical wireless communication, adversarial autoencoder.

I. INTRODUCTION

NOWADAYS, there has been growing demands for underwater applications, such as oceanic studies, marine resource exploration, and underwater rescue [1]. Consequently, it is imperative to establish direct communication across the water-air interface [2]. However, underwater radio frequency signal experiences significant signal absorption, and acoustic signal exhibits high latency [3]. Optical wireless communication (OWC) is a highly promising solution for water-to-air (W2A) communication, offering higher data rates and lower transmission latency at medium to short distances.

In W2A-OWC systems, the water surface fluctuation and refraction effects cause directional jitter of the light beam and alterations in light intensity distribution, which further increase the probability of information leakage [4]. Especially, the communication behaviors can be detected by a watchful adversary (Willie) in outdoor W2A-OWC systems, which threatens the stability of communication links. Therefore, covert communications, also known as low-probability-of-detection communications, are required to hide the existence of wireless communications and effectively improve the system security [5].

Recent researches on covert communication schemes in different scenarios can be characterized into several categories. Several beamforming schemes based on multiple antennas were designed to improve the covert communication performance in millimeter-wave communication systems [6], multi-input multi-output systems [7], and frequency-diverse-array-aided near-field communication systems [8]. Moreover, in uplink non-orthogonal multiple access (NOMA) systems [9] and satellite-related NOMA systems [10], a covert user transmitted the private message assisted by a reliable user transmitting the public message. Artificial noise (AN) or jamming can be adopted by a multi-antenna transmitter [11], a friendly jammer [12], [13], and a cognitive jammer [14] to cover covert information transmission, respectively. Several covert transmission parameters can be optimized under natural noise. For instance, the probability of channel selection [15], the flying location and the transmitted power of unmanned aerial vehicle (UAV) [16], the resource allocation and UAV trajectory [17], the transmission power and the number of tolerable slots [18] were optimized to enhance the communication rates under covertness constraints. Assisted by intelligent reflecting surface (IRS), the signal power received at the legitimate receiver can be strengthened and the signal power received at Willie can be weakened in [19], [20], [21]. Furthermore, several channel inverse power control schemes were adopted in Internet-of-Things systems [22] and communication systems with finite block length [23] to guarantee a constant signal power at the receiver and hide the

Qingqing Hu is with the School of Computer and Information, Hefei University of Technology, Hefei 230601, China, and also with the CAS Key Laboratory of Wireless-Optical Communications, University of Science and Technology of China, Hefei 230052, China (e-mail: ruixihu@mail.ustc.edu.cn).

Nuo Huang and Chen Gong are with the CAS Key Laboratory of Wireless-Optical Communications, University of Science and Technology of China, Hefei 230052, China (e-mail: huangnuo@ustc.edu.cn; cgong821@ustc.edu.cn).

Yuwei Chen is with the Anhui Province Key Laboratory of Advanced Laser Technology, Hefei 230000, China (e-mail: chenyuwei_cas@hotmail.com).

Digital Object Identifier 10.1109/TCCN.2024.3524625

transmitter from Willie. In terms of covert signal design, the information was hidden in the index of spreading codes [24], the correlation coefficients of two consecutive Gaussian sequences [25], and the parameters of stable non-Gaussian noise sequences [26], respectively. Furthermore, different covert signals were designed, such as Faster-than-Nyquist signals [27], chaotic pseudo orthogonal signals [28], and distance-adaptive absorption peak modulation in Terahertz covert communications [29]. Some existing covert schemes require additional antennas, auxiliary users, jammers, IRS, or power, which cannot be applied in resource-limited scenarios. In addition, previous covert schemes based on transmission parameter optimization under nature noise are not suitable for scenarios with relatively low natural noise.

In this work, we propose a novel covert scheme for W2A-OWC systems. Different from the previous covert signal design schemes requiring manual feature extraction, such as the index of spread spectrum codes in [24] and the correlation coefficients in [25], we adopt an adversarial autoencoder (AAE) for nonlinear coding where the covert information is automatically hidden in the generated pseudorandom signals during the coding process. Compared with the covert scheme in our previous work [30] requiring strong ambient radiation to achieve signal covertness, the completely different scheme proposed in this work is more robust to ambient radiation intensity and can also be applied in weak ambient radiation scenarios. In addition, the proposed covert communication scheme can also be adopted for other OWC scenarios with intensity modulation/direct detection (IM/DD) and random channel fading, such as visible light communication and free-space optical systems. After replacing the W2A-OWC channel with another channel model, the scheme can be re-trained and the network parameters will be updated.

The contributions of this work can be summarized as follows:

1) We propose a covert scheme in dynamic W2A-OWC scenarios based on an AAE to generate pseudorandom signals. The proposed scheme hides the covert information in the transmitted signals without needing extra assistance resources, such as antennas, user, and IRS. We develop an optimization algorithm for training the covert scheme, such that Willie cannot distinguish the signal from AN, and Bob can demodulate the covert signal correctly. The proposed scheme takes advantage of GAN and autoencoder to design and generate the covert signal, without the need for additional bandwidth like spreading spectrum codes or manual feature extraction for information hiding.

2) We test the proposed scheme under different receiver noise intensities, block lengths, and code rates, and adopt linear block codes (LBCs) as the performance comparison benchmark. After traversing all LBC groups, the minimum block error rate (BLER) under hard decision and the minimum Jensen-Shannon (JS) divergence between the covert signals and AN of LBCs are both higher than those of the proposed scheme, while the minimum BLER under soft decision with pilot-free channel estimation is close to that of the proposed scheme.
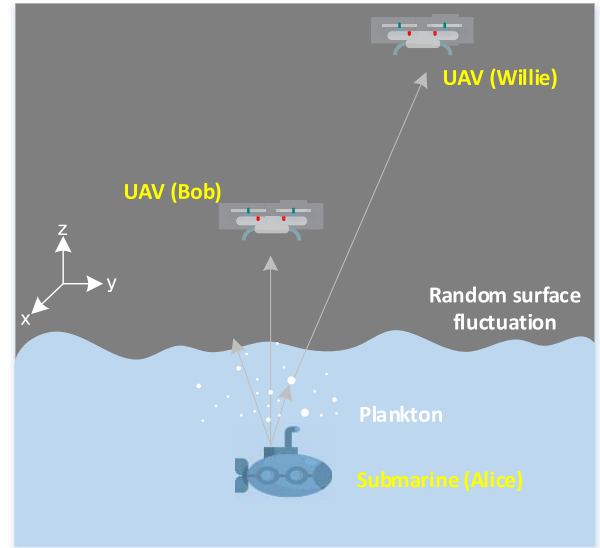

Fig. 1. Illustration of a W2A-OWC system under weak ambient radiation.

3) We establish an experimental W2A-OWC system under weak ambient radiation by adopting an infrared light-emitting diode (LED) to achieve covertness. After conducting experimental channel measurements, we train the proposed scheme and generate the transmitted signals. The experimental results under different wave intensities and position offsets indicate that the demodulation error rate of the proposed scheme is slightly higher than that of LBCs, while the JS divergence of the proposed scheme is significantly lower than that of LBCs.

The remainder of this paper is organized as follows. In Section II, we introduce the W2A-OWC system model. In Section III, we propose a covert scheme based on AAE, and develop the network optimization and training process. In Section IV, we numerically evaluate the proposed scheme on W2A channels, and test the proposed scheme under different receiver noise intensities, block lengths, code rates, and Willie's positions. We further compare the proposed scheme with LBCs as the benchmark. In Section V, we experimentally investigate the detection and covertness performance of the proposed scheme. Finally, we conclude this work in Section VI.

II. SYSTEM MODEL

We consider a W2A-OWC system as shown in Fig. 1. Under weak ambient radiation, an underwater transmitter (Alice) modulates the information bits to the emitted light intensity of an LED, and sends the modulated signal to a legitimate receiver (Bob) in the air, while an adversary receiver (Willie) tries to detect the communication behavior between Alice and Bob. Both Bob and Willie adopt an avalanche photodiode (APD) to receive signals, due to its higher sensitivity, speed, gain and larger detection dynamic range compared to standard PIN photodiode [31]. Given the positions and parameters of the transmitter and receiver, underwater scattering and absorption coefficients, and the slope angle of water surface,
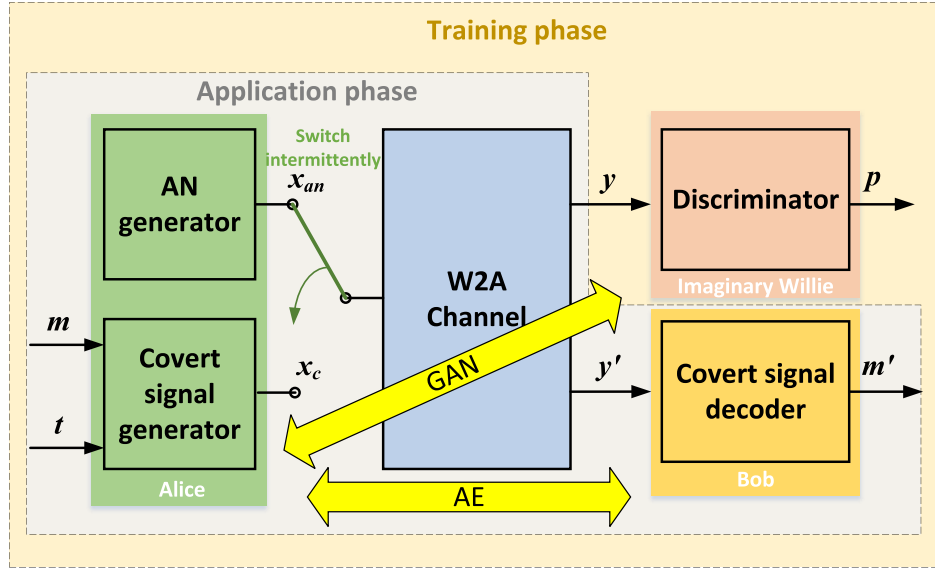
Fig. 2. Illustration of the covert scheme in a W2A-OWC system.

the W2A channel gain h can be calculated from the theoretical model of [4, eq. (20)]. In addition, the channel gain can be measured by processing experimental data in [3]. As a result of random fluctuation of the water surface, the channel gains vary randomly with time, which can be modeled as flat block fading [3], [32]. Considering the nonlinearity and noise of APDs, the received signals in electrical-domain are given by [30]

$$y = \frac{SM_A}{2M_0}\left(\sqrt{1 + \frac{4M_0{}^2\eta_A(h\eta_L x + b)}{S}} - 1\right) + M_A I_{sd} + n_r,$$

(1)

where x is the transmitted signal; h is the W2A channel gain; b is ambient light intensity; η_A is APD quantum efficiency; η_L is the electro-optical conversion efficiency of LED; M_0 is the initial multiplication factor of APDs; S is the designed parameter for achieving an expected gain; M_A is the amplification factor of the circuit; n_r is the additive white Gaussian noise (AWGN). Detailed parameters of the W2A-OWC system will be provided in the subsequent simulations of Section IV and experiments of Section V.

