Optimal Probabilistic Constellation Shaping for Covert Communications

Shuai Ma, Yunqi Zhang, Haihong Sheng, Hang Li, Member, IEEE, Jia Shi, Member, IEEE, Long Yang, Member, IEEE, Youlong Wu, Member, IEEE, Naofal Al-Dhahir, Fellow, IEEE, and Shiyin Li

Abstract—In this paper, we investigate the optimal probabilistic constellation shaping design for covert communication systems from a practical view. Different from conventional covert communications with equiprobable constellations modulation, we propose non-equiprobable constellations modulation schemes to further enhance the covert rate. Specifically, we derive covert rate expressions for practical discrete constellation inputs for the first time. Then, we study the covert rate maximization problem by jointly optimizing the constellation distribution and power allocation. In particular, an approximate gradient descent method is proposed for obtaining the optimal probabilistic constellation shaping. To strike a balance between the computational complexity and the transmission performance, we further develop a framework that maximizes a lower bound on the achievable rate where the optimal probabilistic constellation shaping problem can be solved efficiently using the Frank-Wolfe method.

Extensive numerical results show that the optimized probabilistic constellation shaping strategies provide significant gains in the achievable covert rate over the state-of-the-art schemes.

Index Terms—Covert communications, probabilistic constellation shaping, achievable rate.

I. INTRODUCTION

Radio frequency (RF) based wireless communication is inevitably susceptible to eavesdropping due to the broadcast nature of the electromagnetic waves. With the ever-growing Internet of Things (IoT) applications, the security issue has become more and more crucial in the future sixth-generation (6G) wireless networks [1]. For example, in large enterprise buildings, hospitals, factories, communication may be very sensitive to a hostile adversary. The conventional cryptography approaches [2] usually focus on protecting the transmission content or increasing message decoding complexity, while the physical-layer security [3], [4] approaches exploit the intrinsic wireless fading channels properties to minimize the information leakage to the eavesdroppers. In fact, a higher level of security is to hide the existence of the communication, which not only can be applied in all aforementioned application scenarios [5], [6], [7], [8], but also meets more critical demands from military or security agencies. To address this high level security, covert communications [9], [10], which shields the existence of message transmissions against the detection of a warden, are emerging as a cutting-edge wireless communication technique, and have recently attracted significant research attention. Note that, covert communication aims to hide the communication behavior from the eavesdropper, while physical layer security tries to reduce the interception information of the eavesdropper.

The basic idea of covert communications is as follows. The legitimate transmitter (Alice) transmits messages to the paired receiver (Bob), while guaranteeing a low detection probability for a Warden (Willie). Although such idea has been realized by spread-spectrum techniques [11] for several decades, the information-theoretic limits of covert communications, which is also referred to as low probability of detection (LPD) communications in some literature, were only recently derived [10], [12], [13], [14]. In particular, the authors in [12] firstly demonstrated that in additive white Gaussian noise (AWGN) channels, Alice can reliably send at most $O(\sqrt{n})$ bits to Bob in $n$ channel usages under the covert requirement. Such result is also called the square root law (SRL). Subsequent
works have extended this result to various channel models such as binary symmetric channels [15], broadcast communications [16], multiple access channels [17], and interference channels [18].

Although the SRL indicates that the asymptotic achievable rate under the covert requirement approaches zero, many researchers have shown that the SRL limit can be beaten by exploiting additional techniques in the considered covert communication scenario. These methods include: taking advantage of the ignorance of the transmission time at Willie [19]; applying an intelligent reflecting surface [20], [21], [22]; exploring the molecular absorption or scattering feature of the Terahertz spectrum [23], [24]; cooperative jamming [25] or uninformed jamming [26], [27], [28], [29]; jointly optimizing the beam training and data transmission for millimeter-wave communication [30]; exploiting the uncertainty noise power or channel state information (CSI) at Willie [31], [32], [33], [34], [35]; robust beamforming design [36], [37], over a finite number of channel uses [38]; applying a full-duplex transceiver [39], [40], [41], [42]; intermediate relay [43], [44] or exploring unmanned aerial vehicle (UAV) as mobile relay [45], [46].

To be more specific, it is shown in [19] that Alice can covertly transmit $O \left( \min \left\{ n, \sqrt{n \log T(n)} \right\} \right)$ bits to Bob. When Willie lacks the knowledge of his noise power, Alice can reliably transmit $O(n)$ bits [31], [32], [33]. With the aid of an uninformed jammer [26], Alice can also achieve positive transmission rate. With a finite number of channel uses, a uniformly distributed power allocation scheme was proposed [38] to enhance the covert transmission. In [39], the authors showed that, the effective throughput under delay constraints can be improved by adding artificial noise (AN) at the full-duplex receiver. For a one-way relay network, the authors in [43] studied the performance limits of covert communications of an energy harvesting relay. In [44], a multiple-relay network was considered, and both the maximum throughput and the minimum end-to-end delay routing algorithms were developed with multiple Willies. In [36], two probabilistic metrics, called the covert outage probability and the connectivity probability, were analyzed for multi-antenna covert communications with randomly located wardens and interferers. Using Kullback-Leibler divergence, i.e., $D(p_1||p_0)$ or $D(p_0||p_1)$ to measure the covertness, a Gaussian input distribution was shown to be optimal for the covert metric $D(p_1||p_0)$, and not optimal for the covert metric $D(p_0||p_1)$ [9], where $p_0$ and $p_1$ represent Willie’s received signal distributions when covert communications occur and not occur, respectively.

The aforementioned research advances in covert communications mainly make the assumption of a Gaussian input distribution at the transmitter side, which can hardly be realized in practical communication systems. In fact, the information symbols in practical communication systems are realized in the form of discrete constellation points, i.e., finite alphabet inputs, such as pulse amplitude modulation (PAM) and multiple-quadrature amplitude modulation (M-QAM). In [47], both the lower bound of covert transmission probability and throughput maximization have been analyzed with discrete constellation inputs. The discrete constellation points are assumed to be equally likely, which is not optimal for practical covert communications, especially for high-order modulation schemes. So far, the optimal discrete constellation inputs of covert communication are still not well discussed in the literature.

Motivated by the above background, we develop information-theoretic limits of covert communications with probabilistic constellation shaping. First, we derive the achievable rate expression of the system with the discrete constellation input signals, rather than the Gaussian inputs adopted in most of the existing works. Then, we investigate the performance with the optimal input distribution. Our results provide a practical design framework for covert communication systems. The main contributions of this paper are summarized as follows:

- Generally, the inputs of practical communications systems follow a finite-set discrete distribution rather than a Gaussian distribution. To evaluate performance, we derive the achievable rate expressions for an arbitrary discrete distributed input. Comparing to the existing rate expressions with equiprobable discrete constellation points, the derived expressions are more general and practical. Since the derived rate expression is not in closed-form, we further derive both lower and upper bounds. All these results can be used as performance metrics for the considered covert communication system.

- Furthermore, we design optimal discrete constellation inputs to maximize the exact covert rate under the covertness constraints, transmit power limitations, and the signal distribution requirements, which is a challenging problem since neither the exact covert rate nor the covertness constraint has an analytical expression. To efficiently solve it, we conservatively transform the covertness constraint into its upper bound with closed-form expression. Then, we adopt the numerical integration method to approximate the covert rate objective function and its gradient. Afterwards, the optimal probability distributions of the discrete constellation are calculated by the approximate gradient descent method, where the step sizes are calculated by the backtracking line search.

- To reduce the computation complexity of the design problem, we further adopt the derived lower bound as the covert rate performance metric. To overcome the non-convexity challenge, this problem is iteratively solved by the proposed Frank-Wolfe method.

The rest of this paper is organized as follows. The system model and the derivation of Bob’s achievable rate are presented in Section II. The optimal probabilistic constellation shaping design for covert communications is provided in Section III. The probabilistic constellation shaping design and its approximations are presented in Section IV. In Section V, we evaluate the proposed probabilistic constellation shaping design using numerical results. Finally, we conclude the paper in Section VI.

