Covert Communication System Based on Walsh Modulation and Noise Carriers

Zhijiang XU [1,2], *Suo ZHANG* [1], *Zhewei LIU* [1], *Jiliang LIN* [1], *Xiaoshuo HUANG* [1], *Jingyu HUA* [3]

[1] School of Automation, Zhejiang Polytechnic University of Mechanical and Electrical Engineering,
Binwen Rd. 528, 310053 Hangzhou, China
[2] Guangdong Provincial Key Laboratory of Advanced Wireless Communications,
Xueyuan Rd. 1088, 518055 Shenzhen, China
[3] School of Information & Electronic Engineering, Zhejiang Gongshang University,
Xuezheng Street 18, 310018 Hangzhou, China

{xuzhijiang, zhangsuo, liuzhewei, linjiliang, huangxiaoshuo}@zime.edu.cn, eehjy@zjut.edu.cn

Submitted March 4, 2024 / Accepted August 12, 2024 / Online first September 30, 2024

Abstract. *This study proposes a covert communication system, in which a non-zero-mean normally distributed random process is used as a carrier, and its mean is modulated by a Walsh code carrying M covert bits. The number of combinations is so huge that it is difficult for malicious parties to decode the covert information, even if they are aware of the existence of the transmitting signal. The received signal is multiplied at the receiving end with each Walsh code, and the mean value is computed. The Walsh code corresponding to the largest mean value is the transmitter's modulation code, thus recovering the transmitted covert bits. The system's theoretical symbol error rate and bit error rate are derived under additive white Gaussian noise and quasi-static fading channels, respectively. Simulation results are very consistent with the theoretical derivation. Compared with other existing schemes, the proposed scheme has good security, flexibility, and BER performance, and is very suitable for IoT devices with limited resources and low transmission rate but high concealability requirements.*

Keywords

Normal distribution, Walsh code, bit error rate, covert communication, IoT

1. Introduction

Wireless communication networks have been one of the important infrastructures of modern society, the carrier of mobile internet, Internet of Things (IoT), intelligent manufacturing, and other information industries. However, wireless communications are more vulnerable to security threats than wired counterparts owing to the open nature of wireless channels [1]. Information security-related problems are increasingly becoming a critical issue that cannot be ignored [2]. With the advent of the IoT era, the number of terminal devices is increasing rapidly, and the issue of protection of private information closely related to individuals is becoming more important. While security protection via steganography [3] or physical-layer security (PLS) [4] has been studied heavily in the past, users' privacy concerns have not been fully alleviated. For the message encryption at the application layer, a more complex encryption protocol can be established to secure the user's privacy [5]. However, traditional cryptography methods for network security cannot solve all security problems in IoT systems [6], due to the limited resources of low-cost IoT terminal devices.

Recently, significant efforts have been made to pave the way for the integration of PLS into contemporary and future networks [7–10]. The primary motivation for deploying PLS is its low complexity and ability to provide secure transmission of information theory, which alleviates the burden of complexity associated with implementing sophisticated encryption schemes. Unlike cryptographic schemes, the PLS techniques including channel coding and precoding take full advantage of the physical characteristics of the wireless channel (e.g., noise, fading, interference) to obtain secure and efficient data transmission. The quality of the eavesdropping channel can be reduced and the security rate of the wireless communication system can be increased by the precoding technique, thus providing a lightweight secure transmission scheme for resource-limited devices by utilizing PLS technology. Furthermore, many physical-layer sources such as power allocation, beamforming, and polarization are jointly optimized to guarantee security while providing excellent transmission performance. The steganography and PLS can only hide the information in a confusing order to prevent the malicious parties from decoding the legitimate information from the received signals. However, they cannot prevent the malicious parties from acquiring the transmitted information by improving their decoding techniques. In many cases, the mere presence of communications or a change in the way

DOI: 10.13164/re.2024.0552

they are communicated is enough to arouse the suspicion of a malicious party and reveal the onset of events [11].

To prevent the communication between legitimate users from being detected by the malicious party and to ensure the security of the wireless network, it is recommended to use covert communication technique [12], also known as low detection probability. The basis of covert communication is to hide the covert information in a noisy background, and covert communication has many practical applications in military, national security, and commerce [13], [14]. Research on IRS-assisted covert communication has been carried out recently [15–17]. For low-cost IoT devices that won't be configured with MIMO RF modules and won't have the superb computational power to implement deep reinforcement learning, these solutions may be less appropriate.

Wireless covert channel based on physical layer (WCC-P) can be categorized into two types: coding-based wireless covert channels (C-WCC) and modulation-based wireless covert channels (M-WCC) [18]. Currently, research on WCC-P focuses on constellation-modulated covert channels belonging to M-WCC, such as "dirty constellation". This approach hides the information in "dirty" constellations that mimic the noise typically caused by hardware defects and channel conditions [19]. In [20], the authors investigated the optimal non-equiprobable constellations modulation schemes for covert communication systems from a practical view. The authors investigated the optimal design of a wireless-powered covert communication (WP-CC) system and proposed a probabilistic accumulate-then-transmit (ATT) protocol to achieve a higher level of covertness in [21]. A random process is used as a carrier and one of the characteristic parameters of the carrier is modulated by a covert bit for covert communication. The main references in this area are the following. In [22], [23], the covert bits modulate the characteristic exponent α or skewness β of the alpha-stable distribution. Xu et al. proposed a covert communication scheme in [24] where the correlation coefficients of two consecutive Gaussian sequences are modulated by a covert bit. In [25], Xu et al proposed another scheme that utilizes a non-zero mean Gaussian sequence as a carrier, and they modulated the mean value of the carrier signal with a covert bit. The main difference between this study and [25] is that this study allows for multi-bit transmissions, while [25] can only allow for 1-bit transmissions. The concept of noise modulation (NoiseMod) is introduced in [26] to covert information transmission.

In this study, we attempt to increase the transmission rate by transmitting multiple covert bits simultaneously, while being highly resistant to brute force decryption. Our main contributions can be summarized as follows:

- A new covert communication method is proposed, which utilizes a Gaussian noise with a non-zero mean value of length L as a carrier. The signal of the carrier multiplied by the bipolar Walsh code of length L called the covert signal, presents a Gaussian noise with a mean value of zero. Gaussian noise is a common channel noise, therefore, the transmitted covert signal is indistinguishable from the background noise in terms of statistical properties. Thus, the eavesdropper cannot easily detect the presence of the covert transmission.

- The proposed covert communication scheme has good security. 2^M Walsh codes are selected from the set of L orthogonal Walsh codes as the modulation code set for M covert bits, i.e., each Walsh code carries M covert bits, where $2^M \ll L$. This, on the one hand, makes multi-bit transmission easy to realize. More importantly, on the other hand, the combination of 2^M Walsh codes selected from L Walsh codes as the modulation code set is $\binom{L}{2^M}$. By choosing the appropriate M and L, the number of combinations is very huge, which makes it more difficult for a malicious party to crack it through brute force. Furthermore, for different M, their corresponding covert signals are indistinguishable both in terms of time-domain waveforms and statistical properties. In other words, it is very difficult for an eavesdropper to guess how many bits are contained in each modulation code.

- The theoretical SER and BER of the system are derived under additive white Gaussian noise and quasi-static fading channels, respectively. Extensive simulations are conducted to evaluate the performance of our proposed covert system's performance. Simulation results are very consistent with the theoretical derivation.

The remaining parts of this study are organized as follows. In Sec. 2, the probability density function (PDF) of a continuous random variable multiplied by an equiprobable discrete random variable, and the covert communication system is presented. In Sec. 3, the expression of the mean estimator and its PDF are derived, as well as the symbol error rate (SER) and the bit error rate (BER) of the proposed covert system. In Sec. 4, extensive numerical calculations and Monte Carlo simulations are presented to evaluate the performance of the proposed scheme in additive white Gaussian noise (AWGN) channel. In Sec. 5, a conclusion is drawn.

2. Covert Communication Scheme

The idea of this study is to use environmental noise to mislead and confuse eavesdroppers in determining the presence or absence of transmitting signals. The eavesdroppers have no intention of revealing the content in the signals sent from the transmitter if they are not aware of the transmission since they assume that the "signals" are merely environmental noise. Hence, the eavesdroppers will not exploit resources to tackle the "signals" and the transmitters can well guarantee their private information.

The environmental noise can be modeled as Gaussian distributed noise in many wireless communication scenarios, i.e., the environmental noise obeys the normal distribution with zero mean. If the transmitted covert signal is also a noise obeying a zero-mean normal distribution, the covert signal is statistically indistinguishable from the environmental noise. This makes it indistinguishable from eavesdroppers, thus the purpose of covert transmission can be achieved.

In this section, we first provide a brief review of the PDF of a continuous random variable multiplied by an equiprobable discrete random variable, then put forward a covert communication system combining non-zero-mean-normally distributed noise and Walsh spread spectrum code.