III. COVERT SCHEME DESIGN

A. Covert Signal Generation Based on an AAE

The covert communication scheme in a W2A-OWC system is shown in Fig. 2. Alice intermittently switches between transmitting covert signals and AN, and the received signals at Bob and Willie suffer random link gain variation and receiver noise.

In the covert scheme training phase, an imaginary Willie is introduced as a reference to the covert signal generation at Alice, and the imaginary Willie's position can be set according to a desired covert region. The AAE-based covert

scheme consists of two parts: a generative adversarial network (GAN) and an autoencoder (AE). The GAN consists of the covert signal generator at Alice and the discriminator at imaginary Willie to improve the signal's covertness. The AE consists of the covert signal generator at Alice and the decoder at Bob to achieve reliable demodulation at Bob. The training process iteratively enhances the Willie's capability of discriminating whether the received signals come from covert data or AN, Alice's capability of generating the covert signals, and Bob's capability of demodulating the received signal, establishing an adversarial interplay between them. In this way, the distribution of the generated signal becomes close to that of noise, as measured by the JS divergence. After the training process, the generated signal exhibits high levels of covertness and demodulation accuracy.

In the application phase, the trained network and parameters are deployed to Alice and Bob. The JS divergence between the received covert signals and AN remains sufficiently small for arbitrary actual Willie within the covert region, indicating that the proposed scheme can guarantee the covertness. It means that the actual Willie is not required to be in the same location or uses the same discriminator network as that in the training phase. The neural networks for Alice, Bob and Willie are shown in Fig. 3.

1) Network of Covert Signal Generator: Alice transforms a covert message of k bits to corresponding one-hot encoding representation $\boldsymbol{m}$. The one-hot code $\boldsymbol{m} = [m_1, m_2, \ldots, m_{2^k}]$ and random vector $\boldsymbol{t} = [t_1, t_2, \ldots, t_n]$ are sent into the covert signal generator network Υ_a to generate the covert signal vector $\boldsymbol{x_c} = [x_{c1}, x_{c2}, \ldots, x_{cn}]$, where n is the size of encoded block. Vector $\boldsymbol{t}$ is adopted to randomize the generation process and produce different $\boldsymbol{x_c}$ for the same message at different time. In Alice's network of covert signal generator, five dense layers with layer normalization and LeakyReLU activation function are utilized to achieve dimension transformation, feature extraction and information encoding. The network
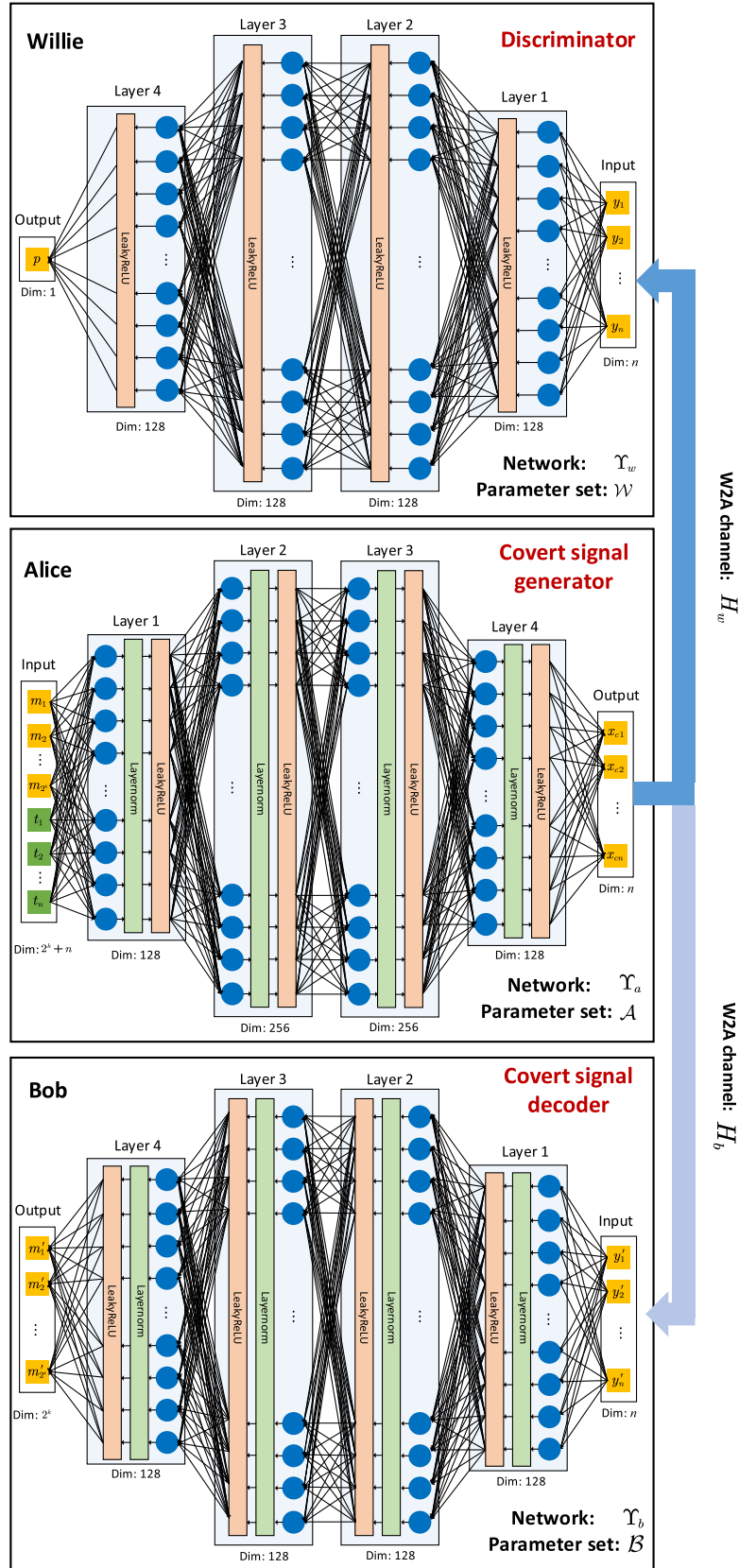
Fig. 3. Structure of neural networks for Alice, Bob and Willie.

topology as well as input and output dimensions of each layer are shown in the second subfigure of Fig. 3. AN vector x_{an} is produced by an AN generator to confuse Willie. The distribution of AN is same as that of receiver noise, such as Gaussian distribution. Alice intermittently switches between transmitting covert signals x_c and AN x_{an}.

TABLE I
SPACE AND TIME COMPLEXITY OF ALICE'S, BOB'S AND WILLIE'S NETWORKS

Network \ Complexity	Space : Number of (Layers, Weights, Biases)	Time : Number of FLOPs
Alice	$\left(5, \left(2^k + 2n\right) \times 128 + 131072, 768 + n\right)$	$2^{k+8} + 512n + 269832$
Bob	$\left(5, \left(2^k + n\right) \times 128 + 49152, 512 + 2^k\right)$	$2^{k+8} + 256n + 103432$
Willie	$\left(5, 128n + 49280, 513\right)$	$256n + 99584$

2) Network of Discriminator: After the signals passing through the W2A channel, Willie detects the received signal vector $\boldsymbol{y} = [y_1, y_2, \ldots, y_n]$ and outputs a scalar p which represents the confidence probability of $\boldsymbol{y}$ belonging to the samples of learned distribution (i.e., AN distribution). The discriminator network Υ_w at Willie is composed of five dense layers with LeakyReLU activation function, and the last layer transforms the feature space to a scalar. The network topology as well as input and output dimensions of each layer are shown in the first subfigure of Fig. 3.

3) Network of Covert Signal Decoder: The decoder network Υ_b at Bob is also composed of five dense layers with layer normalization and LeakyReLU activation function. The network topology as well as input and output dimensions of each layer are shown in the third subfigure of Fig. 3. The received covert signal vector $\boldsymbol{y}' = [y_1', y_2', \ldots, y_n']$ at Bob is demodulated by the decoder network to output the estimated code $\boldsymbol{m}' = [m_1', m_2', \ldots, m_{2^k}']$. Bob eventually predicts the covert message by performing a classification on $\boldsymbol{m}'$.

Since Bob is deployed on a resource-constrained UAV, the complexity of the proposed scheme needs to be taken into consideration. The space complexity and time complexity of Alice's, Bob's and Willie's networks are shown in Table I. The space complexity corresponds to the number of parameters in the network, including layers, weights, and biases. The time complexity can be characterized by the number of floating point operations (FLOPs). The number of FLOPs of a dense layer with input dimension I_{dim} and output dimension O_{dim} is $2I_{dim}O_{dim}$. The number of FLOPs of the layer normalization with dimension I_{dim} is $8I_{dim} + 2$. The number of FLOPs of a LeakyReLU activation function layer with dimension I_{dim} is $2I_{dim}$. The total number of FLOPs can be calculated by summing the number of FLOPs of each layer in the network. Some commercial embedded computing boards are available for machine learning (ML)/deep learning (DL) applications within UAVs. For example, the low-power "Nvidia Jetson" boards are designed for accelerating ML/DL applications, and its version "Jetson Nano" released in 2019 can run 4.72×10^{11} FLOPs per second under power 5 to 10 W [33]. In addition, the drone manufacturer DJI has launched Nvidia Tegra TK1-powered "Manifold" embedded computer for performing ML/DL tasks in 2015 [34]. In recent years, the DL algorithms for object detection [35], weed identification [36], and damage assessment [37] have been investigated for UAV-based systems. It is shown that the FLOP consumption of UAV-deployed DL networks in [35] and [37] is much larger than that of the proposed scheme in this paper, indicating the feasibility of the proposed scheme running on UAVs.

B. Optimization Algorithm for Covert Scheme

During the training process for each batch of size *N*, we develop the optimization algorithm to alternatively optimize Willie's network, Alice's network and Bob's network. The loss function and parameter update of each network are designed to achieve higher stability and faster convergence. The optimization algorithm for training the proposed scheme is summarized in Algorithm 1.