Notations: The vectors and matrices are represented by boldfaced lowercase and uppercase letters, respectively. The notations $(\cdot)^*$, $E(\cdot)$, $\|\cdot\|$, $\text{Tr} (\cdot)$, $\text{Re} (\cdot)$ and $\text{Im} (\cdot)$ represent
the conjugate, the expectation, Frobenius norm, trace, the real part and imaginary part of its argument, respectively. And $\odot$ is Hadamard Product, i.e., $A_{m\times n}[a_{ij}] \odot B_{m \times n}[b_{ij}] = C_{m \times n}[a_{ij}b_{ij}]$. The operator $A \succeq 0$ means $A$ is positive semi-definite. The notation $CN(\mu, \sigma^2)$ denotes a complex-valued circularly symmetric Gaussian distribution with mean $\mu$ and variance $\sigma^2$.

II. SYSTEM MODEL

Consider a typical covert communication scenario as illustrated in Fig. 1, where Alice and Bob are a legitimate communication pair, and Willie is the eavesdropper. Each one of them is equipped with a single antenna. Let $g_b \sim CN(0, \sigma_b^2)$ and $g_w \sim CN(0, \sigma_w^2)$ denote the Rayleigh flat fading channel from Alice to Bob and Willie, respectively [48], where $\sigma_b^2$ and $\sigma_w^2$ are the variances of $g_b$ and $g_w$. Let $x[i]$ denote Alice’s transmitted symbol at the $i$-th channel use, where $i = 1, \ldots, N$, and $N$ is the total number of channel uses.

A. Achievable Rate of Bob

As it is the case in practice, $x[i] \in \Omega$ follows a discrete constellation distribution instead of Gaussian distribution. Here, $\Omega$ denotes a discrete constellation set with $K$ discrete points $\{x_k\}_{1 \leq k \leq K}$, i.e.,

$$\Omega = \left\{ X \mid \Pr(X = x_k) = p_k \geq 0, \sum_{k=1}^K p_k = 1, \sum_{k=1}^K p_k |x_k|^2 \leq P_A, x_k \in \mathbb{C}, k = 1, \ldots, K \right\}, \quad (1)$$

where $x_k$ denotes the $k$th discrete point, $p_k$ denotes the corresponding probability, and $P_A$ denotes the average power.

For the $i$-th channel use, the received signal at Bob $y_b[i]$ is given as

$$y_b[i] = g_b^* x[i] + z_b[i], \quad (2)$$

where $z_b[i] \sim CN(0, \sigma_b^2)$ denotes the received noise at Bob. Since $x[i] \in \Omega$, then the likelihood functions of $y_b[i]$ is given as

$$p(y_b) = \frac{1}{\pi \sigma_b^2} \sum_{k=1}^K p_k \exp\left(-\frac{|y_b - g_b^* x_k|^2}{\sigma_b^2}\right). \quad (3)$$

Therefore, given the discrete constellation, the achievable rate of Bob $R_b$ is given by

$$R_b = I(y_b[i]; x[i]) \quad (4a)$$
$$= h(y_b[i]) - h(z_b[i]) \quad (4b)$$
$$= - \int_{-\infty}^{\infty} p(y_b) \log_2 p_g(y_b) \, dy_b - \log_2 \pi e \sigma_b^2 \quad (4c)$$
$$= - \sum_{k=1}^K p_k \mathbb{E}_{z_b} \left\{ \log_2 \sum_{j=1}^K p_j \exp\left(-\frac{|g_b^* (x_k - x_j) + z_b|^2}{\sigma_b^2}\right) \right\} - \frac{1}{\ln 2}. \quad (4d)$$

where $h(X) = \int f(x) \log f(x) \, dx$ denotes the differential entropy, and $f(x)$ represents the probability density function (PDF).

Based on the achievable rate expression in (4), we will further investigate the optimal probability of discrete constellations for covert communications.

B. Hypothesis Testing

According to the received signals, Willie attempts to decide whether Alice covertly transmits information to Bob or not by performing an optimal statistical hypothesis test (such as the Neyman-Pearson test). Specifically, willie needs to distinguish between two hypotheses: 1) the null hypothesis $H_0$ indicating no transmission; 2) the hypothesis $H_1$ indicating the transmission. Let $y_w[i]$ denote the received signal at Willie in the $i$-th channel use. Under the two hypotheses, the signal received at Willie is given as

$$H_0 : y_w[i] = z_w[i], \quad (5a)$$
$$H_1 : y_w[i] = g_w^* x[i] + z_w[i], \quad (5b)$$

where $z_w[i] \sim CN(0, \sigma_w^2)$ denotes the received noise at Willie. Let $D_1$ and $D_0$, respectively, denote the binary decisions of Willie. Thus, the total detection error probability of Willie is defined as [9], [49], [50]

$$\zeta = \Pr(D_1|H_0) + \Pr(D_0|H_1). \quad (6)$$

Note that, $\Pr(D_1|H_0)$ denotes the false alarm probability that Willie believes $H_1$ when Alice does not transmit, and $\Pr(D_0|H_1)$ denotes the missed detection probability that Willie decides $H_0$ when Alice transmits. Moreover, let $p_{y,0} = f(y_w|H_0)$ and $p_{y,1} = f(y_w|H_1)$ denote the likelihood functions of $y_w$ under $H_0$ and $H_1$, respectively. According to (5), we have

$$p_{y,0} = \frac{1}{\pi \sigma_w^2} \exp\left(-\frac{|y_w|^2}{\sigma_w^2}\right), \quad (7a)$$
$$p_{y,1} = \frac{1}{\pi \sigma_w^2} \sum_{k=1}^K p_k \exp\left(-\frac{|y_w - g_w^* x_k|^2}{\sigma_w^2}\right). \quad (7b)$$

Let $V_T(p_{y,0}, p_{y,1}) = \frac{1}{2} \|p_{y,0} - p_{y,1}\|_1$ denote the total variation distance between $p_{y,0}$ and $p_{y,1}$. According to
Theorem 13.1.1 in [50], the optimal detection error probability of Willie is given as
\[ \xi^{\text{opt}} = 1 - V_T(p_{y,0} \| p_{y,1}) = 1 - \frac{1}{2} \|p_{y,0} - p_{y,1}\|_1. \]  

(8)

However, in general, \( V_T(p_{y,0} \| p_{y,1}) \) is difficult to analyze. To address this issue, we apply Pinsker’s inequality [49] to obtain an upper bound
\[ V_T(p_{y,0} \| p_{y,1}) \leq \frac{1}{2} D(p_{y,0} \| p_{y,1}), \]  
\[ V_T(p_{y,0} \| p_{y,1}) \leq \frac{1}{2} D(p_{y,1} \| p_{y,0}), \]  

(9) (10)

where \( D(p_{y,0} \| p_{y,1}) = \int p_{y,0} \log \frac{p_{y,0}}{p_{y,1}} dP_{y,w} \) denotes the Kullback-Leibler (KL) divergence from \( p_{y,0} \) to \( p_{y,1} \), and \( D(p_{y,1} \| p_{y,0}) = \int p_{y,1} \log \frac{p_{y,1}}{p_{y,0}} dP_{y,w} \) denotes the KL divergence from \( p_{y,1} \) to \( p_{y,0} \).

Based on the likelihood functions of \( y_w \) in (7), \( D(p_{y,0} \| p_{y,1}) \) and \( D(p_{y,1} \| p_{y,0}) \) are, respectively, given as
\[ D(p_{y,0} \| p_{y,1}) = \frac{1}{\ln 2} - \mathbb{E}_{z_w} \left\{ \log_2 \sum_{k=1}^{K} p_k \exp \left( \frac{-|z_w - g_w x_k|^2}{\sigma_w^2} \right) \right\}, \]  
\[ D(p_{y,1} \| p_{y,0}) = \sum_{k=1}^{K} p_k \mathbb{E}_{z_w} \left\{ \log_2 \sum_{j=1}^{K} p_j \exp \left( \frac{-|g_w (x_k - x_j) + z_w|^2}{\sigma_w^2} \right) \right\} + \frac{|g_w^2 p_A|}{(\ln 2) \sigma_w^2} + \frac{1}{\ln 2}. \]  

(11a) (11b)

Covert communication is achieved for a given \( \epsilon \) if the detection error probability \( \xi \) is no less than \( 1 - \epsilon \), i.e.,
\[ \xi \geq 1 - \epsilon, \epsilon \in [0, 1], \]  

(12)

where \( \epsilon \) is a small number determining the required covertness level.