2.1 PDF of Mixed Random Variable

Before introducing the proposed covert communication system, the PDF of a mixed random variable is given in this subsection. Specially, a non-zero-mean-normally distributed random variable multiplied by an equiprobable discrete random variable can be considered to be normally distributed with zero mean, under the condition of $\mu < 0.4\sigma$.

Let $H \in \{+1, -1\}$ be equiprobable discrete random variable, i.e., $\Pr(H = +1) = \Pr(H = -1) = \frac{1}{2}$. Let X be a non-zero-mean-normally distributed random variable and its PDF be $f_X(x)$, i.e., $X \sim \mathcal{N}(\mu, \sigma_\mathrm{c}^2)$. Further, H and X are assumed to be independent of each other. If let $S = HX$, then S is a continuous random variable and its corresponding PDF, $f_S(s)$, is given by

$$\begin{aligned} f_S(s) &= \frac{1}{2}\left(f_X(s) + f_X(-s)\right) \\ &= \frac{1}{2}\left(\frac{1}{\sqrt{2\pi}\sigma_\mathrm{c}}\mathrm{e}^{-\frac{(s-\mu)^2}{2\sigma_\mathrm{c}^2}} + \frac{1}{\sqrt{2\pi}\sigma_\mathrm{c}}\mathrm{e}^{-\frac{(s+\mu)^2}{2\sigma_\mathrm{c}^2}}\right). \end{aligned} \quad (1)$$

In this study, κ is defined as the ratio of the mean μ to the standard deviation σ_c of the random variable X, which we refer to as the mean-to-standard-deviation-ratio (MSR),

$$\kappa = \frac{\mu}{\sigma_\mathrm{c}}. \quad (2)$$

Specifically, when $0 < \kappa \le 0.4$, $f_S(s)$ can be approximated by

$$f_S(s) \simeq \frac{1}{\sqrt{2\pi}\sigma_\mathrm{s}}\mathrm{e}^{-\frac{s^2}{2\sigma_\mathrm{s}^2}} \quad (3)$$

where $\sigma_\mathrm{s}^2 = \mu^2 + \sigma_\mathrm{c}^2$. For a detailed proof process, see [25]. From (3), it can be seen that the random variable S obeys a normal distribution with zero mean, i.e., $S \sim \mathcal{N}(0, \sigma_\mathrm{s}^2)$.

2.2 Proposed Covert Communication System

Figure 1 shows that the covert bitstream in the proposed covert communication scheme is modulated by the Walsh code. Non-zero-mean normally distributed noise is used as the carrier. The transmitted covert signal exhibits the statistical properties of zero-mean-normally distributed, which is indistinguishable from the environmental noise. Using the orthogonality of Walsh coding, a receiver that demodulates a covert signal using the same Walsh coding as the transmitter will have the same statistical properties of the demodulated signal as the transmitter's carrier, i.e., its mean value is not zero. Otherwise, the demodulated signal is a Gaussian noise with zero mean value. At the transmitting end, different Walsh codes are chosen for different bit streams, i.e., Walsh codes carry the covert information. At the receiving end, the mean value of the demodulated signal is used to determine which Walsh code is used at the transmitting end, thus achieving the purpose of transmitting covert information.

Non-zero-mean Gaussian random processes of length L are used as carriers at the transmitter, denoted as $X = \{X_i, i = 0, \cdots, L-1\} \sim \mathcal{N}(\mu, \sigma_\mathrm{c}^2)$, where L is an integer power of 2 (e.g., $L = 1024$), μ is the mean and σ_c^2 is the variance of X. It is well known that there are L Walsh codes in the Walsh code set of length L, denoted as $\{\mathcal{H}_i, i = 0, \cdots, L-1\}$. From which the same 2^M Walsh codes (excluding the Walsh code where all code chips are '1', i.e., $\mathcal{H}_0$) are pre-selected as the modulation codes by both the transmitting and receiving parties. The set of these Walsh modulation codes is denoted as $\{H_i, i = 0, \cdots, 2^M - 1\}$, where M is the length of the transmitted covert bitstream, and satisfies the condition $2^M \ll L$. The selected set of Walsh modulation codes is a subset of the entire code set, i.e., $\{H_i\} \subset \{\mathcal{H}_i\}$. Moreover, the correspondence between H_i and $\mathcal{H}_j$ is agreed upon in advance by the transmitting and receiving parties.

A M-bit stream, denoted as $\{b_i, i = 0, \cdots, M-1\}$, selects a Walsh code from the modulation codes. The mapping of the M-bit covert bits to the 2^M Walsh codes $\{H_i\}$ can be agreed upon by the transceiver and the transmitter, either by natural binary encoding, Gray's code, or an arbitrary one-to-one mapping. For simplicity, this study assumes that the natural binary encoding is used, i.e., the decimal value of the M-bit covert bits, m, corresponds to the H_m. The modulation process at the transmitter is that the covert bitstream $\{b_i\}$ is mapped to obtain modulation code $H_m \in \{H_i\}$, which is then multiplied by the carrier to get the transmitted signal S, i.e.,

$$S = H_m \times X = H_{m,i} \times X_i, \quad i = 0, \cdots, L-1 \quad (4)$$

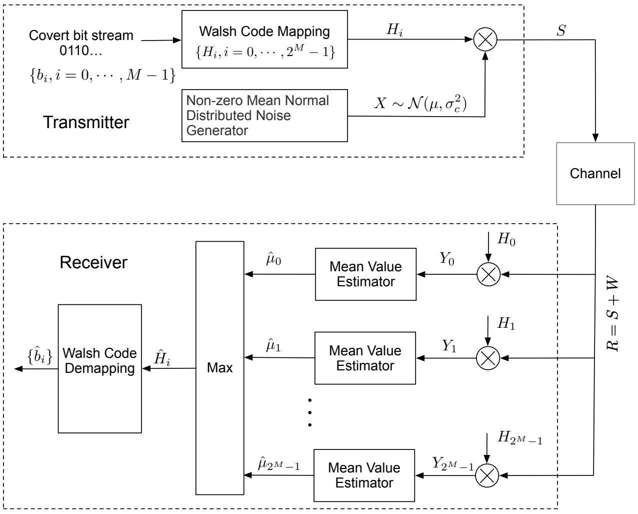
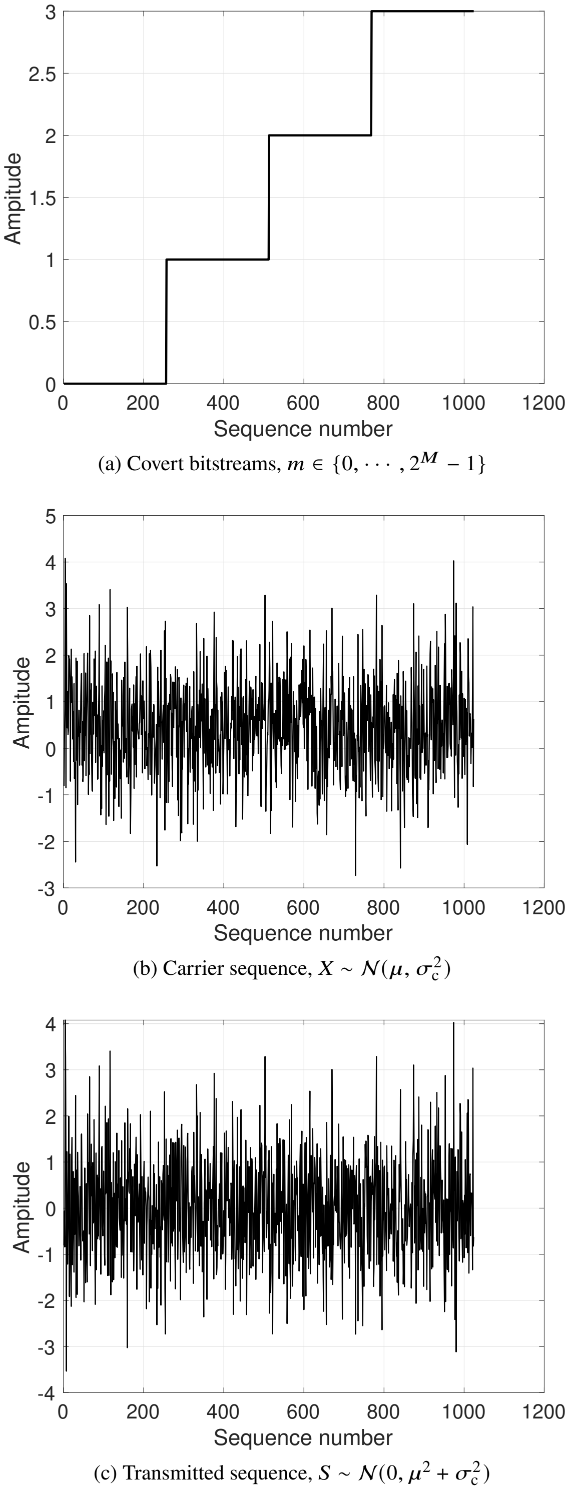


Fig. 1. Block diagram of the proposed communication system.

where $H_{m,i}$ is the i-th chip of the Walsh code H_m, and its value is ±1. The signal S is transmitted to the channel. Furthermore, the signal S can be considered to follow a normal distribution of zero mean under the condition of $\mu \leq 0.4\sigma_\mathrm{c}$ (see (3)). Without loss of generality, assuming that the covert bits are all '0', the modulation code is chosen to be H_0, then the transmitted signal $S = H_0 X$.