1) Optimization of Discriminator: When optimizing Willie's discriminator network Υ_w, we first transform a covert information bit set into a one-hot code set denoted as $\{\boldsymbol{m}^{(i)}\}_{i=1}^N$. We sample a random vector set $\{\boldsymbol{t}^{(i)}\}_{i=1}^N$ and an AN set $\{\boldsymbol{x_{an}}^{(i)}\}_{i=1}^N$ from a Gaussian distribution $\mathcal{N}(0, 1)$. After passing through the W2A channel between Alice and Willie (The channel mapping is denoted as H_w), the *i*-th covert signal arrived at Willie is denoted as $H_w(\Upsilon_a(\boldsymbol{m}^{(i)}, \boldsymbol{t}^{(i)}))$, and the *i*-th AN arrived at Willie is denoted as $H_w(\boldsymbol{x_{an}}^{(i)})$, where Υ_a is Alice's generator network. The outputs of Willie's discriminator network for the *i*-th covert signal and the *i*-th AN are $\Upsilon_w(H_w(\Upsilon_a(\boldsymbol{m}^{(i)}, \boldsymbol{t}^{(i)})))$ and $\Upsilon_w(H_w(\boldsymbol{x_{an}}^{(i)}))$, respectively. We adopt the loss function L_w of Wasserstein GAN with gradient penalty (WGAN-GP) in [38] as

$$L_w = \frac{1}{N} \sum_{i=1}^N \begin{bmatrix} \Upsilon_w\Big(H_w\Big(\Upsilon_a\Big(\boldsymbol{m}^{(i)}, \boldsymbol{t}^{(i)}\Big)\Big)\Big) - \\ \Upsilon_w\Big(H_w\Big(\boldsymbol{x_{an}}^{(i)}\Big)\Big) + \\ \lambda\Big(\Big\|\nabla_{\boldsymbol{c}^{(i)}} \Upsilon_w\Big(\boldsymbol{c}^{(i)}\Big)\Big\|_2 - 1\Big)^2 \end{bmatrix}, \quad (2)$$

where $\boldsymbol{c}^{(i)} = \epsilon^{(i)} H_w(\boldsymbol{x_{an}}^{(i)}) + (1 - \epsilon^{(i)}) H_w(\Upsilon_a(\boldsymbol{m}^{(i)}, \boldsymbol{t}^{(i)}))$, and $\epsilon^{(i)}$ follows a uniform distribution between 0 and 1, denoted as $U(0, 1)$. WGAN-GP is an improved version of GAN, which shows faster convergence and higher stability of optimization process. After the gradient backpropagation, parameter set $\mathcal{W} = \{w_1, w_2, \ldots\}$ of the discriminator network is updated with the Adam optimizer, where w_i is the *i*-th weight or bias in the network.

2) Optimization of Covert Signal Generator: When optimizing Alice's generator network Υ_a, we sample a random vector set $\{\boldsymbol{t}^{(i)}\}_{i=1}^N$ and send it to the covert signal generator along with $\{\boldsymbol{m}^{(i)}\}_{i=1}^N$. After covert signals pass through the W2A channel and arrive at receivers, the *i*-th output of Willie's discriminator is $p = \Upsilon_w(H_w(\Upsilon_a(\boldsymbol{m}^{(i)}, \boldsymbol{t}^{(i)})))$ and the *i*-th output of Bob's decoder network Υ_b is $\boldsymbol{m}'^{(i)} = \Upsilon_b(H_b(\Upsilon_a(\boldsymbol{m}^{(i)}, \boldsymbol{t}^{(i)})))$, where H_b is the W2A channel mapping between Alice and Bob. The loss function L_a of

Algorithm 1: Covert Scheme Training

Input: Batch size N; number of discriminator iterations per generator iteration n_{disc}; gradient penalty coefficient λ; cross entropy loss function f_{CEloss}; W2A channel mapping functions H_w and H_b; weights λ_b and λ_w; initial parameter set $\mathcal{W}$ of discriminator; initial parameter set $\mathcal{A}$ of generator; initial parameter set $\mathcal{B}$ of decoder; Alice's network Υ_a; Bob's network Υ_b; Willie's network Υ_w.

1 **while** $\mathcal{W}, \mathcal{A}, \mathcal{B}$ *have not converged* **do**
2   **for** $j = 1, \ldots, n_{disc}$ **do**
3   Input a one-hot code set $\left\{ \boldsymbol{m}^{(i)} \right\}_{i=1}^{N}$;
4   Sample a random vector set $\left\{ \boldsymbol{t}^{(i)} \right\}_{i=1}^{N} \sim \mathcal{N}(0,1)$;
5   Sample a AN set $\left\{ \boldsymbol{x_{an}}^{(i)} \right\}_{i=1}^{N} \sim \mathcal{N}(0,1)$;
6     $\boldsymbol{c}^{(i)} \leftarrow \epsilon^{(i)} H_w \left(\boldsymbol{x_{an}}^{(i)} \right) + \left(1 - \epsilon^{(i)} \right) H_w \left(\Upsilon_a \left(\boldsymbol{m}^{(i)}, \boldsymbol{t}^{(i)} \right) \right), \epsilon^{(i)} \sim U(0,1)$;
7     $L_w \leftarrow$
$$\frac{1}{N} \sum_{i=1}^{N} \begin{bmatrix} \Upsilon_w \left(H_w \left(\Upsilon_a \left(\boldsymbol{m}^{(i)}, \boldsymbol{t}^{(i)} \right) \right) \right) - \\ \Upsilon_w \left(H_w \left(\boldsymbol{x_{an}}^{(i)} \right) \right) + \\ \lambda \left(\left\| \nabla_{\boldsymbol{c}^{(i)}} \Upsilon_w \left(\boldsymbol{c}^{(i)} \right) \right\|_2 - 1 \right)^2 \end{bmatrix};$$
8     $\mathcal{W} \leftarrow \text{Adam}(\nabla_{\mathcal{W}} L_w, \mathcal{W})$;
9   **end**
10   Sample a random vector set $\left\{ \boldsymbol{t}^{(i)} \right\}_{i=1}^{N} \sim \mathcal{N}(0,1)$;
11   $L_a \leftarrow$
$$\frac{1}{N} \sum_{i=1}^{N} \begin{bmatrix} -\lambda_w \Upsilon_w \left(H_w \left(\Upsilon_a \left(\boldsymbol{m}^{(i)}, \boldsymbol{t}^{(i)} \right) \right) \right) + \\ \lambda_b f_{CEloss} \left(\boldsymbol{m}'^{(i)}, \boldsymbol{m}^{(i)} \right) \end{bmatrix};$$
12   $\mathcal{A} \leftarrow \text{Adam} \left(\nabla_{\mathcal{A}} L_a, \mathcal{A} \right)$;
13   Sample a random vector set $\left\{ \boldsymbol{t}^{(i)} \right\}_{i=1}^{N} \sim \mathcal{N}(0,1)$;
14   $L_b \leftarrow$
$$\frac{1}{N} \sum_{i=1}^{N} \left[f_{CEloss} \left(\Upsilon_b \left(H_b \left(\Upsilon_a \left(\boldsymbol{m}^{(i)}, \boldsymbol{t}^{(i)} \right) \right) \right), \boldsymbol{m}^{(i)} \right) \right];$$
15   $\mathcal{B} \leftarrow \text{Adam}(\nabla_{\mathcal{B}} L_b, \mathcal{B})$.
16 **end**

generator is derived from the weighted sum of the two terms as

$$L_a = \frac{1}{N} \sum_{i=1}^{N} \begin{bmatrix} -\lambda_w \Upsilon_w \left(H_w \left(\Upsilon_a \left(\boldsymbol{m}^{(i)}, \boldsymbol{t}^{(i)} \right) \right) \right) + \\ \lambda_b f_{CEloss} \left(\boldsymbol{m}'^{(i)}, \boldsymbol{m}^{(i)} \right) \end{bmatrix}, \quad (3)$$

where λ_w and λ_b are weights, and $f_{CEloss}(\boldsymbol{m}'^{(i)}, \boldsymbol{m}^{(i)}) = -\sum_{q=1}^{2^k} m_q^{(i)} \log(m_q'^{(i)})$ is the cross-entropy loss function. The first term in Eq. (3) is the loss function for generator making the distribution of covert signals close to that of AN, while the second term is the loss function for decoder improving the demodulation performance of covert signals. After the gradient backpropagation, parameter set $\mathcal{A} = \{a_1, a_2, \ldots\}$ of the generator network is updated with the Adam optimizer.

3) Optimization of Covert Signal Decoder: When optimizing Bob's decoder network Υ_b, we sample a random vector set $\{\boldsymbol{t}^{(i)}\}_{i=1}^{N}$ and send it to the covert signal generator along with $\{\boldsymbol{m}^{(i)}\}_{i=1}^{N}$. The i-th output of Bob's decoder network Υ_b is $\boldsymbol{m}'^{(i)} = \Upsilon_b(H_b(\Upsilon_a(\boldsymbol{m}^{(i)}, \boldsymbol{t}^{(i)})))$. The loss function L_b of Bob's decoder network is calculated as

$$L_b = \frac{1}{N} \sum_{i=1}^{N} \left[f_{CEloss} \left(\boldsymbol{m}'^{(i)}, \boldsymbol{m}^{(i)} \right) \right]. \quad (4)$$

Parameter set $\mathcal{B} = \{b_1, b_2, \ldots\}$ of the decoder network is updated with the Adam optimizer after gradient backpropagation.

According to [39], the optimization of discriminator determines the accuracy of signal generation in training GAN. Therefore, the discriminator is updated n_{disc} times, while the covert signal generator and decoder are updated once for each batch. The covert scheme is trained for multiple batches and epochs until the parameters converge.

IV. NUMERICAL RESULTS

k message bits are encoded to a block of size n, denoted as (n, k). We adopt a training set $\{\boldsymbol{m}^{(i)}\}_{i=1}^{10240}$ of one-hot code of size 10240 with 10 batches to train the proposed scheme with a learning rate of 2×10^{-4}, and a validation set of one-hot code with size 10240 to calculate Bob's BLER and JS divergence at each epoch. In our proposed covert scheme, the transmission of covert signals and AN is intermittently switched, which effectively embeds the covert signal within noise. Note that excessively large covert signal block length can increase the difference in the distributions of the covert signal and noise, since it has been demonstrated that Kullback-Leibler (KL) divergence of sequence distributions grows linearly with the sequence length [40]. Therefore, the covert signals with short-to-medium block length are considered, and BLER is introduced as the performance metric in finite block length regime, which has been adopted in [41] and [42] to measure the communication reliability between Alice and Bob in AE.

As shown in Fig. 2, the information bits are structured in blocks of message, and the direct input and output of the autoencoder network are message blocks. BLER is defined as the average probability that the decoded message m' at Bob is not identical to the original message m transmitted by Alice. Consequently, BLER serves as a direct and interpretable metric for network optimization in terms of the overall accuracy of the transmission, which is consistent with how the AE works. JS divergence is adopted to measure the difference between the distributions of covert signal and AN, since minimizing the JS divergence has been proven to be the optimization objective for the generator in GAN [39].

