SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •

September 2021, Vol. 64 192303:1–192303:15
https://doi.org/10.1007/s11432-019-2919-5

Covert communication with beamforming over MISO channels in the finite blocklength regime

Xinchun YU, Yuan LUO* & Wen CHEN

School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

Received 3 December 2019/Revised 25 February 2020/Accepted 13 May 2020/Published online 18 August 2021

Abstract This paper considers beamforming techniques for covert communication over multiple input single output (MISO) channels with finite blocklength. We first show that the optimal input distribution for covert communication over complex Gaussian channels is circular symmetric complex Gaussian. By reviewing our previous results on the throughput of Gaussian random coding over additive white Gaussian noise (AWGN) channels, the achievability and converse bounds on the attainable throughput over MISO channels are analyzed. Then, the optimal beamforming strategies and their relationship with the transmitting power are thoroughly investigated in a variety of situations in terms of several types of channel state information available at both ends of the system. We reveal the fact that the maximal allowable transmit power is not constrained by covertness requirement when there is full channel information of the adversary, while there is an upper bound of the transmit power which is based on beamforming and an outage-based covertness criterion if there is only partial channel information of the adversary. Finally, numerical results are presented to show that the throughput of covert communication can be increased notably by adopting a proper beamforming strategy in MISO channels in comparison with the single antenna case.

Keywords beamforming, circular symmetric complex Gaussian, covert communication, MISO channels, power control

Citation Yu X C, Luo Y, Chen W. Covert communication with beamforming over MISO channels in the finite blocklength regime. Sci China Inf Sci, 2021, 64(9): 192303, https://doi.org/10.1007/s11432-019-2919-5

1 Introduction

1.1 Background and related work

As multi-antenna systems became more and more prevalent in the past decade, a lot of research has been done on Gaussian multiple input multiple output (MIMO) wiretap channels, such as [1–5]. In these results, the transmitters are all assumed to have full channel state information (CSI) about the legitimate receiver and the eavesdropper. Through the feedback channel, the transmitter is assumed to be able to track the channel variations. When the transmitter is assumed to have only partial CSI for the legitimate receiver or the eavesdropper, the results can be found in [6–8]. Although beamforming is widely discussed in various wiretap channel models, their application in covert communication has not received much attention. In this paper, we are interested in the application of beamforming techniques in covert communication, in which the adversary cannot effectively detect the transmission between the sender and the legitimate receiver. The theoretical limit of covert communication was first investigated in [9] over additive white Gaussian noise (AWGN) channels and in [10, 11] over discrete memoryless channels (DMCs), and later in [12] over MIMO channels. It has been shown that covert communication has square root law (SRL): the transmitter is allowed to send $O(\sqrt{n})$ information bits with n channel uses when a lower bound of the noise level of the adversary is known, and the quantity decreases to $o(\sqrt{n})$ when the lower bound is unknown. That is, the channel coding rate is asymptotically zero. Other discussions on covert communication systems can be found in [13, 14]. As the conclusion of SRL is pessimistic, a lot of proposals have been put forward to improve the transmission efficiency and even a positive rate can be

*Corresponding author (email: yuanluo@sjtu.edu.cn) obtained in some scenarios. For example, in [15,16], the covert communication was discussed in a scenario where the adversary has uncertainty about his noise variance over AWGN and MIMO channels, and it is possible to achieve $O(n)$ bits with n channel uses. In [17,18], the transmitter's uncertainty about the noise level and the channel of the adversary were investigated. The authors in [19] investigated jammer-aid covert communication and also proved that a positive rate is achievable. In [20], the transmitter has an alliance that produces artificial noise to aid covert communication when there are a group of uniformly distributed adversaries. Under a random network situation, a comprehensive analysis and optimization framework for the covert throughput of the system from the stochastic geometry theory is provided in [21] by considering both centralized and distributed antenna systems.

The first and second order asymptotics of covert communication is firstly characterized in [22, 23] over a discrete memoryless channel. In [24], the throughput and maximal allowable power are investigated over a slow fading channel. The upper bound of the transmit power of each symbol as a linear function of $\frac{1}{\sqrt{n}}$ and the channel throughput have been determined, both of which incorporate the channel coefficient and follow SRL. The achievability and converse bounds on the throughput of covert communication over AWGN channels are investigated by us in [25, 26] by revisiting the techniques in [27]. Based on these results, we go a step further to consider the achievability and converse bounds of covert communication over multiple input single output (MISO) channels in this work by considering beamforming technologies, which will decrease or even remove the effect of SRL depending on the quality of channel state information (of the adversary) available at the transmitter.

1.2 Main contributions

The main contributions of our work are listed as follows.

- To maximize the mutual information of the legitimate channel under a covert constraint, which is imposed in terms of K-L divergence between distributions perceived at the adversary depending on transmitter's states, we prove that the optimal signaling for covert communication over complex Gaussian channel is circular symmetric complex Gaussian.
- The application of our previous results over AWGN channels on the bounds of the throughput of covert communication over MISO channels is briefly introduced, which is important for numerical testification.
- When the transmitter has full CSI for both the legitimate receiver and the adversary, an optimal beamforming strategy is proposed to maximize the transmission efficiency in terms of the throughput with finite blocklength at the expense of some amount of covertness in order to attain better transmission efficiency than that of the zero-forcing strategy. The throughput is surprisingly not constrained by SRL.
- When the transmitter has partial CSI of the adversary but full CSI of the legitimate receiver, a sharp upper bound on the transmit power per channel use is determined with an outage-based covertness criterion. Under this bound, the optimal beamforming strategy is further investigated for maximizing the transmission efficiency with the corresponding covertness criterion. As a special case, when there is only statistical information of the adversary's channel state, it is shown that the unique choice of beamforming strategy is maximal ratio transmission.
- Numerical results are presented to testify the superiority of the optimal beamforming over the zero-forcing beamforming and the single antenna scenario in terms of achievability and converse bounds of the throughput. In addition, when there is only partial information of the adversary's channel, the available regions for beamforming with varying parameters are also illustrated.

1.3 Notations

Throughout the paper, the following notations are utilized if not stated otherwise. X, Y, Z are generic random variables of X^n, Y^n, Z^n, and x, y, z are the sample values of X, Y, Z, respectively. We use bold lower and upper letters to denote vectors and matrices. Vectors are always column vectors. $\mathbb{R}$ and $\mathbb{C}$ represent the field of real numbers and the field of complex numbers. We denote $\boldsymbol{A}^{\mathrm{H}}$ and $\boldsymbol{A}^{\mathrm{T}}$ as Hermitian transpose and the transpose of a matrix $\boldsymbol{A}$, respectively. $|\cdot|$ and $\|\cdot\|$ denote the absolute value of a scalar and the L_2-norm of a vector, respectively. $\mathbf{I}$ stands for the identity matrix. $\Pi_{\boldsymbol{A}}^{\perp}$ is the orthogonal projector onto the orthogonal complement of the column space of $\boldsymbol{A}$, i.e., $\Pi_{\boldsymbol{A}}^{\perp} = \mathbf{I} - \Pi_{\boldsymbol{A}}$ with $\Pi_{\boldsymbol{A}} = \boldsymbol{A}(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{H}}$. We denote E as the expectation of random variables. $\mathcal{H}(\cdot)$ describes the entropy function. The relative entropy or K-L divergence between two distributions $\mathbb{P}_0$ and $\mathbb{P}_1$ (or between $\mathbb{P}_1$ and $\mathbb{P}_0$) is denoted by $D(\mathbb{P}_0||\mathbb{P}_1)$ (or $D(\mathbb{P}_1||\mathbb{P}_0)$). The log function in this paper is of base 2.

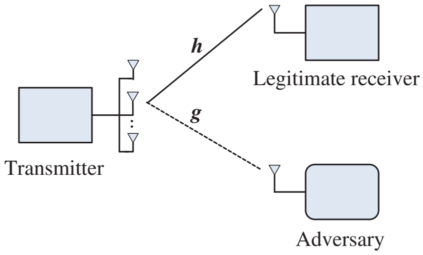


Figure 1 MISO channel model of covert communication in Subsection 2.1.

2 Channel model and the adversary's hypothesis testing

2.1 Channel model

In this subsection, we present the channel model of covert communication over MISO channels as shown in Figure 1. The transmitter is communicating to a legitimate receiver Bob in the presence of an adversary Willie. The adversary is assumed to be passive, which means that it only receives signals but does not transmit. An $(n, 2^{nR})$ code for covert communication systems consists of a message set $\mathcal{W} = \{1, \ldots, 2^{nR}\}$, an encoder at the transmitter Alice, a decoder at the legitimate user Bob and a detector is at the adversary Willie. The encoder is defined as $f_n : \mathcal{W} \to \mathbb{C}^{m\times n}, w \mapsto \boldsymbol{w}\boldsymbol{x}^{\mathrm{T}}$, where $\boldsymbol{w}$ is a beamforming vector:

$$\boldsymbol{w} \in \mathbb{C}^m, \quad \|\boldsymbol{w}\| = 1, \tag{1}$$

and $\boldsymbol{x}$ is a column vector which represents a codeword $[x_1, x_2, \ldots, x_n]^{\mathrm{T}}$. The decoder is defined as $g_n : \mathbb{C}^n \to \mathcal{W}, y^n \to \hat{w}$. The detector is defined as $h_n : \mathbb{C}^n \to \{0, 1\}, z^n \to 0/1$. Bob wants to decode the message $\hat{w}$ with a small error probability P_e^n and Willie wants to determine whether Alice is communicating ($h_n = 1$) or not ($h_n = 0$) by statistical hypothesis test. Note that the adversary in covert communication is different from that of secret communication scenarios. In the latter case, the adversary always wants to decode the message transmitted by Alice, but here the adversary Willie's aim is to detect Alice's communication. Alice is equipped with m antennas while both Bob and Willie are equipped with a single antenna. The channel gain vector between Alice and Bob is $\boldsymbol{h}$, while the channel vector between Alice and Willie is $\boldsymbol{g}$. Both of them are complex vectors with length m. When the effect of path-loss is omitted, the channel model is formulated by

$$y_j = \sqrt{P}\boldsymbol{h}^{\mathrm{H}}\boldsymbol{w}x_j + \xi_j, \quad j = 1, \ldots n, \tag{2}$$

$$z_j = \sqrt{P}\boldsymbol{g}^{\mathrm{H}}\boldsymbol{w}x_j + \nu_j, \quad j = 1, \ldots n. \tag{3}$$

In the above equations, P denotes the power of the transmitting signal, $x_j \in \mathcal{X}$ is the complex-valued input signal with unit power. y_j and z_j are the complex channel outputs at Bob and Willie. The scalars ξ_j and ν_j are the white Gaussian noises at Bob and Willie, which are independent and identically circularly symmetric distributed with zero mean and variances δ_b^2 and δ_w^2, respectively. In the following analysis, we denote $\rho_b = \frac{P}{\delta_b^2}$ and $\rho_w = \frac{P}{\delta_w^2}$.