(a) Covert bitstreams, $m \in \{0, \cdots, 2^M - 1\}$

(b) Carrier sequence, $X \sim \mathcal{N}(\mu, \sigma_\mathrm{c}^2)$

(c) Transmitted sequence, $S \sim \mathcal{N}(0, \mu^2 + \sigma_\mathrm{c}^2)$

Fig. 2. Covert bitstreams, carriers, and transmitted signals in the time domain, where $M = 2$, $L = 256$, $\mu = 0.4$ and $\sigma_\mathrm{c} = 1$.

In the above proposed covert transmitter, the normally distributed carrier is set to have a mean of 0.4 and a variance of 1, i.e., $\mu = 0.4, \sigma_\mathrm{c} = 1$. Walsh codes are generated from `comm.WalshCode` function in MATLAB, whose parameter "`Index`" take values from $L/2$ to $L/2 + 2^M - 1$, corresponding to the covert bitstreams (symbols) 0 to $2^M - 1$, and whose parameters "`Length`" and "`SamplesPerFrame`" are both set to L. Comparing subfigures (b) and (c) in Fig. 2, it can be seen that the carrier sequence has a significant deviation from 0, while the transmitted sequence after being multiplied by a Walsh code has an obvious symmetry about 0, which exhibits the same statistical properties as the environmental noise. In other words, the mean of the transmitted signal is still 0, no matter what the bitstream is.

At the receiving end, the received signal is multiplied with each Walsh code H_m in the selected Walsh code set $\{H_i\}$ to obtain 2^M demodulated signals Y_i. Noting that the transmitter has chosen H_0 as the modulation code, it is clear that, using the orthogonality of Walsh codes, in the case of an ideal channel, Y_0 is a stochastic process that is correctly demodulated and $Y_i (i \neq 0)$ is a stochastic process that is incorrectly demodulated. In particular, under ideal and Gaussian channels, Y_0 is a Gaussian stochastic process with non-zero mean, and $Y_i (i \neq 0)$ is a Gaussian stochastic process with zero mean. In this study, assuming that the mean μ of the carrier in the transmitter is greater than 0. Correspondingly, the estimated value of the mean $\hat{\mu}_0$ of Y_0 is greater than 0, and the mean estimate $\hat{\mu}_i$ of $Y_i (i \neq 0)$ is 0. Thus, the Walsh modulation code corresponding to the maximum value of the output of all mean estimators is the estimate of the Walsh modulation code of the transmitter, which recovers the transmitted covert bit stream.

3. System Performance Analysis

In this section, we will derive the expression of the mean estimator $\{\hat{\mu}_i\}$ and its PDF. Combined with the hard decision, the SER and BER of the proposed communication system are obtained.

3.1 Mean Estimator and its PDF

After an additive white Gaussian channel, the received signal is

$$R = S + W = H_0 X + W \tag{5}$$

where W is the AWGN with zero mean and σ_n^2 variance, i.e., $W \sim \mathcal{N}(0, \sigma_\mathrm{n}^2)$. In this study, the signal-to-noise ratio (SNR), ρ, is defined as the ratio of the average power of the transmitted signal to the average power of the additive white Gaussian noise, i.e.,

$$\rho = \frac{\sigma_\mathrm{s}^2}{\sigma_\mathrm{n}^2} = \frac{\mu^2 + \sigma_\mathrm{c}^2}{\sigma_\mathrm{n}^2} = \frac{(1 + \kappa^2)\sigma_\mathrm{c}^2}{\sigma_\mathrm{n}^2}. \tag{6}$$

Notice that any two Walsh codes in the Walsh code set are multiplied together to make another Walsh code, i.e.,

$$H_jH_k = \begin{cases} \mathcal{H}_0, & j = k, \\ \mathcal{H}_l, & j \neq k. \end{cases} \tag{7}$$

It is possible that $\mathcal{H}_l$ may not be in the set of the Walsh codes $\{H_i\}$ chosen by the transmitting and receiving parties, but it must be in the entire Walsh codes set $\{\mathcal{H}_i\}$. Importantly, this does not affect the specific derivation process below as well as the correctness of the conclusions.

By the orthogonality of the Walsh code, it follows that the demodulated signal, Y_i is

$$Y_i = RH_i = XH_0H_i + WH_i = \begin{cases} X\mathcal{H}_0 + WH_0, & i = 0, \\ X\mathcal{H}_k + WH_i, & i \neq 0. \end{cases} \tag{8}$$

From (7), we can obtain that $X\mathcal{H}_0 = X \sim \mathcal{N}(\mu, \sigma_\mathrm{c}^2)$, $WH_0 \sim \mathcal{N}(0, \sigma_\mathrm{n}^2)$, $X\mathcal{H}_k \sim \mathcal{N}(0, \sigma_\mathrm{s}^2)$ and $WH_i \sim \mathcal{N}(0, \sigma_\mathrm{n}^2)$. Furthermore, since the stochastic processes X and W are two Gaussian distributions independent of each other, there are $\mathbb{E}[(X\mathcal{H}_0)(WH_0)] = \mathbb{E}[X]\mathbb{E}[W]\mathbb{E}[H_0] = 0$ that hold, as well as $\mathbb{E}[(X\mathcal{H}_k)(WH_i)] = \mathbb{E}[X]\mathbb{E}[W]\mathbb{E}[\mathcal{H}_l] = 0$, where $\mathbb{E}[X]$ is the expectation of the random variable X. This means that $X\mathcal{H}_0$ and WH_0, as well as $X\mathcal{H}_k$ and WH_i, are independent of each other. Thus

$$Y_i \sim \begin{cases} \mathcal{N}(\mu, \sigma_\mathrm{c}^2) + \mathcal{N}(0, \sigma_\mathrm{n}^2) = \mathcal{N}(\mu, \sigma_\mathrm{c}^2 + \sigma_\mathrm{n}^2), & i = 0, \\ \mathcal{N}(0, \sigma_\mathrm{s}^2) + \mathcal{N}(0, \sigma_\mathrm{n}^2) = \mathcal{N}(0, \sigma_\mathrm{s}^2 + \sigma_\mathrm{n}^2), & i \neq 0. \end{cases} \tag{9}$$

This means that the output $\{Y_i\}$ at the receiving end obeys a Gaussian distribution. Moreover, there is

$$\begin{aligned} \mathbb{E}[Y_0Y_i] &= \mathbb{E}[X^2\mathcal{H}_k] + \mathbb{E}[XWH_i] + \mathbb{E}[XWH_0\mathcal{H}_k] \\ &\quad + \mathbb{E}[W^2H_0H_i] = 0 = \mathbb{E}[Y_0]\mathbb{E}[Y_i],\ i \neq 0 \end{aligned} \tag{10}$$

holds, as well as:

$$\begin{aligned} \mathbb{E}[Y_iY_j] &= \mathbb{E}[(X\mathcal{H}_k)(X\mathcal{H}_l)] + \mathbb{E}[(X\mathcal{H}_k)(WH_j)] \\ &\quad + \mathbb{E}[(X\mathcal{H}_l)(WH_i)] + \mathbb{E}[(WH_i)(WH_j)] \\ &= 0 = \mathbb{E}[Y_i]\mathbb{E}[Y_j],\ i \neq j \neq 0. \end{aligned} \tag{11}$$

Combining (9), (10) and (11), we get that $\{Y_i\}$ are mutually independent normally distributed stochastic processes.

Without loss of generality, let $\sigma_\mathrm{c} = 1$, and combine (6) and (9) to get

$$Y_i \sim \begin{cases} \mathcal{N}(\kappa, \frac{\rho+1+\kappa^2}{\rho}) & i = 0, \\ \mathcal{N}(0, \frac{(\rho+1)(1+\kappa^2)}{\rho}) & i \neq 0. \end{cases} \tag{12}$$

In this study, the MSR is taken as 0.2, i.e., $\kappa = 0.2$. When the SNR ρ is varied from −10 dB to 30 dB, the relative error between the variance of the correctly demodulated and incorrectly demodulated stochastic processes is

$$\begin{aligned} \eta &= \frac{\mathrm{var}(Y_i) - \mathrm{var}(Y_0)}{\mathrm{var}(Y_i)} = 1 - \frac{\rho + 1 + \kappa^2}{(\rho+1)(1+\kappa^2)} \\ &= 1 - \frac{1 + \frac{\kappa^2}{1+\rho}}{1+\kappa^2} < \frac{\kappa^2}{1+\kappa^2} < \kappa^2 = 4\%. \end{aligned} \tag{13}$$

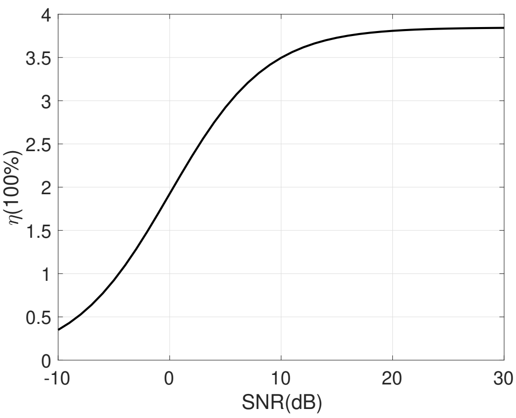


Fig. 3. Relative error between the variance of the correctly demodulated and incorrectly demodulated, when $\kappa = 0.2$.