Assume that Alice and Bob locate at $(0, 0, -2)$ m and $(0, 0, 1)$ m, respectively. The position of imaginary Willie is set to be the same as that of Bob to ensure the signal covertness at Bob after training. Assume that underwater absorption and scattering coefficients are 0.04 /m and 0.06 /m, respectively. Assume that wave slope follows truncated logistic distribution with parameters $u_w = 0$ and $\sigma_w = 5$ [4]. Assume that a collimated LED with $\eta_L = 1$ is adopted and the receiving aperture of APDs is 1 mm. The distribution of W2A channel
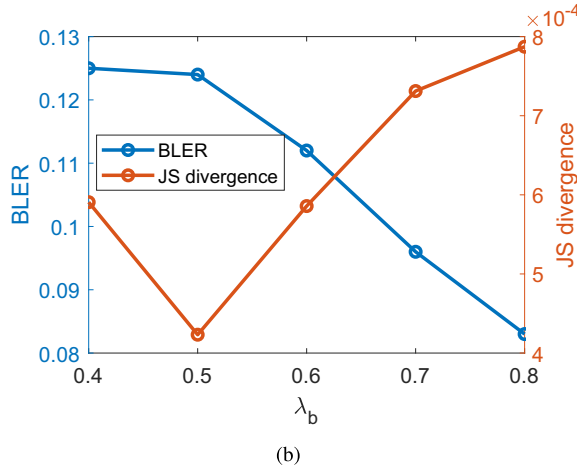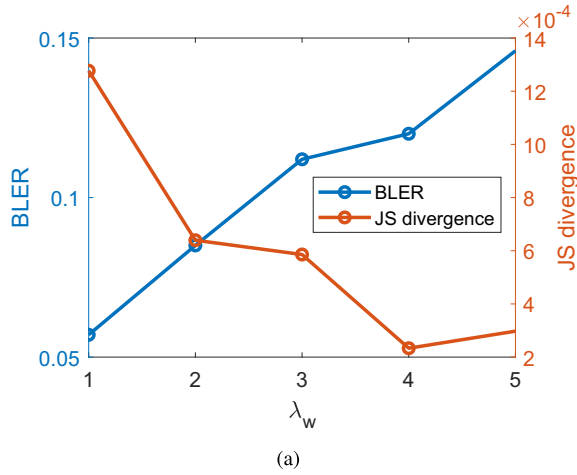
Fig. 4.   BLER and JS divergence on the validation set versus (a) λ_w ($\lambda_b = 0.6$) and (b) λ_b ($\lambda_w = 3$).

Fig. 5.   (a) BLER and (b) JS divergence on the validation set versus epoch under W2A channel.

gain can be calculated from the theoretical model in [4] following lognormal distribution with parameters $u = -13.15$ and $\sigma = 0.33$. The APD parameters are set to typical values in [30], i.e., $S = 1.1538$ A, $M_0 = 100$, $\eta_A = 1$, $I_{sd} = 10^{-9}$ A. The transmitted signal power in electrical-domain is 1 W, the amplification factor of the circuit $M_A = 1250$, and the ambient light intensity $b = 0$ W. In this case, the noise variance is mainly dominated by thermal noise due to the low incident light power, and different noise variances are adopted to represent different channel conditions in the subsequent parts.

To determine the values of λ_w and λ_b, we compare the covertness performance and demodulation performance under different weights after training 2000 epoches with $(n, k) = (8, 4)$, as shown in Fig. 4. In Fig. 4(a), the BLER increases with λ_w, while the JS divergence decreases with λ_w. It indicates that a larger λ_w leads to poorer demodulation performance and better covertness performance. Similarly, it is seen from Fig. 4(b) that a larger λ_b leads to better demodulation performance and poorer covertness performance. Therefore, we set $\lambda_w = 3$ and $\lambda_b = 0.6$ to balance the demodulation and covertness performance.
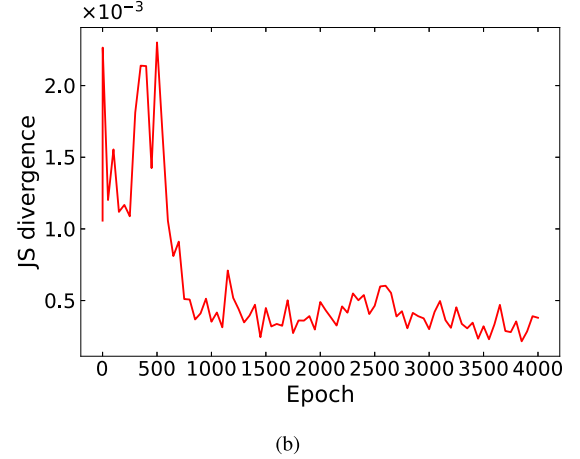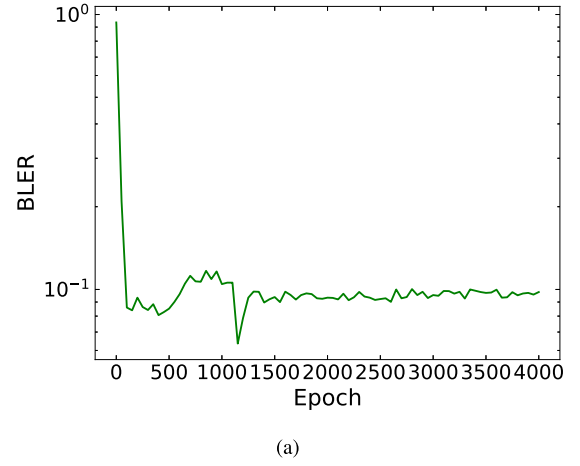
A. Validation Set

We set $(n, k) = (8, 4)$ and receiver noise power of -15 dB (relative to the signal power of 1).

In Fig. 5(a), Bob's BLER decreases from 0.9360 to 0.0980 as the epoch increases from 0 to 4000. The JS divergence between the covert signals and AN decreases from 1.058×10^{-3} to 3.800×10^{-4}, as shown in Fig. 5(b). It indicates that Bob's demodulation performance and signal's covertness performance are gradually enhanced with the training process.

Figure 6 shows the scatter diagrams on the first two elements $y[0]$ and $y[1]$ in received vector $\boldsymbol{y}$ at epoch 0 and 4000. It is seen that the scatter distributions of covert signals and AN are obviously different before training, as shown in Fig. 6(a). After training as shown in Fig. 6(b), the covert signals are sufficiently divergent and close to the AN.

In Fig. 7, the t-SNE (t-distributed Stochastic Neighbor Embedding) is adopted to visualize high-dimensional signals into a two-dimensional space. It is seen from Fig. 7(a) that the covert signals and AN are separated from each other at epoch 0. At epoch 4000, the distributions of covert signals and AN are sufficiently close and coincident, as shown in Fig. 7(b).

To verify the scheme effectiveness in training the signal covertness, we only train the Alice and Bob's AE network without discriminator network. It is seen from Fig. 8(a) that
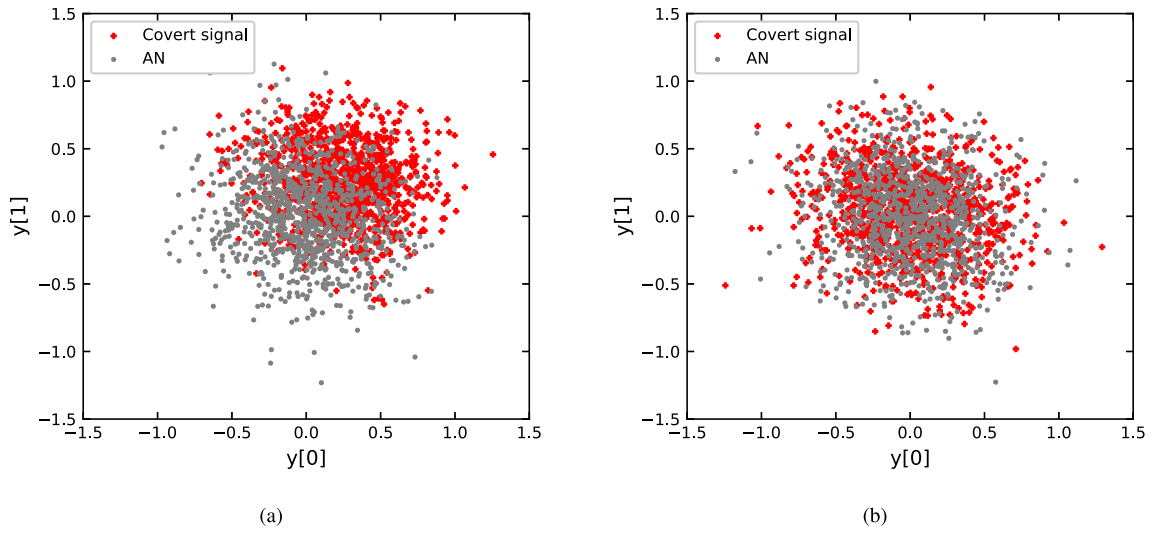
Fig. 6. Scatter diagrams on $\{y[0], y[1]\}$ for received covert signals and AN at (a) epoch 0 and (b) epoch 4000 under W2A channel.
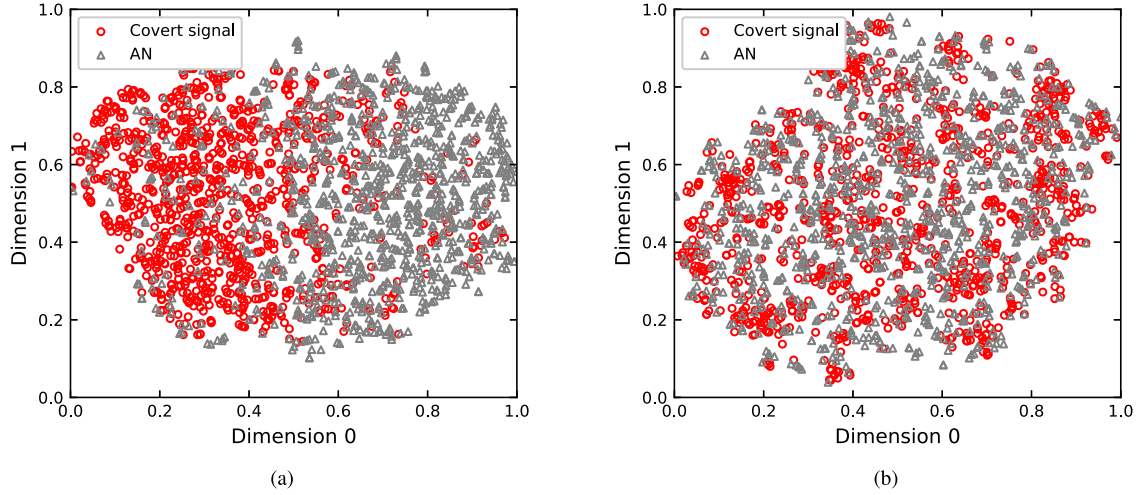

Fig. 7. 2-dimensional T-SNE diagrams for received covert signals and AN at (a) epoch 0 and (b) epoch 4000 under W2A channel.
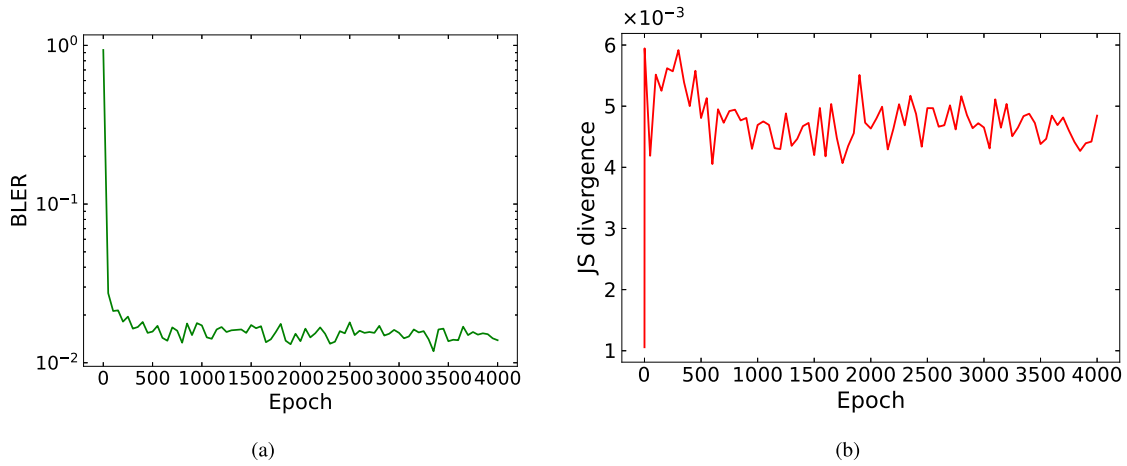

Fig. 8. (a) BLER and (b) JS divergence on the validation set versus epoch under W2A channel without discriminator.