2.2 Assumptions and Willie's hypothesis testing

Firstly, it is assumed that the channels considered here are under quasi-static block flat fading, so the same channel realization and beamforming vector $\boldsymbol{w}$ appear in the same codeword. Thus, information-theoretic analysis can be deployed within each block. The second assumption is that the adversary node runs an optimal binary hypothesis test to minimize its detection error probability about the presence or absence of legitimate transmission. It is also assumed that the transmitter has full knowledge of the limit of the adversary's ability, and is active about her choice of transmitting strategy. Finally, the adversary is known about the coding scheme being Gaussian random coding [26], but she does not have the specific codebook. In addition, we assume that the vector $\boldsymbol{h}$ has different values in different blocks, which are independent of each other. The same assumption is applicable for $\boldsymbol{g}$. Each of these two receivers, Bob and Willie, has full knowledge of his own channel realizations, i.e., $\boldsymbol{h}$ and $\boldsymbol{g}$. The transmitter has full knowledge of $\boldsymbol{h}$, while the knowledge of $\boldsymbol{g}$ depends on different situations. The details will be illustrated later. The aim of Alice is to optimize the transmission of the main channel and at the same time to confuse Willie successfully.

For computational convenience, the relative entropy (K-L divergence) $D(\mathbb{P}_1\|\mathbb{P}_0)$ is employed as a covert metric in this paper as in the literature. Alice is certain that once she obtains the covertness criterion $D(\mathbb{P}_1\|\mathbb{P}_0) \leqslant \delta$, it is impossible for Willie to distinguish between the two states of her.

3 The optimal signaling and problem formulation

3.1 The optimal signaling

In this subsection, we concern a basic problem — the optimal signaling for covert communication over MISO channels. After the introduction of these preliminaries, the major concerns of this paper on beamforming strategy and power selection will be presented in Section 4. The optimal signaling for covert communication has been investigated in [10,28] over AWGN channels and has been shown to be Gaussian. This problem was also discussed in [12] over MIMO channels where circular symmetric Gaussian distribution was proved to minimize the K-L divergence at the adversary. Here, for covert communication over MISO channels, we provide a detailed proof about the optimality of circular symmetric complex Gaussian signaling in the maximization of the mutual information over the MISO channel under the constraint $D(\mathbb{P}_1||\mathbb{P}_0) \leqslant \delta$.

As the channels are under quasi-static block flat fading, the assumption of memoryless channel is applicable. In fact, the study of blocklength n can be reduced to the study of each component. The channel model is represented by

$$y = \sqrt{P}\boldsymbol{h}^{\mathrm{H}}\boldsymbol{w}x + \xi, \tag{4}$$

$$z = \sqrt{P}\boldsymbol{g}^{\mathrm{H}}\boldsymbol{w}x + \nu. \tag{5}$$

Therefore, the channels defined above can be reduced to complex Gaussian channels, and we seek for a distribution of x which both maximizes the mutual information between x and y, and meanwhile minimizes K-L divergence between x and z. For the first maximization problem, we have the following fact.

Fact 1. Given $\boldsymbol{w}$, $\boldsymbol{g}$ and P, circularly symmetric complex Gaussian input distributions maximize $I(X;Y)$ over complex Gaussian channels.

The proof of the above fact can be found in [29] and is not included here. For the second minimization problem, we have the following lemma.

Lemma 1. Fix $\boldsymbol{w}$, $\boldsymbol{g}$ and P. Let the second moment of X be 1. $X \sim \mathcal{CN}(0,1)$ minimizes $D(f_1||f_0)$, where f_1 is the pdf of Z and f_0 is the pdf of ν.

Proof. From (4), we have $f_0 \sim \mathcal{CN}(0, \delta_w^2)$. Since the signal x and the noise ν are independent, and the second moment of X is 1, we have $E(|Z|^2) = \delta_w^2 + P|\boldsymbol{g}^{\mathrm{H}}\boldsymbol{w}|^2$. The K-L divergence is formulated as follows:

$$\begin{aligned}
D(f_1||f_0) = E_{f_1}\left[\log\frac{f_1}{f_0}\right] &= -\mathcal{H}(Z) + E_Z\left(\log \pi\delta_w^2 e^{\frac{|Z|^2}{\delta_w^2}}\right)\\
&= -\mathcal{H}(Z) + E_Z(\log \pi\delta_w^2) + E_Z\left(\frac{|Z|^2}{\delta_w^2}\log e\right)\\
&= -\mathcal{H}(Z) + \log \pi\delta_w^2 + \frac{P|\boldsymbol{g}^{\mathrm{H}}\boldsymbol{w}|^2 + \delta_w^2}{\delta_w^2}\log e\\
&\geqslant -\log\left(\pi e\left(P\left|\boldsymbol{g}^{\mathrm{H}}\boldsymbol{w}\right|^2 + \delta_w^2\right)\right) + \log \pi\delta_w^2 + \frac{P|\boldsymbol{g}^{\mathrm{H}}\boldsymbol{w}|^2 + \delta_w^2}{\delta_w^2}\log e\\
&= \rho_w\left|\boldsymbol{g}^{\mathrm{H}}\boldsymbol{w}\right|^2\log e - \log\left(\rho_w\left|\boldsymbol{g}^{\mathrm{H}}\boldsymbol{w}\right|^2 + 1\right),
\end{aligned} \tag{6}$$

where e is the Euler number. Note that the third equality holds because the noise and the signal are independent and both of them follow zero mean circular complex Gaussian distribution. The inequality follows from the fact that the circularly symmetric complex Gaussian distribution maximizes differential entropy among distributions of the same covariance matrix [29].

3.2 The bounds on the throughput over MISO channels

In this subsection, we review some thorough analysis over AWGN channels carried out by us in [25, 26][1] on the coding scheme with maximal power constraint where we have obtained both achievability and converse bounds on the throughput when Gaussian codewords are adopted. With given transmitting power and beamforming strategy, these results are further applied in MISO channels. In this way, the throughput of MISO systems can be sufficiently estimated once the power and beamforming vector are determined. Since the throughput is discussed in the finite blocklength regime (the asymptotic capacity is zero), the testification of the optimality of these beamforming strategies in numerical results depends heavily on the results in this part.

3.2.1 *Codebook generation and bounds over AWGN channels*

The generation of our codebook is described as follows. Firstly, a set of candidates are generated. Each coordinate of these candidates is drawn from i.i.d normal distribution of variance $\mu P(n)$. Secondly, each codeword is randomly chosen from a subset of these candidates: $\mu^2 nP(n) \leqslant \|\boldsymbol{x}\|^2 \leqslant nP(n)$. It is obvious that each element of the subset satisfies the maximal power constraint $P(n)$. The resulting codebook can be regarded as Gaussian generated if μ is properly chosen with moderate blocklength. Especially, we have the following achievability and converse bounds for the attainable throughput.

- When n is sufficiently large,

$$\log M_m^*(n, \epsilon, P(n)) \geqslant nC_\mu(n) - \sqrt{nV_\mu(n)}Q^{-1}(\epsilon) + \frac{1}{2}\log n + O(1) \tag{7}$$

with $C_\mu(n) = \frac{1}{2}\log(1 + \mu P(n))$, $V_\mu(n) = V(n) \cdot (\frac{2\mu P+P^2}{2P+P^2})$ where $V(n) = (\frac{P(n)\log e}{2(1+P(n))})^2(\frac{4}{P(n)} + 2)$ denotes channel dispersion with power $P(n)$.

-

$$\log M_m^*(n, \epsilon, P(n)) \leqslant nC(n) - \sqrt{nV(n)}Q^{-1}(\epsilon) + \frac{1}{2}\log n + O(1). \tag{8}$$

The power $P(n)$ can be selected under the covert constraint.