For simplicity, it can be seen from Fig. 3 and (13) that the variance in the case of incorrect demodulation can be considered as being the same as the variance in the case of correct demodulation. Therefore, in all the following sections, we consider $\mathrm{var}(Y_0)$ to be equal to $\mathrm{var}(Y_i)$.

For a Gaussian random process Y_i with length L, the mean of Y_i, $\hat{\mu}_i$ is estimated as

$$\hat{\mu}_i = \frac{1}{L}\sum_{j=0}^{L-1} Y_{i,j} = \begin{cases} \hat{\mu}_0 \sim \mathcal{N}(\kappa, \sigma^2), & i = 0, \\ \hat{\mu}_i \sim \mathcal{N}(0, \sigma^2), & i \neq 0 \end{cases} \tag{14}$$

where $Y_{i,j}$ is the j-th sample of Y_i, and the variance of $\hat{\mu}_i$ is defined as

$$\sigma^2 = \frac{\rho + 1 + \kappa^2}{\rho L}. \tag{15}$$

Let $\hat{\mu}_0$ be the mean-estimated random variable when correctly demodulated and $\hat{\mu}_i$ be the mean-estimated random variable when incorrectly demodulated, then the decision is wrong when $\hat{\mu}_0 < \hat{\mu}_i$ for all $i \in \{1, \cdots, 2^M - 1\}$. Therefore, the symbol error rate, r_s, is defined as

$$r_\mathrm{s} = \Pr\left(\hat{\mu}_0 < \max_{i=1}^{2^M-1}\{\hat{\mu}_i\}\right). \tag{16}$$

We will derive the expressions of the symbol error rate and the bit rate error in the following subsection.

3.2 SER when $M = 1$

We first derive the SER at $M = 1$. In this case, $\max_i\{\hat{\mu}_i\}$ simplifies to $\hat{\mu}_1$. Since Y_0 and Y_1 are mutually independent Gaussian stochastic processes, their mean value estimates $\hat{\mu}_0$ and $\hat{\mu}_1$ are also independent of each other, i.e., $\mathbb{E}[\hat{\mu}_0\hat{\mu}_1] = \mathbb{E}[\hat{\mu}_0]\mathbb{E}[\hat{\mu}_1] = 0$. Let $\Lambda = \hat{\mu}_0 - \hat{\mu}_1$, and since $\hat{\mu}_0$ and $\hat{\mu}_1$ are two Gaussian random variables independent of each other, then Λ is also a Gaussian random variable. Its mean value is the difference between the means of $\hat{\mu}_0$ and $\hat{\mu}_1$, and its variance is the sum of their variances, i.e., $\Lambda \sim \mathcal{N}(\kappa, 2\sigma^2)$. Therefore, we can easily obtain the SER in this case as

$$
\begin{aligned}
r_{\mathrm{s}} = \Pr(\Lambda < 0) &= \int_{-\infty}^{0} \frac{1}{\sqrt{4\pi}\sigma} \mathrm{e}^{-\frac{(\lambda-\kappa)^2}{4\sigma^2}} \mathrm{d}\lambda \\
&= \frac{1}{2}\mathrm{Erfc}\left(\frac{\kappa}{2\sigma}\right) = \frac{1}{2}\mathrm{Erfc}\left(\frac{\kappa}{2}\sqrt{\frac{\rho L}{1+\kappa^2+\rho}}\right)
\end{aligned} \tag{17}
$$

where the complementary error function is defined as $\mathrm{Erfc}(z) = \frac{2}{\sqrt{\pi}}\int_z^{\infty} \mathrm{e}^{-t^2}\mathrm{d}t$. When the SNR ρ increases to infinity, i.e., for an ideal channel, the SER still exists and is

$$
r_{\mathrm{s}} = \frac{1}{2}\mathrm{Erfc}\left(\frac{\kappa}{2}\sqrt{L}\right). \tag{18}
$$

3.3 SER when $M > 1$

For multi-bit covert bit stream with $M > 1$, the derivation process of SER is a bit more complicated than that for $M = 1$, since $\max_{i=1}^{2^M-1}\{\hat{\mu}_i\}$ no longer obeys a normal distribution. For brevity, the detailed derivation of the SER is given in Appendix A, and the result is presented directly here. The expression SER is

$$
\begin{aligned}
r_{\mathrm{s}} &= \int_{-\infty}^{\infty} \frac{2^M-1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{z^2}{2\sigma^2}}\left(1-\frac{1}{2}\mathrm{Erfc}\left(\frac{z-\kappa}{\sqrt{2}\sigma}\right)\right) \\
&\qquad \times\left(1-\frac{1}{2}\mathrm{Erfc}\left(\frac{z}{\sqrt{2}\sigma}\right)\right)^{2^M-2}\mathrm{d}z \\
&= 1-\int_{-\infty}^{\infty}\left(1-\frac{1}{2}\mathrm{Erfc}\left(\frac{z+\kappa}{\sqrt{2}\sigma}\right)\right)^{2^M-1}\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{z^2}{2\sigma^2}}\mathrm{d}z.
\end{aligned} \tag{19}
$$

The integration of the above equation does not lead to a closed-form expression. However, when M, L, and κ are given, we can obtain the SER at a certain SNR by numerical integration.

3.4 BER

For a covert bitstream of length M, the number of 1-bit errors is $\binom{M}{1}$ and the number of i-bit errors is $\binom{M}{i}$, and so on. At the receiving end, H_0 is the correct demodulation code and $\{H_i, i = 1, \cdots, 2^M - 1\}$ are the incorrect demodulation codes, and they correspond to the outputs of the mean estimator $\hat{\mu}_0$ and $\{\hat{\mu}_i, i \neq 0\}$, respectively. Furthermore, the outputs of the mean estimators, $\{\hat{\mu}_i, i \neq 0\}$, all obey the $\mathcal{N}(0, \sigma^2)$ distribution and are independent of each other. Therefore, they have the same probability of causing demodulation errors, which is $\frac{r_{\mathrm{s}}}{2^M-1}$ in each case. Based on the above discussion, the BER r_{b} is

$$
r_{\mathrm{b}} = \frac{\frac{r_{\mathrm{s}}}{2^M-1}\sum_{i=1}^{M} i\binom{M}{i}}{M} = \frac{r_{\mathrm{s}}}{M(2^M-1)}\sum_{i=1}^{M} i\binom{M}{i}. \tag{20}
$$

By the Binomial Theorem, $(1+x)^M = \sum_{i=0}^{M}\binom{M}{i}x^i$, and taking the derivatives of the variable x on both sides of the equation, we get

$$
M(1+x)^{M-1} = \sum_{i=1}^{M} iC_M^i x^{i-1} \xrightarrow{x=1} M2^{M-1} = \sum_{i=1}^{M} iC_M^i. \tag{21}
$$

Combining (20) and (21), the expression for BER is obtained as

$$
r_{\mathrm{b}} = \frac{2^{M-1}}{2^M-1}r_{\mathrm{s}}. \tag{22}
$$

When the number of covert bitstream bits M is relatively large, the BER approximates half of the SER. This is more interesting because it is different from the relationship between SER and BER in conventional multi-digit modulation communication systems.