Bob's BLER decreases to 0.0140 as the epoch increases to 4000, which is lower than that in the proposed scheme with discriminator network. In Fig. 8(b), the JS divergence increases to 4.845×10^{-3}, which is 12.75 times that in the proposed scheme with discriminator network. Since the network without discriminator only tends to improve
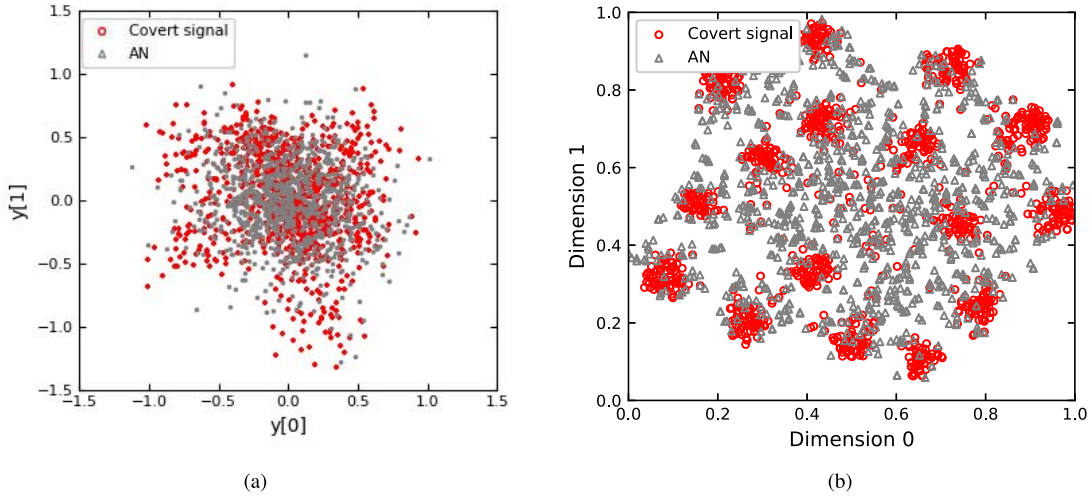
Fig. 9. (a) Scatter diagrams on $\{y[0], y[1]\}$ and (b) 2-dimensional T-SNE diagrams for received covert signals and AN at epoch 4000 under W2A channel without discriminator.
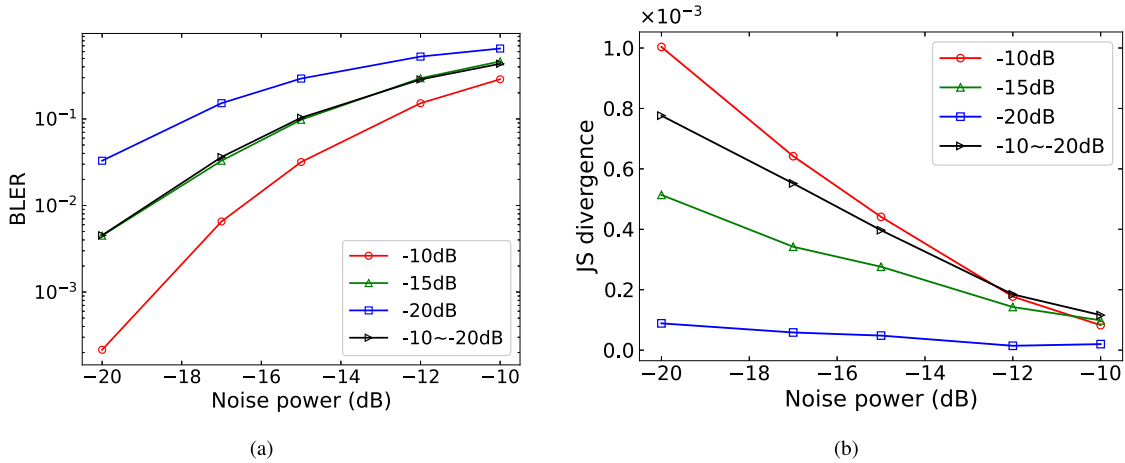

Fig. 10. (a) BLER and (b) JS divergence on the test set versus receiver noise power under different receiver noise powers during training process.

Bob's demodulation performance, leading to poor covertness performance of signal.

At epoch 4000, it is seen from Fig. 9(a) that the covert signals on $y[0]$ and $y[1]$ axes exhibit irregular divergence, which is different from the distribution of AN. The t-SNE diagram in Fig. 9(b) shows that the covert signals are composed of 16 separate clusters. Compared with the proposed scheme with discriminator network, the demodulation performance improves while covertness performance deteriorates significantly in AE network without discriminator network. It indicates that there exists a trade-off between the covertness performance and demodulation performance.

B. Test Set

A test set of one-hot code with size 307200 is adopted with $(n, k) = (8, 4)$. We train the proposed scheme under receiver noise powers of -10 dB, -15 dB, -20 dB, and random values in $[-10, -20]$ dB, respectively. Figs. 10(a) and 10(b) show the BLER and JS divergence on the test set versus receiver noise power under different receiver noise

powers during training process. It is seen that the scheme trained under lager receiver noise power yields lower BLER and larger JS divergence on the test set, since the covertness performance of received signals is relatively high due to the large receiver noise, and the training process tends to improve the demodulation performance. The model trained under random receiver noise power in $[-10, -20]$ dB shows a similar BLER as that trained under -15 dB, and the JS divergence is higher than that trained under -15 dB.

Then, we train the scheme of different block lengths (code rate is fixed at 0.5) under receiver noise power of -15 dB. Figs. 11(a) and 11(b) show the BLER and JS divergence on the test set versus receiver noise power under different block lengths. The number of total transmitted information bits is fixed at 1228800, corresponding to test sets of size 307200 with $(n, k) = (8, 4)$, 245760 with $(n, k) = (10, 5)$, 204800 with $(n, k) = (12, 6)$, respectively. The BLERs under block length from 8 to 12 are close to each other due to the same code rate. The JS divergence shows larger difference for receiver noise power lower than -15 dB, since the schemes after training show great difference in covertness performance
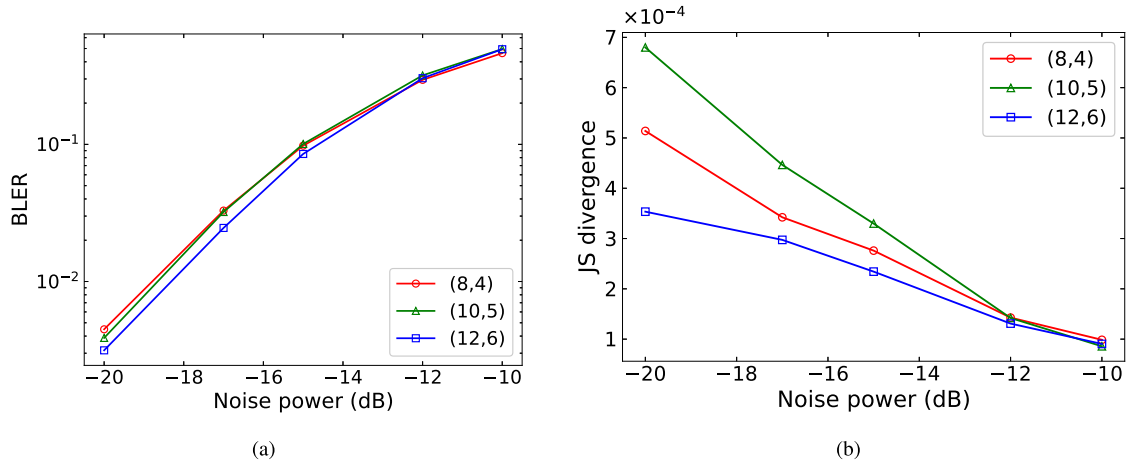
Fig. 11. (a) BLER and (b) JS divergence on the test set versus receiver noise power under different block lengths.
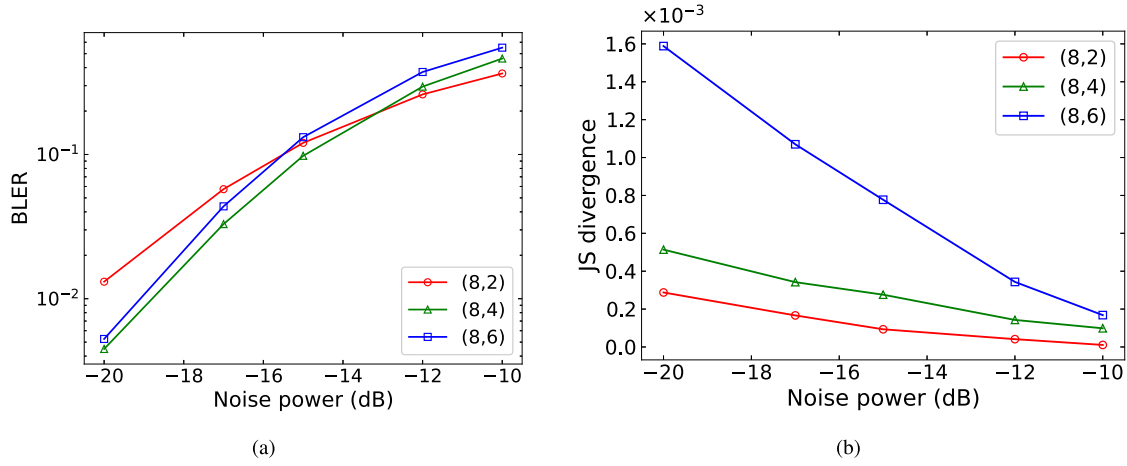

Fig. 12. (a) BLER and (b) JS divergence on the test set versus receiver noise power under different codes rates.

under low noise power. In our scheme, the covert signal generator is pitted against the discriminator. Thus, the JS divergence and block lengths are not be linearly related.

Moreover, we train the scheme of different code rates (block length is fixed at 8) under receiver noise power of -15 dB. Figs. 12(a) and 12(b) show the BLER and JS divergence on the test set versus receiver noise power under different codes rates, where the length of transmitted information bits is fixed at 1228800. It is seen that larger code rate leads to higher BLER at receiver noise power higher than -12 dB. Larger code rate also yields larger JS divergence under different receiver noise powers, since an information block with more information bits is more different from a noise block.