3.2.2 *Random Gaussian coding over MISO channels and bounds on throughput*

From previous analysis, the covertness constraint can be quantized as $nD(f_1\|f_0) \leqslant \delta$ with finite blocklength n over MISO systems. Under MISO channels, the above inequalities (7) and (8) can be applicable to bound the maximal throughput by n channel uses under covert constraints $D(f_1||f_0) \leqslant \delta$ with finite blocklength n and decoding error probability ϵ. We shall explain it as follows. First, as we have proved, circular symmetric complex Gaussian signaling is optimal for covert communication over MISO systems. Second, since a circular symmetric Gaussian random variable has i.i.d zero-mean real and imaginary components (Appendix 1.3 in [31]), the real and imaginary parts of both the signal and the noise can be regarded independently and the complex channel is divided into two AWGN channels with the same signal-to-noise ratio (SNR). Therefore, our above random Gaussian coding scheme is applicable for both the real and imaginary parts of the complex signals. In this way, the only concern is SNR at Willie $\frac{P|\boldsymbol{g}^{\mathrm{H}}\boldsymbol{w}|^2}{\delta_w^2}$ or the power of the received signal. If we denote $P^R(n,\delta) = \frac{P|\boldsymbol{g}^{\mathrm{H}}\boldsymbol{w}|^2}{\delta_w^2}$, we just need to maximize $P^R(n,\delta)$ under the covert constraint $nD(f_1\|f_0) \leqslant \delta$ to maximize the throughput. In the follow-on analysis, we shall investigate the optimal transmission strategy of beamforming that maximizes $P^R(n,\delta) = \rho_b|\boldsymbol{h}^{\mathrm{H}}\boldsymbol{w}|^2$ with given n and δ.

3.3 Transmission optimization

Under the assumption that each component of the signal adopts Gaussian random coding, what Alice can do is to choose an optimal transmission strategy under which the SNR of the main channel is maximal and the K-L divergence at the adversary is under a given threshold at the same time. From the analysis of Subsection 3.2, the optimization problem is as follows:

$$\max_{\boldsymbol{w}\in\mathbb{C}^m, \|\boldsymbol{w}\|=1} \rho_b|\boldsymbol{h}^{\mathrm{H}}\boldsymbol{w}|^2 \quad \text{s.t.} \quad n \cdot D(f_1||f_0) \leqslant \delta, \tag{9}$$

[1] The interested reader can refer to [30] as a full version including these two studies.

where $\rho_b = \frac{P}{\delta_b^2}$ is the ratio of the power of the sending signal and the noise at the legitimate receiver. Note that there are n channel uses in the constraint inequality because a quantitative covertness criterion should be defined with finite blocklength. In practice, the transmitter has her choice to determine the vector $\boldsymbol{w}$ and P to decrease the quantity $D(f_1||f_0)$. However, it will also affect the throughput for the legitimate receiver at the same time. Thus, the power P and the direction of $\boldsymbol{w}$ should be considered together in the transmission strategy.

4 Optimal transmission for MISO channels

In this section, an optimal transmission including several beamforming & power selection strategies for covert communication over MISO channels is characterized. With respect to different channel feedback models including full CSI model and partial CSI model, different covert metrics are adopted. As we shall see, the covert metrics in different channel state models put forward different power requirements and corresponding beamforming strategies. When there is full channel information for both the legitimate receiver and the adversary, the beamforming is characterized in Theorem 1. When there is only partial information of the adversary, the main conclusion can be found in Theorems 2 and 3.

4.1 Full CSI for Bob and Willie

The assumption that the full CSI of both Bob and Willie is available at the transmitter is widely used for the calculation of secrecy capacity over wiretap channels. In the scenario of covert communication, it is also assumed that the channel states between the transmitter and the two receivers are perfectly attainable for the former, which is reasonable in the situation where the adversary is part of the communication system.

As mentioned before, the major difference between the MISO channel and single input single output (SISO) channel is the impact of beamforming. Consequently, the transmitter has an alternative choice other than reducing the transmit power directly to decrease the signal power at the adversary. This will decrease or even remove the effect of SRL according to the channel information of the adversary at the transmitter. As a result, a positive rate can be obtained by zero-forcing beamforming. Furthermore, if some allowable covert leakage is sacrificed, an even larger rate is achievable by optimal beamforming.

We first define three beamforming vectors as follows:

$$\boldsymbol{w}_{\mathrm{MRT}} = \frac{\boldsymbol{h}}{\|\boldsymbol{h}\|}, \quad \boldsymbol{w}_{\mathrm{ZF}} = \frac{\Pi_{\boldsymbol{g}}^{\perp}\boldsymbol{h}}{\|\Pi_{\boldsymbol{g}}^{\perp}\boldsymbol{h}\|}, \quad \boldsymbol{w}_{\mathrm{ZF}}^{\perp} = \frac{\Pi_{\boldsymbol{g}}\boldsymbol{h}}{\|\Pi_{\boldsymbol{g}}\boldsymbol{h}\|}. \tag{10}$$

The first vector is the maximal ratio transmission (MRT) beamforming vector. The second vector is the zero-forcing (ZF) beamforming vector. The third one lies in the same two-dimensional subspace of $\boldsymbol{w}_{\mathrm{MRT}}$ and $\boldsymbol{w}_{\mathrm{ZF}}$, but it is orthogonal to $\boldsymbol{w}_{\mathrm{ZF}}$. As any vector $\boldsymbol{w}$ that satisfies $\boldsymbol{w} \perp \boldsymbol{g}$ will null out the signal at the adversary, the transmission power per channel use can be arbitrarily large. In the following analysis, we assume that the average transmitting power is a constant P and the attention is on the beamforming. With the above assumptions, the problem is formulated in (9). Note that from (6), the covertness constraint inequality could be formulated as

$$\rho_w \left|\boldsymbol{g}^{\mathrm{H}}\boldsymbol{w}\right|^2 \log e - \log\left(\rho_w \left|\boldsymbol{g}^{\mathrm{H}}\boldsymbol{w}\right|^2 + 1\right) \leqslant \delta/n. \tag{11}$$

Optimal beamforming. Letting x denote $\rho_w|\boldsymbol{g}^{\mathrm{H}}\boldsymbol{w}|^2$, the above inequality is reformulated as

$$x - \ln(1+x) \leqslant \delta \ln 2/n. \tag{12}$$

The derivative of the left part is $1 - \frac{1}{1+x}$, which means the function $x - \ln(1+x)$ is monotonically increasing when $x > 0$. Hence Eq. (12) has solution $x \leqslant \tilde{x}(\delta)$ with $\tilde{x}(\delta) > 0$ being the solution of the above inequality with equality $x - \ln(1+x) = \delta \ln 2/n$. Therefore, the covert constraint can be rewritten as

$$\left|\boldsymbol{g}^{\mathrm{H}}\boldsymbol{w}\right| \leqslant \sqrt{\frac{\tilde{x}(\delta)}{\rho_w}}. \tag{13}$$

The following theorem characterizes the optimal beamforming vector.

Theorem 1. The optimal solution $\boldsymbol{w}$ for the problem (9) is given by

$$\boldsymbol{w}(\gamma) = \sqrt{1-\gamma}\boldsymbol{w}_{\rm ZF} + \sqrt{\gamma}\boldsymbol{w}_{\rm ZF}^{\perp} \tag{14}$$

with some $\gamma \in (0, 1)$.

Proof. Expand the vectors $\boldsymbol{w}_{\rm ZF}$ and $\boldsymbol{w}_{\rm ZF}^{\perp}$ in their orthogonal complement space to get an orthonormal basis $\boldsymbol{w}_{\rm ZF}, \boldsymbol{w}_{\rm ZF}^{\perp}, \boldsymbol{w}_3, \ldots, \boldsymbol{w}_m$ for $\mathbb{C}^m$. A beamforming vector $\boldsymbol{v}$ is expressed as

$$\boldsymbol{v} = \gamma_1\boldsymbol{w}_{\rm ZF} + \gamma_2\boldsymbol{w}_{\rm ZF}^{\perp} + \sum_{i=3}^{m}\gamma_i\boldsymbol{w}_i \tag{15}$$

with $\sum_{i=1}^{m}|\gamma_i|^2 = 1, \gamma_i \in \mathbb{C}$. The power allocated to the directions other than $\boldsymbol{w}_{\rm ZF}$ and $\boldsymbol{w}_{\rm ZF}^{\perp}$ has no impact on the quantities $|\boldsymbol{h}^{\rm H}\boldsymbol{v}|^2$ and $|\boldsymbol{g}^{\rm H}\boldsymbol{v}|^2$ because they are orthogonal to the subspace spanned by $\boldsymbol{h}$ and $\boldsymbol{g}$, so $\gamma_i = 0$ for $i = 3, \ldots, m$. From $\sum_{i=1}^{m}|\gamma_i|^2 = 1$, the vector $\boldsymbol{w}$ is expressed as $\boldsymbol{w} = \sqrt{1-\gamma}{\rm e}^{{\rm i}\phi_1}\boldsymbol{w}_{\rm ZF} + \sqrt{\gamma}{\rm e}^{{\rm i}\phi_2}\boldsymbol{w}_{\rm ZF}^{\perp}$ with $\gamma \in (0, 1)$ and $\phi_1, \phi_2 \in [0, 2\pi]$. The values of ϕ_1, ϕ_2 need to be solved. Since these vectors are in a complex vector space, the values of ϕ_1, ϕ_2 have no impact on the value of $|\boldsymbol{h}^{\rm H}\boldsymbol{w}|$ and $|\boldsymbol{g}^{\rm H}\boldsymbol{w}|$. Consequently, $\phi_1 = 0,\ \phi_2 = 0$ are assigned and the conclusion is obtained.

From the above conclusion, the optimal beamforming vector lies in the same two-dimension subspace of $\boldsymbol{h}$ and $\boldsymbol{g}$, and could be expressed as a linear combination of any two of the three beamforming vectors. Now consider the formula for the beamforming vector $\boldsymbol{w}$. There are two cases, based on the relationship between the vectors $\boldsymbol{h}$ and $\boldsymbol{g}$. Denote $|\boldsymbol{h}^{\rm H}\boldsymbol{g}| = \|\boldsymbol{h}\|\|\boldsymbol{g}\|\cos\theta, 0 \leqslant \theta \leqslant \pi/2$.