Therefore, to achieve covert communication with the given \( \epsilon \), i.e., \( \xi \geq 1 - \epsilon \), the KL divergences of the likelihood functions should satisfy one of the following constraints:
\[ D(p_{y,0} \| p_{y,1}) \leq 2 \epsilon^2, \]  
\[ D(p_{y,1} \| p_{y,0}) \leq 2 \epsilon^2. \]  

(13a) (13b)

Both of the above constraints can meet the requirements of covert communication, but these two constraints are not exactly the same, as we discuss next.

### III. Optimal Signaling Design Under Covert Constraints

In this section, we investigate the design of optimal probability of discrete constellation points for covert transmission with covertness constraints \( D(p_{y,0} \| p_{y,1}) \leq 2 \epsilon^2 \) or \( D(p_{y,1} \| p_{y,0}) \leq 2 \epsilon^2 \) [9], [10], [12], [13].

#### A. Case of \( D(p_{y,0} \| p_{y,1}) \leq 2 \epsilon^2 \)

For the probabilistic constellation shaping scheme, we aim to maximize the achievable rate of Bob \( R_b \) by optimizing the distribution of discrete constellation inputs, while satisfying both the covert transmission constraint and the discrete distribution constraint. Mathematically, the covert discrete constellation input optimization problem can be formulated as follows
\[ \max_{\{p_k\}} R_b \]  
\[ \text{s.t.} \quad D(p_{y,0} \| p_{y,1}) \leq 2 \epsilon^2, \]  
\[ \Pr(X = x_k) = p_k \geq 0, \]  
\[ \sum_{k=1}^{K} p_k |x_k|^2 \leq P_A, \]  
\[ \sum_{k=1}^{K} p_k = 1, k = 1, \ldots, K. \]  

(14a) (14b) (14c) (14d) (14e)

Since the KL-divergence \( D(p_{y,0} \| p_{y,1}) \) in (11b) is not an analytical expression, constraint (14b) is intractable. To circumvent this, we derive an explicit upper bound for the KL-divergence \( D(p_{y,0} \| p_{y,1}) \), which is given by
\[ D_U(p_{y,0} \| p_{y,1}) = -\log_2 \sum_{k=1}^{K} p_k \exp \left( -\frac{|g_w x_k|^2}{\sigma_w^2} \right). \]  

(15)

The details of derivations for (15) can be found in Appendix A. Based on (15), problem (14) can be rewritten as
\[ \max_{\{p_k\}} R_b \]  
\[ \text{s.t.} \quad -\log_2 \sum_{k=1}^{K} p_k \exp \left( -\frac{|g_w x_k|^2}{\sigma_w^2} \right) \leq 2 \epsilon^2, \]  
(16a) (16b)

In order to put problem (16) in a more concise form, we define the following variables
\[ x \equiv [x_1, \ldots, x_K]^T, \]  
\[ p \equiv [p_1, \ldots, p_K]^T, \]  
\[ q \equiv [\log_2 p^T \hat{q}_1, \ldots, \log_2 p^T \hat{q}_K]^T, \]  
\[ \hat{q} \equiv \left[ \exp \left( -\frac{|g_w^* (u_k - x_k)|^2}{\sigma_b^2} \right), \ldots, \exp \left( -\frac{|g_w^* (u_k - x_K)|^2}{\sigma_b^2} \right) \right]^T, \]  
\[ t \equiv \mathbb{E}_{z_w} \left( p^T q \right), \]  
\[ \phi (p) \equiv \mathbb{E}_{z_w} \left( p^T q \right). \]  

(17a) (17b) (17c) (17d) (17e) (17f)

Furthermore, the rate of Bob \( R_b \) and the upper bound of the KL-divergence \( D_U(p_0 \| p_1) \) can be, respectively, rewritten as follows
\[ R_b = -\phi (p) - \frac{1}{\ln 2}, \]  

(18a)
\[ D_U (p_{y,0} \| p_{y,1}) = -\log_2 p^T t. \] (18b)

Since \( \frac{1}{m^2} \) is constant and maximizing \( -\phi (p) \) is equivalent to minimizing \( \phi (p) \), problem (16) can be reformulated as

\[
\begin{align*}
\min_p & \quad \phi (p) \\
\text{s.t.} & \quad -\log_2 p^T t \leq 2\varepsilon^2, \\
& \quad p^T 1_K = 1, \\
& \quad p^T (x \otimes x) \leq P_A, \\
& \quad p \geq 0.
\end{align*}
\] (19)

where \( 1_K \) is a \( K \times 1 \) vector with all elements equal to 1.

Problem (19) is now a convex problem, and we adopt the gradient projection method to solve it. Specifically, let \( \nabla \phi (p) \) denote the gradient of the objective function (19a), which is given by

\[ \nabla \phi (p) = E_{z_b} [q + Qp] = \int_{-\infty}^{\infty} f_{z_b} (z_b) (q + Qp) dz_b. \] (20a)

Here, \( f_{z_b} (z_b) = \frac{1}{\pi \sigma_b^2} \exp \left( \frac{-|z_b|^2}{\sigma_b^2} \right) \) denotes the probability density function of \( z_b \), \( Q \triangleq \begin{bmatrix} Q_{ij} \end{bmatrix} \), where \( Q_{i,j} = \frac{q_i q_j}{q_i p_{in,2}} \), and \( e_i \) is the unit vector where the \( i \)th element is 1 and the other elements are 0.

However, neither the objective function \( \phi (p) \) or the gradient \( \nabla \phi (p) \) has an analytic expression. To tackle this challenge, we adopt the numerical integration method to approximate \( \phi (p) \) and \( \nabla \phi (p) \), i.e.,

\[
\begin{align*}
\tilde{\phi} (p) &= \int_{-r_1}^{r_1} f_{z_b} (z_b) \left( p^T q \right) dz_b, \\
\nabla \tilde{\phi} (p) &= \int_{-r_2}^{r_2} f_{z_b} (z_b) (q + Qp) dz_b,
\end{align*}
\] (21)

where \( [-r_1, r_1] \) and \( [-r_2, r_2] \) denote the integration intervals, by defining \( r_1 > 0, r_2 > 0 \). Furthermore, \( \phi (p) \) and \( \nabla \phi (p) \) denote the approximations of the objective function and its gradient, respectively.

Furthermore, let \( p_0 \) denote a feasible starting point, and \( p_n \) denote the \( n \)th iteration feasible point. With approximate gradient \( \nabla \tilde{\phi} (p) \), the gradient descent iteration step is given as

\[
\hat{p}_{n+1} = p_n - \alpha_n \nabla \tilde{\phi} (p_n),
\] (22)

where \( \alpha_n \in (0, 1] \) is the stepsize of the \( n \)th iteration. To choose a proper step size \( \alpha_n \), with a sufficient decrease, we adopt the backtracking line search algorithm, given in Algorithm 1.

Then, we project \( \hat{p}_{n+1} \) in the feasible region of problem (19). Specifically, when \( \hat{p}_{n+1} \) satisfies constraints (19b)-(19e), \( \hat{p}_{n+1} = \bar{p}_{n+1} \) is in the feasible region. Otherwise, we need to find the closest point \( p_{n+1} \) in the feasible region as the projection of \( \hat{p}_{n+1} \).