Finally, we investigate the covertness performance of the trained scheme when actual Willie is located in different positions. For ease of illustration in the figure, we fix Willie's y-coordinate at 0 m and vary the x-coordinate and z-coordinate. In the location variation, the x-coordinate cannot be 0 since Willie cannot block Bob's light path to avoid exposing his eavesdropping behavior, as shown in Fig. 13. The model is trained under receiver noise power of -15 dB and imaginary Willie's position (0, 0, 1) m same as Bob's. A test set of one-hot code with size 307200 is adopted
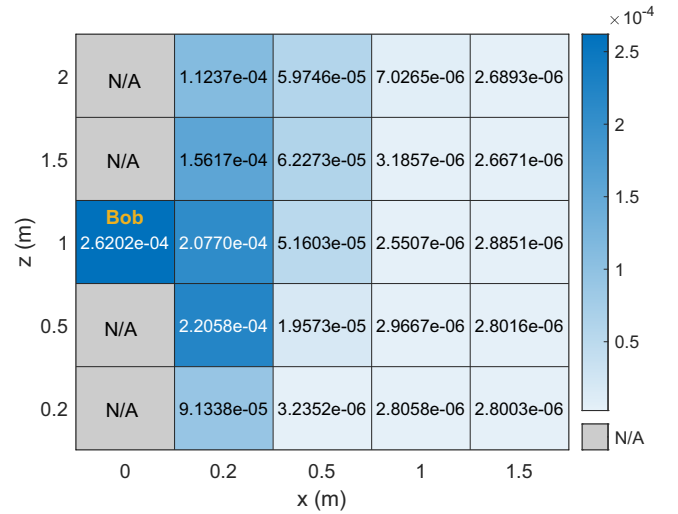

Fig. 13. JS divergence on the test set versus Willie's position.

with $(n, k) = (8, 4)$. It is seen from Fig. 13 that the signal covertness holds even when the actual Willie is closer to Alice than Bob at (0.2, 0, 0.2) m. In the practical application of the covert scheme, due to the unknown position of Willie, a covert

TABLE II
BLER AND JS DIVERGENCE AMONG DIFFERENT (8, 4) LBCS

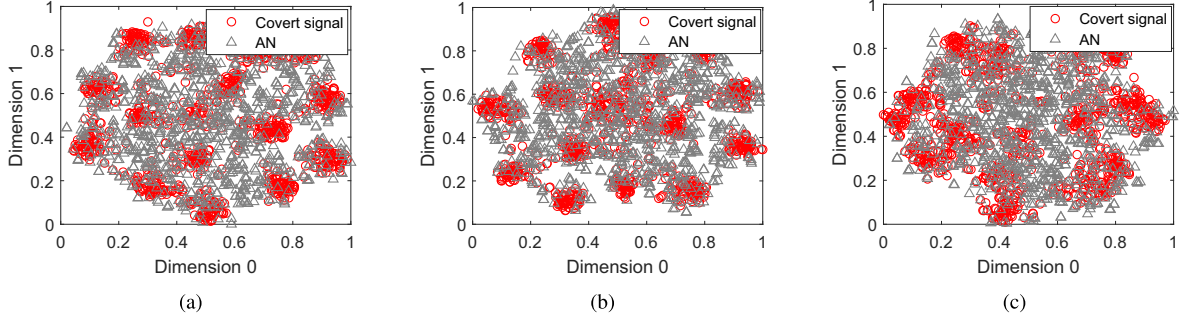| Performance \ Scheme	G_1	G_2	G_3	Proposed scheme
$\mathrm{BLER_{hard}}$	0.1370	0.1630	0.4070	0.0979
$\mathrm{BLER_{soft\&accurate}}$	0.0180	0.0128	0.0828	
$\mathrm{BLER_{soft\&blind}}$	0.0912	0.1340	0.2950	
JS divergence	1.490×10^{-2}	2.031×10^{-2}	3.403×10^{-3}	2.761×10^{-4}

Fig. 14. T-SNE diagrams with generator matrices (a) G_1, (b) G_2 and (c) G_3.

region can be set up to ensure that Willie cannot effectively detect within this area.

C. Comparison With LBC

We further compare the proposed scheme with LBC in demodulation performance and covertness performance. We traverse all generator matrices of $(n, k) = (8, 4)$ LBCs, and calculate Bob's BLER and JS divergence of each coded sequence under W2A channel with receiver noise power of -15 dB. The hard decision at Bob is performed with syndrome decoding. The soft decision with accurate channel information is the maximum-likelihood decoding with perfect channel fading coefficients and APD parameters, which is an upper bound on the demodulation performance. Since the transmitted signals of the proposed scheme do not contain the pilot signals, a pilot-free channel estimation algorithm based on Lloyd-Max algorithm in [43] is adopted for performance comparison. The soft decision with pilot-free channel estimation corresponds to the maximum-likelihood decoding after performing pilot-free channel estimation.

Table II shows the minimum BLER after hard decision or soft decision, as well as the minimum JS divergence among different (8,4) LBCs. The minimum BLER after hard decision, denoted as $\mathrm{BLER_{hard}}$, is 0.1370, and the corresponding generator matrix is

$$G_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}. \tag{5}$$

For data generated by G_1, the BLER after soft decision with accurate channel information, denoted as $\mathrm{BLER_{soft\&accurate}}$, is 0.0180, the BLER after soft decision with pilot-free channel

estimation, denoted as $\mathrm{BLER_{soft\&pilot-free}}$, is 0.0912, and the JS divergence is 1.490×10^{-2}.

The minimum $\mathrm{BLER_{soft\&accurate}}$ is 0.0128, and the corresponding generator matrix is

$$G_2 = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 \end{bmatrix}. \tag{6}$$

For data generated by G_2, $\mathrm{BLER_{hard}}$ is 0.1630, $\mathrm{BLER_{soft\&pilot-free}}$ is 0.1340, and the JS divergence is 2.031×10^{-2}.

The minimum JS divergence is 3.403×10^{-3}, and the corresponding generator matrix is

$$G_3 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{7}$$

For data generated by G_3, $\mathrm{BLER_{hard}}$ is 0.4070, $\mathrm{BLER_{soft\&accurate}}$ is 0.0828, and $\mathrm{BLER_{soft\&pilot-free}}$ is 0.2950.

It is seen that the minimum $\mathrm{BLER_{hard}}$ and the minimum JS divergence among all LBCs are larger than those (BLER: 0.0979, JS divergence: 2.761×10^{-4}) on the test set in the proposed scheme, while $\mathrm{BLER_{soft\&accurate}}$ is lower since the prior information including channel fading and APD parameters are known to Bob. The minimum $\mathrm{BLER_{soft\&pilot-free}}$ among G_1, G_2, and G_3 is close to that in the proposed scheme. The t-SNE diagrams in Fig. 14 show that the distributions of received signals coded by G_1, G_2, and G_3 are obviously different from those of AN, which indicates that LBCs can not achieve good signal covertness.

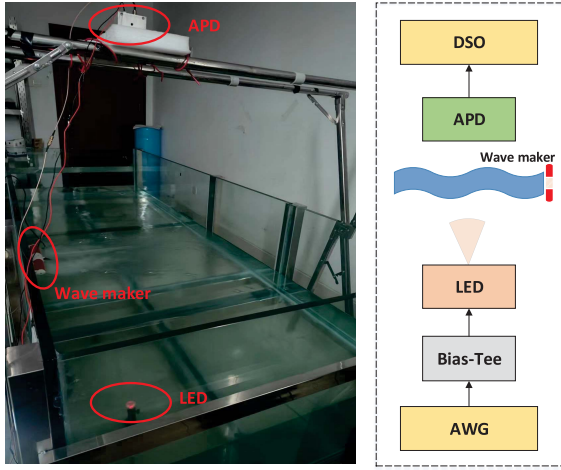Further investigation shows that the bit error rate (BER) of the proposed scheme is 2.61×10^{-2}, comparable to that of the

Fig. 15. The experimental system of W2A-OWC under weak ambient radiation.

Fig. 16. The PDF of the channel gain under the wave of index 3.

optimum linear block code without pilot signals 4.46×10^{-2}. Such BER can be corrected by a length-10000 LDPC code of rate 0.5. To further reduce the BLER of the proposed scheme under low signal-to-noise ratio (SNR) conditions, we can increase the transmission signal power or lower the code rate.

V. EXPERIMENTAL RESULTS

To verify the effectiveness of the proposed covert scheme, a W2A-OWC system under weak ambient radiation is established as shown in Fig. 15. The ambient light power measured by an optical power meter (Thorlabs, PM100D with sensor S130C) at central wavelength 520 nm is $0.112 \ \text{W/m}^2$. An infrared LED (RUIBAO 3535, 785~790 nm) is placed under depth 0.30 m as the transmitter with coordinate $(0, 0, -0.30)$ m. An APD (Hamamatsu, S2385) is placed directly above the transmitter at $(0, 0, 0.85)$ m as the receiver. A wave generator (Yujang CX-W3) is placed in a water tank to generate wave. The wave intensity can be adjusted by switching the power mode of the wave generator, where the wave intensity increases with the wave type index from 1 to 5. The transmitted signal from an arbitrary waveform generator (AWG, keysight 33600A) and the direct current (DC) bias (DC = 2.2 V) from a DC power supply (Rigol DP832A) are used to drive the LED with a Bias-Tee. After passing through the W2A channel, the received signal is sampled by a digital storage oscilloscope (DSO, Agilent MSO-X 6004A) with sampling rate 20 MSa/s for offline signal processing.

Firstly, the channel parameters are measured to train the proposed covert scheme. A sinusoidal wave (Vpp = 3 V) of frequency 500 kHz is transmitted to obtain the channel gain. The channel can be modeled as a flat block fading, since the temporal dispersion caused by scattering can be neglected if the transmission rate is lower than Gbps [3], [32]. We extract the gain from the amplitudes of received signals and transmitted signals, and further obtain the statistical distribution of channel gain under the wave of index 3 produced by the wave generator, as shown in Fig. 16. It is seen that the data can
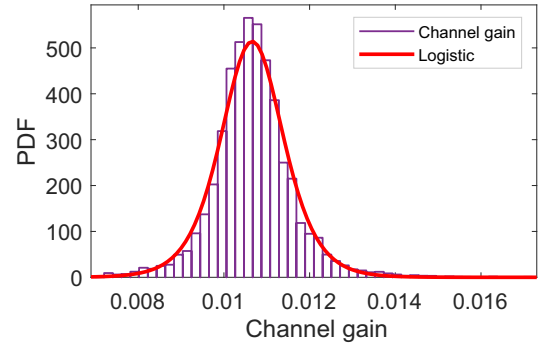
be well fitted by a logistic distribution with parameters $\mu = 1.07 \times 10^{-2}$ and $\sigma = 4.86 \times 10^{-4}$ by adopting the Distribution Fitter in MATLAB, i.e.,

$$f(x) = \frac{\exp\left(\frac{x-u}{\sigma}\right)}{\sigma \left[1 + \exp\left(\frac{x-u}{\sigma}\right)\right]^2}, \tag{8}$$

In addition, the receiver noise variance is calculated to be $\sigma_n^2 = 3.20 \times 10^{-7}$ without signal transmission. These channel parameters are adopted to train the covert scheme with a training set of size 10240 and a validation set of size 10240. We set $(n, k) = (8, 4)$. Figure 17 shows that Bob's BLER on validation set decreases from 0.9330 to 0.0160 as the epoch increases from 0 to 4000, while the JS divergence between the covert signals and AN decreases from 2.189×10^{-3} to 3.733×10^{-4}.