3.5 SER under Quasi-Static Fading Channels

To further evaluate the potential impact of channel impairments on the system's performance, a frequency-flat Rayleigh fading channel without Doppler shift is considered in this study. The channel gain (or fading coefficient), h, remains unchanged during a symbol interval T_{s} and obeys a Rayleigh distribution with parameter γ, i.e.,

$$
f_h(h) = \frac{h}{\gamma^2}\exp\left(-\frac{h^2}{2\gamma^2}\right). \tag{23}
$$

Under this assumption, the derivation procedure is the same as under an AWGN channel, and we obtain the output of the mean value estimator as

$$
\hat{\mu}_i = \begin{cases} \hat{\mu}_0 \sim \mathcal{N}(h\kappa, \sigma_h^2), & i = 0, \\ \hat{\mu}_i \sim \mathcal{N}(0, \sigma_h^2), & i \neq 0 \end{cases} \tag{24}
$$

where $\sigma_h^2 = \frac{1+\kappa^2+h^2\rho}{\rho L}$. Further, the SER at $M = 1$ is

$$
r_{\mathrm{s}} = \int_0^{\infty} \frac{1}{2}\mathrm{Erfc}\left(\frac{\kappa h}{2}\sqrt{\frac{\rho L}{1+\kappa^2+h^2\rho}}\right)\frac{h}{\gamma^2}\mathrm{e}^{-\frac{h^2}{2\gamma^2}}\mathrm{d}h. \tag{25}
$$

Similarly, the SER for $M > 1$ is rewritten directly from (19) as

$$
\begin{aligned}
r_{\mathrm{s}} = \int_0^{\infty} &\left[\int_{-\infty}^{\infty} \frac{2^M-1}{\sqrt{2\pi}\sigma_h}\mathrm{e}^{-\frac{z^2}{2\sigma_h^2}}\left(1-\frac{1}{2}\mathrm{Erfc}\left(\frac{z-h\kappa}{\sqrt{2}\sigma_h}\right)\right)\right. \\
&\left.\times\left(1-\frac{1}{2}\mathrm{Erfc}\left(\frac{z}{\sqrt{2}\sigma_h}\right)\right)^{2^M-2}\mathrm{d}z\right]\frac{h}{\gamma^2}\mathrm{e}^{-\frac{h^2}{2\gamma^2}}\mathrm{d}h.
\end{aligned} \tag{26}
$$

For comparison with the BER under AWGN channels, the fading average power is set to 1 so that the assumed fading channel follows a Rayleigh distribution with parameter $\gamma = \sqrt{1/2}$. Then, the SER at $M = 1$ under the fading channel channel is rewritten as

$$
r_{\mathrm{s}} = \int_0^{\infty} \mathrm{Erfc}\left(\frac{\kappa h}{2}\sqrt{\frac{\rho L}{1+\kappa^2+h^2\rho}}\right)h\mathrm{e}^{-h^2}\mathrm{d}h. \tag{27}
$$

3.6 Security

The security of the communication system proposed in this study is mainly manifested in the following aspects, which are explained as follows.

First of all, as can be seen from the transmitter of the proposed communication system shown in Fig. 1, the non-zero-mean Gaussian noise is used as a carrier, and after multiplication with the Walsh code, it becomes a zero-mean Gaussian noise, whose statistical properties are no different from the environmental noise, and which is not easily detected by the malicious party.

Secondly, for covert M bitstream, 2^M Walsh codes $\{H_i\}$ are selected from the set of L orthogonal Walsh codes $\{\mathcal{H}_i\}$ as the modulation codes. This makes multi-digit transmission easy to implement on one hand; more importantly, on the other hand, the combination of 2^M codes selected from the set of L codes is $\binom{L}{2^M}$. For example, if let $L = 1024$ and $M = 2$, i.e., 4 Walsh codes are selected from the 1023 Walsh codes (excluding $\mathcal{H}_0$ with all '+1' chips) as the modulation code for the 2 covert bits, there are $\binom{1023}{4} \approx 4.5 \times 10^{10}$ combinations. Furthermore, when M increases to 4, it increases sharply to $\binom{1023}{16} \approx 6.1 \times 10^{34}$ combinations. This huge number of combinations greatly increases the difficulty for a malicious party to break it through brute force.

Thirdly, at the receiving end of the proposed communication system, the PDFs of the output of the mean estimator ($\{\hat{\mu}_i\}$) for incorrect demodulation are all the same and $\{\hat{\mu}_i\}$ are independent of the number of bits M. Therefore, it is difficult for a malicious party to guess how many bits are contained in each modulation code, i.e., it is not possible to guess the size of M, which makes the cracking more difficult.

Finally, even if the modulation code set $\{H_i\}$ chosen by the legitimate users is fixed and illegally obtained by a malicious party in advance, the legitimate user continues to transmit covert data by using a method similar to frequency hopping spread spectrum, although at the expense of reduced transmission rates. Specifically, the transmission rate is reduced by a factor of M. In other words, the number of bits carried by each Walsh modulation code is reduced from M to 1 bit. The Walsh modulation code corresponding to bit '0' is unchanged and remains H_0. However, the modulation code corresponding to bit '1' is selected from the $\{H_1, \cdots, H_{2^M-1}\}$ set in an order predetermined by the legitimate parties. In this scenario, if the malicious party still uses the entire $\{H_i\}$ modulation code set to demodulate, the SER is relatively high, while the legitimate receiver uses 2 modulation codes to demodulate, and its SER is relatively low. This can be seen in Fig. A.1 in Appendix A, where the PDF curve of Z at $M = 4$ is very close to the PDF curve of $\hat{\mu}_0$, whereas the PDF curve of Z at $M = 1$ is relatively far away from the PDF curve of $\hat{\mu}_0$. This can be interpreted as a high SNR at the legitimate receivers' end, while a low SNR at the malicious receivers. Therefore, there exists a security capacity that guarantees the ability to transmit covertly.

4. Simulation and Analysis

In this section, we provide numerical and MATLAB simulation results to examine the SER and BER performances of the proposed covert communication system. From (17) and (22), it can be seen that the SER/BER is closely related to the parameters the number of covert bits per Walsh code M, the length of the carrier random sequence L, the MSR κ and the SNR ρ. In all simulations, we consider an MSR of $\kappa = 0.2$, the variance of the carrier of $\sigma_c = 1$, and vary M or L. Walsh codes as a set of modulation codes are generated from "`comm.WalshCode`" function in MATLAB, whose the parameter "`Index`" take values from $L/2$ to $L/2 + 2^M - 1$, corresponding to the covert symbols (bitstreams) 0 to $2^M - 1$, and whose parameters "`Length`" and "`SamplesPerFrame`" are both set to L.

The BERs of the simulation and theoretical derivation are shown in Fig. 4. From the figure, it can be seen that our theoretical derivation is in good agreement with the simulation, and the correctness of the theoretical derivation is also verified. When the SNR is as low as −4 dB, there is still a BER of less than 1% (in the case $L = 1024$).

In [25], the authors proposed another scheme that utilizes a non-zero mean Gaussian sequence as a carrier, and the polarity of the carrier's mean is modulated by the covert bits. The SER/BER of this covert system is

$$r_s = r_b = \frac{1}{2}\mathrm{Erfc}\left(\frac{\kappa}{2}\sqrt{\frac{2\rho L}{1+\kappa^2+\rho}}\right). \tag{28}$$

Because Erfc(x) decreases as the variable x increases, the BER of work [25] is smaller than that of this study. Indeed, such a comparison may not be fair. From the point of view of the constellation diagram, the constellation points for sending binary bits in work [25] are at μ and $-\mu$, respectively, whereas those in this study are at 0 and μ, respectively. This means that the distance between the constellation points in this study is exactly half of that of work [25]. Therefore, in this sense, the BER is the same for both. More importantly, the proposed covert system in this study can easily realize multi-bit transmission with stronger security and no significant increase in complexity. In the works [22], [23], the characteristic exponent α or skewness β of the α-stable distribution needs to be estimated to transmit the covert bits. We know that the estimation of the mean of a Gaussian-distributed stochastic process requires only a smaller number of samples to obtain a more accurate estimate. In contrast, the estimation of the parameters of the α-stable distribution requires much more samples. Thus, the complexity of our system is reduced and the transmission rate is increased.

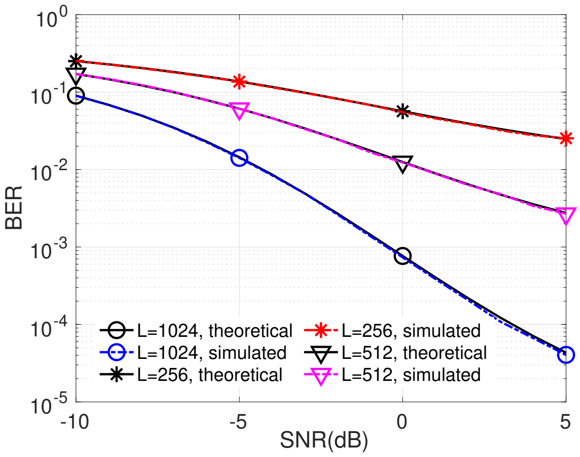


Fig. 4. The theoretical BER (solid line) and the simulated BER (dotted line), where $M = 1$, and $\kappa = 0.2$.

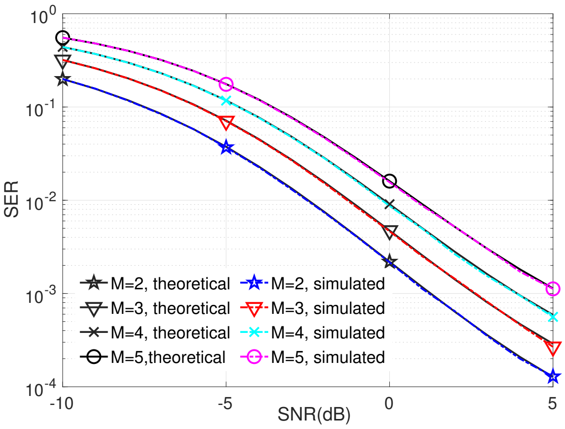


Fig. 5. Comparison of the theoretical SER obtained by numerical integration (solid line) and the SER obtained by simulation (dotted line) when M goes from 2 to 5, where $L = 1024$ and $\kappa = 0.2$.