(1) When $\theta = \pi/2$, the optimal beamforming vector should be $\boldsymbol{w} = \frac{\boldsymbol{h}}{|\boldsymbol{h}|}$. That is, MRT beamforming is adopted. Meanwhile, we have $\boldsymbol{g}^{\rm H}\boldsymbol{w} = 0$ and $\boldsymbol{w}_{\rm MRT} = \boldsymbol{w}_{\rm ZF}$. The K-L divergence requirement is naturally satisfied.

(2) When $\theta \neq \pi/2$, from (13) and (14),

$$|\boldsymbol{g}^{\rm H}\boldsymbol{w}(\gamma)| = \sqrt{\gamma}\cdot\boldsymbol{g}^{\rm H}\boldsymbol{w}_{\rm ZF}^{\perp} \leqslant \sqrt{\frac{\tilde{x}(\delta)}{\rho_w}} \overset{\rm (a)}{\Longleftrightarrow} \sqrt{\gamma}\cdot\|\boldsymbol{g}\| \leqslant \sqrt{\frac{\tilde{x}(\delta)}{\rho_w}} \Longleftrightarrow \gamma \leqslant \gamma_0 = \frac{\tilde{x}(\delta)}{\rho_w\|\boldsymbol{g}\|^2}, \tag{16}$$

where (a) follows from $\boldsymbol{w}_{\rm ZF}^{\perp} = \frac{\boldsymbol{g}}{\|\boldsymbol{g}\|}$. The above γ will make the K-L divergence of the adversary satisfy the requirement. The objective function is to maximize $\log(1 + \rho_b|\boldsymbol{h}^{\rm H}\boldsymbol{w}|^2)$, which is equivalent of maximizing $|\boldsymbol{h}^{\rm H}\boldsymbol{w}|^2$. Hence, the optimization problem is reformulated as

$$\max_{\boldsymbol{w}\in\mathbb{C}^m, \|\boldsymbol{w}\|=1} |\boldsymbol{h}^{\rm H}\boldsymbol{w}|^2 \quad \text{s.t.} \quad \boldsymbol{w} = \sqrt{1-\gamma}\boldsymbol{w}_{\rm ZF} + \sqrt{\gamma}\boldsymbol{w}_{\rm ZF}^{\perp},\ \gamma \leqslant \gamma_0. \tag{17}$$

There is the following conclusion from Appendix D of [6]:

(a)

$$|\boldsymbol{h}^{\rm H}\boldsymbol{w}(\gamma)|^2 = \gamma\|\Pi_{\boldsymbol{g}}\boldsymbol{h}\|^2 + (1-\gamma)\|\Pi_{\boldsymbol{g}}^{\perp}\boldsymbol{h}\|^2 + 2\sqrt{\gamma}\sqrt{1-\gamma}\|\Pi_{\boldsymbol{g}\boldsymbol{h}}\|\|\Pi_{\boldsymbol{g}}^{\perp}\boldsymbol{h}\|; \tag{18}$$

(b)

$$\frac{\partial^2}{\partial^2\gamma}\|\boldsymbol{h}^{\rm H}\boldsymbol{w}\|^2 = -\frac{1}{2}\frac{\|\Pi_{\boldsymbol{g}}\boldsymbol{h}\|\|\Pi_{\boldsymbol{g}}^{\perp}\boldsymbol{h}\|}{\sqrt{\gamma^3}\sqrt{(1-\gamma)^3}}; \tag{19}$$

(c) $\frac{\partial}{\partial\gamma}\|\boldsymbol{h}^{\rm H}\boldsymbol{w}(\gamma)\|^2 = 0$ has solution $\gamma^* = \frac{\|\Pi_{\boldsymbol{g}}\boldsymbol{h}\|^2}{\|\boldsymbol{h}\|^2}$, which corresponds to the MRT beamforming $\boldsymbol{w}(\gamma^*) = \boldsymbol{w}_{\rm MRT}$.

From the above facts, $|\boldsymbol{h}^{\rm H}\boldsymbol{w}|^2$ is a strictly concave function in γ and obtains its maximum at $\gamma^* = \frac{\|\Pi_{\boldsymbol{g}}\boldsymbol{h}\|^2}{\|\boldsymbol{h}\|^2}$. Consequently, the value of γ could be determined by comparing γ^* and γ_0.

- If $\gamma^* > \gamma_0$, $\gamma_0 = \frac{\tilde{x}(\delta)}{\rho_w\|\boldsymbol{g}\|^2}$ is the optimal solution for the above problem, so $\boldsymbol{w} = \boldsymbol{w}(\gamma_0)$.
- If $\gamma^* \leqslant \gamma_0$, γ^* is the optimal solution. The optimal beamforming vector is expressed as $\boldsymbol{w}(\gamma^*) = \boldsymbol{w}_{\rm MRT}$.

Zero-forcing beamforming. If we sacrifice some transmission efficiency by letting $\gamma = 0$, then we get $\boldsymbol{w}_{\rm ZF}$. The power of the received signal with zero-forcing beamforming is expressed as follows:

$$P_{\rm ZF} = \rho_b|\boldsymbol{h}^{\rm H}\boldsymbol{w}_{\rm ZF}|^2. \tag{20}$$

Note that in above equation, $\rho_b = \frac{P}{\delta_b^2}$ can be arbitrarily large since the covertness constraint is satisfied by $|\boldsymbol{g}^{\mathrm{H}}\boldsymbol{w}| = 0$ in (11).

From the above analysis, it is obvious that the covertness criterion $\delta = 0$ in (11) will imply the zero-forcing beamforming is the only solution. However, the optimal beamforming maximizes the received SNR at the legitimate receiver under a positive covertness criterion δ by allocating more power to the legitimate receiver's antenna with better gain. Alice is able to extract larger transmission efficiency from the covertness threshold δ by the optimal beamforming than zero-forcing beamforming when there is perfect channel feedback of the receiver and the adversary. Hence, zero-forcing beamforming is a suboptimal strategy. The transmission efficiency of optimal beamforming will also not conform with SRL since the power can be arbitrarily large.

4.2 The mean estimate of $\boldsymbol{g}$ for Willie

In this subsection, we consider the situation that perfect CSI is available for the main channel, but only partial CSI of the adversary's channel, i.e., the mean estimate of the channel vector, is available at the transmitter. This is the most common situation in practice. In addition, an initial value of the sending signal's average power is given as P and it can be adjusted if necessary. In this case, the transmitter can cancel out only a part of the transmit power by choosing a beamforming vector orthogonal to the mean estimate of $\boldsymbol{g}$. Therefore, SRL will still be effective with an upper bound of the transmit power, which is similar as the conclusion in [24].

The mean feedback model corresponds to the situation where the transmitter has a rough estimate of the adversary's channel. In this study, we adopt the mean feedback model in [6, 8], which is written as

$$\boldsymbol{g} = \sqrt{1-\alpha}\hat{\boldsymbol{g}} + \sqrt{\alpha}\tilde{\boldsymbol{g}}, \tag{21}$$

where $\hat{\boldsymbol{g}}$ is the estimate of the channel vector $\boldsymbol{g}$, $\tilde{\boldsymbol{g}}$ represents the estimation error and $\alpha \in (0,1)$ indicates the accuracy of the estimation. It is assumed that $\tilde{\boldsymbol{g}}$ follows from $\mathcal{CN}(0, \delta_{\boldsymbol{g}}^2\boldsymbol{I})$ and $\hat{\boldsymbol{g}}$ and $\tilde{\boldsymbol{g}}$ are independent of each other. The transmitter is aware of the values of the vectors $\hat{\boldsymbol{g}}$, $\boldsymbol{h}$, the scale α and the estimation error variance $\delta_{\boldsymbol{g}}^2$. In this situation, the received signal of the adversary is formulated as follows if the beamforming vector $\boldsymbol{w}$ is determined:

$$z = \sqrt{P}\boldsymbol{g}^{\mathrm{H}}\boldsymbol{w}x + \nu = \sqrt{(1-\alpha)\cdot P}\hat{\boldsymbol{g}}^{\mathrm{H}}\boldsymbol{w}x + \sqrt{\alpha\cdot P}\tilde{\boldsymbol{g}}^{\mathrm{H}}\boldsymbol{w}x + \nu. \tag{22}$$

Now we analyze the terms of the above formula. Since $x \sim \mathcal{CN}(0,1)$, the first term of the right hand side is from $\mathcal{CN}(0, (1-\alpha)P|\hat{\boldsymbol{g}}^{\mathrm{H}}\boldsymbol{w}|^2)$. For the second term, since $\tilde{\boldsymbol{g}}$ is a zero-mean complex circular symmetric Gaussian vector, the inner product between $\tilde{\boldsymbol{g}}$ and a constant vector $\boldsymbol{w}$ with $\|\boldsymbol{w}\| = 1$ is also complex circular symmetric Gaussian variable. The second term is thus the product of two zero-mean complex circular symmetric Gaussian random variables. The third term represents the noise. As mentioned above, we denote the pdfs of ν and z as f_0 and f_1, respectively.