After training the proposed scheme, we generate a test set of size 10240 and send it into the covert signal generator to obtain the transmitted signals. In addition, AN is generated by the Gaussian noise generator. After pulse shaping by a root-raised-cosine filter (RRC) with roll-off factor 0.25 and upsampling factor 10, the transmitted signals are loaded into the AWG with Vpp = 3 V and frequency 5 MHz. At the receiver, the electrical signals sampled by DSO are filtered by an RRC and further sent to the covert signal decoder. Under the same wave of index 3 as the training process, Bob's BLER on the test set is 0.0540 and JS divergence is 5.005×10^{-4}. The demodulation and covertness performance on the test set degrades slightly compared with the results in the training process, due to the difference of the water surface fluctuation and the measurement imperfection of channel parameters.

For comparison, we traverse all generator matrices of $(n, k) = (8, 4)$ LBCs on the channel model measured in the experiment, and demodulate the codes by hard decision or soft decision with pilot-free channel estimation. The sequences coded by generator matrices G_{BLERmin} and G_{JSmin} corresponding to the minimum BLER and minimum JS divergence are adopted for transmission on the experimental link, respectively. The transmission process and parameters are the same as those of the proposed scheme above. Under the same wave of index 3, Bob's BLER for sequence coded by G_{BLERmin} is 0.0223 after soft decision with pilot-free channel estimation, and the JS divergence between the sequence coded by G_{JSmin} and AN is 3.620×10^{-2}. The results indicate that the demodulation performance of LBC is in the same order as that
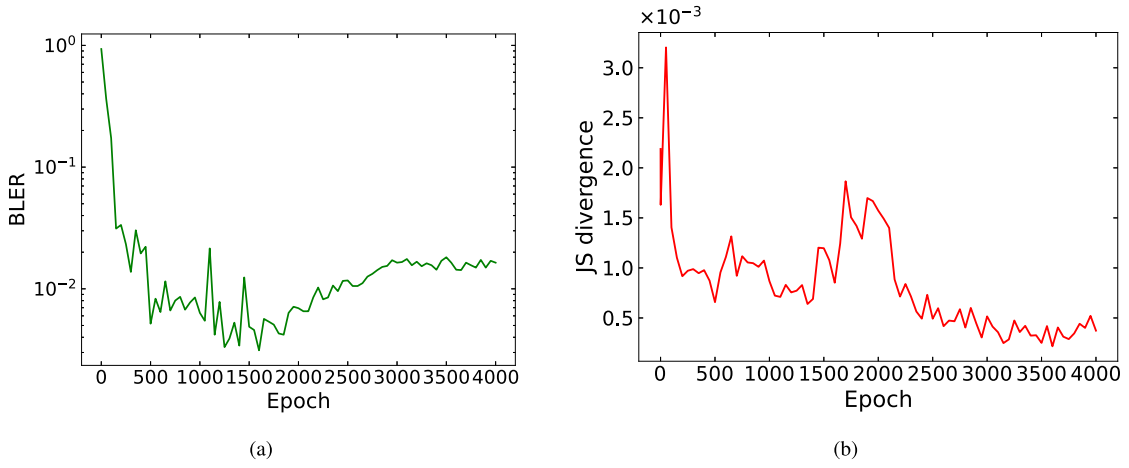
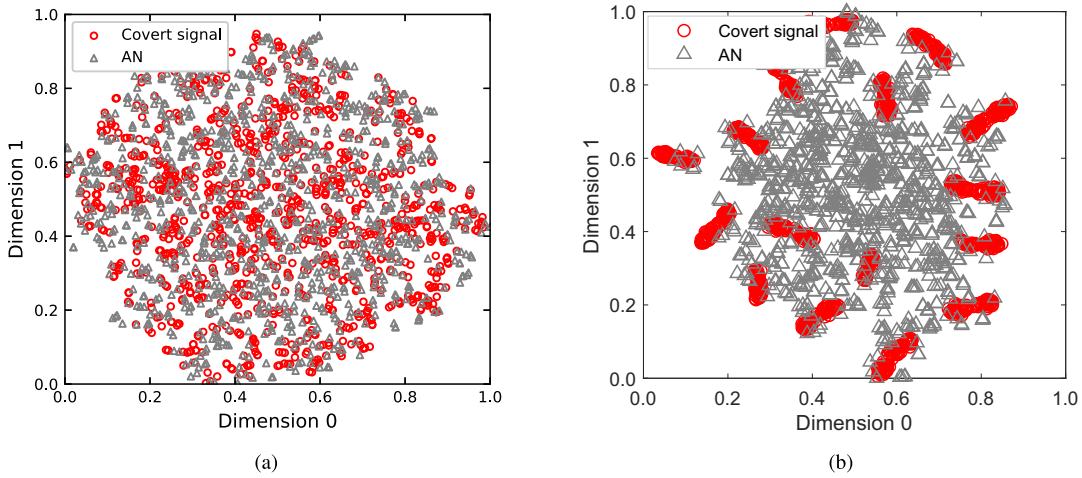Fig. 17. (a) BLER and (b) JS divergence on the validation set versus epoch under experimental channel.

Fig. 18. 2-dimensional T-SNE diagrams for received covert signals and AN of (a) proposed scheme and (b) LBC sequences coded by G_{JSmin}.
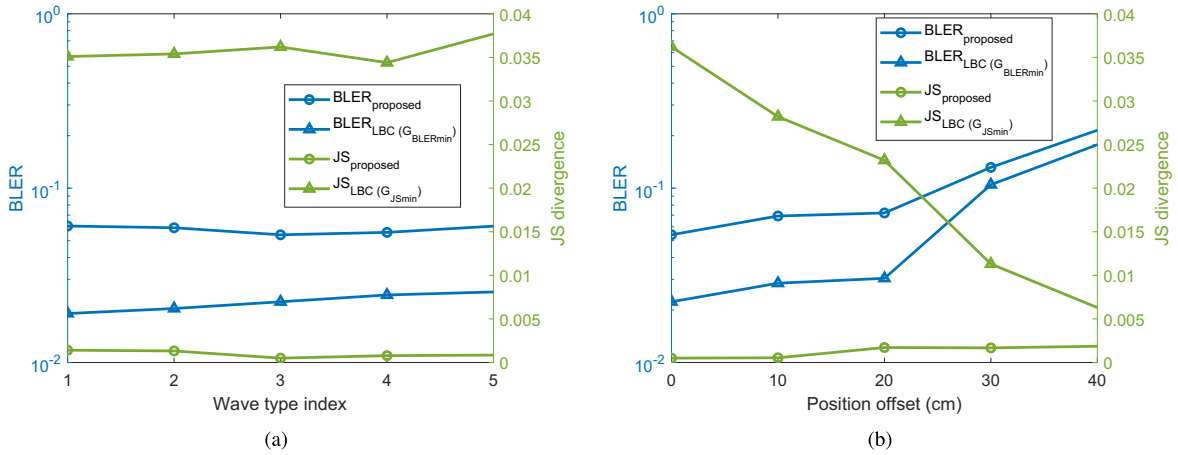

Fig. 19. The BLER and JS divergence (a) under different wave types and (b) different position offsets.

of the proposed scheme, while the covertness performance of LBC is much worse than that of the proposed scheme.

Figure 18(a) shows the distribution of received signals generated by proposed scheme is sufficiently close to that of AN in the t-SNE diagram, while the LBC sequences coded by G_{JSmin} show 16 separate clusters in Fig. 18(b).

To investigate the scheme robustness to the wave intensity and position offset, we calculate the BLER and JS divergence under different wave types (index from 1 to 5) and position offsets, as shown in Fig. 19. The wave intensity increases with the wave type index. It is seen from Fig. 19(a) that among all wave types, the BLER of proposed scheme and

LBC sequences coded by G_{BLERmin} are around 0.0500 and 0.0200, respectively. Due to imperfect channel modeling during training process, the demodulation performance of proposed scheme is slightly worse than that of LBC. However, the JS divergence of LBC sequences coded by G_{JSmin} is about 70 times that of the proposed scheme.

In Fig. 19(b), as the horizontal offset of the receiver increases from 0 cm to 40 cm, the BLER of proposed scheme increases from 0.0540 to 0.2149, and the BLER of LBC sequences coded by G_{BLERmin} increases from 0.0223 to 0.1776, since lager position offset leads to weaker signal power and further larger BLER. The JS divergence of LBC sequences coded by G_{JSmin} decreases from 3.620×10^{-2} to 6.300×10^{-3}, since the proportion of noise in the received signal increases with the horizontal offset, thus reducing the JS divergence. The JS divergence of proposed scheme changes slightly around 1.000×10^{-3}.

In the experiment, the BLER of the proposed scheme is slightly higher than that of LBCs, while the JS divergence of the proposed scheme is significantly lower than that of LBCs. The primary goal in covert communication is to achieve signal covertness, so slightly sacrificing the demodulation performance is worthwhile.

VI. CONCLUSION

In this work, we have proposed a covert scheme for W2A-OWC systems based on an AAE to generate covert signals without manual feature extraction for information hiding. The neural network structure and related optimization have been explored to achieve signal distribution close to AN distribution. The visualization results in two-dimensional space through t-SNE show similar distributions of covert signals and AN after training under W2A channels. Furthermore, the proposed scheme has been tested under different receiver noise intensities, block lengths, code rates, and Willie's positions. The comparison results with LBC indicate that the proposed scheme shows lower BLER and JS divergence compared with those of LBCs under hard decision, as well as similar BLER and lower JS divergence compared with those of LBCs under soft decoding with pilot-free channel estimation. In the experiments, the proposed AAE-based scheme shows slightly higher BLER but significantly lower JS divergence, compared with LBCs under soft decoding with pilot-free channel estimation.

REFERENCES

[1] L.-K. Chen, Y. Shao, and Y. Di, "Underwater and water-air optical wireless communication," *IEEE J. Lightw. Technol.*, vol. 40, no. 5, pp. 1440–1452, Mar. 1, 2022.

[2] H. Luo, J. Wang, F. Bu, R. Ruby, K. Wu, and Z. Guo, "Recent progress of air/water cross-boundary communications for underwater sensor networks: A review," *IEEE Sensors J.*, vol. 22, no. 9, pp. 8360–8382, May 2022.

[3] T. Lin, N. Huang, C. Gong, J. Luo, and Z. Xu, "Preliminary characterization of coverage for water-to-air visible light communication through wavy water surface," *IEEE Photon. J.*, vol. 13, no. 1, pp. 1–13, Feb. 2021.

[4] Q. Hu, C. Gong, T. Lin, J. Luo, and Z. Xu, "Secrecy performance analysis for water-to-air visible light communication," *IEEE J. Lightw. Technol.*, vol. 40, no. 14, pp. 4607–4620, Jul. 15, 2022.

[5] X. Chen et al., "Covert communications: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 25, no. 2, pp. 1173–1198, 2nd Quart., 2023.