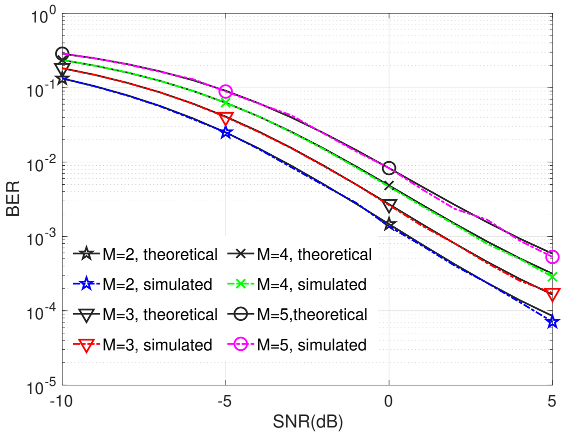


Fig. 6. Comparison of the theoretical BER obtained by numerical integration (solid line) and the BER obtained by simulation (dotted line) when M goes from 2 to 5, where $L = 1024$ and $\kappa = 0.2$.

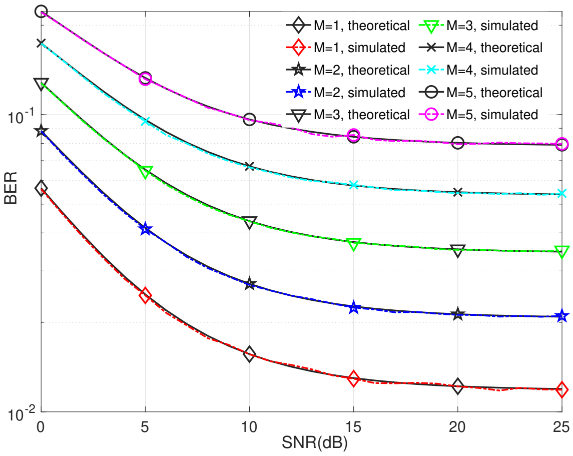


Fig. 7. Comparison of the theoretical BER obtained by numerical integration (solid line) and the BER obtained by simulation (dotted line) when M goes from 1 to 5, where $L = 256$ and $\kappa = 0.2$.

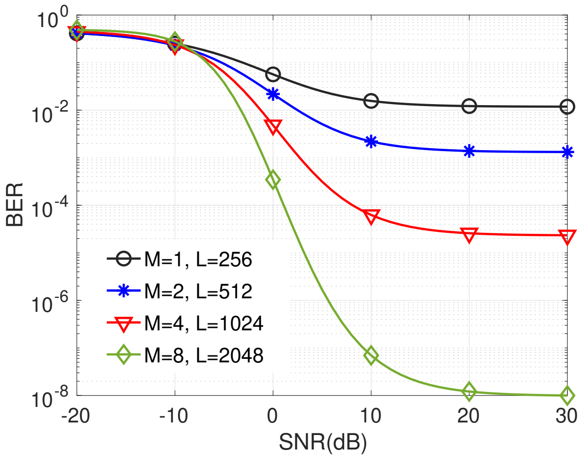


Fig. 8. Comparison of theoretical BER while keeping the transmission rate $\frac{M}{L}$ constant, where $\kappa = 0.2$.

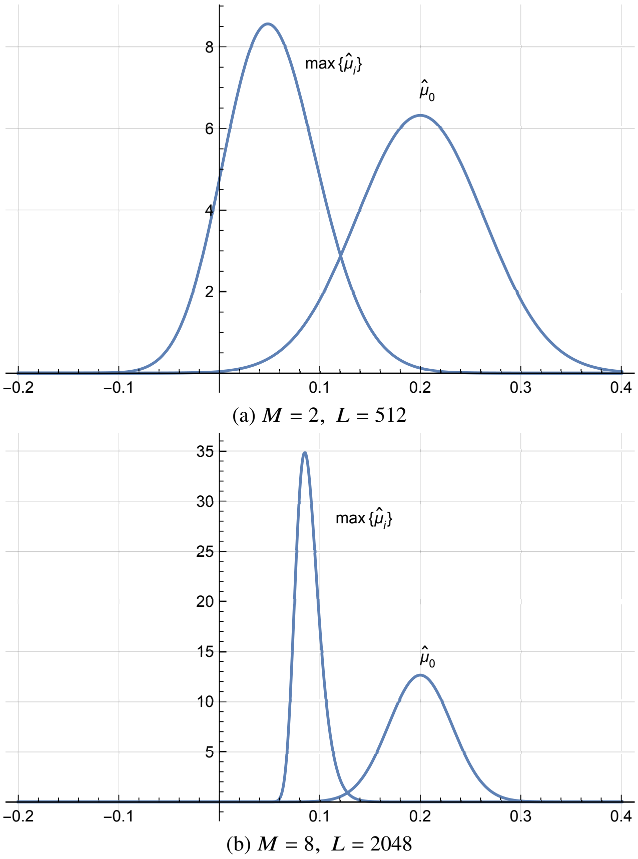
As shown in Fig. 5, the SER obtained from the theoretical derivation is still very close to the SER obtained from the simulation. Notice that as the number of bits contained in each modulation code increases (e.g., $M = 5$ indicates that a modulation code contains 5 covert bits), the SER becomes significantly larger, indicating an increased probability of error. However, it can be seen from this figure that at a higher SNR, e.g., 5 dB, the SER is about 0.1%. The SNR is low enough to still meet the needs of certain communication scenarios.

As shown in Fig. 6, the BER obtained from the theoretical derivation is very close to that obtained from the simulation, which verifies the correctness of the theoretical derivation. When M increases, its BER increases significantly, indicating an increase in the probability of bit error. At the same time, however, an increase in M indicates that more bits can be transmitted in the same period. As can be seen from the figure, at a higher SNR, say 5 dB, $M = 5$ corresponds to an erroneous bit rate of about 0.05%, which is about half of the SER, which indicates that (22) is correctly derived.

Other simulation conditions remain unchanged, only L is reduced from 1024 to 256, and the BER results obtained from the simulation and theory are shown in Fig. 7. Compared with Fig. 6, it is clear that when L decreases, its BER increases significantly, indicating an increase in the probability of bit errors. For example, with $M = 5$, the BER is about 10%, even though the SNR is as high as 25 dB. Despite reducing M from 5 to 1, its BER is still as high as about 1%. Thus, the length L has a significant effect on the error bit rate.

To compare the BER of different lengths of L, to make the comparison fair, the same number of covert bits are required to be transmitted at the same time, that is, $\frac{M}{L}$ remains unchanged. As shown in Fig. 8, the same four error bit rate curves for $\frac{M}{L} = 256$ are listed. It is easy to see from the figure that when the SNR is greater than -10 dB, the performance of using a long L (large M) is significantly higher than that of a short L (small M) by several orders of magnitude. This means that the parameter L has a greater effect on BER than the parameter M. This is because, as shown in Fig. 9, the variance of the random variable $\hat{\mu}_i$ decreases by $\sqrt{L}$ as L increases, making the PDF of the mean value estimate $\hat{\mu}_0$ when correctly demodulated more clearly distinguishable from that of the mean value estimate $\hat{\mu}_i$ when incorrectly demodulated. Thus, the BER is greatly reduced.

The BER performance curves under quasi-static fading channels are given in Fig. 10. We only simulate the BERs under a Rayleigh fading channel at $M = 1$. Comparison of the theoretical BER obtained from the numerical integration of (27) (solid line) with the BER obtained from simulations (dotted line) shows that the simulations and theoretical derivations are in perfect agreement. As in the previous simulations, increasing L quickly reduces the BER of the system. Further, comparing Fig. 4 and Fig. 10, we see that fading has a very large effect on the system, which also conforms to our intuitive knowledge. The reason for this is that when the fading coefficient $h < 1$, it leads to a decrease in the SNR. Conversely, it increases the SNR when $h > 1$. Further, for a Rayleigh fading distribution variable with parameter $\gamma = \sqrt{1/2}$, the probability of $h < 1$ is 63% and the probability of $h < 1/2$ is 22%. Even worse, $\mathrm{Erfc}(x)$ increases rapidly as the variable x decreases, resulting in the support of the integrated function in (27) to be on $[0, 1]$, except when the SNR is extremely low, e.g., -10 dB. As a result, fading leads to a dramatic increase in BER.



(a) $M = 2,\ L = 512$

(b) $M = 8,\ L = 2048$

Fig. 9. Comparison of the PDF of the mean estimate $\hat{\mu}_0$ for correct demodulation and that of the mean estimate $\max\{\hat{\mu}_i\}$ for incorrect demodulation, with the SNR of $\rho = 1$.