Mathematically, the projection operation of \( \hat{p}_{n+1} \) can be formulated as follows

\[
\begin{align*}
\min_{p_{n+1}} & \quad \| p_{n+1} - \hat{p}_{n+1} \| \\
\text{s.t.} & \quad -\log_2 p^T_{n+1} t \leq 2\varepsilon^2, \\
& \quad p^T_{n+1} 1_K = 1,
\end{align*}
\] (23)

where \( \bar{p}_{n+1} (x \otimes x) \leq P_A, \) (23d)

\[ p_{n+1} \geq 0. \] (23e)

Algorithm 1: Backtracking Line Search for Stepsize \( \alpha_n \)

1: **Input**: \( p_n, \tilde{\phi} (p_n), \nabla \tilde{\phi} (p_n) \) and \( \alpha > 0, \rho, \epsilon \in (0, 1); \)
2: **Update** \( \hat{p}_{n+1} = p_n - a_n \nabla \tilde{\phi} (p_n); \)
3: **if** \( \hat{p}_{n+1} \) satisfies constraints (19b)-(19e)\n4: **then** \( \hat{p}_{n+1} = \bar{p}_{n+1}; \)
5: **else**
6: **solve** problem (23) over \( \hat{p}_{n+1} \) to obtain \( p_{n+1}; \)
7: **end**;
8: **While** \( \tilde{\phi} (p_{n+1}) \leq \tilde{\phi} (p_n) + c\alpha \nabla \tilde{\phi} (p_n)^T (p_{n+1} - p_n); \)
9: \( \tilde{\alpha} \leftarrow \rho \tilde{\alpha}; \)
10: **Update** \( \hat{p}_{n+1} = p_n - a_n \nabla \tilde{\phi} (p_n); \)
11: **if** \( \hat{p}_{n+1} \) satisfies constraints (19b)-(19e)\n12: **then** \( \hat{p}_{n+1} = \bar{p}_{n+1}; \)
13: **else**
14: **solve** problem (23) over \( \hat{p}_{n+1} \) to obtain \( p_{n+1}; \)
15: **end**;
16: **end**;
17: **return** \( \alpha_n = \tilde{\alpha}. \)

Therefore, we propose the approximate gradient descent projection method to efficiently solve problem (19), which is summarized in Algorithm 2.

Algorithm 2: Inexact Gradient Descent Projection Method

1: **Input**: choose \( K \geq 2 \) and choose a random starting point \( p_0 \) which satisfies constraints (19b)-(19e), set \( c_2 \) as the stopping parameter and \( n = 0; \)
2: **Repeat**
3: \( n \leftarrow n + 1; \)
4: **Update** \( \tilde{\phi} (p_{n-1}) = \int_{-r_1}^{r_1} f_{z_b} (z_b) \left( p^T_{n-1} q \right) dz_b; \)
5: **Update** \( \nabla \tilde{\phi} (p_{n-1}) = \int_{-r_2}^{r_2} f_{z_b} (z_b) (q + Qp_{n-1}) dz_b; \)
6: Compute stepsize \( a_{n-1} \) by Algorithm 1;
7: **Update** \( \hat{p}_n = p_{n-1} - a_{n-1} \nabla \tilde{\phi} (p_{n-1}); \)
8: **if** \( \hat{p}_n \) satisfies constraints (19b)-(19e)\n9: \( p_n = \hat{p}_n; \)
10: **else**
11: **solve** problem (23) over \( \hat{p}_n \) to obtain \( p_n; \)
12: **end**;
13: **Until** \( \| p_n - p_{n-1} \| \leq c_2; \)
14: **Output** \( p^{opt} = p_n. \)

B. Case of \( D (p_{y,1} \| p_{y,0}) \leq 2\varepsilon^2 \)

In this subsection, we further consider the other covert constraint \( D (p_{y,0} \| p_{y,0}) \leq 2\varepsilon^2, \) and the corresponding covert rate optimization problem can be formulated as

\[
\begin{align*}
\max_{p_{y,0}} & \quad R_b \\
\text{s.t.} & \quad D (p_{y,1} \| p_{y,0}) \leq 2\varepsilon^2, \\
& \quad \Pr (X = x_k) = p_k \geq 0,
\end{align*}
\] (24)
\[
\sum_{k=1}^{K} p_k |x_k|^2 \leq P_A, \quad (24d)
\]
\[
\sum_{k=1}^{K} p_k = 1, k = 1, \ldots, K. \quad (24e)
\]

To handle intractable constraint \((24b)\), we first derive an upper bound on \(D(p_{y,1} \mid p_{y,0})\) denoted by \(DU(p_{y,1} \mid p_{y,0})\), which is given by
\[
DU(p_{y,1} \mid p_{y,0}) = \sum_{k=1}^{K} p_k \log_2 \sum_{j=1}^{K} p_j \exp \left( -\frac{|g_w^* (x_k - x_j)|^2}{2\sigma_w^2} \right) + \frac{1}{\ln 2} + \frac{|g_w^2 P_A|}{(\ln 2) \sigma_w^2} - 1. \quad (25)
\]

The details of derivations for \((25)\) are given in Appendix B. Then, problem \((24)\) can be reformulated as
\[
\max_{\{p_k\}} R_b \quad \text{s.t. } DU(p_{y,1} \mid p_{y,0}) \leq 2\epsilon^2, \quad (24c), \quad (24d), \quad (24e).
\]

By defining the following equations to simplify \(DU(p_{y,1} \mid p_{y,0})\)
\[
s_{w,k} \overset{\Delta}{=} \left[ \exp \left( -\frac{|g_w^* (x_{k-1} - x_k)|^2}{2\sigma_w^2} \right), \ldots, \exp \left( -\frac{|g_w^2 (x_{k-1} - x_k)|^2}{2\sigma_w^2} \right) \right]^T, \quad (27a)
\]
\[
v_w(p) \overset{\Delta}{=} \left[ \log_2 p^T s_{w,1}, \ldots, \log_2 p^T s_{w,K} \right]^T, \quad (27b)
\]
we obtain
\[
DU(p_{y,1} \mid p_{y,0}) = p^T v_w(p) + \frac{1}{\ln 2} + \frac{|g_w^2 P_A|}{(\ln 2) \sigma_w^2} - 1. \quad (28)
\]

Unfortunately, \(DU(p_{y,1} \mid p_{y,0})\) is non-convex in \(p\), and the covert constraint is also non-convex. To handle this issue, we apply the first order Taylor expansion to \(DU(p_{y,1} \mid p_{y,0})\). Specifically, the derivative of \(DU(p_{y,1} \mid p_{y,0})\) at \(\bar{p}_w\) is given by
\[
\nabla DU(p_{y,1} \mid p_{y,0}) \big|_{p=\bar{p}_w} = v_w(\bar{p}_w) + \nabla v_w(\bar{p}_w), \quad (29)
\]
where \(\nabla v_w = \left[ \frac{s_{w,1}^T}{\frac{p_{w,1}}{\sigma_w}}, \ldots, \frac{s_{w,K}^T}{\frac{p_{w,K}}{\sigma_w}} \right]_{K \times K} \). Then, the first order Taylor expansion of \(p^T v_w(p)\) is given as follows
\[
L(p) \approx p^T v_w(\bar{p}_w) + (v_w(\bar{p}_w) + \nabla v_w(\bar{p}_w))^T (p - \bar{p}_w). \quad (30)
\]
Then, constraint \((26b)\) can be recast to a convex form as
\[
L(p) + \frac{1}{\ln 2} + \frac{|g_w^2 P_A|}{(\ln 2) \sigma_w^2} - 1 \leq 2\epsilon^2. \quad (31)
\]
Thus, problem \((24)\) can be reformulated as
\[
\min_{p} \phi(p) \quad \text{s.t. } L(p) + \frac{1}{\ln 2} + \frac{|g_w^2 P_A|}{(\ln 2) \sigma_w^2} - 1 \leq 2\epsilon^2, \quad (32a)
\]
\[
p^T 1_K = 1, \quad (32b)
\]
\[
p^T (x \otimes x) \leq P_A, \quad (32d)
\]
\[
p \geq 0, \quad (32e)
\]
which is convex.

Similarly, problem \((32)\) can be efficiently solved by the inexact gradient descent projection method. The details are omitted due to space limitation.

IV. SIGNALING DESIGN WITH APPROXIMATE COVERT RATE EXPRESSION

Due to the expectation operation, the achievable rate in \((4)\) does not have a closed-form expression, and can only be computed numerically using the approximate gradient descent method at the expense of high computational complexity.

To strike a balance between complexity and performance, we further derive analytical upper bound and lower bound on the achievable rate in \((4)\).

Lemma 1: An upper bound \(R^U_b\) on the rate \(R_b\) is given by
\[
R^U_b = -\sum_{k=1}^{K} p_k \log_2 \sum_{j=1}^{K} p_j \exp \left( -\frac{|g_w^* (x_k - x_j)|^2}{\sigma_w^2} \right), \quad (33)
\]
while a lower bound \(R^L_b\) on the rate \(R_b\) is given as
\[
R^L_b = -\sum_{k=1}^{K} p_k \log_2 \sum_{j=1}^{K} p_j \exp \left( -\frac{|g_w^* (x_k - x_j)|^2}{2\sigma_w^2} \right) - \frac{1}{\ln 2} + 1. \quad (34)
\]
Please find the derivation in Appendices C and D.