Now let us consider the covertness criterion, i.e., the K-L divergence between the distribution of Willie's received signal z and the noise ν. From the definition of the K-L divergence, it seems that we should calculate it between the distribution of the variables z and ν. However, it is complicated and involves the distribution of the product of two complex Gaussian random variables. In fact, the channel is assumed to be under quasi-static block flat fading, so the channel vectors $\boldsymbol{h}$ and $\boldsymbol{g}$ can be viewed as constant vectors. In other words, the difference between the distributions of Willie's received signal and the noise is independent of the estimation error. Therefore, on one hand, the K-L divergence is irrelative with the estimation error from the adversary's perspective; on the other hand, from the transmitter's viewpoint, she is aware of the existence of the estimation error, and what she should do is to make the probability of a large K-L divergence at the adversary small enough. Hence the covertness criterion is expressed as

$$P\{nD(f_1||f_0) > \delta\} \leqslant \kappa, \tag{23}$$

where there is randomness in pdf f_1 and κ is a small constant satisfying $0 \leqslant \kappa < 1$. Therefore, the security criterion is that the probability that the K-L divergence is greater than a lower bound should be kept small. The optimization problem is formulated as follows:

$$\max_{\boldsymbol{w}\in\mathbb{C}^m, \|\boldsymbol{w}\|=1} \rho_b|\boldsymbol{h}^{\mathrm{H}}\boldsymbol{w}|^2, \quad \text{s.t.} \quad \rho_b = \frac{P}{\delta_b^2}, \quad P\{nD(f_1||f_0) > \delta\} \leqslant \kappa. \tag{24}$$

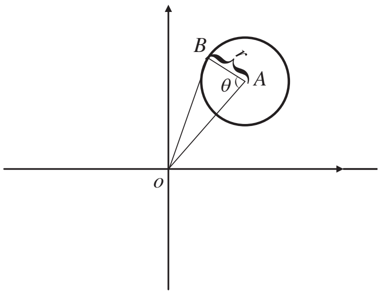


Figure 2 The model to compute the random variable $|\boldsymbol{g}^{\mathrm{H}}\boldsymbol{w}|^2$.

The probability $P\{nD(f_1||f_0) > \delta\}$ is denoted as the outage probability on covertness. It is obvious that smaller δ and κ will lead to more covertness. From the analysis about (11) and (12), the inequality in (23) can be reformulated as

$$P\left\{\left|\boldsymbol{g}^{\mathrm{H}}\boldsymbol{w}\right|^2 > \frac{\tilde{x}(\delta)}{\rho_w}\right\} \leqslant \kappa, \tag{25}$$

where $x(\delta)$ is the solution of (12) with equality.

Now assume the beamforming vector $\boldsymbol{w}$ is determined, we focus on the random variable $|\boldsymbol{g}^{\mathrm{H}}\boldsymbol{w}|$. From

$$\boldsymbol{g}^{\mathrm{H}}\boldsymbol{w} = \sqrt{1-\alpha}\hat{\boldsymbol{g}}^{\mathrm{H}}\boldsymbol{w} + \sqrt{\alpha}\tilde{\boldsymbol{g}}^{\mathrm{H}}\boldsymbol{w}, \tag{26}$$

we plot the relationship of these variables in Figure 2. Denote $\overrightarrow{\mathrm{OA}}$ as the complex constant $\sqrt{1-\alpha}\hat{\boldsymbol{g}}^{\mathrm{H}}\boldsymbol{w}$ and its magnitude as c. From Lemma A1 in Appendix A, the second term $\sqrt{\alpha}\tilde{\boldsymbol{g}}^{\mathrm{H}}\boldsymbol{w}$ in (26) is a circular symmetric complex Gaussian random variable, which is denoted as the vector $\overrightarrow{\mathrm{AB}}$ in the complex plane. The angle θ is uniformly distributed in $(0, 2\pi]$ from the property of circular symmetry. Then, $\overrightarrow{\mathrm{OB}}$ is the random variable $\boldsymbol{g}^{\mathrm{H}}\boldsymbol{w}$ and the equality $|\mathrm{OB}|^2 = |\mathrm{OA}|^2 + |\mathrm{AB}|^2 - 2\,|\mathrm{OA}|\,|\mathrm{AB}|\cos\theta$ is valid by the law of cosines. Let $|\mathrm{AB}| = r$, which is a Rayleigh distributed random variable. We have

$$P\{|\mathrm{OB}|^2 < y\} = P\{c^2 + r^2 - 2cr\cos\theta < y\} = P\left\{\cos\theta > \frac{1}{2c}\left(r + \frac{c^2 - y}{r}\right)\right\}. \tag{27}$$

Note that the parameter r has pdf

$$f(r) = \begin{cases} \dfrac{r}{\alpha\delta_{\boldsymbol{g}}^2}\mathrm{e}^{\frac{-r^2}{2\alpha\delta_{\boldsymbol{g}}^2}}, & r \geqslant 0, \\ 0, & \text{otherwise.} \end{cases} \tag{28}$$

Hence the probability in (27) can be rewritten as

$$P\left\{\cos\theta > \frac{1}{2c}\left(r + \frac{c^2 - y}{r}\right)\right\} \overset{\text{(a)}}{=} E_r\left[\frac{\arccos\frac{1}{2c}(r + \frac{c^2-y}{r})}{\pi}\right] = \int_0^{\infty} \frac{\arccos\frac{1}{2c}(r + \frac{c^2-y}{r})}{\pi}\frac{r}{\alpha\delta_{\boldsymbol{g}}^2}\mathrm{e}^{\frac{-r^2}{2\alpha\delta_{\boldsymbol{g}}^2}}\mathrm{d}r, \tag{29}$$

where equation (a) is from the fact that θ is uniformly distributed in $(0\ 2\pi]$. From the security criterion in (25), a proper beamforming vector $\boldsymbol{w}$ should satisfy the following inequality:

$$\int_0^{\infty} \frac{\arccos\frac{1}{2c}(r + \frac{c^2 - \frac{\tilde{x}(\delta)}{\rho_w}}{r})}{\pi}\frac{r}{\alpha\delta_{\boldsymbol{g}}^2}\mathrm{e}^{\frac{-r^2}{2\alpha\delta_{\boldsymbol{g}}^2}}\mathrm{d}r \geqslant 1 - \kappa. \tag{30}$$

- The situation $\kappa = 0$ should be paid attention, which corresponds to that the variable $\left|\boldsymbol{g}^{\mathrm{H}}\boldsymbol{w}\right|$ is definitely less than $\frac{\tilde{x}(\delta)}{\rho_w}$. We further have that $\alpha = 0$ and $|\hat{\boldsymbol{g}}^{\mathrm{H}}\boldsymbol{w}| < \frac{\tilde{x}(\delta)}{\rho_w}$, i.e., Alice has full CSI about Willie's channel and the power is low enough. In this case, the attainable throughput will conform with SRL.

- In the covertness criterion pair (δ, κ), $\kappa < 1$ is necesary since the inequality (25) with $\kappa = 1$ is naturally satisfied.

Now we focus on beamforming and define the following vectors:

$$\hat{\boldsymbol{w}}_{\mathrm{ZF}} = \frac{\Pi_{\hat{\boldsymbol{g}}}^{\perp}\boldsymbol{h}}{\|\Pi_{\hat{\boldsymbol{g}}}^{\perp}\boldsymbol{h}\|}, \quad \hat{\boldsymbol{w}}_{\mathrm{ZF}}^{\perp} = \frac{\Pi_{\hat{\boldsymbol{g}}}\boldsymbol{h}}{\|\Pi_{\hat{\boldsymbol{g}}}\boldsymbol{h}\|}. \tag{31}$$

Note that the above vectors are similar to (10) except that $\boldsymbol{g}$ is replaced by $\hat{\boldsymbol{g}}$. Consider zero-forcing beamforming with arbitrary power P in this case. As Alice only knows an estimation of the adversary's channel vector, the zero-forcing vector $\hat{\boldsymbol{w}}_{\mathrm{ZF}}$ will eliminate the estimation part of $\hat{\boldsymbol{g}}$ but not the error part $\tilde{\boldsymbol{g}}$. Therefore, it is very probable that the latter part will lead to the excess of the K-L divergence to the covertness criterion.

Theorem 2. In covert communication with covert criterion (δ, κ) over MISO channels, if the transmitter has a mean estimation of the adversary's channel and full CSI of the legitimate receiver's channel, then there is some threshold P_0 depending on κ and δ such that if the average power P of the signal is larger than P_0, the transmitter cannot obtain effective beamforming based on the known channel information to satisfy the covertness criterion.

Proof. The transmitter has the information of α, $\hat{\boldsymbol{g}}$ and $\delta_{\boldsymbol{g}}$. From Alice's perspective, the direction of $\boldsymbol{w}$ has no effect on $\tilde{\boldsymbol{g}}^{\mathrm{H}}\boldsymbol{w}$ because of the circular symmetry of $\tilde{\boldsymbol{g}}$, so the direction of the beamforming vector $\boldsymbol{w}$ only affects the constant term $\hat{\boldsymbol{g}}$. To decrease the K-L divergence at the adversary, the best strategy for Alice is adopting zero-forcing beamforming $\hat{\boldsymbol{w}}_{\mathrm{ZF}}$ which satisfies $\hat{\boldsymbol{g}}^{\mathrm{H}}\hat{\boldsymbol{w}}_{\mathrm{ZF}} = 0$, which will eliminate the effect of the known channel vector $\hat{\boldsymbol{g}}$. Now we have $\boldsymbol{g}^{\mathrm{H}}\hat{\boldsymbol{w}}_{\mathrm{ZF}} = \sqrt{\alpha}\tilde{\boldsymbol{g}}^{\mathrm{H}}\hat{\boldsymbol{w}}_{\mathrm{ZF}}$. As a function of the random variable $|\tilde{\boldsymbol{g}}^{\mathrm{H}}\hat{\boldsymbol{w}}_{\mathrm{ZF}}|^2$, the K-L divergence of one-shot at the adversary is $D(f_1||f_0) = \rho_w\alpha|\tilde{\boldsymbol{g}}^{\mathrm{H}}\hat{\boldsymbol{w}}_{\mathrm{ZF}}|^2\log e - \log(\rho_w\alpha|\tilde{\boldsymbol{g}}^{\mathrm{H}}\hat{\boldsymbol{w}}_{\mathrm{ZF}}|^2 + 1)$, which is a random variable from Alice's perspective. The covertness criterion $P\{nD(f_1||f_0) > \delta\} < \kappa$ is equivalent to the inequality $P\{|\tilde{\boldsymbol{g}}^{\mathrm{H}}\hat{\boldsymbol{w}}_{\mathrm{ZF}}|^2 > \frac{\tilde{x}(\delta)}{\alpha\rho_w}\} < \kappa$. From Corollary A1 in Appendix A, the random variable $|\tilde{\boldsymbol{g}}^{\mathrm{H}}\hat{\boldsymbol{w}}_{\mathrm{ZF}}|^2$ is exponentially distributed, so there is a $\tau_0(\kappa)$ which satisfies

$$P\left\{\left|\tilde{\boldsymbol{g}}^{\mathrm{H}}\hat{\boldsymbol{w}}_{\mathrm{ZF}}\right|^2 > \tau_0(\kappa)\right\} = \kappa. \tag{32}$$