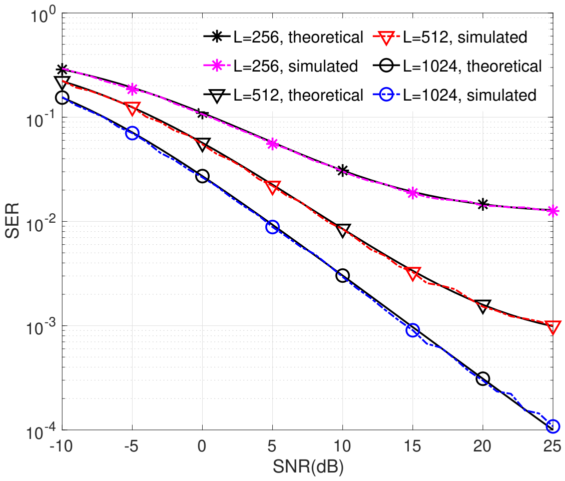


Fig. 10. SER performance under quasi-static fading channel in the case $M = 1$, where $\kappa = 0.2$ and $L = 256, 512, 1024$, respectively.

5. Conclusion

In this study, we propose a covert communication system in which a non-zero-mean normally distributed random process is used as a carrier, and a Walsh code carrying a covert bit stream is used as a modulation code, so that the statistical properties of the transmitted signal are indistinguishable from the environmental noise, thus achieving the purpose of covert communication. The demodulation process at the receiver is also very simple. The covert bit stream is recovered by multiplying the received signal with the Walsh modulation code, calculating the mean value, and finding the maximum mean value. The proposed system, working at the physical layer, has good security, flexibility, and BER performance, and is very suitable for IoT devices with limited resources and low transmission rate requirements but high concealability requirements. The disadvantage of this study is that it requires significant changes in the signal processing of the existing IoT devices at the physical layer (PHY). As a next step, we intend to use a software-defined radio platform like “HackRF One” to implement and validate the proposed covert system.

Acknowledgments

The authors thank to anonymous reviewers for their time and effort spent in evaluating our manuscript and providing us with constructive comments to improve the quality of the manuscript. This work was supported in part by “Pioneer” and “Leading Goose” R&D Program of Zhejiang (Grant No. 2023C01030), in part by the National Natural Science Foundation of China (Grant No. 62271445), and part by the open project of Guangdong Provincial Key Laboratory of Advanced Wireless Communications (Southern University of Science and Technology, Shenzhen 518055, China).

References

[1] SHEN, Y., ZHANG Y., JIANG, X. *Secrecy, Covertness and Authentication in Wireless Communications: Physical Layer Security Approach*. Cham (Switzerland): Springer, 2023. ISBN: 9783031384646

[2] YANG, N., WANG, L., GERACI, G., et al. Safeguarding 5G wireless communication networks using physical layer security. *IEEE Communications Magazine*, 2015, vol. 53, no. 4, p. 20–27. DOI: 10.1109/MCOM.2015.7081071

[3] MENG, R., CUI, Q., ZHOU, Z., et al. A steganography algorithm based on CycleGAN for covert communication in the Internet of Things. *IEEE Access*, 2019, vol. 7, p. 90579–90584. DOI: 10.1109/ACCESS.2019.2920956

[4] YENER, A., ULUKUS, S. Wireless physical-layer security: Lessons learned from information theory. *Proceedings of the IEEE*, 2015, vol. 103, no. 10, p. 1814–1825. DOI: 10.1109/JPROC.2015.2459592

[5] CHENG, X., AN, J., XIONG, Z., et al. Covert communications: A comprehensive survey. *IEEE Communications Survey & Tutorials*, 2023, vol. 25, no. 2, p. 1171–1198. DOI: 10.1109/COMST.2023.326391

[6] LIU, Z. H., LIU, J. J., ZENG, Y., et al. Covert wireless communications in IoT systems: Hiding information in interference. *IEEE Wireless Communications*, 2018, vol. 25, no. 6, p. 46–52. DOI: 10.1109/MWC.2017.1800070

[7] ILLI, E., QARAQE, M., ALTHUNIBAT, S., et al. Physical layer security for authentication, confidentiality, and malicious node detection: A paradigm shift in securing IoT networks. *IEEE Communications Surveys & Tutorials*, 2024, vol. 26, no. 1, p. 347–388. DOI: 10.1109/COMST.2023.3327327

[8] PAKRAVAN, S., CHOUINARD, J. Y., LI, X. W., et al. Physical layer security for NOMA systems: Requirements, issues, and recommendations. *IEEE Internet of Things Journal*, 2023, vol. 10, no. 24, p. 21721–21737. DOI: 10.1109/JIOT.2023.3296319

[9] JAISWAL, N., PANDEY, A., YADAV, S., et al. Physical layer security performance of NOMA-aided vehicular communications over Nakagami-*m* time-selective fading channels with channel estimation errors. *IEEE Open Journal of Vehicular Technology*, 2023, vol. 4, p. 72–100. DOI: 10.1109/OJVT.2022.3222187

[10] SHU, F., WANG, J. Z. *Intelligent Reflecting Surface-Aided Physical-Layer Security*. Cham (Switzerland): Springer, 2023. ISBN: 9783031418112

[11] ZANDER, S., ARMITAGE, G., BRANCH, P. A survey of covert channels and countermeasures in computer network protocols. *IEEE Communications Surveys & Tutorials*, 2007, vol. 9, no. 3, p. 44–57. DOI: 10.1109/COMST.2007.4317620

[12] AZERANG, S. F., SAMSAMI KHODADAD, F., FOROUZESH, M. Covert communication based on energy harvesting and cooperative jamming. *Wireless Networks*, 2022, vol. 28, p. 3729–3738. DOI: 10.1007/s11276-022-03082-x

[13] TA, H. Q., PHAM, Q. V., KHUONG, H. V., et al. Covert communication with noise and channel uncertainties. *Wireless Networks*, 2022, vol. 28, p. 161–172. DOI: 10.1007/s11276-021-02828-3

[14] SHEN, Y. L., ZHANG, Y. Y., JIANG, X. H. *Secrecy, Covertness and Authentication in Wireless Communications: Physical Layer Security Approach*. Cham (Switzerland): Springer, 2023. ISBN: 9783031384646

[15] CHEN, X., ZHENG, T. X., DONG, L. M., et. al. Enhancing MIMO covert communications via intelligent reflecting surface. *IEEE Wireless Communications Letters*, 2022, vol. 11, no. 1, p. 33–37. DOI: 10.1109/LWC.2021.3119687

[16] HU, L. T., BI, S. J., LIU, Q. J., et. al. Intelligent reflecting surface aided covert wireless communication exploiting deep reinforcement learning. *Wireless Networks*, 2023, vol. 29, p. 877–899. DOI: 10.1007/s11276-022-03037-2

[17] LI, M., TAO, X. F., LI, N., et. al. Energy-efficient covert communication with the aid of aerial reconfigurable intelligent surface. *IEEE Communications Letters*, 2022, vol. 26, no. 9, p. 2101–2105. DOI: 10.1109/LCOMM.2022.3183637

[18] HUANG, S. H., LIU, W. W., LIU, G. J., et al. A correlation-based approach to detecting wireless physical covert channels. *Computer Communications*, 2021, vol. 176, p. 31–39. DOI: 10.1016/J.COMCOM.2021.05.017

[19] HU, M. Y., QIAO, S., JI, X. P. Wireless covert communication with polarization dirty constellation. *Applied Sciences*, 2023, vol. 13, no. 6, p. 1–17. DOI: 10.3390/APP13063451

[20] MA, S., ZHANG, Y. Q., SHENG, H. H., et. al. Optimal probabilistic constellation shaping for covert communications. *IEEE Transactions on Information Forensics and Security*, 2022, vol. 17, p. 3165–3178. DOI: 10.1109/TIFS.2022.3203310

[21] WANG, Y. D., YAN, S. H., YANG, W. W., et. al. Probabilistic accumulate-then-transmit in wireless-powered covert communications. *IEEE Transactions on Wireless Communications*, 2022, vol. 21, no. 12, p. 10393–10406. DOI: 10.1109/TWC.2022.3183892

[22] CEK, M. E., SAVACI, F. A. Stable non-Gaussian noise parameter modulation in digital communication. *Electronics Letters*, 2009, vol. 45, no. 24, p. 1256–1257. DOI: 10.1049/EL.2009.2280

[23] CEK, M. E. Covert communication using skewed α-stable distributions. *Electronics Letters*, 2015, vol. 51, no. 1, p. 116–118. DOI: 10.1049/EL. 2014.3323.