In this section, we adopt the upper bound and lower bound on the achievable rate \(R_b\) in our following analysis.

A. Maximizing \(R^U_b\)

In this subsection, we consider the upper bound on the achievable rate for Bob \(R^U_b\) as the objective function to find the optimal probability of discrete constellation points set. Specifically, we study beamforming design with the objective of maximizing \(R^U_b\), subject to the covert transmission constraint, and the discrete constellation set.

1) \(DU(p_{y,0} \mid p_{y,1}) \leq 2\epsilon^2\): Finding the optimal probability of discrete constellation set can be equivalently written as the following optimization problem
\[
\max_{\{p_k\}} R^U_b \quad \text{s.t. } DU(p_{y,0} \mid p_{y,1}) \leq 2\epsilon^2, \quad (35a)
\]
\[
p_k (X = x_k) = p_k \geq 0, \quad (35b)
\]
\[
\sum_{k=1}^{K} p_k |x_k|^2 \leq P_A, \quad (35c)
\]
\[
\sum_{k=1}^{K} p_k = 1, k = 1, \ldots, K. \quad (35e)
\]

In order to solve problem \((35)\), we first define the following variables
\[
r_k \overset{\Delta}{=} \left[ \exp \left( -\frac{|g_w^* (x_k - x_1)|^2}{\sigma_b^2} \right), \ldots, \exp \left( -\frac{|g_w^* (x_k - x_L)|^2}{\sigma_b^2} \right) \right]^T, \quad (36a)
\]
\[
\mathbf{u}(\mathbf{p}) \triangleq \left[ \log_2 \mathbf{p}^T \mathbf{r}_1, \ldots, \log_2 \mathbf{p}^T \mathbf{r}_K \right]^T.
\]  

(36b)

In this case, we can obtain

\[
R_b^U = -\mathbf{p}^T \mathbf{u}(\mathbf{p}).
\]  

(37)

Therefore, problem (35) can be reformulated as follows

\[
\begin{align*}
\min_{\mathbf{p}} & \quad f_U(\mathbf{p}) \\
\text{s.t.} & \quad -\log_2 \mathbf{p}^T \mathbf{t} \leq 2\varepsilon^2, \\
& \quad \mathbf{p}^T \mathbf{1}_K = 1, \\
& \quad \mathbf{p}^T (\mathbf{x} \otimes \mathbf{x}) \leq P_A, \\
& \quad \mathbf{p} \geq 0,
\end{align*}
\]  

(38a)

(38b)

(38c)

(38d)

(38e)

where \( f_U(\mathbf{p}) = \mathbf{p}^T \mathbf{u}(\mathbf{p}) \). Since the Frank-Wolf method is an algorithm for solving linearly-constrained problems, it makes a linear approximation of the objective function, obtains the feasible descending direction by solving the linear programming, and conducts a one-dimensional search in the feasible region along this direction. Therefore, we will apply the Frank-Wolf method to solve the optimization problem.

We use Taylor's expansion to make a linear approximation of the objective function \( f_U(\mathbf{p}) \). The first order Taylor expansion at \( \mathbf{p}_n \) is as follows

\[
f_U(\mathbf{p}) \approx \mathbf{p}_n^T \mathbf{u}(\mathbf{p}_n) + \nabla f_U(\mathbf{p}_n)^T (\mathbf{p} - \mathbf{p}_n),
\]  

(39a)

\[
\nabla f_U(\mathbf{p}_n) = \mathbf{u}(\mathbf{p}_n) + \nabla \mathbf{u}(\mathbf{p}_n),
\]  

(39b)

where \( \nabla \mathbf{u} = \left[ \frac{r_1}{p_1^{r_1^*}}, \ldots, \frac{r_K}{p_K^{r_K^*}} \right]_{K \times K} \) and \( \mathbf{p}_n \) denotes the current iteration point. Then, we reformulate the optimization problem of (38) as follows

\[
\min_{\mathbf{p}} \quad \nabla f_U(\mathbf{p}_n)^T \mathbf{p}
\]  

(39c)

(39d)

(39e)

(39f)

(39g)

By applying the Frank-Wolf method, the detailed procedures for solving (40) are summarized in Algorithm 3. Note that \( \lambda_n \) is the stepsize of the \( n \)th iteration and \( \mathbf{d}_n \) denotes the feasible descending direction of the \( n \)th iteration.

**Algorithm 3**: Solving (40) by Frank-Wolf Method

1. **Initialization**: Choose a feasible starting point \( \mathbf{p}_0 \), set \( \delta > 0 \) as the stopping parameter, let \( n = 0 \);
2. **While** \( \| \nabla f(\mathbf{p}_n) \| \leq \delta \);
3. Solve the linear programming problems (40) and obtain optimal solution \( \mathbf{p}_{n+1} \);
4. Construct the feasible descending direction \( \mathbf{d}_n = \mathbf{N}_n - \mathbf{p}_n \);
5. Obtain optimal solution \( \lambda_n = \arg \min f(\mathbf{p}_n + \lambda \mathbf{d}_n); \)
6. Let \( \mathbf{p}_{n+1} = \mathbf{p}_n + \lambda_n \mathbf{d}_n, \ n \leftarrow n + 1; \)
7. **end**
8. **Output** \( \mathbf{p}_n \)

2) \( D_U(\mathbf{p}_y,1\|\mathbf{p}_y,0) \leq 2\varepsilon^2 \): Furthermore, we consider the optimal probabilistic constellation shaping for covert communications with covert constraint \( D_U(\mathbf{p}_y,1\|\mathbf{p}_y,0) \leq 2\varepsilon^2 \), such that

\[
\begin{align*}
\max_{\mathbf{p},1} & \quad R_b^U \\
\text{s.t.} & \quad D_U(\mathbf{p}_y,1\|\mathbf{p}_y,0) \leq 2\varepsilon^2, \\
& \quad \mathbf{p}^T \mathbf{1}_K = 1, \\
& \quad \mathbf{p}^T (\mathbf{x} \otimes \mathbf{x}) \leq P_A, \\
& \quad \mathbf{p} \geq 0,
\end{align*}
\]  

(41a)

(41b)

which is non-convex.

Similar to problem (26), by replacing constraint (41b) by constraint (31), we can obtain the optimization problem

\[
\begin{align*}
\min_{\mathbf{p}} & \quad \nabla f_U(\mathbf{p}_n)^T \mathbf{p} \\
\text{s.t.} & \quad L(\mathbf{p}) + \frac{1}{\ln 2} \left[ \frac{\sigma_w^2 P_A}{(\ln 2) \sigma_\mu^2} - 1 \right] \leq 2\varepsilon^2, \\
& \quad \mathbf{p} \geq 0,
\end{align*}
\]  

(42a)

(42b)

(42c)

(42d)

Then, we apply the Frank-Wolf method to solve problem (35), and the details are omitted since the corresponding algorithm is similar to Algorithm 3.

**B. Maximizing \( R_b^L \)**

In this subsection, we further study the lower bound beamforming design for covert communication by maximizing the lower bound \( R_b^L \), while satisfying the covert transmission requirement and the discrete constellation set with \( K \).