If the average power P satisfies $\frac{\tilde{x}(\delta)}{\alpha\rho_w} < \tau_0(\kappa)$, then the covertness criterion is not satisfied by zero-forcing beamforming, and it will not be satisfied by any other beamforming strategies since the zero-forcing is the best strategy Alice could do to merely decrease the K-L divergence at the adversary based on the known information. From (32), if the transmitter wants to satisfy the covertness criterion, he should limit the average power of the signal as follows:

$$\frac{\tilde{x}(\delta)}{\alpha\rho_w} \geqslant \tau_0(\kappa)$$

which results in $P \leqslant \frac{\tilde{x}(\delta)\delta_w^2}{\alpha\tau_0(\kappa)}$. Therefore, the upper bound of the transmit power is

$$P_0 = \frac{\tilde{x}(\delta)\delta_w^2}{\alpha\tau_0(\kappa)}. \tag{33}$$

When the average power of the signal satisfies $P < P_0$, the transmitter could adjust $\boldsymbol{w}$ to ensure the security requirement and increase the attainable throughput further more. In addition, as α indicates the accuracy of the estimation vector, a larger α will lead to a smaller P_0 and hence larger communication efficiency of the main channel. Especially, if $\alpha = 0$, there will be no upper bound of the power, which is just the conclusion of Subsection 4.1.

We will discuss the problem in the following three cases.

- When $P > \frac{\tilde{x}(\delta)\delta_w^2}{\alpha\tau_0(\kappa)}$, the transmitter could not ensure that the K-L divergence of the adversary satisfies the covert requirement no matter what beamforming vector is adopted. The only solution is to decrease the power of the sending signal P until $P \leqslant \frac{\tilde{x}(\delta)\delta_w^2}{\alpha\tau_0(\kappa)}$ is satisfied. Thus, power control is necessary.
- When $P = \frac{\tilde{x}(\delta)\delta_w^2}{\alpha\tau_0(\kappa)}$, the K-L divergence requirement can be satisfied by $\boldsymbol{w}\perp\hat{\boldsymbol{g}} = 0$. The measured part $\hat{\boldsymbol{g}}$ makes no contribution to the K-L divergence, and only the uncertain part $\tilde{\boldsymbol{g}}$ of $\boldsymbol{g}$ contributes to it.

- When $P < \frac{\tilde{x}(\delta)\delta_w^2}{\alpha\tau_0(\kappa)}$, it means the power P and the covertness criterion (δ, κ) permit an optimal beamforming $\boldsymbol{w}$. The optimal beamforming vector $\boldsymbol{w}$ will be characterized in the following theorem.

Theorem 3. When there is a mean estimation of the adversary's channel $\boldsymbol{g}$, the optimal beamforming vector $\boldsymbol{w}$ with proper power $P < P_0$ and covertness criterion (δ, κ) is given by

$$\boldsymbol{w} = \sqrt{1-\hat{\gamma}}\hat{\boldsymbol{w}}_{\mathrm{ZF}} + \sqrt{\hat{\gamma}}\hat{\boldsymbol{w}}_{\mathrm{ZF}}^{\perp}, \quad \hat{\gamma} \in (0,1). \tag{34}$$

Proof. We consider the problem directly from the inequality (30), the inequality is directly related to $c = |\sqrt{1-\alpha}\hat{\boldsymbol{g}}^{\mathrm{H}}\boldsymbol{w}|$ but not $r = |\sqrt{\alpha}\tilde{\boldsymbol{g}}^{\mathrm{H}}\boldsymbol{w}|$, whose distribution is fixed. Hence, the security criterion is related to the direction of the vector $\hat{\boldsymbol{g}}$ but not the direction of $\tilde{\boldsymbol{g}}$. The remainder of the proof is similar to Theorem 1 and is not included here owing to space limitation.

From the above theorem, the optimal beamforming in the third case above is calculated as follows. First, solve the equation with the variable c:

$$\int_0^{\infty} \frac{\arccos\frac{1}{2c}(r + \frac{c^2 - \frac{\tilde{x}(\delta)}{\rho_w}}{r})}{\pi} \frac{r}{\alpha\delta_{\boldsymbol{g}}^2} \mathrm{e}^{\frac{-r^2}{2\alpha\delta_{\boldsymbol{g}}^2}} \mathrm{d}r = 1 - \kappa. \tag{35}$$

Note that the integrand is the product of a variable in $[0,1]$ and the pdf of a Rayleigh distributed random variable, a solution c_0 is obtainable from the analysis before (27) since the power P and the covertness criterion (δ, κ) are proper determined. Second, solve the equations with variable γ:

$$\begin{cases} c_0 = \sqrt{1-\alpha}\hat{\boldsymbol{g}}^{\mathrm{H}}\boldsymbol{w}, \\ \boldsymbol{w} = \sqrt{1-\hat{\gamma}}\hat{\boldsymbol{w}}_{\mathrm{ZF}} + \sqrt{\hat{\gamma}}\hat{\boldsymbol{w}}_{\mathrm{ZF}}^{\perp} \end{cases} \tag{36}$$

to get $\hat{\gamma}_0$. Thus, from the covertness criterion, $\hat{\gamma}$ should satisfy $\hat{\gamma} \leqslant \hat{\gamma}_0$. Third, consider the transmission efficiency of the main channel, and the optimization problem is rewritten as

$$\max_{\boldsymbol{w} \in \mathbb{C}^m, \|\boldsymbol{w}\|=1} |\boldsymbol{h}^{\mathrm{H}}\boldsymbol{w}|^2 \quad \text{s.t.} \quad \boldsymbol{w} = \sqrt{1-\hat{\gamma}}\hat{\boldsymbol{w}}_{\mathrm{ZF}} + \sqrt{\hat{\gamma}}\hat{\boldsymbol{w}}_{\mathrm{ZF}}^{\perp}, \quad \hat{\gamma} \leqslant \hat{\gamma}_0, \tag{37}$$

which is almost the same as (17). Note that the covertness criterion is in the form of probability, and it is possible that the covertness criterion is violated at the adversary. However, the probability could be limited by setting proper parameters κ.

4.3 Statistical information about $\boldsymbol{g}$

In this subsection, it is assumed that perfect CSI of the main channel is available. However, only statistical information, i.e., the covariance matrix $\Sigma_{\boldsymbol{g}}$ of the complex Gaussian random vector $\boldsymbol{g}$ is available at the transmitter. The covertness criterion is defined by $P\{nD(f_1||f_0) > \delta\} < \kappa$ as before.

An intuition for this situation is that the direction of the beamforming vector will have little effect on the adversary's K-L divergence because the adversary's channel vector could be arbitrary. In fact, if it is further assumed that $\Sigma_{\boldsymbol{g}} = \delta_{\boldsymbol{g}}^2\boldsymbol{I}$, the above conclusion can be checked from (26) and (27). When $\alpha = 1$, i.e., only the distribution of the channel vector $\boldsymbol{g}$ is known, the point A will be located at O in Figure 2. Therefore, the inequality (27) could be rewritten as

$$P\left\{|\mathrm{OB}|^2 < y\right\} = P\{r^2 < y\}.$$

From Corollary A1 in Appendix A, r^2 is exponentially distributed. Therefore, the covertness criterion is expressed as

$$\int_0^{\frac{\tilde{x}(\delta)}{\rho_w}} \frac{r}{\delta_{\boldsymbol{g}}^2} \mathrm{e}^{-\frac{r^2}{2\delta_{\boldsymbol{g}}^2}} \mathrm{d}r > 1 - \kappa. \tag{38}$$

From the above inequality, it is obvious that beamforming will have no effect on covertness. The only solution is to decrease the power P so that the quantity ρ_w is decreased and the upper limit of the integral is increased. Thus, the inequality (38) implies an upper bound on the transmit power. This strategy will permit a covert outage probability of κ, which is the consequence of the uncertainty about the vector $\boldsymbol{g}$. For the main channel, the MRT beamforming $\boldsymbol{w}_{\mathrm{MRT}}$ is the optimal solution to maximize the throughput of the main channel.