[24] XU, Z. J., GONG, Y., WANG, K., et al. Covert digital communication systems based on joint normal distribution. *IET Communications*, 2017, vol. 11, no. 8, p. 1282–1290. DOI: 10.1049/iet-com.2016.1333

[25] XU, Z. J., LU, W. D., GONG, Y., et al. A covert communication system using non-zero mean normal distributions. *Radioengineering*, 2020, vol. 29, no. 3, p. 580–588. DOI: 10.13164/RE.2020.0580

[26] BASAR, E. Noise modulation. *IEEE Wireless Communications Letters*, 2024, vol. 13, no. 3, p. 844–848. DOI: 10.1109/LWC.2023.3346471

About the Authors . . .

Zhijiang XU (corresponding author) was born in 1973. He received his Ph.D. degree in Information and Communication Engineering in 2005, from Zhejiang University, China. He held an appointment as Associate Professor in the College of Information Engineering at Zhejiang University of Technology, China, from 2007 to 2019. Since 2019, he has joined the Zhejiang Institute of Mechanical & Electrical Engineering in the School of Automation and was promoted to full Professor in 2020. His research interests include digital communications over fading channels, channel modeling, coding, digital synchronization, etc.

Jingyu HUA was born in Zhejiang province, China in 1978. He received his B.S. and M.S. degrees in Electronic Engineering from the South China University of Technology, Guangzhou, China, in 1999 and 2002. Then in 2006, he received a Ph.D. degree in Electronic Engineering from Southeast University, Nanjing, China. In 2006, he joined the Zhejiang University of Technology as an Associate Professor in the Electronic Engineering Department and was promoted to full Professor in 2012. Since 2019, he has been with Zhejiang Gongshang University as a distinguished Professor. He is the author of more than 200 articles and more than 20 inventions. His research interests include the areas of parameter estimation, channel modeling, wireless localization, and digital filtering in wireless communications.

Appendix A: SER when $M > 1$

For $M > 1$, the symbol decision is error when $\hat{\mu}_0 < \max\{\hat{\mu}_i, i = 1, \cdots, 2^M - 1\}$, so the probability of error is $\Pr\left(\hat{\mu}_0 < \max_{i=1}^{2^M-1}\{\hat{\mu}_i\}\right)$. Let $Z = \max_i\{\hat{\mu}_i\}$, and noting that $\{\hat{\mu}_i\}$ are independent of each other, then we have

$$\begin{aligned}\Pr(Z \le z) &= \Pr\left(\hat{\mu}_1 \le z, \cdots, \hat{\mu}_{2^M-1} \le z\right)\\ &= \prod_{i=1}^{2^M-1} \Pr(\hat{\mu}_i \le z) = \left(\Pr(\hat{\mu}_1 \le z)\right)^{2^M-1}\\ &= \left(1 - \frac{1}{2}\mathrm{Erfc}\left(\frac{z}{\sqrt{2}\sigma}\right)\right)^{2^M-1}.\end{aligned} \tag{A1}$$

Thus, derivation of (A1) in terms of the variable z yields the PDF of Z,

$$\begin{aligned}f_Z(z) &= (2^M - 1) f_{\mu_1}(z)\left(1 - \frac{1}{2}\mathrm{Erfc}\left(\frac{z}{\sqrt{2}\sigma}\right)\right)^{2^M-2}\\ &= \frac{2^M - 1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{z^2}{2\sigma^2}}\left(1 - \frac{1}{2}\mathrm{Erfc}\left(\frac{z}{\sqrt{2}\sigma}\right)\right)^{2^M-2}.\end{aligned} \tag{A2}$$

Clearly, when $M = 1$, $Z = \hat{\mu}_1$, and the PDF of Z degenerates to $f_Z(z) = \frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{z^2}{2\sigma^2}}$, which is the PDF of $\hat{\mu}_1$.

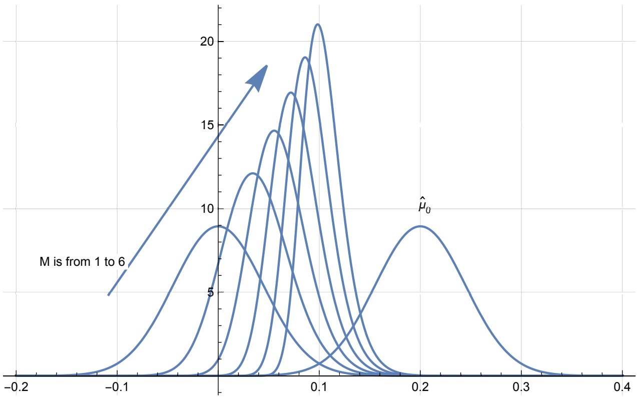


Fig. A.1. The PDFs of $\hat{\mu}_0$ and Z when M ranges from 1 to 6, $L = 1024$, $\kappa = 0.2$, and $\rho = 1$.

Figure A.1 shows the PDF of Z as M goes from 1 to 6, as well as the PDF of $\hat{\mu}_0$ for correct demodulation. It can be seen from the figure that as M gets larger, the PDF of Z gets closer and closer to the PDF of $\hat{\mu}_0$. Therefore, the SER must increase.

As with the derivation of the SER when $M = 1$, let $\Lambda = \hat{\mu}_0 - Z$, then the SER is $\Pr(\Lambda < 0)$. Since $\hat{\mu}_0$ and Z are independent of each other, the PDF of Λ is

$$\begin{aligned}f_\Lambda(\lambda) &= \int_{-\infty}^{\infty} f_{\hat{\mu}_0}(\lambda + z) f_Z(z)\mathrm{d}z\\ &= \int_{-\infty}^{\infty} \frac{2^M - 1}{2\pi\sigma^2}\left(1 - \frac{1}{2}\mathrm{Erfc}\left(\frac{z}{\sqrt{2}\sigma}\right)\right)^{2^M-2}\mathrm{e}^{-\frac{(z+\lambda-\kappa)^2+z^2}{2\sigma^2}}\mathrm{d}z.\end{aligned} \tag{A3}$$

Thus, the SER is

$$\begin{aligned}r_\mathrm{s} &= \int_{-\infty}^{0} f_\Lambda(\lambda)\mathrm{d}\lambda = \int_{-\infty}^{0}\int_{-\infty}^{\infty} f_{\hat{\mu}_0}(\lambda + z) f_Z(z)\mathrm{d}z\mathrm{d}\lambda\\ &= \int_{-\infty}^{\infty} f_{\hat{\mu}_0}(\lambda + z)\left[\int_{-\infty}^{0} f_{\hat{\mu}_0}(\lambda + z)\mathrm{d}\lambda\right]\mathrm{d}z\\ &= \int_{-\infty}^{\infty} \frac{2^M - 1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{z^2}{2\sigma^2}}\left(1 - \frac{1}{2}\mathrm{Erfc}\left(\frac{z-\kappa}{\sqrt{2}\sigma}\right)\right)\\ &\qquad\times\left(1 - \frac{1}{2}\mathrm{Erfc}\left(\frac{z}{\sqrt{2}\sigma}\right)\right)^{2^M-2}\mathrm{d}z.\end{aligned} \tag{A4}$$

An alternative derivation of the SER is given below. Let the probability of a correct demodulation be $r_\mathrm{c} = \Pr(\hat{\mu}_i \le \hat{\mu}_0, \forall i)$. We notice that $\{\hat{\mu}_i\}$ are independent of each other, but $\hat{\mu}_i \le \hat{\mu}_0$ and $\hat{\mu}_j \le \hat{\mu}_0$ are not independent of each other. Applying a bit of trickery in probability calculations and a little abuse of notation, we have

$$\begin{aligned}r_\mathrm{c} &= \Pr(\hat{\mu}_i \le \hat{\mu}_0, \forall i)\\ &= \sum_{\hat{\mu}_0}\left(\prod_{i=1}^{2^M-1}\Pr\left(\hat{\mu}_i \le \hat{\mu}_0 | \hat{\mu}_0 = z\right)\right)\Pr(\hat{\mu}_0 = z)\\ &= \int_{\hat{\mu}_0}\left(\prod_{i=1}^{2^M-1}\Pr(\hat{\mu}_i \le z)\right) f_{\hat{\mu}_0}(z)\mathrm{d}z\\ &= \int_{\hat{\mu}_0}\left(\Pr(\hat{\mu}_1 \le z)\right)^{2^M-1} f_{\hat{\mu}_0}(z)\mathrm{d}z\\ &= \int_{-\infty}^{\infty}\left(1 - \frac{1}{2}\mathrm{Erfc}\left(\frac{z}{\sqrt{2}\sigma}\right)\right)^{2^M-1}\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(z-\kappa)^2}{2\sigma^2}}\mathrm{d}z\\ &= \int_{-\infty}^{\infty}\left(1 - \frac{1}{2}\mathrm{Erfc}\left(\frac{z+\kappa}{\sqrt{2}\sigma}\right)\right)^{2^M-1}\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{z^2}{2\sigma^2}}\mathrm{d}z,\end{aligned} \tag{A5}$$

then the SER is $r_\mathrm{s} = 1 - r_\mathrm{c}$. Numerical calculations show that (A4) and (A5) yield identical results.