1) \( D_U(\mathbf{p}_y,0\|\mathbf{p}_y,1) \leq 2\varepsilon^2 \): Under the covert constraint \( D_U(\mathbf{p}_y,0\|\mathbf{p}_y,1) \leq 2\varepsilon^2 \), the optimal probabilistic constellation shaping for covert communications is formulated as follows

\[
\begin{align*}
\max_{\mathbf{p}_k} & \quad R_b^L \\
\text{s.t.} & \quad D_U(\mathbf{p}_y,0\|\mathbf{p}_y,1) \leq 2\varepsilon^2, \\
& \quad \Pr(X = x_k) = p_k \geq 0, \\
& \quad \sum_{k=1}^K p_k |x_k|^2 \leq P_A, \\
& \quad \sum_{k=1}^K p_k = 1, k = 1, \ldots, K.
\end{align*}
\]  

(43a)

(43b)

(43c)

(43d)

(43e)

To solve the problem, we define the following equations

\[
\begin{align*}
\mathbf{s}_{b,k} & \triangleq \left[ \exp \left( -\frac{|s_k(x_k-x)|^2}{2\sigma_b^2} \right), \ldots, \exp \left( -\frac{|s_k(x_k-x)|^2}{2\sigma_b^2} \right) \right]^T, \\
\mathbf{v}_b(\mathbf{p}) & \triangleq \left[ \log_2 \mathbf{p}^T \mathbf{s}_{b,1}, \ldots, \log_2 \mathbf{p}^T \mathbf{s}_{b,K} \right]^T,
\end{align*}
\]  

(44a)

(44b)

and then the lower bound \( R_b^L \) can be transformed to

\[
R_b^L = -\mathbf{p}^T \mathbf{v}_b(\mathbf{p}) - \frac{1}{\ln 2} + 1.
\]  

(45)

Thus, the covert optimization problem is recast as follows

\[
\begin{align*}
\max_{\mathbf{p}} & \quad -\mathbf{p}^T \mathbf{v}_b(\mathbf{p}) \\
\text{s.t.} & \quad -\log_2 \mathbf{p}^T \mathbf{t} \leq 2\varepsilon^2, \\
& \quad \mathbf{p}^T \mathbf{1}_K = 1, \\
& \quad \mathbf{p}^T (\mathbf{x} \otimes \mathbf{x}) \leq P_A, \\
& \quad \mathbf{p} \geq 0,
\end{align*}
\]  

(46a)

(46b)

(46c)

(46d)

(46e)

Similar to problem (35), we can apply the Frank-Wolf method to solve problem (43), and the details are omitted.
2) $D_{U}(p_{x,1} \parallel p_{y,0}) \leq 2\epsilon^2$: With the covert constraint $D_{U}(p_{x,1} \parallel p_{y,0}) \leq 2\epsilon^2$, the optimal probabilistic constellation shaping for covert communications is given as

$$
\max_{\{p_{1}\}} R_{b}^{L}
$$

s.t. $D_{U}(p_{x,1} \parallel p_{y,0}) \leq 2\epsilon^2$, (43c), (43d), (43e).

By replacing the constraint (47b) with (31), we obtain the optimization problem

$$
\max_{\mathbf{p}} - \mathbf{p}^{T} \mathbf{v}_{b}(\mathbf{p})
$$

s.t. $L(\mathbf{p}) + \frac{1}{\ln 2} + \left[ \frac{g_{w}^{2}P_{A}}{(\ln 2)\sigma_{w}^{2}} \right] - 1 \leq 2\epsilon^2$, (46c), (46d), (46e).

Similar to problem (41), we can apply a similar method to solve problem (47).

V. NUMERICAL RESULTS

In this section, we present and discuss numerical results to assess the performance of the proposed probabilistic constellation shaping designs. In our simulations, the discrete constellation input is QAM modulation, the total transmit power of Alice is $P_{A} = 10W$, the noise variance of Willie is $\sigma_{w}^{2} = 1W$, and the noise variance of Bob is $\sigma_{b}^{2} = \frac{P_{b}}{10^{proberm}}$. Moreover, we assume that all channels experience Rayleigh flat fading, and $\sigma_{1} = \sigma_{2} = 1$ [48].

We first compare the proposed optimal probabilistic constellation shaping design with the equiprobable design, starting from the empirical CDF of KL divergence and the rate comparison.

Fig. 2 shows the empirical CDF of the achieved $D(p_{y,0} \parallel p_{y,1})$ and $D(p_{x,1} \parallel p_{y,0})$, respectively, for both the proposed optimal probabilistic constellation shaping design and the equiprobable design with SNR = 10dB and $K = 8$, where the covertness threshold is $2\epsilon^2 = 0.02$, i.e., $D(p_{y,0} \parallel p_{y,1}) \leq 0.02$ and $D(p_{x,1} \parallel p_{y,0}) \leq 0.02$. As observed from Fig. 2, the proposed optimal probabilistic constellation shaping design satisfies the covertness constraint. On the other hand, the equiprobable design cannot satisfy the covertness constraints.

Fig. 3 (a) and (b) show the achievable covert rate of Bob $R_{b}$ for the proposed optimal probabilistic constellation shaping design and the equiprobable design with $D(p_{y,0} \parallel p_{y,1})$ and $D(p_{x,1} \parallel p_{y,0})$, respectively. We observe that the optimal probabilistic constellation shaping design is superior when the signal-to-noise ratio is low. Therefrom, in practical applications, the proposed optimal probabilistic constellation shaping design is advantageous in medium and low SNR, and the equiprobable design is suitable for high SNR.

Next, we evaluate the performance of the proposed optimal probabilistic constellation shaping design.

Fig. 4 shows the optimal probability distribution of input $\{p_{x,1}\}$ with SNR = 12dB of the proposed optimal probabilistic constellation shaping design, for $D(p_{y,0} \parallel p_{y,1})$ in Fig. 4 (a), and for $D(p_{x,1} \parallel p_{y,0})$ in Fig. 4 (b). As it can be seen from Fig. 4, for the proposed optimal probabilistic constellation shaping design, the optimal probability distribution is not equiprobable, and the symmetrical points have equal probabilities. Specifically, when the number of discrete constellation points is sixteen, the probability of each constellation point for $D(p_{y,0} \parallel p_{y,1})$ and $D(p_{x,1} \parallel p_{y,0})$ is given in Table I. From the table, we can clearly see that when the coordinates of the constellation points are symmetrical, their probabilities are the same, and vice versa.

Fig. 5 considers the proposed optimal probabilistic constellation shaping design and depicts, the achievable covert rate of Bob $R_{b}$ versus SNR with different number of constellation points $K = 2, 4, 8, 16$ for $D(p_{y,0} \parallel p_{y,1})$. It can be seen from Fig. 5 that when SNR increases, the covert rate of Bob $R_{b}$ increases. In addition, we observe that larger number of points
Table I

| $\{p_{i,j}\}$ | $\text{Re}\{z_{i,j}\}$ | $D\left(p_{y,0} \| p_{y,1}\right)$ | $D\left(p_{y,1} \| p_{y,0}\right)$ |
|---------------|----------------|-------------------------------|-------------------------------|
| $-3$          | 0.0484        | 0.0524                        | 0.0484                        |
| $-1$          | 0.0524        | 0.0968                        | 0.0524                        |
| $1$           | 0.0524        | 0.0968                        | 0.0524                        |
| $3$           | 0.0484        | 0.0552                        | 0.0484                        |

$K$ results in higher covert rate of Bob $R_b$, especially for high SNRs. Thus, as the modulation order increases, the rate of covert communication increases.

Fig. 3. The covert rate of (a) $D\left(p_{y,0} \| p_{y,1}\right)$ and (b) $D\left(p_{y,1} \| p_{y,0}\right)$ for the proposed optimal probabilistic constellation shaping design and the equiprobable design.

Fig. 4. The optimal probability of discrete constellation points with (a) $D\left(p_{y,0} \| p_{y,1}\right)$ and (b) $D\left(p_{y,1} \| p_{y,0}\right)$ for the proposed optimal probabilistic constellation shaping design.

Fig. 6 shows the covert rate versus $\varepsilon$ for the proposed optimal probabilistic constellation shaping design for $D\left(p_{y,0} \| p_{y,1}\right)$ and $D\left(p_{y,1} \| p_{y,0}\right)$, where $K = 8$. 
Fig. 5. The achievable covert rate of Bob versus SNR with different number of points $K$ for $D(p_{y,0} \parallel p_{y,1})$.

Fig. 6. The covert rate versus $\varepsilon$ for the proposed optimal probabilistic constellation shaping design with the cases of $D(p_{y,0} \parallel p_{y,1})$ and $D(p_{y,1} \parallel p_{y,0})$.

For $\text{SNR} = 6\text{dB}$. It can be observed from the figure that as $\varepsilon$ increases, the covert constraint becomes loose, resulting in a covert rate increase. The rate of the covert constraint $D(p_{y,0} \parallel p_{y,1}) \leq 2\varepsilon^2$ is higher than that of the case $D(p_{y,1} \parallel p_{y,0}) \leq 2\varepsilon^2$. This is because $D(p_{y,1} \parallel p_{y,0})$ is less than $D(p_{y,1} \parallel p_{y,0})$ for the same probability distribution, and thus $D(p_{y,1} \parallel p_{y,0}) \leq 2\varepsilon^2$ is more stringent than $D(p_{y,0} \parallel p_{y,1}) \leq 2\varepsilon^2$.