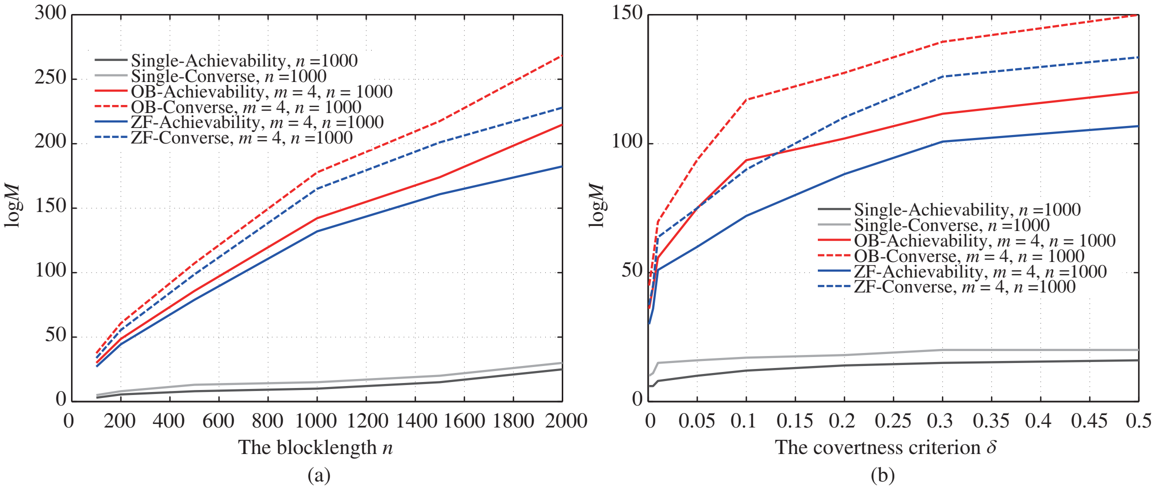


Figure 3 (Color online) Comparison of the bounds with different beamforming strategies. (a) $\delta = 0.2$; (b) $n = 1000$.

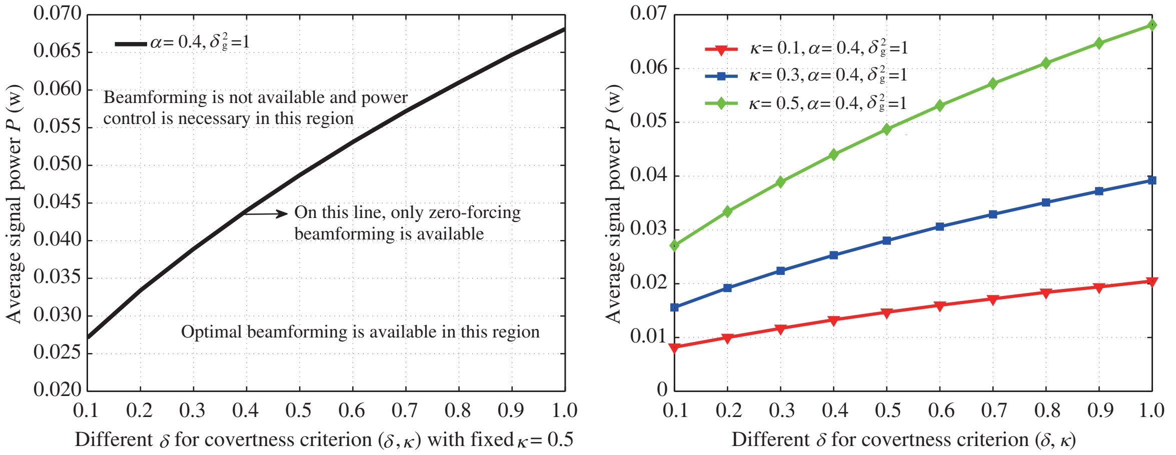
5 Numerical results

In this section, numerical results are presented to examine the effect of beamforming and illustrate the proper region for beamforming. In the following experiments, the number of antennae at the transmitter is 4 and the variances of the background noise at Bob and Willie are both 1 ($\delta_w^2 = \delta_b^2 = 1, m = 4$).

Firstly, the situation when the transmitter has full CSI about the legitimate receiver and the adversary is illustrated. We generate 5000 pairs of complex Gaussian vectors randomly to represent channel vectors $\boldsymbol{h}$ and $\boldsymbol{g}$ and calculate these beamforming vectors following the methods in Subsection 4.1. In general, the vectors $\boldsymbol{h}$ and $\boldsymbol{g}$ are randomly generated, and it is probable that $\gamma_0 \geqslant \gamma^*$, and then $\boldsymbol{w} = \boldsymbol{w}_{\mathrm{MRT}}$ is obtained. Whatever the case is, our method ensures that the power of the received signal is maximal and meanwhile the covertness criterion is not violated, which is especially better than the rate obtained by the zero-forcing strategy. The average numbers of the throughput in different situations are compared. The right picture in Figure 3 illustrates the achievability and converse bounds of the throughput $\log M(n, \epsilon, \delta)$ of the main channel versus the covertness criterion δ when the power of the signal is set to be $2P_{\mathrm{MRT}}$ with fixed blocklength n. The left hand side of Figure 3 illustrates $\log M(n, \epsilon, \delta)$ versus n with fixed $\delta = 0.2$ and $\epsilon = 0.01$. The average number of throughput obtained by optimal beamforming exceeds that of zero-forcing beamforming by more than dozens in Figure 3 when the covertness criterion δ is more than 0.2 and the blocklength is more than 1000, respectively. When there is only a single antenna at the transmitter, no beamforming strategy is available. The covertness should be granted by decreasing the signal power. In this situation, the attainable throughput follows from SRL and is far less than the situations with multiple antennae.

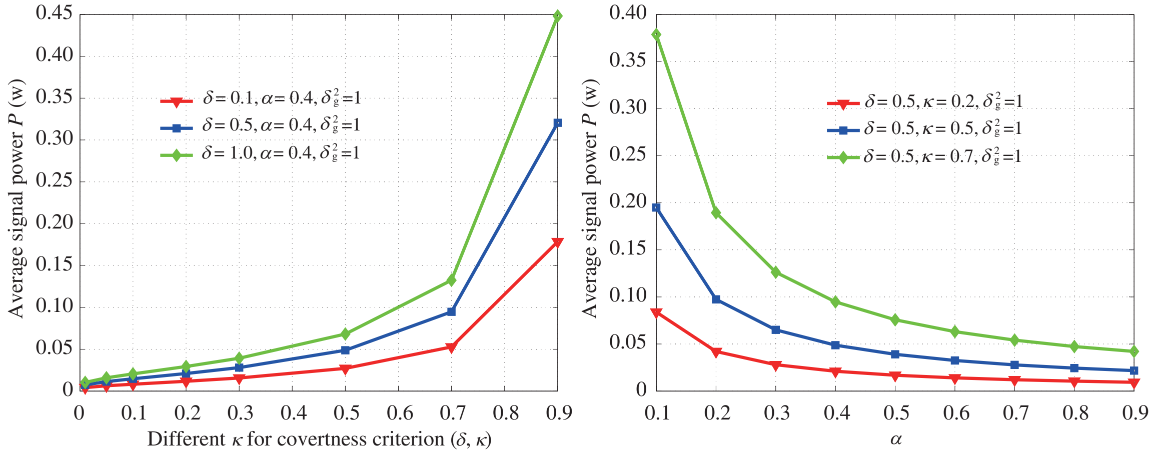
Secondly, we illustrate the availability of these beamforming strategies when there is only partial information of $\boldsymbol{g}$. The availability depends on the feedback parameters α, $\delta_{\boldsymbol{g}}$ and the covertness criterion pair (δ, κ). In our setting, the parameter $\delta_{\boldsymbol{g}}$ is always 1. To get the critical value of P_0, Eq. (12) for a given δ is solved. Then $\tau_0(\kappa)$ is computed by resolving (32). Finally, the critical P_0 is calculated by (33). The critical power P_0 for different values δ are plotted in Figure 4, in which the regions available for optimal and zero-forcing beamforming are illustrated. The line is the boundary of the region and only zero-forcing beamforming is available on the line. If the pair (δ, P) lies above the line, power control is necessary to ensure the covertness. If the pair (δ, P) is under the line, optimal beamforming is available. The regions corresponding to different values of α, δ and κ are depicted in Figure 5. From these figures, we can see that if the transmitter has less accurate information about $\boldsymbol{g}$, it is more probable that beamforming is unavailable and the power should be adjusted. Moreover, if the covertness requirement is stricter, then it is more likely that the power should be adjusted.

Finally, the achievability and converse bounds with optimal beamforming and zero-forcing beamforming are compared with a proper signal power P, the blocklength $n = 1000$ and covertness parameter pair $(\delta, \kappa) = (0.5, 0.2)$ which satisfy $P < P_0 = \frac{\tilde{x}(\delta)\delta_w^2}{\alpha\tau_0(\kappa)}$. Zero-forcing beamforming only eliminates the effect of the mean estimation of the adversary's channel vector, and we compare these two strategies in Figure 6. It is shown that the optimal beamforming obtained by Subsection 4.2 has better behavior than zero-forcing



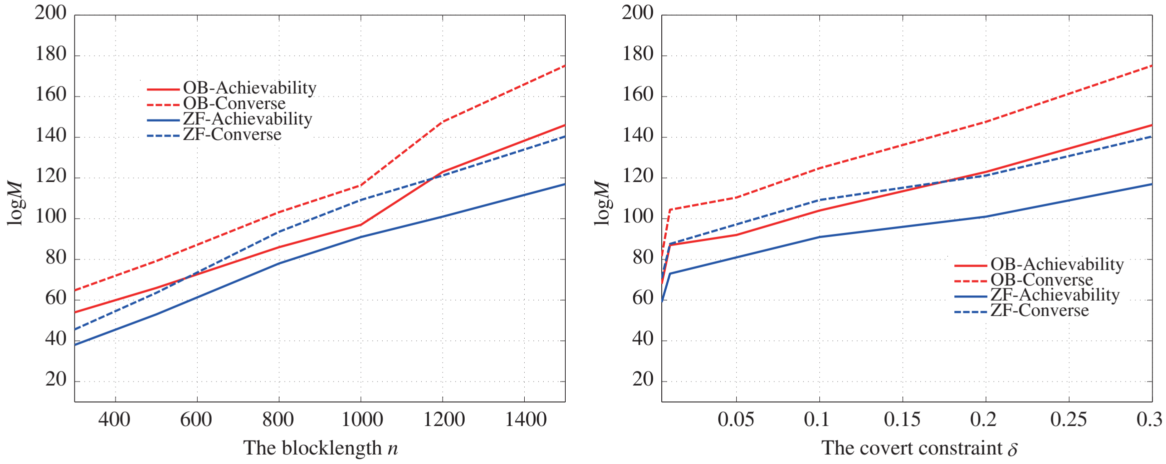
(a) (b)

Figure 4 (Color online) (a) Available region of the pair (P, δ) for beamforming. (b) Available regions for beamforming when δ varies from 0.1 to 1.