[6] R. Ma, W. Yang, X. Guan, X. Lu, Y. Song, and D. Chen, "Covert mmWave communications with finite blocklength against spatially random wardens," *IEEE Internet Things J.*, vol. 11, no. 2, pp. 3402–3416, Feb. 2024.

[7] H. Zhu, H. Wu, and X. Jiang, "Jamming with zero-forcing beamforming for covert communication in MIMO systems," in *Proc. Int. Conf. Netw. Netw. Appl.*, 2023, pp. 122–126.

[8] Y. Zhang, Y. Zhang, J. Wang, S. Xiao, W. Ni, and W. Tang, "Robust beamfocusing for FDA-aided near-field covert communications with uncertain location," in *Proc. IEEE Int. Conf. Commun. Workshops*, 2023, pp. 1349–1354.

[9] Z. Duan, X. Yang, Y. Gong, D. Wang, and L. Wang, "Covert communication in uplink NOMA systems under channel distribution information uncertainty," *IEEE Commun. Lett.*, vol. 27, no. 5, pp. 1282–1286, May 2023.

[10] B. Kang, N. Ye, and B. Qi, "Comparisons on covert performances of NOMA in satellite Internet of Things," in *Proc. Comput., Commun. IoT Appl.*, 2021, pp. 318–322.

[11] L. Huang, J. Lei, and Y. Huang, "Spatial modulation covert communication assisted by artificial noise," in *Proc. IEEE/CIC Int. Conf. Commun. China*, 2023, pp. 1–6.

[12] L. Zhang, G. Li, J. Chen, H. Wang, W. He, and J. Yi, "Joint transmit power and trajectory optimization for UAV covert communication assisted by artificial noise," in *Proc. Int. Conf. Ubiquitous Commun.*, 2023, pp. 46–51.

[13] W. He et al., "Optimal transmission probabilities of information and artificial noise in covert communications," *IEEE Commun. Lett.*, vol. 26, no. 12, pp. 2865–2869, Dec. 2022.

[14] W. Xiong, Y. Yao, X. Fu, and S. Li, "Covert communication with cognitive jammer," *IEEE Wireless Commun. Lett.*, vol. 9, no. 10, pp. 1753–1757, Oct. 2020.

[15] B. Che, C. Gao, R. Ma, X. Zheng, and W. Yang, "Covert wireless communication in multichannel systems," *IEEE Wireless Commun. Lett.*, vol. 11, no. 9, pp. 1790–1794, Sep. 2022.

[16] S. Yan, S. V. Hanly, and I. B. Collings, "Optimal transmit power and flying location for UAV covert wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 11, pp. 3321–3333, Nov. 2021.

[17] X. Jiang, Z. Yang, N. Zhao, Y. Chen, Z. Ding, and X. Wang, "Resource allocation and trajectory optimization for UAV-enabled multi-user covert communications," *IEEE Trans. Veh. Technol.*, vol. 70, no. 2, pp. 1989–1994, Feb. 2021.

[18] R. He et al., "When the warden does not know transmit power: Detection performance analysis and covertness strategy design," *IEEE Trans. Commun.*, vol. 72, no. 6, pp. 3184–3198, Jun. 2024.

[19] J. Si et al., "Covert transmission assisted by intelligent reflecting surface," *IEEE Trans. Commun.*, vol. 69, no. 8, pp. 5394–5408, Aug. 2021.

[20] J. Liu, J. Yu, R. Zhang, S. Wang, K. Yang, and J. An, "Intelligent reflecting surface-aided covert ambient backscatter communication," *IEEE Trans. Commun.*, vol. 72, no. 6, pp. 3558–3571, Jun. 2024.

[21] L. Hu et al., "RIS-assisted integrated sensing and covert communication design," *IEEE Internet Things J.*, vol. 11, no. 9, pp. 16505–16516, May 2024.

[22] J. Hu, S. Yan, X. Zhou, F. Shu, and J. Wang, "Covert communications without channel state information at receiver in IoT systems," *IEEE Internet Things J.*, vol. 7, no. 11, pp. 11103–11114, Nov. 2020.

[23] R. Ma, X. Yang, G. Pan, X. Guan, Y. Zhang, and W. Yang, "Covert communications with channel inversion power control in the finite blocklength regime," *IEEE Wireless Commun. Lett.*, vol. 10, no. 4, pp. 835–839, Apr. 2021.

[24] X. Yao, P. Yang, J. Fu, Z. Liu, Y. Xiao, and S. Li, "A hybrid multi-domain index modulation for covert communication," *IEEE Wireless Commun. Lett.*, vol. 11, no. 1, pp. 8–12, Jan. 2022.

[25] Z.-J. Xu, Y. Gong, K. Wang, W.-D. Lu, and J.-Y. Hua, "Covert digital communication systems based on joint normal distribution," *Inst. Eng. Technol. Commun.*, vol. 11, no. 8, pp. 1282–1290, May 2017.

[26] F. Savaci, "Stable non-gaussian noise parameter modulation in digital communication (mobile communication)," *Inst. Eng. Technol. Electron. Lett.*, vol. 45, no. 24, pp. 1256–1257, Nov. 2009.

[27] Y. Li et al., "Performance analysis for covert communications under faster-than-Nyquist signaling," *IEEE Commun. Lett.*, vol. 26, no. 6, pp. 1240–1244, Jun. 2022.

[28] X.-Y. Hu, C. Bai, and H.-P. Ren, "A chaotic pseudo orthogonal covert communication system," in *Proc. Int. Conf. Commun. Inf. Syst.*, 2022, pp. 61–65.

[29] W. Gao, Y. Chen, C. Han, and Z. Chen, "Distance-adaptive absorption peak modulation (DA-APM) for Terahertz covert communications," *IEEE Trans. Wireless Commun.*, vol. 20, no. 3, pp. 2064–2077, Mar. 2021.

[30] Q. Hu, T. Lin, T. Wei, N. Huang, Y.-J. Zhu, and C. Gong, "Covert transmission in water-to-air optical wireless communication systems," *IEEE Trans. Inf. Forensics Security*, vol. 19, pp. 4432–4447, 2024.

[31] M. M. Hossain et al., "Low-noise speed-optimized large area CMOS avalanche photodetector for visible light communication," *IEEE J. Lightw. Technol.*, vol. 35, no. 11, pp. 2315–2324, Jun. 1, 2017.

[32] T. Lin, C. Gong, J. Luo, and Z. Xu, "Dynamic optical wireless communication channel characterization through air-water interface," in *Proc. IEEE/CIC Int. Conf. Commun. China*, 2020, pp. 173–178.

[33] (NVIDIA, Santa Clara, CA, USA). *NVIDIA Jetson Nano*. 2024. [Online]. Available: https://www.nvidia.com/en-us/autonomous-machines/embedded-systems/jetson-nano/product-development/

[34] B. Nemire (NVIDIA, Santa Clara, CA, USA). *DJI Launches GPU-Based High Performance Embedded Computer for Drones*. 2015. [Online]. Available: https://developer.nvidia.com/blog/dji-launches-gpu-based-high-performance-embedded-computer-for-drones-2/

[35] P. Zhang, Y. Zhong, and X. Li, "SlimYOLOv3: Narrower, faster and better for real-time UAV applications," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops*, Oct. 2019, pp. 1–9.

[36] N. A. Mohidem et al., "How can unmanned aerial vehicles be used for detecting weeds in agricultural fields?" *Agriculture*, vol. 11, no. 10, p. 1004, Oct. 2021.

[37] X. Yang, E. del Rey Castillo, Y. Zou, and L. Wotherspoon, "UAV-deployed deep learning network for real-time multi-class damage detection using model quantization techniques," *Autom. Constr.*, vol. 159, Mar. 2024, Art. no. 105254.

[38] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of Wasserstein GANs," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–20.

[39] I. Goodfellow et al., "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 27, 2014, pp. 1–9.

[40] B. A. Bash, D. Goeckel, and D. Towsley, "Limits of reliable communication with low probability of detection on AWGN channels," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 9, pp. 1921–1930, Sep. 2013.

[41] J. N. Njoku, M. E. Morocho-Cayamcela, and W. Lim, "BLER performance evaluation of an enhanced channel autoencoder," *Comput. Commun.*, vol. 176, pp. 173–181, Aug. 2021.

[42] K.-L. Besser, P.-H. Lin, C. R. Janda, and E. A. Jorswieck, "Wiretap code design by neural network autoencoders," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 3374–3386, 2020.

[43] O. Dizdar and A. O. Yilmaz, "Blind channel estimation based on the Lloyd-Max algorithm in narrowband fading channels and partial-band jamming," *IEEE Trans. Commun.*, vol. 60, no. 7, pp. 1986–1995, Jul. 2012.

Qingqing Hu received the B.S. degree from the Dalian University of Technology, Dalian, China, in 2016, and the M.S. and Ph.D. degrees from the University of Science and Technology of China, Hefei, China, in 2019 and 2024, respectively. She is currently a Lecturer with the School of Computer and Information, Hefei University of Technology, Hefei. Her research interests include information security, signal processing, and visible light communication.

Nuo Huang received the B.S. degree in electronics and information engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2012, and the Ph.D. degree in information and communication engineering from the National Mobile Communications Research Laboratory, Southeast University, Nanjing, China, in 2019. He is currently an Associate Professor with the Department of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei, China. From December 2015 to June 2017, he was a visiting student with the Department of Electrical Engineering, Columbia University, New York, NY, USA. His research interests include resource allocation and transceiver design in wireless (optical) communications. He was selected by the National Postdoctoral Program for Innovative Talents in 2019.

Yuwei Chen received the B.E. and M.Sc. degrees from the Department of Information Science and Electronics Engineering, Zhejiang University, China, in 1999 and 2002, respectively, the Ph.D. degree in circuits and systems from the Shanghai Institute of Technical Physics, Chinese Academy of Science in 2005, and the Doctor of Technology degree in telecommunication software from Aalto University, Finland, in 2020. He attended the development of the echo detecting sensor for the laser range finder of Chang'e—China first moon-explore satellite which proved a successful mission in 16-months on-orbit operation and prototyped China first airborne pushbroom laser scanner. He is currently working in the Advanced Laser Technology Anhui Province as the Acting Director General. He is also a Guest Professor of Academy of Opto-Electronics, Chinese Academy of Science and Zhejiang University. He holds 16 patents and has authored and co-authored more than 200 scientific papers and book chapters. His research interests include hyperspectral LiDAR, radar, and navigation and positioning.

Chen Gong (Senior Member, IEEE) received the B.S. degree in electrical engineering and mathematics (minor) from Shanghai Jiaotong University, Shanghai, China, in 2005, the M.S. degree in electrical engineering from Tsinghua University, Beijing, China, in 2008, and the Ph.D. degree from Columbia University, New York City, NY, USA, in 2012. He was a Senior Systems Engineer with the Qualcomm Research, San Diego, CA, USA, from 2012 to 2013. He is currently a Faculty Member with the University of Science and Technology of China. His research interests in wireless communications, optical wireless communications, and signal processing. He was selected by the Young 1000 Talent Program of China Government in 2014, and awarded by Hongkong Qiushi Outstanding Young Researcher Award in 2016.