Finally, we compare the performance and complexity of three objective functions.

Fig. 7 depicts the achievable rate of Bob $R_b$ with the proposed optimal probabilistic constellation shaping design, as well as the objective functions $R_b^L$ and $R_b^U$ versus the SNR for the case of $D(p_{y,0} \parallel p_{y,1})$ and $D(p_{y,1} \parallel p_{y,0})$, respectively. It can be observed that the mutual information of Bob $R_b$ increases as the SNR increases, while $R_b$ of the proposed optimal probabilistic constellation shaping design is between the objective functions $R_b^L$ and $R_b^U$, and the objective function $R_b^U$ is higher than the objective function $R_b^L$.

From Fig. 7, we observe that the proposed optimal probabilistic constellation shaping design is between the objective functions $R_b^L$ and $R_b^U$. Here, we compare the computational complexity of the three designs by computational time in Table II, and all simulations of the three methods are performed using MATLAB 2016b with 2.30GHz, 2.29GHz dual CPUs and a 128GB RAM, where $K = 8$.

Specifically, Table II shows that the computational time of objective functions $R_b$, $R_b^L$ and $R_b^U$ for the covert constraint condition $D(p_{y,0} \parallel p_{y,1})$ is 55.04, 2.845 and 3.012 seconds,
respectively. The computational times of the latter two cases is approximately 95 percent shorter than that of the probabilistic constellation shaping design. Under the covert constraint condition $D (p_{y,0} \| p_{y,1})$ the computational time of objective functions $R_b$, $R_b^L$ and $R_b^U$ is 107.77, 3.219 and 3.438 seconds, respectively. The computational times of the latter two cases is approximately improved by 97 percent compared to that of the probabilistic constellation shaping design. Moreover, the computational time of the design for $D (p_{y,0} \| p_{y,1})$ is less than that of $D (p_{y,1} \| p_{y,0})$.

### VI. Conclusion

In this paper, we propose an optimal probabilistic constellation shaping design for covert communications, where Alice covertly sends a message to Bob while avoiding being discovered by Willie. We derive the achievable rate expressions of the covert communications system, and we study the covert rate maximization problem via optimizing the constellation distribution. In addition, to strike a balance between the computational complexity and the transmission performance, we further develop a framework that maximizes the upper and lower bounds of the achievable rate. Numerical results quantify the gains of the proposed beamformers design over state-of-the-art schemes in terms of the achievable covert rate.

### APPENDIX A

#### DERIVATION OF THE FORMULATION (15)

The upper bound $D_U (p_{y,0} \| p_{y,1})$ on the KL divergence $D (p_{y,0} \| p_{y,1})$ is derived as follows

$$D (p_{y,0} \| p_{y,1})$$

$$= - \log_2 \sum_{k=1}^K p_k \exp \left( - \frac{|g^*_w x_k|^2}{\sigma_w^2} \right)$$

where inequality (49a) holds due to Jensen’s Inequality.

$$D_U (p_{y,0} \| p_{y,1})$$

is given as

$$D_U (p_{y,0} \| p_{y,1})$$

$$= - \log_2 \sum_{k=1}^K p_k \exp \left( - \frac{|g^*_w x_k|^2}{\sigma_w^2} \right)$$

where inequality (49a) holds due to Jensen’s Inequality.

#### APPENDIX B

#### DERIVATION OF THE FORMULATION (25)

The upper bound $D_U (p_{y,1} \| p_{y,0})$ of the KL divergence $D (p_{y,1} \| p_{y,0})$ is given as

$$D (p_{y,1} \| p_{y,0})$$

$$= \frac{1}{ln 2} \ln 2 \frac{g^*_w P_A}{\sigma_w^2}$$

where inequality (49a) holds due to Jensen’s Inequality.

Equality (50b) holds because of the definitions $z_{w,R} \equiv \Re \{z_w\}$
and \( z_{w,R} \triangleq \text{Im} \{ z_{w} \} \), \( z_{w,R}, z_{w,I} \sim \mathcal{N} \left( 0, \frac{1}{2} \sigma_{b}^{2} \right) \), and \( c_{R} \triangleq \text{Re} \left( g_{w}^{*} (x_{k} - x_{j}) \right) \), \( c_{I} \triangleq \text{Im} \left( g_{w}^{*} (x_{k} - x_{j}) \right) \).

**APPENDIX C**

**DERIVATION OF THE FORMULATION (33)**

The upper bound \( R_{b}^{U} \) of the covert rate \( R_{b} \) is derived as follows:

\[
R_{b}^{U} \leq - \sum_{k=1}^{K} p_{k} \log_{2} \sum_{j=1}^{K} p_{j} \times \exp \left( - \mathbb{E}_{cb} \left[ \frac{|g_{k}^{*} (x_{k} - x_{j}) + z_{b}^{2}|}{\sigma_{b}^{2}} \right] - \frac{1}{\ln 2} \right) \tag{51a}
\]

\[
= - \sum_{k=1}^{K} p_{k} \log_{2} \sum_{j=1}^{K} p_{j} \exp \left( - \mathbb{E}_{cb} \left[ \frac{|g_{k}^{*} (x_{k} - x_{j})|^{2}}{\sigma_{b}^{2}} \right] - \frac{1}{\ln 2} \right) \tag{51b}
\]

\[
= - \sum_{k=1}^{K} p_{k} \left( \log_{2} \sum_{j=1}^{K} p_{j} \exp \left( - \mathbb{E}_{cb} \left[ \frac{|g_{k}^{*} (x_{k} - x_{j})|^{2}}{\sigma_{b}^{2}} \right] \right) - \frac{1}{\ln 2} \right) \tag{51c}
\]

\[
= - \sum_{k=1}^{K} p_{k} \log_{2} \sum_{j=1}^{K} p_{j} \exp \left( - \mathbb{E}_{cb} \left[ \frac{|g_{k}^{*} (x_{k} - x_{j})|^{2}}{\sigma_{b}^{2}} \right] \right) \exp(-1) \tag{51d}
\]

\[
= - \frac{1}{\ln 2} - \sum_{k=1}^{K} p_{k} \log_{2} \sum_{j=1}^{K} p_{j} \exp \left( - \frac{|g_{k}^{*} (x_{k} - x_{j})|^{2}}{\sigma_{b}^{2}} \right) \tag{51e}
\]

where inequality (51a) is true due to Jensen’s Inequality.

**APPENDIX D**

**DERIVATION OF THE FORMULATION (34)**

The lower bound \( R_{b}^{L} \) on the covert rate \( R_{b} \) is given as

\[
R_{b}^{L} \geq - \sum_{k=1}^{K} p_{k} \log_{2} \sum_{j=1}^{K} p_{j} \times \mathbb{E}_{cb} \left[ \exp \left( - \frac{(g_{k}^{*} (x_{k} - x_{j}) + z_{b})^{2}}{\sigma_{b}^{2}} \right) \right] - \frac{1}{\ln 2} \tag{52a}
\]

\[
= - \sum_{k=1}^{K} p_{k} \log_{2} \sum_{j=1}^{K} p_{j} \times \left[ \int_{-\infty}^{\infty} d z_{b,R} \exp \left( - \frac{(a_{R} + z_{b,R})^{2} + z_{b,R}^{2}}{\sigma_{b}^{2}} \right) \right] \tag{52b}
\]

\[
= - \sum_{k=1}^{K} p_{k} \log_{2} \sum_{j=1}^{K} p_{j} \exp \left( - \frac{\left| g_{k}^{*} (x_{k} - x_{j}) \right|^{2}}{2 \sigma_{b}^{2}} \right) \tag{52c}
\]

\[
= - \frac{1}{\ln 2} - \sum_{k=1}^{K} p_{k} \log_{2} \sum_{j=1}^{K} p_{j} \exp \left( - \frac{\left| g_{k}^{*} (x_{k} - x_{j}) \right|^{2}}{2 \sigma_{b}^{2}} \right) \tag{52d}
\]

where inequality (52a) is true due to Jensen’s inequality, equality (52b) holds because of the definitions \( z_{b,R} \triangleq \text{Re} \{ z_{b} \} \), \( z_{b,R}, z_{b,I} \sim \mathcal{N} \left( 0, \frac{1}{2} \sigma_{b}^{2} \right) \), and \( a_{R} \triangleq \text{Re} \left( g_{k}^{*} (x_{k} - x_{j}) \right) \), \( a_{I} \triangleq \text{Im} \left( g_{k}^{*} (x_{k} - x_{j}) \right) \).