(a) (b)

Figure 5 (Color online) Available regions for beamforming when (a) κ varies from 0 to 0.9 and (b) α varies from 0.1 to 0.9.



(a) (b)

Figure 6 (Color online) Comparison of different beamforming strategies. (a) $(\delta, \kappa) = (0.5, 0.2)$; (b) $n = 1000$ and $\kappa = 0.2$.

beamforming. More accurate information will lead to more gain for zero-forcing beamforming, while the gain is less than that of optimal beamforming.

6 Conclusion

In this study, we investigated the throughput of finite blocklength with optimal beamforming which utilizes the covertness criterion and the channel feedback information to maximize the throughput and meanwhile satisfy the covert requirement. The achievability and converse bounds under optimal beamforming are shown to be better than zero-forcing beamforming over MISO channels.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant No. 61871264).

References

1 Shafiee S, Liu N, Ulukus S. Towards the secrecy capacity of the Gaussian MIMO wire-tap channel: the 2-2-1 channel. IEEE Trans Inform Theor, 2009, 55: 4033–4039

2 Khisti A, Wornell G W. Secure transmission with multiple antennas I: the MISOME wiretap channel. IEEE Trans Inform Theor, 2010, 56: 3088–3104

3 Khisti A, Wornell G W. Secure transmission with multiple antennas-part II: the MIMOME wiretap channel. IEEE Trans Inform Theor, 2010, 56: 5515–5532

4 Zhang L, Wu G, Li S Q. Capacity bounds of transmit beamforming over MISO time-varying channels with imperfect feedback. Sci China Inf Sci, 2010, 53: 1417–1430

5 Liu T, Shamai S. A note on the secrecy capacity of the multiple-antenna wiretap channel. IEEE Trans Inform Theor, 2009, 55: 2547–2553

6 Gerbracht S, Scheunert C, Jorswieck E A. Secrecy outage in MISO systems with partial channel information. IEEE Trans Inform Forensic Secur, 2012, 7: 704–716

7 Rezki Z, Khisti A, Alouini M S. On the secrecy capacity of the MISO wiretap channel under imperfect channel estimation. In: Proceedings of IEEE Global Communications Conference (GLOBECOM'2014), Austin, 2014. 1602–1607

8 Zhou X, Rezki Z, Alomair B, et al. Achievable rates of secure transmission in Gaussian MISO channel with imperfect main channel estimation. IEEE Trans Wirel Commun, 2016, 15: 4470–4485

9 Bash B A, Goeckel D, Towsley D. Limits of reliable communication with low probability of detection on AWGN channels. IEEE J Sel Areas Commun, 2013, 31: 1921–1930

10 Wang L, Wornell G W, Zheng L. Fundamental limits of communication with low probability of detection. IEEE Trans Inform Theor, 2016, 62: 3493–3503

11 Bloch M R. Covert communication over noisy channels: a resolvability perspective. IEEE Trans Inform Theor, 2016, 62: 2334–2354

12 Abdelaziz A, Koksal C E. Fundamental limits of covert communication over MIMO AWGN channel. In: Proceedings of 2017 IEEE Conference on Communications and Network Security (CNS), Las Vegas, 2017. 1–9

13 Wang K, Gong Y, Zhou G M, et al. A novel covert communication system based on symmetric α-stable distribution (in Chinese). Sci Sin Inform, 2017, 47: 374–384

14 Zhu Y, Yu M Y, Hu H X, et al. Efficient construction of provably secure steganography under ordinary covert channels (in Chinese). Sci Sin Inform, 2013, 43: 657–669

15 Lee S, Baxley R J, Weitnauer M A, et al. Achieving undetectable communication. IEEE J Sel Top Signal Process, 2015, 9: 1195–1205

16 Che P H, Bakshi M, Jaggi S. Reliable deniable communication: hiding messages in noise. In: Proceedings of IEEE International Symposium on Information Theory (ISIT2013), Istanbul, 2013. 2945–2949

17 He B, Yan S H, Zhou X Y, et al. On covert communication with noise uncertainty. IEEE Commun Lett, 2017, 21: 941–944

18 Shahzad K, Zhou X, Yan S. Covert communication in fading channels under channel uncertainty. In: Proceedings of IEEE 85th Vehicular Technology Conference (VTC Spring), Sydney, 2017. 1–5

19 Sobers T V, Bash B A, Goeckel D, et al. Covert communication with the help of an uninformed jammer achieves positive rat. In: Proceedings of the 49th Asilomar Conference on Signals, Systems and Computers, Pacific Grove, 2015. 625–629

20 Soltani R, Goeckel D, Towsley D, et al. Covert wireless communication with artificial noise generation. IEEE Trans Wirel Commun, 2018, 17: 7252–7267

21 Li J, Petropulu A P. On ergodic secrecy rate for Gaussian MISO wiretap channels. IEEE Trans Wirel Commun, 2011, 10: 1176–1187

22 Tahmasbi M, Bloch M R. Second-order asymptotics of covert communications over noisy channels. In: Proceedings of IEEE International Symposium on Information Theory (ISIT), Barcelona, 2016. 2224–2228

23 Tahmasbi M, Bloch M R. First- and second-order asymptotics in covert communication. IEEE Trans Inform Theor, 2019, 65: 2190–2212

24 Tang H, Wang J, Zheng Y R. Covert communication with extremely low power under finite block length over slow fading. In: Proceedings of IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), Honolulu, 2018. 657–661

25 Yu X, Wei S, Luo Y. One-shot achievability and converse bounds of Gaussian random coding in AWGN channels under covert constraint. In: Proceedings of the 57th Annual Allerton Conference on Communication, Control, and Computing (Allerton), Monticello, 2019

26 Yu X, Wei S, Luo Y. Finite blocklength analysis of Gaussian random coding in AWGN channels under covert constraints. IEEE Transactions on Information Forensics & Security, 2019. doi: 10.1109/TIFS.2020.3032292

27 Polyanskiy Y, Poor H V, Verdu S. Channel coding rate in the finite blocklength regime. IEEE Trans Inform Theor, 2010, 56: 2307–2359

28 Yan S H, Cong Y R, Hanly S V, et al. Gaussian signalling for covert communications. IEEE Trans Wirel Commun, 2019, 18: 3542–3553

29 Telatar I E. Capacity of multi-antenna Gaussian channels. Eur Trans Telecommun, 1999, 10: 585–595

30 Yu X, Wei S, Luo Y. Finite blocklength analysis of Gaussian random coding in AWGN channels under covert constraints. 2019. ArXiv: 1909.11324

31 Tse D, Viswanath P. Fundamentals of Wireless Communication. Cambridge: Cambridge University Press, 2005

Appendix A

Lemma A1. If $\tilde{\boldsymbol{g}}$ is subject to $\mathcal{CN}(0, \delta_{\boldsymbol{g}}^2 \boldsymbol{I})$ and $\boldsymbol{w}$ is a constant vector with the same dimension, $\tilde{\boldsymbol{g}}^{\mathrm{H}}\boldsymbol{w}$ is a zero-mean complex circularly symmetric Gaussian random variable with variance $\|\boldsymbol{w}\|^2\delta_{\boldsymbol{g}}^2$.

Proof. Since $\tilde{\boldsymbol{g}}$ follows $\mathcal{CN}(0, \delta_{\boldsymbol{g}}^2 \boldsymbol{I})$, the complex conjugate $\tilde{\boldsymbol{g}}_i^*$ $(i = 1, \ldots, n)$ are zero-mean complex circularly symmetric Gaussian with variance $\delta_{\boldsymbol{g}}^2$ and independent of each other. $\tilde{\boldsymbol{g}}^{\mathrm{H}}\boldsymbol{w} = \sum_{i=1}^{n} \tilde{\boldsymbol{g}}_i^* \boldsymbol{w}_i$ is a linear combination of these $\tilde{\boldsymbol{g}}_i^*$. The ith term in the right side of the equation is zero-mean complex Gaussian with variance $|\boldsymbol{w}_i|^2\delta_{\boldsymbol{g}}^2$. Hence, the sum of them is zero-mean complex circularly symmetric Gaussian with variance $\|\boldsymbol{w}\|^2\delta_{\boldsymbol{g}}^2$.

Corollary A1. With the same condition as Lemma A1, $|\tilde{\boldsymbol{g}}^{\mathrm{H}}\boldsymbol{w}|$ is subject to Rayleigh distribution with probability density function:

$$f(x) = \frac{x}{\|\boldsymbol{w}\|^2\delta_{\boldsymbol{g}}^2} \mathrm{e}^{-\frac{x^2}{2\|\boldsymbol{w}\|^2\delta_{\boldsymbol{g}}^2}}, x \geqslant 0.$$

$|\tilde{\boldsymbol{g}}^{\mathrm{H}}\boldsymbol{w}|^2$ is subject to chi-squared distribution with two degrees of freedom, i.e., exponential distribution with pdf

$$g(x) = \frac{1}{2\|\boldsymbol{w}\|^2\delta_{\boldsymbol{g}}^2} \mathrm{e}^{-\frac{x}{2\|\boldsymbol{w}\|^2\delta_{\boldsymbol{g}}^2}}, x \geqslant 0.$$