**REFERENCES**

[1] V.-L. Nguyen, P.-C. Lin, B.-C. Cheng, R.-H. Hwang, and Y.-D. Lin, “Security and privacy for 6G: A survey on prospective technologies and challenges,” IEEE Commun. Surv. Tuts., vol. 23, no. 4, pp. 2384–2428, 4th Quart., 2021.

[2] X. Chen, D. W. K. Ng, W. Gerstacker, and H. H. Chen, “A survey on multiple-antenna techniques for physical layer security,” IEEE Commun. Surv. Tuts., vol. 19, no. 2, pp. 1027–1053, 2nd Quart., 2016.

[3] M. Bloch and J. Barros, Physical-Layer Security: From Information Theory to Security Engineering, Cambridge, U.K.: Cambridge Univ., 2011.

[4] H.-M. Wang, T.-X. Zheng, J. Yuan, D. Towsley, and M. H. Lee, “Physical layer security in heterogeneous cellular networks,” IEEE Trans. Commun., vol. 64, no. 3, pp. 1204–1219, Mar. 2016.

[5] H. Du, D. Niyato, Y.-A. Xie, Y. Cheng, J. Kang, and D. I. Kim, “Performance analysis and optimization for jammer-aided multi-antenna UAV covert communication,” 2022, arXiv:2202.00973.

[6] H. Du, J. Kang, T. D. Niyato, J. Zhang, and D. I. Kim, “Reconfigurable intelligent surface-aided joint radar and covert communications: Fundamentals, optimization, and challenges,” IEEE Veh. Technol. Mag., vol. 17, no. 3, pp. 54–64, Sep. 2022.

[7] H. Du, D. Niyato, Y.-A. Xie, Y. Cheng, J. Kang, and D. I. Kim, “Performance analysis and optimization for jammer-aided multi-antenna UAV covert communication,” 2022, arXiv:2202.00973.

[8] Y.-A. Xie et al., “Securing federated learning: A covert communication-based approach,” IEEE Netw., early access, Jul. 25, 2022, doi: 10.1109/MINET.17.2200060.

[9] S. Yan, Y. Cong, S. V. Hanly, and X. Zhou, “Gaussian signalling for covert communications,” IEEE Trans. Wireless Commun., vol. 18, no. 7, pp. 3542–3553, Jul. 2019.

[10] M. R. Bloch, “Covert communication over noisy channels: A resolvability perspective,” IEEE Trans. Inf. Theory, vol. 62, no. 5, pp. 2334–2354, May 2016.

[11] M. K. Simon, J. K. Omura, R. A. Scholtz, and B. K. Levitt, Spread Spectrum Communications Handbook. New York, NY, USA: McGaw-Hill, Apr. 1994.

[12] 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.

[13] L. Wang, G. W. Wornell, and L. Zheng, “Fundamental limits of communication with low probability of detection,” IEEE Trans. Inf. Theory, vol. 62, no. 6, pp. 3493–3503, Jun. 2016.

[14] H. Wu, X. Liao, Y. Dang, Y. Shen, and X. Jiang, “Limits of covert communication on two-hop AWGN channels,” in Proc. Int. Conf. Netw. Netw. Netw. Appl. (NAN), Oct. 2017, pp. 42–47.
Shuai Ma received the B.S. and Ph.D. degrees in communication and information systems from Xidian University, Xi’an, China, in 2009 and 2016, respectively. From 2014 to 2015, he was a Vis-iting Scholar with the Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX, USA. He has been an Asso-ociate Professor with the School of Information and Control Engineering, China University of Mining and Technology, Xuzhou, China. His research interests include visible light communication, wireless communications, and network information theory.

Yunqi Zhang received the B.S. degree from Nanjing Agricultural University, Nanjing, China, in 2018. She is currently pursuing the M.E. degree in electronic and communication engineering with the School of Information and Control Engineering, China University of Mining and Technology, Xuzhou, China. Her research interests include visible light communication, wireless communications, and network information theory.

Haihong Sheng received the B.S. degree from the Suzhou Institute of Technology, Jiangsu Uni-versity of Science and Technology, Suzhou, China, in 2020. She is currently pursuing the M.E. degree in electronic and communication engineering with the School of Information and Control Engineer-ing, China University of Mining and Technology, Xuzhou, China. Her research interests include visi-ble light communication, wireless communications, and network information theory.
Hang Li (Member, IEEE) received the B.E. and M.S. degrees from Beihang University, Beijing, China, in 2008 and 2011, respectively, and the Ph.D. degree from Texas A&M University, College Station, TX, USA, in 2016. He was a Post-Doctoral Research Associate with Texas A&M University from September 2016 to August 2017 and the University of California at Davis from September 2017 to March 2018. Since April 2018, he has been a Visiting Research Scholar with the Shenzhen Research Institute of Big Data, Shenzhen, China. His current research interests include wireless networks, stochastic optimization, and applications of machine learning.

Jia Shi (Member, IEEE) received the M.Sc. and Ph.D. degrees from the University of Southampton, U.K., in 2010 and 2015, respectively. He was a Research Associate with Lancaster University, U.K., from 2015 to 2017. Then, he became a Research Fellow with 5GIC, University of Surrey, U.K., from 2017 to 2018. Since 2018, he has been with Xidian University, China, where he is now an Associate Professor with the National Key Laboratory of Integrated Services Networks (ISN). His current research interests include resource allocation in wireless systems, covert communications, physical layer security, mmWave communications, and satellite communications. He is now serving as an Associate Editor for Electronic Letters and an Editor for International Journal of Communication Systems. He is serving as a Guest Editor for China Communications.

Long Yang (Member, IEEE) received the B.E. and Ph.D. degrees from Xidian University, Xian, China, in 2010 and 2015, respectively. Since 2015, he has been a Faculty Member with Xidian University, where he is currently an Associate Professor with the State Key Laboratory of Integrated Services Networks. From 2017 to 2019, he was a Post-Doctoral Fellow with the Department of Electrical and Computer Engineering, University of Alberta, Canada. His current research interests include non-orthogonal multiple access, cooperative communications, and wireless physical-layer security.

Youlong Wu (Member, IEEE) received the B.S. degree in electrical engineering from Wuhan University, Wuhan, China, in 2007, the M.S. degree in electrical engineering from Shanghai Jiao Tong University, Shanghai, China, in 2011, and the Ph.D. degree from Telecom ParisTech, Paris, France, in 2014. In December 2014, he worked as a Post-Doctoral Researcher at the Institute for Communication Engineering, Technical University Munich (TUM), Munich, Germany. In 2017, he joined the School of Information Science and Technology, ShanghaiTech University. His research interests in communication theory, information theory, and its applications, e.g., coded caching, distributed computation, and machine learning. He received the TUM Fellowship in 2014 and is an Alexander von Humboldt Research Fellow.

Naofal Al-Dhahir (Fellow, IEEE) received the Ph.D. degree from Stanford University. He was a Principal Member of Technical Staff at the GE Research Center and AT&T Shannon Laboratory from 1994 to 2003. He is currently an Erik Jonsson Distinguished Professor and the ECE Associate Head at UT-Dallas. He is a co-inventor of 43 issued patents, the coauthor of about 500 papers, and the co-recipient of five IEEE best paper awards. He is a fellow of the National Academy of Inventors. He received the 2019 IEEE SPCC Technical Recognition Award and the 2021 Qualcomm Faculty Award. He served as the Editor-in-Chief for IEEE TRANSACTIONS ON COMMUNICATIONS from January 2016 to December 2019.

Shiyin Li received the Ph.D. degree in information and communication engineering from the China University of Mining and Technology, Xuzhou, China, in 2010. Since 2010, he has been a Professor with the School of Information and Control Engineering, China University of Mining and Technology, where he is the Head of the Department of Information Engineering. His research interests include wireless communication and network congestion control.